

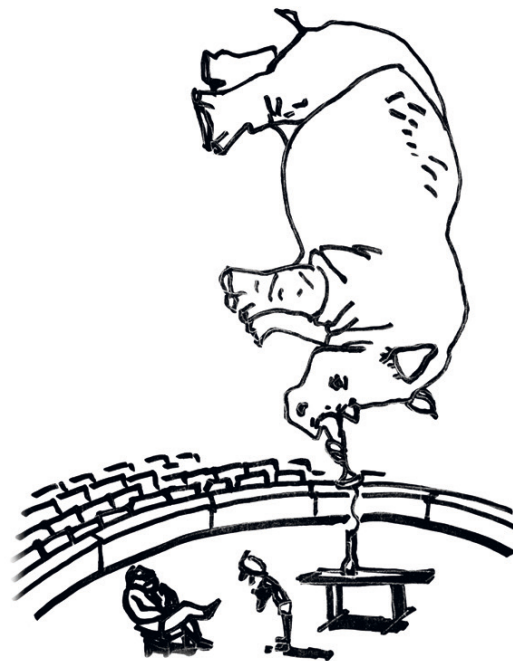
WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

Unterrichtskompetenz erfassen und entwickeln

Eine qualitative Studie über die Entwicklung eines
Bewertungskonzepts für die zweite Phase der Lehrerbildung

David Luhr

Die Lehrprobe



„...und warum haben Sie keinen
Elefanten genommen?“

Fachbereich

Erziehungswissenschaft

Unterrichtskompetenz erfassen und entwickeln

**Eine qualitative Studie
über die Entwicklung
eines Bewertungskonzepts
für die zweite Phase der Lehrerbildung**

Inaugural-Dissertation zur Erlangung des Doktorgrades
der Philosophischen Fakultät
der Westfälischen Wilhelms-Universität
zu Münster (Westf.)

vorgelegt von

David Luhr

2015

Tag der mündlichen Prüfung: 12.10.2015

Hauptfach Erziehungswissenschaft, Nebenfach Soziologie

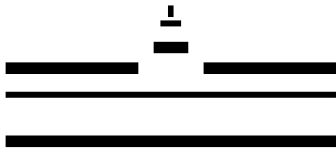
Prüfungskommission: Prof. Dr. Ewald Terhart (Erstgutachter),
Prof. Dr. Friedhelm Brüggem (Zweitgutachter),
Prof. Dr. Matthias Grundmann (Nebenfachprüfer)

David Luhr

Unterrichtskompetenz erfassen und entwickeln



MV WISSENSCHAFT



**WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER**

Wissenschaftliche Schriften der WWU Münster

Reihe VI

Band 15

David Luhr

Unterrichtskompetenz erfassen und entwickeln

Eine qualitative Studie über die Entwicklung eines Bewertungskonzepts
für die zweite Phase der Lehrerbildung

Wissenschaftliche Schriften der WWU Münster

herausgegeben von der Universitäts- und Landesbibliothek Münster

<http://www.ulb.uni-muenster.de>

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Dieses Buch steht gleichzeitig in einer elektronischen Version über den Publikations- und Archivierungsserver der WWU Münster zur Verfügung.

<http://www.ulb.uni-muenster.de/wissenschaftliche-schriften>

David Luhr

„Unterrichtskompetenz erfassen und entwickeln. Eine qualitative Studie über die Entwicklung eines Bewertungskonzepts für die zweite Phase der Lehrerbildung“

Wissenschaftliche Schriften der WWU Münster, Reihe VI, Band 15

© 2016 der vorliegenden Ausgabe:

Die Reihe „Wissenschaftliche Schriften der WWU Münster“ erscheint im Verlagshaus Monsenstein und Vannerdat OHG Münster

www.mv-wissenschaft.com

Dieses Werk ist unter der Creative-Commons-Lizenz vom Typ 'CC BY-SA 3.0 DE'

lizenziert: <http://creativecommons.org/licenses/by-sa/3.0/de/>

Von dieser Lizenz ausgenommen sind Abbildungen, welche sich nicht im Besitz des Autors oder der ULB Münster befinden.



ISBN 978-3-8405-0134-0

(Druckausgabe)

URN urn:nbn:de:hbz:6-26249655785

(elektronische Version)

direkt zur Online-Version:

© 2016 David Luhr

Alle Rechte vorbehalten

Satz:

David Luhr

Titelbild:

Tobias Buss

Umschlag:

MV-Verlag

Druck und Bindung:

MV-Verlag



Danksagung

Mit der Veröffentlichung dieser Arbeit geht ein fünf-jähriger Arbeitsprozess zu Ende, in dessen Verlauf mich zahlreiche Personen unterstützt haben.

An erster Stelle gilt mein Dank den Ausbildern und Auszubildenden des Zentrums für schulpraktische Lehrerausbildung Bonn, die sich an der Entwicklung und Erprobung des Bewertungskonzepts beteiligt haben. Ohne ihre Bereitschaft, sich auf neue Ideen und Herangehensweisen der Bewertung einzulassen, hätte das Projekt nicht realisiert werden können. Aus Gründen der Anonymität muss ich auf den persönlichen, namensbezogenen Dank verzichten. Ebenso bin ich der ehemaligen Leiterin des ZfsL Bonn, Fr. Barbara Utz, zu tiefem Dank verpflichtet. Sie hat die notwendigen Impulse gegeben, damit sich ausreichend Beteiligte für eine Erprobung finden ließen und das Projekt stets unterstützt. Durch die Ratschläge meines Doktorvaters Prof. Dr. Ewald Terhart konnte ich das Projekt in einem Rahmen halten, der in einer Promotion zu bewältigen war. Ohne seine wissenschaftliche Expertise, mit der er mir immer zur Seite stand, hätte ich mich mehr als einmal übernommen. Dafür gebührt ihm mein großer Dank! Meinem Kollegen Tobias Buss möchte ich für die Gestaltung des Titelbildes danken.

Meiner Familie, allen voran meinen Eltern und meiner Frau, möchte ich ebenfalls meinen Dank aussprechen. Ohne ihre großzügige und bedingungslose Unterstützung wäre es mir nicht möglich gewesen den kraftraubenden Weg einer berufsbegleitenden Promotion über einen langen Zeitraum von fünf Jahren zu beschreiten. Danke!

Inhalt

Einleitung	1
1 Die Zweite Phase der Lehrerbildung in der Bundesrepublik Deutschland	5
1.1 Rechtliche Grundlagen, Gemeinsamkeiten und Unterschiede der Ausbildung im föderalen Bildungssystem	5
1.2 Das Referendariat als ‚problematische Ausbildungspraxis‘ – Ausbildungs- und Bewertungspraxis aus Sicht der Auszubildenden	10
1.2.1 Die Differenz von ‚Regel und Praxis‘	12
1.2.2 Problematische Abhängigkeitsverhältnisse zwischen Mentoren und Referendaren	15
1.2.3 ‚In der Lehrprobe da machst du ne Show‘ – die ambivalente Rezeption der Unterrichtsbesuche	18
1.3 Die Bewertungspraxis im Referendariat aus Sicht der Wissenschaft	20
1.4 Anforderungen an das zu entwickelnde Bewertungsinstrument	22
2 Konzeptentwicklung – was ist ‚guter‘ Unterricht und wie lässt sich Unterrichtsqualität ‚messen‘?	25
2.1 Was ist ‚guter‘ Unterricht? – ein kurzer Überblick über die Unterrichtsforschung	27
2.1.1 Unterrichtsforschung und Didaktik – ‚fremde Schwestern‘?	27
2.1.2 Was kann die Unterrichtsforschung leisten?	29
2.1.3 Konzepte guten Unterrichts	32
2.1.4 Kritik an den Konzepten guten Unterrichts	39
2.1.5 Relevante Merkmale guten Unterrichts für die Ausbildung in der zweiten Phase der Lehrerbildung	48
2.1.5.1 Klarheit und Strukturiertheit	51
2.1.5.2 Hoher Anteil echter Lernzeit	53
2.1.5.3 Kompetenzorientierung	54
2.1.5.4 Aktivierung	55
2.1.5.5 Umgang mit Heterogenität	56
2.1.5.6 Lernförderliches Klima	56
2.1.5.7 Sinnstiftendes Kommunizieren	58
2.1.5.8 Transparente Leistungserwartungen	60
2.1.5.9 Vorbereitete Umgebung	62
2.1.5.10 Lernzuwachs	63
2.1.5.11 Fehlende Merkmale guten Unterrichts	64

2.2	Kompetenzbegriffe, Kompetenzmodelle und Erfassung von Kompetenzen....	65
2.2.1	Kompetenzbegriff und Kompetenzmodelle	66
2.2.2	Die Messung und Erfassung von beruflicher Handlungskompetenz	72
2.2.3	Unterrichtskompetenz – ein heuristisches Kompetenzmodell und die Erfassung der Unterrichtskompetenz.....	74
2.2.4	Kompetenzen als Zielkategorie von Lernprozessen – eine sinnvolle Perspektive pädagogischen Handelns und erziehungswissenschaftlicher Forschung im Bereich der Lehrer(aus)bildung?	85
2.3	Kompetenzen entwickeln – Kompetenzentwicklung erfassen: pre-Alpha-Version des Bewertungskonzepts.....	97
2.3.1	Der Beobachtungsbogen.....	97
2.3.2	Der Dokumentationsbogen für die Nachbesprechung.....	99
2.3.3	Die Sammlung über das Repertoire.....	99
3 	Forschungsdesign und theoretische Fundierung des Projektes	101
3.1	Paradigmatische Einordnung des Projektes – Handlungsforschung und qualitative Sozialforschung	101
3.2	Projekttablauf und Forschungsmethoden	109
3.2.1	Gruppendiskussionen unter den Ausbildern und Auszubildenden	110
3.2.2	Erprobung der Instrumente in der Ausbildungspraxis und Datenerhebung durch problemzentrierte Interviews.....	118
3.2.3	Auswertung im Sinne qualitativer Evaluationsforschung	121
3.3	Theoretischer Rahmen – Widerstände in institutionellen Veränderungsprozessen.....	124
4 	Auswertung der Gruppendiskussionen	127
4.1	Das hierarchische Kategoriensystem der Gruppendiskussionen.....	128
4.2	Die Logik der Ausbildungssituation/Rahmenbedingungen von Unterrichtsbesuchen	129
4.2.1	Die Sichtweise der Ausbilder	135
4.2.2	Die Sichtweise der Auszubildenden.....	154
4.2.3	Zusammenfassende Interpretation ‚Logik der Ausbildungssituation/ Rahmenbedingungen von Unterrichtsbesuchen‘	179
4.3	Standardisierung der Bewertungsprozesse.....	181
4.3.1	Die Sichtweise der Ausbilder	183

4.3.2	Die Sichtweise der Auszubildenden.....	196
4.3.3	Zusammenfassende Interpretation ‚Standardisierung der Bewertungsprozesse‘	203
4.4	Design des Bewertungskonzepts/der Instrumente	204
4.4.1	Leitende Konstruktionsprinzipien	207
4.4.2	Überarbeitung der Instrumente	222
4.5	Von der pre-Alpha- zur Alpha-Version des Bewertungskonzepts.....	233
4.6	Abschließende Interpretation der Gruppendiskussionen – axiales Codieren.....	238
5 	Auswertung der Interviews	243
5.1	Das hierarchische Kategoriensystem der Interviews.....	243
5.2	Konzept Einsatz der Ausbilder	244
5.2.1	Einsatz bei A2 (Ausbilder 2)	246
5.2.2	Einsatz bei A5.....	251
5.2.3	Einsatz bei A6.....	256
5.2.4	Zusammenfassende Interpretation ‚Konzept Einsatz Ausbilder‘	263
5.3	Konzeptbewertung.....	267
5.3.1	Konzeptbewertung bei A2	271
5.3.2	Konzeptbewertung bei A5.....	295
5.3.3	Konzeptbewertung bei A6.....	313
5.3.4	Passung mit fachlichen Qualitätsstandards.....	341
5.4	Abschließende Interpretation der Interviews – axiales Codieren.....	348
5.5	Von der Alpha- zur Beta-Version	353
6 	Abschließendes Fazit – Was leistet ein standardisiertes Bewertungskonzept in der zweiten Phase der Lehrerbildung?.....	361
	Literatur- und Quellenverzeichnis.....	367
	Zitierte Verordnungen	378

Anhang	379
Matrix der Qualitätsbereiche guten Unterrichts in Anlehnung an Helmke und Meyer (pre-Alpha-Version).....	380
pre-Alpha-Version des Bewertungskonzeptes	384
<i>Beobachtungsbogen ,A‘ mit Kompetenzskala (pre-Alpha-Version)</i>	384
<i>Beobachtungsbogen ,B‘ mit Ratingskala (pre-Alpha-Version)</i>	394
<i>Dokumentationsbogen für die Nachbesprechung (pre-Alpha-Version)</i>	396
<i>Sammlung über das Repertoire (pre-Alpha-Version)</i>	398
Alpha-Version des Bewertungskonzeptes	411
<i>Operationalisierung der Studententypen</i>	411
<i>Matrix der Qualitätsbereiche guten Unterrichts (in Anlehnung Helmke und Meyer)</i> <i>(AlphaVersion)</i>	413
Dokumentationsbogen für die Nachbesprechung (Alpha-Version)	418
Sammlung über das Repertoire (Alpha-Version).....	423
Leitfaden für die problemzentrierte Gruppendiskussion zur Auswahl zentraler Teilkompetenzen und der Rezeption der Instrumente mit Fachleitern	429
Leitfaden für die problemzentrierte Gruppendiskussion zur Rezeption der Bewertungsinstrumente mit Referendaren	435
Leitfaden für das problemzentrierte Interview zur Erprobung der Bewertungsinstrumente mit Fachleitern	440
Leitfaden für das problemzentrierte Interview zur Erprobung der Bewertungsinstrumente mit Referendaren.....	446
Transkriptionsregeln	452
„Bonner Qualitätstableau“	454
„Bubbles“ des ZfsL Bonn	456
Dokumentationsbogen für die Nachbesprechung des ZfsL Bonn	457
Matrix der Qualitätsbereiche guten Unterrichts, reduzierte Version (Beta-Version).....	458
Qualitätsbereiche guten Unterrichts	461

Einleitung

Im Zuge der internationalen Schulvergleichsstudien wie PISA oder TIMSS zeigte sich, dass Schulsysteme, die eine systematische Qualitätssicherung auf der Basis standardbasierter Kompetenzen betreiben, eher dazu in Lage sind, ‚kompetente‘¹ Schüler aus der Schule in das Leben zu entlassen, als solche Schulsysteme, die keine systematische standardbasierte Qualitätssicherung betreiben (vgl. z.B. Klieme 2002). In diesem Kontext entstanden bildungspolitische Bestrebungen in Deutschland, ein System der standardbasierten Qualitätssicherung einzuführen. Zu den ersten praktisch erfahrbaren Ergebnissen dieses Reformprozesses gehören u.a. die nationalen Bildungsstandards für Fächer im Primarbereich und die mittleren Schulabschlüsse, die Lernstandserhebungen in den Jahrgangsstufen drei und acht sowie das nationale Bildungspanel. Diese Instrumente sollen es ermöglichen, einen landes- bzw. bundesweiten Vergleich der Leistungsfähigkeit der Schüler und damit auch der jeweiligen Schulsysteme der Bundesländer zu erhalten (vgl. z.B. Köller 2010). Die in diesem Kontext stattfindenden Reformbestrebungen und Entwicklungen werden in dieser Arbeit unter dem Begriff der ‚standardbasierten Bildungsreform‘ verhandelt und bilden den größeren Kontext, in den sich die Promotion einordnet.

Auch auf die Organisation der Lehrerbildung hat die standardbasierte Bildungsreform Einfluss, z.B. in der Formulierung von ‚Standards für die Lehrerbildung‘ durch die Kultusministerkonferenz. Mit der Verabschiedung der ‚Standards für die Lehrerbildung: Bildungswissenschaften‘ im Dezember 2004 kamen die Bundesländer überein „[...] die vorgelegten Standards für die Lehrerbildung zu implementieren und anzuwenden [...] und] die Lehrerbildung auf der Grundlage der gemeinsam vereinbarten Standards zu evaluieren“ (Kultusministerkonferenz 2004). Im Jahr 2008 ergänzte die KMK diese ersten Standardisierungsbestrebungen in der Lehrerbildung durch ‚Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung‘ (vgl. Kultusministerkonferenz 2008).

¹ Siehe zur Diskussion um den Kompetenzbegriff ausführlich Kapitel 2.2 dieser Arbeit.

Aus der scientific community werden zeitgleich allerdings Klagen laut, dass zuverlässige empirische Daten auf die sich diese Bildungsreform im Bereich der Lehrerbildung stützen könnte, nur sehr eingeschränkt vorliegen (vgl. Allemann-Ghionda & Terhart 2006: 7, Blömeke 2004, Blömeke 2007: 13–16, Lacher & Oelkers 2004: 129, Schulte 2008: 15–24, Strietholt & Terhart 2009: 622, Terhart et al. 2011: 345–447). Zu diesem Mangel an empirisch begründeter Forschung gesellt sich auf der Umsetzungsebene der Ausbildung, also den Ausbildungsinstitutionen der Zweiten Phase der Lehrerbildung, ein Mangel an geeigneten Verfahren und Methoden, um die Kompetenzentwicklung der angehenden Lehrer systematisch und gehaltvoll zu erfassen (vgl. Baumert & Kunter 2006: 469 & 485, Maag Merki & Werner 2011: 576, Terhart 2007: 40–41). Strietholt und Terhart fassen die diesbezüglichen Ergebnisse einer Studie zu den in der zweiten Phase der Lehrerausbildung genutzten Beurteilungsinstrumenten wie folgt zusammen:

„Aus kompetenzdiagnostischer Sicht muss die Qualität dieses Prozesses [des Beurteilungsprozesses, D. L.] selbst jedoch zumindest als problematisch, wenn nicht in manchen Varianten als unzureichend beurteilt werden“ (Strietholt & Terhart 2009: 643).

Der Nutzen von Ausbildungsstandards, KMK-Übereinkünften und Verwaltungsverordnungen über interne Evaluation von Studienseminaren wird allerdings so lange gering sein und keine oder nur marginale Auswirkungen auf die Ausbildung und damit auch auf die tatsächlich abrufbaren Kompetenzen des Lehrpersonals haben, wie es an qualitativ hochwertigen empirischen Instrumenten mangelt, welche in der Ausbildung direkt eingesetzt werden können. Die konkrete Gestaltung praktisch nutzbarer Diagnose- und Evaluationsinstrumente stellt demnach den sprichwörtlichen ‚Flaschenhals‘ für den Erfolg der standardbasierten Bildungsreform im Bereich der Lehrerbildung dar. Der Mangel an empirisch abgesichertem Wissen über die Wirkung von Standardisierungsbestrebungen im Bereich von Bewertungsprozessen bildet ein Hindernis für die Entwicklung eben dieser nötigen Diagnostikinstrumente.

An diesem ‚Flaschenhals‘ der oben beschriebenen bildungspolitischen Reform setzt die Promotion an: *Ein Ziel des Promotionsprojektes ist es, mit Hilfe wissenschaftlicher Methoden aus der Ausbildungspraxis heraus ein Diagnostikinstrument für den zentralen Bereich der Fähigkeit zu unterrichten zu entwickeln, das den Ansprüchen von Ausbildern, Auszubildenden und Wissenschaftlern gerecht wird und die Auszubildenden bei der Entwicklung ihrer persönlichen ‚Unterrichtskompetenz‘ unterstützt.* In diesem Prozess kommt den Betroffenen der Reform, den Ausbildern und den Auszubildenden zu der Frage eine zentrale Rolle zu: Ihre Rückmeldungen, welche Instrumentarien geeignet sind, um den Prozess der Standardisierung produktiv und gewinnbringend in die Praxis zu übertragen, sollen die entscheidenden Determinanten für die Entwicklung von konkreten Bewertungsinstrumenten sein. Damit versucht diese Arbeit die Forderung von Oelkers und Reusser einzuholen wenn diese fordern:

„Was politisch intendiert ist, muss zusammen/gemeinsam mit den stets auch ausgehend von eigenen Deutungen handelnden Akteuren konstruiert und auf Verständigung hin kommuniziert werden“ (Oelkers & Reusser 2008: 45–46).

Das Projekt folgt somit einem ‚bottom-up-Ansatz‘ der Institutionsentwicklung, in dem es die betroffenen Akteure in die Entwicklung des Bewertungskonzepts einbindet. Weitergehend soll mit Hilfe dieser Promotion ein *Beitrag zur Theoriebildung über die Standardisierung von Bewertungsprozessen* geleistet werden. Im Rahmen rekonstruktiver Sozialforschung werden die besonderen Anforderungen, die eine Standardisierung von Bewertungsprozessen mit sich bringt, aus den qualitativ erhobenen Daten herausgearbeitet, um die *Grundsätze einer gegenstandsbezogenen Theorie der Standardisierung von Bewertungsprozessen in der Zweiten Phase der Lehrerbildung zu entwickeln.*

Kapitel eins kommt im Rahmen der Arbeit die Aufgabe zu, die Rahmenbedingungen der zweiten Phase der Lehrerbildung in Deutschland darzustellen. Aktuelle Herausforderungen und Probleme der Ausbildung und hierbei insbesondere der Bewertungsprozesse stehen im Mittelpunkt des Interesses. Kapitel

zwei widmet sich der Dokumentation der Entwicklung des Bewertungskonzepts, während in Kapitel drei die wissenschaftliche Methodologie der Arbeit ausführlich dargelegt wird. Kapitel vier analysiert nach Maßgabe der zuvor in Kapitel drei entwickelten Methodologie die qualitativ gewonnenen Daten und fasst die Entwicklungsschritte des Bewertungskonzepts, die sich aus den Daten ableiten ließen zusammen. Kapitel fünf zieht schlussendlich ein Fazit des Promotionsprojektes und fasst die wesentlichen Erkenntnisse, die sich aus dem Projekt für die Frage produktiver Bewertungsprozesse in der zweiten Phase der Lehrerbildung und hierbei insbesondere der Rolle von Standardisierung für diese, gewinnen ließen, zusammen.

Die Arbeit ordnet sich einem qualitativen, rekonstruktiven Wissenschaftsverständnis zu, aber hierbei ist zu erwähnen, dass die Arbeit aus den persönlich erlebten Herausforderungen gelingender Lehrerbildung in der zweiten Phase entstanden ist. Zahlreiche geschilderte Aspekte der Ausbildungspraxis entspringen demnach der persönlichen Erfahrung des Autors oder den Gesprächen mit Akteuren der Ausbildung und stellen implizites Wissen dar, das so nirgendwo formalisiert erfasst ist, für die Bearbeitung des hier vertretenen Erkenntnisinteresses aber immanent wichtig ist. Dies betrifft z.B. die Erwartungen, die von den Akteuren an die Unterrichtsbesuche oder unterrichtspraktische Prüfungen, gestellt werden. Die Arbeit versucht größtmöglich Transparenz über die Herkunft dieses Wissens herzustellen, z.B. indem Kapitel eins den Forschungsstand zur zweiten Phase der Lehrerbildung aufarbeitet. Aber an einzelnen Stellen der Arbeit werden erlebte Rahmenbedingungen, die sich den Akteuren, welche die Ausbildung durchlaufen haben, schnell als „praktisches Wissen“ offenbaren, als vorerst nicht wissenschaftlich abgesicherte Erkenntnisgrundlage der Arbeit stehen bleiben müssen.

1 | Die Zweite Phase der Lehrerbildung in der Bundesrepublik Deutschland

1.1 Rechtliche Grundlagen, Gemeinsamkeiten und Unterschiede der Ausbildung im föderalen Bildungssystem

Die Zweite Phase der Lehrerbildung² in Deutschland gestalten entsprechend dem Bildungsföderalismus die Bundesländer in eigener Verantwortung. Die Ausbildungsordnungen ähneln sich hinsichtlich der Dauer des Vorbereitungsdienstes, der Ausbildungselemente und der zu erreichenden Ausbildungsziele, wobei im Detail durchaus Unterschiede festzustellen sind (vgl. dazu: Schubarth et al. 2007: 41–51, Schulte 2008: 28–41, Walke 2007: 10–12). Während z.B. in Nordrhein-Westfalen die Ausbildung an den sogenannten ‚Zentren für schulpraktische Studien‘ (ZfsL)³ und ihnen zugeordneten Schulen stattfindet, erleben die Auszubildenden⁴ in Bayern das erste halbe Jahr der Ausbildung an sog. ‚Seminarschulen‘, um die darauf folgenden zwölf Monate der Ausbildung an ihren ‚Einsatzschulen‘ zu verbringen. Erste sind Schulen, denen ein Seminar für Lehrerausbildung angeschlossen ist, zweite ‚normale‘ staatliche und private Schu-

² Wenn im Folgenden vom ‚Vorbereitungsdienst‘, der ‚Ausbildung‘, der ‚Zweiten Phase der Lehrerbildung‘ oder dem ‚Referendariat‘ gesprochen wird, ist in allen Fällen die zweite Ausbildungsphase zum Lehrerberuf gemeint, die sich an die universitäre Ausbildung anschließt. Die folgenden Ausführungen beziehen sich i.W. auf die Ausbildung für das Lehramt an Gymnasien und Gesamtschulen, wobei die Vorbereitungsdienste für andere Schulformen i.W. den gleichen Richtlinien folgen, sich lediglich die Ausbildungsschwerpunkte unterscheiden (vgl. z.B. für die Primarstufe Englert et al. 2006: 40–44; Pille 2013: 21–22). Vgl. zur zweiten Phase der Lehrerbildung in Deutschland auch Lenhard 2004, zu der Bedeutung von Berufsvorbereitungsprogrammen für Lehrer in internationaler Perspektive Abs & Anderson-Park 2014.

³ Im Folgenden werden die Lehrerbildungsinstitute der Zweiten Phase der Einfachheit halber durchgängig als ‚Studienseminare‘ bezeichnet, auch wenn diese im Einzelnen nicht die offizielle Bezeichnung der Institutionen in den verschiedenen Bundesländer sein mag. Beziehen sich die Ausführungen auf eine konkrete Institution, wird der treffende Titel gewählt.

⁴ Die Begriffe der ‚Referendare‘, ‚Auszubildenden‘ und ‚Lehramtsanwärtern‘ werden im Folgenden synonym genutzt. Aus Gründen der Lesbarkeit wird im gesamten Text ausschließlich die männliche Form genutzt.

len. Der Vorbereitungsdienst in Nordrhein-Westfalen und Baden-Württemberg dauert 18 Monate, wohingegen jener in Bayern 24 Monate umfasst. Die Referendare in Baden-Württemberg sind verpflichtet ihren Ausbildern vor dem zweiten Staatsexamen lediglich sechs Unterrichtsstunden vorzuführen, während ihre Referendarskollegen in Bremen mindestens 18 Hospitationen vorweisen müssen (vgl. OVP 2011, ZALG 1992, AprOGymn 2004, Ausbildungs- und Prüfungsordnung für Lehrämter 2008).

Neben den Detailunterschieden in der Gestaltung der Zweiten Phase der Lehrerbildung weisen die Ausbildungs- und Prüfungsordnungen aller 16 Bundesländer doch auch wichtige Gemeinsamkeiten auf. Alle Bundesländer setzen auf eine Mischung aus schulpraktischer Ausbildung in Schulen und theoretischer Ausbildung in Pädagogik, pädagogischer Psychologie und Fachdidaktik in fachlich angeleiteten Seminaren. Bundesweit müssen Referendare im Unterricht ihrer Kollegen und Ausbilder hospitieren, unter Anleitung von Fachlehrern unterrichten, später auch eigene Klassen und Kurse leiten und sollen weitergehend alle Bereiche des Berufsfeldes ‚Lehrer‘ kennenlernen, in denen sie sich später bewähren müssen. Kurzum sie sollen ‚Kompetenzen‘⁵ entwickeln, die wichtig sind um später in der Schule einen guten Job zu machen und bis zum Ende ihrer Dienstzeit bestehen zu können.

Im Jahr 2004 vereinbarte die Kultusministerkonferenz (KMK) in den sog. *Standards für die Lehrerbildung* grundlegendes Wissen und zu beherrschende Fähigkeiten, die Lehrer im Zug ihrer Ausbildung erlangen sollen. Diese Standards lassen sich dem Dokument folgend in vier Kompetenzbereiche einteilen:

- Unterrichten
- Erziehen
- Beurteilen
- Innovieren

⁵ Eine genauere Auseinandersetzung um den erziehungswissenschaftlichen Begriff der Kompetenz erfolgt in Kapitel 2.2.

Diese Kompetenzbereiche weisen jeweils zahlreiche Teilkompetenzen auf, die ihrerseits theoretisches Wissen und praktisches Können erfordern. Abbildung 1.1 zeigt als Beispiel die Operationalisierung des Kompetenzbereichs ‚Unterrichten‘, dem im hier diskutierten Kontext besondere Bedeutung zukommt⁶.

Abbildung 1.1: Kompetenzbereich Unterrichten der KMK-Standards für die Lehrerbildung – Lehrerinnen und Lehrer sind Fachleute für das Lehren und Lernen (vgl. KMK 2004a)

Kompetenz 1:	
Lehrerinnen und Lehrer planen Unterricht unter Berücksichtigung unterschiedlicher Lernvoraussetzungen und Entwicklungsprozesse fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.	
Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p><i>Die Absolventinnen und Absolventen...</i></p> <ul style="list-style-type: none"> ▪ kennen die einschlägigen Erziehungs- und Bildungstheorien, verstehen bildungs- und erziehungstheoretische Ziele sowie die daraus abzuleitenden Standards und reflektieren diese kritisch. ▪ kennen allgemeine und fachbezogene Didaktiken und wissen, was bei der Planung von Unterrichtseinheiten beachtet werden muss. ▪ kennen unterschiedliche Unterrichtsmethoden und Aufgabenformen und wissen, wie man sie an forderungs- und situationsgerecht einsetzt. ▪ kennen Konzepte der Medienpädagogik und -psychologie und Möglichkeiten und Grenzen eines anforderungs- und situationsgerechten Einsatzes von Medien im Unterricht. ▪ kennen Verfahren für die Beurteilung von Lehrleistung und Unterrichtsqualität. 	<p><i>Die Absolventinnen und Absolventen...</i></p> <ul style="list-style-type: none"> ▪ können aus den einschlägigen Erziehungs- und Bildungstheorien Zielperspektiven und Handlungsprinzipien ableiten. ▪ verknüpfen fachwissenschaftliche und fachdidaktische Argumente und planen und gestalten Unterricht auch unter der Berücksichtigung der Leistungsheterogenität. ▪ wählen Inhalte und Methoden, Arbeits- und Kommunikationsformen unter Bezug auf Curricula und ggf. individuelle Förderpläne aus. ▪ integrieren moderne Informations- und Kommunikationstechnologien didaktisch sinnvoll und reflektieren den eigenen Medieneinsatz. ▪ überprüfen die Qualität des eigenen Lehrens und reflektieren die Passung zu den Lernvoraussetzungen und Lernbedürfnissen der Schülerinnen und Schüler.

⁶ Detailliertere Ausführungen zu den KMK-Standards und deren Umsetzung in den einzelnen Bundesländern finden sich z.B. bei Walke 2007: 17- 23; 38–42.

Kompetenz 2:	
Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schülerinnen und Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu nutzen.	
Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p><i>Die Absolventinnen und Absolventen...</i></p> <ul style="list-style-type: none"> ▪ kennen Lerntheorien und Formen des Lernens. ▪ kennen Grundlagen und Ergebnisse der Schul- und Unterrichtsforschung und können diese anwendungsbezogen reflektieren. ▪ wissen, wie man Lernende aktiv in den Unterricht einbezieht und Verstehen und Transfer unterstützt. ▪ kennen Theorien der Lern- und Leistungsmotivation und Möglichkeiten, wie sie im Unterricht angewendet werden. 	<p><i>Die Absolventinnen und Absolventen...</i></p> <ul style="list-style-type: none"> ▪ regen unterschiedliche Formen des Lernens an und unterstützen sie. ▪ gestalten Lehr-Lernprozesse unter Berücksichtigung der Erkenntnisse über den Erwerb von Wissen und Fähigkeiten. ▪ stärken bei Schülerinnen und Schülern Lern- und Leistungsbereitschaft. ▪ führen und begleiten Lerngruppen.
Kompetenz 3:	
Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.	
Standards für die theoretischen Ausbildungsabschnitte	Standards für die praktischen Ausbildungsabschnitte
<p><i>Die Absolventinnen und Absolventen...</i></p> <ul style="list-style-type: none"> ▪ kennen Lern- und Selbstmotivationsstrategien, die sich positiv auf Lernerfolg und Arbeitsergebnisse auswirken. ▪ kennen Methoden der Förderung selbstbestimmten, eigenverantwortlichen und kooperativen Lernens und Arbeitens. ▪ wissen, wie sie weiterführendes Interesse und Grundlagen des lebenslangen Lernens im Unterricht entwickeln. ▪ vermitteln und fördern Lern- und Arbeitsstrategien. ▪ vermitteln den Schülerinnen und Schülern Methoden des selbstbestimmten, eigenverantwortlichen und kooperativen Lernens und Arbeitens. 	<p><i>Die Absolventinnen und Absolventen...</i></p> <ul style="list-style-type: none"> ▪ vermitteln und fördern Lern- und Arbeitsstrategien. ▪ vermitteln den Schülerinnen und Schülern Methoden des selbstbestimmten, eigenverantwortlichen und kooperativen Lernens und Arbeitens.

Die sog. ‚KMK-Standards‘ bilden die Grundlage für die Lehrerausbildung in Universität und in der Zweiten Phase der Lehrerbildung in allen Bundesländern. Sie wird durch Ausbildungsverordnungen der Bundesländer aufgegriffen und konkretisiert. Die Ordnung des Vorbereitungsdienstes in Nordrhein-Westfalen

definiert z.B. im sog. ‚Kerncurriculum‘ des Vorbereitungsdienstes (siehe Abbildung 1.2) sechs Handlungsfelder von Lehrern, die sich an den Standards der Lehrerbildung der KMK orientieren (vgl. OVP 2011).

Abbildung 1.2: Handlungsfelder des Vorbereitungsdienstes in NRW laut Kerncurriculum (vgl. OVP 2011)

Handlungsfeld (Nummer)	Zugeordnete Kompetenzen
Unterricht gestalten und Lernprozesse anregen (1)	Lehrerinnen und Lehrer planen Unterricht fach- und sachgerecht und führen ihn sachlich und fachlich korrekt durch.
	Lehrerinnen und Lehrer unterstützen durch die Gestaltung von Lernsituationen das Lernen von Schülerinnen und Schülern. Sie motivieren Schülerinnen und Schüler und befähigen sie, Zusammenhänge herzustellen und Gelerntes zu nutzen.
	Lehrerinnen und Lehrer fördern die Fähigkeiten von Schülerinnen und Schülern zum selbstbestimmten Lernen und Arbeiten.
Den Erziehungsauftrag in Schule und Unterricht wahrnehmen (2)	Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung.
	Lehrerinnen und Lehrer vermitteln Werte und Normen und unterstützen selbstbestimmtes Urteilen und Handeln von Schülerinnen und Schülern.
	Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht.
Leistungen herausfordern, erfassen, rückmelden, dokumentieren und beurteilen (3)	Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern.
	Lehrerinnen und Lehrer erfassen Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe.
Schülerinnen und Schüler und Eltern beraten (4)	Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern (siehe auch Handlungsfeld 3).
Vielfalt als Herausforderung annehmen und Chancen nutzen (5)	Lehrerinnen und Lehrer kennen die sozialen und kulturellen Lebensbedingungen von Schülerinnen und Schülern und nehmen im Rahmen der Schule Einfluss auf deren individuelle Entwicklung (siehe auch Handlungsfeld 2).
Im System Schule mit allen Beteiligten entwicklungsorientiert zusammenarbeiten (6)	Lehrerinnen und Lehrer sind sich der besonderen Anforderungen des Lehrerberufs bewusst. Sie verstehen ihren Beruf als ein öffentliches Amt mit besonderer Verantwortung und Verpflichtung.
	Lehrerinnen und Lehrer verstehen ihren Beruf als ständige Lernaufgabe.
	Lehrerinnen und Lehrer beteiligen sich an der Planung und Umsetzung schulischer Projekte und Vorhaben.

Entsprechend der Zielsetzung des Vorbereitungsdienstes, zukünftige Lehrer an ihre Aufgaben heranzuführen und sie bei der Entwicklung zentraler berufsbezogener Kompetenzen zu unterstützen, beziehen die *Beurteilungsprozesse*, wie sie in den gültigen Rechtsgrundlagen konkretisiert werden, die verschiedenen Kompetenzbereiche des Lehrerhandelns ein. Alle Bundesländer legen der Gesamtnote des zweiten Staatsexamens die Urteile verschiedener Personen mit spezifischen Blickwinkeln zu Grunde. Die Schulleiter der Ausbildungsschulen beurteilen das gesamte dienstliche Verhalten der angehenden Lehrer sowie die unterrichtlichen Fähigkeiten, wie sie sie in Unterrichtshospitationen erlebt haben. Mündliche Prüfungen zu allgemein- und fachdidaktischen Fragestellungen, in einigen Bundesländern auch zu Fragen des Schulrechts, erweitern die Überprüfung des Ausbildungserfolges um eine theoretische Dimension. Unterrichtspraktische Prüfungen sollen schließlich Aufschluss über die Kompetenzen im Kernbereich des beruflichen Handelns von Lehrern, eben dem Unterrichten, geben. Die Gewichtung der verschiedenen Einzelnoten ist in den Bundesländern verschieden.

1.2 Das Referendariat als ‚problematische Ausbildungspraxis‘ – Ausbildungs- und Bewertungspraxis aus Sicht der Auszubildenden

Das Referendariat⁷ an sich besitzt in der Öffentlichkeit und v.a. bei Lehramtsstudenten keinen guten Ruf, es wird gelegentlich als „die schlimmste Zeit des Lebens“ bezeichnet, so z.B. die Überschrift eines Onlineforums der Süddeutschen Zeitung (vgl. Gerstenberg 2007). Der Wechsel von der theoretischen Auseinandersetzung mit Erziehungsprozessen und didaktischen Theorien an der Universität zur praktischen Umsetzung eben dieser und die damit verbundenen neuen Herausforderungen sorgen bei einem Großteil der Referendare für Stress, der das

⁷ Auch in diesem Fall liegt der Fokus der Darstellung auf dem Vorbereitungsdienst für das Lehramt an Gymnasium und Gesamtschulen. Im Vorbereitungsdienst für andere Schulformen zeigen sich jedoch nahezu gleichgelagerte Probleme (vgl. z.B. für die Primarstufe Englert et al. 2006, insb.: 53–54; 315–328; 359–361; 379–427; 440–447).

Leben während der Ausbildungszeit zu einem erheblichen Maß bestimmt (vgl. dazu: Dietrich 2014: 1–2). Diese Negativbewertung durch Erfahrungsberichte von Referendaren in populären Medien decken sich nur teilweise mit den Ergebnissen der Forschung zur zweiten Phase der Lehrerbildung. Neben der dort ebenfalls zu findenden, teils erheblichen, Kritik am Vorbereitungsdienst, kristallisieren sich aber auch deutlich positive Aspekte des Referendariats heraus (vgl. dazu: Werner-Bentke 2010: 219–298, Schubarth et al. 2007: 84–89, Schulte 2008: 120–125).

Die folgenden Ausführungen stützen sich auf eine Auswahl der zur Verfügung stehenden Studien zur Zweiten Phase der Lehrerbildung in Deutschland. Der Forschungsstand zur Zweiten Phase ist insgesamt als prekär zu beschreiben, auch wenn seit dem Jahr 2000 eine verstärkte Forschungsaktivität festzustellen ist (vgl. zu dieser Einschätzung: Dietrich 2014: 7, Schulte 2008: 25, Werner-Bentke 2010: 17). Schulte führt in ihrer systematischen Zusammenfassung des Forschungsstandes zur Zweiten Phase der Lehrerbildung für die Zeitperiode von 1970 bis 2008 aus, dass insbesondere Studien, die repräsentative Aussagen für den Vorbereitungsdienst in der gesamten Bundesrepublik ermöglichen, seit den 1970er Jahren nicht mehr veröffentlicht worden sind (vgl. Schulte 2008: 20). Die Forschung wird Schulte folgend von Studien dominiert, die sich auf einzelne Bundesländer, Unterrichtsfächer oder auf einzelne Studienseminare beziehen, was die Generalisierbarkeit der Ergebnisse deutlich erschwert (vgl. dies.: 23). In einem kumulativen Sinne (vgl. Blömeke 2007: 14) baut diese Promotion auf den zur Verfügung stehenden Ergebnissen der Lehrerbildungsforschung auf. Um eine solide Analyse der Ausgangsbedingungen für die Entwicklung des Bewertungsinstruments leisten zu können, werden jene Studien berücksichtigt, welche die Wahrnehmung des Vorbereitungsdienstes seitens der Referendare oder der Ausbilder systematisch analysieren. Hierbei wird in diesen nach verallgemeinerbaren Erkenntnissen, hinsichtlich der für den hier diskutierten Kontext wichtigen Ausbildungssituationen und Rahmenbedingungen der Ausbildung gesucht. Nicht jede der zitierten Studien kann zu jeder der unten als bedeutsam ausgemachten Ausbildungsstrukturen Belege liefern. Einzelne Studien, v.a. die mit einem sehr spezifischen Blickwinkel – jene von Dietrich, Storr und Pille (vgl. Dietrich 2014, Pille 2013, Storr 2006) – erhalten bei der Rekonstruktion einzelner

Strukturmerkmale der zweiten Phase besondere Bedeutung. Berücksichtigt wurden im Folgenden die Studie von Schulte (vgl. Schulte 2008), die ‚Lehramtskandidaten-Studie Potsdam‘ von Schubarth et al. (vgl. Schubarth et al. 2006 & 2007), eine Diplomarbeit von Storr zu den besonderen Eigenschaften der Lehrproben (vgl. Storr 2006), die Dissertation von Werner-Bentke, welche die generelle Rezeption des Vorbereitungsdienstes durch Referendare und Ausbilder analysiert (vgl. Werner-Bentke 2010), eine Arbeit von Pille, in der die sozialen Praktiken im Sinne Bourdieu analysiert werden, mit denen die Referendare Eingang in den Beruf des Lehrers finden (vgl. Pille 2013) und schließlich die Promotion von Dietrich, der das Referendariat als Abfolge von Professionalisierungskrisen deutet (vgl. Dietrich 2014).

In Bezug auf das hier zu entwickelnde Bewertungskonzept, lassen sich aus den oben erwähnten Studien zwei zentrale Herausforderungen herausarbeiten, denen sich Referendare stellen und die sie bewältigen müssen, um den Vorbereitungsdienst als erfolgreich und gewinnbringend zu erleben. Nachdem diese im Folgenden vorgestellt werden, schließt sich eine vertiefte Auseinandersetzung über die Rezeption der Unterrichtsbesuche und der anschließenden Auswertungsgespräche an, da diese Ausbildungssituation der Dreh- und Angelpunkt des hier zu erarbeitenden Bewertungskonzepts ist.

1.2.1 Die Differenz von ‚Regel und Praxis‘

Wie oben beschrieben spielt sich die Ausbildung der angehenden Lehrer an zwei grundsätzlich verschiedenen Institutionen ab, dem Studienseminar und den Ausbildungsschulen. Beide Institutionen leisten im Optimalfall einen je spezifischen Beitrag zur Entwicklung von professioneller Handlungskompetenz. Im schlechteren Fall blockieren sich die unterschiedlichen Logiken, mit denen die Angehörigen der beiden Institutionen auf das Unterrichten und ‚Lehrersein‘ blicken, gegenseitig. Werner-Bentke zeigt in einer qualitativen Studie mit Referendaren aus dem Studienseminaren Halle/Saale und Dessau, dass die Schule als primäre ‚Initiationsinstanz‘ für den Einstieg in den Beruf des Lehrers dient, während das Studienseminar seinen Wert als Ort der Reflexion von eigenem unterrichtlichen Handeln mit professionellen Ausbildern und anderen Referendaren hat

(Werner-Bentke 2010: 331–343). Der ständige Wechsel zwischen diesen beiden Institutionen wird von vielen Referendaren allerdings als belastend erlebt, weil er erstens einen ständigen Rollenwechsel vom Lehrenden zum Lernenden erfordert und die Referendare zweitens mit sehr unterschiedlichen Interpretationsmustern und Konzepten von Schulwirklichkeit im Allgemeinen und ‚gutem Unterricht‘ im Besonderen konfrontiert werden. Diese Diskrepanz, jene zwischen den Anforderungen des Seminars an Unterricht und der tagtäglichen Unterrichtspraxis in der Schule, stellt laut Storr eine der grundlegenden Strukturen des Vorbereitungsdienstes dar, welche die Entwicklung einer persönlichen Berufskultur der Referendare maßgeblich prägt (vgl. Storr 2006). In ihrer Studie mit Berufsschul- und Gymnasialreferendaren zeigt sich, dass die wahrgenommene Differenz von normativ als gut erachteten ‚Regelunterricht‘ einerseits und defizitärem ‚Praxisunterricht‘ andererseits, die Referendare einem erheblichen Stress aussetzt. In den offenen, ein- bis zweistündigen Interviews der Studie suchen die Auszubildenden nach Rechtfertigungen für die wahrgenommene Diskrepanz zwischen ‚Regel‘ und ihrer eigenen ‚defizitären Handlungspraxis‘. Sie müssen angestrengt daran arbeiten, eine Ursache für die defizitäre Unterrichtspraxis außerhalb ihrer eigenen Person oder der eigenen Fähigkeiten zu finden. Einige der Referendare rechtfertigen sich in den Interviews für ihre ‚defizitäre Praxis‘ quasi vor sich selbst, da ein Bestehen im Referendariat ohne eine „[...]fremdreferenzielle Ursachenzuschreibung [...]“ (Storr 2006: 29) gar nicht möglich sei.

„Die Formulierungen «man muss [...]»(sic.) aufpassen» und «muss das so sehen» verdeutlichen, dass es zwingend notwendig ist, diese Perspektive einzunehmen“ (dies.: 26).

Vor dem Hintergrund der Zielsetzung des Referendariats, den angehenden Lehrern die Entwicklung jener Kompetenzen zu ermöglichen, mit denen sie erfolgreich in der Berufspraxis bestehen können, ist ein solcher ‚geheimer Lehrplan‘ des Vorbereitungsdienstes eine problematische Angelegenheit. Konzentrieren sich die Auszubildenden zu stark auf intervenierende Fremdeinflüsse, die eine ‚gute Unterrichtspraxis‘ erschweren oder mitunter auch unmöglich machen, dann wird die Entwicklung einer positiven Lehreridentität, die sich der Selbstwirksamkeit für

Unterrichtsprozesse bewusst ist, negativ beeinflusst. Wünschenswerter wäre ein gelassener Umgang mit den alltäglichen Hindernissen, welche die Realisierung des ‚guten Regelunterrichts‘ erschweren und eine Fokussierung auf die Frage, wie trotz aller Hindernisse eine pragmatische Definition von ‚gutem Unterricht‘ dauerhaft realisiert werden kann. Eine solche Perspektive würde die Entwicklung einer stabilen Lehrerpersönlichkeit, die sich auf eine adäquate Selbstwirksamkeitsüberzeugung stützen kann, positiv beeinflussen.⁸ Pille beschreibt die Reaktion der Auszubildenden auf diese erlebten Unterschiede der eigenen Praxis und der vermeintlichen Anforderungen des ‚Regelunterrichts‘ als die „Suche nach dem Greifbaren“ (Pille 2013: 114–121), in der die frischen Novizen nach Richtlinien suchen, um in ihrem Unterricht die ‚Regel‘ zu realisieren, ohne ein Gefühl für die „[...] Unschärfe der Praxis [...]“ entwickelt zu haben, die eine vollkommene Umsetzung des ‚Regelunterrichts‘ ohnehin nicht zulassen würden. Diese bei Pille angedeutete Fokussierung der notwendigen Entwicklung der Referendare, um zwischen den Polen der ‚Regel‘ und ‚Praxis‘ vermitteln zu können, bringt Dietrich noch deutlicher in den Diskurs ein. Entsprechend der Perspektivierung seiner Dissertation, die Krisenerfahrungen der Referendare als ‚Professionalisierungskrisen‘ zu begreifen, die immanenter Teil der Ausbildung in der zweiten Phase der Lehrerbildung sind, deutet er den erlebten Widerspruch zwischen dem geforderten ‚Regelunterricht‘ und der eigenen Unterrichtspraxis als Professionalisierungsimpuls an die Referendare. Sie würden auf die „[...]“ Entscheidungsoffenheit [...]“ der praktischen Unterrichtssituationen mit einer „[...]“ Kontrollerwartung[...]“ im Sinne des ‚Regelunterrichts‘ reagieren, welche die Krise erst hervorrufe, anstatt die prinzipielle Offenheit der Praxissituationen als solche anzunehmen (vgl. Dietrich 2014: 467–469).

Die Unterrichtsbesuche der Ausbilder spielen bei der negativen Wahrnehmung der ‚Differenz von Regel und Praxis‘ eine große Rolle und sind darüber hinaus für die Zielsetzung dieser Promotion von besonderer Bedeutung. Sie werden deswegen weiter unten (siehe Kapitel 1.2.3) ausführlich thematisiert.

⁸ Vgl. zur Bedeutung der Selbstwirksamkeit für die Herausbildung einer stabilen und positiven Lehrerpersönlichkeit Baumert & Kunter: 2006: 502–503, Hermann & Hertrampf 2002: 207–215, zum Einfluss von Selbstwirksamkeit auf professionelles Handlungswissen Schulte et al. 2008: 282–284.

1.2.2 Problematische Abhängigkeitsverhältnisse zwischen Mentoren und Referendaren

Neben den oben umrissenen unterschiedlichen Interpretationsmustern über ‚guten Unterricht‘ sorgen v.a. die zahlreichen Abhängigkeitsverhältnisse, in denen sich die Auszubildenden wiederfinden, für ein erhebliches Ausmaß an Stress. In der Schule müssen sich die Referendare mit Ausbildungslehrern, in deren Unterricht sie hospitieren wollen, gut stellen. V.a. die Beziehung zum Schuldirektor, dessen Gutachten in einigen Bundesländern ein Drittel der Note des zweiten Staatsexamens ausmacht (vgl. AprOGym 2004), ist von einer starken Abhängigkeit geprägt. In den Studienseminaren begegnen die Referendare Ausbildern, von denen sie sich einerseits vertrauensvoll beraten lassen sollen, die andererseits aber eben auch einen erheblichen Teil der Examensnote festlegen. Zusammen mit den unterschiedlichen Sichtweisen auf Schule und Unterricht durch die Angehörigen von Studienseminaren und Schulen finden sich die Referendare in einem Netzwerk von Abhängigkeiten und Widersprüchen wieder, in dem sie zwischen den verschiedenen Sichtweisen, Ansprüchen und Erwartungen vermitteln müssen. Wolfram bezeichnet ihre diesbezüglichen Erfahrungen aus dem Referendariat gar als ‚Vierfrontenkrieg‘, den sie zwischen den Ansprüchen von Schülern, Kollegen, Eltern und Seminar bestehen musste (vgl. Wolfram 2009). Folgende, von Wolfram beschriebene, Situation bringt die Probleme der ‚Differenz von Praxis und Regel‘ sowie des Netzwerks von Abhängigkeiten drastisch auf den Punkt:

„Während wir im Seminar immer neuen theoretischen ‚Input‘ bekamen, den wir in der ‚schulischen Realität‘ mit Leben füllen sollten, wurde die Situation an der Schule immer absurder. Eines Tages hatte ich einige Ordner mit der Aufschrift ‚Stationenlernen Ägypten‘ und ‚Lerntheke ss/ß‘ auf meinem Platz stehen. In einem stillen Moment wurde ich vom Schulleiter beiseite genommen und gebeten, die Ordner doch bitte unter den Tisch zu stellen. Auf meinen verstörten Blick hin erklärte er mir, dass sich einige Kollegen beschwert hätten, die sich durch solche Ordner unter Druck gesetzt sähen. Wenngleich alles in mir rebellierte, war mir nur allzu bewusst, dass von Seiten der Schule noch ein Gutachten anstünde, mit hohem Gewicht für die Endnote. Also stellte ich die Ordner unter den Tisch!“ (Wolfram 2009: 20).

Auch wenn dieser Bericht eine fast schon aberwitzige und so sicher nicht alltägliche Situation schildert, so spiegelt er doch sowohl die Konflikte, welche die Referendare kanalisieren müssen, als auch die bisweilen problematischen Auswirkungen der Bewertungsprozesse im Referendariat plakativ wieder. Während die Ausbilder in den Seminaren der angehenden Lehrerin entsprechend ihrer Rolle den Zugang zu vielfältigen Arten der Unterrichtsinszenierung eröffnen, nehmen die ‚Praktiker‘ an der Schule diese Interventionen in ihren Alltag als Bedrohung wahr. Die Referendarin sieht sich nun in der misslichen Lage, zwischen diesen Polen zu vermitteln, weil sie von beiden bewertet wird. Der Ausbilder wird in einer Lehrprobe natürlich erwarten, dass er zumindest teilweise die vom ihm als ‚gut‘ erachteten und deswegen an die Referendarin weitergegeben Unterrichtsmethoden auch wiederfinden kann. Berücksichtigt sie dies nicht, wird er kritisch nachfragen, warum dies so ist und der Umstand, dass er seine Ausbildung u.U. nicht wertgeschätzt sieht, wird möglicherweise seine Bewertung beeinflussen. Gleichzeitig muss die angehende Lehrerin dem Schulleiter gegenüber Loyalität zeigen. Tut sie dies nicht, sind negative Auswirkungen auf ihre Bewertung nicht auszuschließen.

In den Studien zur Zweiten Phase der Lehrerbildung zeigt sich, dass die von Wolfram geschilderten Erfahrungen durchaus verallgemeinerbar sind. In der Studie von Schulte wird deutlich, dass die Beziehung zum Schulleiter für eine nicht unerhebliche Anzahl von Referendaren als belastend erlebt wird. In den Interviews der Studie berichten sie über den Druck, den der Schulleiter durch seine Bewertungsposition aufbaue und sie dazu nötige, Aufgaben anzunehmen, die über das im Rahmen der Ausbildung zu leistende Maß hinausgehe (vgl. Schulte 2008: 240–241). In ihrer Studie zur Zweiten Phase der Lehrerbildung an den Studienseminaren in Brandenburg weisen Schubarth et al. nach, dass ein Großteil der Referendare die Beziehung zu den Seminarleitern als gewinnbringend und positiv erlebt. Allerdings trübt sich diese positive Einschätzung durch die als ambivalent erlebten Bewertungsprozesse im Anschluss an die Unterrichtsbesuche ein (vgl. Schubarth et al. 2007: 85–89; 97–98). Werner-Bentke kommt in seiner Promotion zu ähnlichen Ergebnissen wie Schubarth et al.. Die Auszubildenden schätzen die fachliche Expertise der Ausbilder nahezu durchgehend als positiv und gewinnbringend, aber die Beziehung zwi-

schen beiden wird durch die Abhängigkeit von den Bewertungen der Ausbilder und den dadurch ausgelösten ‚zwanglosen Zwang‘, ihren Empfehlungen zu folgen, teils deutlichen Belastungen ausgesetzt (vgl. Werner-Bentke 2010: 270–271; 276–282). Insbesondere die weiter unten thematisierten Unterrichtsbesuche der Ausbilder mit anschließenden Auswertungsgesprächen sorgen bei den Auszubildenden für erheblichen Unmut.

Die Beziehungen zwischen den Referendaren und ihren Mentoren werden darüber hinaus immer wieder durch Rollenkonflikte herausgefordert, denen sich die Referendare stellen müssen. Sie sind noch keine ‚fertigen Lehrer‘, müssen aber im eigenverantwortlichen Unterricht als solche handeln und sich vor den Schülern behaupten. Gleichzeitig gehört es zu der Ausbildung, quasi an letzter Stelle der ‚Hackordnung‘ in der Schule zu stehen und im Studienseminar immer wieder in die Rolle des Lernenden schlüpfen zu müssen (vgl. Pille 2013: 94–104). Für den Entwicklungsprozess der Auszubildenden optimal wäre eine professionelle Coaching-Beziehung zwischen ihnen und ihren Mentoren, aber die Ergebnisse der qualitativen Forschung zeigen andersartige Erfahrungen der Referendare. So berichten Teilnehmer der Studie von Werner-Bentke, dass sie in den Protokollen schulischer Gremien nicht als Kollegen, sondern als ‚Gäste‘ vermerkt wurden, oder Konflikte um das pädagogische Handeln von Referendaren hinter deren Rücken zwischen Mentoren und Eltern ausgetragen wurden, ohne dass sie zu dem Sachverhalt selbst hätten Stellung beziehen können (vgl. Werner-Bentke 2010: 241–244). Werner-Bentke fasst die Sichtweise der Referendare auf die Beziehungen zu den Seminaarausbildern wie folgt zusammen:

„Als Fazit lassen sich unterschiedliche Sichtweisen festhalten. Einerseits wurden die Seminarleiter als fachlich kompetent bewertet. Wenn die fachliche Kompetenz mit Praxiserfahrung und sozial-kommunikativen Fähigkeiten einherging, waren sie besonders geschätzt. Andererseits ergab sich aus der personenzentrierten Ausbildungssituation, die einen ausgesprochen subjektiven Charakter trug, im Zusammenhang mit dem existenziellen Notendruck ein Abhängigkeitsverhältnis, das als belastend empfunden wurde und die Entwicklung der eigenen Lehrerpersönlichkeit und des Lehrerstils durchaus einengen konnte“ (Werner-Bentke 2010: 282).

Die Auswertungsgespräche zwischen Auszubildenden und Ausbildern im Anschluss an die Unterrichtsbesuche spielen auch im Kontext dieses problematischen Rollenkonfliktes eine entscheidende Rolle. Sie werden im folgenden Abschnitt ausführlich analysiert.

1.2.3 ‚In der Lehrprobe da machst du ne Show‘ – die ambivalente Rezeption der Unterrichtsbesuche

Der in der Überschrift zitierte Titel von Storrs Studie zum Referendariat in Berlin bringt die Wahrnehmung der Unterrichtsbesuche seitens der Referendare auf den Punkt (vgl. Storr 2006). Sie werden als hochgradig künstliche Situationen empfunden, die mit der tatsächlichen Unterrichtspraxis wenig bis nichts gemeinsam haben und deren Bewertung durch die Ausbilder stark subjektiv geprägt ist. Gleichwohl gewinnen viele Referendare den Unterrichtsbesuchen und den sich anschließenden Auswertungsgesprächen positive Seiten für ihre eigene Professionalisierung ab (vgl. z.B. Schulte 2008: 104, Werner-Bentke 2010: 266–269), aber der Ablauf der anschließenden Auswertungsgespräche und Bewertungsprozesse, sowie der Inszenierungszwang der Unterrichtsbesuche ist Ziel teils vehementer Kritik.

Der oben umrissene Problemkontext der ‚Differenz von Regel und Praxis‘ zeigt sich bei den Unterrichtsbesuchen besonders deutlich. Die Referendare sehen den im Zuge der Unterrichtsbesuche gezeigten Unterricht als künstlich und alltagsfern an. Während einige Referendare die Bedeutung der Vorbereitung und Durchführung von ‚gutem Regelunterricht‘ für ihre Professionalisierung im Sinne des exemplarischen Prinzips anerkennen (vgl. z.B. Werner-Bentke 2010: 268–269), kritisieren andere Referendare gerade diese ‚Künstlichkeit‘ und sehen keinen Sinn darin für die Unterrichtsbesuche Unterricht vorzubereiten, den sie aus der ‚Praxis‘ so nicht kennen und von dem sie auch nicht glauben, dass er Teil ihrer Unterrichtspraxis werden wird (vgl. Pille 2013: 244, Schubarth et al. 97–98, Schulte 2008: 273, Storr 2006: 25–44, Werner-Bentke 2010: 271–273).

In der bereits mehrfach zitierten Studie von Schubarth et al. bewerten alle beteiligten Gruppen (Auszubildende, Ausbilder an Schule und Seminar) den

Vorbereitungsdienst überwiegend positiv, aber in Bezug auf die Wahrnehmung der Unterrichtsbesuche zeigen sich deutliche Unterschiede in der Wahrnehmung der Ausbilder und der Auszubildenden. Während die Ausbilder die Bewertungsprozesse nahezu einstimmig als transparent bezeichnen und immerhin ca. zwei Drittel der Befragten einheitliche, d.h. seminarübergreifend geltende Bewertungskriterien für die Unterrichtsbesuche postulieren, sehen die Referendare die Bewertungsrealität deutlich kritischer. Allerdings wurden die Referendare in der Studie gar nicht zu ihrer Wahrnehmung der Bewertungsprozesse befragt. Vielmehr ziehen Schubarth et al. Ergebnisse aus einer früheren Studie heran, in welcher die Referendare v.a. die Uneinheitlichkeit und die geringe Transparenz der Leistungserwartungen, sowie utopische Anforderungen an den zu zeigenden Unterricht kritisierten (vgl. Schubarth et al. 2007: 95–98). Folgendes Zitat aus den im Rahmen einer Studie aus dem Jahr 2005 durchgeführten Interviews verdeutlicht die problematische Wahrnehmung der Bewertungsprozesse durch die Auszubildenden.

„Was ich aber sagen wollte, die Fachleiter beurteilen ja nicht nur, also eigentlich beurteilen sie gar nicht so sehr den Unterricht, sondern wie wir sind. [...] (sic.) Das ist nämlich nicht festgelegt. [...] (sic.) Also, wo sind denn die Kriterien? Ich habe mich schon richtig angelegt und habe gesagt, ‚Wo sind die Kriterien?‘ (Zustimmung) ‚Die gibt es nicht.‘ Und ‚Das entscheide ich nach dem Gesamtbild.‘ wurde mir dann gesagt“ (Schubarth et al. 2007: 97).

Die in der Studie von Werner-Bentke interviewten Referendare zeigen eine ähnliche Einschätzung. Die Referendare äußern in dieser Studie zwar keinen expliziten Unmut über fehlende transparente Bewertungskategorien, aber sie fühlen sich abhängig von den persönlichen Vorlieben der Ausbilder und versuchen, diesen auf Grund der Bewertungssituationen gerecht zu werden, obwohl dies mitunter ihren eigenen Vorstellungen von ‚gutem Unterricht‘ widerspricht (vgl. Werner-Bentke 2010: 281–282). Zu einem gleichen Fazit wie Werner-Bentke kommt auch Schulte, wenn Sie schreibt, dass die Referendare in der Vorbereitung der Unterrichtsbesuche eher Überlegungen anstellen, was ein bestimmter Fach-

leiter für Erwartungen an Unterricht habe, als dass sie die Lernvoraussetzungen der Lerngruppe verstärkt in den Blick nähmen (vgl. Schulte 2008: 273). Die geringe Transparenz der Bewertungskriterien wird von den Referendaren in der Studie von Schulte explizit kritisiert. Von den 15 Referendaren, die sich insgesamt zu diesem Kontext geäußert haben, sagen nur ca. 50%, dass ihnen klar sei, ‚was der Fachleiter sehen will‘ (vgl. Schulte 2008: 205)⁹. Diese negative Einschätzung einer erheblichen Zahl von Studienteilnehmern, der Bewertungsprozesse im Allgemeinen und der Unterrichtsbesuche im Besonderen seitens der Referendare lassen sich, wie im kommenden Unterkapitel gezeigt werden soll, durch Forschungsstand zur zweiten Phase der Lehrerbildung bestätigen.

1.3 Die Bewertungspraxis im Referendariat aus Sicht der Wissenschaft

Die Promotion verfolgt das Ziel, ein Bewertungsinstrument zu entwickeln, das die Bewertungsprozesse im Vorbereitungsdienst dezidiert verbessert. Dies erfordert, dass neben den Klagen aus der Praxis auch die Einschätzungen aus der Wissenschaft, wie es um die Ausbildungs- und Bewertungsprozesse steht, berücksichtigt werden.

Im Rahmen einer explorativen Studie haben Strietholt und Terhart eine Auswertung der eingesetzten Bewertungsinstrumente an allen 257 Studienseminaren in Deutschland durchgeführt. Datengrundlage der Studie sind 201 Bewertungsinstrumente aus der Ausbildungspraxis, die den Forschern bereitgestellt wurden (vgl. Strietholt & Terhart 2009: 626). Die Schlüsse, zu denen sie nach der Auswertung dieser Instrumente kommen, sind teils sehr ernüchternd. Sie stellen fest,

⁹ Wobei an dieser Stelle anzumerken ist, dass allein durch die Annahme, dass ein Referendar weiß, ‚was der Fachleiter sehen will‘, noch kein hinreichendes Maß an Transparenz sichergestellt ist, das einem transparenten und gewinnbringenden Bewertungsprozess gerecht wird. Das ‚was ein Fachleiter sehen will‘, kann u.U. höchst kontraproduktiv für die Planung eines spezifischen Unterrichts sein. Wie Transparenz über die Bewertungskriterien auf eine sinnvolle Art und Weise hergestellt werden kann, wird weiter unten gezeigt (siehe Kapitel 2.1).

dass einem erheblichen Teil der in der Ausbildungspraxis eingesetzten Instrumente keine klaren Bewertungskategorien zu Grunde liegen, sondern oft eher pauschal ‚die Kompetenz‘ beurteilt wird, ohne dass eine klare Operationalisierung vorliegt, wie sich diese denn ausdrückt (vgl. dies.: 634; 642). In etwa der Hälfte der Bewertungsinstrumente ist unklar, auf welche Art und Weise die Bewertung erfolgen soll. Die Entscheidung für Freitextkommentare, Ratingskalen oder Kompetenzskalen erfolgt demnach nicht systematisch, sondern unterliegt dem Gutdünken der einzelnen Ausbilder (vgl. dies.: 641). Die Einheitlichkeit der Bewertungsprozesse innerhalb der Studienseminare muss ebenfalls als defizitär bezeichnet werden. Etwa ein Drittel der Instrumente wird lediglich von einzelnen Ausbildern eingesetzt, Instrumente, die innerhalb einer gesamten Institution eingesetzt werden, existieren praktisch nicht, dies deutet auf einen deutlichen Mangel an verbindlichen Absprachen in den Studienseminaren hin (vgl. dies.: 642). Ein gravierendes Problem der Bewertungsinstrumente ist der überwiegend mangelhafte oder teils sogar fehlende Bezug auf die gültigen Ausbildungsgrundlagen, sprich die KMK-Standards. 20% der Instrumente weisen keinen Bezug zu den KMK-Standards, 42% nur zu einem der in den Standards erwähnten Kompetenzbereiche auf. Hinsichtlich der in der Ausbildungspraxis eingesetzten Beurteilungsinstrumente kommen Strietholt und Terhart demnach zu folgendem Fazit:

„Ingesamt wird deutlich, dass hinsichtlich der Qualität der Beurteilungsprozesse und -instrumente, die in der Zweiten Phase der Lehrerbildung eingesetzt werden, ein dringender Entwicklungsbedarf besteht“ (dies.: 622).

Die negative Einschätzung der Bewertungsprozesse durch die Referendare, wie sie weiter oben vorgestellt wurde, deckt sich demnach mit der Defizitbeschreibung derselben, wie Strietholt und Terhart sie formulieren.

Aus diesen Defizitbeschreibungen der Bewertungsprozesse im Vorbereitungsdienst aus wissenschaftlicher Sicht und den im Kapitel 1.2 herausgearbeiteten Defiziten des Referendariats aus Sicht der Referendare sollen nun Ansprüche an

das zu entwickelnde Bewertungsinstrument abgeleitet werden, die erfüllt sein müssen, damit das Instrument sein Ziel, auf die drängenden Probleme der Ausbildungspraxis zu reagieren, diese zu entschärfen und für einen Beitrag zur gewinnbringenden Implementation der standardbasierten Bildungsreform in die Ausbildungspraxis der Zweiten Phase der Lehrerbildung zu sorgen, erfüllen kann.

1.4 Anforderungen an das zu entwickelnde Bewertungsinstrument

Die im vorangegangenen Unterkapitel zusammengefassten Problemfelder der Zweiten Phase der Lehrerausbildung spitzen sich, wie gezeigt, in den bewerteten Unterrichtsbesuchen zu. Die Unterschiede zwischen ‚Regel‘- und ‚Praxisunterricht‘, sowie die Abhängigkeitsverhältnisse zwischen Auszubildenden und Ausbildern und die Rollenkonflikte der Referendare treten hier besonders deutlich zu Tage. Gleichzeitig müssen die Bewertungsprozesse des Vorbereitungsdienstes im Allgemeinen und der Unterrichtsbesuche im Besonderen aus einer wissenschaftlichen Perspektive als defizitär bezeichnet werden. Demnach bietet diese Ausbildungssituation aber eben auch ein großes Entwicklungspotential, die teils negative Rezeption der Ausbildung durch die Auszubildenden zu verbessern. Das zu entwickelnde Bewertungskonzept soll dies leisten, indem es die oben umschriebenen Problemfelder der Praxis aufgreift und Möglichkeiten aufzeigt, wie die einzelnen Konflikte entschärft werden können. Zeitgleich sollen aus der Defizitbeschreibung von Strietholt und Terhart wissenschaftlich fundierte Qualitätskriterien für ein zeitgemäßes Bewertungskonzept der Unterrichtsbesuche gewonnen werden. Zusammenfassend lassen sich an dieser Stelle fünf Ansprüche an das Konzept formulieren:

- *Das Konzept soll* zu einer Steigerung der Transparenz der Bewertungsgrundlagen führen, damit der wahrgenommenen Beliebigkeit und Subjektivität der Beurteilungen entgegengewirkt wird und die Referendare schriftlich fixierte Bewertungsmaßstäbe haben, auf die sie sich im Konfliktfall berufen können.

- *Das Konzept soll* einen deutlichen und transparenten Bezug zu allgemein anerkannten Kriterien ‚guten Unterrichts‘ aufweisen und zu erlangende Einzelkompetenzen einer allgemeinen ‚Unterrichtskompetenz‘ transparent und eindeutig operationalisieren.
- *Das Konzept soll* der Individualität der Lehrerpersönlichkeit der Auszubildenden gerecht werden, denn es gibt ebenso wenig ‚den‘ guten Lehrer, wie es ‚den‘ guten Unterricht gibt¹⁰.
- *Das Konzept soll* eine in der Praxis nutzbare, aber zeitgleich wissenschaftlichen Ansprüchen genügende Erfassung von Unterrichtskompetenz ermöglichen.
- *Das Konzept soll* auf verschiedene Unterrichtsfächer übertragen werden können, damit es innerhalb eines gesamten Studienseminars eingesetzt werden kann.

Mit diesen Ansprüchen an das Bewertungsinstrument wird ein zentraler Defizitbereich, den Strietholt und Terhart ausgemacht haben, nicht berücksichtigt. Sie beklagen, dass ein Großteil der Bewertungsprozesse des Vorbereitungsdienstes auf das Unterrichten zugeschnitten ist, obwohl das Berufsfeld der Lehrer durch zahlreiche andere Handlungsfelder geprägt ist (vgl. KMK 2004a). Dieser Kritik ist sicherlich zuzustimmen, und demnach hat das hier zu entwickelnde Bewertungsinstrument auch nicht den Anspruch, alle Probleme der Bewertungsprozesse im Vorbereitungsdienst gänzlich aus der Welt zu schaffen. Es hat vielmehr zum Ziel die Bewertungsprozesse im eng umgrenzten Bereich der Unterrichtsbesuche gemäß den oben formulierten Qualitätskriterien zu verbessern.

¹⁰ Siehe dazu auch Kapitel 2.1.

2 | Konzeptentwicklung – was ist ‚guter‘ Unterricht und wie lässt sich Unterrichtsqualität ‚messen‘?

Die im vorangegangenen Kapitel formulierten Ansprüche legen den Rahmen fest in dem die Instrumente zur Bewertung des Unterrichts in der Zweiten Phase der Lehrerbildung entwickelt werden sollen. Übergeordnetes Ziel der Zweiten Phase der Lehrerbildung ist es natürlich, ‚gute‘ Lehrer aus den Ausbildungsinstitutionen zu entlassen (vgl. Schulte 2008: 42), wobei mit dieser Aussage noch nicht geklärt ist, was einen ‚guten‘ Lehrer ausmacht. Diese Arbeit verfolgt nicht das vermessene Ziel, diese Frage zu beantworten – sofern sie denn in dieser Allgemeinheit überhaupt zu beantworten ist –, sondern sie setzt sich ein klarer definiertes und enger eingegrenztes Ziel:

Ziel der Arbeit ist es, ein Bewertungskonzept zu entwickeln, mit dessen Hilfe der Erwerb von ‚Unterrichtskompetenz‘ in den Unterrichtsfächern der angehenden Lehrer unterstützt wird, indem eine transparente und systematische Diagnose des Entwicklungsstandes unabdingbarer, basaler Teilkompetenzen von Unterrichtskompetenz ermöglicht wird.

Das Konzept eignet sich demnach *nicht* dazu, zu beurteilen, ob jemand ein ‚guter Lehrer‘ ist. Es geht lediglich darum, zu beurteilen, ob über einen längeren Beobachtungszeitraum eine Entwicklung der Unterrichtskompetenz, wie sie in den aufgeführten Teilkompetenzen näher operationalisiert ist, zu beobachten ist.

Die Fähigkeit, ‚guten Unterricht‘¹¹ zu erteilen, ist nicht hinreichend, um ein ‚guter‘ Lehrer¹² zu sein, aber sicherlich eine notwendige Bedingung.

Der Weg zum Ziel der Lehrerbildungsinstitutionen, ‚gute Lehrer‘ in die Schulen zu bringen, geht also über die Frage, was ‚guten Unterricht‘ ausmacht, hinaus. Diese Frage steht im Zentrum dieses Kapitels der Arbeit. In einem ersten Schritt sollen einige einführende Worte über die Unterrichtsforschung erfolgen, auf denen aufbauend die Konzeption ‚guten Unterrichts‘, die dieser Arbeit zu Grunde liegt, entwickelt wird (Kapitel 2.1). Es schließt sich eine Auseinandersetzung mit dem Begriff der Kompetenz als Zielkategorie von Lernprozessen an (Kapitel 2.2). Diese beiden Aspekte werden dann in der eigentlichen Entwicklung von einer ersten Version des angestrebten Bewertungskonzepts zusammengeführt (Kapitel 2.3). Um Klarheit in die folgenden Ausführungen zu bringen ist es an dieser Stelle notwendig, schon einen ersten Blick auf die zu entwickelnden Bewertungsinstrumente zu werfen. Diese setzten sich in einer ersten Version des Bewertungskonzepts, die in Anlehnung an die Entwicklungsstadien von Software als ‚pre-Alpha-Version‘ bezeichnet wird, aus drei Einzelinstrumenten zusammen: Zentrales Instrument ist ein *standardisierter Beobachtungsbogen*, der für die Bewertung der Unterrichtsbesuche eingesetzt werden sollte. Ergänzt wird dieses Instrument um einen *Dokumentationsbogen* für die Nachbesprechung der Unterrichtsbesuche, sowie die sog. *Sammlung über das Repertoire*, welche das im Rahmen der Lehrproben gezeigte Kompetenzrepertoire der Auszubildenden dokumentieren sollte.

¹¹ Der Begriff des ‚guten‘ Unterrichts ist hier durch Anführungsstriche hervorgehoben, weil noch nicht näher operationalisiert ist, was unter diesem zu verstehen ist. Diese Relativierung wird aufgehoben, sobald der Begriff mit konkreten Konzepten guten Unterrichts verknüpft wird.

¹² Die Frage, welche Eigenschaften einen guten Lehrer ausmachen, bzw. welche Fähigkeiten er mitbringen soll, versuchen die bereits zitierten Bildungsstandards für die Lehrerbildung der KMK zu beantworten (vgl. KMK 2004a). Zeichner fasst ähnliche Ansätze aus den USA zusammen (vgl. Zeichner 2004: 98–103). Bei Goodwin und Oyler finden sich weitere Hinweise auf Indikatoren für ‚gute Lehrer‘ (vgl. Goodwin & Oyler 2008, insb. 472–481). Die American Education Research Association lieferte 2004 eine kurze Übersicht über Forschungsergebnisse zum Einfluss von ‚guten Lehrern‘ auf Lernerfolge von Schülern (vgl. AREA 2004). Zumwalt und Craig ermöglichen schlussendlich einen Überblick über die Forschung zu Qualitätskriterien für ‚gute Lehrer‘ aus einer englischsprachigen Perspektive (vgl. Zumwalt & Craig 2005). Cochran-Smith kommt zu dem lapidaren Schluss: „[...] Everybody likes teacher quality and wants more of it. The problem is, that there is no consensus about, what it is“ (Cochran-Smith 2005: 6).

2.1 Was ist ‚guter‘ Unterricht? – ein kurzer Überblick über die Unterrichtsforschung

Der Begriff des Unterrichts bedarf eigentlich keiner Klärung, denn was unter diesem zu verstehen ist, ist Teil des Wissens und der Erfahrungen nahezu jedes Menschen – zumindest in den ‚entwickelten‘ Industrieländern. Dementsprechend verfügt auch ein Großteil der Menschen über subjektive Theorien, welche Faktoren den Unterricht prägen und was ‚guten‘ Unterricht ausmacht. Lehrern wird im öffentlichen Diskurs gemeinhin eine große Verantwortung für die Realisierung ‚guten‘ Unterrichts zugewiesen, während an anderer Stelle die limitierenden Faktoren erfolgreichen Unterrichts durch vermeintlich schlecht ausgestattete Bildungssysteme betont werden oder die mangelnde ‚Bildungsfähigkeit‘ der Kinder beklagt wird (vgl. Paulsen 2012, Lehmann 2012).

Einige dieser ‚populären‘ Erklärungen lassen sich durchaus durch die Erkenntnisse erziehungswissenschaftlicher Forschung bestätigen, andere müssen relativiert werden. Der Fokus der folgenden Auseinandersetzung mit den Forschungsergebnissen zur Frage, was denn ‚guten‘ Unterricht ausmacht, liegt auf den Arbeiten von zwei Autoren (Andreas Helmke sowie Hilbert Meyer), deren Konzeptionen ‚guten‘ Unterrichts Grundlage des hier zu entwickelnden Bewertungskonzepts sind. Die Darstellung der Überlegungen dieser beiden Autoren soll ergänzt werden um einige einleitende Ausführungen zur Unterrichtsforschung und ihren Erkenntnismöglichkeiten, um die Möglichkeiten, aber auch Grenzen des hier zu entwickelnden Bewertungskonzepts zu umreißen¹³.

2.1.1 Unterrichtsforschung und Didaktik – ‚fremde Schwestern‘?

Es mag etwas verwunderlich scheinen, aber im Rahmen einer Darstellung zentraler Ergebnisse der Unterrichtsforschung im deutschen Sprachraum, macht es Sinn, mit einer Problematisierung des Begriffes der ‚Unterrichtsforschung‘ an

¹³ Zur vertieften Auseinandersetzung mit der Unterrichtsforschung sei auf folgende Literatur und die dort weiter angegebenen Quellen verwiesen: Arnold 2009: 27–32, Kiel 2010, Lüders & Rauin 2004, weitere Konzepte ‚guten‘ Unterrichts werden in Kapitel 2.1.3 ausgeführt.

sich zu beginnen¹⁴. Auf den ersten Blick scheint allein die Begrifflichkeit der ‚Unterrichtsforschung‘ ihr die wissenschaftliche Deutungshoheit über die Frage, was ‚guten Unterricht‘ ausmacht, zu bescheren. Allerdings herrscht im deutschen Sprachraum¹⁵ ein Konflikt zwischen der *Unterrichtsforschung* als empirisch ausgerichteter Wissenschaft, die versucht Bedingungsfaktoren erfolgreichen Unterrichts zu ermitteln und der *Didaktik*, als „[...] geisteswissenschaftlich geprägtem Nachdenken über Unterricht [...]“ (Kiel 2010: 781), die Theorien und Modelle über Unterricht erstellt und die Lehre an der Hochschule, wie Unterricht auszusehen habe, dominiert¹⁶. Aus dieser (künstlichen) Trennung der beiden Unterdisziplinen der Erziehungswissenschaft resultierten zahlreiche Auseinandersetzungen über die Frage, was ‚guten Unterricht‘ ausmacht. Während der Didaktik immer wieder eine versteckte normative Ausrichtung ihrer Empfehlungen vorgeworfen wurde sowie eine mangelnde empirische Absicherung ihrer Modelle und Handlungsanweisungen (vgl. Terhart 2002: 81), erschien die empirische Unterrichtsforschung als abgehobene Wissenschaft, die sich vornehmlich mit technologisch ausgerichteten Prozessmodellen von Unterricht abgab, deren Erkenntnisse aber keine Hilfen für Bildungspraktiker bereitstellen würden (vgl. ders.: 80). Das Verhältnis der beide Unterdisziplinen scheint sich auch bis heute nicht entspannt zu haben. So schreiben Arnold u.a. im Vorwort zu einem im Jahr 2009 erschienenen Sammelband, dass die von Terhart gewählte Bezeichnung der ‚fremden Schwestern‘ (vgl. Terhart 2002), „[...] fast noch als beschönigend zu bewerten [...]“ (Arnold et al. 2009: 9) sei. Gleichwohl existieren Synopsen der

¹⁴ Der Begriff der Unterrichtsforschung wird hier synonym zum ansonsten auch gebräuchlichen Begriff der ‚Lehr- Lernforschung‘ genutzt.

¹⁵ Auch in anderen Ländern zeichnen sich Konflikte zwischen den verschiedenen Traditionen erziehungswissenschaftlicher Forschung und Theoriebildung zum Unterricht ab. So beklagt Shirley für die USA eine zu starke Konzentration auf quantitativ ausgerichtete Forschungsansätze und einen Mangel an geisteswissenschaftlicher didaktischer Tradition (vgl. Shirley 2009).

¹⁶ Diese Unterscheidung von Unterrichtsforschung, bzw. Lehr- Lernforschung und Didaktik ist sicherlich nicht die einzig mögliche Lesart der Beziehungen dieser beiden Unterdisziplinen der Erziehungswissenschaften, sie gehört aber im postmodernen Sinne zu den Metanarrativen der Erziehungswissenschaft als Disziplin und wird hier so hingenommen. Für eine weitergehende Auseinandersetzung des Verhältnisses von Didaktik und Unterrichtsforschung vgl. Arnold et al. 2009, Baumert & Köller 2000: 271, Helmke et al. 2007: 66–68, Kiel 2010: 781–782; 786, Terhart 2002, zum Verhältnis von Bildungstheorie und empirischer Bildungsforschung i.A. vgl. Koller 2012.

Expertisen aus beiden Sphären, die es ermöglichen, empirisch fundierte Aussagen darüber zu treffen, welche Faktoren ‚guten Unterricht‘ bestimmen und wie dieser gestaltet werden kann. Im deutschen Sprachraum haben sich v.a. Hilbert Meyer und Andreas Helmke um eine Verbindung von Ergebnissen der empirischen Unterrichtsforschung und konkreten Empfehlungen, wie ‚guter Unterricht‘ zu gestalten sei, verdient gemacht (vgl. Helmke 2010, Meyer 2007a). Beide, Meyer ehemals Volksschullehrer und ‚Didaktiker‘, als auch Helmke, von Haus aus Psychologe und ‚Unterrichtsforscher‘, haben Kataloge mit Kriterien ‚guten Unterrichts‘ ausgearbeitet und die Bedeutsamkeit der Kriterien jeweils mit zahlreichen Belegen aus der empirischen Unterrichtsforschung fundiert¹⁷ (vgl. Helmke 2010: 168–267, Meyer 2007a: 23–126), sodass in ihren Arbeiten die oben umschriebene Trennung zwischen Didaktik und Unterrichtsforschung teilweise aufgehoben ist. Helmke führt diesbezüglich aus:

„Obwohl zwischen den Sichtweisen der Schulpädagogen (von denen Hilbert Meyer sicher einer der prominentesten Vertreter ist) und der empirisch forschenden Pädagogischen Psychologie gelegentlich Welten liegen, ist dies bei der Betrachtung der Unterrichtsqualität keineswegs der Fall. Trotz unterschiedlicher Semantik ist klar erkennbar, dass es aus beiden Perspektiven einen Kernbereich der Unterrichtsqualität gibt“ (Helmke u.a. 2007: 57).

2.1.2 Was kann die Unterrichtsforschung leisten?

Die empirische Absicherung der Arbeiten von Helmke und Meyer muss durch die Erkenntnismöglichkeiten der empirischen Unterrichtsforschung kritisch betrachtet werden. V.a. Helmke führt die (derzeitigen) Grenzen wissenschaftlicher Erkenntnis hinsichtlich der Frage, welche (Einzel-)Merkmale ‚guten Unterricht‘

¹⁷ Wobei dazu angemerkt werden muss, dass zwischen den Empfehlungen zur Unterrichtsgestaltung die v.a. Meyer gibt und den als Beleg aufgeführten empirischen Ergebnissen bestenfalls ein sehr ‚loser‘ Zusammenhang besteht und Meyer bei einigen Merkmalen seines Unterrichts sehr selbstbewusst „behauptet“, dass diese wichtig seien, auch wenn er dafür keine empirischen Belege finden könne (vgl. z.B. Meyer 2007a: 121, zu dieser Kritik auch: Schilmöller 2006: 79). Siehe zu der Kritik am wissenschaftlichen Vorgehen von Meyer und Helmke Kapitel 2.1.4.

ausmachen, ausführlich aus (vgl. Helmke et al. 2007: 61–66, Helmke 2010: 22–35). Ausgehend von dem *variablenzentrierten Ansatz*¹⁸ der Unterrichtsforschung, dessen Ziel die Aufdeckung statistischer Zusammenhänge zwischen Merkmalen von Unterricht und dem Lernerfolg der Nachfrager des Unterrichts (=Schüler) ist, müssen klare Grenzen der Erkenntnismöglichkeiten anerkannt werden. U.a. muss berücksichtigt werden, dass die traditionellen Zusammenhangsmaße der Statistik im Bereich der Erforschung von Unterrichtsqualität an ihre Grenzen stoßen, denn die hier zu *findenden Zusammenhänge zwischen Merkmalen des Unterrichts und Lernerfolg entwickeln sich nicht linear*. Ein ‚guter Unterricht‘ zeichnet sich nicht durch die maximale Ausprägung eines bestimmten Kriteriums¹⁹, wie etwa der Strukturierung des Unterricht aus, sondern durch ein Optimum. Helmke zu diesem Zusammenhang:

„So leuchtet es ein, dass ein völliges Fehlen jeglicher Strukturierung des Unterrichts für den Lernerfolg ebenso ungünstig ist wie eine exzessive, übertriebene Strukturierung“ (Helmke 2010: 33).

Entsprechend reicht es nicht aus nur zu messen ob ein bestimmtes Merkmal stark ausgeprägt ist, die Frage ist auch, in welchem Bereich einer Merkmalskala das Optimum zu verorten ist und diese Frage ist nicht ohne weiteres klar zu beantworten, sondern muss ihrerseits durch statistische Modelle und deren Test bestätigt werden. An diese Frage schließt sich ein weiteres Problem der empirischen Absicherung ‚guten Unterrichts‘ an, *denn verschiedene Merkmale ‚guten Unterrichts‘ ergänzen sich nicht additiv, sondern multiplikativ*, d.h., dass geringe Ausprägungen eines Merkmals im gewissen Rahmen durch eine starke Ausprägung

¹⁸ Neben dem variablenzentrierten Ansatz führt Helmke weiterhin den personenzentrierten Ansatz der Unterrichtsforschung aus. Dieser sucht nicht nach statistischen Korrelationen zwischen bestimmten Merkmalen ‚guten‘ Unterrichts und dem Lernerfolg, sondern versucht aus der Beobachtung von nachweislich ‚erfolgreichen‘ Personen(-gruppen) Eigenschaften ‚guten Unterrichts‘ abzuleiten. Auch dieser Ansatz ist mit spezifischen Einschränkungen hinsichtlich der empirischen Absicherung seiner Ergebnisse behaftet (vgl. Helmke et al. 2007: 64–66, Helmke 2010: 26–28). Die Darstellung konzentriert sich hier auf die Darstellung des variablenzentrierten Ansatzes, da ein Großteil der von Helmke und Meyer zur Absicherung ihrer Merkmale angeführten empirischen Ergebnisse auf diesem Ansatz beruht (vgl. Helmke 2010: 171).

¹⁹ Siehe zu den Einzelmerkmalen ‚guten Unterrichts‘ im Sinne von Helmke und Meyer, Kapitel 2.1.5.

eines anderen Merkmals (Helmke verweist auf die Merkmale der Klassenführung und des Lernklimas) ausgeglichen werden können, aber eben nur bis zu einem gewissen Maß⁴. Helmke seinerseits zu diesem Problem:

„Weist nur ein [Herv. i. Org.] Faktor die Ausprägung null auf (Totalausfall) oder fällt er unter einen kritischen Schwellenwert, dann ist der Gesamtertrag ebenfalls gleich null (Multiplikation mit null ergibt null). Die dahinterstehende Frage, unter welchen Umständen man in der Unterrichtsforschung von Kopplung oder Kompensation von Merkmalen der Unterrichtsqualität (oder Mischformen) sprechen kann, ist theoretisch schwierig und methodisch anspruchsvoll und deswegen bisher kaum untersucht worden“ (ders.: 33).

Darüber hinaus hängen starke oder schwache Ausprägungen einzelner Merkmale teilweise stark zusammen, was dazu führt, dass das *statistische Problem der Multikollinearität* auftritt. Statistikprogramme können mit den eingegebenen Daten keine sinnvollen Berechnungen anstellen oder kommen zu dem Ergebnis, dass keines der Merkmale eines ‚guten Unterrichts‘ einen signifikanten Zusammenhang mit dem Lernerfolg aufweist, schlichtweg weil die Einflussfaktoren eines einzelnen Merkmals zu gering sind. Dieses Problem lässt sich laut Helmke kaum vollständig lösen, sondern durch geschickte und komplexe Studiendesigns bestenfalls entschärfen (vgl. Helmke et al. 2007: 63).

Ein weiteres Problem ist die Tatsache, dass es *erhebliche Wechselwirkungen zwischen den Merkmalen des Unterrichts und den Merkmalen der Schüler* einer Lerngruppe gibt. Ein Unterricht, der viele der im Rahmen empirischer Forschung abgesicherten Merkmale ‚guten Unterrichts‘ berücksichtigt, kann der einen Hälfte einer Lerngruppe gerecht werden und diese zu hohen Leistungen anspornen, während er die andere Hälfte frustriert und ausbremst. Wiederum Helmke zu diesem Punkt:

„So hat die Forschung gezeigt, dass ein hoch strukturierter, lehrerzentrierter Unterricht mit vielen förderlichen Hinweisen und unterstützenden Maßnahmen von selbstbewussten, selbstsicheren und leistungsstärkeren Schülerinnen und Schülern als bevormundend und demotivierend

empfundener wird, während diese Art von «direkter Instruktion» von leistungsängstlichen Schülerinnen und Schülern geschätzt wird und für ihren Lernfortschritt geradezu nötig ist“ (Helmke 2010: 34).

Es gibt demnach nicht ‚den‘ einen guten Unterricht, der durch empirische Ergebnisse aufgedeckt werden kann, sondern nur *passenden Unterricht für bestimmte Lerngruppen oder Schüler*²⁰.

Entsprechend dieser Ausführungen muss in Bezug zu der in der Überschrift dieses Unterkapitels formulierten Frage festgehalten werden, dass die im Folgenden vorgestellten Konzepte ‚guten Unterrichts‘ von Helmke und Meyer keinesfalls als objektive Wahrheit des guten Unterrichts an sich aufzufassen sind. Wie Helmke und Meyer beide explizit anmerken, handelt es sich vielmehr um heuristische Ausführungen, die Lehrkräften bei der Orientierung helfen sollen, welche Aspekte ihren Unterrichts einen möglichen Lernerfolg positiv beeinflussen können (vgl. Helmke 2006: 44–45, Helmke: 2010: 169–170, Meyer 2007b: 170–171)²¹. Der Beitrag, den die Unterrichtsforschung hier leisten kann, liegt darin, im Rahmen ihrer durchaus eingeschränkten Erkenntnismöglichkeiten abgesichertes Wissen darüber zu erzeugen, welche Merkmale empirisch nachweisbar zu einem ‚lernwirksamen‘ Unterricht beitragen.

2.1.3 Konzepte guten Unterrichts

Helmke und Meyer haben auf Basis ihrer jahrelangen Forschung und universitären Lehre zur Lehrerbildung und zum ‚gutem Unterricht‘ Zusammenfassungen erstellt, welche Merkmale schulischer Unterricht aufweisen sollte, damit er als ‚gut‘ zu gelten hat. Beide Autoren kommen zu Katalogen von jeweils zehn Merkmalen, die in Abbildung 2.1 stichpunktartig aufgelistet werden (vgl. Helm-

²⁰ Entsprechend unsinnig ist es pauschal zu behaupten „Frontalunterricht macht klug“, wie es Kloepfer in der FAZ im Dezember 2012 mit Bezug auf eine Metastudie des Münchener IFO Instituts tut (vgl.: Kloepfer 2012). Vgl. zur Kritik an den Hoffnungen die an die empirische Bildungsforschung zur Verbesserung der Bildungsqualität gelegt werden auch Tenorth 2013.

²¹ Siehe zu der Frage welchen Sinn solche ‚Kataloge‘ von Merkmalen ‚guten‘ Unterrichts machen auch das Kapitel 2.1.4.

ke 2010: 168–267, Meyer 2007a: 23–126)²². Unterricht, der sich an diesen Kriterien orientiert, kann im Sinne der beiden Autoren als „guter Unterricht“ (vgl. Meyer 2007b: 167) bzw. Unterricht mit „hoher Güte“ (Helmke 2010: 22) angesehen werden.

Ein solcher Unterricht wird von den Autoren als „gut“ angesehen, da er jene Kriterien bündelt, die nach den Erkenntnissen der empirischen Forschung dafür sorgen, dass Unterricht „lernwirksam“ ist (vgl. Helmke 2010: 168–170, Meyer 2007b: 169), es also ermöglicht, dass die Schüler durch den Besuch des Unterrichts einen Lernfortschritt erreichen²³.

²² Die Merkmale werden an dieser Stelle im einzelnen nicht näher erläutert. Im Abschnitt 2.1.5 dieses Kapitels erfolgt im Rahmen Entwicklung der Bewertungsinstrumente eine kurze Auseinandersetzung mit ausgewählten Merkmalen guten Unterrichts.

²³ An dieser Stelle ist auf Helmkes Angebots-Nutzungs-Modell der Wirkung von Unterricht zu verweisen. Guter Unterricht ist nicht per se lernwirksamer Unterricht, aber guter Unterricht stellt ein Angebot an die Schüler dar, das diese vor dem Hintergrund weiterer Wirkfaktoren (z.B. Lernpotenzial der Schüler, familiäres Umfeld, schulischer und regionaler Kontext des Unterrichts) nutzen können um Lernfortschritte zu erreichen (vgl. Helmke 2010: 71–104). Über den Einfluss eines guten Unterricht auf die Lernerfolge konnte die empirische Unterrichtsforschung bisher keine eindeutigen Erkenntnisse gewinnen. Der Einfluss des Lehrerhandelns auf Lernerfolge variiert bei verschiedenen Studien zwischen 3–45 % wobei sich v.a. erhebliche Unterschiede zwischen einzelnen Schulfächern zeigen (vgl. Meyer 2007b: 171, Meyer 2010a: 171–173; Terhart 2007: 55). In seiner Metaanalyse von mehr als 800 einzelnen Metaanalysen, die auf mehr als 50.000 Einzelstudien beruhen, differenziert Hattie zu der Frage des Einflusses des Lehrers auf den Lernerfolg zehn unterschiedliche Kategorien und weist schwache bis sehr starke Einflüsse des Lehrers in diesen einzelnen Kategorien nach (vgl. Hattie 2009: 108–128).

Abb. 2.1: Kriterien guten Unterrichts nach Helmke und Meyer²⁴

Aspekte bzw. Merkmale guten Unterrichts nach Andreas Helmke	Merkmale guten Unterrichts nach Hilbert Meyer
Klassenführung	Klare Strukturierung des Unterrichts
Klarheit und Strukturiertheit	Hoher Anteil echter Lernzeit
Konsolidierung und Sicherung	Lernförderliches Klima
Aktivierung	Inhaltliche Klarheit
Motivierung	Sinnstiftendes Kommunizieren
Lernförderliches Klima	Methodenvielfalt
Schülerorientierung	Individuelles Fördern
Kompetenzorientierung	Intelligentes Üben
Umgang mit Heterogenität	Transparente Leistungserwartungen
Angebotsvielfalt	Vorbereitete Lernumgebung

Sowohl Meyer als auch Helmke erwähnen aber, dass die Merkmale guten Unterrichts nicht einfach aus der Unterrichtsforschung deduziert werden könnten, sondern immer normativ begründet werden müssten und ein gewisser Eklektizismus bei der Zusammenstellung eines solchen Kataloges nicht zu vermeiden sei (vgl. Helmke 2010: 169, Meyer 2007b: 167–168). Helmke negiert hierbei, dass es eine Bildungstheorie geben könne, mit der sich ein derartiger Katalog begründen ließe. Er verweist vielmehr auf den heuristischen Zweck seines Konzepts, Lehrkräften dabei zu helfen, lernwirksamen Unterricht anzuleiten (vgl. Helmke 2010: 170). Meyer postuliert hingegen, dass sein Katalog durchaus auf einer Bildungstheorie beruhe, ohne diese aber klar auszuarbeiten. Zwar findet sich in seiner Monographie „Was ist guter Unterricht“ ein Kapitel, das den „Theorierah-

²⁴ Meyer spricht hierbei von „Merkmalen guten Unterrichts“, Helmke von „Bereichen, Merkmalen und Prinzipien“ (vgl. Helmke 2010: 168, Meyer 2007a: 23–24). Meyer nimmt darüber hinaus eine Differenzierung zwischen „Merkmalen“ und „Kriterien“ guten Unterrichts vor, wobei erste empirisch erforschten Ausprägungen entsprechen sollen, die Lernwirksamkeit sicher stellen, während die zweiten für die auf Basis einer Bildungstheorie ausgewählten Merkmale stehen, die einen „guten Unterricht“ ausmachen (vgl. Meyer 2007a: 20). Wirklich stringent hält Meyer diese Unterscheidung selber nicht durch, weswegen hier im Bezug auf Meyer vereinfachend von ‚Merkmalen‘ gesprochen wird.

men“ seines Merkmalskataloges bilden soll (vgl. Meyer 2007a: 153–173), dieses enthält aber additiv aneinandergefügt Ausführungen über die Vorgehensweisen der Unterrichtsforschung (vgl. ders.: 154–162) um dann auf drei Seiten mit einer „didaktischen Landkarte“ ein „Ordnungsangebot“ zu machen (vgl. ders.: 162–165). Eine konsistente Bildungstheorie zur Begründung des Merkmalskataloges sieht anders aus²⁵.

Die Kataloge von Helmke und Meyer sind nicht die einzigen Arbeiten zu der Frage, was qualitativ guten, weil lernwirksamen, Unterricht ausmacht. Weitere Beispiele sollen hier nicht vertieft ausgeführt, der Vollständigkeit halber jedoch zumindest erwähnt werden²⁶. Ditton nennt für den deutschen Sprachraum (vgl. Ditton 2006: 240) etwa Modelle von Baumert sowie Klieme, die beide deutlich abstraktere Prinzipien guten Unterrichts formulieren als Helmke und Meyer es tun. Hierbei gibt es teils deutliche Überschneidungen hinsichtlich der als Kernaspekte von gutem Unterricht angesehenen Merkmale, wie etwa einem strukturierenden „Classroom Management“, eine Anpassung an die individuellen Lernvoraussetzungen der Schüler oder die Bedeutung motivationaler Aspekte (vgl. Baumert & Köller 2000, Klieme 2002). Kliemes „Grunddimensionen“ der Qualitätsmerkmale von Unterricht, einer strukturierenden Unterrichts- und Klassenführung, einer individualisierenden Schülerorientierung und kognitiven Aktivierung der Schüler durch anspruchsvolle Aufgaben, wurden im Rahmen der TIMSS Studien Ende der 1990er Jahre entwickelt und beziehen sich vornehmlich auf die Qualität mathematischen Unterrichts (vgl. Klieme 2002: 99, Klieme & Rakoczy 2003: 335–337). Baumert und Köller nutzen ebenfalls die im Rahmen der TIMSS-Studie generierten Daten, um nach erfolgversprechenden „Unterrichtsstrategien“ im Mathematik- und Physikunterricht zu suchen.

²⁵ Siehe zu diesem Punkt auch die Kritik an den Arbeiten von Helmke und v.a. Meyer von Gruschka und Schilmöller, die in Kapitel 2.1.4.

²⁶ Vgl. für weitergehende Literaturhinweise zur Modellen und Konzepten der Unterrichtsqualität: Ditton 2006: 237–241, Helmke et al. 2007: 55–57, sowie das Friedrich Jahresheft 2007, S. 64–65. Die aktuelle Diskussion um den guten Unterricht aus verschiedenen Perspektiven kann z.B. in dem Sammelband von Jürgens und Standop 2010, sowie jenem von Fischer und Schilmöller 2010 nachvollzogen werden.

„Verständnisorientierte Unterrichtsstrategien“ zeigen hierbei systematische Zusammenhänge mit der Erreichung kognitiver und motivationaler Ziele des Mathematikunterrichts (vgl. Baumert & Köller 2000: 305).

Im deutschsprachigen Kontext erwähnenswert sind sicherlich auch die Ergebnisse der SCHOLASTIK-Studie von Helmke und Weinert, sowie des COACTIV-Projektes von Kunter, Baumert u.a.. Ausgehend von Untersuchungen in 54 Grundschulklassen wurde im Rahmen des SCHOLASTIK-Projektes ein Katalog von sechs Merkmalen guten Unterrichts erarbeitet, der deutliche Parallelen sowohl zu Meyers Katalog zeigt, als auch die Grundlage der späteren Weiterentwicklung von Helmkes Katalog bildet (vgl. Helmke & Weinert 1997: 241–251). Die COACTIV-Studie zu kognitiv aktivierendem Mathematikunterricht und der Entwicklung der professionellen Kompetenz von Lehrkräften benennt drei Dimensionen eines qualitativ hochwertigen Unterrichts, nämlich die Effizienz der Klassenführung, das Potenzial zur kognitiven Aktivierung der Schüler und die konstruktive Unterstützung derselben (vgl. Kunter & Voss 2011: 86–90).

Für den internationalen Forschungskontext verweist Helmke (vgl. Helmke et al. 2007: 61, Helmke 2010: 168) auf die Publikationen der International Academy of Education/International Bureau of Education, wobei v.a. die Beiträge von Brophy, sowie von Grouws und Cebulla zum Bereich der Merkmale guten Unterrichts bedeutsam seien (vgl. Brophy 1999, Grouws & Cebulla 2000). Während Brophy hierbei einen Katalog von zwölf Merkmalen eines allgemeinen guten Unterrichts anbietet, der denjenigen von Helmke und Meyer nicht unähnlich ist, erarbeiten Grouws und Cebulla einen Katalog von zehn Merkmalen eines guten *Mathematik*unterrichts. Zu den verbreitetsten Konzepten im angloamerikanischen Sprachraum gehören darüber hinaus die effective teaching methods von Borich (vgl. Helmke 2010: 168). Wiederum Ditton erwähnt im internationalen Kontext das sog. QAIT-Modell von Slavin als weit verbreitetes Modell guten Unterrichts, das seinerseits nur vier relativ abstrakte Kriterien guten Unterrichts formuliert (vgl. Ditton 2006: 240).

Demnach kann als Zwischenfazit dieses Abschnitts zusammengefasst werden, dass es zwar verschiedene Modelle guten Unterrichts gibt, diese aber einen Kern von Merkmalen, bzw. Prinzipien teilen, die empirisch (im Rahmen der Erkenntnisgrenzen der Unterrichtsforschung) abgesichert sind und deren Realisierung einen lernwirksamen Unterricht im Sinne des Angebots-Nutzung-Modells von Helmke wahrscheinlich macht.

Die Entscheidung für die vorliegende Arbeit, die Konzepte von Helmke und Meyer zu nutzen, lässt sich auf verschiedene Art und Weise begründen. *Erstens* sind diese beiden Konzepte guten Unterrichts, v.a. jenes von Meyer, weit verbreitet und gemeinhin bekannt. Sie bilden die Grundlage verschiedener in der Praxis erprobter Bewertungsinstrumente für Unterricht. Helmkes Konzept findet etwa Anwendung wie im Evaluationsprogramm EVIT der Qualitätsagentur für Schulen der Landes Schleswig-Holstein und dem Unterrichtsbeobachtungsbogen ‚Einblick in die Lehr-Lern-Situation‘ der Qualitätsagentur AQS der Landes Rheinland-Pfalz (vgl. Helmke 2010: 169). Meyers Merkmale guten Unterrichts bilden die Grundlage des Unterrichtsbeobachtungsbogens der Qualitätsanalyse des Landes Nordrhein-Westfalen (vgl. Lennartz 2010: 165). Dies dürfte im Sinne des Ziels der Arbeit, ein möglichst konsensfähiges Bewertungskonzept zu entwickeln, förderlich sein. Die Konzepte sind weiterhin mit den gültigen Vorgaben für die Lehrerbildung vereinbar, wie sie z.B. in NRW im „Kerncurriculum für die Ausbildung im Vorbereitungsdienst“ im „Handlungsfeld Unterricht gestalten und Lernprozesse nachhaltig anlegen“ festgehalten sind (vgl. OVP 2011). *Zweitens* sind die Konzepte im Gegensatz zu anderen oben vorgestellten Modellen guten Unterrichts detaillierter ausgearbeitet und mit zahlreichen konkreten Handlungsempfehlungen versehen, welche im Kontext der Ausbildung angehenden Lehrkräfte unerlässlich sind²⁷.

²⁷ Darüber hinaus kann eine erhebliche Ähnlichkeit zwischen der im folgenden entwickelten Matrix der Qualitätsbereiche guten Unterrichts und der von Baumert und Kunter ausgemachten „Kompetenzfacetten“ eines „theoretisch begründeten Tableaus von Lehrerkompetenzen“ postuliert werden (vgl. Baumert & Kunter 2006: 486–489).

Im Folgenden werden die oben zitierten Merkmalskataloge von Helmke und Meyer in einer gemeinsamen Matrix zusammengeführt. Diese Matrix verbindet die in den Arbeiten von Helmke und Meyer aufgeführten Merkmale guten Unterrichts zu gemeinsamen ‚Qualitätsbereichen‘²⁸. Wenn im Folgenden von ‚Qualitätsbereichen‘ die Rede ist, dann ist damit die in der Matrix erstellte Synopse aus Helmkes und Meyers Arbeiten gemeint. Wird hingegen von ‚Merkmalen‘ guten Unterrichts gesprochen, so beziehen sich diese Ausführungen auf einzelne von Meyer, bzw. Helmke ausgemachte Teileaspekte der Qualitätsbereiche, später werden mit ‚Merkmalen‘ auch einzelne ‚Merkmale‘ der umfassenden ‚Qualitätsbereiche‘ der Matrix guten Unterrichts bezeichnet.

Den einzelnen Qualitätsbereichen werden in dieser Matrix kurze Erklärungen angefügt, die auf den Ausführungen von Helmke und Meyer beruhen. Weitergehend wird eine erste Operationalisierung der einzelnen Merkmale der umfassenderen Qualitätsbereiche vorgenommen, sodass ihnen beobachtbare Handlungen oder beobachtbares Verhalten zugeordnet werden können. Vor dem Hintergrund der spezifischen Anforderungen der Ausbildungssituationen der zweiten Phase der Lehrerbildung, wird dann eine Analyse unternommen, welche dieser Qualitätsbereiche guten Unterrichts Grundlage der Bewertungsinstrumente für die Unterrichtsbesuche sein sollen, die dann in eine erste Version des Bewertungskonzepts für das Referendariat münden. Der letzte in der Matrix aufgeführte Qualitätsbereich guten Unterrichts, der Lernzuwachs, findet sich so explizit allerdings weder bei Helmke noch bei Meyer. Zwar sehen beiden Autoren das Ziel guten Unterrichts darin, lernwirksam zu sein, aber sie führen keine Bilanzierung der Lernwirksamkeit des Unterrichts, als eigenes Merkmal auf. Vielmehr gehen sie davon aus, dass Unterricht, der ihren Merkmalskatalogen folgt, prinzipiell lernwirksam sei. Im Sinne eines zielorientierten Unterrichts kann dieser aber nur als ‚gut‘ gelten, wenn am Ende einer Unterrichtseinheit, sei dies eine Unterrichtsstunde oder eine Unterrichtsreihe, ein klar zu benennender Lernfortschritt festzustellen ist. Unterricht ist nur

²⁸ Diese Matrix befindet sich im Anhang.

dann gut, wenn Lernziele erreicht werden²⁹. Entsprechend bezieht die hier vertretene Qualitätsauffassung guten Unterrichts den Lernfortschritt als eigenes, für sich zu beobachtendes Merkmal, mit ein. Die Operationalisierung dieses Kriteriums orientiert sich an den Unterrichtsbeobachtungsbögen der Qualitätsanalyse in Nordrhein-Westfalen.

Die Matrix dient als Grundlage für die folgende Auswahl zentraler Merkmale guten Unterrichts, die Eingang in die Bewertungsinstrumente für die Zweite Phase der Lehrerbildung finden sollen. Die Zusammenfassung der Merkmale guten Unterrichts von Helmke und Meyer in der Matrix erfolgt vor dem Hintergrund der Ähnlichkeit der Merkmale der beiden Autoren. Für die Konstruktion dieser Matrix gelten die gleichen Bedingungen wie für die Kataloge an sich, in Helmkes Worten:

„Jede Klassifikation von Merkmalen der Unterrichtsqualität – auch diese – ist eine individuelle Konstruktion; Anzahl (Herv. i. Org.) und Bezeichnung der Kategorien sind bis zu einem gewissen Grad willkürlich und beliebig. Mit guter Begründung könnte man auch für eine geringere Anzahl von Qualitätsbereichen plädieren (...) oder für eine höhere (...)“ (Helmke 2010: 169).

2.1.4 Kritik an den Konzepten guten Unterrichts

Nachdem die Konzepte guten Unterrichts von Helmke und Meyer einführend dargestellt wurden, sollen kritische Stimmen zu diesen Konzeptionen zu Wort kommen, bevor mit der Entwicklung der Bewertungsinstrumente begonnen wird. Eine pointierte Kritik an Merkmalskatalogen guten Unterrichts wurde v.a. von Gruschka und Schilmöller formuliert (vgl. Gruschka 2007, Schilmöl-

²⁹ Natürlich kann aus lerntheoretischer Perspektive oder aus der Sichtweise einer konstruktivistischen Didaktik Kritik an solch einer unterstellten Zielsetzung von Unterricht geäußert werden. Hier wird jedoch die pragmatische und durch Erfahrungswissen abgesicherte Position vertreten, dass guter Unterricht sehr wohl dazu in der Lage ist Lernziele zu definieren und deren erfolgreiches Erreichen ex post abzu prüfen. Zu einer Kritik an dieser Sichtweise vgl. Kösel 2002, Luhmann 1986, Luhmann & Schorr 1982, Reich 2008, zur Kritik an diesen konstruktivistischen Ansätzen vgl. Terhart 1999.

ler 2006; 2010)³⁰. Beide argumentieren aus der Perspektive einer geisteswissenschaftlichen Pädagogik gegen die vermeintlich reduktionistische Antwort einer empirisch ausgerichteten Erziehungswissenschaft auf die Frage was ‚guten Unterricht‘ ausmache (vgl. Gruschka 2007: 10, Schilmöller 2010: 99).

Schilmöller gliedert seine Kritik in sechs Oberpunkte (vgl. Schilmöller 2006: 78). Meyer listet seinerseits ebenfalls sechs Kritikpunkte auf, in denen die beiden oben genannten Autoren seine Arbeiten zu den Merkmalen guten Unterrichts kritisieren würden (vgl. Meyer 2010b: 32). Die folgende Darstellung orientiert sich lose an diesen Strukturierungen, benennt aber drei wesentliche Oberpunkte, unter die sich die Kritik an Helmke und Meyer unterordnen lässt³¹. Dies sind *erstens wissenschaftstheoretische bzw. paradigmatische Einwände, zweitens der Vorwurf, dass die Konstruktion der Merkmalskataloge keinem sauberen wissenschaftlichen Vorgehen entspreche* und diese deshalb auf ‚*handwerklichen Fehlern*‘ beruhten und *drittens dem Vorwurf eines verzerrten Theorie-Praxis-Verhältnisses*, welchem Helmke und Meyer folgen würden.

Aus einer *wissenschaftstheoretischen bzw. paradigmatischen* Perspektive bemängeln Gruschka und Schilmöller das szientifisch-positivistische Wissenschaftscredo, dem Helmke und Meyer folgen würden. Die Zergliederung eines komplexen Sinnsystems wie Unterricht, das von der Teilhabe verschiedener Menschen geprägt sei, in einzelne Merkmale und deren Operationalisierung durch Indikatoren, werde dem Wesenskern des Unterrichts nicht gerecht, ja können diesen gar

³⁰ Weitere Kritik an den Katalogen von Helmke und Meyer findet sich z.B. bei Becker 2008: 33–36.

³¹ Interessanterweise arbeiten sich sowohl Gruschka als auch Schilmöller v.a. an Meyers Merkmalskatalog ab, verlieren aber kaum explizit kritische Worte über frühe Versionen der Kriterienkataloge von Helmke, wie etwa jenem aus der SCHOLASTIK-Studie abgeleiteten, obwohl dieser jenem von Meyer sehr ähnlich ist. Gruschka setzt sich zwar explizit mit dem Angebots-Nutzungsmodell von Helmke auseinander, seinen Merkmalskatalog spart er in seiner Kritik aber aus. Man mag den Grund für diese Fokussierung auf Meyer mit Reh als „[...] gegenüber dem Erfolg des Buchautors Meyer neidvoll vorgebrachten Kritik [...]“ (Reh 2005) verstehen, kann aber auch die These aufstellen, dass sich Schilmöller und Gruschka v.a. auf Meyers Arbeiten konzentrieren, da dieser sich wie oben bereits angerissen, hinsichtlich der Bedeutung seines Merkmalskataloges deutlich weiter aus dem Fenster lehnt, als Helmke es tut. Während Helmke sich der Frage, was Merkmalskataloge guten Unterrichts nicht können, ausführlich und explizit widmet (vgl. Helmke 2010: 169–171), geht Meyer auf diesen Aspekt immer wieder nur am Rande ein, während er das, was seine Kataloge (vermeintlich) können, sehr pointiert ausführt (vgl. Meyer 2003: 37, Meyer 2007a: 156, Meyer 2007b: 171).

nicht erfassen, denn mit ihrem Drängen auf Objektivität verfehle sie schlechthin den Kernbereich dessen, was einen ‚guten und gelingenden‘ Unterrichte ausmache (Vgl. Gruschka 2007: 10; 18, Schilmöller 2006: 79). Schilmöller grenzt hierbei die „Pädagogik“ als Disziplin explizit von einer „empirischen Erziehungswissenschaft“ ab:

„Der Pädagogik ist das Engagement für den Erfolg und das Gelingen pädagogischer Praxis insofern inhärent. Einer empirisch ausgerichteten, auf Wertneutralität und Distanz bedachten «Erziehungswissenschaft» (Herv. i. Org.) ist dieses erkenntnisleitende Interesse fremd“ (Schilmöller 2010: 99).

In diesem Kontext wirft Gruschka Meyer vor, Forschungsergebnisse zur Frage des guten Unterrichts nur selektiv zu rezipieren und qualitative Studien im Sinne eines rekonstruktiven Wissenschaftsparadigmas systematisch zu ignorieren (vgl. Gruschka 2007: 16). An Helmkes Angebots-Nutzungs-Modell von Unterricht kritisiert Gruschka die vermeintliche Selbstreferentialität der empirischen Unterrichtsforschung. Das Modell sei mehr oder weniger willkürlich konstruiert und beziehe verschiedene Einflussfaktoren des Unterrichts in eklektischer Art und Weise aufeinander (vgl. Gruschka 2007: 28–32). Im Fazit zum Sinn dieses Modells heißt es:

„Das Modell ist vor allem eines, das den Modus einer bestimmten Unterrichtsforschung reflektiert. Auf diese Weise wird die Sache an eine durchgespielte Forschungsmaschine angepasst und die wiederum erlaubt vielfältige neue Studien zu generieren“ (Gruschka 2007: 32).

Nicht das Modell diene also dazu, den Unterricht besser zu verstehen um damit die pädagogische Praxis anzuregen, nein, es diene vielmehr der empirischen Unterrichtsforschung als Zweck an sich, um Forschung über die Forschung zu betreiben³².

³² Aus Sicht der Systemtheorie eine wenig überraschende Erkenntnis: Das autopoetische System der (Erziehungs-)Wissenschaft funktioniert nach einer anderen Funktionslogik, einem anderen binären Code als das Erziehungssystem. Systembestand ist eben immer nur selbstreferentiell möglich, Systembeziehungen nur im Sinne struktureller Kopplung, als ‚Beobachtung‘ nach Maßgabe der jeweiligen Systemstruktur (vgl. Baecker 2004: 41–166).

Den Arbeiten von Helmke und Meyer lassen sich zu diesem Kritikpunkt teils explizite Entgegnungen entnehmen. Meyer führt aus, dass rekonstruktive Studien deutlich weniger Forschungsergebnisse vorzuweisen hätten als empirische-analytische Studiendesigns und er sie deshalb nicht berücksichtigt hätte (vgl. Meyer 2010b: 33). Helmke merkt an, dass eine Zerlegung des Unterrichts in separate Prinzipien und Variationen von diesen für viele „Pädagogen“ ungewohnt sei, allerdings sei sie ebenso unumgänglich, „[...] um ein differenziertes Bild des Unterrichts, seiner Eigenheiten, Stärken und Schwächen zu gewinnen“ (Helmke 2010: 170). Was sich hier zeigt, ist ein Disput verschiedener wissenschaftstheoretischer Strömungen innerhalb der Erziehungswissenschaft, der sich nicht auflösen lässt. Vielmehr ist diese paradigmatische Vielfalt als Wesensmerkmal der Erziehungswissenschaft wie auch anderer Sozialwissenschaften auszumachen, der hinzunehmen und produktiv zu gestalten ist (vgl. Teller 2008, Terhart 2012).

Eine Ebene unterhalb dieses paradigmatischen ‚Schulstreits‘ der Erziehungswissenschaftler – oder ‚Pädagogen‘? – angesiedelt sind die Vorwürfe von Gruschka und Schilmöller, Helmke und v.a. Meyer hätten *wissenschaftlich unsaubere Arbeit* geleistet. Die Zusammenstellung der Merkmale in den Katalogen sei beliebig, ebenso die inhaltliche Füllung der einzelnen Merkmale. Das methodische Vorgehen beim Entwerfen der Kataloge verbleibe im Dunkeln (vgl. Gruschka 2007: 16–28, Schilmöller 2006: 78–79):

„Offensichtlich können Merkmale also entfallen oder sind austauschbar, so dass der Eindruck der Beliebigkeit und Willkür entsteht“ (Schilmöller 2006: 78).

Auch die von Meyer angegebenen „Konstruktionsregeln“ (vgl. Meyer 2007a: 18) bei der Formulierung der Merkmale würden nicht weiterhelfen, da diese weder wissenschaftlichen Kriterien standhalten würden, noch den Konstruktionsvorgang tatsächlich transparent machten. Vielmehr sei ein Prozess, dessen Gütekriterium danach bemessen ist, auf „[...] gründlichen Absprachen mit Theoretikern und Praktikern [...]“ (Meyer, zit. n. Gruschka 2007: 16) zu beruhen, höchst willkürlich. Weiterhin kritisieren Gruschka und Schilmöller eine fehlende theoretische Fundierung der Merkmalskataloge. Allein aus den Ergebnissen der

empirischen Unterrichtsforschung über Zusammenhänge zwischen Merkmalen einzelner Unterrichtsstunden und der Lernwirksamkeit dieser Stunden ließe sich keine Begründung generieren, warum diese Merkmale jetzt guten Unterricht ausmachen würden (vgl. Gruschka: 2007: 12–14). Schilmöller wirft Meyer einen naturalistischen Fehlschluss vor, wenn er aus dem empirischen Nachweis über Wirkungszusammenhänge eine Qualitätsdefinition von Unterricht mache:

„Das Sollen ist aus dem Sein nicht ableitbar. Auf den vorliegenden Fall angewendet heißt das, dass unterrichtliche Qualitätskriterien nicht einfach und umstandslos aus Ergebnissen empirischer Forschung ableitbar sind. Dass sich etwas erfolgreich bewerkstelligen lässt, macht es noch nicht wünschenswert, ist kein ausreichender Grund, es anzustreben“ (Schilmöller 2010: 77).

Vielmehr sei eine bildungstheoretische Fundierung der Konzeption guten Unterrichts notwendig, welche die gefundenen Merkmale, die Lernwirksamkeit bedingen, in einen didaktischen Begründungszusammenhang einordne (vgl. Schilmöller 2006: 79, Gruschka 2007: 14). Schilmöller führt mit Rückbezug auf Pleines eine Differenzierung zwischen „erfolgreichem“ und „gelingendem“ Unterricht ein. Erfolgreicher Unterricht sei jener, der im Sinne von Helmke und Meyer lernwirksam hinsichtlich kognitiver Kompetenzfortschritte der Schüler sei. Diese Bedingung sei aber eben nicht erschöpfend, um von ‚gelingendem‘ Unterricht zu sprechen, denn über die Aufgabe der Wissensvermittlung hinaus sei guter Unterricht eben auch immer „erziehender Unterricht“³³, durch den der Bildungsauftrag der Schule, die Schüler zur Mündigkeit heranzuführen, erst erreicht werden könne. Diese theoretische Fundierung im besten Sinne der didaktischen Tradition fehle den Merkmalskatalogen. Gruschka geht so weit, Meyer die „[...] Substitution von Theorie durch didaktische Spielerei [...]“ vorzuwerfen (Gruschka 2007: 19). Darüber hinaus bezweifeln sowohl Gruschka als auch Schilmöller, dass

³³ Vgl. zum Konzept des erziehenden Unterrichts Benner 2001: 231–285. Schilmöller entwickelt in Schilmöller 2010 ein eigenes Konzept eines „gelingenden erziehenden Fachunterrichtes“. In diesem Konzept vereinigt er empirische Ergebnisse der SCHOLASTIK-Studie von Weinert und Helmke mit didaktischen Überlegungen in herbartscher Tradition (vgl. Schilmöller 2010: 79–96).

die zur Verfügung stehenden Daten aus der empirischen Unterrichtsforschung es überhaupt zulassen würden, verlässliche Aussagen über Ursache-Wirkungszusammenhänge zwischen bestimmten Unterrichtsmerkmalen und Lernwirksamkeit zu belegen. Viele der Daten würden nur Zusammenhänge für einzelne Fächer oder Schulformen belegen, aber nicht für allgemeine Konzepte guten Unterrichts, wie Helmke und Meyer sie postulieren. So würden v.a. Ergebnisse über Wirkzusammenhänge aus dem Mathematikunterricht herangezogen, die so z.B. in keiner Weise auf den Deutschunterricht zu übertragen seien (vgl. Gruschka 2007: 19, Schilmöller 2006: 79–80)³⁴.

Auch hinsichtlich dieser Kritikpunkte finden sich Kommentare von Helmke und Meyer in ihren Arbeiten wieder. Auf die Frage, ob Qualitätsmerkmale aus empirischen Forschungsergebnissen zu deduzieren seien, entgegnet Meyer, dass diese Ableitung natürlich immer normativ sei (Meyer 2007b: 167). Er erkennt die Grenzen der empirischen Unterrichtsforschung also durchaus an und verweist auch darauf, dass eine konsistente Theorie für deren systematische Herleitung in der Tat fehle (vgl. Meyer 2010b: 33). Zur kritisierten Deduktion von Qualitätsmerkmalen von Unterricht aus der empirischen Forschung äußert sich Helmke, zumindest bei den dieser Arbeit zu Grunde gelegten Aufsätzen und Monographien, nur an wenigen Stellen, führt aber aus, dass die Konstruktion von Merkmalskatalogen immer zu einem guten Stück willkürlich und beliebig sei und in diesem Sinne keine reine Deduktion aus den Forschungsergebnissen, sondern vielmehr eine pragmatische Zusammenstellung von Aspekten der Lernwirksamkeit (vgl. Helmke 2010: 170). Die begrenzte Aussagekraft der Daten wird sowohl von Helmke als auch von Meyer wiederholt angesprochen (vgl. Helmke et al. 2007: 61–64, Meyer 2007a: 155–162).

Ein dritter Oberpunkt der Kritik an den Protagonisten der Merkmalskataloge guten Unterrichts bezieht sich auf das *Verhältnis zwischen Theorie über Unterricht und der Umsetzung in konkreten Unterricht*. In der ersten Veröffentlichung seines

³⁴ Die Kritik von Schilmöller wird allerdings fragwürdig, wenn er bei der Konzeption seines „gelingenden erziehenden Fachunterrichts“ Merkmale ‚erfolgreichen‘ Unterrichts aus der SCHOLASTIK-Studie ableitet und konstatiert, dass diese Merkmale eine „erwiesene Bedeutung für den Leistungszuwachs von Schülern“ hätten. Seine eigene Kritik holt ihn hier ein (vgl. Schilmöller 2010: 85).

Merkmalskataloges unterstellt Meyer einen direkten Zusammenhang zwischen der Realisierung der von ihm ausgemachten Merkmale guten Unterrichts und dem Lernerfolg von Schülern.

„These 1: Wer dafür sorgt, dass die zehn Merkmale des Katalogs in seinem Unterricht stark ausgeprägt sind, macht guten Unterricht und verhilft dadurch seinen Schülerinnen und Schülern zu dauerhaft hohen Lernerfolgen“ (Meyer 2003: 37).

Schilmöller kritisiert diese Verdinglichung des Unterrichts als „Technik“ als unrealistisch in Bezug auf die pädagogische Praxis. Unterricht sei keine „Technik“ im engen Sinne, denn dieser würde unterstellen, dass es ein fixes Repertoire an Unterrichtstechniken gebe, die ein Lehrer zu lernen zu habe, um dann dauerhaft guten Unterricht anbieten zu können (vgl. Schilmöller 2006: 70–75). Dies sei aber keineswegs so. Vielmehr entfalte sich guter Unterricht immer nur unter dem Wirken schulischer Kontextfaktoren in einem institutionellen Rahmen, die jeweils unterschiedliche Anforderungen an guten Unterricht bedingen würden.

„«Die Schule» (Herv. i. Org.), so das Fazit an dieser Stelle, gibt es so wenig wie «den Unterricht» (Herv. i. Org.). Es gibt immer nur Konzepte von Unterricht und Konzepte von Schule. Eine perspektivlose Sicht ist nicht möglich. Also muss die Perspektive auch berücksichtigt und thematisiert werden“ (Schilmöller 2010: 78).

Meyer würde demnach eine Verallgemeinerung seiner Qualitätskriterien postulieren, die nicht einzulösen sei und verschleierte darüber hinaus das ihnen zu Grunde gelegte Qualitätsverständnis. Gruschka bezieht diesen Argumentationsstrang auf das Angebot-Nutzungs-Modell des Unterrichts von Helmke. Es sei fraglich, inwieweit man Unterricht als reines ‚Angebot‘ des Lehrers modellieren könne. Gehöre nicht vielmehr der einzelne Schüler zum Unterricht dazu, denn ohne Schüler kein Unterricht? Gleiches gelte für die von Helmke durchaus erwähnten Einflussfaktoren des Unterricht, etwa des ‚Klassenkontexts‘ oder des ‚fachlichen Kontexts‘. Gruschka fragt, was der Unterricht denn nun sei: Eine unabhängige, durch den Lehrer und die Umsetzung der Qualitätsmerkmale be-

dingte Variable, oder eine vom Kontext abhängige Variable (vgl. Gruschka 2007: 28–32)? So oder so, das Modell lege sich teleologisch, auf den „[...] motivierten, interessierten Schüler [...]“ (Gruschka 2007: 33) als Adressaten des Unterrichtsangebots des Lehrers fest. Und was sei, wenn die Schüler das Angebot des Lehrers nicht nutzen wollten? Seien sie dann selber schuld, sei der Unterricht trotzdem gut? Schlussendlich lande man bei einer Tautologie, mit der dem Praktiker, der Unterricht in einem konkreten Kontext anleiten soll, wenig gedient sei:

„Guter Unterricht ist, wenn die Wirkfaktoren für guten Unterricht zutreffen; wenn die Wirkungen zutreffen, ist guter Unterricht“ (Gruschka 2007: 35).

Helmkes Entgegnung auf diese Kritik führt letzten Endes zum oben ausgeführten ersten Oberpunkt der Kritik zurück, den der unterschiedlichen paradigmatischen Sichtweisen der Wissenschaftler auf die Frage, was guten Unterricht ausmache. Helmke wehrt sich gegen die Unterstellung „Unterrichtsrezepte“ anbieten zu wollen, es handele sich vielmehr um „Steuerungswissen“ über Unterricht (vgl. Helmke 2010: 170). Die Kenntnisse über die Wirkungsweise der Merkmale guten Unterrichts sei „[...] eine hilfreiche und förderliche, aber für sich genommen noch nicht hinreichende Voraussetzung für erfolgreichen Unterricht [...]“ (Helmke 2010: 170). Die Merkmale dienen also als Wissen, mit dem man als Praktiker Unterricht in einem gewissen Rahmen ‚steuern‘ und in diesem Sinne die Wahrscheinlichkeit der Lernwirksamkeit erhöhen kann. Man könnte hier mit Schilmöller auf die Diskussion um das von Luhmann und Schorr ausgemachte Technologiedefizit der Pädagogik verweisen, dass es eben im Bereich der Erziehung und Bildung keine Technik gebe, die eine mechanische Erfüllung des Erziehungs- und Bildungsauftrags ermögliche und demnach der Anspruch der ‚Steuerung‘ an sich verfehlt sei (Vgl. Schilmöller: 2006: 72). Man könnte darüber hinaus die Diskussion führen, ob ein auf ‚Mündigkeit‘ zielender, ‚erziehender Unterricht‘ es nicht auch moralisch verbiete, auch nur eine ‚Technologieersatztechnologie‘ zu nutzen bzw. ob dies mit dem Ziel der Mündigkeit nicht unvereinbar wäre (vgl. Teller 2009)? Man könnte – und wäre damit wieder bei bei dem Schulstreit zwischen einer positivistischen-szientifisch und einer geisteswissen-

schaftlich-hermeneutisch orientierten Erziehungswissenschaft – oder Pädagogik? – angekommen.

Zieht man aus der Kritik in Bezug auf das hier bearbeitete Anliegen, ein praxistaugliches Bewertungsinstrument für die Zweite Phase der Lehrerbildung zu entwickeln, das den Referendaren im Sinne formativer Evaluation hilft ihre eigene Lehrerpersönlichkeit zu entwickeln, ein Fazit, kann der oben zitierte Anspruch von Meyer erneut herangezogen werden.

„Wer dafür sorgt, dass die zehn Merkmale des Katalogs in seinem Unterricht stark ausgeprägt sind, macht guten Unterricht und verhilft dadurch seinen Schülerinnen und Schülern zu dauerhaft hohen Lernerfolgen“ (Meyer 2003: 37).

Geht Meyer zu weit, wenn er dies postuliert? Wahrscheinlich! Ist sein Merkmalskatalog, und damit verbunden jener von Helmke deswegen untauglich, um als Referenzmaßstab für die Bewertung von Referendaren im Sinne formativer Evaluation herangezogen werden? Nein! Denn Referendare müssen sich im Rahmen der Zweiten Phase der Lehrerbildung mit einer sehr komplexen neuen Aufgabe zurechtfinden, für deren Bewältigung sie Orientierung benötigen. Für diesen Zweck sind Merkmalskataloge, wie Helmke und Meyer sie anbieten, in einem heuristischen Sinne geeignet. Tiefgehende didaktische und bildungstheoretische Überlegungen, wie Schilmöller und Gruschka sie anstellen, sind wichtig, aber von den alltäglichen Anforderungen des Referendarsdaseins weit entfernt. Die Entwicklung von Routinen der Unterrichtsplanung und -durchführung im Rahmen bewährter Praktiken bildet einen wichtigen, vielleicht den wichtigsten Fixpunkt der Ausbildung. Es geht, um mit Meyer zu sprechen, um die Aneignung des „[...] Handwerkszeugs des Unterrichtens [...]“ (Meyer 2010b. 34), denn ohne dieses ist ein Referendar wie auch später ein Lehrer im Berufsalltag verloren. Wie bereits oben ausgeführt, ist die Beherrschung dieses Handwerkszeugs weder hinreichend, um ein guter Lehrer zu sein, noch um guten Unterricht anzuleiten, aber ohne ein erlerntes Repertoire an Praktiken, um Unterricht mit dem Ziel einer gewissen Lernwirksamkeit zu planen und durchzuführen, wird die Realisierung von gutem Unterricht nahezu unmöglich, ob man diesen nun als ‚gelingenden‘,

‚erziehenden‘ oder wir auch immer gearteten Unterricht titulierte. Die Arbeiten von Helmke und Meyer bieten hierfür einen möglichen Bezugspunkt unter vielen; es lassen sich aber gute Gründe anführen, warum die Wahl auf sie fiel. Die Arbeit nimmt in diesem Sinne auch den Impuls von Meyer auf, solche Modelle guten Unterrichts, wie er und Helmke sie entwickelt haben, aufzugreifen und im Sinne der Aktionsforschung zu nutzen, um konkrete Herausforderungen der pädagogischen Praxis zu bewältigen (vgl. Meyer 2010b: 35).

2.1.5 Relevante Merkmale guten Unterrichts für die Ausbildung in der zweiten Phase der Lehrerbildung

Es wäre sicherlich wünschenswert, dass jeder Referendar im Rahmen der Ausbildung lernt, alle aufgeführten Merkmale im eigenen Unterricht umzusetzen und im Anschluss an die Ausbildung realisieren zu können. Realistisch gesehen ist dieser Anspruch allerdings zu hoch gegriffen. Gegen eine solche Auffassung spricht erstens das breite Spektrum, in dem sich verschiedene Lehrerpersönlichkeiten ausbilden, die jeweils bestimmte der in dem Katalog aufgeführten Merkmale besonders gut, andere dafür weniger gut werden umsetzen können. Als *erstes Kriterium* für die Auswahl einzelner Merkmale aus der Matrix guten Unterrichts für einen Beobachtungs- und Bewertungsbogen kann folglich gelten, dass *nur wirklich grundlegende Merkmale ausgewählt werden sollten, die so essentiell für die Realisierung guten Unterrichts sind, dass jeder Lehrer die Fähigkeit beherrschen sollte, sie zu realisieren.*

Weitergehend muss die besondere Form des Unterrichts, eben die der Unterrichtsbesuche, berücksichtigt werden, die Grundlage der Bewertungsprozesse in der Zweiten Phase der Lehrerbildung ist. Die Ausbilder besuchen die Auszubildenden in einzelnen Unterrichtsstunden und bewerten primär den in dieser Stunde gesehenen Unterricht³⁵. Einzelne der in dem Katalog aufgeführten Merkmale lassen sich aber nur schwer an einzelnen, beobachtbaren Indikato-

³⁵ Zwar bezieht sich das Urteil auch auf weitere Kompetenzen als die in der unmittelbaren Stunde gezeigten, wie etwa die Fähigkeit eine sinnvolle Unterrichtsreihe zu planen, die ihren Ausdruck in der schriftlichen Planung der Referendare finden soll, aber der Fokus der Bewertung liegt klar auf der einzelnen Unterrichtsstunde.

ren fest machen, sondern zeigen sich erst über einen längeren Unterrichtszeitraum. Hier wäre z.B. der sinnvolle Einsatz von Ritualen zur Strukturierung des Unterrichts zu nennen. Rituale in Lerngruppen einzuführen und zu einer wirklichen Selbstverständlichkeit werden zu lassen, sodass sie ihren Sinn als strukturierende Rituale erfüllen, benötigt ein nicht unerheblich Zeitspanne. Eine Beobachtung ein- und derselben Lerngruppe über solche lange Zeiträume ist im Rahmen der Bewertungsprozesse des Vorbereitungsdienstes aber nicht möglich. Entsprechend lässt sich als *zweites Auswahlkriterium festhalten, dass jene Merkmale im Rahmen der Bewertungsprozesse Berücksichtigung finden sollten, die in einer einzelnen Unterrichtsstunde auf Basis operationalisierter Indikatoren beobachtbar sind, damit eine an Gütekriterien³⁶ orientierte Feststellung von Kompetenzen möglich ist.*

Schlussendlich muss bei der Auswahl der Merkmale die generelle Ausbildungssituation Berücksichtigung finden, wie sie in Kapitel 1.2 umrissen wurde. Hierbei ist insbesondere der bisweilen schwierigen und von Widersprüchen geprägten Beziehung zwischen Ausbildern und Auszubildenden Rechnung zu tragen. Diese legt nahe, jene Merkmale guten Unterrichts, die eng mit Persönlichkeitsmerkmalen, wie z.B. Humor, gekoppelt sind nicht in ein standardisiertes Bewertungsinstrument aufzunehmen³⁷. Vielmehr wird als *drittes Auswahlkriterium herangezogen, dass die mit den ausgewählten Merkmalen guten Unterrichts in Verbindung stehenden Fähigkeiten im Rahmen einer 18–24 monatigen Ausbildung zumindest partiell erlernbar sein müssen.* Dies schließt Merkmale aus, die mit einer u.U. erheblichen Persönlichkeitsveränderung verbunden sind.

³⁶ Siehe zu den relevanten Gütekriterien das Kapitel 2.2.2.

³⁷ Dies soll nicht heißen, dass es im Rahmen der Ausbildung u.U. geboten sein kann, Auszubildende auf die notwendige Entwicklung bestimmter Persönlichkeitsmerkmale hinzuweisen. Allerdings ist dies zwingend mit einer beratenden Rolle des Ausbilders verbunden und sollte nicht Grundlage von Bewertungsprozessen sein. Die Herausforderungen und Probleme die für die Ausbilder aus diesem sensiblen Bereichen der Professionalisierung entstehen, werden ausführlich in der Auswertung der Gruppendiskussionen diskutiert (siehe Kapitel 4).

Zusammenfassend können als Kriterien für die Auswahl einzelner Merkmale aus der allg. Matrix von Merkmalen guten Unterrichts die drei folgenden herangezogen werden:

- Die mit dem Merkmal guten Unterrichts in Verbindung stehenden Kompetenzen müssen so grundlegende Aspekte guten Unterrichts betreffen, dass jeder Lehrer sie beherrschen sollte.
- Die Realisierung der Merkmale guten Unterrichts müssen im Rahmen einer einzelnen Unterrichtsstunde auf Basis klar operationalisierter Anweisungen beobachtbar sein.
- Die mit den Merkmalen guten Unterrichts in Verbindung stehenden Kompetenzen müssen im Rahmen einer 18–24 monatigen Ausbildung zumindest partiell erlernbar sein, ohne dass dies einer substantiellen Veränderung der (Lehrer-)Persönlichkeit bedarf.

Die Festlegung auf ausgewählte Merkmale aus der Matrix der Qualitätsbereiche guten Unterrichts bildet den *ersten Schritt zur Entwicklung der pre-Alpha-Version des Bewertungskonzepts*. Die Auswahl dieser Merkmale kann nur gut begründet erfolgen. Gleichwohl verbleibt bei der Auswahl immer auch eine normative Komponente, warum bestimmte Merkmale als essentiell wichtig, andere hingegen als weniger wichtig eingestuft werden. Diese normative Gewichtung hat u.a. mit der Bevorzugung eines bestimmten Unterrichtsstils zu tun, von der sich niemand frei machen kann. Dieser normative Faktor ist bei der Entwicklung eines praxistauglichen Bewertungsinstruments nicht zu verhindern. Um diesen Prozess jedoch mit einer größtmöglichen Transparenz zu versehen, wird die Auswahl der einzelnen Merkmale vor dem Hintergrund der Anforderungen der Ausbildung, wie sie oben umrissen wurde, begründet. Das besondere Setting zur der Auswahl einzelner Kriterien aus der Matrix guten Unterrichts reagiert genau auf diesen normativen Faktor: Indem die pre-Alpha-Version des Bewertungskonzepts und die ihr zu Grunde liegende Merkmalsauswahl in einer Gruppendiskussion thematisiert wird, soll eine konsensfähige Auswahl der Kriterien erfolgen, die auch

im Kontext einer gesamten Ausbildungsinstitution und darüber hinaus Zustimmung erhalten kann³⁸.

Die erste Version des Beobachtungsbogens für die Unterrichtsbesuche umfasst 23 Einzelmerkmale, die zehn der in der Matrix guten Unterrichts aufgeführten Qualitätsbereiche zu zuordnen sind. Die Auswahl dieser Kompetenzen sowie die der Auswahl zu Grunde liegenden Überlegungen sollen hier vorgestellt werden. Der Sinn von vielen der im Folgenden zu findenden operationalisierten Merkmale guten Unterrichts wird erst klar, wenn die Formulierungen der verschiedenen möglichen Kompetenzstufen hinzugezogen werden, wie sie sich aus der pre-Alpha-Version des Beobachtungsbogens ablesen lassen, der sich im Anhang finden lässt.

2.1.5.1 Klarheit und Strukturiertheit

Den größten Block von Einzelmerkmalen im Beobachtungsbogen für Unterrichtsbesuche nimmt der Qualitätsbereich *Klarheit und Strukturiertheit* ein. Dieser umfasst je zwei ‚Merkmale‘ guten Unterrichts nach Helmke (Klassenführung und Klarheit und Strukturiertheit, vgl. Helmke 2010: 172–200) und Meyer (klare Strukturierung und inhaltliche Klarheit, vgl. Meyer 2007a: 25–38; 55–66). In diesem Qualitätsbereich des Katalogs sind Aspekte guten Unterrichts zusammengefasst, welche einerseits den möglichst reibungslosen Ablauf des Unterrichts gewährleisten und andererseits dafür sorgen sollen, dass er inhaltlich möglichst ‚gehaltvoll‘ ist. Meyer folgend ist ‚Klarheit‘ ein zentrales Kriterium dieses Qualitätsbereiches, es geht um ‚klar‘ gestellte Aufgaben, eine ‚klar‘ auszumachende Rhythmisierung der Unterrichtsschritte oder um ‚klare‘ Regeln, die den Unterricht strukturieren. Eine stimmige Verbindung von Unterrichtszielen, -methoden und -inhalten bildet zusammen mit einer durchdachten Auswahl der Inhalte die Grundlage für erfolgreiche Lernprozesse. Helmke und Meyer zitieren zahlreiche empirische Studien, welche die große Bedeutung einer klaren Strukturierung für den Erfolg von Unterricht belegen (vgl. Helmke 2010: 174–175; , Meyer 2007a: 32–35).

³⁸ Siehe zur Rolle der Gruppendiskussion im Forschungsdesign der Arbeit Kapitel 3.2.1.

Einzelne Merkmale dieses Bereiches lassen sich im Rahmen von Unterrichtsbesuchen nur sehr eingeschränkt beobachten, bzw. bewerten. Wie in Kapitel 1.2 ausgeführt handelt es sich bei den Lehrproben um künstliche Lehrsituationen, die sich vom Unterrichtsalltag massiv unterscheiden, sodass z.B. die Wirkung von Ritualen für die Strukturierung einer Unterrichtsstunde nicht wirklich überprüft werden kann. Erstens werden sich die Schüler auf Grund der Tatsache, dass ‚ihr Lehrer‘ eine Prüfung hat, anders verhalten als im sonstigen Unterricht, sodass es fragwürdig ist, ob in der Lehrprobe gezeigte Rituale tatsächliche Routine in einer Lerngruppe sind oder nur die sprichwörtliche ‚Show‘, die in der Lehrprobe abgehalten wird. Weiterhin absolvieren die Referendare einen erheblichen Teil ihrer Unterrichtsbesuche nicht in eigenen Klassen oder Kursen, sondern verweilen in diesen Kursen im Rahmen des Ausbildungsunterrichts nur eine gewisse Zeit als ‚Gastlehrer‘. Der Einbau von Ritualen in den Unterricht wird unter dem Vorzeichen dieser Bedingungen kaum möglich sein.

Basierend auf den oben formulierten Auswahlkriterien für einzelne Merkmale guten Unterrichts wurden für diesen Qualitätsbereich fünf für guten Unterricht essentielle Merkmale ausgewählt. Dies sind die folgenden Merkmale eins bis fünf des Beobachtungsbogens³⁹:

- 1) Es ist eine transparente und stimmige Phasierung des Unterrichts erkennbar, einzelne Unterrichtsschritte werden klar markiert.
- 2) Die Schüler verstehen die Aufgabenstellungen und können erläutern was sie gerade tun.
- 3) Unterrichtsziele, Inhalte und Methoden stehen in einem sinnvollen Verhältnis zueinander.
- 4) Die Inhalte der Stunden werden didaktisch sinnvoll inszeniert und knüpfen an zentrale fachdidaktische Prinzipien an.

³⁹ Zur Frage wie diese einzelnen Merkmale ‚erfasst‘ werden, siehe Kapitel 2.2.2.

- 5) Die Äußerungen des Lehrers zeichnen sich durch akustische Verständlichkeit und sprachliche Klarheit aus.

Merkmal eins operationalisiert die Fähigkeit der Referendare, einer Unterrichtsstunde eine klare Struktur zu geben und für einen ‚runden‘ Ablauf der Stunde zu sorgen, dem die Schüler folgen können. Merkmal zwei erfasst die Fähigkeit, Klarheit über zu bearbeitende Inhalte und Lernziele herzustellen. Die Merkmale drei und vier operationalisieren eine sinnvolle (fach-)didaktische- und methodische Inszenierung, welche den Schülern vor dem Hintergrund ihrer Entwicklung und ihres Kenntnisstandes gerecht wird. Das Merkmal fünf greift einen Aspekt auf, der nahezu alle anderen Merkmale guten Unterrichts beeinflusst: Ist die Lehrersprache nicht akustisch verständlich und inhaltlich klar, werden nahezu alle anderen Bemühungen, guten Unterricht zu realisieren, konterkariert.

Einzelne Aspekte, die in der Matrix guten Unterrichts im Qualitätsbereich Klarheit und Strukturierung aufgeführt werden, tauchen in anderen Qualitätsbereichen wieder auf und werden dort operationalisiert. So finden sich z.B. die Merkmale ‚Klarheit und Verbindlichkeit der Ergebnissicherung‘ später im Bereich des Lernzuwachses wieder und werden auch an dieser Stelle in den Beobachtungsbogen aufgenommen.

2.1.5.2 Hoher Anteil echter Lernzeit

Der Qualitätsbereich des *hohen Anteils echter Lernzeit* taucht in dieser Formulierung nur bei Meyer auf (vgl. Meyer 2007a: 39–46). Seine Ausführungen werden hier ergänzt durch einige Überlegungen, die Helmke im Merkmal ‚Klassenführung‘ verortet (vgl. Helmke 2010: 184–190). Unterricht mit einem hohen Anteil ‚echter‘ Lernzeit wird im hier im Sinne Meyers verstanden als ein Unterricht, in dem ein Großteil der Zeit für die Arbeit an den geplanten Inhalten und Aufgaben verwendet wird und nicht für Disziplinierungen von Unterrichtsstörungen, nachträglichen Aufgabenerklärungen oder -veränderungen (vgl. Meyer 2007a: 39). Im Rahmen von Unterrichtsbesuchen lässt sich dieses Merkmal guten Unterrichts klar erfassen und bewerten, da Planungskorrekturen und notwendige Nachbesserungen bei Aufgabenstellungen für alle Beteiligten ohne weiteres zu

beobachten sind. Die Merkmale sechs bis sieben des Beobachtungsbogens sollen diesen Qualitätsbereich erfassen:

- 6) Die meisten Schüler sind nicht gelangweilt oder überfordert.
- 7) Die Stunde weist einen hohen Anteil ‚echter Lernzeit‘ auf.

Merkmal sechs erfasst die Fähigkeit des Lehrers, Unterricht so zu gestalten, dass er die Schüler aktiviert, sich mit den angebotenen Inhalten zu beschäftigen und sie so zu erfolgreichen Lernprozessen anzuregen. Dieses Merkmal steht in enger Verbindung zum Merkmal elf, das die Fähigkeit, auf heterogene Lerngruppen zu reagieren erfassen, soll. Die Operationalisierung des Merkmals sieben erscheint hier zunächst als Leerformel. Erst die genaue Operationalisierung der einzelnen Kompetenzstufen, wie sie sich im Beobachtungsbogen im Anhang finden lassen, macht die Intention dieses Merkmals, das Verhältnis zwischen der für ‚Lernen‘ und ‚lernfremde Dinge‘ aufgewendeten Zeiten zu erfassen, wirklich deutlich.

Weitere Aspekte dieses Qualitätsbereichs, die von Meyer und Helmke angeführt werden, wie etwa das Entstehen reichhaltiger Arbeitsergebnisse oder zur Verfügung stehende Lernhilfen, werden in weiteren Qualitätsbereichen (‚Lernzuwachs‘, bzw. ‚Umgang mit Heterogenität‘) aufgegriffen und operationalisiert.

2.1.5.3 Kompetenzorientierung

Der Begriff der Kompetenz steht nicht nur im konzeptionellen Zentrum dieser Arbeit, sondern stellt auch einen zentralen Bezugspunkt des ‚post-PISA Unterrichts‘ dar. Wie in der Einleitung ausgeführt wurde, war eine der maßgeblichen Reaktionen auf das schlechte Abschneiden der deutschen Schüler in internationalen Schulvergleichsstudien die Einführung von bundesweit geltenden Bildungsstandards für die Unterrichtsfächer. Im Sinne Helmkes ist ein kompetenzorientierter Unterricht derjenige, „[...] der bewusst und geplant die Förderung der in den Bildungsstandards formulierten Kompetenzen zum Gegenstand hat“ (Helmke 2010: 234). Die in der Einleitung umrissene ‚standardbasierte Bildungsreform‘

ist ein weitreichendes Reformprojekt, das bis in die Planung und Gestaltung der einzelnen Unterrichtsstunden hineinreicht, sodass auch im Rahmen von einzelnen Unterrichtsstunden und -reihen eine deutliche Orientierung an den in den Bildungsstandards festgeschriebenen Kompetenzen zu finden sein sollte⁴⁰. Das so nur bei Helmke zu findende Prinzip guten Unterrichts wird in den Beobachtungsbögen operationalisiert durch die Merkmale acht und neun:

- 8) Die Aufgaben zeigen eine klare Berücksichtigung der Kompetenzorientierung.
- 9) Im Kontext der Unterrichtsstunde/-reihe ist eine klare Progression über die Anforderungsbereiche zu erkennen.

2.1.5.4 Aktivierung

Für eine erfolgreiche Umsetzung der im vorherigen Abschnitt angesprochenen Kompetenzorientierung ist der Einbezug des Qualitätsbereichs der Aktivierung nahezu unerlässlich. Kompetenzorientierung, z.B. im Erreichen von ‚Handlungskompetenz‘ in verschiedenen fachlichen Kontexten, bedingt, dass die Schüler aktive Gestalter der Lernprozesse werden und ihnen die Möglichkeit gegeben wird, sich aus der passiven Rolle von reinen Rezipienten von Unterricht zu emanzipieren. Für Helmke, der dieses Prinzip guten Unterrichts ausarbeitet, gehört zur Aktivierung v.a. die Einbindung der Schüler in den Lernprozess. Dies beginnt bei dem Einsatz kooperativer Arbeitsformen und dem Einbezug der Schüler in die Planung des Unterrichts und geht bis zu selbstgesteuerten und eigenverantwortlichen Lernformen. Der (ernsthafte) Einbezug von Schülern in die Planung von Unterrichtsreihen und -stunden wird im Rahmen von Unterrichtsbesuchen während des Vorbereitungsdienstes nur in Ausnahmefällen zu beobachten sein. Die Gründe dafür werden später für den Qualitätsbereich des sinnstiftenden Kommunizierens ausgeführt. Der gewinnbringende Einsatz von Elementen kooperativen Arbeitens und selbstgesteuerten Lernens

⁴⁰ Dies ‚sollte‘ ist so natürlich eine normative Feststellung. Zur Kritik an dem Kompetenzbegriff und der standardbasierten Bildungsreform siehe Kapitel 2.1.4 und 2.2.4.

sollte inzwischen zum Grundrepertoire von Lehrern gehören. Deswegen wird der Qualitätsbereich der Aktivierung über das Merkmal 10 wie folgt operationalisiert:

- 10) Kooperative oder selbstgesteuerte Arbeitsformen werden gewinnbringend eingesetzt.

2.1.5.5 Umgang mit Heterogenität

Der Umgang mit heterogenen Lerngruppen gewinnt im Schulalltag, nicht zuletzt auf Grund der Forderung nach einem inklusiven Schulsystem, immer mehr an Bedeutung. Der Qualitätsbereich *Umgang mit Heterogenität* orientiert sich in seiner Bezeichnung an Helmke (vgl. Helmke 2010: 244–259) und integriert Aspekte, die Meyer unter dem Titel des „individuellen Förderns“ aufgreift (vgl. Meyer 2007a: 86–103⁴¹). Auf Grund der Tatsache, dass jede Lerngruppe hinsichtlich der Fähigkeiten und Lernausgangsvoraussetzungen ihrer Mitglieder heterogen zusammengesetzt ist und nur das Ausmaß der Heterogenität in Frage steht, kann die Fähigkeit, mit dieser Heterogenität umzugehen als eine essentielle Kompetenz eines jeden Lehrers angesehen werden. Die Einschätzung, wie gut angehende Lehrer auf die unterschiedlichen Lernvoraussetzungen und Fähigkeiten ihrer Schüler reagieren, kann in einzelnen Unterrichtsbesuchen dadurch überprüft werden, ob sie bei der Gestaltung der Aufgaben oder des Lernmaterials auf die unterschiedlichen Ausgangsbedingungen ihrer Schüler reagieren. Merkmal 11 operationalisiert diesen Qualitätsbereich demnach wie folgt:

- 11) Eine Binnendifferenzierung reagiert sowohl auf die Bedürfnisse schwächerer als auch leistungsstärkerer Schüler.

2.1.5.6 Lernförderliches Klima

Eine ‚gute Atmosphäre‘ im Unterricht herzustellen ist ein Kriterium, dass über eine reine Output-Orientierung des Unterrichts hinausgeht. Die Gestaltung der Schule als einem ‚menschlichen Lebensraum‘, in dem sich die Beteiligten

⁴¹ Streng genommen handelt es sich dabei um die Ausführungen von Ela Eckert, denn sie hat das betreffende Kapitel in Meyers Buch geschrieben.

wohlfühlen oder zumindest nicht unwohl, ist eine essentielle Voraussetzung für eine positive Lernumgebung, in der guter Unterricht realisiert werden kann (vgl. Meyer 2007a: 53). Zwar merken sowohl Helmke als auch Meyer an, dass die empirischen Belege bezüglich der Bedeutsamkeit dieses Kriteriums für erfolgreiches Lernen in der Schule nicht so deutlich ausfallen wie für andere der bei ihnen aufgeführten Merkmale, aber beide betonen die wichtige Rolle, die ein ‚lernförderliches Klimas‘ an sich für Bildungsinstitutionen spielt und nehmen dieses Kriterium unter gleicher Bezeichnung in ihre Merkmalslisten guten Unterrichts auf (vgl. Helmke 2010: 220–221, Meyer 2007a: 49–53). Da der Unterricht den größten Teil der Zeit ausmacht, den sowohl Schüler als auch Lehrer in der Schule verbringen, bildet ein positives, lernförderliches Klima das Fundament einer positiven Atmosphäre in der gesamten Schule. Alle Lehrer sollten dazu in der Lage sein, dieses zu fördern. Aspekte eines lernförderlichen Klimas lassen sich Helmke und Meyer folgend u.a. an einer Atmosphäre ausmachen, die von gegenseitigem Respekt sowohl unter Schülern als auch in der Schüler-Lehrerbeziehung gekennzeichnet ist (vgl. Helmke 2010: 226–227). Weiterhin betont Meyer die Bedeutung einer gemeinsam von Schülern und Lehrern geteilten Verantwortung für eine positive Lernatmosphäre und für den Lernerfolg (vgl. Meyer 2007a: 48).

Die Auswahl von Merkmalen aus diesem Qualitätsbereich gestaltet sich schwieriger als jene aus den bisher aufgeführten Bereichen, da die Bedeutung von individuellen Persönlichkeitsmerkmalen der angehenden Lehrer in diesem Bereich eine größere Rolle spielt als in den vorherigen. So betonen sowohl Helmke als auch Meyer die Bedeutung des ‚richtigen‘ Grads von Humor auf Seiten des Lehrers für die Realisierung eines lernförderlichen Klimas (vgl. Helmke 2010: 225, Meyer 2007a: 49). Die Aufnahme eines solchen Aspektes in einen standardisierten Bewertungsbogen verbieten allerdings mindestens zwei Aspekte: Erstens ist es fraglich, ob es überhaupt möglich ist, Menschen im Rahmen einer Ausbildung dazu anzuleiten humorvoller zu sein, und zweitens muss die ethische Frage in den Raum gestellt werden, ob diese Absicht nicht einer illegitimen, versteckten Überwältigung der Auszubildenden gleichkommt. Demnach wurden lediglich jene Merkmale guten Unterrichts dieses Qualitätsbereichs in den Beobachtungsbogen übernommen, die zu realisieren sind

ohne eine Veränderung von substantiellen Persönlichkeitsmerkmalen anstreben zu müssen. Der Bereich wurde auf Basis von zwei Merkmalen operationalisiert.

12) Der Lehrer reagiert angemessen auf Unterrichtsstörungen.

13) Ein großer Teil der Lerngruppe bringt sich aktiv in den Unterricht ein.

Das Merkmal zwölf operationalisiert die Verantwortung des Lehrers, für einen möglichst reibungslosen Ablauf des Unterrichts zu sorgen, ohne dabei die notwendige Wertschätzung für die Schüler vermissen zu lassen. Die Verantwortung der Schülerseite für ein lernförderliches Klima wird durch das Merkmal 13 operationalisiert.

Einzelne Aspekte dieses Qualitätsbereiches, wie der von Helmke angemerkte „konstruktive Umgang mit Fehlern“ (vgl. Helmke 2010: 22–223), finden ihren Niederschlag in anderen Qualitätsbereichen. Eine konstruktive ‚Fehlerkultur‘ wird etwa im Bereich der transparenten Leistungserwartungen aufgegriffen.

2.1.5.7 Sinnstiftendes Kommunizieren

Kommunikation findet in jedem Unterricht statt und zwar durchgängig. Es ist aber die Frage, zwischen wem diese Kommunikation stattfindet und welche Inhalte die Kommunikation hat, schlussendlich: mit welchem ‚Sinn‘ die stattfindende Kommunikation aufgeladen wird. Mit ‚Sinn‘ ist in diesem Qualitätsbereich gemeint, dass die Schüler den behandelten Inhalten einen persönlichen Sinn geben können, dass sie die Bedeutsamkeit der behandelten Inhalte und zu erlernenden Kompetenzen für ihr eigenes Leben als ‚sinnvoll‘ erachten. Meyer führt aus, dass dieser Sinnstiftungsprozess durch Metakommunikation über den Unterricht und die dort behandelten Inhalte angeregt werden kann, ebenso wie durch eine Feedbackkultur, die sowohl Schüler-Lehrer-Feedback als auch Schüler-Schüler-Feedback umfasse (vgl. Meyer 2007a: 68–73). Helmke erwähnt die Bedeutung des Feedbacks in seinem Prinzip der Schülerorientierung, wo er ebenfalls die Rolle der Schülerbeteiligung bei der Auswahl von Unterrichtsinhalten und -methoden, für die Motivation der Schüler betont (vgl. Helmke 2010: 232–233), die Meyer ebenfalls in seinem Merkmal des sinnstiftenden Kommunizierens

verortet. Demnach verbindet der Qualitätsbereich sinnstiftendes Kommunizieren der Matrix das Prinzip des sinnstiftenden Kommunizierens von Meyer und der Schülerorientierung von Helmke. Meyer begründet die Bedeutsamkeit des Merkmals damit, dass eine erhöhte Sinnggebung durch Beteiligung der Schüler bei der Planung, metakommunikative Methoden und eine Feedbackkultur eine erhöhte Lernmotivation und fachliche und überfachliche Interessenbildung bewirke. Beide Autoren führen wenige empirisch abgesicherte Erkenntnisse für die Bedeutsamkeit dieses Merkmals guten Unterrichts an (vgl. Helmke 2010: 230–233, Meyer 2007a: 69–70). Empirische Belege für den Zusammenhang zu höherer Lernmotivation bleibt Meyer schuldig und liefert lediglich Hinweise darauf, dass bestimmte Unterrichtsmethoden die Interessenbildung bei Schülern fördern können, wobei unklar bleibt, welchen Bezug diese Aussage zu den von ihm zuvor aufgeführten Einzelaspekten des Merkmals des sinnstiftenden Kommunizierens aufweist.

Der Qualitätsbereich im Beobachtungsbogen greift die Ausführungen von Helmke und Meyer im Gegensatz zu den bisher operationalisierten Merkmalen nur lose auf und formuliert eine eigenständige Definition von sinnstiftendem Kommunizieren. Hier soll der Prozess der individuellen Sinnggebung der Inhalte durch die Schüler im Mittelpunkt stehen, der bei Meyer und v.a. bei Helmke nur implizit behandelt wird. Der Qualitätsbereich wird durch die Merkmale 14–16 operationalisiert:

- 14) Die Schüler werden dazu angeregt, die Unterrichtsinhalte in ihren eigenen Sinnhorizont einzuordnen.
- 15) Das Gelernte wird mit anderen Unterrichtsinhalten verknüpft.
- 16) Die Schüler geben sich untereinander Rückmeldung über Inhalte oder Lernfortschritte.

Das Merkmal 14 erfasst, inwieweit der vom Referendar inszenierte Unterricht dazu geeignet ist, den Schülern individuelle Sinnggebungsprozesse zu ermögli-

chen. Die Formulierung der verschiedenen Kompetenzstufen dieses Merkmals⁴² machen deutlich, dass hierbei die Verbindung von abstrakten Konzepten mit praktischen Erfahrungen eine Rolle spielt. Eine ‚systemische Unterrichtsarchitektur‘ hat das Merkmal 15 im Blick, da es bewertet, inwieweit der Unterricht auf bereits erarbeiteten Lernergebnissen aufbaut oder einen Transfer der Lernergebnisse ermöglicht. Der Bedeutung von Schüler-Schüler-Feedback für erfolgreiche Lernprozesse, sowie für die Verantwortungsübernahme auf Seiten der Schüler wird durch Merkmal 16 Rechnung getragen.

Auf Grund der bereits mehrfach ausgeführten Besonderheiten der Unterrichtsbesuche (künstliche Unterrichtssituation, eingeschränkte Zeit der Referendare in der Lerngruppe, ...) finden einige der von Helmke und Meyer angemerkten Aspekte keinen Niederschlag in der Operationalisierung dieses Qualitätsbereichs. Dies betrifft u.a. die Planungsbeteiligung von Schülern. Es mag zwar in einzelnen Fällen vorkommen, dass Referendare Unterrichtsstunden zeigen, in denen sie die Schüler an der Planung der weiteren Unterrichtsschritte beteiligen, aber diese Stunden werden vor dem Hintergrund der oben genannten Kontexte der Unterrichtsbesuche selten vorkommen und keine ‚typische‘ Lehrprobenstunde darstellen. Die Planbarkeit von Unterrichtseinheiten und erst recht von Unterrichtsstunden nimmt erheblich ab, wenn man den Schülern die Planungskompetenz in die Hand gibt. Dadurch kann ein Effekt entstehen, der konträr zu den Anforderungen von Unterrichtsbesuchen steht: an einem im Vorfeld festgelegten Termin eine bestimmte Unterrichtsstunde zu zeigen.

2.1.5.8 Transparente Leistungserwartungen

Erwartungen an den Unterricht und an die am Unterricht teilnehmenden Akteure bilden eine wesentliche Determinante von erfolgreich realisierten Lernanlässen. Hierbei spielen die Erwartungen, die der Lehrer an die Leistung der Schüler hat, eine große Rolle, sie stellen Lernanlässe dar und geben den Schülern Orientierung hinsichtlich der Frage, ‚was‘ sie lernen sollen, wohin ein erfolg-

⁴² Siehe den Beobachtungsbogen mit Kompetenzskala im Anhang.

reicher Lernweg sie bringen soll⁴³. Die transparente Kommunikation über die Lernziele des Unterrichts und die Frage, wie bewertet wird, bilden bei Meyer das Fundament dieses Merkmals (vgl. Meyer 2007a: 114). Darüber hinaus greift Meyer bei diesem Merkmal Aspekte auf, die weiter oben bereits in dem Qualitätsbereich Klarheit und Strukturiertheit in die Operationalisierung eingegangen sind, nämlich die Berücksichtigung der gültigen Bildungsstandards, den Stand der (fach-)didaktischen Diskussion und die Lernvoraussetzungen der Schüler bei der Gestaltung der Lernangebote. Helmke führt einen Teilaspekt der transparenten Leistungserwartungen in seinem Prinzip des lernförderlichen Klimas auf, nämlich die Frage nach einem konstruktiven Umgang mit Fehlern (vgl. Helmke 2010: 222–225). Hinsichtlich der empirischen Belege für die Bedeutung dieses Merkmals führt Meyer an, dass es eine positive Korrelation zwischen „hohen Leistungserwartungen“ der Lehrer und tatsächlich erbrachten Leistungen der Schüler gebe (vgl. Meyer 2007a: 117 –118). Helmke verweist auf die Forschungen von Weinert hinsichtlich der Bedeutung einer konstruktiven Fehlerkultur für erfolgreiche Lernprozesse (vgl. Helmke 2010: 22–223).

In der Matrix guten Unterrichts finden diese Überlegungen von Helmke und Meyer Niederschlag in drei Merkmalen, die v.a. die Überlegungen von Helmke zur Fehlerkultur und die von Meyer zur Zieltransparenz aufgreifen. Andere Teilaspekte dieses Qualitätsbereichs wurden, wie bereits erwähnt, schon im Qualitätsbereich Klarheit und Strukturiertheit in Einzelmerkmalen operationalisiert. Die Merkmale 17–19 des Beobachtungsbogens sollen dazu dienen, die Ausprägung der Zieltransparenz und die Realisierung einer positiven Fehlerkultur zu erfassen:

17) Der Lehrer gibt deutliches Feedback, ob die Beiträge richtig waren.

⁴³ Auch an dieser Stelle ließe sich aus der Perspektive einer konstruktivistisch argumentierenden Didaktik Kritik an der hier vorgetragenen Sichtweise formulieren, dass Unterricht den Schülern v.a. die Möglichkeiten zur Entwicklung eigener ‚Sinnwelten‘ ermöglichen sollte. Hier sei noch einmal auf die bereits in Anmerkung 29 vorgetragene Position verwiesen, dass in dieser Arbeit die Auffassung vertreten wird, dass es Aufgabe des schulischen Unterrichts ist, Welt- und Orientierungswissen zu vermitteln und dass dieser ‚Vermittlungsweg‘ in einem pragmatischen Sinne möglich ist.

18) Der Lehrer unterstützt die Schüler, Fehler selbstständig zu beheben.

19) Dem Lehrer gelingt es Klarheit über die Lernziele herzustellen.

Durch diese Operationalisierungen nicht erfasst werden die Bemühungen des Lehrers, die Leistungen transparent zu bewerten und mit „gerechten“ Noten zu versehen, wie Meyer es als einen Teilaspekt des Merkmals aufführt. Diese Aspekte wurden auch in diesem Fall auf Grund der besonderen Ausbildungssituation außen vor gelassen, stehen aber keinesfalls für die Geringschätzung einer transparenten Leistungsbewertung. Die Ausbildungssituation einer singulären Stunde liefert aber im Regelfall keine systematischen Hinweise darauf, wie transparent die formale Leistungsbewertung durch den Lehrer erfolgt, da dies wiederum Beobachtungen der Lerngruppe über einen längeren Zeitraum erfordern würde, wie sie im Vorbereitungsdienst nicht vorgesehen sind.

2.1.5.9 Vorbereitete Umgebung

Der Qualitätsbereich der *vorbereiteten Umgebung* knüpft an die Überlegungen von Meyer zu dem von ihm ebenfalls so betitelten Merkmals guten Unterrichts an. Helmke lässt Ausführungen zu diesem Aspekt des Unterrichts vermissen. Meyer orientiert sich bei der Operationalisierung dieses Merkmals an der Qualität der Räumlichkeiten, in denen das Lernen realisiert werden soll. Er grenzt den Begriff der ‚Lernumgebung‘ so auf die tatsächliche räumliche Umgebung ein und grenzt sich explizit von (konstruktivistischen) Didaktikern ab, die den Begriff der ‚Lernumgebung‘ weiter fassen und darunter u.a. vorbereitetes Lernmaterial, etc. verstehen. Für ihn geht es um die Realisierung ordentlicher Unterrichtsräume mit einer funktionalen Ausstattung, die von den Schülern und Lehrern als ‚ihre‘ Lernumgebung wahrgenommen werden (vgl. Mayer 2007: 120 –123). Die empirischen Belege für die Bedeutung dieses in Meyers Sinne operationalisierten Merkmals sind sehr dünn, um nicht zu sagen de facto nicht existent. Meyer führt diesen Mangel an empirischen Belegen selbst auf, „behauptet“ aber, dass ein positiver Effekt ansprechender Räumlichkeiten auf Lernprozesse zu erwarten sei (vgl. Meyer 2007a: 121). Die hier gewählte Operationalisierung des Qualitätsbe-

reichs der vorbereiteten Umgebung erweitert den Begriff der Lernumgebung im Vergleich zur definitorischen Engführung von Meyer. Entgegen der Positionierung von Meyer werden zur ‚Lernumgebung‘ hier auch die eingesetzten Medien, Arbeitsmittel und Materialien gezählt. Dies geschieht, weil in den Unterrichtsbesuchen der Referendare relativ oft selbst gestaltete Unterrichtsmaterialien (Arbeitsblätter, Stationenlernen, etc.) zum Einsatz kommen, die zur ‚Lernumgebung im weiteren Sinne‘ gezählt werden können und die Auswahl oder der Entwurf dieser Materialien einen wesentlichen Teil der späteren Bewertung ausmacht. Auf die räumlichen Gegebenheiten der Ausbildungsschule haben die Referendare nur sehr beschränkten Einfluss. Sie können lediglich dafür sorgen, dass die Unterrichtsräume sauber und aufgeräumt sind und dass die Raumgestaltung der ‚Logik‘ der gezeigten Unterrichtsstunde entspricht. Entsprechend werden auch nur die zuletzt genannten Bereich der räumlichen Gestaltung bei der Operationalisierung berücksichtigt. Diese Überlegungen lassen sich in der Operationalisierung der Merkmale 20 und 21 des Beobachtungsbogens zusammenfassen:

- 20) Die Arbeitsmittel und Medien werden zielführend eingesetzt und sorgen für Verständnis
- 21) Die Sitzordnung und die Raumgestaltung stehen in einem sinnvollen Verhältnis zu Zielen, Inhalten und Methoden des Unterrichts

2.1.5.10 Lernzuwachs

Der Qualitätsbereich des Lernzuwachses geht über die Arbeiten von Helmke und Meyer hinaus. Mehr implizit als explizit konzentrieren sich beide Autoren, v.a. Meyer, bei ihren Ausführungen stark auf die Prozessdimension des Unterrichts und unterstellen quasi, dass ein Unterricht, der den von ihnen entwickelten Leitlinien folgt, Lernerfolge, also positiven ‚Output‘, ermögliche. Bei Meyer tauchen feststellbare Lernerfolge in seiner Definition, was guten Unterricht ausmacht, noch auf. Hier bezieht er sich explizit auf den Kompetenzfortschritt der Schüler als einem Kriterium eines guten Unterrichts, aber ein eigenständiges Merkmal, das es zu erfassen gelte, wird dem Lernzuwachs nicht zugedacht (vgl. Meyer 2007a:

13; 23–127). Helmke widmet dem Lernzuwachs in dem bereits oben genannten Prinzip der Kompetenzorientierung einigen Raum, konzentriert sich aber auf längerfristige Prozesse des Kompetenzerwerbs, die nicht den Kontextbedingungen der Unterrichtsbesuche in der zweiten Phase der Lehrerbildung entsprechen (vgl. Helmke 2010: 234–244). Entsprechend der starken Gewichtung einer einzelnen Unterrichtsstunde bei den Bewertungsprozessen im Vorbereitungsdienst, greift die Operationalisierung dieses Qualitätsbereichs auf Überlegungen der Qualitätsanalyse von Schulen in NRW und den dort genutzten Beobachtungsbogen für einzelne Stunden zurück⁴⁴. Der Qualitätsbereich eines erfolgreichen Lernzuwachses wird mit den Merkmalen 22 und 23 operationalisiert:

- 22) Am Ende der Stunde steht ein klar zu benennender Erkenntnisfortschritt der verbindlich fixiert wird, oder die Schüler formulieren selbstständig Aufgabenlösungen und stellen diese vor.
- 23) Die Schüler können den erzielten Erkenntnisfortschritt für weitere Lernprozesse nutzen.

2.1.5.11 Fehlende Merkmale guten Unterrichts

Die Matrix guten Unterrichts umfasst drei weitere Qualitätsbereiche, die nicht als eigenständige Qualitätsbereiche in den Beobachtungsbögen berücksichtigt wurden. Dies sind i.E.:

- Methodenvielfalt
- Konsolidierung/Sicherung
- Motivierung

Der Bereich der Methodenvielfalt findet seinen Niederschlag in einem anderen Instrument, das weiter unten vorgestellt wird, der sog. ‚Sammlung über das Re-

⁴⁴ Die entsprechenden Beobachtungsbögen für den Unterricht sind zu finden unter: http://www.schulministerium.nrw.de/QA/Tableau/X_Downloads/Unterrichtsbeobachtungsbogen.pdf (abgerufen am 16.08.2012).

pertoire⁴⁵. Dem Merkmal der Konsolidierung und Sicherung von Helmke (vgl. Helmke 2010: 200–204) wird im Qualitätsbereich des Lernzuwachses Rechnung getragen, wenn dort die verbindliche Fixierung eines Erkenntnisfortschritts am Ende einer Unterrichtsstunde verlangt wird. Die von ihm in diesem Merkmal aufgeführten weitergehenden Aspekte des Übens, die sich so auch in Meyers Merkmal des intelligenten Übens wiederfinden (vgl. Meyer 2007a: 104–112), haben keinen Eingang in die Beobachtungsbögen gefunden, da Unterrichtsstunden, die einen Schwerpunkt auf das wiederholende Üben legen, im Rahmen von Unterrichtsbesuchen sehr selten gezeigt werden und sie somit in einem allgemeinen nutzbaren Bewertungsinstrument keinen Niederschlag finden sollten. Das ebenfalls von Helmke formulierte Merkmal der Motivierung bezieht sich auf eine motivierende Gestaltung des Unterricht, die lebensweltliche Erfahrungen der Schüler aufgreift oder intrinsische und extrinsische Motivationskanäle ausnutzt, um die Schüler zur Mitarbeit zu bewegen (vgl. Helmke 2010: 214–220). Als eigener Qualitätsbereich hat dieses Merkmal ebenfalls keinen Eingang in den Beobachtungsbogen gefunden, aber mittelbar schlägt sich in den Operationalisierungen der Qualitätsbereiche des sinnstiftenden Kommunizierens (Anknüpfung des Unterrichts an die Lebenswelt der Schüler) und der Klarheit und Strukturierung (Berücksichtigung von didaktischen Prinzipien, wie z.B. der Schüler- oder Problemorientierung) nieder.

2.2 Kompetenzbegriffe, Kompetenzmodelle und Erfassung von Kompetenzen

Der Begriff der ‚Kompetenz‘ ist ein zentrales Konzept der in der Einleitung umrissenen, standardbasierten Bildungsreform und bildet den Dreh- und Angelpunkt der mit diesem Forschungsprojekt verbundenen Zielsetzung, ein Bewertungskonzept zu entwickeln, das eine transparente Diagnose der Fähigkeit angehenden Lehrer, zu unterrichten, ermöglichen soll. Die *Unterrichtskompetenz* kann im

⁴⁵ Siehe dazu Kapitel 2.3.3.

Alltagsverständnis des Begriffes verkürzt als *Fähigkeit zu unterrichten* übersetzt werden. Allerdings bedingt eine wissenschaftlich fundierte Auseinandersetzung mit dem Begriff der Unterrichtskompetenz, der dem hier zu entwickelnden Bewertungskonzept zu Grunde liegt, differenziertere Überlegungen hinsichtlich der Definition des Begriffes, der Einzeldimensionen einer umfassenden Unterrichtskompetenz, der Erfassung bzw. Messung einer solchen Kompetenz sowie der Einbettung dieser Kompetenz in den größeren Rahmen der Lehrerbildung im Speziellen und den Trends im Bildungssystem im Allgemeinen. Diese grundlegenden Eigenschaften des Begriffes der Unterrichtskompetenz werden in diesem Kapitel vor dem Hintergrund des spezifischen Anforderungskontextes, in dem sich die Bewertungsprozesse der Zweiten Phase der Lehrerbildung abspielen, erörtert und um eine kritische Auseinandersetzung mit dem Für und Wider der Kompetenzorientierung im Bildungswesen und in der Lehrerbildung ergänzt. Aus diesen Ausführungen leitet sich dann der *zweite Schritt zur Entwicklung der pre-Alpha-Version* des Bewertungskonzepts ab, nämlich die Formulierung eines heuristischen Kompetenzmodells der Unterrichtskompetenz sowie von Leitlinien zur Erfassung derselben.

2.2.1 Kompetenzbegriff und Kompetenzmodelle

Der Begriff der Kompetenz ist kein neuer Begriff, weder in der Alltagssprache noch in den Erziehungs- und Sozialwissenschaften, er ist, wie Klieme und Hartig schreiben, allerdings ein „Modebegriff“, dessen Nutzung in den Sozial- und Erziehungswissenschaften gewissen Konjunkturzyklen unterliege. Mit den Reformtendenzen im deutschen Bildungssystem, die in der Einleitung unter dem Stichwort der ‚standardbasierten Bildungsreform‘, verhandelt wurden, habe der Begriff der Kompetenz einen konjunkturellen Aufschwung, wenn nicht sogar Boom erlebt (vgl. Klieme & Hartig 2007: 11–13). Der Begriff der Kompetenz ist in den Sozial- und Erziehungswissenschaften keineswegs einheitlich definiert, vielmehr gibt es verschiedene Traditionen sprachwissenschaftlicher, psychologischer und auch erziehungswissenschaftlicher Definitionen (vgl. dies.: 14–20, Frey 2006: 31), die unter dem Begriff der Kompetenz keineswegs

immer deckungsgleiche Konzepte verstehen⁴⁶. Diese Arbeit schließt sich einer Kompetenzdefinition an, die im Rahmen der aktuellen deutschsprachigen erziehungswissenschaftlichen Forschung einen breiten Konsens beanspruchen kann. Klieme u.a. definieren in Ihrer „Expertise zur Entwicklung nationaler Bildungsstandards“ Kompetenz als

„[...] die bei Individuen verfügbaren oder von ihnen erlernbaren kognitiven Fähigkeiten und Fertigkeiten, bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Klieme et al. 2003: 72).

Ein erster wichtiger Aspekt dieser Definition für das hier bearbeitete Vorhaben liegt in der Annahme, dass Kompetenzen grundlegend erlernbar sind und eben

„[...] nicht angeborene Talente oder Persönlichkeitsstrukturen, sondern abgrenzbare Fähigkeiten und Kenntnisse, die ihrerseits im Rahmen geeigneter Lernprozesse – wie der Lehramtsausbildung – vermittelt und vertieft werden können“ (Kunter: 2010: 307)⁴⁷.

Weitergehend erhält die Feststellung, dass Kompetenzen eben ‚mehr‘ als Wissen sind, nämlich eben die Fähigkeit, Problemlösungen in verschiedenen Kontexten erfolgreich einzusetzen, eine hohe Bedeutung. Unterrichtskompetenz zeigt sich nicht nur durch Wissen über unterrichtliche Zusammenhänge, sondern in der erfolgreichen Anwendung dieses Wissens. Ein dritter Aspekt dieser Definition wird, wie weiter unten ausführlicher dargelegt⁴⁸, in dem hier entwickelten Bewertungskonzept auf Grund der Beschränkungen des Anwendungskontextes bestenfalls implizit aufgegriffen, nämlich die Annahme, dass Kompetenzen im-

⁴⁶ Vgl. zu vertiefenden Ausführungen über den Kompetenzbegriff in den Sozial- und insb. den Erziehungswissenschaften: Frey 2006, Frey & Jung 2011, Klieme 2004, Klieme & Hartig 2007, sowie die weitere dort angegebene Literatur.

⁴⁷ Vgl. zur Diskussion dieser Annahme, sowie empirischen Ergebnissen: Kunter et al. 2011, sowie Abel & Faust 2010.

⁴⁸ Siehe Kapitel 2.2.3.

mer durch motivationale und volitionale Faktoren beeinflusst werden und nicht bloß durch kognitiv-rationale Aspekte.

Kompetenzen stehen in enger Verbindung mit *Standards* für die Gestaltung von Lernprozessen. Diese formulieren Ziele, die einem hohen Ausprägungsgrad zu erlernender Kompetenzen entsprechen, und zwar auch im hier diskutierten Kontext der beruflichen Kompetenz von Lehrern. Bezugsstandards für die hier bearbeitete Thematik sind die bereits zitierten Standards für die Lehrerbildung (KMK 2004a) sowie die in den Ausbildungs- und Prüfungsordnungen (vgl. z.B. OVP 2011) festgehaltenen Standards, die im Rahmen der Ausbildungsprozesse zu erreichen sind.⁴⁹

Ausgehend von dieser Definition des Kompetenzbegriffes werden in *Kompetenzmodellen* die Teilkompetenzen einer allgemeinen Kompetenz sowie verschiedene Stufen der Kompetenzerreichung näher definiert. Diese Modelle nehmen nach Abs die Struktur von Kompetenzen an sich in den Blick und dienen als Grundlage, Kompetenzen und deren Ausprägungen in ein Messmodell überführen zu können (vgl. Abs 2007: 73). Für den Bereich der beruflichen Handlungskompetenz von Lehrern existieren verschiedene Modellierungen, für die stellvertretend hier jene von Frey vorgestellt werden soll, weil das im Folgenden zu entwickelnde Bewertungskonzept auf diesem Modell aufbaut⁵⁰. Frey unterscheidet verschiedene Teilkompetenzen einer allgemeinen „Handlungskompetenz“, die notwendig sind, um letztere zu realisieren. Diese Überlegungen gehen über den hier diskutierten Kontext des Berufsbildes des Lehrers zunächst hinaus, beschreiben sie doch eine allgemeine Systematik „beruflicher Handlungskompetenz“, werden aber später in Bezug auf das Berufsbild von Lehrpersonen kon-

⁴⁹ Vgl. zur Verbindung von Standards und Kompetenzen z.B. Baumert & Kunter 2006: 478–479, Oser & Renold 2005, Terhart 2006. Für andere Standardmodelle als die KMK-Standards Baumert & Kunter 2006: 478–479, Frey & Jung 2011: 554–560.

⁵⁰ Vgl. zu anderen Modellen der beruflichen Handlungskompetenz von Lehrern einfürend: Frey & Jung 2011: 544–554, Terhart 2007: 45–48. Für eine ausführliche Darstellung von Überlegungen zur Kompetenzmodellierung, auch vor dem Hintergrund des englischsprachigen Forschungsstandes: Baumert & Kunter 2006: 478–505.

kretisiert⁵¹. Frey führt in diesem Kontext vier *Kompetenzklassen* auf, welche in ihrer Summe die Handlungskompetenz bestimmen (vgl. Frey 2006: 33; Frey & Jung 2011: 551–553). Die *Fachkompetenzklasse* umfasst berufsdomänenspezifische Fähigkeiten, ohne die eine bestimmte berufliche Tätigkeit nicht erfolgreich ausgeführt werden kann. Diese werden ergänzt durch die Klasse der *Methodenkompetenz*, in der jene Fähigkeiten subsumiert sind, welche eine Person benötigt, um in einem spezifischen Kontext denk- und handlungsfähig zu sein, wie etwa über Handlungsvorgänge zu reflektieren und diese ggf. zu modifizieren. Fähigkeiten der Konfliktlösung und Kommunikation in Zusammenarbeit mit anderen Personen werden durch die *Sozialkompetenzklasse* zusammengefasst, wohingegen primär nach innen gewandte Aspekte des selbstverantwortlichen Handelns und der dauerhaften Motivation mit der *Personalkompetenzklasse* umschrieben werden. Die Unterscheidung dieser Kompetenzklassen ist im hier diskutierten Kontext wichtig, da ein Instrument, das eine umfassende Diagnostik der beruflichen Handlungskompetenz einer Person vornehmen möchte, diese verschiedenen Kompetenzklassen in einem Bewertungsprozess berücksichtigen oder zumindest Transparenz darüber schaffen muss, welche dieser Kompetenzklassen in der Diagnostik Berücksichtigung finden und welche nicht. Diese Kompetenzklassen bilden eine zweite Hierarchieebene unter einer allgemeinen beruflichen Handlungskompetenz in Freys „hierarchischem Strukturmodell der Handlungskompetenz“ (vgl. Frey & Jung 2011: 551–552). Den Kompetenzklassen ordnet Frey weitergehend „Fähigkeitsdimensionen“ zu und konkretisiert diese auf einer untersten Ebene durch konkrete „Fertigkeiten“. Gerade diese Modellierung verschiedener Hierarchieebenen einer allgemeinen beruflichen Handlungskompetenz bietet sich als Blaupause für das hier zu entwickelnde Modell von ‚Unterrichtskompetenz‘ an, wie im Folgenden gezeigt wird. Abbildung 2.2 zeigt das so entstehende Modell einer beruflichen Handlungskompetenz von Lehrpersonen. Freys Modell der beruflichen Handlungskompetenz von Lehrern geht von einer allgemeinen beruflichen Kompetenz der Lehrer aus, ohne die verschiedenen beruflichen Anforderungen, denen sich Lehrer stellen müssen, wie sie etwa

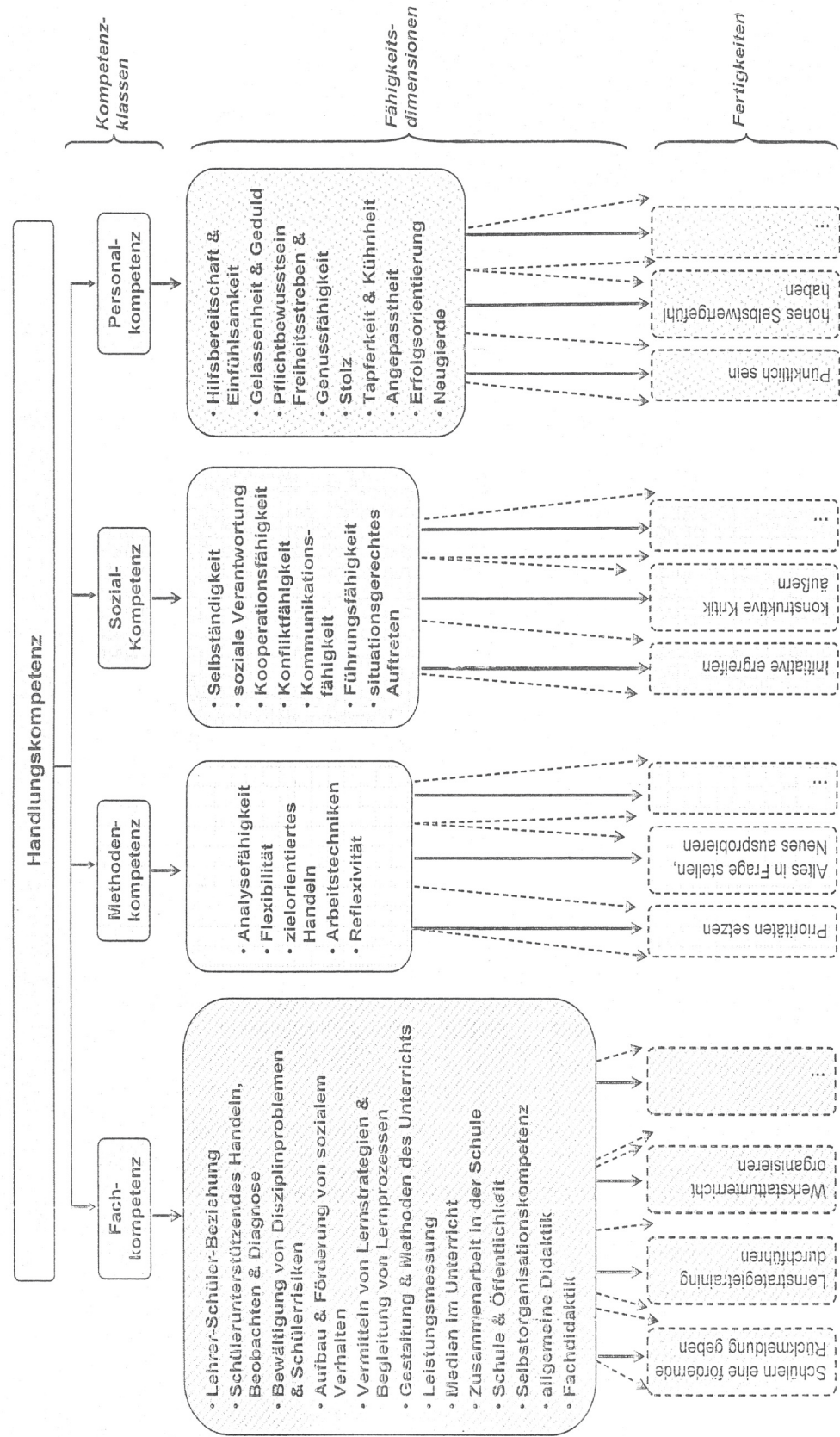
⁵¹ Ein Aspekt, den Baumert und Kunter an Freys Modell kritisieren, vgl. Baumert & Kunter 2006: 479.

die „Standards für die Lehrerbildung: Bildungswissenschaften“ formulieren (vgl. KMK 2004a), innerhalb dieses Modells näher zu differenzieren. Insofern agiert das hier diskutierte Forschungsvorhaben auf einer anderen Ebene, da es mit der ‚*Unterrichtskompetenz*‘ von Lehrern, einen spezifischen Bereich einer allgemeinen beruflichen Handlungskompetenz erfassen will. Demnach schließt sich das Forschungsvorhaben auf dieser Betrachtungsebene anderen Forschungsvorhaben an, welche einzelne Tätigkeitsbereiche von Lehrkräften isoliert in den Blick nehmen und diese modellieren. So geschehen z.B. für die Beratungskompetenz (vgl. Bruder et al. 2010), die diagnostische Kompetenz (vgl. Abs 2007, Gräsel et al. 2010), oder die Urteilskompetenz über Unterricht (Seidel & Prenzel: 2007, Seidel et al. 2010).⁵²

⁵² Allerdings muss hierbei festgestellt werden, dass diese Forschungsprojekte einem durchweg anderen Forschungsdesign, als der hier vertretene Ansatz folgen und die Art der Modellierung demnach gänzlich anders verläuft.

Abb. 2.2: Hierarchisches Strukturmodell der Handlungskompetenz nach Frey

Quelle: Frey & Jung 2011: 552



2.2.2 Die Messung und Erfassung von beruflicher Handlungskompetenz

Forschungsprojekte zur Erfassung beruflicher Handlungskompetenz von Lehrkräften greifen auf verschiedene Instrumente zurück, mit denen diese gemessen oder erfasst werden sollen. Ausgangspunkt des Instrumentendesigns sind die Überlegungen, auf welche Art und Weise sich Kompetenzen zeigen und wie sie sich messen lassen. Bereits in der oben zitierten Kompetenzdefinition von Klieme u.a. wurde deutlich, dass Kompetenzen nicht mit Wissen gleichzusetzen sind, sondern immer in einem Anwendungskontext stehen. Folglich sind reine Wissenstests unzureichend, um eine befriedigende Kompetenzdiagnostik zu betreiben. Vielmehr muss überprüft werden, inwieweit das Wissen in konkreten Handlungssituationen zielführend eingesetzt werden kann. Für die Erfassung der Handlungskompetenz kommen verschiedenen Methoden und Instrumente in Frage, begonnen bei einfachen Wissenstests und psychologischen Tests emotionaler und sozialer Aspekte, über Selbstbeurteilungen durch die Probanden, Beobachtungen in konkreten Handlungssituationen durch Diagnostiker oder die Anfertigung von Entwicklungsportfolios durch die Probanden (vgl. Frey 2006: 34–35, Kunter & Klausmann 2010:74- 78, Terhart 2007: 52–54). Jede dieser Methoden bringt spezifische Vor- und Nachteile mit sich. So muss etwa die Validität von Selbstauskünften kritisch hinterfragt werden, wohingegen sie zeitökonomisch eingesetzt und ausgewertet werden können (vgl. Abs 2007: 70–71, Kunter & Klausman 2010: 78, Terhart 2007: 53). Entwicklungsportfolios nehmen, sowohl für die Erstellung als auch in der Bewertung viel Zeit in Anspruch, erfassen dafür aber vielfältige Einzelelemente einer allgemeinen, beruflichen Handlungskompetenz, wohingegen ihr Bestand vor Testgütekriterien wiederum fragwürdig ist (vgl. Frey 2006: 34–35).⁵³

⁵³ Weitere Ausführungen zu den Möglichkeiten der Kompetenzmessung finden sich z.B. bei: Bruder et al. 2010: 277, Klieme 2004: 13, Maag Merki & Werner: 576–584, Oser et al. 2010: 135–138. Frey bietet in Frey 2006 eine Sammlung aller zwischen 1991 und 2005 publizierten Instrumente an, die sich für die Kompetenzdiagnostik angehender Lehrkräfte eignen.

In Bezug auf die Erfassung der diagnostischen Kompetenz führt Abs aus, dass eine anspruchsvolle Kompetenzmessung über die oftmals aus praktischen Erwägungen eingesetzten Selbsteinschätzungen hinausgehen müsste (vgl. Abs 2007: 71). Zentrale Überlegungen, die er bezüglich der Erfassung der diagnostischen Kompetenz ausführt, werden für die Gestaltung eines (heuristischen) Modells der Unterrichtskompetenz im Folgenden aufgegriffen und sollen hier deswegen kurz dargestellt werden. Abs führt aus, nach welchen Logiken Kompetenzen in Modellen erfasst werden können und in welchen hierarchischen Zusammenhängen sie sich ordnen lassen. Er findet diese u.a. in den Kompetenzzusammenhängen der Graduierung, der Expertise und des Repertoires (vgl. Abs 2007: 73–74).⁵⁴ Der *Zusammenhang als Graduierung* drückt aus, dass verschiedene Schwierigkeitsstufen bei der Bewältigung von Aufgaben ausgemacht werden, die mit verschiedenen Kompetenzstufen in Verbindung gebracht werden können. Hierbei zu unterscheiden sind Graduierungen, die sich durch einen *kumulativen Aufbau* von niedrigen zu hohen Kompetenzstufen auszeichnen und dabei die ehemals niedrigeren Kompetenzstufen in die jeweils höheren integriert werden, und einem *strukturellen Wandel* derselben, bei dem niedrige Kompetenzstufen durch höherwertige ersetzt werden.

Der *Zusammenhang als Expertise* drückt aus, dass das Wissen von Experten anders organisiert ist als jenes von Novizen. Erste können die Informationen effizienter verarbeiten und flexibler auf Fragestellungen reagieren. Sie verbalisieren ihr den Handlungen zu Grunde liegendes Wissen anders. In konkreten Handlungssituationen kann es sein, dass eine Vielzahl von Handlungsmöglichkeiten sinnvolle Wege eröffnen könnte. Der *Zusammenhang als Repertoire* trägt diesem Gedankengang Rechnung, in dem er eine hohe Kompetenz dadurch definiert, dass einem Handelnden verschiedene Handlungsoptionen zur Verfügung stehen, aus denen er unter Abwägung der jeweiligen Handlungsfolgen wählen kann (vgl.

⁵⁴ Er führt weiterhin die Zusammenhänge der ‚Erweiterung‘ (eine höhere Kompetenz liegt dann vor, wenn eine größere Anzahl an Aufgaben bewältigt werden kann) und der ‚Typenbildung‘ (es werden Konstellationen und Kontexte unterschieden, in denen gehandelt werden soll) aus, die nicht ohne weiteres auf den hier diskutierten Anwendungskontext übertragen werden können und deswegen nicht weiter thematisiert werden.

ders.: 73–74). Für die Modellierung einer konkreten beruflichen Handlungskompetenz, wie etwa der diagnostischen Kompetenz von Lehrkräften, sollten die Instrumente diese Kompetenzzusammenhänge aufgreifen und operationalisieren. Die so entstehenden „hypothetischen Modelle“ bedürften nach ihrer Formulierung eines empirischen Tests, um die hypothetisch unterstellten Kompetenzzusammenhänge zu bestätigen und den empirischen Gütekriterien zu entsprechen (vgl. ders.: 75–81)⁵⁵.

Was hier sehr verkürzt dargestellt wurde, ist ein Forschungsprogramm für die Kompetenzmodellierung im Rahmen quantitativer empirischer Forschung, die im Sinne einer summativen Evaluation den Erreichungsgrad einer bestimmten Kompetenzstufe feststellen, genauer: ‚messen‘ will. Hiervon muss erstens der hier gewählte Kontext der Kompetenzdiagnostik unterschieden werden, denn es geht um Evaluation bzw. Diagnostik im Rahmen von Ausbildungsprozessen, womit in Verbindung steht, dass zweitens die Zielsetzung der Diagnostik eine formative und keine summative ist. Im Rahmen von Ausbildungsprozessen steht nicht nur die Feststellung eines Ist-Standes im Mittelpunkt des Interesses, sondern immer auch das Aufzeigen von Entwicklungsmöglichkeiten. Entsprechend werden die Überlegungen von Abs zur Kompetenzmodellierung aufgegriffen und an den Kontext angepasst. In diesem Zusammenhang werden die Ansprüche an die klassischen Gütekriterien der Testtheorie für ‚Messungen‘, wie zu zeigen sein wird, nicht eingehalten. Deswegen wird im Folgenden nicht von der ‚Messung‘ von Kompetenzen, sondern von deren ‚Erfassung‘ gesprochen.

2.2.3 Unterrichtskompetenz – ein heuristisches Kompetenzmodell und die Erfassung der Unterrichtskompetenz

Ausgehend von den Überlegungen von Frey und Abs soll in diesem Unterkapitel ein heuristisches Modell der Unterrichtskompetenz dargelegt werden, das dem zu entwickelnden Bewertungskonzept zu Grunde liegt. Heuristisch deshalb, weil es den hohen wissenschaftlichen Ansprüchen an Modellbildung, die im Rahmen

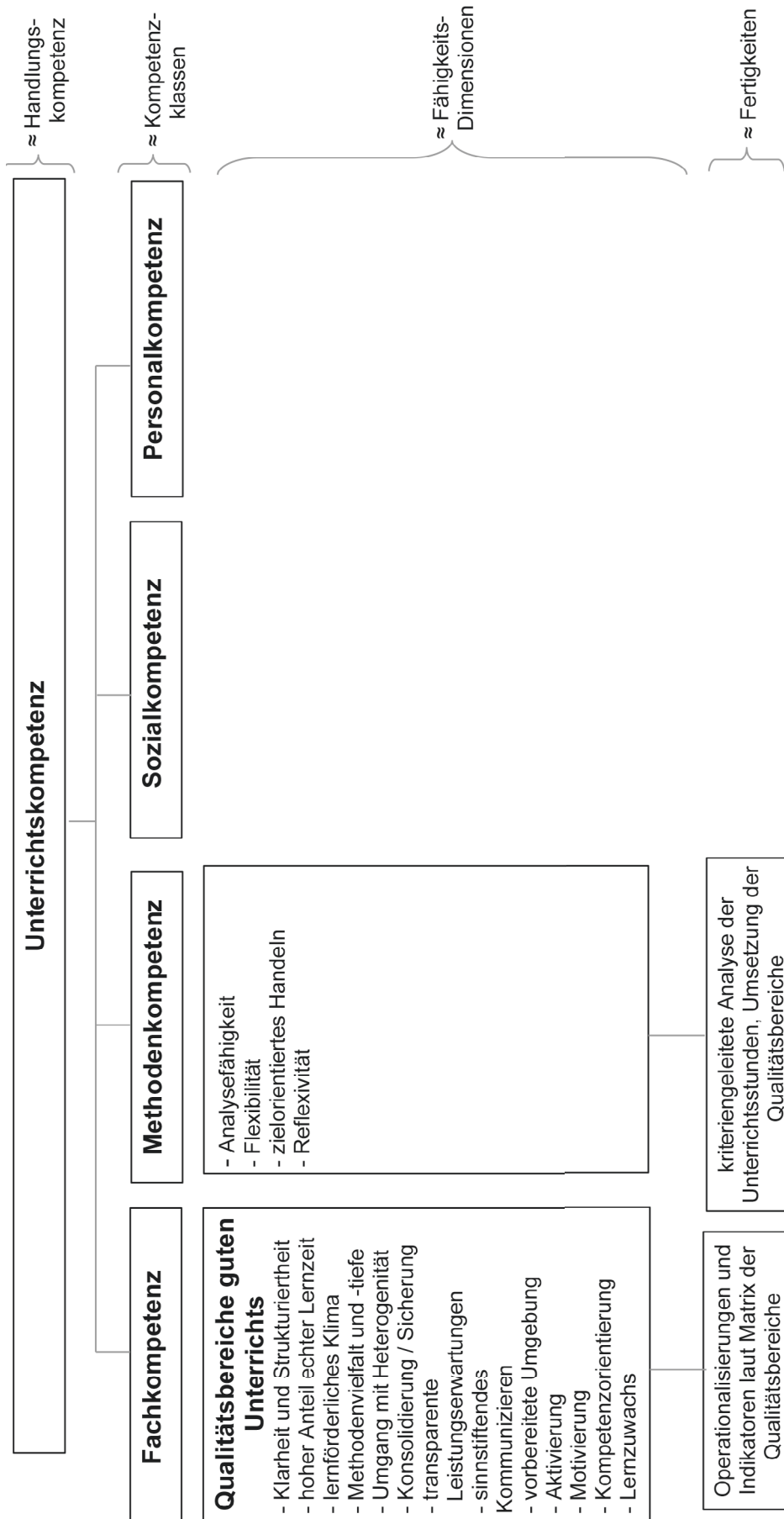
⁵⁵ Ähnlich gehen Bruder u.a. bei der Modellierung der diagnostischen Kompetenz vor (vgl. Bruder et al. 2010: 276–279).

quantitativer empirischer Forschung aufgestellt werden, nicht genügen kann. Das Modell lässt sich nicht – zumindest nicht im Rahmen dieser Promotion – durch empirische Test validieren. Für die angehenden Lehrer und Ausbilder, die im Rahmen der Zweiten Phase der Lehrerbildung einerseits Unterricht planen und durchführen und diesen andererseits bewerten müssen, macht es allerdings keinen großen Unterschied, ob sich das Konstrukt der ‚Unterrichtskompetenz‘ empirisch validieren lässt oder nicht. Sie müssen stillschweigend davon ausgehen, dass diese Kompetenz existiert und auch in Ausbildungsprozessen entwickelt werden kann, denn sonst wäre der Vorbereitungsdienst ohne großen Sinn. Das im Folgenden entwickelte Modell soll, eben im Sinne von Heuristik, auf der Basis begrenzten Wissens über die Kompetenzstrukturierung und -messung, wie sie in den vorangegangenen Unterkapiteln vorgestellt wurde, Transparenz über die hier unterstellte Struktur und Erfassbarkeit von ‚Unterrichtskompetenz‘ herstellen.

Ausgehend von Freys hierarchischem Strukturmodell der beruflichen Handlungskompetenz kann die Struktur der hier zu modellierenden Unterrichtskompetenz, wie folgt, dargestellt werden (vgl. Abb. 2.3). Wie die Abbildung zeigt, wird die Unterrichtskompetenz hier auf der Ebene der beruflichen *Handlungskompetenz* angesiedelt, die sich in die vier von Frey erwähnten *Kompetenzklassen* aufgliedert. Von diesen vier Kompetenzklassen werden von dem hier entwickelten Bewertungskonzept allerdings lediglich zwei abgedeckt, nämlich die Kompetenzklassen der Fach- und Methodenkompetenz. Die *Fähigkeitsdimensionen*, die die Fachkompetenzklasse näher definieren, finden sich in den Qualitätsbereichen guten Unterrichts, wie die Matrix der Qualitätsbereiche sie formuliert. Fachkompetenz liegt demnach dann vor, wenn die Qualitätsbereiche von einer Lehrperson im konkreten Unterricht realisiert werden können. Die *Fertigkeiten*, welche die Fähigkeitsdimension der Fachkompetenz konkretisieren, liegen im Rahmen der in der Matrix geleisteten Operationalisierung der einzelnen Qualitätsbereiche vor, die konkrete Handlungserwartungen formulieren (z.B. die Anpassung der Inhalte an die Lernausgangslage der Lernenden durch die Lehrperson). Die Fähigkeitsdimensionen der Methodenkompetenz werden, bis auf die „Arbeits-techniken“, von Frey übernommen. Sie konkretisieren sich durch die Fertigkeit-

ten einer kriteriengeleiteten Analyse des gezeigten Unterrichts (Analysefähigkeit und Reflexivität), die Umsetzung der Qualitätsbereiche in den Unterrichtsstunden (zielorientiertes Handeln) und ggf. in der Abweichung von der ursprünglich intendierten Planung, falls dies im Rahmen des Unterrichts sinnvoll sein sollte (Flexibilität). Die Klassen der Sozial- und Personalkompetenz finden in dem hier zu entwickelnden Bewertungskonzept und den es zusammensetzenden Bewertungsinstrumenten keinen Niederschlag. Damit werden auch die oben erwähnten motivationalen und volitionalen Aspekte des Kompetenzbegriffes in dem Bewertungskonzept nicht explizit aufgegriffen.

Abb. 2.3: Struktur der Unterrichtskompetenz in Anlehnung an Freys hierarchisches Strukturmodell beruflicher Handlungskompetenz



Die Instrumente des Bewertungskonzepts haben den Anspruch, die ausgewählten Kompetenzklassen, Fähigkeitsdimensionen und Fertigkeiten zu erfassen. Der hohe Anspruch der Messung von Kompetenzniveaus, der Grundlage von Kompetenzdiagnostik im Rahmen quantitativer empirischer Forschung ist und dessen Qualitätsstandard die Gütekriterien der klassischen Testtheorie, Validität, Objektivität und Reliabilität sind, wird hier nicht angestrebt. Stattdessen wird hinsichtlich der Gütekriterien des Erfassungsprozesses von ‚Unterrichtskompetenz‘ ein Weg eingeschlagen, den Bovet und Frommer für die Bewertung im Rahmen der Zweiten Phase der Lehrerbildung vorgeschlagen haben (vgl. Bovet & Frommer 2009: 74–77). Bevor diese Gütekriterien, auf denen die Erfassung der Kompetenz beruht, näher erläutert werden, sollen, basierend auf Abs (s.o.), die Kompetenzzusammenhänge, die der Erfassung der Unterrichtskompetenz zu Grunde liegen, genauer dargestellt werden.

Die drei oben ausgeführten Kompetenzzusammenhänge der Kompetenz als Expertise, Kompetenz als Repertoire und Kompetenz als Graduierung bilden die Grundlage des späteren Instrumentendesigns des Bewertungskonzepts. Wie bereits oben ausgeführt, erfolgt im Rahmen der hier entworfenen Diagnostikinstrumente keine anspruchsvolle Modellierung der verschiedenen Kompetenzstufen, etwa im Sinne einer theoriebasierten Entwicklung derselben oder im Rahmen eines post-hoc-Tests (vgl. Abs 2007: 78–80), die sich auf einen der Kompetenzzusammenhänge beziehen würde. Stattdessen finden alle drei oben erwähnten Kompetenzzusammenhänge im Rahmen der Bewertungsinstrumente ihren Niederschlag. Der *Kompetenzzusammenhang als Graduierung* findet in dem Bewertungskonzept dahingehend Berücksichtigung, dass die Instrumente im Rahmen der Ausbildung einen *strukturellen Wandel der Analysefähigkeit* der Auszubildenden abbilden sollen. Mit Hilfe von Ratingskalen soll die Reflexionsfähigkeit der Auszubildenden hinsichtlich verschiedener Kriterien erfasst werden, die im Fall einer erfolgreichen Kompetenzentwicklung im Rahmen der Ausbildung einen Wandel der Art und Weise abbilden sollen, wie die Auszubildenden über ihren Unterricht sprechen und diesen analysieren⁵⁶. Dabei wird von einer additiv-beschreibenden Art und Weise

⁵⁶ Siehe dazu die Ausführungen zu der pre-Alpha-Version des Bewertungskonzepts in Kapitel 2.3 und die Instrumente im Anhang.

der Reflexion, die Unterricht chronologisch darstellt und typisch für Novizen ist ausgegangen. Am Ende steht ein Expertenstatus, der sich durch eine kriteriengeleitete Analyse einzelner neuralgischer Punkte der Unterrichtsplanung und -durchführung auszeichnet. Diese Skalen sollen gleichzeitig den *Kompetenzzusammenhang als Expertise* erfassen, der sich, wie in Unterkapitel 2.2.2 ausgeführt ja genau durch die unterschiedliche Art und Weise, wie die angehenden Lehrer über ihren Unterricht sprechen, auszeichnet.⁵⁷ Der *Kompetenzzusammenhang als Repertoire* wird seinerseits auch durch eine Ratingsskala erfasst, mit der eingeschätzt werden soll, inwieweit die Auszubildenden in der Unterrichtsnachbesprechung dazu in der Lage waren, selbstständig nach Alternativen für didaktisch-methodische Entscheidungen zu suchen, die sich im Rahmen des Unterrichts als nicht sinnvoll erwiesen haben. Kompetenz im Sinne des Repertoires zeigen sie eben dann, wenn Sie für die konkreten Handlungssituationen, in denen sich eine bestimmte Entscheidung als nicht sinnvoll erwiesen hat, alternative Handlungsoptionen benennen und deren vermutliche Auswirkungen abschätzen können. Der Kompetenzzusammenhang als Repertoire soll weiterhin mit einem eigenen Instrument, der ‚Sammlung über das Repertoire‘ erfasst werden⁵⁸. Alle Ausbilder eines Auszubildenden tragen hier einen Teil ihrer Unterrichtsbeobachtungen ein, die sie im Rahmen der Unterrichtsbesuche getätigt haben, u.a. bestimmende Merkmale der Unterrichtsinszenierungen. Das Instrument gibt Auskunft darüber, wie vielfältig und variabel die Unterrichtsinszenierungen eines Auszubildenden sind: Kann er ein breites Repertoire an Unterrichtsinszenierungen anbieten oder finden sich nur wenige Variationen in seinem Unterrichtsrepertoire wieder, und hat er deshalb im Sinne des Kompetenzzusammenhangs als Repertoire als Experte oder als Novize zu gelten?⁵⁹

⁵⁷ Ähnliche Überlegungen zur Erfassung der „Wissenstypen und ihrer mentalen Repräsentationen“ von Novizen und Experten stellen Baumert und Kunter an, vgl. dies: 2006: 483–484.

⁵⁸ Eine ähnliche Idee für das Design eines Bewertungskonzepts formuliert Dollase (vgl. Dollase 2007).

⁵⁹ Diese Arbeit muss davon absehen, eine genauere Unterscheidung verschiedener Entwicklungsstufen auf dem Weg vom Novizen zum Experten zu definieren (vgl. dazu Baumert & Kunter 2006: 506–507). Sicherlich können auch erfolgreiche Referendare, die in ihrem Vorbereitungsdienst eine beeindruckende Entwicklung ihrer beruflichen Handlungskompetenzen vorzuweisen haben, am Ende des Referendariats noch nicht als „Expertenlehrer“ gelten, dafür fehlt ihnen einfach die notwendige Berufserfahrung. Die Instrumente des hier zu entwickelnden Bewertungskonzepts vertreten den Anspruch, eine Entwicklung in Richtung eines Expertenstatus näherungsweise abbilden zu können.

Das hier zu entwickelnde Instrumentarium zur Erfassung der Unterrichtskompetenz genügt somit einem höheren Qualitätsanspruch, als ihn einfache Selbstauskünfte der Lehrkräfte über die Qualität ihres Unterrichts oder Wissenstest über Unterrichtsqualität, entsprechend könnten. Allerdings entspricht es den hohen messtheoretischen Ansprüchen einer anspruchsvollen Kompetenzdiagnostik bei Lehrern, wie sie z.B. Maag-Merki und Werner sie formulieren, nur eingeschränkt. Diese konstatieren,

„[...] dass die Erfassung von Kompetenzen von Lehrpersonen mehrdimensional zu erfolgen hat, indem nicht einzig Professionswissen, sondern auch die weiteren Dimensionen wie Einstellungen, Motivationen etc. als Inhaltsdimension in den Beurteilungsprozess gebunden sein müssen“ (Maag-Merki & Werner 2011: 576).

Wie weiter oben ausgeführt, werden die Kompetenzdimensionen des motivationalen und volitionalen Bereichs durch das Konzept nicht oder bestenfalls implizit erfasst.

An dieser Stelle muss die Frage aufgeworfen werden, inwiefern das Konzept dann wirklich die Kompetenz von Lehrkräften, also eine abstrakte, in vielen Kontexten realisierbare Fähigkeit zu unterrichten, erfasst oder lediglich *Performanz* in Einzelsituationen. Der Begriff der Performanz umschreibt ein kompetentes Verhalten in einzelnen Situationen, das nicht gleichzusetzen ist mit einer generalisierten Kompetenz. Vielmehr kann das in Einzelfällen gezeigte kompetente Verhalten z.B. auch auf optimalen Kontextbedingungen beruhen, die es begünstigt haben, aber schon eine marginale Verschlechterung der Bedingungen es einer Person nicht mehr ermöglichen würden, dieses Verhalten erneut zu zeigen. Mit dem Einbezug der motivationalen und volitionalen Kompetenzdimensionen, die bei Frey in der Sozial- bzw. Personalkompetenzklasse verortet sind, versuchen Kompetenzmodelle, diesen Aspekt zu erfassen und in Instrumente zur Messung von Kompetenz zu übersetzen. Die Übertragung eines singular realisierten erfolgreichen Verhaltens auf andere Kontexte bedingt die Fähigkeit zur Reflexion über die Bedingungen des Handlungserfolges oder Misserfolges in einer konkreten Situation sowie den Willen, den Erfolg auf andere Situationen zu übertragen

(Volition) und die dauerhafte Aufrechterhaltung dieser Haltung (Motivation) (vgl. Maag Merki & Werner 2011: 576–577). Während das hier entwickelte Diagnostikkonzept den Bereich der Reflexionsfähigkeit erfasst, werden die Bereiche der Volition und Motivation ausgespart, sodass das Konzept wesentliche Aspekte einer ‚Unterrichtskompetenz‘ nicht erfasst. Da im Rahmen der Unterrichtsbesuche Auszubildende über einen längeren Ausbildungszeitraum beobachtet werden und in diesem Zeitraum bestimmte Unterrichtssituationen, in denen ein spezifisches kompetentes Verhalten gezeigt werden muss (z.B. eine angemessene fachdidaktische Reduktion der Inhalte, Moderieren eines Unterrichtsgesprächs, Anleiten kooperativer Arbeitsformen, etc.) immer wieder auftauchen, kann allerdings davon ausgegangen werden, dass die Beobachtung von realisierter Kompetenz gegenüber der Performanz *näherungsweise* darüber abgegrenzt werden kann, dass bestimmte Fähigkeiten in unterschiedlichen Lerngruppen und zu unterschiedlichen Zeiten mehrfach gezeigt wurden.

Die Diagnostik der Kompetenzzusammenhänge erhebt, wie oben angerissen, nicht den Anspruch, eine ‚Messung‘ im Sinne quantitativer empirischer Forschung zu sein. Stattdessen beansprucht das Konzept eine ‚Erfassung‘ der Unterrichtskompetenz, an die andere Gütekriterien als jene der klassischen Testtheorie als Qualitätsstandards angelegt werden sollen. Bovet und Frommer haben im Rahmen eines Forschungsprojektes zu Bewertung von Lehrproben und Unterricht im Rahmen von Ausbildungsprozessen, sowie der Beförderung im Schuldienst alternative Gütekriterien für Qualität von Unterrichtsbewertungen vorgeschlagen, welche dem zu entwickelnden Bewertungskonzept zu Grunde gelegt werden sollen (vgl. Bovet & Frommer 2009: 74–77). Sie führen aus, dass im Kontext von Unterrichtsbewertungen nicht die Messgenauigkeit das entscheidende Qualitätskriterium sei, sondern die „Angemessenheit“ des Urteils. Die Angemessenheit im Kontext von Unterrichtsbewertungen, insbesondere der Endnote des Vorbereitungsdienstes, sehen sie als „angemessen“ an, wenn folgende Bedingung erfüllt ist:

„Die Beurteilung muss der einstellenden Behörde zu erkennen geben, in welchem Maße ein Lehrer z.B. für eine Funktionsstelle geeignet ist, sie muss den Grad der tatsächlichen Eignung wiedergeben, der beim Lehrer vorliegt“ (dies.: 74).

In Bezug auf die Unterrichtskompetenz bedeutet dies, dass eine ‚angemessene Beurteilung‘ durch das Bewertungskonzept die tatsächliche Fähigkeit der angehenden Lehrer, im Berufsalltag guten Unterricht im oben operationalisierten Sinne, anzuleiten, erfassen muss. Diese Angemessenheit des Urteils soll über alternative Gütekriterien erreicht werden, welche diejenigen der Validität, Reliabilität und Objektivität ersetzen.

Anstatt den Anspruch der *Validität* an die Erfassung des Unterrichtskompetenz zu stellen, wird hier in Anlehnung an Bovet und Frommer v.a. der Weg einer kommunikativen Urteilsvalidierung gewählt. Die *kommunikative Validierung* zeichnet sich dadurch aus,

„[...] dass man bei der Datenanalyse (Unterrichtsberatung und -beurteilung) in den Dialog mit den Subjekten (Lehrern) eintritt und gesprächsweise klärt, ob man ihre Intentionen und ihr Handeln richtig verstanden hat“ (dies.: 75).

Eine *gültige Erfassung* liegt demnach dann vor, wenn ein Urteil im Rahmen von Kommunikationsprozessen begründet werden kann und auch Nachfragen und kritischen Blicken aus anderen Perspektiven standhält. Die kommunikative Validierung soll im Rahmen des hier zu entwickelnden Bewertungskonzepts in den Unterrichtsnachbesprechungen sichergestellt werden, wobei das Design der Bewertungsinstrumente die kommunikative Validierung gezielt unterstützen soll⁶⁰.

Darüber hinaus wird das Prinzip der ökologischen Validierung zumindest teilweise als Grundlage der Beurteilungsprozesse herangezogen. Die *ökologische Validierung* meint hier die Berücksichtigung des konkreten Kontextes, in dem sich der Unterricht entfaltet, dass eine Analyse der Beobachtung

„[...] die Bedingungen des jeweiligen Lebensraums (Fach, Klasse, Raum) und die des Subjekts (Konzept, eigene Ideen und Absichten, persönliche Stärken und Schwächen des Lehrers) mit in Betracht zieht“ (dies.: 75).

⁶⁰ Siehe dazu die Ausführungen zur pre-Alpha Version des Bewertungskonzepts in Kapitel 2.3.

Allerdings verzichtet das Bewertungskonzept auf Grund der gegebenen Ausbildungsbedingungen auf andere Teilaspekte einer ökologischen Validierung, nämlich der Untersuchung im ‚natürlichen Lebensraum‘ der Subjekte, also dem alltäglichen Unterricht. Die Lehrproben sind als Prüfungssituationen definiert und als solche in den Prüfungsordnungen festgeschrieben; sie stellen keinen ‚natürlichen‘ Handlungskontext eines Lehrers dar, sondern sind, wie in Kapitel 1.2 ausgeführt, hochgradig künstliche Situationen.

Da jede Unterrichtsstunde eine Sache sui generis darstellt, deren spezielle Kontextbedingungen so nie wieder hergestellt werden können, muss der Anspruch von *Reliabilität* von Bewertungsprozessen fallen gelassen werden. Die Zuverlässigkeit einer Unterrichtsbewertung wird sich nicht dadurch sicher stellen lassen, dass die gleiche Unterrichtsstunde zu einem anderen Zeitpunkt ein zweites Mal bewertet wird, da sie sich eben nicht ein zweites Mal realisieren lässt. Bei der Bewertung von Unterricht gilt es vielmehr, die Ungenauigkeit und die Personen- und Kontextgebundenheit jedweder Bewertung als dem Vorgehen immanente Eigenheiten zu akzeptieren, so Bovet und Frommer (vgl. dies: 76; Maag Merki & Werner 2011: 583). An die Stelle der Reliabilität tritt der *Realitätsgehalt* der Erfassung, der sicher stellen soll, dass die Bewertung von Unterricht oder eben die Erfassung der Unterrichtskompetenz, in möglichst vielen verschiedenen Beobachtungskontexten geschieht, die den tatsächlichen Anforderungen des Berufsalltags entsprechen. Der Realitätsgehalt korrespondiert demnach stark mit der ökologischen Validität und hierbei auch mit den Aspekten, die durch die Situierung der Bewertungsprozesse in Prüfungssituationen für das hier vorliegende Konzept nicht oder nur teilweise erfüllt sind. Gleichwohl wird dem Realitätsgehalt dadurch Rechnung getragen, dass die Bewertungsinstrumente des Konzepts die Vielfalt der Kontexte in denen Unterricht gezeigt wurde, erfassen. Im Fazit der Arbeit wird diskutiert, inwiefern die Erfahrungen der Akteure, die sich an der Erprobung des Bewertungskonzepts beteiligt haben, darauf hindeuten, dass diese alternativen Gütekriterien eingehalten werden konnten.

Objektivität in der Bewertung von Unterricht sicherzustellen wird von vielen Auszubildenden eingefordert, ist aber kaum vollständig zu realisieren. Im Rah-

men von Unterrichtsbewertungen wird weder eine Standardisierung der Erhebungssituationen (Durchführungsobjektivität) noch eine Standardisierung der Dateninterpretation (Interpretationsobjektivität) zu gewährleisten sein. Inwieweit eine (teilweise) Standardisierung der Datenerhebung und -auswertung (Auswertungsobjektivität) im Rahmen von Unterrichtsbeobachtungen in der Zweiten Phase der Lehrerbildung sinnvoll ist, ist eine der zentralen Fragestellungen dieser Arbeit. Allerdings wird der Umfang der Standardisierung nicht den Ansprüchen der klassischen Testtheorie genügen, um ein objektives Erhebungsinstrument herzustellen⁶¹. Bovet und Frommer schlagen vor, die Objektivität auf Grund der einschränkenden Eigenschaften der Unterrichtsbesuche durch die Intersubjektivität und Stimmigkeit der Bewertungen zu ersetzen (vgl. dies.: 76–77). Die *Intersubjektivität* umschreibt den Prozess, in dem verschiedene Bewerter ihre subjektiven Einzelurteile zu einem kommunikativ generierten und intersubjektiv nachvollziehbaren Gesamturteil zusammenfassen und in diesem Prozess das Zustandekommen ihres subjektiven Urteils reflektieren und ggf. korrigieren. Die *Stimmigkeit* verlangt, dass die Urteile eine innere Stimmigkeit aufweisen, also ohne *innere* Widersprüche aufeinander bezogen werden können. Weitergehend verlangt eine *äußere* Stimmigkeit, dass die formulierten Erkenntnisse mit solchen, die in anderen Kontexten generiert worden sind, vereinbar sind. Ein Aspekt, der die Stimmigkeit eines Urteils betreffen würde, wäre demnach z.B. die Frage, ob ein Auszubildender grundsätzliche Probleme hat, den Qualitätsbereich der Klarheit und Strukturiertheit zu realisieren, worauf eine unstrukturierte und chaotische Unterrichtsstunde hinweisen könnte, oder ob im Sinne der äußeren Stimmigkeit schon Beobachtungen anderer Unterrichtsstunden existieren, die diesem Urteil widersprechen.

Ausgehend von den Ausführungen in diesem Unterkapitel kann das dem Bewertungskonzept zu Grunde gelegte Modell von Unterrichtskompetenz wie in Abbildung 2.4 charakterisiert werden.

⁶¹ Siehe dazu auch die Ausführungen zu dem Grad der Standardisierung des vorliegenden Bewertungskonzepts in Kapitel 4.3.

Abb. 2.4: Eigenschaften des heuristischen Modells von ‚Unterrichtskompetenz‘

Kompetenzstruktur	Hierarchische Struktur von Unterrichtskompetenz über <ul style="list-style-type: none"> ▪ Kompetenzklassen ▪ Fähigkeitsbereiche ▪ Fertigkeiten <i>(in Anlehnung an Frey, vgl. Frey & Jung 2011: 552)</i>
Kompetenzerfassung	Erfassung über Kompetenzzusammenhang <ul style="list-style-type: none"> • als Graduierung • als Expertise • als Repertoire <i>(in Anlehnung an Abs, vgl. Abs 2007: 78–80)</i> Erfassung durch Beobachtung von Unterricht im Rahmen von Lehrproben Gütekriterien der Erfassung von Unterrichtskompetenz <ul style="list-style-type: none"> • kommunikative und ökologische Validierung • Realitätsgehalt • Intersubjektivität • innere und äußere Stimmigkeit <i>(in Anlehnung an Bovet und Frommer, vgl. Bovet & Frommer 2009: 74–77)</i>

2.2.4 Kompetenzen als Zielkategorie von Lernprozessen – eine sinnvolle Perspektive pädagogischen Handelns und erziehungswissenschaftlicher Forschung im Bereich der Lehrer(aus)bildung?

Die bisherigen Ausführungen in diesem Kapitel haben den Trend im Bildungswesen und der scientific community hin zu einer stärkeren Kompetenzorientierung bei Lernprozessen kommentarlos hingenommen und auf den Bereich der Lehrerbildung übertragen. Innerhalb des (deutschsprachigen) erziehungswissenschaftlichen Diskurses gibt es allerdings auch kritische Stimmen, welche den Trend zu einer stärkeren Kompetenzorientierung hinterfragen und deren Sinnhaftigkeit zur Disposition stellen. Diese Kritik soll hier aufgegriffen werden, um die Frage zu stellen, inwiefern die Kompetenzorientierung eine sinnvolle und gewinnbringende Perspektive für formative Evaluation in der Lehrerausbildung ist.

Die Kritik an der Kompetenzorientierung bettet sich ein in eine generelle Kritik an der standardbasierten Bildungsreform, wobei sich ein Kern der Kritiker im

Jahr 2005 mit den „Frankfurter Einsprüchen gegen die technokratische Umgestaltung des Bildungswesens“ ein erstes Mal organisiert zu Wort meldete⁶². Zentrale Thesen wurden u.a. im Rahmen von Tagungen im Jahr 2005, 2010 und 2012 formuliert und in drei Ausgaben der Vierteljahresschrift für wissenschaftlichen Pädagogik präsentiert⁶³. Ohne hier eine umfassende Aufarbeitung der Kritik leisten zu können, lässt sich diese grob auf zwei Argumentationslinien zuspitzen⁶⁴: Erstens lasse die Kompetenzorientierung die Fundierung in einer geschlossenen Bildungstheorie vermissen und folge einem technokratisch verkürzten Verständnis von Lernprozessen, und zweitens lege die standardbasierte Bildungsreform und mit ihr die Kompetenzorientierung einen totalitären Anspruch an den Tag, welche den Zielen pädagogischen Handelns zuwider laufe⁶⁵.

Zum ersten Kritikpunkt schreibt etwa Ladenthin, dass es im Rahmen der Kompetenzorientierung im Bildungssystem zu einer „McDonaldisierung“ komme, einer Umstellung des Bildungssystems auf das Erzeugen von immer gleichen „Produkten“, wobei die Kompetenzen eben jene Produkte seien, die das Bildungssystem hervorzubringen habe, ohne dass im Rahmen dieses Diskurses geklärt werden könnte, was denn nun übergeordnetes Ziel der Lernprozesse sein soll (vgl. Ladenthin 2010: 347). Eine Pädagogik, die sich einseitig auf die Optimierung von Lernprozessen fokussiere und formale Kompetenzen schule,

„[...] kann nicht mehr unterscheiden zwischen Kultur, zwischen moralischem und unmoralischem Handeln, zwischen Humanität und Barbarei. [...]. An einem extremen Beispiel ausgedrückt: Ein Trainingslager der Neonazis ist unter kompetenztheoretischen Gesichtspunkten von einer Ausbildung in der Altenpflege nicht zu unterscheiden“ (ders.: 354).

⁶² Vgl. <http://www.uni-frankfurt.de/fb/fb04/initiativen/einsprueche/>, abgerufen am 4.6.2013.

⁶³ Vgl. die Ausgaben 3/2010, 3/2012 sowie das Sonderheft Frost 2006.

⁶⁴ Vgl. zu einer detaillierteren Aufschlüsselung der Kritik im deutschen Sprachraum Terhart 2006: 36–38.

⁶⁵ Um den Kritikern der standardbasierten Bildungsreform gerecht zu werden, muss allerdings angeführt werden, dass diese sich primär mit der Umsetzung der Kompetenzorientierung im schulischen Bereich auseinandersetzen und die Veränderungen, die dieser durch die neue pädagogische Perspektive erfahren hat, wesentlich stärker sind als die Änderungen, die sich bisher in der Lehrerbildung niedergeschlagen haben, um die es hier im folgenden gehen soll.

Auch Gruschka sieht die Bildungstheorie ausgeschlossen, wenn er konstatiert, dass die Frage nach der richtigen Bildung und die mit dieser Frage verbundenen innerdisziplinären und gesellschaftlichen Konflikte, was ‚Bildung‘ ausmache, abgelöst würden durch eine rein positivistisch gewendete Machbarkeitswissenschaft, der es in Bezug auf schulische Bildungsprozesse, nur noch darum gehe,

„[...] wie man das Gegebene so operationalisiert, dass die deutschen Schüler bei zukünftigen OECD-Messungen wieder zur Spitzengruppe vorstoßen“ (Gruschka 2006: 140)⁶⁶.

‚Kompetenz‘ als Begriff, analytisches Instrument und Zielkategorie von Lernprozessen würde beliebig und unscharf, denn auf die genuinen bildungstheoretischen Fragen der Pädagogik,

„[...] welche Kompetenzen denn eigentlich notwendig sind, um das Potential zur vernünftigen Gesellschaft in den Gegenständen der Kultur zu entbinden und anschließend, was davon und wie in der öffentlichen Erziehung vollzogen werden könnte“ (Gruschka 2013: 80),

bliebe eine kompetenztheoretische Sichtweise Antworten schuldig. Koch führt aus, dass das fehlende qualitative Außenkriterium der Kompetenzen im Rahmen einer Bildungstheorie zu bildungstheoretischen Tautologien führe:

„Qualität ist, was der Qualitätstest misst. Noch einfacher ausgedrückt: Schulqualität ist das Ergebnis des kompetenzdefinierten Qualitätstest. Qualitätssicherung ist der Versuch, befriedigende Testergebnisse auf Dauer zu stellen; Qualitätssteigerung das Bestreben, bessere Testergebnisse zu erzielen“ (Koch 2010: 324–325).

Verbunden mit der mangelnden Begründbarkeit der Kompetenzen und Standards aus einer bildungstheoretischen Perspektive erkennen die Kritiker in der

⁶⁶ Insofern zeigen sich hier, wie auch im weiteren Verlauf der Kritik von Gruschka und den anderen im Verlauf zitierten Autoren, deutliche Parallelen zu der ebenfalls von Gruschka vorgetragene Kritik an den Merkmalskatalogen guten Unterrichts, bei der er ebenfalls auf eine mangelnde bildungstheoretische Fundierung dieser hinwies (siehe Kapitel 2.1.4).

Aufgliederung der Kompetenzen in Unter- und Teilkompetenzen, wie sie in den Kompetenzmodellen und darauf aufbauenden formulierten Standards vorgenommen wird, eine „Taylorisierung des Menschen“ (Ladenthin 2010: 350). Sie kritisieren, dass sich der Lernprozess zur Erlangung von Fähigkeiten oder Wissen nicht durch die Zerlegung von komplexen Handlungen und Fähigkeiten in ihre Einzelkomponenten verstehen lasse, geschweige denn, dass ein solches Vorgehen dazu geeignet sei, einen Lernprozess zum Erlangen einer komplexen Fähigkeit anzuleiten⁶⁷:

„Kompetenzen beschreiben psychische Grundvermögen, benennen «nicht-mehr-teilbare» Denkopoperationen und zwar in zu allen Teilhandlungen zerlegten Arbeitsschritten, aber aus ihnen lässt sich nicht Bedeutung konstruieren. Selbst wer alle Operationen im Einzelnen beherrscht, muss sie nicht zur Synthese bringen können“ (Ladenthin 2010: 352).

„Wie soll man kommunikative Kompetenz unterrichten? Etwa indem man die diversen Teilkompetenzen nach einer Inventarliste einübt? Kommt auf diese Weise das Ganze einer sprachlichen Bildung zustande oder nur eine Kompetenzaddition, kein Haus, sondern ein Stein- und Bretterhaufen“ (Koch 2010: 330)?

In der Umsetzung der standardbasierten Bildungsreform in Kerncurricula für die Schulen macht Gruschka entsprechend „[...] ein entropisches sich verlieren in immer weiteren Unterscheidungen, mit denen der Wald vor lauter Bäumen nicht mehr bestimmt wird [...]“ aus (Gruschka 2013: 78)⁶⁸.

Rekonstruiert man die Kritik an der standardbasierten Bildungsreform und der Kompetenzorientierung nun weiter, so finden mehrere Kritiker in ihnen den Kern einer totalitären Auffassung von Bildungsprozessen wieder. So etwa wieder-

⁶⁷ Waldow zeigt auf, dass eine frühe Form der standardbasierten Bildungsreform, das „Social Efficiency Movement“ zu Beginn des 20. Jahrhunderts in den USA, in der Tat durch die Überlegungen Frederick Taylors beeinflusst wurde (vgl. Waldow 2012: 162–163).

⁶⁸ Vgl. zur Position der Befürworter der standardbasierten Bildungsreform, an der Kritik einer mangelnden Bildungstheorie z.B. Köller 2010: 540–541.

rum Ladenthin, für den die Kompetenzorientierung als Sozialtechnologie aufzufassen ist, die er eindeutig von einer ‚Bildung‘ abgrenzt, die im klassischen pädagogischen Sinne auf die Emanzipation des Individuums abziele:

„Kompetenzschulung bildet den Menschen nicht, sondern sie modelliert den Menschen. [...] Sie (Anm.: die Anhänger der Kompetenzorientierung, D. L.) wollen nicht die Bildung des Selbst, sie wollen die Formung des Selbst, ja die Erschaffung des Selbst erreichen“ (Ladenthin: 350, Herv. i. Org.).

Gerade der Einbezug von motivationalen Aspekten in die Kompetenzformulierungen, die im Rahmen der Standardisierung damit auch zum Bewertungsaspekt werden, stellt für Ladenthin eine totalitäre Vorstellung von Lernprozessen dar:

„Eine solche Verhaltensmodifikation zielt massiv auf das, was zu schützen alle bisherigen Bildungstheorien angestrebt hatten, nämlich den freien Willen und damit das Wollen des Selbst, das uns selbst bestimmt“ (ders.: 349:)

Koch schließt sich dieser Perspektive an, wenn er konstatiert, dass der Mensch als „Ressource instrumentalisiert“ werde und er sich in einem plumpen „Funktionalismus“ der „Standardisierung der Wollens“ einordnen müsse (vgl. Koch 2010: 326). Der Rolle der Lehrer drohe unter einem solchen Standardisierungsregime die „Depotenzierung zu Standard-Verwaltern“ (vgl. ders.: 325).

Generell ließe sich auf dieses Argument mit Luhmann einwenden, dass es der Pädagogik und der Erziehung und Bildung *nie* um Freiheit, sondern immer um das Gegenteil gegangen sei und immer gehen werde. Denn das Ziel von Erziehungsprozessen sei eben grundsätzlich nicht die Entfaltung der vollkommenen Freiheit aller Individuen, sondern die Einordnung in einen bestimmten Horizont erwartbarer Verhaltensweisen und Denkmuster. Systemtheoretisch gesprochen geht es eben um die *Reduktion von Komplexität* (vgl. Luhmann 2002: 102–110, Luhmann 2004: 14–17). Ladenthins eigenes Argument umgedreht müsste eine Pädagogik, welche das „Wollen des Selbst“ als höchstes und einziges Ziel ihrer Arbeit sieht, auch ein Trainingslager von Nazis in seiner Sinnhaftigkeit anerkennen, wenn dieses Tun eben

der Entfaltung ihres (böartigen) Selbst entspricht. Postmodern gewendet ist auch die ‚Erziehung und Bildung zur Mündigkeit‘, wiederum mit Luhmann gesprochen, eine Reduktion von Komplexität, im Sinne des (systemtheoretisch wahrscheinlich nicht erfolgreichen) Versuchs, erwünschte Verhaltensweisen und Denkmuster wahrscheinlicher und unerwünschte unwahrscheinlicher zu machen. Auch die Legitimation demokratischer Handlungs- und Denkmuster als Zielsetzung erzieherischer und bildnerischer Prozesse ist Herrschaftslegitimation. Es gibt keine metaphysische Instanz, die die ‚Richtigkeit‘ oder ‚Wahrheit‘ dieser Absichten, gegenüber anderen feststellen kann (vgl. Teller 2009: 583–586), auch nicht „die Vernunft“, auf die Ladenthin zur Abwehr der postmodernen Herausforderung rekurriert (vgl. Ladenthin 2005: 11). Dieses nicht auflösbare Dilemma der Pädagogik ist kein genuines Phänomen (oder Problem) einer kompetenztheoretisch argumentierenden Sichtweise. Ob die Zielsetzung eines Unterrichtsvorhabens in der Anbahnung von ‚Handlungskompetenz‘ der Lernenden oder in einem ‚affektiven Lernziel‘ verbal verpackt wird, macht hinsichtlich der Frage, inwiefern diese Absichten ‚totalitär‘ sind oder sein könnten, keinen Unterschied. Die pädagogische Absicht, Einstellungen und Verhaltensweisen auf bestimmte Art und Weise zu formen, ist beiden zu eigen⁶⁹.

In Bezug auf die Lehrerbildung⁷⁰ muss aus dieser Kritik die Frage abgeleitet werden, inwiefern die Kompetenzorientierung der Ausbildung zu einem rein formalistischen Verständnis der Unterrichtstätigkeit führt, welches droht, ein totalitäres Regime in der Ausbildung zu errichten, das die angehenden Lehrer zu reinen Erfüllungsgehilfen eines technokratisch verkürzten Verständnisses von

⁶⁹ So ist z.B. die Intention, auf Einstellungen und damit auch auf die Motivation, sowie das Verhalten von Schülern formend Einfluss zu nehmen, sowohl in dem kompetenztheoretisch formulierten Kernlehrplan des Faches Erdkunde (Sek. I) für Gymnasien in Nordrhein-Westfalen, als auch in seinem lernzielorientierten Vorgänger von 1993 zu finden. Im Kernlehrplan heißt es: „Die Schülerinnen und Schüler sind dazu in der Lage, im Sinne nachhaltigen Wirtschaftens im eigenen Umfeld ökologisch verantwortungsvoll zu handeln“ (KLP Erdkunde: 30). Im Lehrplan von 1993: „Der Erdkundeunterricht leistet dazu (Anm.: Zur Sicherung der zukünftigen Lebensgrundlagen, D. L.) einen wichtigen Beitrag, indem er nicht nur Kenntnisse und Einsichten über naturräumliche Systeme vermittelt, sondern auch zu einem verantwortungsbewussten und schonend-behutsamen Umgang mit den natürlichen Lebensgrundlagen erzieht (Lehrplan Erdkunde: 34).

⁷⁰ Vgl. für eine Diskussion des „standard based Movement“ in der Lehrerbildung in den USA Cochran-Smith 2005: 8–9.

Unterricht degradiert, in welchem sie die Schüler zur „Employability“ für die Abnehmersysteme der Wirtschaft formen (vgl. Radtke 2006). Kommt es durch die Fokussierung auf Teilkompetenzen der Unterrichtskompetenz im Rahmen der Bewertungs- und Beratungsprozesse zu einer Engführung der Ausbildung, welche zentrale Aspekte des Lehrerseins, im Sinne der Begleitung eines emanzipatorischen Prozesses von Heranwachsenden, im Rahmen einer demokratischen Gesellschaft, ausklammert? Wie bereits oben im ausgeführt, ist die Beherrschung grundlegender Kompetenzen in Bezug auf die Anleitung von Lernprozessen, in deren Rahmen ein wie auch immer gearteter Fortschritt von Kompetenzen, ein Zuwachs von Wissen oder eine Veränderung von Einstellungen von Schülern stattfindet, eine notwendige, aber keine hinreichende Bedingung, um ein ‚guter Lehrer‘ zu sein. Zur Entwicklung einer Lehrerpersönlichkeit und damit zur Lehrerausbildung gehört ganz sicher auch die Auseinandersetzung mit klassisch bildungstheoretischen Fragen, mit dem Ziel, eine ‚persönlichen Bildungstheorie‘ zu entwickeln, die im klafkischen Sinne den ‚Bildungsgehalt‘ der unterrichteten Inhalte oder Kompetenzen bestimmen kann. Aber auch hier gilt, dass die Referendare im Vorbereitungsdienst primär mit anderen Herausforderungen konfrontiert sind als derjenigen tiefgreifende bildungstheoretische Reflexionen anzustellen⁷¹. Sie müssen grundlegende Routinen entwickeln, um vor einer Schulklasse zu bestehen, um Unterricht anzuleiten, der für die Schüler einen klar erkennbaren Schwerpunkt hat, mit dem sie sich auseinandersetzen sollen usw.. Ohne diese Kompetenzen ist alle bildungstheoretische Reflexion in der Praxis ohne großen Nutzen. Auch die Klagen der Referendare über die utopischen Standards der Bewertung, die in der Ausbildung an sie angelegt werden, sind ein starkes Argument für eine pragmatische Zielsetzung des Vorbereitungsdienstes, dem es darum geht den Auszubildenden *grundlegende, alltagsbezogene* Kompetenzen des Unterrichtens näher zu bringen, ohne damit den grundsätzlichen und wichtigen Stellenwert bildungstheoretischer Reflexionen zu negieren. Mit Gruschka ließe sich nun an dieser Stelle einwenden, dass mit einer so verstandenen Kompetenzorientierung wenig neues gesagt würde, sondern lediglich das, was (Lehrer-) *Bildung*

⁷¹ Was natürlich nicht heißt, dass die Auszubildenden nicht immer wieder bildungstheoretische Reflexionen, u.a. im Rahmen ihrer schriftlichen Kommentare der Unterrichtsbesuche anfertigen müssen.

schon immer gemacht habe, nämlich einen Prozess anzuregen, in dem es „[...] zur Bildung neuerer leistungsfähigerer Fähigkeiten komme“ (Gruschka 2006: 146), durch eine andere Semantik beschrieben würde. Streng genommen kann man dem mit Terhart zustimmen, wenn dieser schreibt:

„Insofern führen Standards für die Lehrerbildung nichts grundsätzlich Neues ins System ein; sie entwickeln das längst vorhandene Element der Definition von Ansprüchen jedoch konsequent in Richtung auf eine Konkretisierung dieser Ansprüche zum Zweck der Überprüfung des tatsächlichen Grades ihrer Erfüllung weiter“ (Terhart 2006: 31).

Eine Kompetenzorientierung in der Praxis der Lehrerbildung, wie sie hier in dieser Arbeit verstanden wird, hat keine Revolution der Ausbildungsprozesse im Sinn, bei der die Ausbildung von Fähigkeiten oder Kompetenzen auf Seiten der Lehrer, möglichst lernwirksamen Unterricht anzuleiten, um ‚optimale Lernergebnisse‘ auf Seiten der Schüler zu erzeugen, alle anderen Ausbildungsziele und -aspekte in den Hintergrund drängt. Es geht dabei auch nicht um einen „[...] radikalen Bruch mit der Vergangenheit traditioneller Lernzielbestimmungen“ (Ladenthin 2006: 246), sondern um die Entwicklung eines Instrumentariums, das sich auf basale Fähigkeiten des Unterrichtens stützt, die für den Praktiker von hoher Relevanz sind, wenn dieser ‚guten Unterricht‘ (im oben definierten Sinne⁷²) anleiten will. Die Fundierung dieses Instrumentariums wird systematisch entwickelt und will somit Transparenz über die Bewertungskriterien herstellen, sowohl auf einer inhaltlichen Ebene, mit Hilfe der Qualitätsbereiche guten Unterrichts, als auch auf der Ebene der Bewertungsmaßstäbe, mit Hilfe der Kompetenzmodellierungen, die in der Praxis bisher offensichtlich zu kurz gekommen ist⁷³.

Allerdings muss an dieser Stelle auch angemerkt werden, dass eine im strengen, wörtlichen Sinne gemeinte Umsetzung der Kompetenzorientierung und der formulierten Standards der Lehrerbildung in der Tat den von den Kritikern be-

⁷² Siehe Kapitel 2.1.

⁷³ Siehe Kapitel 1.2 und 1.3.

fürchteten Totalitarismusverdacht nähren würde. Überträgt man die Entwicklungen der standardbasierten Bildungsreform, wie sie im schulischen Bildungsbereich zu beobachten ist, erwähnt seien die Lernstandserhebungen, zentrale Abschlussprüfungen und das Zentralabitur, auf die Lehrerbildung, so muss die Frage aufgeworfen werden, inwiefern eine hoch individualisierte und auch mit Persönlichkeitsmerkmalen wie etwa Empathie, verknüpfte Tätigkeit durch standardisierte Test ‚abgeprüft‘ werden kann⁷⁴. Eine noch so durchdachte und testtheoretisch abgesicherte summative Diagnostik wird die alltägliche Tätigkeit des Lehrers und seine Eignung, diese dauerhaft erfolgreich und gewinnbringend für alle Beteiligten auszuführen, nicht erfassen können. Eine bundes- oder landesweite, standardisierte Abschlussprüfung der Zweiten Phase der Lehrerbildung würde den Anforderungen, denen Lehrer sich gegenübersehen, um ihre Rolle positiv auszufüllen, nicht gerecht werden. ‚Lehrersein‘, so muss es hier als *Behauptung* stehen bleiben, ist ‚übersummativ‘. Die Füllung dieser Rolle beinhaltet mehr als die erfolgreiche Aneignung zentraler Kompetenzen, wie der diagnostischen Kompetenz, Beratungskompetenz, Unterrichtskompetenz, etc.. Es geht auch um das Erlangen ‚menschlicher Qualitäten‘, das wertschätzende Zuwenden an die erzieherisch Anvertrauten – und das alles vor dem Hintergrund der eigenen Persönlichkeit, der individuellen Geschichte, den eigenen Ängsten und Unzulänglichkeiten. Diese wichtigen, die Tätigkeit des Lehrers in ihrem inneren Kern berührenden Aspekte lassen sich nicht standardisiert abprüfen, aber es ist immens wichtig, sie als Teil des ‚Lehrerwerdens‘ zu verstehen und in die Ausbildung einzubeziehen. Ähnlich sehen dies auch Oser u.a., wenn sie im Kontext der Qualitätsmessung von Lehrerkompetenzen schreiben:

„Ebenfalls verweisen wir auf substanzielle nicht messbare Handlungsteile dieses Berufs, die durch Intuition zustande kommen und kaum quantitativ abgesichert werden sollen“ (Oser et al. 2010: 133).

⁷⁴ Womit an dieser Stelle nicht festgestellt werden soll, dass die Umsetzung der standardbasierten Bildungsreform im schulischen Bereich als totalitär zu bezeichnen ist. Diese Auseinandersetzung kann hier nicht geführt werden, sondern lediglich auf die unterschiedlichen Kontextbedingungen des schulischen Lernens und der Lehrerausbildung verwiesen werden.

Eine ins Extrem getriebene Standardisierung der Ausbildung und v.a. der Bewertung kann diese nicht einfangen und müsste zurecht als ‚totalitär‘ gebrandmarkt werden. In diesem Sinne hat das hier zu entwickelnde Bewertungsinstrumentarium, wie oben erläutert⁷⁵, nicht das Ziel, eine Bewertung darüber abzugeben, ob jemand ‚ein guter Lehrer‘ ist, sondern lediglich, ob eine Entwicklung zentraler Teilkompetenzen einer allgemeinen Unterrichtskompetenz stattgefunden hat.

Es bietet sich weitergehend an, die Frage zu thematisieren, inwiefern ein solch ‚totalitäres Regime‘ der Standardisierung und Kompetenzorientierung im Bereich der Lehrerbildung zu befürchten ist. Oelkers mahnt in Bezug auf die Befürchtungen der Kritiker der standardbasierten Bildungsreform hinsichtlich ihrer Auswirkungen im schulischen Bereich, zu mehr Gelassenheit in der Diskussion (vgl. Oelkers: 2006). Die Gefahr, dass diese Reform tatsächlich totalitäre Auswirkungen entfalte, sei vor dem Hintergrund der starken Eigenlogik von Schulen gering bis nicht existent. Die Intention jedweder Bildungsreform habe sich noch nie eins zu eins in die Praxis übertragen lassen, vielmehr folgten Schulen

„[...] nicht einer Idee der Bildung, sondern ihrer eigenen Praxis, die sich nicht in eine tabula rasa der Reform verwandeln kann. [...] Der wahrscheinlichste Effekt einer Bildungsreform ist, dass mehr Papier produziert wird“ (ders.: 107).

Oelkers führt weiter aus, dass die Aspekte der standardbasierten Bildungsreform, wie etwa die Einrichtung einer systematischen Qualitätssicherung, welche den Erreichungsgrad von Zielen der Institutionen des Bildungssystems erfasse, schlichtweg als die notwendigen Versuche der Verwaltung zu verstehen sind, in einem System, welches auf Steuerungsversuche hochgradig „listig“ reagiere, also einfallsreich darin ist, sich gegen Steuerungsversuche zu behaupten, ein Mindestmaß an externer Steuerung zu realisieren.

Die Frage der Verwaltung, welchen ‚Sinn‘ das in der Schule Gelernte für die Lernenden und die Gesellschaft später einmal habe, und die Überprüfung, wie

⁷⁵ Siehe Kapitel 1.4.

erfolgreich die Institutionen bei der Erreichung dieser Ziele sind, sei keine totalitäre Vereinnahmung des Bildungssystems:

„Es ist ein Irrtum, das Lernen in der Schule ohne anschließende Verwertbarkeit zu kalkulieren. Die Frage, was mit dem Gelernten später angefangen werden soll, ist nicht krude und lästig, sondern muss beantwortet werden“ (ders.: 111).

Die Befürchtungen, es entspanne sich um das Bildungssystem im Rahmen der standardbasierten Bildungsreform ein totalitäres Regime, seien deswegen unbegründet.

„Das System ist, nochmals, robust und hat schon andere Reformattacken weitgehend unbeschadet überstanden“ (ders.: 113).

Auch die Ausführungen von Bellmann und Waldow deuten darauf hin, dass derartige Befürchtungen eher unbegründet sind. Sie weisen nach, dass Kernelemente der standardbasierten Bildungsreform keineswegs so neu sind, wie sie bisweilen im aktuellen wissenschaftlichen Diskurs dargestellt werden, sondern über eine lange erziehungswissenschaftliche Tradition verfügen (vgl. Bellmann 2012: 144–145) und in mehreren Reformwellen versucht wurde, diese im Bildungssystem, v.a. im jenen der USA, zu implementieren (ders.: 148), wobei sich das Bildungssystem als relativ resistent gegenüber einer direkten Umsetzung der Intentionen des sog. ‚Social Efficiency Movements‘ zeigte (vgl. ders.: 156, Waldow 2012: 163).

Bezieht man den obigen Gedankengang von Oelkers auf die Lehrerbildung, lassen sich gute Argumente anführen, warum ein totalitäres, auf die unkritische Umsetzung von Standards fixiertes, Bildungsregime auch in der Lehrerbildung nicht droht. Wie in Kapitel 1.2 gezeigt wurde, ist die Ausbildungssituation für die Auszubildenden durch ein hohes Maß an Abhängigkeit von ihren Ausbildern gekennzeichnet und der Grad an Autonomie, mit der die Ausbilder agieren, ist als sehr hoch einzustufen. Gerade dieses Zusammenspiel von *hoher Autonomie der einzelnen Ausbilder*, bei gleichzeitig *starker Abhängigkeit der Auszubildenden von verschiedenen Ausbildern* mit individuellen Wünschen und

Vorstellungen ‚richtigen‘ pädagogischen Handelns, ist einer der, wenn nicht überhaupt der zentrale Stressor in der Zweiten Phasen der Lehrerausbildung. Ausgehend von der These, dass das System der Lehrerausbildung und die in ihm handelnden Akteure ähnlich ‚listig‘ auf ‚Reformattacken‘ reagieren wie die schulische Bildung, ist *erstens* nicht zu befürchten, dass die Umstellung der Ausbildung auf eine Kompetenzorientierung zu einer revolutionären Umwälzung der Ausbildungsverhältnisse in Richtung eines totalitären Regimes führen wird. Vielmehr ist davon auszugehen, dass die ‚etablierten Problemlösungen‘ des Ausbildungsalltages, die sich die Ausbilder im Laufe der Jahre angeeignet haben, über ein erhebliches Maß an Stabilität verfügen und diese nicht ohne weiteres verändert werden können. *Zweitens*, so die hier vertretene Position, bedeutet eine Standardisierung der Ausbildung mit Hilfe von klar operationalisierten Kompetenzdefinitionen, welche Grundlage der Bewertung sind, nicht eine zunehmende Unfreiheit in der Ausbildung, sondern diese beinhaltet für die Auszubildenden vielmehr ein gewisses emanzipatives Potential, sich aus den als unproduktiv empfundenen Abhängigkeitsverhältnissen zu befreien. Schriftlich fixierte Bewertungsgrundlagen können dazu dienen, einen für alle an der Ausbildung Beteiligten geltenden Kanon von Bewertungskriterien zu formulieren, der die von Auszubildenden beklagte Willkür in der Bewertung zumindest teilweise entschärft.

Aus diesen Überlegungen können zwei Thesen abgeleitet werden, die im Rahmen des empirischen Teils der Arbeit überprüft werden sollen:

- Die von den Ausbildern entwickelten Routinen der Bewertung und Beratung werden sich durch das zu erprobende Bewertungsinstrument nur eingeschränkt verändern lassen. Es wird sich ‚Widerstand‘ gegen den Veränderungsprozess zeigen⁷⁶.
- Die transparente Operationalisierung von zentralen Teilkompetenzen einer allgemeinen Unterrichtskompetenz sorgt dafür, dass die Auszubil-

⁷⁶ Siehe zur genaueren Auseinandersetzung mit dem ‚Widerstand‘ in institutionellen Veränderungsprozessen Kapitel 3.3.

denden Klarheit über die Grundlagen ihrer Bewertung haben, sodass die Klagen der an der Erprobung beteiligten Auszubildenden über mangelnde Transparenz der Bewertung geringer ausfallen sollten, als im Vergleich zu den tradierten Bewertungsverfahren.

In gewisser Weise stehen diese beiden Thesen in einem Konflikt miteinander, denn je stärker die Eigenlogik der Ausbildung ist, umso geringer ist die Steuerungswirkung eines standardisierten Instruments und damit auch dessen emanzipatorisches Potential. Die Überprüfung im Rahmen des empirischen Teils dieser Arbeit wird zeigen, ob die Eigenlogik des Ausbilderhandelns oder dem Veränderungsimpuls des Bewertungskonzept die bedeutendere Rolle zuzuschreiben ist.

2.3 Kompetenzen entwickeln – Kompetenzentwicklung erfassen: pre-Alpha-Version des Bewertungskonzepts

Die Überlegungen, die im Rahmen dieses Kapitels angestellt worden sind, münden in eine erste Version des zu entwickelnden Bewertungskonzepts, das im weiteren Verlauf des Forschungsvorhabens im Rahmen der Fokusgruppendifkussionen von Ausbildern und Auszubildenden einer kritischen Würdigung unterworfen wird. Die Instrumente, aus denen sich diese Version des Bewertungskonzepts zusammensetzt finden sich im Anhang wieder. Im Folgenden wird das Design der Instrumente, sowie deren Rolle im Rahmen der Bewertungsprozesse kurz erläutert.

2.3.1 Der Beobachtungsbogen

Die pre-Alpha-Version des Bewertungskonzepts setzt sich, wie bereits oben kurz angerissen, aus drei Instrumenten zusammen. Das erste Instrument ist ein *standardisierter Beobachtungsbogen* für die Unterrichtsbesuche, der die in Kapitel 2.1.5 ausgewählten Qualitätsbereiche und die dort formulierten Einzelmerkmale, welche den Qualitätsbereich operationalisieren, umfasst. Der Beobachtungsbogen

liegt in zwei Versionen vor, die in den Fokusgruppendifkussionen auf ihre jeweiligen Vor- und Nachteile hin befragt werden sollen. Die *Version A* bietet für die Beobachtung der Qualitätsbereiche im Unterricht eine *Kompetenzskala* an, die versucht, zentrale Beobachtungsaspekte der Einzelmerkmale in Verbindung mit einer qualitativen Abstufung zu operationalisieren⁷⁷. Auf Grund der schwierigen Aufgabe, die Kompetenzstufen so zu formulieren, dass sie deutlich voneinander abgegrenzt sind, beschränkt sich die Skala auf eine Abstufung der Kompetenz in drei Ausprägungsgrade. *Version B* verbindet die Einzelmerkmale guten Unterrichts mit einer vierstufigen Ratingskala, mit welcher die Realisierung eines bestimmten Merkmals erfasst werden soll. Beide Bögen bieten die Möglichkeit anzumerken, ob die einzelnen Indikatoren, bzw. Merkmale in der jeweiligen Unterrichtsstunde relevant waren und ob sie beobachtet wurden. Hintergrund dieser Möglichkeit ist die Idee, dass nicht jeder Aspekt guten Unterrichts in jeder Unterrichtsstunde realisiert werden muss. So wird z.B. der Indikator eins, welcher die ‚stimmige Phasierung‘ einer Unterrichtsstunde in den Blick nimmt, im offenen Unterricht von geringer Relevanz sein, wenn die Schüler tatsächlich vollkommen selbstständig arbeiten. Entsprechend ist der Vermerk, dass ein Merkmal nicht beobachtet wurde, dann nicht weiter problematisch, wenn dieses in der jeweiligen Unterrichtsstunde keine Relevanz hatte.

Grundlegende Intention des Beobachtungsbogens ist es, *Transparenz* über das den Bewertungsprozessen zu Grunde liegende Konzept guten Unterrichts herzustellen und Ausbildern und Auszubildenden eine *gemeinsame Grundlage für eine standardisierte Beobachtung* des Unterrichts und die anschließende *Unterrichtsnachbesprechung* anzubieten, auf die sich beide im Rahmen ihrer qualitativen Einschätzung der Unterrichtsstunde berufen können.

⁷⁷ In diesem Sinne handelt es sich um eine ‚pragmatische‘, wenn man so will auch ‚naive‘, Formulierung von Kompetenzstufen und damit auch der Kompetenzskala an sich. Die Rückbindung an ein Kompetenzmodell, mit dem sich die verschiedenen Kompetenzstufen theoretisch begründen ließen, wie sie für eine anspruchsvolle Kompetenzmessung notwendig wäre, fehlt in diesem Fall (siehe Kapitel 2.2).

2.3.2 Der Dokumentationsbogen für die Nachbesprechung

Das Instrumentarium wird durch den *Dokumentationsbogen für die Nachbesprechung* erweitert, mit dessen Hilfe eine strukturierte Dokumentation der in der Nachbesprechung thematisierten Aspekte vorgenommen werden soll. Er setzt sich aus verschiedenen Teilen zusammen, die für die Bewertung und Beratung der Auszubildenden einen je spezifischen Zweck verfolgen.

Die Dokumentation beginnt mit einer grundlegenden Einordnung des Unterrichts hinsichtlich zentraler didaktischer Entscheidungen, die ihren Sinn im Hinblick auf das dritte Instrument des Konzepts, die im Folgenden noch zu erklärende Sammlung über das Repertoire, erhält. Es schließen sich drei Ratingskalen an, mit denen die Reflexionsfähigkeit und die mit dieser in Verbindung stehenden Kompetenzzusammenhänge als Expertise und als Repertoire erfasst werden sollen. Grundlage für die Verortung der Kompetenz der Auszubildenden bilden ihre Ausführungen zu der eigenen Unterrichtsstunde im Kontext der Unterrichtsnachbesprechung. Abschließend werden zentrale Qualitätsbereiche guten Unterrichts sowie deren Thematisierung im Rahmen der Unterrichtsnachbesprechung in einer dafür vorgesehenen Tabelle festgehalten.

Die in dieser Art und Weise strukturierte Dokumentation der Stunde und der Nachbesprechung verfolgt *die Intention*, eine transparente und kriteriengeleitete Bilanzierung der Unterrichtsnachbesprechung vorzunehmen, die den Auszubildenden eine schriftliche fixierte Einschätzung ihres Leistungs- und Ausbildungsstandes an die Hand gibt und im Zweifelsfall zur Klärung von Missverständnissen und Konflikten mit den Ausbildern herangezogen werden kann.

2.3.3 Die Sammlung über das Repertoire

Die *Sammlung über das Repertoire* vervollständigt als drittes Instrument die pre-Alpha-Version des Bewertungskonzepts. In ihr werden zentrale Aspekte der Unterrichtsbeobachtung und -nachbesprechung festgehalten. Am Ende der Ausbildung liegt somit ein Dokument vor, das eine umfassende Bilanzierung der gezeigten Vielfalt an Unterrichtsinszenierungen vornimmt und die Entwicklung zentraler Kompetenzzusam-

menhänge bei den Auszubildenden abbildet. Folgende Elemente sollen diese Leistung des Instruments sicherstellen: Die im Dokumentationsbogen vorgenommene grundlegende Einordnung des Unterrichts soll nach jedem Unterrichtsbesuch in der Sammlung über das Repertoire vermerkt werden, sodass am Ende der Ausbildung Transparenz darüber geschaffen werden kann, wie vielfältig die Unterrichtsinszenierungen sind, welche ein Auszubildender im Rahmen seiner Ausbildung gezeigt hat und in diesem Sinne: als wie ausgeprägt sein Unterrichtsrepertoire anzusehen ist. In einer Kopie des Beobachtungsbogens⁷⁸ vermerken die Ausbilder, welche der Merkmale guten Unterrichts sie in den einzelnen Qualitätsbereichen beobachtet haben und in welcher qualitativen Ausprägung diese realisiert wurden. So wird durch diese Dokumentation erstens sichtbar, welche Merkmale guten Unterrichts oft und welche selten realisiert wurden, und zweitens lässt sich im Optimalfall über den Zeitverlauf eine Entwicklung zur Realisierung höherer Kompetenzstufen und somit ein erfolgreicher – oder eben nicht erfolgreicher – Ausbildungsverlauf aus dem Instrument ablesen. In der gleichen Art und Weise übertragen die Ausbilder ihre Einschätzungen zur Reflexionsfähigkeit der Auszubildenden in die Sammlung über das Repertoire, sodass dieses Instrument auch für diesen Kompetenzaspekt eine Bilanzierung der stattgefundenen Entwicklung auf Seiten der Auszubildenden leistet.

Der *intendierte Zugewinn* durch die Sammlung über das Repertoire liegt in einer Bilanzierung der Ausbildung ‚auf einen Blick‘, welche wiederum die Transparenz über die Einzelbewertungen und die Abschlussbewertungen für die Auszubildenden erhöhen und den Ausbildern eine Hilfestellung zur Formulierung ihres Endgutachtens bieten soll.

⁷⁸ Je nachdem, welche Version der Beobachtungsbögen bei der Unterrichtsbeobachtung eingesetzt wurde, kann hier sowohl die Version mit Kompetenz- als auch jene mit Ratingskala zum Einsatz kommen.

3 | Forschungsdesign und theoretische Fundierung des Projektes

Wie in der Einleitung ausgeführt verfolgt diese Promotion zwei Zielsetzungen: Es geht *erstens* darum, in der Ausbildungspraxis der Zweiten Phase der Lehrerbildung mit Hilfe wissenschaftlicher Methoden ein Diagnostikinstrument für den Bereich der Unterrichtskompetenz zu entwickeln, das in der Ausbildungspraxis eingesetzt werden kann, und *zweitens* um die Formulierung einer gegenstandsbezogenen Theorie der Standardisierung von Bewertungsprozessen im Bereich der Lehrerbildung. Entsprechend handelt es sich bei der vorliegenden Arbeit um praxisorientierte Forschung *in* und *zur* zweiten Phase der Lehrerbildung. Diese übergeordneten Zielsetzungen bestimmen die Einordnung des Forschungsvorhabens in die verschiedenen Forschungsparadigmen der Erziehungs- bzw. Sozialwissenschaften. Im Folgenden werden die wesentlichen methodologischen Entscheidungen des Forschungsvorhabens erläutert, die einzelnen Arbeitsschritte des Projektes in ihrer Abfolge dargestellt, sowie einige wichtige theoretische Begriffe und Konzepte, auf denen die Arbeit beruht, aufgegriffen.

3.1 Paradigmatische Einordnung des Projektes – Handlungsforschung und qualitative Sozialforschung

Versucht man das vorliegende Forschungsvorhaben dahingehend zu charakterisieren, dass man eine Einordnung in die vorherrschenden Dichotomien der Sozialforschung, wie z.B. qualitative gegen quantitative Forschung, theoriegestützte gegen theoriegenerierende, bzw. deduktive gegen induktive Verfahren, erklärende gegen verstehende Ansätze, Grundlagenforschung gegen anwendungsorientierte Forschung usw. (vgl. Bortz & Döring 2006: 140–143; 296–302, Kromrey 2006: 19–21; 33–36, Schnell et al. 2011: 92–99) vornimmt, so wird keine erhellende Darstellung des hier gewählten Forschungsdesigns möglich sein. So geht es weder allein um die Generierung

neuer Theorien, wie z.B. in den Arbeiten von Pille, Schulte und Werner-Bentke (vgl. Pille 2013, Schulte 2008, Werner-Bentke 2010) noch um die standardisierte Überprüfung bestehender Theorien zur Zweiten Phase der Lehrerbildung.

Der Gegensatz von Grundlagen- und anwendungsorientierter Forschung scheint noch am ehesten geeignet zu sein, das hier zu Grunde gelegte Forschungsparadigma zu beschreiben. *Am Beginn und am Ende des Forschungsprozesses steht die Ausbildungspraxis. Sie ist Ausgangspunkt der Forschung, ihre Veränderung ist das Ziel des Projektes.* Die Ausbildungspraxis wurde durch die in Kapitel eins herangezogenen Studien empirisch erfasst, sodass aus der Perspektive der theoriegenerierenden, rekonstruktiven Studien eine Defizitanalyse der derzeitigen Bewertungspraktiken in der zweiten Phase der Lehrerbildung erfolgte. Diese Defizitanalyse wurde erweitert um eine Auseinandersetzung mit den Ansprüchen der Erziehungswissenschaft, wie eine zeitgemäße Bewertung des Unterrichts in der Zweiten Phase der Lehrerbildung auszusehen habe. Basierend auf dieser Defizitanalyse wurden am Ende von Kapitel eins Ansprüche an das zu entwickelnde Bewertungsinstrumentarium abgeleitet, das im Folgenden auf Grundlage dieser Ansprüche entwickelt wurde⁷⁹. An die Konzeptentwicklung schloss sich die Hauptphase des Forschungsprojektes an, nämlich die Erprobung der Bewertungsinstrumente. Hierbei wurden diese in der konkreten Ausbildungspraxis eingesetzt und von den Betroffenen, den Ausbildern und den Auszubildenden hinsichtlich der Frage bewertet, ob diese Instrumente dabei helfen, die Bewertungsprozesse der Unterrichtsbesuche produktiv zu gestalten. Für diesen Test nutzte das Projekt Methoden der qualitativen Sozialforschung⁸⁰. Die Instrumente wurden auf Basis der kritischen Würdigung der Praktiker überarbeitet und bilden die Grundlage für ein in der Praxis einsetzbares Bewertungskonzept, sodass sich der Zirkel von der Ausbildungspraxis zur Ausbildungspraxis geschlossen hat.

Diese Art einen Forschungsprozess zu gestalten, kann am ehesten mit der Bezeichnung der *Handlungs- oder Aktionsforschung* bezeichnet werden. Zentra-

⁷⁹ Siehe Kapitel 2.

⁸⁰ Siehe dazu ausführlicher Kapitel 3.2.

les Merkmal dieses Forschungsansatzes ist die konkrete Veränderung der Praxis des untersuchten Gegenstandsbereichs sozialer Phänomene (vgl. Altrichter & Feindt 2011: 214, Bortz & Döring 2006: 343). Die Bezeichnungen von Forschungsansätzen, die diesem grundsätzlichen Paradigma folgen, sind, wie so oft, sehr vielfältig. Altrichter und Feindt sprechen von einer „[...] «Familie von Forschungsansätzen», der auch Konzepte wie Handlungsforschung, Praxisforschung, Teamforschung und Schulbegleitforschung, bzw. im englischen Sprachraum Teacher Research, (Participatory) Action Research und Practitioner Research zugerechnet werden können“ (Altrichter & Feindt 2011: 214)⁸¹. Praktiker emanzipieren sich im Rahmen dieser Forschungsansätze von ihren tradierten Rollen als ‚Anwender‘ von Vorgaben oder Theorien hin zu ‚(Er-)Forscher(n) der Praxis‘. Im Rahmen von Schulforschung werden Lehrer so zu ‚Lehrerforschern‘. Die paradigmatische Sichtweise einer solchen Wissenschaftsauffassung im Bereich der Schulforschung fasst Stenhouse wie folgt zusammen:

„The mistake is to see the classroom as a place to apply laboratory findings rather than as a place to refute or confirm them“ (Stenhouse, zitiert nach Altrichter & Feindt 2011: 217).

Im Rahmen der Bildungsforschung erlangen Ansätze der Aktionsforschung u.a. Bedeutung für die Schul- bzw. allgemein für die Institutionsentwicklung (vgl. Altrichter & Feindt 2011: 219), womit einer erster Hinweis für ihre Eignung für das hier bearbeitete Projekt gegeben ist, da es ja das erklärte Ziel ist, aus der Ausbildungspraxis heraus Bewertungsinstrumente zu entwickeln, die wissenschaftlichen Standards genügen und später in der Praxis eingesetzt werden können. Kurzum: Es geht um die Fortentwicklung wichtiger Aspekte der institutionellen Logik der zweiten Phase der Lehrerbildung.

⁸¹ Vgl. zur Tradition der Practitioner Research im englischen Sprachraum z.B. Borko et al 2008: 1029–1033, zur aktuellen Anwendung der Practitioner Research in verschiedenen nationalen Kontexten Saleh & Khine 2011. Ausführungen zu den verschiedenen Ansätzen der Forschung zum Lehrerberuf im deutschsprachigen Raum finden sich z.B. bei Tillmann 2011, aus Perspektive der englischsprachigen Forschung bei Borko et al. 2008, sowie in weiteren Artikeln aus dem Sammelband von Cochran-Smith et al. 2008, auf den Seiten 1005–1193.

Vergleicht man diese Arbeit mit den Charakteristika von Aktionsforschungsansätzen, so zeigen sich deutlich die Gemeinsamkeiten. Lehmeier führt mit Rückbezug auf Benner drei Merkmale der Forschungsmethodologie einer pädagogischen Handlungsforschung aus:

- „Öffnung gegenüber dem Objekt der Forschung als Subjekt,
- Offenheit und prinzipielle Unabschließbarkeit des Forschungsprozesses [...],
- Ermöglichung interindividueller Verständigung über Sein (dialektische Situationsanalyse) und Sollen (dialektische Sinndeutung) menschlicher Praxis mit dem Ziel ihrer vernünftigen Planung und Veränderung“ (Lehmeier 1995: 632).

An Hand der drei von Lehmeier definierten Merkmale einer pädagogischen Handlungsforschung kann gezeigt werden, dass das hier gewählte Forschungsdesign im Wesentlichen diesem methodologischen Paradigma entspricht. Erstens legt das Forschungsdesign dieser Arbeit einen starken Fokus auf die Einbindung der betroffenen Akteure in die Entwicklung der Bewertungsinstrumente. Sowohl die Ausbilder als auch die Auszubildenden sollen mit ihren Erfahrungen und ihrer ‚praktischen Expertise‘ die konkrete Ausgestaltung der Instrumente mitbestimmen. In diesem Sinne werden sie ein gutes Stück von ihrer Rolle als ‚Forschungsobjekt‘ in Richtung eines ‚Forschungssubjektes‘ emanzipiert. Zweitens versteht sich der Forschungsprozess dieses Projektes als ein *pragmatisch* abgegrenztes Unterfangen, sowohl räumlich als auch zeitlich. Die Projekterprobung und anschließende Auswertung des Projektes erstreckte sich über einen Zeitraum von etwa zwei Jahren und erfasste nur das ZfsL Bonn. Diese Einschränkungen waren v.a. den begrenzten, für eine Promotion zur Verfügung stehenden Ressourcen geschuldet. Das Projekt dann für abgeschlossen zu erklären, wenn die aus den qualitativen Daten generierten Rückmeldungen in die Bewertungsinstrumente eingearbeitet sind, ermöglicht ein schlüssiges Ende des Forschungsprozesses, aber es wäre ebenso möglich, und auch wünschenswert, die Ergebnisse und die gewonnenen Instrumente durch die Ausweitung des Forschungsprozesses in weite-

ren Schritten zu verfeinern. Denkbar wäre etwa, die entwickelten Instrumente in anderen Studienseminaren einzusetzen und ihre Rezeption dort zu evaluieren oder zu überprüfen, wie sich die Rezeption der Instrumente in einem längeren Zeitverlauf am ZsFL Bonn verändert. In diesem Sinne kann auch der hier vorgestellte Forschungsprozess als ‚prinzipiell unabschließbar‘ angesehen werden. Drittens steht, wie bereits mehrfach erwähnt, die Veränderung der pädagogischen Praxis im Mittelpunkt dieses Projektes. Einerseits die konkrete Praxis im ZfsL Bonn, andererseits aber auch die Praxis der Lehrerbildung generell. Die Ergebnisse dieses Forschungsvorhabens sollen auf einer Internetseite⁸² veröffentlicht werden und so die Ausbildung in der Zweiten Phase der Lehrerbildung auf möglichst breiter Ebene positiv beeinflussen oder doch wenigstens Diskussionsimpulse zur Weiterentwicklung der Bewertungspraxis geben.

Altrichter und Feindt formulieren ihrerseits sieben Merkmale, die allen Ansätzen der Aktionsforschung gemein seien. Neben den drei von Lehmeier formulierten Merkmalen ergänzen Sie folgende (vgl. Altrichter und Feindt 2011: 215):

- Die Verbindung von Forschung und Entwicklung in einem Prozess
- Einbeziehung verschiedener Perspektiven in den Forschungsprozess, um eine möglichst vielfältige Sichtweise auf die untersuchten sozialen Phänomene zu erhalten
- Ausbildung von ‚professional communities‘ im Rahmen der Forschungsarbeit, sodass das mit dem Forschungsprojekt verbundene Lernen in einen sozialen Kontext eingebunden ist
- Nutzung anderer Gütekriterien als Objektivität, Reliabilität und Validität

Auch diese Merkmale werden im hier durchgeführten Forschungsprojekt berücksichtigt. Die Erforschung der Bewertungsprozesse in der Zweiten Phase der

⁸² www.unterrichtskompetenz.de

Lehrerbildung dient der Entwicklung neuer, praxistauglicher Bewertungsinstrumente, sodass der Prozess zugleich Forschungs- als auch Entwicklungsprozess ist. Das methodologische Setting der Arbeit bezieht die Perspektiven von verschiedenen, an der Ausbildung beteiligten Personengruppen mit ein, um möglichst viele Facetten der Herausforderungen einer ‚guten‘ Bewertung in der zweiten Phase der Lehrerausbildung in die Entwicklung der Instrumente einfließen lassen zu können. Wie in Kapitel 2.2.2 gezeigt, beruhen die Bewertungsinstrumente auf anderen Gütekriterien als jenen der klassischen Testtheorie, nämlich der kommunikativen und ökologischen Validierung, dem Realitätsgehalt und der Intersubjektivität.

Mit diesen Ausführungen ist allerdings auch nur ein Teil des Forschungsprojektes zufriedenstellend beschrieben, nämlich jener, der der Zielsetzung folgt, ein praxistaugliches Bewertungsinstrument für die Unterrichtsbesuche im Rahmen der Zweiten Phase der Lehrerbildung zu entwickeln. Die *Zielsetzung*, aus den erhobenen Daten eine *gegenstandsbezogene Theorie der Standardisierung von Bewertungsprozessen* in der Lehrerbildung, zumindest in einigen Grundzügen, zu entwickeln, ist nicht mit dem Forschungsparadigma der Aktionsforschung zu umschreiben. Vielmehr wird der Aktionsforschungsansatz dieser Arbeit hier durch einen klassischen Ansatz der qualitativen, bzw. rekonstruktiven Sozialforschung ergänzt. Unter der Perspektive der *Grounded Theory* geht es um die Formulierung einer Theorie der Standardisierung im hier diskutierten Kontext auf Basis der *induktiv bzw. im induktiv-deduktiven Wechselspiel* gewonnenen Daten über die Wirkung von standardisierten Bewertungsprozessen (vgl. Lamnek 2005b: 100–117). Ziel ist es, die Grundlagen einer, im Sinne von Glaser und Strauss, ‚angemessenen‘ Theorie der Standardisierung zu entwickeln, welche auf der Ebene des konkreten *Gegenstandsbereiches der Unterrichtsbesuche* in der Zweiten Phase der Lehrerbildung „[...] zutreffende Vorhersagen, Erklärungen, Interpretationen und Anwendungsmöglichkeiten“ (ders.: 102) über die Auswirkungen von bzw. für die standardisierten Bewertungsprozesse(n) abliefern kann. Eine systematische Überprüfung, welche Auswirkungen die ‚standardbasierte Bildungsreform‘ aus der Perspektive der betroffenen Akteure hat, ist bislang kaum diskutiert worden. Die *Grounded Theory* als ein „[...] Forschungsstil zur Erarbeitung von in empirischen Daten gegründeten Theorien [...]“ (Strübing 2008: 14) bietet sich an, um diese For-

schungslücke, zumindest im Bereich der Lehrerbildung, zu verkleinern. Die Wirkungen standardisierter Bewertungsprozesse innerhalb des konkreten hier untersuchten Kontextes, die auf einem induktiven Weg gewonnen werden, bilden eine wichtige Ergänzung zu der aus den Daten von PISA erfolgten Deduktion, dass standardbasierte Qualitätssicherung zu einer Steigerung der Bildungsqualität führe. Hierbei ist wichtig zu erwähnen, dass sich das Forschungsdesign nicht streng an den Vorgaben der Grounded Theory als Forschungsmethodologie orientiert (vgl. dazu z.B. Lamnek 2005b: 100–117; Strübing 2008), aber zentrale Elemente von eben dieser integriert, v.a. das induktiv-deduktive Wechselspiel von Datenerhebung und -auswertung in den verschiedenen Phasen des Forschungsprojektes sowie die Art der Datenanalyse (vgl. Strübing 2008: 14–36).⁸³ Die so gewonnene gegenstandsbezogene Theorie kann einen Baustein einer allgemeinen im Sinne der Grounded Theory, *formalen Theorie* der Standardisierung von Bewertungsprozessen im Bildungsbereich bilden, wobei der Schritt von der gegenstandsbezogenen zur formalen Theorie nicht Teil dieser Promotion sein kann.

Das besondere Forschungsdesign dieser Arbeit, v.a. die konkrete Art der Einbettung in den Praxiskontext, bedingen allerdings, dass die Übereinstimmungen des Projektes mit den beiden Bezugsparadigmen, der Handlungsforschung und der qualitativen Sozialforschung, teilweise durchbrochen werden und *keine ‚orthodoxe‘ Umsetzung der Methodologien* dieser Ansätze erfolgt. Hinsichtlich der konstituierenden Merkmale der Handlungs- bzw. Aktionsforschung müssen Abstriche in Bezug auf die Umsetzung des Ideals dieses Forschungsparadigmas hingenommen werden. Die *Einbindung der Akteure in den Forschungsprozess* endet hier bei dem Einbezug ihrer subjektiven Wertungen in die Entwicklung der Bewertungsinstrumente. Weitreichende Ansätze der Aktionsforschung gehen so weit, auch den Interpretationsvorgang von im Zuge des Forschungsprozesses gewonnenen Daten gemeinsam mit den ‚Forschungssubjekten‘ auszuwerten (vgl. Bortz & Döring: 342). Die *prinzipielle Offenheit des Forschungsprozesses* erfährt Abstriche, wenn das Forschungsprojekt mit der Veröffentlichung als abgeschlossen zu gelten hat. Die Bewertungsinstrumente sind entwickelt und sollen Ein-

⁸³ Siehe dazu Kapitel 3.2.3.

gang in die Ausbildungspraxis finden. Ob sie hier weiteren Anpassungs- und Veränderungsprozessen unterworfen sind und wie diese sich zu den ursprünglich formulierten Zielen des Projektes verhalten, entzieht sich reflektierenden Kontrolle. Schlussendlich müssen die *Möglichkeiten neuer Bewertungsinstrumente*, die in Kapitel eins umschriebenen *Probleme der Bewertungsprozesse der Zweiten Phase der Lehrerbildung zu verbessern, realistisch eingeschätzt* werden. Institutionelle Logiken, die Entwicklungsprozesse blockieren, und zwischenmenschliche Konflikte zwischen Auszubildenden und Ausbildern lassen sich nicht durch ein noch so durchdachtes Instrument aus der Welt schaffen, sodass die Möglichkeiten dieses Projektes, die pädagogische Praxis zu verändern, nicht überschätzt werden sollten. Darüber hinaus geschieht die *Ausbildung von ‚professional communities‘*, die den Forschungsprozess begleiten, nur in einem begrenzten Rahmen. In den Gruppendiskussionen⁸⁴ bilden sich nur temporäre Gruppen, die nach der Diskussion ‚zerfallen‘. Eine kontinuierliche Begleitung des Projektes durch eine Stakeholdergruppe der erforschten Institution findet nicht statt, lediglich eine Auswahl von Ausbildern begleitet das Projekt durch die gesamte Eprobungsphase. Zudem ist der Autor der Arbeit selbst ein ‚Outsider‘. Er ist nicht Teil des Kollegiums des ZfsL; er kennt es lediglich aus seiner eigenen Ausbildung. Wie weiter unten zu zeigen sein wird⁸⁵, ist die Güte der im Zuge des Projektes erhobenen qualitativen Daten eingeschränkt, was v.a. mit problematischen Rekrutierungsprozessen der Forschungsteilnehmer zu tun hat. Damit in Verbindung steht eine nur eingeschränkte Umsetzung der Methodologie der Grounded Theory, v.a. das Fehlen eines qualitativen Samplings und des damit in Verbindung stehenden stetigen Wechsels zwischen Datenerhebung und -auswertung⁸⁶.

Zusammenfassend kann also festgehalten werden, dass sich das Forschungsprojekt in den methodologischen Horizont der Aktionsforschung sowie der qualitativen Sozialforschung, genauer der Grounded Theory, stellt, dies aber mit

⁸⁴ Siehe Kapitel 3.2.1.

⁸⁵ Siehe Kapitel 3.2.1.

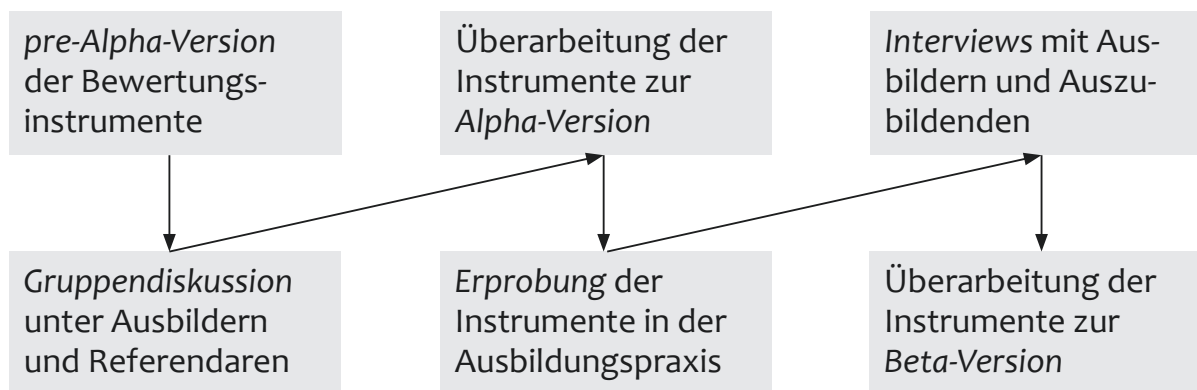
⁸⁶ Siehe Kapitel 3.2.3.

gewissen Abstrichen hinsichtlich der Ansprüche einer ‚orthodoxen‘ Umsetzung dieser Ansätze⁸⁷. Die Arbeit wählt demnach einen ähnlichen Forschungsansatz, wie ihn Bovet und Frommer in einem Forschungsprojekt am Studienseminar Rottweil verfolgten (vgl. Bovet & Frommer 2009: 20–22) und für weitere Forschungsprojekte in der zweiten Phase der Lehrerbildung fordern.

3.2 Projektablauf und Forschungsmethoden

Im Zentrum des Forschungsprojektes steht die Erprobung der neu entwickelten Bewertungsinstrumente in der Ausbildungspraxis, um zu überprüfen, inwieweit diese den Ansprüchen der Stakeholder der Ausbildung (Ausbilder und Auszubildende) gerecht werden. Abbildung 3.1 stellt die Abfolge der einzelnen Erprobungsschritte dar:

Abbildung 3.1: Ablauf der Erprobungsphase



Den Ausgangspunkt der praktischen Erprobung des Bewertungskonzeptes bildet die in Kapitel 2 entwickelte *pre-Alpha-Version* des Bewertungskonzeptes. Diese Version wird in *Gruppendifkussionen* von Ausbildern und Auszubildenden einer kritischen Würdigung, hinsichtlich ihres Nutzens für produktive Be-

⁸⁷ Schulte führt in ähnlicher Art und Weise aus, dass sie in ihrem im Rahmen der Promotion realisierten Forschungsprojekt zur Perspektive der Referendare auf die Ausbildung in der Zweiten Phase der Lehrerbildung auch nur in „[...] Anlehnung an die ethnographische Methode [...]“ vorgegangen ist und diese nicht streng umsetzen konnte (Schulte 2008: 43).

wertungsprozess unterworfen. Auf der Basis dieses Feedbacks wird das Bewertungskonzept zur *Alpha-Version* fortentwickelt, die in der Ausbildungspraxis *erprobt* wird. Mit Hilfe von *leitfadengestützten, problemzentrierten Interviews* wird die Rezeption des Bewertungskonzeptes durch die Ausbilder und Auszubildenden im Anschluss an die Erprobung erhoben. Basierend auf der Auswertung dieser Interviews wird das Bewertungskonzept zur *Beta-Version* überarbeitet, welche auch die Abschlussversion des Bewertungskonzeptes darstellt. Der Einsatz der Forschungsmethoden, wie sie in den einzelnen Phasen des Projektes zum Einsatz kommen, sowie die Begründung ihrer Auswahl sollen im Folgenden erläutert werden.

3.2.1 Gruppendiskussionen unter den Ausbildern und Auszubildenden

In einem ersten Schritt der Konzepterprobung sollte die pre-Alpha-Version des Bewertungskonzeptes von Ausbildern und ehemaligen Auszubildenden⁸⁸ in Gruppendiskussionen einer kritischen Würdigung unterzogen werden. Dieser erste Erprobungsschritt verfolgte drei Zielsetzungen:

- Erstens sollten erfahrene Ausbilder die Bewertungsinstrumente vor dem eigentlichen Einsatz in der Ausbildungspraxis hinsichtlich ihrer Einsatzfähigkeit begutachten, um vermeidbare Fehler schon im Vorfeld zu verbessern und die eigentliche Erprobung möglichst reibungslos ablaufen zu lassen.
- Zweitens diente die Diskussion unter ehemaligen Auszubildenden dazu auszuloten, inwieweit die Bewertungsinstrumente dazu geeignet scheinen, Verbesserungen in den in Kapitel 1 umrissenen Problemfeldern der Bewertung herbeizuführen.
- Drittens dienten die Gruppendiskussionen zur Erhebung von Daten bezüglich der Frage, welchen Beitrag eine Standardisierung für gewinnbringende Bewertungsprozesse im Sinne einer formativen Evaluation leisten

⁸⁸ Hierbei handelte es sich um ehemaligen Referendare, deren zweites Staatsexamen zwischen einem und zwölf Monaten zurücklag.

kann, um darauf aufbauend eine gegenstandsbezogene Theorie der Standardisierung von Bewertungsprozessen in der Lehrerbildung formulieren zu können.

Die Gruppendiskussion bot sich aus verschiedenen Gründen als Methode für diesen Projektschritt an. In der Literatur werden als besondere Stärken der Gruppendiskussion gegenüber anderen Formen der qualitativen Datenerhebung ihr diskursiver Charakter und die damit einhergehende Möglichkeit, die Genese von Meinungsbildungsprozessen nachzuvollziehen, angeführt (vgl. Kühn & Koschel 2011: 36, Lamnek 2005a: 70). Kühn und Koschel fassen diese besondere Fähigkeit der Gruppendiskussion wie folgt zusammen:

„Gruppendiskussionen sind hervorragend dafür geeignet, komplexe Einstellungs-, Wahrnehmungs-, Gefühls-, Bedürfnis-, Orientierungs- und Motivationsgeflechte von Menschen und Gruppen aus bestimmten sozialen Kontexten zu explorieren“ (Kühn & Koschel 2011: 33).

Im Gegensatz zu Einzelinterviews erscheint die Gruppendiskussion für die erste Annäherung an die Rezeption der Bewertungsinstrumente geeigneter, weil gerade die unterschiedlichen Ansichten zwischen Ausbildern und Auszubildenden mit verschiedenen fachlichen Hintergründen, mit verschiedenen Persönlichkeitsmerkmalen und Ausbildungserfahrungen hinsichtlich der Einsatzfähigkeit der Instrumente eine besondere Rolle spielen, wenn es darum geht ein Instrument zu entwickeln, das in einem gesamten Studienseminar eingesetzt werden kann. Die Gruppendiskussionen ermöglichen es, die (fachspezifischen) Einwände einzelner Personen durch andere Angehörige der Institution diskutieren zu lassen und schon in der Gruppendiskussion nach möglichen Lösungen zu suchen oder aber die Einwände entkräften zu lassen. Die Gruppen bilden die Institution des Studienseminars ‚im Kleinen‘ ab, sodass gerade auch gruppendynamische Effekte hinsichtlich der Akzeptanz und Rezeption der Instrumente Berücksichtigung finden. Zeigt sich, dass es innerhalb einer Gruppe nicht möglich ist zu einem Konsens über die Einsatzfähigkeit der Bewertungsinstrumente zu kommen, dann muss es als fragwürdig angesehen werden, ob ein solcher Konsens auf der Ebene eines Studienseminars hergestellt werden

könnte. Dieses Erkenntnispotential würde beim Einsatz von Einzelinterviews statt Gruppendiskussionen nicht genutzt werden.

Die Gruppendiskussion wird weiterhin als Methode angesehen, die sich dazu eignet Forschungsinstrumente einem Pretest zu unterziehen bzw. die Erfolgsaussichten neuer Konzepte, Dienstleistungen Produkte auszuloten (Concept-Screening) (vgl. Kühn & Koschel 2011: 40, Lamnek 2005a: 72). Nun handelt es sich bei den hier zu testenden Instrumenten weder um Forschungsinstrumente im engeren Sinne, wie standardisierte Fragebögen für eine quantitative Untersuchung, noch um Produkte oder Dienstleistungen die auf dem freien Markt angeboten werden, sondern um Bewertungsinstrumente, die in klar abgegrenzten institutionellen Kontexten eingesetzt werden sollen. Gerade deswegen bieten Gruppendiskussionen, die von Mitgliedern eben dieser Institution durchgeführt werden, ein hohes Potential, mögliche Fallstricke auf dem Weg zur Umsetzung im Vorfeld zu erkennen und ggf. durch Änderungen zu entschärfen. Auch hier bietet die ‚kollektive Intelligenz‘ der Gruppe einen Erkenntnisvorteil gegenüber Einzelinterviews, in denen die Impulse, die durch die Beiträge anderer Gruppenmitglieder ausgelöst werden, ausbleiben.

Bei der Frage nach der Anzahl der Gruppendiskussionen, die zu einer bestimmten Fragestellung durchgeführt werden sollen, verweisen Kuhn und Köschel auf das Prinzip der Saturiertheit im Sinne der Grounded Theory (vgl. Kuhn & Köschel 2011: 83–85). In diesem Sinne gelte es, keine bestimmte Anzahl an Gruppendiskussionen im Vorfeld einer Studie festzulegen, sondern so lange weitere Diskussionen durchzuführen, wie diese neue Erkenntnisse generieren würden (vgl. dies.: 83). Ausgehend von ersten Gruppendiskussionen und deren Auswertungen, gilt es im Kontext qualitativer Sozialforschung ein *qualitatives Sampling* durchzuführen. Dies dient dazu, jene Merkmale und Kategorien auszumachen, welche für die Unterschiede im Forschungsfeld verantwortlich sind (vgl. Kelle & Kluge 2010: 41–55), um die Heterogenität des Forschungsfeldes möglichst breit abzubilden und so zu Daten einer hohen Güte zu kommen (vgl. dies.: 41–42). Hierbei gilt es, den Konflikt zwischen der möglichst umfangreichen Abbildung des heterogenen Forschungsfeldes und der entstehenden Datenmenge auszutarie-

ren. Bezüglich der Auswahl von Teilnehmern für die Diskussionen geben Kuhn und Köschel entsprechend die Frage nach der Homogenität bzw. der Heterogenität der Gruppe als leitende Auswahlfragestellung an. Es gehe um „[...] das richtige Spannungsverhältnis [...] (Kuhn & Köschel 2011: 80) zwischen Homogenität und Heterogenität hinsichtlich wichtiger soziodemografischer Merkmale und bestimmter Einstellungen der potentiellen Diskussionsteilnehmer, um eine für die Diskussion möglichst fruchtbare Spannweite der Meinungen und Einstellungen zu erhalten. Ziel müsse es sein, „[...] dass Personen eingeladen werden, die sich durch besondere thematische Betroffenheit auszeichnen und von denen eine konstruktive Diskussion zu erwarten ist“ (dies.: 83).

Der Forschungsplan des Promotionsprojektes sah zunächst die Durchführung von fünf Gruppendiskussionen vor, um einerseits dem Anspruch des qualitativen Samplings gerecht zu werden und andererseits eine im Rahmen der Promotion noch zu bewältigende Datenmenge zu generieren. Entsprechend den oben zitierten Ausführungen von Kuhn und Köschel können als Zielgruppen des Bewertungskonzepts einerseits die Ausbilder und andererseits die Auszubildenden ausgemacht werden. Entsprechend wurden für jede der beiden Zielgruppen zunächst eine Gruppendiskussion geplant, die sich durch eine möglichst große Heterogenität, hinsichtlich der für die Forschungsfrage relevanten Merkmale der Diskussionsteilnehmer auszeichnen sollte. Bei der Rekrutierung der konkreten Teilnehmer der Diskussionen spielten soziodemografische Merkmale eine untergeordnete Rolle, da sich die große Mehrheit der Auszubildenden hinsichtlich dieser Merkmale wie Alter, Bildungs- und Familienstand oder Herkunft durch ein erhebliches Maß an Homogenität auszeichnet. Bei den Ausbildern mögen sich partielle Unterschiede z.B. im Bereich des Familienstandes ergeben, deren Relevanz für die hier diskutierte Fragestellung aber gegen Null tendieren. Das unterschiedliche (Dienst-)Alter der Ausbilder kann hingegen durchaus als Kriterium herangezogen werden, um den Faktor der Gruppenheterogenität zu steuern. Ein unterschiedliches (Dienst-)Alter bringt Unterschiede in der beruflichen Sozialisation durch unterschiedliche generationsspezifische Aspekte, wie führende paradigmatische Ansichten in den Leitwissenschaften der Ausbildung, Forderung

nach Reform bestimmter Aspekte der Ausbildung oder epochenspezifische pädagogische Herausforderungen mit sich. Gleichzeitig verändern sich bestimmte Einstellungsmuster mit zunehmender Erfahrung im Beruf und der Erfahrungsreichtum nimmt zu. Eine deutliche größere Rolle als die soziodemografischen Merkmale bei den Überlegungen zur Rekrutierung spielten die Einstellungen zum Vorbereitungsdienst im Allgemeinen und zur Bewertung innerhalb von diesem im Besonderen. Hierbei gilt es das Spektrum der Einstellungen möglichst breit abzubilden, um der oben formulierten Intention, die Institution des ZfsL in der Gruppendiskussion ‚im Kleinen‘ abzubilden, möglichst gerecht zu werden. Eine besondere thematische Betroffenheit der Diskutanten durch das Thema der Gruppendiskussion, wie Kuhn und Köschel sie als Rekrutierungsbedingung fordern, ist ohnehin bei allen Ausbildern und Auszubildenden anzunehmen.

Nach einem Pretest des Diskussionsleitfadens sollte nun in einem ersten Schritt je eine Gruppe aus Ausbildern und ehemaligen Auszubildenden des ZfsL Bonn die erste Version des Bewertungskonzepts diskutieren. Diese Diskussionen sollten einer Auswertung unterzogen werden, mit dem Ziel ein erstes Kategoriensystem zu entwickeln, mit dessen Hilfe, im Sinne des *theoretical Sampling* nach Glaser und Strauss (vgl. Kelle & Kluge 2010: 47–49, Strübing 2008: 30–33), nach Merkmalen der Diskutanten gesucht werden sollte, welche die Rezeption der Bewertungsinstrumente sowie die Wahrnehmung der Standardisierung der Bewertungsprozesse beeinflussen, um darauf aufbauend eine gezielte Auswahl von Teilnehmern für die weiteren Gruppendiskussionen vornehmen zu können. Neben zwei weiteren Gruppendiskussionen, je eine unter Ausbildern und eine unter ehemaligen Auszubildenden, sollte eine fünfte Diskussion in einer aus beiden Stakeholdergruppen gemischten Diskussionsrunde durchgeführt werden, um die polarisierenden Ansichten dieser beiden Gruppen möglichst pointiert herauszuarbeiten.

Nach diesen Vorüberlegungen wurden erste Rekrutierungsversuche gestartet, um das Teilnehmerpotential für das Forschungsprojekt auszuloten. Diese Versuche waren leider derart ernüchternd, dass eine gezielte, an den oben umrissenen Kriterien orientierte, Auswahl von Teilnehmern im Sinne des qualitativen Samplings ad acta gelegt werden musste. Stattdessen konzentrierte sich die Rekru-

tierung darauf, überhaupt genug Teilnehmer mobilisieren zu können, um die Gruppendiskussionen durchzuführen. Die Rekrutierung der Teilnehmer für die Diskussionen erwies sich allerdings als derart schwierig, dass nicht einmal die anvisierte Zahl der Gruppendiskussionen ohne ein qualitatives Sampling durchgeführt werden konnte. Durch die intensiven Bemühungen am ZfsL Bonn, auch unterstützt durch die persönlichen Kontakte des Autors zu dieser Institution und die tatkräftige Unterstützung der Zentrenleitung, konnten genug Diskutanten für je eine Gruppendiskussion unter Ausbildern und eine unter ehemaligen Auszubildenden rekrutiert werden. Genug Teilnehmer für weitere Diskussionen fanden sich leider nicht. Trotz mehrfacher auch persönlicher, Ansprachen aller sechs ZfsL in der Bezirksregierung Köln konnte keine ausreichende Zahl an Teilnehmern für eine Gruppendiskussion unter Ausbildern gewonnen werden. Gleiche Probleme zeigten sich bei der Rekrutierung von ehemaligen Referendaren für weitere Diskussionen. Andere Studien zur Lehrerbildung berichten von ähnlichen Einschränkungen, die zu einem ‚pragmatischen Vorgehen‘ bei der Rekrutierung von Teilnehmern geführt haben (vgl. Schulte 2008: 26–27, Kunter & Klusmann 2010: 71).

Die Ansprüche an die Güte der generierten Daten müssen durch diese Art der Rekrutierung heruntergeschraubt werden. Damit einhergehend muss an dieser Stelle die Frage aufgeworfen werden, inwieweit das zweite Ziel des Forschungsvorhabens, eine gegenstandsbezogene Theorie der Standardisierung von Bewertungsprozessen in der Lehrerbildung zu formulieren, erreicht werden kann. Zwei zentrale Merkmale der Grounded Theory, der Wechsel, bzw. das Ineinandergreifen von Phasen der Datengenerierung,-auswertung und Theoriebildung, sowie eine theoriegeleitete Auswahl von Teilnehmern, konnte für die erste Empiriephase des Forschungsprojektes nicht eingehalten werden, so viel sei vorweggenommen, sie ließ sich auch für die zweite Empiriephase, in der durch problemzentrierte Interviews die Erfahrungen mit dem Bewertungskonzept evaluiert wurden, nicht einhalten. Folgt man Strübing's Ausführungen, was ‚gute Grounded Theory‘ ausmacht, so muss das theoretische Sampling und der damit in Verbindung stehende Wechsel von Phasen der Datenerhebung und

-auswertung, sowie Theoriebildung als konstituierendes Merkmal dieser Methodologie aufgefasst werden (vgl. Stürbing: 30–33; 87–88). Auch wenn Strübing mit Rückbezug auf Strauss ausführt, dass die Grounded Theory kein „[...] rigides Regelwerk für das analytische Vorgehen [...]“ ist und Strauss die in der „[...] Grounded Theory entwickelten analytischen Verfahren lediglich als Vorschläge verstanden wissen [...]“ (ders.: 17) will, bedeutet das Fehlen eines theoretischen Samplings in der Fallauswahl und ein lediglich in Ansätzen vorhandener Wechsel von Datenerhebung, -auswertung und Theoriebildung⁸⁹, dass in dieser Arbeit *nicht* der Anspruch vertreten werden kann, eine gegenstandsbezogene Theorie der Standardisierung von Bewertungsprozessen im Bereich der Lehrerbildung zu entwickeln. Insbesondere der *Repräsentativitätsanspruch* der Grounded Theory, der durch das theoretische Sampling gesichert wird und sich darin findet, dass möglichst alle Kontexte und Perspektiven eines zu untersuchenden Phänomens, hier eben der Standardisierung im Bereich der Unterrichtsbesuche in der Zweiten Phase der Lehrerbildung, in der Datenerhebung und -analyse Berücksichtigung finden (vgl. ders.: 83), konnte nicht realisiert werden. Mit den im Rahmen der Promotion zur Verfügung stehenden zeitlichen und materiellen Ressourcen können die Ansprüche, die an eine ‚gute‘, in empirischen Daten begründete Theorie gestellt werden, deswegen nicht erfüllt werden. An die Stelle des Anspruchs der Entwicklung einer gegenstandsbezogenen Theorie tritt die Suche nach in den empirischen Daten *begründeten, gegenstandsbezogenen Hypothesen*, welche Rolle die Standardisierung von Bewertungsprozessen im Bereich der Lehrerbildung spielen kann, um damit zumindest einen Teil des Anspruchs der Grounded Theory umzusetzen, eben im Sinne der ‚*Praxistauglichkeit*‘ „[...] auf eine erweiterte Handlungsfähigkeit der Praktiker, also der Menschen in den jeweils untersuchten Handlungsfeldern“ (ders.: 86) hinzuarbeiten.

Entsprechend des abgegrenzten Erkenntnisinteresses des Forschungsprojektes wurde ein Leitfaden zur Fokussierung auf relevante Diskussionsaspekte erstellt.

⁸⁹ An die Gruppendiskussionen schloss sich, wie Abb. 3.1 zeigt, eine Phase der Datenauswertung an, auf die mit den problemzentrierten Interviews im Anschluss an die Konzepterprobung eine weitere Phase der Datenerhebung anschloss.

Die Leitfadententwicklung orientierte sich an den Vorschlägen von Lamnek über die Gestaltung eines differenzierten Leitfadens (questioning route) (vgl. Lamnek 2005a: 96–104). Die Vorschläge von Lamnek wurden ergänzt um einige Überlegungen von Kühn und Koschel, sowie Witzels Hinweise zur Formulierung von Fragen in problemzentrierten Interviews (vgl. Kühn & Koschel 2011: 102–121, Witzel 2000).

Der Leitfaden soll an dieser Stelle nicht detailliert vorgestellt, sondern lediglich die generelle Struktur, sowie die Anwendung, angesprochen werden⁹⁰. Der Leitfaden gliederte sich in vier Bereiche⁹¹:

- eine kurze *Einführung* in die Gruppendiskussion, das Erkenntnisinteresse der Diskussion sowie die Bedeutung der Diskussion für das Forschungsprojekt
- ein *Warm up*, das eine persönlichen Annäherung der Diskutanten an das Thema der Bewertung von Unterrichtsbesuchen in der Zweiten Phase der Lehrerbildung ermöglicht
- einen *Hauptteil*, der sich in drei Frageblöcke unterteilt. Diese Frageblöcke konzentrieren die Diskussion auf folgende Aspekte:
 - Frageblock 1: Wahrnehmung des Instruments und der ‚Logik‘ mit der Unterricht, bewertet werden soll
 - Frageblock 2: Diskussion, ob die Auswahl von basalen Teilkompetenzen aus den Konzepten von Helmke und Meyer schlüssig ist
 - Frageblock 3: Diskussion, was eine ‚gute‘ Bewertung von Unterrichtsbesuchen ausmacht und wie das vorliegende Konzept dazu beitragen kann
- einen *Abschlussteil* in dem die Diskutanten hervorheben können, was ihnen an der Diskussion besonders wichtig war, bzw. welche Aspekte sie vermisst haben

⁹⁰ Die Leitfäden der Gruppendiskussionen finden sich im Anhang.

⁹¹ Der Hauptteil des Leitfadens für die Gruppendiskussion der Referendare umfasste lediglich zwei Frageblöcke. Frageblock zwei wurde aus der Diskussion ausgeklammert.

Im Zuge der Gruppendiskussionen wurden nicht zwangsläufig alle im Leitfaden vermerkten Fragen gestellt. Wenn das Erkenntnisinteresse der Fragen durch Beiträge bereits geklärt werden konnte, wurden sie ebenso ausgelassen wie für den Fall, dass im Zuge der Diskussion zu erwarten war, dass die Frage so keine zusätzlichen projektrelevanten Erkenntnisse generieren würde.

3.2.2 Erprobung der Instrumente in der Ausbildungspraxis und Datenerhebung durch problemzentrierte Interviews

Im Anschluss an die Gruppendiskussionen, deren Auswertung und der anschließenden ersten Überarbeitung der Instrumente wurden diese in der Ausbildung eingesetzt. Drei Ausbilder des Ausbildungsjahrgangs, der im November 2012 die Ausbildung am ZfsL Bonn begann, beteiligten sich an der Erprobung. Die Erprobung des Konzepts erfolgte in den Fächern Deutsch, Erdkunde und Sport und deckte somit eine relativ breite Spannweite verschiedener fachlicher Kontexte ab. In diesen Seminaren wurden jeweils drei Referendare mit den neuen Instrumenten bewertet, insgesamt beteiligten also neun Auszubildende an der Erprobung. Ursprünglich sollte die Erprobung ebenfalls im Fachseminar für katholische Theologie erfolgen, aber leider verstarb der Ausbilder vor dem Beginn des Erprobungszeitraums.

Auch die Auswahl der Teilnehmer der Erprobung sollte, wie die Rekrutierung für die Gruppendiskussionen, im Sinne eines qualitativen Samplings geschehen, aber es zeigten sich die gleichen Probleme, sodass auch in diesem Fall das freiwillige Engagement der Teilnehmer das leitende Teilnahmekriterium wurde, mit einer entsprechend eingeschränkten Güte der aus der Erprobung generierten Daten.

Für die Auswertung der Erprobung wurden *leitfadengestützte, problemzentrierte Interviews* durchgeführt. Trotz dieser Bezeichnung der Interviews folgt die Durchführung nicht streng der von Witzel entwickelten Systematik des

problemzentrierten Interviews (vgl. Witzel 2000)⁹². Vielmehr waren die Interviews in dem Sinne ‚problemzentriert‘, als sich das Interview um das ‚Problem‘ der Bewertungsprozesse in der Zweiten Phase der Lehrerbildung drehte. Dies deckt sich i.W. mit der von Witzel intendierten Problemzentrierung, wenn er diese wie folgt definiert :

„Die Problemzentrierung kennzeichnet die Orientierung an einer gesellschaftlich relevanten Problemstellung und charakterisiert die Organisation des Erkenntnis- oder Lernprozesses (Vorinterpretation): Der Interviewer nutzt die vorgängige Kenntnisnahme von objektiven Rahmenbedingungen der untersuchten Orientierungen und Handlungen, um die Explikationen der Interviewten verstehend nachzuvollziehen und am Problem orientierte Fragen bzw. Nachfragen zu stellen.“

Entsprechend der in Kapitel eins hergeleiteten ‚objektiven Rahmenbedingungen‘ der Bewertungsprozesse in der Ausbildung wurde ein Leitfaden erstellt, der das Interview hinsichtlich des Erkenntnisinteresses strukturieren sollte. Zentrales Erkenntnisinteresse der Interviews war, inwieweit die Bewertungsinstrumente die am Ende von Kapitel eins formulierten Ansprüche erfüllen oder nicht, sowie die Erfahrungen der Auszubildenden und der Ausbilder mit der Standardisierung der Bewertungsprozesse. Wie bereits im Fall der Gruppendiskussion der Ausbilder soll der Leitfaden hier nicht en detail, sondern nur seine generelle Struktur

⁹² Die passgenaue Bezeichnung von halbstandardisierten, bzw. -strukturierten Interviews erscheint vor der Fülle der verschiedenen Definitionen und Varianten qualitativer Einzelbefragungen nahezu unmöglich, wenn man nicht genau den Vorgaben eines einzelnen Autors folgt. So grenzt Schorn das themenzentrierte- vom problemzentrierten Interview dahingehend ab, dass das zuerst genannte ‚tiefer‘ liegende Sinngehalte des Gesagten erfasse. Während das problemzentrierte Interview lediglich den manifesten, subjektiv gemeinten Sinn der Interviewten erfasse, erhebe ein themenzentriertes Interview auch latente und abgewehrte Sinngehalte (vgl. Schorn 2000). Flick nennt das themenzentrierte Interview in seiner Einführung in die qualitative Sozialforschung nicht explizit, grenzt dafür das problemzentrierte Interview in der Tradition von Witzel vom fokussierten, halbstandardisierten Experten- und ethnographischen Interview ab (vgl. Flick 2005: 118–139), Przyborski und Wohlrab-Sahr wiederum sehen im Dilemmainterview nach Reinshagen eine besondere Form des fokussierten Interviews, während Bortz und Döring diese beiden Formen des Interviews explizit voneinander abgrenzen (vgl. Bortz & Döring 2006: 315, Przyborski & Wohlrab-Sahr 2009: 146). Die hier gewählte Bezeichnung des Interviews als ‚problemzentriert‘ folgt demnach v.a. heuristischen Zwecken, um es von sehr offenen Interviewformen wie z.B. dem narrativen Interview abzugrenzen.

vorgestellt werden⁹³. Für die Interviews mit den Ausbildern sowie mit den Auszubildenden wurden jeweils eigene Leitfäden erstellt. Auch die Leitfäden der Interviews gliederten sich wie der Leitfaden der Gruppendiskussion in vier Bereiche:

- eine kurze *Einführung* in das Interview, das Erkenntnisinteresse des Interviews, sowie die Bedeutung des Interviews für das Forschungsprojekt
- ein *Warm up*, das einen Erzählimpuls hinsichtlich der Rezeption der Bewertungsprozesse (Auszubildende) bzw. der Wirkung der neuen Bewertungsinstrumente (Ausbilder) generieren sollte
- einen *Hauptteil*, der sich in mehrere Frageblöcke unterteilt: Diese Frageblöcke konzentrieren das Interview auf folgende Aspekte;
 - Frageblock 1: Ist das Konzept so, wie es gestaltet ist, tauglich für den Einsatz in der Ausbildungspraxis?
 - Frageblock 2: Bietet das Bewertungskonzept Hilfestellungen für die Planung und Analyse von Unterricht und regt es dazu an, über den Unterricht „ins Gespräch zu kommen“?
 - Frageblock 3: Kann das Bewertungskonzept dazu beitragen die Transparenz der Bewertungsprozesse zu erhöhen?
- einen *Abschlussteil*, in dem die Interviewten hervorheben können, was ihnen an der Diskussion besonders wichtig war, bzw. welche Aspekte sie vermisst haben

Die Frageformulierungen berücksichtigen Witzels Anforderungen an ein problemzentriertes Interview. So fungierte eine vorformulierte, offene Einstiegsfrage bezüglich der Bewertungsprozesse in den Unterrichtsnachbesprechungen in der Einführungsphase der Interviews erzählgenerierend, während weitere ad-hoc Fragen bereitstanden für den Fall, dass die Interviewten bestimmte Aspekte der Bewertungsprozesse nicht ansprachen (vgl. Witzel 2000).

⁹³ Auch die Interviewleitfäden finden sich im Anhang.

3.2.3 Auswertung im Sinne qualitativer Evaluationsforschung

Die Auswertung der Daten, die im Rahmen der Gruppendiskussionen und der problemzentrierten Interviews erhoben wurden, sollte ursprünglich, entsprechend dem Anspruch, eine gegenstandsbezogene Theorie der Standardisierung von Bewertungsprozessen im Bereich der Lehrerbildung zu formulieren, im Sinne der Grounded Theory stattfinden. Auf Grund der oben ausgeführten Probleme bei der Rekrutierung von Teilnehmern für das Forschungsprojekt und der damit in Verbindung stehenden eingeschränkten Güte der generierten Daten erscheint eine genaue Umsetzung der Analysemethoden der Grounded Theory (offenem-, axialen- und selektivem Codieren, Dimensionalisieren, Methode der ständigen Vergleiche, theoretisches Samplen der Daten, Verfassen von Memos, vgl. Bortz & Döring 2006: 332–333, Kuckartz 2010: 73–83, Strübing 2008: 18–36) zu hoch gegriffen. Es wäre das sprichwörtliche ‚mit Kanonen auf Spatzen schießen‘. Stattdessen wird im Bereich der Datenauswertung ein Weg gegangen, der zentrale Auswertungsschritte im Sinne der Grounded Theory nutzt, ohne den Anspruch zu erheben, diese Analysemethodik voll umzusetzen⁹⁴. Hinweise zu dieser Art einer ‚kategorienbasierten Auswertung‘ finden sich v.a. bei Kuckartz (vgl. Kuckartz et al. 2008 : 33–49; Kuckartz 2010: 57–71), der ausführt, dass im Rahmen einer solchen Auswertung Kategorien als „[...] Werkzeuge zur Phänomenklassifizierung mit der Möglichkeit der Bildung von Unterklassen begriffen werden“ (Kuckartz 2010: 62), ohne die teils verwirrende Vielfalt von Kategorienbezeichnungen der Grounded Theory umzusetzen. Einer solchen Vorgehensweise, die sich an der Grounded Theory orientiert, schließen sich auch Kühn und Koschel in Ihrer Empfehlung für die Auswertung von Gruppendiskussionen an (vgl. dies. 2011: 173–216)⁹⁵. Die Auswertung folgt in dem Sinne dem Vorgehen

⁹⁴ Die damit einhergehenden Einschränkungen hinsichtlich der Genauigkeit und Vollständigkeit der Analyseschritte schlagen sich auch in anderen Forschungsprojekten nieder und werden in der qualitativen Evaluationsforschung, wie Flick schreibt, unter dem Stichwort der „Abkürzungsstrategien“ verhandelt (vgl. Flick 2006: 21–22, vgl. zur Diskussion um diese Abkürzungsstrategien z.B. Kuckartz 2006).

⁹⁵ Wohingegen Lamnek ein solches Vorgehen als „Cut-and-Paste-Technik“ von theoretisch begründeten Verfahren der qualitativen Datenanalyse explizit abgrenzt und ihm lediglich ein beschränktes Erkenntnispotential jenseits deskriptiver Aussagen zugesteht (vgl. Lamnek 2005a: 183–191).

„qualitativer Evaluationsforschung“, da sie Methoden der qualitativen Sozialforschung für die Evaluation des Konzeptes nutzt und aus dieser Evaluation Erkenntnisse über das Design des zu entwickelnden Bewertungskonzepts, sowie über die Rolle, die Standardisierung für produktive Bewertungsprozesse in der zweiten Phase der Lehrerbildung spielt, abzuleiten. Diese grundlegenden Ideen der Datenauswertung wurden im Rahmen dieses Promotionsprojektes wie folgt umgesetzt:

Die Gruppendiskussionen und Interviews wurden komplett transkribiert, um für eine systematische Textanalyse zur Verfügung zu stehen. Die *Transkription* richtete sich nach den Transkriptionshinweisen, die Dresing und Pehl, sowie Kuckartz et al. formulieren (vgl. Dresing & Pehl 2012: 20–32, Kuckartz et al. 2008: 27–28). Eine genaue Auflistung der Transkriptionsregeln findet sich im Anhang.

In einem *ersten, induktiv orientierten*⁹⁶ *Auswertungsschritt* wurden die transkribierten Gruppendiskussionen und Interviews einer explorativen Lektüre unterzogen, in deren Rahmen jenen Textstellen codiert wurden, die zu der hier diskutierten Thematik–der Entwicklung eines Bewertungsinstrumentes für die Unterrichtsbesuche in der zweiten Phase der Lehrerbildung, sowie der Frage, welche Rolle eine Standardisierung der Bewertungsprozesse für zufriedenstellende Bewertung spielen kann – Erkenntnismöglichkeiten eröffnen. Im Sinne des *offenen Codierens* wurden zunächst zahlreiche Kategorien vergeben, ohne diese in ein Kategoriensystem einzuordnen oder nach Kategorientypen zu unterscheiden. So folgte der erste Auswertungsschritt einem Vorgehen, wie Mayring es für die induktive Kategorienbildung empfiehlt (vgl. Kuckartz 2010: 202). Die so gewonnenen Kategorien wurden in einem *zweiten Schritt in ein hierarchisches Kategoriensystem* eingeordnet, das in *weiteren, deduktiv arbeitenden Auswertungsschritten* für *weitere Analysedurchgänge* der Transkripte genutzt wurde. Die

⁹⁶ Induktiv ist dieser Analyseschritt nur insofern, als keine fertigen Kategorien existieren, in welche die Textstellen eingefügt werden. Natürlich leiten aber die bereits generierten Erkenntnisse zu der Thematik und die Eindrücke aus der Moderation der Gruppendiskussionen zu einer Fokussierung auf bestimmte Aspekte, sodass der Zugang hier auch als Mischung aus einer induktiven und deduktiven Perspektive verstanden werden kann (vgl. dazu auch: Kuckartz 2010: 200–202).

zunächst vergebenen Kategorien aus dem offenen Codieren wurden durch die neu gebildeten Kategorien und Subkategorien aus dem Kategoriensystem ersetzt. Neue Textstellen, die sich nun in einzelne oder mehrere der neu gebildeten Kategorien einordnen ließen, wurden entsprechend codiert. Im Verlauf der Analyse entstanden somit *verschiedene Versionen hierarchischer Kategoriensysteme*, welche im Sinne der theoretischen Sättigung eine immer genauere Ordnung der in dem Datenmaterial vorliegenden Erkenntnisse zu dem hier diskutierten Thema ermöglichten.

Begleitet wurden die Analyseschritte durch das *Schreiben von Memos*, in denen Auswertungsideen, Deutungsalternativen für codierte Textstellen, etc. festgehalten und ggf. in den Auswertungsprozess einbezogen wurden. Das *ständige Vergleichen der unterschiedlichen Deutungen und Wertungen* der Diskutanten und interviewten Personen, und hierbei v.a. die unterschiedlichen Perspektiven der Ausbilder und der Auszubildenden über die Phänomene, die im Zusammenhang mit der Bewertung von Unterrichtsbesuchen berichtet wurden, bildeten den Dreh- und Angelpunkt für die Entwicklung von Kategorien und die Interpretation des Sinngehalts der einzelnen Kategorien. Bei der Auswertung der Gruppendiskussionen wurde weiterhin ein Blick auf die *Gruppendynamik* gelegt und hierbei insb. auf die Frage, ob sich unter den Diskutanten ein Konsens herstellen ließ oder ob ein Dissens zu diskutierten Fragen bestehen blieb.

Die Auswertung folgt den Schritten des Forschungsprojektes und wertet die Gruppendiskussionen und die Interviews getrennt voneinander aus. Die Auswertung der Gruppendiskussionen erfolgte vor der Durchführung der Interviews, so dass im Sinne des theoretischen Samplings Erkenntnisse aus den Gruppendiskussionen in die Konstruktion der Interviewleitfäden einfließen konnten. In einem abschließenden Fazit werden die Erkenntnisse aus der Analyse der beiden Auswertungsphasen zusammengeführt. Leitende Perspektive bei der Auswertung ist eine Fokussierung auf die Interaktions- und Kommunikationslogiken, von denen die Akteure berichten. Der Begriff des *Interaktions- und Kommunikationssystems* wird in dem Sinne genutzt, dass die Ausbildung immer aus sozialen Situationen besteht, in welche die Ausbilder und Auszubildenden mit je individuellen Erwartungen, Hoffnungen und Ängsten hineingehen, aber mit bestimmten wiederkehrenden

Rollenerwartungen des Gegenübers konfrontiert werden. Die Interaktions- und Kommunikationslogiken, die sich aus diesen Begegnungen ergeben, haben systemischen Charakter in dem Sinne, dass sie übersummativ sind, also nicht aus der Summe ihrer Teile (den Eigenschaften der Akteure) erklärt werden können und sich durch komplexe und nicht-lineare Rückkopplungsschleifen auszeichnen, so dass Ursache und Wirkung einzelner Handlungen nicht klar voneinander getrennt werden können (vgl. Baecker 2004: 12–65 , Watzlawick et al. 1986: 53). Im Rahmen der kategorienbasierten Auswertung lassen sich verschiedene Zustände dieser Interaktionssysteme *idealtypisch* gegeneinander abgrenzen und hinsichtlich ihres Beitrags zu einer gewinnbringenden Ausbildung qualitativ bewerten.

3.3 Theoretischer Rahmen – Widerstände in institutionellen Veränderungsprozessen

Die Entwicklung des Bewertungskonzepts und seine Erprobung geschehen in der Ausbildungspraxis des ZfsL Bonn. Insofern ist die Entwicklung des Konzepts in die institutionellen Logiken des ZfsL eingebunden. Die individuelle Bewertungspraxis von Unterrichtsbesuchen der an der Erprobung beteiligten Akteure, werden durch die tradierten Vorgehensweisen die an dieser Institution gelten beeinflusst sein. Entsprechend kann die Erprobung des Bewertungskonzeptes in dem ‚laufenden Ausbildungsbetrieb‘ als ein *Institutionsentwicklungsprojekt* verstanden werden. Die an dem Projekt beteiligten Ausbilder sollen sich in ihrer Bewertungspraxis von einem neuen Bewertungskonzept beeinflussen lassen und ihre gewachsenen Handlungsroutinen verändern. Hierbei müssen sie ihre Bewertungspraxis nicht nur nach Maßgabe eigener Vorstellungen und Wertungen hinsichtlich der Fragen von Qualitätskriterien guten (Fach-)Unterrichts oder sinnvolle Gestaltung der Unterrichtsnachbesprechungen verändern, sondern sich an extern festgelegten Kriterien und Prozessen orientieren. Im Sinne des von dem Projekt verfolgten bottom-up-Ansatzes der Institutionsentwicklung waren die Ausbilder zwar in die Entwicklung des Bewertungskonzepts eingebunden, aber es bleibt zu einem erheblichen Teil ein von außen kommender Impuls, der ihre gewachsene Bewertungspraxis beeinflussen und unter Umständen auch verändern soll.

Aus den Theorien zur Organisationsentwicklung im Allgemeinen und jenen der Schulentwicklungsforschung⁹⁷ im Speziellen, lassen sich aus diesem Setting bestimmte Rahmenbedingungen hinsichtlich der Frage ableiten, welche Erfolgsaussichten dem intendierten Veränderungsprozess der Bewertungspraxis zu attestieren sind. *Erstens* ist davon auszugehen, dass das Projekt auf Grund seines bottom-up-Ansatzes und der freiwilligen Beteiligung der Akteure in einem *innovationsfreudigen Rahmen* stattfindet (Reh 2010: 292, Stegmaier: 316–317, Antoni 2007). Auf Seiten der Ausbilder ist von einem nicht unerheblichen Maß an intrinsischer Motivation für die Entwicklung innovativer Bewertungskonzepte auszugehen und auch bei den beteiligten Auszubildenden kann eine gewisse Innovationsfreudigkeit vorausgesetzt werden. Die Erprobungsgruppe des Projekts, so kann man mit Terhart hier schlussfolgern, wird sich überwiegend aus den „[...] Engagierten und Überzeugten [...]“ (Terhart 2013: 79) zusammensetzen, wohingegen die in der täglichen Praxis relevanten Gruppen der „[...] Ablehner, [und] der Unbeteiligten [...]“ eher unterrepräsentiert sein dürften. Öffentliche Kritik an dem Vorhaben wurde nicht geäußert, vielmehr wurde es von diversen Personen mit viel Wohlwollen und positiven Wünschen unterstützt. Das Projekt war außerdem eng mit der Leitung des ZfsL Bonn, sowie der Bereichsleitung der Lehrerausbildung für Gymnasien und Gesamtschulen am ZfsL abgestimmt und wurde von diesen explizit begrüßt. Insofern lag auf einer institutionellen Ebene ausreichend Rückendeckung für die Erprobung des Bewertungskonzepts vor. Nichtsdestotrotz muss im Rahmen der Erprobung *zweitens* mit *Widerständen bei der Anwendung* des neuen Bewertungskonzepts gerechnet werden, wie sie im Rahmen von Entwicklungsprozessen auf institutioneller Ebene typischerweise auftauchen. Das Entstehen von Widerständen in Entwicklungsprozessen gegen Neuerungen ist nichts ‚Anormales‘, vielmehr sind diese Widerstände als Teil jedes Entwicklungsprozesses zu begreifen, sowohl auf der individuellen, persönlichen Ebene, als auch auf der Ebene einer Institution (vgl. Bohnsack 1995: 22–23, Reh 2010, Rolff et al. 2011: 186–203, Schiersmann & Thiel 2011: 417–418, Ter-

⁹⁷ Vgl. einführend zur Organisationsentwicklung z.B. Elke 2007, Schiersmann & Thiel 2011, Stegmaier 2007, zur Schulentwicklungsforschung im Allgemeinen Holtappels & Rolff 2010, Rolff 2007, zu verschiedenen Paradigmen der Schulentwicklungsforschung Bohl et al. 2010 79–129.

hart 2013: 77–79). Institutionelle Änderungen beeinflussen und verändern die gewachsenen Alltagspraktiken, die aus dem individuell geprägten Umgang mit Herausforderungen und Schwierigkeiten der Berufsalltags gewachsen sind, Veränderung oder ‚Change‘ zuzulassen ist immer eine Herausforderung, eine „[...] Verunsicherung des eigenen Kompetenzerlebens [...]“ (Terhart 2013: 77). Neben rational begründbaren Einwänden gegen die zu implementierenden Neuerungen spielen bei Widerständen gegen Veränderungsprozesse *emotionale Aspekte* die größere Rolle. Der Widerstand richtet sich nicht gegen die konkrete Veränderung, sondern gegen Veränderungen der gewohnten Arbeitsprozesse an sich, er kann „[...] als Abwehr von Ängsten, vor allem vor Kontrollverlust interpretiert werden (Reh 2007: 292). In der Datenauswertung und dem abschließenden Fazit der Arbeit ist deswegen ein genauerer Blick auf mögliche Widerstände gegen den Einsatz des Bewertungskonzepts zu werfen, die sich im Rahmen der Erprobung möglicherweise zeigen. Im Sinne des Erkenntnisinteresses der Arbeit sollen im Folgenden drei Arten von Widerstand, die sich gegen den Einsatz des Bewertungskonzepts richten, unterschieden werden:

- *(Emotionaler) Widerstand innerhalb von Veränderungsprozessen*, der im Sinne der Theorien institutionellen Wandels als gegeben zu akzeptieren ist, aus dem aber nur bedingt Erkenntnisse für die hier diskutierten Fragestellungen gezogen werden können.
- *Widerstand gegen das Design* des Bewertungskonzepts, aus dem Impulse für die Überarbeitung der Bewertungsinstrumente des Konzepts gezogen werden können.
- *Widerstand gegen standardisierte Bewertungsprozesse* an sich, aus dem Erkenntnisse hinsichtlich des generellen Nutzens von Standardisierung für Bewertungsprozesse in der zweiten Phase der Lehrerbildung gezogen werden können.

4 | Auswertung der Gruppendiskussionen

Die Gruppendiskussionen sollten in der ursprünglichen Intention überwiegend der mehr oder weniger pragmatischen Überprüfung der pre-alpha-Version des Bewertungskonzepts dienen. Beide Gruppendiskussionen entwickelten sich aber schnell zu Diskussionen über grundsätzliche Fragen der Bewertung von Unterrichtsbesuchen und den Möglichkeiten guten Unterricht zu operationalisieren und standardisiert zu bewerten, sodass aus der Dynamik der Diskussionen schon ein erhebliches Erkenntnispotential für die eigentlich erst im weiteren Verlauf des Forschungsprojektes anvisierte Frage resultierte, welche Rolle die Standardisierung für eine gewinnbringende Bewertung im Sinne der formativen Evaluation spielen kann. In diesem Unterkapitel werden folglich sowohl die konkreten Änderungen der pre-Alpha-Version des Bewertungskonzepts diskutiert und vorgestellt, wie sie sich aus der Analyse der Gruppendiskussionen ableiten ließen, als auch wesentliche Erkenntnisse über den Nutzen von Standardisierung der Bewertungsprozesse für formative Evaluation dargelegt.

Die Dokumentation der Ergebnisse beschränkt sich auf die dritte und damit letzte Version des hierarchischen Kategoriensystems, sowie die aufbauend darauf geleistete axiale Codierung. Die Auswertung folgt in ihrer Struktur dem Kategoriensystem und interpretiert die prägenden Erkenntnisse der einzelnen Kategorien für das hier diskutierte Thema und belegt diese durch ausgewählte Zitate. Innerhalb der Analyse einzelner Kategorien, zeigen Absätze den Beginn der Auswertung neuer Subkategorien an. Diese Darstellung ist nicht rein beschreibend, sondern eben *interpretativ*. In dem Sinne, wie Kuckartz et al. es formulieren, wenn sie sagen, dass die aus den Textstellen gewonnenen Zitate, die einer Kategorie zugeordnet werden konnten, in einen größeren Kontext eingeordnet werden (vgl. Kuckartz et al. 2008: 48). Diesen Kontext bildet die Frage, wie eine gewinnbringende Bewertung der Unterrichtsbesuche in der zweiten Phase der Lehrerbildung aussehen kann und welche spezifische Bedeutung eine Standardisierung der Bewertungsprozesse dabei spielt. Der besondere Fokus der *Interpretation* liegt auf

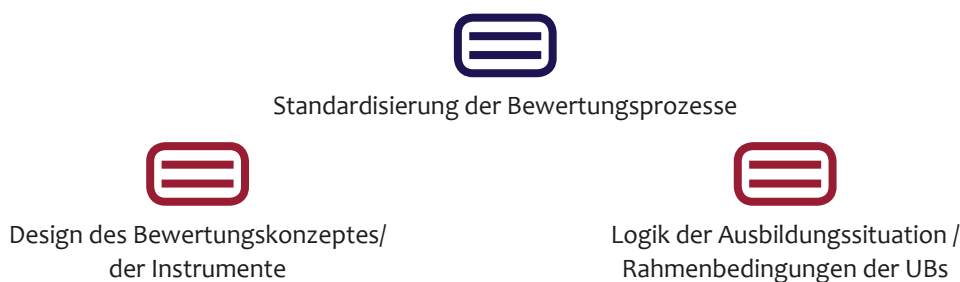
der Herausarbeitung wiederkehrender Argumentationsmuster in der Diskussion, die ein Erkenntnispotential für das oben genannte Erkenntnisinteresse der Arbeit aufweisen. Hierbei interessieren sowohl konsensuell verhandelte Aspekte, aus denen sich relativ abgesichertes Wissen über die Strukturbedingungen der Bewertungsprozesse ableiten lassen, als auch Diskussionsaspekte, über die kein Konsens hergestellt werden kann, weil diese den Blick auf offensichtlich ungeklärte, vielleicht auch gar nicht klärbare Herausforderungen richten und Hinweise für weitere Auswertungs- und Forschungsschritte geben.

Aus Gründen der Anonymisierung werden die Ausbilder und Ausbilderinnen geschlechtlich neutralisiert, es wird pauschal nur von den Ausbildern gesprochen, da ansonsten auf Grund des Fachbezugs und dem Geschlecht der Ausbilder und Ausbilderinnen auf konkrete Personen geschlossen werden könnte.

4.1 Das hierarchische Kategoriensystem der Gruppendiskussionen

Abbildung 4.1 zeigt die dritte Version des hierarchischen Kategoriensystems, das sich aus der Analyse der Gruppendiskussionen ergeben hat, ohne die Subkategorien der drei Hauptkategorien.

Abbildung 4.1: Hierarchisches Kategoriensystem der Gruppendiskussionen (Gamma-Version)



Die drei Oberkategorien, die das System strukturieren, umfassen einerseits die konkrete Fragestellung, auf welche Art und Weise das Bewertungskonzept dieser Arbeit modifiziert werden sollte, damit es in der Ausbildungspraxis in der Erprobung eingesetzt werden kann (*Design des Bewertungskonzepts/der Instrumente*⁹⁸); andererseits ordnen die beiden weiteren Kategorien Aspekte zum größeren Kontext der Bewertungsprozesse in der zweiten Phase der Lehrerbildung. Die Kategorie der *Logik der Ausbildungssituation/Rahmenbedingungen der UBs* stellt zentrale Erkenntnisse über die Besonderheiten der sozialen Interaktionssysteme, die sich im Rahmen der Unterrichtsnachbesprechungen⁹⁹ ergeben, sowie die realisierten Akteurstrategien innerhalb dieser Interaktionssysteme dar. Erkenntnisse über den Einfluss von Standardisierungsprozessen auf Bewertungssituationen in der zweiten Phase der Lehrerbildung werden in der Kategorie *Standardisierung der Bewertungsprozesse* verhandelt. Im Folgenden wird zunächst die Kategorie der *Logik der Ausbildungssituation* mit ihren Subkategorien vorgestellt, da sie die wesentlichen Kontextbedingungen der sozialen Interaktionssysteme umreißt, in die sich sowohl die generelle Frage der Standardisierung von Bewertungsprozessen, als auch die Frage des konkreten Designs des hier zu entwickelnden Bewertungskonzepts einordnen muss.

4.2 Die Logik der Ausbildungssituation/Rahmenbedingungen von Unterrichtsbesuchen

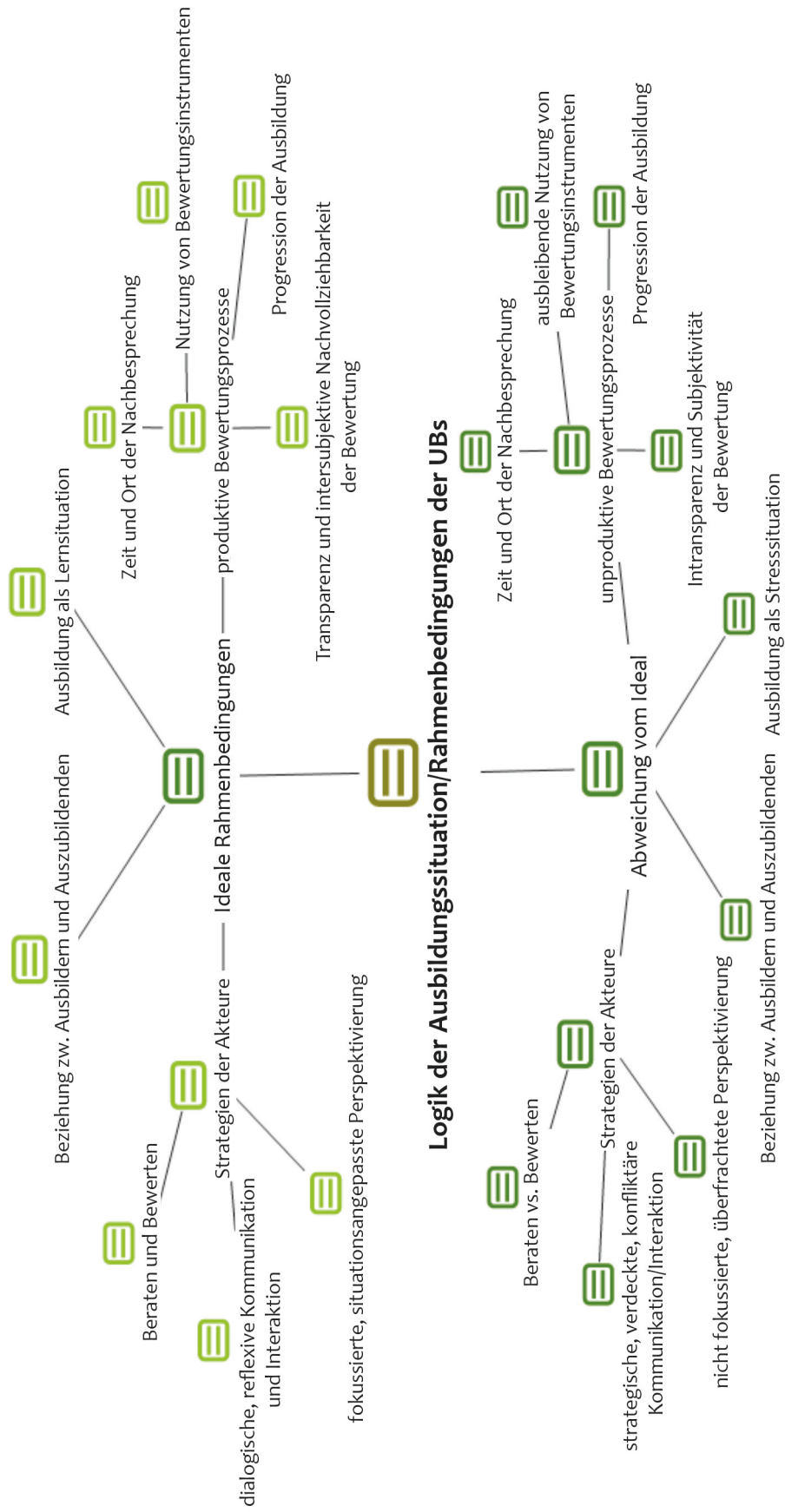
Abbildung 4.2 zeigt die Kategorie mit ihren Subkategorien, wobei die Kategorie die codierten Textstellen hinsichtlich einer grundsätzlichen qualitativen Einschätzung ordnet, und zwar anhand der Frage, ob die erlebten Ausbildungssituationen und die damit in Verbindung stehenden Strukturmerkmale der Interaktionssysteme den gewünschten und für eine Entwicklung von unterrichtsbezogenen

⁹⁸ In der folgenden Auswertung werden die Auswertungskategorien kursiv und in einfache ‚Anführungsstriche‘ gesetzt.

⁹⁹ Im Folgenden mit der, bei den Akteuren gängigen Abkürzungen „UB“ dargestellt.

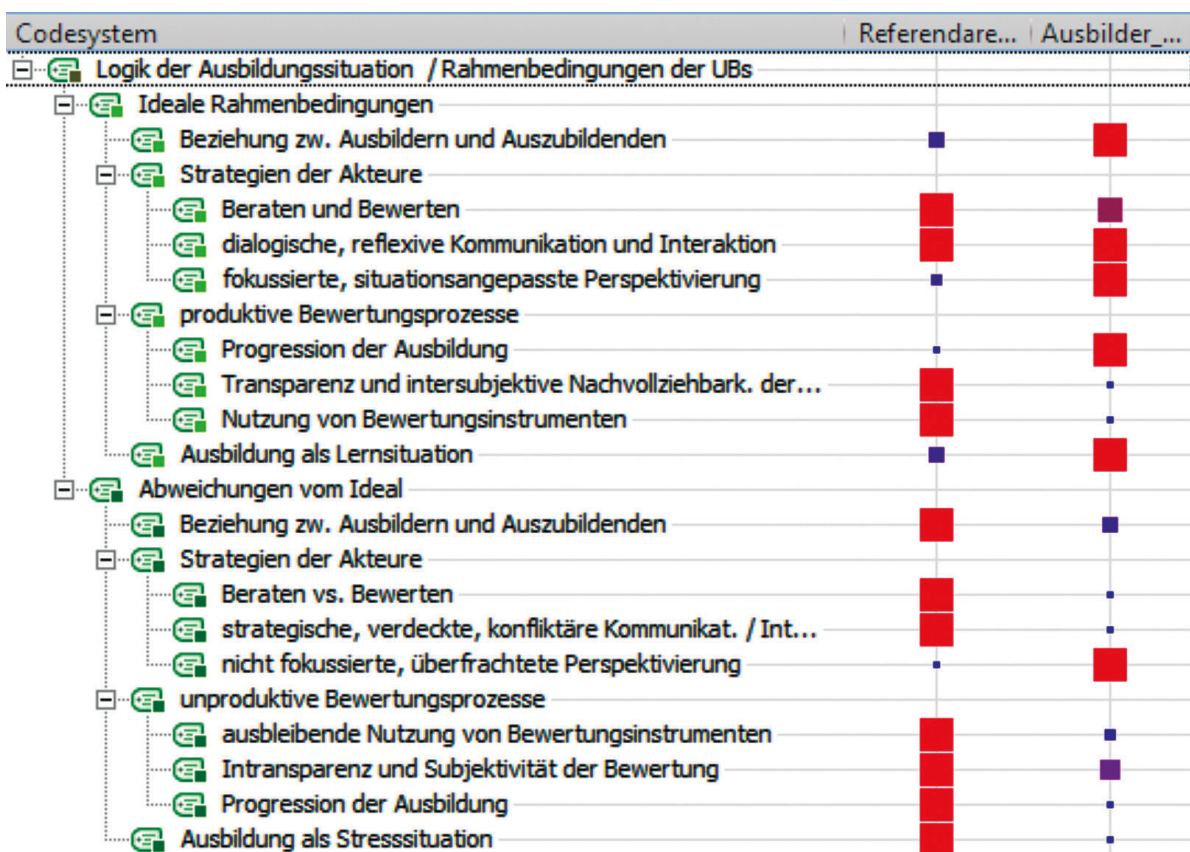
Kompetenzen förderlichen Merkmalen entsprechen und in diesem Sinne als ‚*ideale Rahmenbedingungen*‘ einzuschätzen sind oder ob es sich um Berichte über Erfahrungen und Strukturmerkmale handelt, die eine ‚*Abweichung vom Ideal*‘ darstellen. Die Sichtweisen der Auszubildenden und Ausbilder werden in der folgenden Auswertung, mit Hilfe des Code-Matrix-Browser von MaxQDA, zunächst in ihrem rein quantitativen Auftreten deskriptiv gegenübergestellt, um dann getrennt voneinander interpretativ ausgewertet zu werden. Ein abschließendes Fazit führt beide Perspektiven hinsichtlich ihrer Gemeinsamkeiten und Unterschiede zusammen.

Abbildung 4.2: Subkategorien der Kategorie ‚Logik der Ausbildungssituation/Rahmenbedingungen der UBs‘



Ein erstes Fazit der Gruppendiskussionen lässt sich aus dem Code-Matrix-Browser ziehen, der die Codierhäufigkeiten einzelner Kategorien in der Diskussion der Ausbilder der Häufigkeit in der Diskussion der Auszubildenden gegenüberstellt, also die relative Codierhäufigkeit zeilenweise abbildet. Hierbei ist festzustellen, dass in der ca. 25% längeren Gruppendiskussion der Ausbilder insgesamt lediglich 78 Textstellen mit dieser Kategorie codiert wurden, wohingegen es bei den Auszubildenden 103 Textstellen waren, was darauf hindeutet, dass die Auszubildenden ein deutlich stärker ausgeprägtes Bedürfnis hatten, über die Rahmenbedingungen der Unterrichtsbesuche zu sprechen als die Ausbilder.

Abbildung 4.3: Code-Matrix der Kategorie ‚Logik der Ausbildungssituation/Rahmenbedingungen von UBs‘



Die generell höhere Zahl der codierten Textstellen in der Diskussion der Auszubildenden erklärt im Code-Matrix-Browser die Ausschläge zu Gunsten der Auszubildenden teilweise, da diese, wie gesagt, prinzipiell mehr zu dieser Kategorie beigetragen haben und entsprechend auch die relativen Häufigkeiten der

Subcodes dieser Kategorie stärker auf Seiten der Auszubildenden codiert wären, wenn beide Diskussionsgruppen die gleichen Einschätzungen zu den erfragten Themen abgegeben hätten. Aus den reinen Häufigkeiten der codierten Subkategorien lassen sich darüber hinaus keine qualitativen Einschätzungen ableiten, da aus dem Code-Matrix-Browser nicht ersichtlich ist, in welcher Art und Weise die jeweiligen Akteure die Subkategorien thematisieren, z.B. ob sie über positive Erfahrungen berichten oder vielmehr Wünsche formulieren, wie die Ausbildung sein sollte. Aber auch vor dem Hintergrund dieser relativierenden Aussagen zeigt die Auswertung mit Hilfe des Code-Matrix-Browsers deutliche Unterschiede in den Codierungen der beiden Diskussionen, die auf eine je spezifische Perspektive der Auszubildenden und der Ausbilder hindeuten. Nur die Subkategorie der *„dialogischen, reflexiven Kommunikation und Interaktion“* weist mit elf codierten Passagen eine ausgewogene Thematisierung in beiden Gruppendiskussionen auf, *alle* (!) anderen Subkategorien werden in den beiden Diskussionen mit unterschiedlichem Schwerpunkt diskutiert.

Positive *„Beziehungen zwischen den Ausbildern und den Auszubildenden“* werden von den Ausbildern häufiger angesprochen als von den Auszubildenden. Die Relationen im Code-Matrix-Browser zeigen an, dass die Auszubildenden wesentlich stärker auf die problematischen Aspekte von Beziehungen fokussieren als die Ausbilder. Auch bei der Berücksichtigung der absoluten Codierhäufigkeiten bestätigt sich dieses Bild (codierte Passagen auf der positiven/negativen Seite der Subkategorie: Ausbilder 4/4 vs. Auszubildende 2/10).

Bei den *„Strategien der Akteure“* zeigt sich ebenfalls eine abweichende Einschätzung zwischen den beiden Gruppen. Die Ausbilder thematisieren deutlich häufiger als die Auszubildenden (11 vs. 4 codierte Passagen) die Bedeutung einer *„fokussierten, situationsangepassten Perspektivierung“*, aber nahezu ebenso häufig (12 codierte Passagen) sprechen die Ausbilder die Problematik einer *„nicht fokussierten, überfrachteten Perspektivierung“* an, wohingegen die Auszubildenden auch den negativen Aspekt deutlich seltener (2 codierte Passagen) thematisieren, was auf eine gering ausgeprägte Sensibilisierung der Auszubildenden für diesen Aspekt produktiver Unterrichtsnachbesprechungen hindeutet. Gleich häufig the-

matisieren die beiden Gruppen die Bedeutung einer ‚*dialogischen, reflexiven Kommunikation und Interaktion*‘ (jeweils 11 codierte Passagen), aber die Auszubildenden betonen ungleich häufiger als die Ausbilder die Rolle ‚*strategischer, verdeckter, konfliktärer Kommunikation und Interaktion*‘ (20 vs. 2 codierte Passagen), woraus sich ein Hinweis auf die dominante Strategie der Auszubildenden, mit der sie versuchen, in den Unterrichtsnachbesprechungen zu bestehen, ableiten lässt. Welche Rolle die Wahrnehmung einer Beraterfunktion durch die Ausbilder für positive Unterrichtsnachbesprechungen spielt, wird von beiden Gruppen nahezu gleich häufig thematisiert (Auszubildende 7 vs. Ausbilder 5 codierte Passagen), aber die Auszubildenden berichten deutlich häufiger (6 vs. 1 codierte Passage) als die Ausbilder über die Probleme ausbleibender Beratung oder jener, die aus einer Vermischung der Beratungs- und Bewertungsfunktion der Ausbilder entstehen. In diesem Sinne lässt sich zusammenfassend sagen, dass die Auszubildenden häufiger als die Ausbilder die problematischen ‚*Strategien der Akteure*‘ thematisieren und beide Gruppen offensichtlich je spezifisch für problematische Strategien sensibilisiert sind.

Ein ähnliches Bild zeigt sich bei der Subkategorie der ‚*(un-)produktiven Bewertungsprozesse*‘, in der die Akteursgruppen einem je individuellen Fokus folgen. Auffällig ist, dass die Auszubildenden in fünf von sechs Subkategorien mehr codierte Passagen vorzuweisen haben als die Ausbilder, also insgesamt öfter über produktive als auch unproduktive Bewertungsprozesse berichten. Insofern ist besonders die Subkategorie der Bedeutung der ‚*Progression der Ausbildung*‘ für produktive Bewertungsprozesse auffällig, weil lediglich dieser Aspekt von den Ausbildern häufiger thematisiert wird als von den Auszubildenden, und dies sehr deutlich mit sieben gegenüber einer einzigen codierten Passage bei den Auszubildenden. Hier zeigt sich wieder ein gruppenspezifischer Fokus auf die Herausforderungen, die die Bewertungsprozesse mitbringen.

In einer ersten zusammenfassenden Übersicht der Auswertung der Gruppendiskussionen können die obigen Ausführungen in der Subkategorie der ‚*Ausbildung als Lernsituation, bzw. Stresssituation*‘ gebündelt werden. Die Auszubildenden betonen deutlicher die Qualität der Ausbildungssituationen als ‚*Stresssituation*‘ (7

codierte Passagen), wohingegen die Ausbilder in ihrer Diskussionen einen stärkeren Fokus auf die ‚Ausbildung als Lernsituation‘ (9 codierte Passagen) werfen. Die Auszubildenden die Ausbildungsbedingungen also offenbar deutlich negativer bewerten, als die Ausbilder dies tun.

4.2.1 Die Sichtweise der Ausbilder

Alle¹⁰⁰ Ausbilder kennen aus ihrer Arbeitspraxis Ausbildungssituationen, die in ihren situativen Kontextbedingungen als ‚ideale Rahmenbedingungen‘ zu kennzeichnen sind und sich dadurch auszeichnen, dass die Ausbilder subjektiv wahrnehmen, dass die Auszubildenden Verständnis für zentrale Herausforderungen des Unterrichts entwickeln, dies an die eigene Unterrichtspraxis zurückbinden können und Offenheit dafür zeigen, die Impulse der Ausbilder zum eigenen Kompetenzaufbau anzunehmen. Lassen sich ‚ideale Rahmenbedingungen‘, wie sie im weiteren Verlauf dieses Unterkapitels dargestellt werden, realisieren, dann resultiert daraus ein soziales Interaktionssystem, das als ‚Lernsituation‘ charakterisiert werden kann. Diese ‚Lernsituation‘, das wird durch die Statements der Ausbilder im Rahmen der Gruppendiskussion sehr deutlich, ist hochgradig von den Beziehungen geprägt, in deren Rahmen sie sich entfaltet; es muss eine vertrauensvolle Atmosphäre zwischen den Auszubildenden und den Ausbildern gegeben sein, denn sonst wird „der Motor der Entwicklung“, wie es ein Ausbilder formuliert, nicht anspringen:

A3: „[...] Aber ich denke, gute Ausbilder und Ausbilderinnen, ich benutze dieses Klischee jetzt einfach mal, obwohl es ein Klischee ist, haben, also davon bin ich fest überzeugt, haben immer schon die Ressourcen und die Reflexion, die Selbstreflexion angeregt. Und ich glaube, nur die bringt Menschen wirklich weiter. Also ich glaube, es ist ein entscheidender Punkt in der Ausbildung. Wenn ich den, wenn ich diesen Motor für Entwicklung, den jeder Mensch hat, wenn ich den nicht sehe oder zu fassen kriege, dann glaube ich, habe ich ganz schlechte Chancen. Also dann macht jemand was,

¹⁰⁰ Bei den quantifizierenden Aussagen zur Häufigkeit der hier präsentierten Ergebnisse folge ich i. W. den Vorschlägen von Kuckartz et al. und unterscheide: „Alle Diskutanten/Interviewten...“, „Fast alle..“, „Mehr als die Hälfte...“ und „Einzelne...“ (vgl. Kuckartz et al. 2008: 48).

weil der A3 das gerne will, der sieht gerne, was weiß ich, aber der macht das nicht, weil er seine Ressourcen kennenlernt (A1: «mhm» (bejahend)), sondern ich glaube, das ist entscheidend für erfolgreiche Ausbildung: Mehr über sich zu erfahren. [...]“

Unterrichtsnachbesprechungen werden demnach dann zu einer Situation in der die Auszubildenden etwas lernen, sie werden zu einer ‚Lernsituation‘, wenn die Auszubildenden bereit sind, Selbstreflexion zu üben und Verständnis für die Relevanz der Impulse der Ausbilder entwickeln. Das Zitat von A3 zeigt darüber hinaus, dass er offensichtlich auch andere Formen der Interaktion und Kommunikation mit Referendaren kennt, wenn diese nicht etwas tun, weil sie von dem Sinn und Zweck überzeugt sind, sondern „[...] weil der A3 das gerne will [...]“. Hier offenbart sich eine Interaktionsstrategie der Akteure, die verhindert, dass eine Unterrichtsnachbesprechung zu einer ‚Lernsituation‘ wird, sondern vielmehr zu dem Gegenpart, nämlich einer hemmenden ‚Stresssituation‘, wie weiter unten näher ausgeführt wird. Betrachtet man die Eigenschaften der ‚Lernsituation‘ weiter, dann zeigt sich, dass die Fähigkeit und der Wille Selbstreflexion zu üben, zentrale Kriterien für gelingende Unterrichtsnachbesprechungen sind:

A1: „[...] Aber letztendlich, und das finde ich ist wirklich schon auch ein Bereich der Erwachsenenenerziehung, muss das aus dem Menschen selber kommen (A3: Ja sicher) und ansonsten kann der sich da auch in die Richtung nicht entwickeln (A3: Ja, A4 «mhm» (bejahend)), das ist für mich auch ganz klar (A3: Genau). Und dann kann ich noch so oft sagen, ok/[...]“

A3: „[...] Aber für mich ist ganz klar, dass die Ausbildung anregen sollte, sich mit persönlichen Merkmalen auseinander zu setzen, so würde ich es formulieren(A6: «mhm» (bejahend)). Und ohne das, glaube ich, kann man nichts bewegen. Wenn ich mich nicht auseinandersetze mit dem, was mein Verhalten (A4: Kritisch), dann werde ich mich nicht entwickeln können (A1: Ja, A6: «mhm» (bejahend)), sondern ich muss wissen: Was löst mein Verhalten aus und mir darüber bewusst werden. Das meine ich, kann Ausbildung leisten, aber auch nicht mehr.“

A3: „[...] Und jeder Mensch will seine Wirkung erfahren, also das ist ja was, Gott sei Dank, dieses bringen sie ja mit. Und das zu nutzen und nicht zu missbrauchen, also das bedeutet ja nicht zu bewerten. Ich bewerte das nicht, sondern ich unterstütze eigentlich nur das Bewusstwerden, mehr nicht. Dann ist das sehr erfolgreich.“

Zeitgleich scheint das Respektieren von zentralen Persönlichkeitsaspekten der Auszubildenden ein zentrales Merkmal zu sein, um eine vertrauensvolle Atmosphäre aufzubauen. Der folgende Ausschnitt der Diskussion zeigt, auch durch die emotionale Art des Disputs, wie er im sonstigen Verlauf der Diskussion eher selten zu beobachten war, dass sich ein Konfliktfeld aus Nähe und Distanz für die Ausbilder aufspannt. A4 reagiert auf einen Impuls von A2, dass Humor eine wichtige Eigenschaft eines Lehrers sei:

A4: „[...] Ich bin als Theologe bitterernst immer (ironisierender Tonfall).

A3: Das stimmt ja auch nicht.

A2: Wir müssen das nicht übertreiben, Nein.

A4: Nein, nein ich will es nur sagen. Ich wollte/Unsere verschiedenen Naturells, das gilt natürlich für unsere Referendare genau so (A2, A3, A6 «mh» (bejahend)). Das kann ich nicht einfach umpolen. Also ich kann nicht einfach sagen: „Seien Sie doch mal ein bisschen locker (A3: Nein, genau) und jetzt machen sie mal ein Witzchen“, darum geht es auch gar nicht.

A2 (A1, A2, A4 durcheinander): Nein. Soll ich jetzt einen Vortrag halten darüber, wie man das macht, da wo die Grenzen liegen? Potential ist trotzdem da, bei jedem, individuell.

A4 (A1, A2, A4 durcheinander): Nein, nein, nein, ist mir schon klar.

A1 (A1, A2, A4 durcheinander): Aber ich finde die Persönlichkeit (unv.)

A4: Die Ressourcen ansprechen und sie herausfordern.

A1: Ressourcen ansprechen, genau.

A4: Das ist deshalb insofern durchaus/Wir können nicht biegen und es geht auch nicht darum, dass ich ein Bild, ein Idealbild vom Lehrer habe (A2, A3, A6: Ja, genau), das ich dann erzeugen möchte (A3: Nee, genau geht gar nicht! A6: Ätzend!) im Sinne der Lehrerpersönlichkeit. Wobei ich schon sage: Wertschätzung gegenüber den Schülerinnen und Schülern ist für mich etwas Unabdingbares, das sind aber, da kommen wir an die Grundwerte unserer Demokratie (A3, A6: «mhm» (bejahend)) natürlich jetzt auch meines Glaubens (A6: Unseres Dienstherrn auch), des Dienstherrn, die da eingefordert werden. In welcher Ausprägung ist dann noch immer eine zweite Sache, aber das ist eine Aufgabe, die wir ein Leben lang haben und nicht nur während des Referendariats. Und darüber sich aber Rechenschaft zu geben, dass man auch diesbezüglich immer wieder sich einen Spiegel vorhalten soll, halte ich für essentiell (A3: «mhm» (bejahend)).“

A2 berichtet von einem anderen Fall, der deutlich macht, welche große Rolle eine vertrauensvolle Beziehung zwischen den Ausbildern und den Auszubildenden dabei spielt, soziale Kontexte zu erschaffen, in denen Raum für Selbstreflexion ist:

„A2: Also ich hatte auch Fälle, wo eine Referendarin z.B. zu mir kam im Hauptseminar und formuliert hat, mit Tränen in den Augen, dass sie keinen Kontakt zur Klasse aufbauen kann, zum Kurs. Die sind bei anderen Referendaren an der gleichen Schule viel netter zu denen, und sie ist ständig da in einem Kampf mit denen drin, und sie versteht nicht, woran das liegt. Das ist so eine Situation, (A1: Ja typisch) wo wir dann auch gefragt sind. Und wir haben dann drüber gesprochen und ich habe ihr, wir haben dann versucht Lösungen gemeinsam zu finden: Wo liegt es am Verhalten vielleicht von ihr, wie sie mit Schülern umgeht? An ihrem Auftreten? Und vielleicht ist sie zu stark am Plan immer orientiert und geht gar nicht auf die Schüler ein, hört nicht aktiv zu, usw.? Das war dann das Thema, aber ich glaube, da ist auch Voraussetzung eine gute Beziehungsebene zwischen Fachleiter oder Hauptseminarleiter und Referendar oder Referendarin

(A3: Ja, genau). Wir selber bauen auch eine Beziehungsebene oder sollten das (A4: Ja aber hundertprozentig, A6: Ja ohne Zweifel, A3: Ja (unv.)), das ist meine Überzeugung. [...]“

In all diesen Zitaten, aber insbesondere in dem letzten, wurde die Bedeutung einer vertrauens- und respektvollen ‚*Beziehung zw. Ausbildern und Auszubildenden*‘ für die Realisierung von sozialen Kontexten, die als ‚*Lernsituation*‘ wahrgenommen werden, deutlich. Betrachtet man die Herausforderungen, die sich an die Ausbilder stellen, so findet sich ein zentrales Konfliktfeld von Nähe und Distanz: Die Ausbilder müssen, umgangssprachlich gesprochen, nahe genug an die Auszubildenden heran kommen, damit sie die zentralen Baustellen der Professionalisierung ansprechen können und die Referendare diese Impulse aufnehmen. Hierbei spielen Persönlichkeitsmerkmale nicht nur hinsichtlich der Frage eine Rolle, ob diese Reflexionsimpulse angenommen werden, sondern darüber hinaus geht es immer wieder ganz konkret um bestimmte Persönlichkeitsmerkmale der Auszubildenden, die im Rahmen ihrer Professionalisierung eine Herausforderung darstellen (Hier: Der vorhandene oder eben nicht vorhandene Humor und der wertschätzende Umgang mit anderen Menschen, bzw. Schülern). Gleichzeitig bedarf es in dieser Interaktion einer hohen Sensitivität der Ausbilder, damit sie den Referendaren und ihren je spezifischen Persönlichkeitsmerkmalen gerecht werden und zu diesen Distanz wahren, um nicht der Versuchung zu erliegen, den Auszubildenden den jeweils individuellen Umgang mit herausfordernden Persönlichkeitsmerkmalen aufzuoktroieren. Das richtige Maß von Nähe und Distanz ist offensichtlich nicht ohne weiteres zu klären (siehe Meinungsunterschiede von A2 und A4), es scheint eine zentrale Aufgabe der Ausbilder zu sein, das jeweils für sich angemessene Maß zu finden, aber dass die Ausbilder ‚nah‘ an die Auszubildenden heran kommen müssen, scheint in der Gruppe Konsens zu sein.

Das Erleben der ‚*Ausbildung als Lernsituation*‘ hängt eng mit den Interaktions- und Kommunikationsmustern zusammen, welche von den Akteuren in den Ausbildungssituationen gezeigt werden. Diese ‚*Strategien der Akteure*‘ scheinen sowohl Folge der erlebten Interaktions- und Kommunikationsmuster zu sein, als auch eben diese hinsichtlich ihrer konkreten Ausgestaltung zu formen. Ge-

lingt es den Akteuren, das Interaktionssystem im Sinne einer ‚*Lernsituation*‘ zu gestalten, wobei die Kommunikationsstrategien sowohl Ursache als auch Folge dieser Gestaltung sind, dann legen sie Strategien an den Tag, die durch die drei Subkategorien einer ‚*dialogischen und reflexiven Kommunikation und Interaktion*‘, ‚*einer fokussierten und situationsangepassten Perspektivierung*‘ und schließlich einer diskursiven Trennung von ‚*beraten und bewerten*‘ charakterisiert werden können. A4 berichtet in diesem Kontext von seiner Idealvorstellung der Kommunikationsmuster der Auszubildenden, in der es um die reflexive und im wahrsten Sinne (sich) selbstbewusste(r) Form der Unterrichtsnachbesprechung geht:

A4: „[...] Und hatte ich keine Bedenken, und ich würde immer, ich ziehe das auch vor, weil für mich eigentlich, ich nie an meinem Referendar, meiner Referendarin eigentlich vorbei reden kann, wenn sie selber sich da einbringt, in dem Sinne, dass sie sagt, was für sie bedeutsam ist (A3: «mhm» (bejahend)), dann habe ich, bin ich immer näher dran und habe Hoffnungen, dass mein Beitrag besser verstanden werden kann.“

Das Zitat zeigt, dass die Ausbilder für eine gelingende Ausbildung offensichtlich auf das Zutun der Auszubildenden angewiesen sind, weil deren Einbringen der eigenen Bedürfnisse quasi eine eingebaute Qualitätssicherungsmaßnahme gelingender Ausbildung ist. Ausbildung kann offensichtlich nur im Dialog zwischen Ausbilder und Auszubildenden gelingen, nicht durch die einseitige Übernahme von Expertise. Hierbei geht es den Ausbildern nicht um eine Selbstkasteiung der Auszubildenden, sondern um reflexive Auseinandersetzung mit den Entwicklungsaufgaben, die sich ihnen je individuell stellen, wie der folgende kurze Diskussionsausschnitt zeigt:

A4 (A2, A3, A4 durcheinander): „Ja, ja, und wichtig ist mit dabei, dass die nicht im Sinne hier: Oh me culpa! Ich habe da was falsch gemacht, also da (A3: Nee (stimmt A4 zu)) das zulsondern, dass Sie das nachweis/

A3: Einen Ansatzpunkt finden für die Weiterarbeit.

A4: Ja.“

Eng in Verbindung mit dieser ‚*dialogischen und reflexiven Kommunikation und Interaktion*‘ steht die ‚*fokussierte und situationsangepasste Perspektivierung*‘ während der Unterrichtsbesuche und der Unterrichtsnachbesprechungen. Die hohe Komplexität des Unterrichts und der Professionalisierungsabläufe der Auszubildenden bedingen, dass die Beobachtung während der Ausbildungssituationen nicht versucht, die Komplexität in Gänze zu erfassen, sondern auf bestimmte Ausschnitte der Ausbildungsrealität fokussiert und situationsangepasst bewertet. Dies kann nur im Dialog zwischen Ausbildern und Auszubildenden geleistet werden, wie die folgenden Zitate verdeutlichen:

A6: „[...]Am idealsten ist, finde ich, ist mir ein Referendar, der vorher zu mir kommt und im Entwurf schon geschrieben hat: „Mir wäre am liebsten, wenn sie guckten auf Punkt x und y“ (A1: Lachen). Das sind die Punkte, an denen ich derzeit am meisten arbeite, die mir am meisten Probleme haben. Dann kann ich darauf auch ganz speziell meine Aufmerksamkeit fokussieren. Sonst, wenn der Referendar hinterher fragt und sagt: «Was ist Ihnen eigentlich noch mal zu dem Punkt aufgefallen, da haben wir vor drei Unterrichtsbesuchen mal drüber gesprochen», dann kann ich nur sagen: «UPPS, das war mir jetzt gar nicht im Blick, weil ich so auf das geguckt habe».“

A3: „Aber das finde ich keinen Nachteil, sondern ich finde es sogar einen Vorteil. Also ich arbeite da ja auch relativ häufig mit, wirklich nur genau dieses Kriterium zu nehmen, wo war die spannendste, wo war die langweiligste Stelle in Ihrem Unterricht? Eben diese zwei Dinger und dann markiert der Referendar oder die Referendarin das und dann kommen wir wirklich über die entscheidenden Aspekte des Unterrichts ins Gespräch, und daraus entwickeln sich dann die, die dann darunter liegen. Also ich finde das gar keinen Nachteil, nur ein Kriterium zunächst in den Blick zu nehmen (A6: Ich auch nicht.), sondern es ist ja oft ganz schwierig, sonst haben wir u.U. dieses Additive (wendet sich an A2, A2: nicken) und haben das Wesentliche gar nicht erfasst.“

Auch hier zeigt sich wieder die Bedeutung dialogischer, selbstreflexiver Verständigung zwischen den Ausbildern und Auszubildenden. Ein Aspekt, der hingegen

v.a. die Ausbilder stark in die Pflicht nimmt, ist ein konstruktiver Umgang mit der problematischen Doppelfunktion, gleichzeitig Beratender und Bewertender zu sein. *Mehr als die Hälfte* der Ausbilder berichten von Situationen, in denen sie ihre Bewertungsfunktion zu Gunsten rein beratender Tätigkeit temporär außer Kraft gesetzt haben, da es in den spezifischen Situationen im Sinne des Ausbildungsfortschritts keinen Sinn gemacht hätte, weiter auf einer Bewertung zu beharren. Das folgende Zitat zeigt abermals die große Bedeutung einer stimmigen Beziehung zwischen Auszubildenden und Ausbildern, damit der Spagat zwischen Beraten und Bewerten geleistet werden kann:

A5: „Ich glaube, es ist auch ganz wichtig in so Unterrichtsnachbesprechungen, wenn man merkt, dass dem Referendar oder Referendarin was auf dem Herzen liegt, dass man dann mal alle Merkmale zur Seite nimmt (A3: Ja) und dass man dann offen spricht und was ist los und die meisten!“

A2: Das haben wir selber auch schon praktiziert mal zusammen, das weiß ich noch.

A5: Genau, die meisten Referendare nehmen das auch sehr, sehr dankbar an (A3: Ja). Und das ANDERE dann, das ist ja dann jetzt wirklich sekundär, ja dann geht es erst mal darum, die Persönlichkeit wieder ein bisschen zu stabilisieren (A2, A3: «mhm» (bejahend)).“

Das „Andere“, also die Bewertung des gesehenen Unterrichts, tritt in einzelnen Situationen hinter die Beratungsfunktion der Ausbilder zurück. Diese Bedürfnisse der Referendare in den Situationen zu erkennen ist entscheidend, die Bewertung wird in solchen Situationen offenbar „sekundär“, so die Sichtweise der Diskutanten.

Bei der Realisierung der ‚idealen Rahmenbedingungen‘ spielen die Bewertungsprozesse der Unterrichtsnachbesprechungen eine große Rolle. Idealbild sind ‚produktive Bewertungsprozesse‘, die durch ‚Transparenz und intersubjektive Nachvollziehbarkeit der Bewertung‘ gekennzeichnet sind. Die Ausbilder, das wird im weiteren Verlauf dieses Unterkapitels deutlich, sehen die großen Herausforderun-

gen, die dieser Anspruch an sie stellt und so thematisieren sie kaum die Möglichkeit, objektiv zu bewerten. Vielmehr wird der dialogische Charakter, wie er oben bereits aus dem Datenmaterial herausgearbeitet wurde, zum zentralen Kriterium möglichst nachvollziehbarer Bewertungsprozesse. Eine zentrale Herausforderung hinsichtlich produktiver Bewertungen bezieht sich auf den größeren Kontext der ‚*Progression der Ausbildung*‘. Es scheint einen Konsens unter den Ausbildern zu geben, dass die Bewertungen in den verschiedenen Phasen der Ausbildung berücksichtigen müssen, über welche Kompetenzen die Auszubildenden realistischerweise schon verfügen können. A2 und A6 thematisieren in den beiden folgenden Zitaten die Grenzen von standardisierten Bewertungsgrundlagen für Bewertungsprozesse, die sich durch die Ausbildungsprogression ergeben:

A2: „Mir ist ein Punkt wichtig bei der Standardisierung, dass wir die Referendare nicht erschlagen am Anfang der Ausbildung, durch die Fülle dieser Indikatoren, die ihnen dann möglicherweise direkt, komplett auch mal vorgestellt werden. Und dass sie dann sozusagen ein Ziel vor sich sehen, dass für sie unerreichbar ist. Und da müssen wir unheimlich aufpassen, dass wir hier Ängste reduzieren und diesen Katalog wirklich so vorstellen, dass eben nicht alles zur gleichen Zeit erfüllt werden soll, sondern dass das schrittweise sich entwickelt (A4: 18 Monate Zeit hat). [...]“

A6: „Aber das ist natürlich auch noch eine Sache, da kommt jetzt nochmal, finde ich, ein ganz großes zusätzliches Problem rein, was wir an anderer Stelle schon angesprochen haben, die Dynamik der Ausbildung (A4: Ja, A1: Die Komplexität). Das ist natürlich was, was/ Am Anfang der Ausbildung wird kaum ein Referendar die relevanten Merkmale, wie du (A2) das genannt hast, vollständig nennen und beschreiben können und erst recht analysieren können. [...]“

Spannend wird die Diskussion unter den Ausbildern, wenn dieser abstrakte Konsens, dass die Progression der Ausbildung eine bedeutende Rolle für ‚*produktive Bewertungsprozesse*‘ spielt, konkretisiert werden muss. In dieser Frage sind sich die Ausbilder dahingehend einig, dass es offenbar keine klare Leitlinie geben kann, wie diese Progression konkret auszusehen hat:

A1 (A1, A2 durcheinander): „Aber ich glaube auch da, da müssen wir schon ganz klar sagen: Wir können keine Chronologie aufmachen, was ist hier gerade als Erstes wichtig, was braucht jeder Referendar (A3, A6: «mh» (bejahend)) als Erstes? Das ist so nicht, das ist genauso individualisieren/Ja nee, also ich glaube, das kann man nicht machen, das ist Individualisierung der Ausbildung!

A2 (A1, A2 durcheinander): Es gibt kein lineares Aufbauen da, es ist eher System, systemisch (A4: Ja, ja).“

Aus Sicht der Ausbilder gibt es offenbar kein klares „lineares Aufbauen“ einzelner Ausbildungsbausteine oder Kompetenzen, vielmehr muss man diese Progression „individualisieren“ und „systemisch“ betrachten. Hier offenbart sich ein entscheidender Punkt der Bewertungsprozesse, der später bei der Kritik der Referendare deutlich wird: Die Ausbildungsprogression und deren Berücksichtigung ist ein von den Ausbildern stark betontes Element produktiver Bewertungsprozesse, die dem Ziel der Ausbildung dienen. Aber für die Referendare bleibt die Frage, worin sich die Progression denn genau definiert, wann welche Aspekte des Unterrichtens Grundlage der Bewertung werden und wann sie es noch nicht sind, nebulös, wie weiter unten gezeigt wird. Der Verweis auf die „individualisierten Ansprüche“ und das „systemische“ Denken verdeckt die Unklarheit dieses zentralen Bewertungsaspektes. Eine Klärung solch unklarer Aspekte in den Bewertungsprozessen, dies führt *ein Ausbilder* aus, kann durch ‚Nutzung von Bewertungsinstrumenten‘ zumindest teilweise geleistet werden:

A4: „Wir haben uns mit Meyer auseinandergesetzt und haben versucht über unsere Fachgrenzen hinaus ein gemeinsames Qualitätstableau¹⁰¹, so nennen wir das, zu benennen und haben auch noch ein paar Ergänzungen reingebracht. Also die sind jetzt, decken sich zum Teil mit dem (Anm.: Der Matrix der Qualitätsbereiche guten Unterrichts) und haben auch

¹⁰¹ Das sog. „Bonner Qualitätstableau“, das in der weiteren Auswertung noch öfter eine Rolle spielen wird. Es ist im Anhang dokumentiert. Bei diesem „Qualitätstableau“ handelt es sich um die tabellarische Ausarbeitung der zentralen Qualitätsbereiche guten Unterrichts, die Grundlage der Bewertungsprozesse im ZfSL Bonn sein soll.

versucht Indikatoren zu benennen. Und genau das ist eigentlich unsere gemeinsam vereinbarte, ja, Grundlage des Betrachtens, des Beobachtens von Unterricht. Inwiefern die tauglich ist, inwiefern die eingebracht wird in unsere Hospitation, wäre, könnte genau überprüft werden, denn wir könnten genau gucken, wo wird das eigentlich mal thematisiert?“

Der letzte Satz dieses Zitates deutet aber schon die praktische Bedeutung an die solche Bewertungsinstrumente offensichtlich bisher in der Ausbildungspraxis haben, nämlich eine äußerst marginale, wie im späteren Verlauf dieses Unterkapitels noch deutlicher gezeigt wird.

Zusammenfassend repräsentiert diese interpretative Darstellung der Gruppendiskussion unter Ausbildern die Charakteristika ‚idealer Rahmenbedingungen‘ der Ausbildung und der in ihrem Rahmen ablaufenden Bewertungsprozesse. Als *Zwischenfazit* kann die große Bedeutung dialogischer, selbstreflexiver Verständigung zwischen den Ausbildern und Auszubildenden betont werden, die sich immer deutlicher als der Dreh- und Angelpunkt gelingender Ausbildung herauskristallisiert und die Ausbilder, aber v.a. auch die Auszubildenden in die Pflicht nimmt: Ohne das aktive Gestalten der Ausbildungssituationen in ihrem Sinne kann Ausbildung nicht erfolgreich sein, so die Sichtweise der Ausbilder.

Es ist wenig verwunderlich, dass *alle Ausbilder* über Erfahrungen berichten, in denen Rahmenbedingungen der Ausbildung durch eine ‚Abweichung vom Ideal‘ gekennzeichnet sind. Das soziale Interaktionssystem, das sich durch diese Abweichung ausbildet, grenzt sich durch seine Eigenschaften deutlich von der oben umrissenen ‚Lernsituation‘ ab. Im idealtypischen Sinne ergibt sich v.a. für Auszubildende eine ‚Stresssituation‘, die es hinter sich zu bringen gilt, ohne dass ein Kompetenz- oder Wissensaufbau tatsächlich im Zentrum dieses Interaktionssystems stehen würde. Auf Seiten der Ausbilder wird zwar implizit deutlich, dass die Art der Interaktion in der ‚Stresssituation‘ wenig professionelle Befriedigung hervorruft, sie aber nicht im gleichen Maße unter dieser Interaktionslogik leiden wie die Auszubildenden, wie es weiter unten in diesem Kapitel dargestellt werden wird. Stellvertretend dafür steht, dass bei den Ausbildern keine einzige Interviewsequenz explizit mit der Subkategorie ‚Stresssituation‘ codiert wurde, wohingegen

es im nur etwas kürzeren Interview der Auszubildenden sieben Passagen waren. Die ‚*Beziehung zwischen Ausbildern und Auszubildenden*‘ spielt, wie es sich vor dem Hintergrund der bisherigen Analyse erwarten lässt, eine große Rolle für die Realisierung eines bestimmten Typus des Interaktionssystems. Kommt es zu Störungen auf der Beziehungsebene, dann tendieren die Interaktionssysteme dazu, sich in Richtung der ‚*Stresssituation*‘ zu entwickeln. Zentrales Problem in diesem Bereich sind die aus der Ausbildungs- und Bewertungssituation resultierenden hierarchischen Beziehungen zwischen den Ausbildern und den Auszubildenden, die dazu führen, dass die Rollen, in der sich Ausbilder und Auszubildende begegnen, dem Ziel der oben umrissenen selbstreflexiven Interaktion entgegenstehen:

A4: „[...] Ja und das geschieht natürlich auch in der Auseinandersetzung mit dem Referendar, wobei ich in der Regel, eben halt oft, nicht den Referendar, leider nicht oft als einen Gesprächspartner erleben kann, der mir wagt zu widersprechen. Also meine Äußerung wird zum Gesetz. [...]“

Dadurch, dass die Einschätzung der Fachleiter „zum Gesetz“ wird, wie A4 es nennt, werden Kommunikationsprozesse, durch die die Auszubildenden selbstreflexiv ihre Professionalisierung vorantreiben, torpediert. Sie „trauen sich nicht“ selbstbewusst aufzutreten, wie A6 im folgenden Zitat ausführt, mit entsprechenden Folgen für die Kommunikation mit den Ausbildern, wenn diese als ideale Rollenerwartung selbstbewusste Auszubildende für einen selbstreflexiven Diskurs wünschen:

A6: „[...] Also wenn die Referendare sich trauen, sind wir gut dran, denke ich. Die meisten trauen sich nicht (A3: nicken, «mhm» (bejahend)), das ist der bedauerliche Punkt, aber das kann natürlich dann leicht passieren, dann ist die Kommunikation schwieriger in dem Fall, das glaube ich schon.“

Einerseits Folge, aber zugleich Ursache dieser Beziehungsmuster sind die ‚*Strategien der Akteure*‘, mit denen sie im Rahmen der ‚*Stresssituation*‘ in die Interaktionen eintreten. Die Interaktionssysteme, die sich aus dieser Logik ergeben, führen dazu, dass aus der Perspektive der Ausbilder beide Akteursgruppen, v.a.

aber die Auszubildenden, eine ‚*strategische, verdeckte, konfliktäre Kommunikation und Interaktion*‘ an den Tag legen, wie A3 es in dem bereits oben genutzten Zitat verdeutlicht:

A3: „[...] Also dann macht jemand was, weil der A3 das gerne will, der sieht gerne, was weiß ich, aber der macht das nicht, weil er seine Ressourcen kennenlernt (A1: „«mhm» (bejahend)), [...].“

Ergründende Selbstreflexion und Kompetenzfortschritte im Bereich des Unterrichts werden durch diese Form der Interaktion und Kommunikation nicht gefördert, vielleicht sogar eher verhindert, da die Ausbilder eben nicht ‚nah‘ an die Auszubildenden herankommen, wie es aus der Perspektive der Ausbilder für eine gewinnbringende Ausbildung nötig wäre, sondern diese die Ausbilder – vielleicht aus Selbstschutz?- auf Distanz halten und sich den Ausbildern in ihren Professionalisierungskrisen nicht offenbaren. An die Stelle zunehmender Professionalisierung auf Seiten der Auszubildenden tritt die Suche nach den persönlichen Vorlieben des Ausbilders, die es zu erfüllen gilt. Eine auch für die Ausbilder wenig befriedigende, wie A6 es ausdrückt, „bedauerliche“ Erfahrung, die aber scheinbar zur alltäglichen Praxis der Ausbildung gehört.

Wie oben dargelegt wurde, ist es eine besondere Herausforderung für die Ausbilder, die Unterrichtsbeobachtung und die folgende Nachbesprechung im Dialog mit dem Auszubildenden auf die wesentlichen Punkte zu fokussieren und situativ zu bewerten. Gelingt dies nicht, kommt es zu einer ‚*nicht fokussierten, überfrachteten Perspektivierung*‘ in der Unterrichtsbeobachtung und -nachbesprechung, aus der Unklarheiten in der Bewertung und Reflexion entstehen. Die Probleme der Ausbilder stehen v.a. mit der Frage in Verbindung, wie der Unterricht in seiner Komplexität zu beobachten ist:

A6: „[...] Die Frage ist auch, wie viele Karten können wir eigentlich zumuten, das finde ich auch eine ganz, ganz entscheidende Grundsatzfrage.

A3: Ja (nicken).

A4: Ja, definitiv (nicken).

A6: Also der Hilbert Meyer war, glaube ich nicht so schlecht beraten, es auf zehn zu begrenzen.

A3: Ist viel, sehr viel.

A4: Jo. “

Die 10 Merkmale guten Unterrichts von Hilbert Meyer im Rahmen eines Unterrichtsbesuchs als Beobachtungsgrundlage zu nehmen, scheint offensichtlich schwierig, weil dies „viel, sehr viel“ sei, das es zu beobachten gelte. A6 führt zu dieser Begrenzung der Beobachtungsfähigkeit weiter aus:

A6: „[...] Und man muss, finde ich, den Referendaren auch sehr klar machen, dass wir als Beobachter uns überfordern und sie sich überfordern, wenn man auf mehr als zwei, drei Punkte, die die derzeitige Steuerung der Stunde angehen, fokussieren will, wenn wir gleichzeitig noch den Untergrund mit wahrnehmen wollen. Also niemand ist so perfekt (A3: «mhm» (bejahend)). Da müssten wir eine Stunde fünf, sechs mal auf Video angucken um die bewerten zu können. [...]“

Eine Einschätzung, die von *allen Fachleitern* im Verlauf der Diskussion geteilt wird. Die Beobachtungsfähigkeit ist eingeschränkt, sodass eine fehlende Perspektivierung im Vorfeld der Stunde, z.B. dadurch, dass der Auszubildende keine Hinweise gibt, welche Aspekte für ihn besonders im Fokus stehen, den Keim dafür zu legen scheinen, dass es in der folgenden Nachbesprechung schwieriger wird, eine angemessene Fokussierung zu erreichen. Werden zentrale Beobachtungskriterien im Vorfeld nicht geklärt, beinhaltet dies Hindernisse für eine reflexive, die Professionalisierung der Auszubildenden fördernde, Kommunikation, wie A6 in der Diskussionspassage weiter ausführt:

A6: „[...] Sonst, wenn der Referendar hinterher fragt und sagt: „Was ist Ihnen eigentlich noch mal zu dem Punkt aufgefallen? Da haben wir vor drei Unterrichtsbesuchen mal drüber gesprochen“, dann kann ich nur sagen: „UPPS, das war mir jetzt gar nicht im Blick, weil ich so auf das geguckt habe.“ Also so diese Begrenztheit der Wahrnehmung, finde ich,

spielt eine ganz große Rolle auf beiden Seiten (A1, A2, A3, A4, A5: «mhm» (bejahend), nicken) und das muss man auf jeden Fall kritisch im Blick haben.“

Kommunikations- und Beziehungsstörungen können ebenfalls aus dem Rollenkonflikt der Ausbilder entstehen, wenn diese ihre Bewertungs- und Beratungsfunktionen vermischen, also in der Ausbildungslogik ‚Beratung gegen Bewertung‘ steht. Auch hier finden sich in der Gruppendiskussion der Auszubildenden deutlich mehr Passagen, welche diese Problematik thematisieren (sechs gegenüber einer bei den Ausbildern), aber A4 spricht dieses Thema direkt in der folgenden Passage an, in der er auf die geänderten Aufgaben der Kernseminarleiter in der neuen Ausbildungsverordnung in Nordrhein Westfalen anspielt, die diesen nur noch eine beratende, aber keine bewertende Funktion mehr zuweist:

A4: „Also ich glaube, dass das bisher in meinen Augen viel zu kurz kam. Unter den jetzigen Bedingungen unserer Ausbildung, also der Ausbildung von zukünftigen Lehrern. Und dass wir durchaus jetzt ein paar Chancen eröffnet bekommen, in denen das zum Thema werden könnte. Die personorientierte Beratung ist in meinen Augen genau der Punkt (A2, A3: «mhm» (bejahend)), wo ich diesen Raum haben kann. Geschützt, weil nicht bewertet, also weil derjenige, der da, jetzt also Coach oder wie auch immer wir ihn bezeichnen wollen, nicht bewertet, und da sind genau diese Chancen. Das in den jetzigen Strukturen war immer nur bedingt möglich, und ich kann jetzt einfach nur[...]“

A3: Also, da würde ich jetzt gerne widersprechen. Also da sehe ich überhaupt keinen Unterschied. Ich finde nur jetzt, also ich verstehe was Sie (A4) meinen, und das will ich auch unterschreiben. Trotzdem glaube ich, dass die Chancen genauso da waren, und sie werden jetzt benannt irgendwo, sie werden offener gehandhabt (A6: «mhm» (bejahend)), das schon.“

A4 sieht einen zentralen Konflikt zwischen der Beratungs- und Bewertungsfunktion, wie er in dieser Arbeit schon an vielen Stellen thematisiert worden ist. Er beklagt einen fehlenden „geschützten Raum“, in dem es zu Beratungsgesprä-

chen kommen kann, die natürlich unerlässlich sind, wenn es Ziel der Ausbildung ist, an die Fähigkeit zur Selbstreflexion der Auszubildenden, wie A3 es weiter oben gesagt hat, „den Motor der Entwicklung“ anzudocken. Interessanterweise widerspricht ihm A3 an dieser Stelle und konstatiert, dass diese Chancen schon immer da waren, diese durch die Ausbildungsreform nur offensichtlicher würden. Die Ausbilder scheinen die Bedeutung dieses Rollenkonflikts als weit weniger problematisch einzustufen, als es die Auszubildenden tun. Eine adäquate Reaktion auf diese Probleme wird dadurch erschwert, wie es weiter unten gezeigt wird. Hinsichtlich der Frage, durch welche Mechanismen ‚*unproduktive Bewertungsprozesse*‘ im Referendariat entstehen, spielt die Frage um die Objektivität oder in abgeschwächter Form der intersubjektiven Nachvollziehbarkeit von Bewertungen für die Ausbilder eine große Rolle. *Alle Ausbilder* berichten über die große Bedeutung von ‚*Intransparenz und Subjektivität*‘ im Rahmen der Bewertungen, die scheinbar eine feste strukturelle Kontextbedingung der Bewertungsprozesse ist. Frappierend ist das Verhältnis von sieben codierten Diskussionspassagen, die sich um den negativ besetzt Codewert hinsichtlich der Transparenz der Bewertungen drehen, gegenüber lediglich einer einzigen codierten Textstelle, die den positiven Gegenpart thematisiert. A4 führt die strukturelle Unmöglichkeit des Loslösens von den eigenen, in der individuellen beruflichen Sozialisation gewachsenen, Bewertungsmaßstäben als zu akzeptierende grundlegende Voraussetzung eines jeden Bewertungsprozesses an, und reprozessiert hier implizit das von Luhmann postulierte Verstehensdefizit der Pädagogik (vgl. Luhmann 1986):

A4: „[...] Ich habe keine Plausibilität außerhalb meiner Plausibilitätsstrukturen, d.h. ich kann nur das gut finden, was ich gut finde und habe manchmal Schwierigkeiten, was andere, für andere gut ist. Ich beziehe das ja letztendlich immer wieder auf mich und sage: «Das ist jetzt in Ordnung gewesen, das ist sinnstiftend, das ist ok, das ist lernförderlich.» Aber so unterschiedlich die Schüler sind, die ich gar nicht konkret kenne, kann ich gar nicht immer sagen, ich kann immer das, immer wieder auf mich adaptieren. [...]“

Wie weitreichend diese subjektiven Elemente der Bewertung sind, und welche fundamentalen Auswirkungen sie auf die umfassende Einschätzung der Professionalisierungsmöglichkeiten von Auszubildenden haben, spricht A6 an, der auf die diametral entgegengesetzten Einschätzung verschiedener Ausbilder hinsichtlich derselben Referendare verweist:

A6: „Und der ist auch dann, finde ich, fallen mir jetzt mehrere Fälle ein, wenn ein Ausbilder sagt, oder eine Ausbilderin: «Ich kann mir den gut vorstellen, der hat Entwicklungspotential.» Und die anderen sagen: «Das kann ich mir ÜBERHAUPT NICHT vorstellen, dass das mal irgendwas wird!» Also da fallen mir jetzt mehrere Fälle ein. Und ja also, das finde ich wirklich eine schwere Situation.“

Man kann sich vorstellen, dass dies für Ausbilder „eine wirklich schwere Situation“ ist, wenn es mit der Benotung der Auszubildenden auch um deren Einstellungschancen und berufliche Zukunft geht, die Situation der Nachbesprechung also auch für die Ausbilder zur ‚*Stresssituation*‘ wird. Eine Zielsetzung, die an das in dieser Arbeit zu entwickelnde Bewertungskonzept gerichtet ist, besteht darin, die subjektiven Anteile der Bewertung durch kommunikative Validierung der Bewertungsmaßstäbe zu reduzieren. Ausgehend von den Äußerungen von A6, sowie weiterer Diskussionspassagen, muss die Frage aufgeworfen werden, inwiefern dieser Anspruch tatsächlich erreichbar ist. Insbesondere hinsichtlich der Frage der jeweiligen Bedeutung der Einzelmerkmale guten Unterrichts für guten Unterricht als komplexes System aus diesen Merkmalen, scheint kaum ein Konsens herstellbar. Der folgende Ausschnitt zeigt, zu welchen kontroversen Diskussionen die Frage um die Hierarchie der einzelnen Merkmale guten Unterrichts führen, die sich im Laufe der Diskussion nicht auflösen ließen. Die Diskutanten blieben bei ihren jeweiligen Positionierungen und brachen die Diskussionen ohne Konsens ab. Stattdessen kann die Bemerkung von A5 („das ist auch vom Fachleiter abhängig“) als Fazit hinsichtlich der strukturell bedingten subjektiven Anteile in den Bewertungen festgehalten werden:

A2: „Trotzdem bin ich, trotzdem glaube ich, die Reduktion und die inhaltliche klare Struktur (A1: Reduktion schon, aber I) ist am Anfang

extrem wichtig. Sachanalyse, Sachstruktur, damit überhaupt mal rein kommen (A4: Ja aber I), in den Unterricht hineinkommen (A6: Und dennoch, ich I). Und auch ein gewisses Klassenmanagement mit Ritualen, das muss auch am Anfang gewährleistet sein. Und dann kommt alles andere schrittweise.

A3: Das ist Ihre (A2) Sicht!

A2: (Lacht)

A5: Ja genau, also ich persönlich/Ich glaube, das ist dann auch vom Fachleiter abhängig. Sollte es vielleicht nicht, aber bei mir ist es ganz wichtig, wahrscheinlich auch eben weil ich eben die Voraussetzung habe, wenn ich jetzt diese drei Intensivtage habe, wenn die Neuen kommen, habe ich das schon alles organisiert. Die werden zugeknallt bei uns an der Schule mit gemeinsamem Unterricht, mit Inklusion.“

Die ‚Progression der Ausbildung‘ und deren Berücksichtigung in den Bewertungsprozessen stellt aus Ausbilderperspektive kein großes Problem dar, trotz der oben umschriebenen und in der eben zitierten Passage wiederum ausbleibenden Klärung, wie diese Progression denn nun genau auszusehen habe. Lediglich *ein Ausbilder* merkt (selbst-)kritisch an, dass Entwicklungsaufgaben, die Referendaren im Rahmen von Unterrichtsnachbesprechung gestellt wurden, ggf. – selten?, oft? – in den folgenden Lehrproben gar nicht mehr thematisiert werden, sondern das dies nur „potentiell möglich“ sei:

A4: „Das ist nicht per se gegeben, sondern ist potentiell möglich. Denn was in einer Stunde, in einer Lerngruppe, in einer Jahrgangsstufe notwendig war, muss absolut nicht in der nächsten Hospitation (A1: nicken), in einer anderen Lerngruppe, in einer anderen Jahrgangsstufe wieder anknüpfbar eingebracht werden. Es gibt solche Dinge, wie Gesprächsführung kann unter Umständen da wieder Thema werden, aber das muss es nicht. Insofern bietet so eine Dokumentation (Anm.: Gemeint ist der Dokumentationsbogen für die Nachbesprechung, D. L.), bietet eine Chance, daran anzuknüpfen, vielleicht zweite Hospitation, aber vielleicht auch NIE wieder in einer der Hospitationen, die noch ausstehen.“

In dem Zitat klingen schon die intendierte Verbesserung der Bewertungsprozesse durch die zu entwickelnden Bewertungsinstrumente an, wenn A4 vermutet, dass eine schriftliche Dokumentation der Nachbesprechung eine „Chance bietet“, an Entwicklungsaufgaben anzuknüpfen. Betrachtet man die Einschätzung eines anderen Ausbilders zu der Frage der ‚*ausbleibenden Nutzung von Bewertungsinstrumenten*‘ in der Ausbildungspraxis, keimt allerdings zunächst wenig Hoffnung auf, dass es zu einer weit verbreiteten Nutzung der Instrumente kommen kann.

A3: „[...] Also so verstehe ich auch das, die Funktion dieser Zusammenstellung (Anm.: Gemeint ist diejenige die schon im ZfsL Bonn eingesetzt wird, D. L.)¹⁰² ist für die Kommunikation über Unterricht im Kollegium sehr hoch, für die Ausbildung im Referendariat Sie (A4) wissen ja was ich denke, außer bei A4 und noch ein paar ausgewählten Leuten, die hier vielleicht auch sitzen, sehr klein.“

A3 berichtet hier über die Nutzung eines bereits entwickelten Bewertungsinstruments des ZfsL, das aber „außer von ein paar ausgewählten Leuten“ de facto nicht eingesetzt wird.

Es sollte bei den Ausführungen deutlich geworden sein, dass die an den Tag gelegten Interaktions- und Kommunikationsstrategien der Akteure die entscheidenden Faktoren sind, die zu einer ‚*Abweichung vom Ideal*‘ der Ausbildungssituationen führen und diese zeitgleich durch die Abweichung selber geprägt werden. Aus der Perspektive der Ausbilder scheinen allerdings auch strukturelle Gegebenheiten zu existieren, wie etwa die subjektiven Elemente der Bewertung, die schlichtweg als Bedingung der real-existierenden Bewertungsprozesse hinzunehmen sind, wohingegen sie für andere, aus Sicht der Auszubildenden problematischen Aspekte wie etwa die Klärung, wodurch sich die Ausbildungsprogression konkret ausdrückt, nur eine eingeschränkte Sensibilität mitbringen. Diese im folgenden Unterkapitel dargestellte Perspektive der Auszubildenden wird einerseits Unterschiede, aber auch Gemeinsamkeiten der Rezeption zwischen Ausbildern und Auszubildenden zu Tage fördern.

¹⁰² Hier wir das bereits erwähnte „Bonner Qualitätstableau angesprochen.

4.2.2 Die Sichtweise der Auszubildenden

Alle Auszubildenden, bis auf eine Person, schildern aus ihren Erfahrungen heraus Ausbildungssituationen, die in ihren situativen Kontextbedingungen als ‚ideale Rahmenbedingungen‘ zu kennzeichnen sind und deren oben dargestellten Eigenschaften entsprechen. Die Auszubildenden berichten demnach von Ausbildungserfahrungen, die sich als ‚Lernsituation‘ charakterisieren lassen. In diesen Ausbildungskontexten nehmen die Auszubildenden Rollen ein, in denen sie selbstbewusst kommunizieren und Impulse zur Selbstreflexion aufnehmen, wie es etwa in dem folgenden Zitat von R2 deutlich wird, in dem sie über den Ablauf der Unterrichtsnachbesprechungen berichtet:

I: „Hast du¹⁰³ das (Anm.: Gemeint ist der Bezug auf schriftlich fixierte Bewertungskategorien, D. L.) als gewinnbringend erlebt?

R2: Klar, dadurch, dass man, dass erst mal sozusagen die Überschrift des nächsten Gesprächsabschnitts (R5: nicken), dann einfach mal klar war.

I: Ok und das war geordnet auch nach den Kriterien?

R2: Also wenn ich irgendwie mein Statement sozusagen angefangen hab, habe ich gesagt: «Ich würde gerne das, das und das erwähnen» (R3: «mhm» (bejahend), R5: nicken).

R4: Das habe ich auch gemacht.

R2: Und dann kam eigentlich fast immer von den Fachleitern: Ok und dann sprechen wir danach noch über das und das, zum Beispiel über den Einstieg oder was weiß der Kuckuck. Oder über Gelenkstellen oder weiß ich nicht, und dann war eigentlich relativ klar was passierte in der Nachbesprechung, also nicht in jeder, aber relativ häufig habe ich das so erlebt.“

¹⁰³ Mehrere der Diskussionsteilnehmer waren dem Autor der Arbeit aus Arbeitskontexten bekannt. In einem Vorgespräch ergab sich das „Du“ als gängige Anrede unter der Diskutanten von selbst.

„Nicht jede Nachbesprechung“ ist so abgelaufen, aber „häufig“ hat R2 einen solchen, offensichtlich durchaus „gewinnbringenden“ Ablauf erlebt, und R4 pflichtet diesen Erfahrungen bei. Auch R1 hat durchaus positive Erinnerungen an die Unterrichtsnachbesprechungen und v.a. an die darauf folgende reflexive Auseinandersetzung mit Entwicklungsaufgaben im Fachseminar:

R1: „[...] Also wir waren ein sehr kleines Seminar, und dann kam immer einer und meinte, ja ich hatte Lehrprobe und ich soll jetzt hier nochmal vorstellen, weil das und das war schlecht, und jetzt habe ich das überarbeitet, also so wurde das dann auch für die Seminarsitzungen noch genutzt. Das fand ich schon gut. Also es war natürlich nochmal Zeit, die man investieren musste, aber dann hat man nochmal an dieser Stunde gearbeitet. Klar, war es dann wieder nicht unbedingt übertragbar auf eine andere, aber zumindest hat man dann gesehen: Ok, ich stelle jetzt da nochmal vor, was habe ich mir überlegt, was könnte man anders machen und dann wird es auch vom Seminarleiter nochmal kommentiert. [...]“

Über die Bedeutung der ‚*Beziehung zw. Ausbildern und Auszubildenden*‘ für die Realisierung einer ‚*Lernsituation*‘ tauschen sich die Auszubildenden kaum aus, aber mehr als die Hälfte der Auszubildenden, die ihre Ausbildung explizit als überwiegend positiv und gewinnbringend wahrgenommen haben, sprechen ihre im positiven Sinne erlebten Beziehungen zu den Ausbildern zumindest an:

R5: „Es ist halt eben auch, ich meine, die Referendare und die Fachleiter sind beide Menschen und deswegen hat man da einfach natürlich, muss die Chemie auch irgendwie stimmen und man muss irgendwie miteinander klarkommen. Man kann es nicht komplett von jetzt, der einzelnen Persönlichkeit trennen (R4: Richtig). Ich hatte jetzt zum Beispiel absolutes Glück, ich kam mit allen Dreien wunderbar zurecht und hatte nie irgendwelche Probleme und konnte auch wirklich die Nachbesprechungen auch häufig wirklich dafür nutzen, das auch mal wirklich durchzudiskutieren und meine Meinung dazu vertreten. Und/also manchmal war es wirklich so, dass es einfach eine wunderbare Diskussion war. Gar nicht so, wie man eine Nachbesprechung kennt, einfach weil es auf der zwischenmenschlichen Ebene gut lief.“

Zwischen Ausbilder und Auszubildenden „muss die Chemie stimmen“, wie es R5 lapidar auf den Punkt bringt und R4 ihr beipflichtet. Bemerkenswert an diesem Zitat ist der letzte Satz, in dem R5 ihre eigenen positiven Erfahrungen relativiert, in dem sie diese gegenüber dem „normalen Verlauf“ einer Unterrichtsnachbesprechung abgrenzt. Offenbar erscheint es ihr ‚anormal‘, dass es bei Unterrichtsnachbesprechungen auf der „zwischenmenschlichen Ebene gut läuft“, was umgekehrt bedeutet, dass eine durch zwischenmenschliche Konflikte geprägte Nachbesprechung für R5 also der Normalfall sein muss, ihre eigenen positiven Erfahrungen also die Ausnahme darstellen. Ein deutlicher Hinweis auf die Wirkmächtigkeit der eingprägten Mythen und Ängste, mit denen die Auszubildenden in die Interaktionssysteme eintreten.

Diese existierenden Überzeugungen prägen auch auf Auszubildendenseite ihre ‚Strategien als Akteure‘ innerhalb der Interaktionssysteme und sorgen, wie weiter unten gezeigt werden wird, für eine Form der Interaktion und Kommunikation die den ‚idealen Ausbildungsbedingungen‘ oft zuwider läuft. Gleichwohl berichten die Auszubildenden auch über eine ‚dialogische, reflexive Kommunikation und Interaktion‘ mit den Ausbildern, die sie durchaus als gewinnbringend im Sinne der ‚Lernsituation‘ erleben. R2 berichtet von ihrer Herangehensweise an die Unterrichtsnachbesprechungen, die offensichtlich einerseits der idealen Rollenerwartung der Auszubildenden, selbstbewusst und selbstreflexiv aufzutreten, entspricht und andererseits auch deutliche Übereinstimmung mit der Forderung nach einer ‚fokussierten, situationsangepassten Perspektivierung‘ aufweist:

R2: „Aber es ist dann auch utopisch über alles sprechen (R4: absolut). Also ich habe das schon erlebt, dass man selber gesagt hat: «Dadadada darüber würde ich sprechen» und dann sagt der Fachleiter: «Und dann nehmen wir noch bitte das dazu, weil da habe ich jetzt Mehreres zu aufgeschrieben.» Also das habe ich schon anhand dieser Bubbles¹⁰⁴ tatsächlich so erlebt.“

¹⁰⁴ Die ‚Bubbles‘ sind ein weiteres Instrument des ZfSL Bonn, das das Bonner Qualitätstableau in verkürzter Form wiedergibt. Die Bubbles sind im Anhang dokumentiert.

Die Bedeutung der Hierarchieverhältnisse zwischen Ausbildern und Auszubildenden für die Realisierung einer ‚*dialogischen, reflexiven Kommunikation und Interaktion*‘ lässt sich aus dem folgenden Zitat von R4 herauslesen, in dem es um eine selbstbewusste Offenbarung der eigenen Sichtweise der Auszubildenden geht:

R4: „[...] Und wenn das so ist wie bei dir (R5), wie du es erlebt hast und ich teilweise auch, muss ich sagen, dass man möglichst auf Augenhöhe zumindest diskutieren kann, (R5: nicken) das ist, das finde ich, sollte der Ideall, das wäre der Idealfall und ich denke, dass würde auch auf jeden Fall dazu beitragen, dass man auch offener miteinander umgeht, dass man sich auch traut, mal zu sagen: „Kann ich jetzt hier nicht verstehen“/[...]“

„Zumindest diskutieren“ sollte man auf Augenhöhe können, und „teilweise“ hat R4 dies in den Nachbesprechungen auch erlebt. Aber auffällig ist an dieser Stelle natürlich die im Konjunktiv gehaltene Formulierung: „Das würde auf jeden Fall dazu beitragen, dass man auch offener miteinander umgeht“, auch offen Uneinigkeit mit dem Ausbilder ausdrückt. Diese Art der Kommunikation ist für R4 „das Ideal“, aber inwiefern dieses Ideal eine normative Wunschvorstellung oder tatsächlich erlebte Ausbildungspraxis ist, wird nicht endgültig klar. Den Beitrag, den diese Art der Kommunikation für gelingende Bewertungsprozesse leistet, zeigen auch die Erfahrungen von R5 auf, die für ihre Art der Gestaltung der Nachbesprechung immer „ganz gute Rückmeldungen“ bekommen hat.

R5: „[...] Und also ich hatte es mir auch immer in den Nachbesprechungen, war für mich auch wirklich dieser Punkt ganz wichtig, dass ich immer gesagt habe: „So das war gut, das war schlecht und HIER hätte ich dieses oder jenes so und so machen können“, das war für mich immer so meine Abfolge an der ich mich entlang gehandelt habe. Und dafür habe ich auch immer ganz gute Rückmeldungen dann bekommen.“

Auf den gewinnbringenden Beitrag einer ‚*fokussierten, situationsangepassten Perspektivierung*‘ verweist wiederum R2:

R2: „Die anderen Mitreferendare von mir, wo man das auch beobachten konnte. Ich habe aber auch tatsächlich solche Sachen mitgekriegt: «Sie

hatten jetzt in ihrer Stunde hier viel zu wenig Reflexionszeit am Ende, sie haben irgendwie, quasi die Früchte ihrer Arbeit da nicht ernten können, beim nächsten Mal 15 Minuten einplanen, das machen sie beim nächsten Mal.» Zack, bumm (R5: nicken). Da macht die Frau R2 und dann wurde es auch besser, also es war relativ klar, was passieren sollte, aber das ist vielleicht auch so die Luxusversion gewesen, auch nicht immer 27 Sachen, sondern das ist so der Hauptaugenmerk für das nächste Mal, und das machen Sie dann erst mal.“

„Dann macht die Frau R2 das und dann wurde es auch besser“. Die Hilfestellungen durch die Ausbilder, bestimmte Aspekte betonend, herauszugreifen und in den folgenden Unterrichtsbesuchen einzufordern fördert die Professionalisierung, „ist vielleicht so die Luxusversion“ der Ausbildung. Demgegenüber stehen Forderungen von Ausbildern „27 Sachen“ auf einmal beachten zu müssen. Der durchaus lenkenden, aus der Situation heraus geborenen Perspektivierung durch die professionelle Expertise der Ausbilder kommt für gefühlte Lernerfolge auf Seiten der Auszubildenden offenbar eine große Bedeutung zu, wie auch R5 ausführt und R4 bekräftigt.

R5: [...] Und wenn man dann sagt: «So jetzt handeln wir den Bereich ab, den Bereich ab, den Bereich ab», dann kann man noch Verknüpfungen ziehen zwischen zwei Bereichen, kann aber auch immer drauf zeigen und sagen: «Hier und dort.» Und damit glaube ich, würde man als Referendar auch leichter folgen können. Gerade auch in der Anfangszeit, wo man völlig überfordert ist, von all dem, was man erzählt bekommt.

R4: Stimme ich dir vollkommen zu. Also hilfreich, definitiv. [...]“

Die Rollenerwartungen an die Ausbilder sind nicht verwunderlich: Sie sollen durch ihre Expertise Orientierung geben und, v.a. am Anfang der Ausbildung, Durchblick in der unüberschaubaren Vielfalt des Unterrichts schaffen. Aber auch hier findet sich wieder eine Formulierung im Konjunktiv („würde man als Referendare leichter folgen können“), die fraglich werden lässt, ob die Ausbilder von R5 diesen Anspruch tatsächlich eingelöst haben oder es das von R4 zitierte vor-

nehmlich normative „Ideal“ ist. Für die Referendare ist weiterhin die beratende Funktion der Ausbilder entscheidend, wenn es um die Realisierung einer ‚*Lernsituation*‘ geht. Die Ausbilder müssen ‚*beraten und bewerten*‘ und v.a. Alternativen zu den aus ihrer Sicht nicht gelungenen Aspekten des Unterrichts eines Referendars aufzeigen können. Dies wird v.a. aus den weiter unten thematisierten Erfahrungen der Auszubildenden deutlich, in denen die Ausbilder dieser Doppelrolle nicht gerecht werden konnten. Auf der positiven Seite dieser Kategorie berichten *mehr als die Hälfte der Auszubildenden* über hilfreiche Impulse ihrer Ausbilder. Neben den oben bereits zitierten Erfahrungen von R2 zeigen die folgenden Zitate von R4 und R5 die positive Rolle, die beratungsorientierte Gespräche mit den Ausbildern für ihre Professionalisierung hatten:

R4: „Mhm, bei mir persönlich, teils, teils, also beim einen Fachseminar tatsächlich immer sehr auf den Punkt gebracht auch, was war gut, was war schlecht, klar, aber dann auch tatsächlich gesagt: «Wie könnten wir das, also auch gemeinsam überlegen (R2: nicken), was gäbe es für Möglichkeiten, wie könnten wir das, wie könnte man das anders lösen?» Natürlich überlegt man dann auch selber und bringt vielleicht seine Ideen, dann kommt aber, also kam zumindest dann von Seiten der Fachleiterin auch wirklich Tipps, die ich nachvollziehen konnte und Hinweise, die ich nachvollziehen konnte. [...]“

R5: „Ich hatte sehr viele, ja konstruktiv-kritische Gespräche, wo wirklich auch Alternativen erarbeitet wurden. Hatte dann aber wieder das Problem, dass ich dann in der nächsten Nachbesprechung irgendwas so gemacht habe, wie es dort als besserer Weg aufgezeigt wurde und es dann hieß: «Nein, nein» und jetzt wieder komplett zurück zu dem, was ich vorher gemacht habe, und dass, da fehlte mir dann wiederum die Transparenz, aber Handlungsoptionen habe ich da auf jeden Fall aufgezeigt bekommen.“

Neben den positiven Aspekten zeigen diesen beiden Zitate aber gleichzeitig auch die Grenzen der Beratung. Offensichtlich war die Beratung bei R5 nicht so deutlich, dass es zu einer klaren Verständigung über zu erreichende Professio-

nalisierungsschritte kommen konnte, und R4 hat die positive Bedeutung nur in einem Fachseminar erlebt.

Ähnlich wie bei der Diskussion der Ausbilder finden sich wenige Passagen, in denen die Auszubildenden über ‚*Transparenz und intersubjektive Nachvollziehbarkeit*‘ oder gar Objektivität der Bewertungsprozesse berichten. R3 schildert allerdings Erfahrungen mit dem Beobachtungsinstrument einer Ausbilderin, das zu einer gewissen Transparenz der Bewertungsgrundlage geführt hat:

R3: „Also es war ein DINA 4 Blatt, sie meinte nämlich unfassbar wichtig ist, dass es übersichtlich ist und das man wirklich so auf einem Blatt halt sieht. Und da waren dann verschiedene Kriterien aufgelistet, wie aktive Bewegungszeit, wie was für den Sportunterricht wichtig ist, kognitive Phasen. Also Kriterien waren links aufgezählt, dann hatte sie, glaube ich, so drei Kategorien, inwiefern das sehr erfüllt ist oder nicht so sehr erfüllt und sich selber dann rechts noch eine Spalte gelassen, in der sie dann immer schnell einträgt, was ihr dann noch halt an Notizen noch einfällt, was man in der Stunde besonders gesehen hat oder was verbesserungswürdig wäre. Also relativ übersichtlich, was eigentlich andererseits ganz angenehm war, aber dafür vielleicht auch etwas undifferenziert.

I: Fandest Du das gewinnbringend?

R3: Das vorher gesehen zu haben? Dass sie das so bewertet?

I: Unter anderem, die ganze Art der Bewertung, so mit so einem standardisierten oder halb-standardisierten Bogen vorzunehmen.

R3: Ja doch. Das wäre wieder das – ehrlich gesagt – was wir eben gesagt haben, schnellstmöglich herauszufinden, was der Fachleiter von einem möchte, worauf es ankommt und das vor Augen zu haben, auch in einer übersichtlichen Art und Weise, hilft auf jeden Fall.“

Das Vorliegen schriftlich fixierter Bewertungsmaßstäbe führt vordergründig zu mehr Transparenz und intersubjektiv nachvollziehbaren Bewertungsprozessen.

sen, aber auf einer anderen Ebene, als es dem oben herausgearbeiteten Ideal der Rollenerwartungen an die Adresse der Auszubildenden entspricht. Die offen gelegten Bewertungsmaßstäbe sorgen nicht für eine Klärung des Verständnisses der Eigenschaften guten Unterrichts, sondern – „ganz ehrlich“ – für die Möglichkeit, das eigene Handeln strategisch an den Erwartungen des Ausbilders anzulegen. In eine ähnliche Richtung weisen die Berichte von R2, deren Ausbilder offenbar klare Vorgaben zu denen im Rahmen der Unterrichtsbesuche zu zeigenden Stunden gemacht hat:

R2: „[...]Ich lief gar nicht die Gefahr eventuell die gleiche Sache ein paarmal zu machen, weil wir einfach in Musik verschiedene Studententypen einfach abarbeiten mussten. Da hatte der seine Liste drüber, Ende im Gelände, also da hatte ich das Problem sowieso nicht.“

Klare Vorgaben und schriftlich fixierte Bewertungsgrundlagen kennen also einige der Auszubildenden, aber Berichte über wirkliche ‚*Transparenz und intersubjektive Nachvollziehbarkeit der Bewertungen*‘ im Sinne einer für die Auszubildenden begründeten Klärung der Bewertungsmaßstäbe finden sich nicht. Gleichwohl werden teilweise durchaus ‚*produktive Bewertungsprozesse*‘ erlebt, aber v.a. dann, wenn die Auszubildenden die Maßstäbe der Ausbilder für sich genommen akzeptieren können. Es gibt Vorgaben und „Ende im Gelände“, wie R2 es betitelt. Ein Zusammenhang, der sich auch im folgenden Zitat von R3 widerspiegelt, die darüber berichtet, auf welche Art und Weise die Auszubildenden eines der Bewertungsinstrumente des ZfsL, die bereits oben genannten ‚Bubbles‘, nutzen sollten:

R3: „«Bei einem Statement darf es nicht darum gehen, eure Stunde von hinten bis vorne einfach nur der Reihe nach durchzugehen. Ihr sollte euch nach unseren Bubbles orientieren. Nehmt euch drei Bubbles und das ist, analysiert kriteriengeleitet.» Das wurde einmal so was von klar gesagt und da ist man irgendwie auch relativ schnell hinter gekommen, dass man dann also nicht der Stunde nachgeht und sagt: «Am Anfang habe ich ja das gemacht und so», Sondern halt wirklich diese drei Dinger hatte.“

Die Auszubildenden sind „dahinter gekommen“, wie sie ihre Reflexion zur Stunde aufbauen sollen, aber ob sich ihnen der Sinn einer kategorialen Reflexion, in Abgrenzung von einer additiv-chronologischen im Sinne der Professionalisierung offenbart hat, verbleibt unklar. Die Auszubildenden scheinen sich mit den Ausbildern allerdings dahingehend einig zu sein, dass die ‚*Progression der Ausbildung*‘ Niederschlag in ‚*produktiven Bewertungsprozessen*‘ finden sollte, gerade vor dem Hintergrund der Überforderung mit der Komplexität des Unterrichtsgeschehens am Beginn der Ausbildung.

R2: „Also ich würde mir, also ich glaube ich könnte, würde das als gewinnbringend dann empfinden, wenn man bestimmte Sachen raus sucht (R4: Genau, R5 «mhm» (bejahend)), also wenn man irgendwie nicht alle 23 Punkte bespricht (R5: nicken.), sondern irgendwie, weiß ich nicht, im Seminar dann sagt: «Am Anfang achten Sie besonders auf die Punkte ddddmmmmm», und da gucken wir dann mal nach und wenn das in Richtung UPP-Standard¹⁰⁵ oder was auch immer wo drüber gesprochen wird, gehen könnte, dann könnte man mehr Instrumente oder mehr Punkte noch genauer dazunehmen. Dann würde es mir, glaube ich, helfen.“

Allerdings wird weiter unten deutlich, dass für die Auszubildenden eine nur in diffusen Begriffen gefasste Progression deutlich problematischer ist, als dies die Ausbilder sehen. Liegen Bewertungsinstrumente oder auch nur schriftlich fixierte Bewertungsmaßstäbe vor, so scheinen diese von den Auszubildenden durchaus als gewinnbringend angesehen zu werden. Die ‚*Nutzung der Bewertungsinstrumente*‘ sorgt für Klärungsprozesse hinsichtlich der an die Auszubildenden gerichteten Erwartungen, wie z.B. in den ganz zu Beginn dieses Unterkapitels zitierten Erfahrungen von R2, in der sie über die Bedeutung der vom Seminar genutzten ‚Bubbles‘ berichtet und auf die klare Strukturierung der Nachbesprechung durch dieses Instrument verweist. Eine ähnliche Hilfestellung bot dasselbe Instrument für R3, die ansonsten überwiegend kritisch über die Unterrichtsbesuche und -nachbesprechungen, sowie die Ausbildung generell berichtet. Die ‚Bubbles‘ scheinen aber auch ihr die notwendige Orientierung in stressigen Situationen gegeben zu haben:

¹⁰⁵ UPP = unterrichtspraktische Prüfung im Rahmen des zweiten Staatsexamens

R3: „Fürs Statement war es wirklich in dem Sinne allein schon mal, weil man nach der Stunde oftmals vollkommen erschöpft aus dem UB rauskam, also hatte man endlich hinter sich und da war auch wieder diese Übersichtlichkeit: Ein DIN A 4 Blatt mit ein Paar Unterpunkten und ich konnte nochmal echt sagen: «Boah, in dieser Stunde ist mir das richtig gut gelungen, aber das richtig schwer», weil allein nur so vier Unterpunkte, und du wusstest: Aha genau Strukturierung, das ist wichtig (R2, R4: «mhm» (bejahend)), hat das heute bei mir gut geklappt, auf jeden Fall, das kann ich gut ins Statement einbringen oder das hat halt überhaupt nicht geklappt (R5: «mhm» (bejahend), nicken), das muss ich wahrscheinlich erwähnen. Das war unfassbar hilfreich, fand ich, immer nach diesem UBs, nachdem man auch echt platt und fertig war (R5: «mhm» (bejahend)), nicken), was ganz Übersichtliches und Einfaches zu haben und da die Schwerpunkte zu suchen, zu sagen: «Hier die drei suche ich mir jetzt raus.»“

Auch R5 bekräftigt den Nutzen dieses Instruments. Auch wenn sie „Listen lieber mag“, haben die fixierten Merkmale guten Unterrichts dabei geholfen auch im Vorfeld von Unterrichtsbesuchen zu überprüfen, ob sie zentrale Merkmale guten Unterrichts berücksichtigt hatte:

R5: „Ich mochte lieber Listen, aber vorher war es ganz angenehm, wenn man seine Stunde geplant hat, als Vorbereitung so. Man konnte daran auch nochmal seine Stunde checken und schauen, welche Punkte habe ich denn hier in der Stunde von vornherein versucht zu berücksichtigen (R4: «mhm» (bejahend)), konnte sich die schon mal markieren und dann später im Statement gucken: Ok sind jetzt das auch die, die ich im Statement wirklich ansprechen will oder schmeiße ich einen raus, nehme ich einen anderen rein, konnte also anhand dieser Punkte auch nochmal seinen Plan durchgehen und das später nochmal für sich selbst überprüfen. Also das war schon wirklich sinnvoll. (...)“

Inwiefern die Auszubildenden die Instrumente genutzt haben, um ihre individuelle Professionalisierung im Sinne des von den Ausbildern vertretenen Verständnisses guten Unterrichts voranzutreiben und sich ggf. auch gegen dieses abzugrenzen, wenn es ihrer eigenen Lehrerpersönlichkeit widersprach, oder diese

doch eher im Sinne von „Checklisten“ genutzt haben, die an sie gerichteten Erwartungen zu erfüllen, muss an dieser Stelle offen bleiben.

Betrachtet man diese Perspektive der Auszubildenden auf die ‚*idealen Rahmenbedingungen*‘ der Unterrichtsbesuche und Unterrichtsnachbesprechungen, zeigen sich mit den Ausbildern geteilte Erfahrungen und Wünsche, aber auch strukturell bedingtes Konfliktpotential, das aus den gegenseitigen Rollenerwartungen der beiden Akteursgruppen resultiert, die nicht immer komplementär sind. Eine ‚*dialogische, selbstreflexive Verständigung*‘ birgt auch aus der Perspektive der Auszubildenden das größte Potential für die eigene Professionalisierung, wie aus den Berichten hervorgeht, allerdings wird aus der Diskussion nicht deutlich, wie oft die ‚*Beziehungen zwischen den Ausbildern und den Auszubildenden*‘ diese Art der Interaktion zu lassen oder die Auszubildenden diese Interaktion nicht viel mehr als Ideal sehen, welches auf Grund der hierarchischen Beziehungen zu selten in die Realität umgesetzt werden kann. Aus den Ausführungen der Auszubildenden lässt sich eine relativ deutliche Rollenerwartung an einen ‚idealen Ausbilder‘ formulieren: Dieser kann klar zwischen seinen Funktionen des Bewertens und Beratens unterscheiden und gibt den Referendaren durch klar umrissene Entwicklungsaufgaben und Hilfestellungen Orientierung für das Erreichen des Ausbildungsziels.

Die teils eher skeptisch vorgetragene Erfahrungen der Auszubildenden hinsichtlich der Realisierung der ‚*idealen Rahmenbedingungen*‘ der Ausbildung weisen auf die Strukturen und Prozesse hin, die aus ihrer Sicht zu einer ‚*Abweichung vom Ideal*‘ führen, in deren Kontext Interaktionssysteme entstehen, die als ‚*Stresssituation*‘ charakterisiert werden können. *Mit einer Ausnahme* berichten *alle Auszubildenden* von Erfahrungen, die den Eigenschaften der ‚*Stresssituation*‘, wie sie hier dargestellt wird, entsprechen. Prägend für diese Art des Interaktionssystems sind für die Auszubildenden Frustration, Ängste und Erschöpfung, die sie aus den Unterrichtsbesuchen und -nachbesprechungen mitnehmen, teilweise gepaart mit sehr grundlegenden Zweifeln an der eigenen Lehrbefähigung, wie es z.B. R4 im folgenden Zitat berichtet:

R4: „[...] Da kommt auch dieses negative Gefühl zustande, dass man sich tatsächlich, ja man sich irgendwie, dass man rausgeht aus einer Besprechung und sich total, ja, schlecht fühlt, weil man eben denkt: «Ok, du kannst ja scheinbar doch, bist ja scheinbar doch kein guter Lehrer, vielleicht doch nicht dein Beruf», und so. Und das geht, glaube ich, nur dann, wenn man auch so dieses, ja, diese Bewertung dann bekommt (R2: nicken, «mhm» (bejahend)).“

Die – gefühlt – fehlende Würdigung für bereits erreichte Kompetenzfortschritte durch die Ausbilder scheint einen relativen großen Anteil am Entstehen von Frustration und Selbstzweifeln der Auszubildenden zu haben, wie ebenfalls R4 ausführt:

R4: „[...] Weil, seien wir ehrlich, jede Stunde, auch wenn es vielleicht eine schlechte Stunde ist, hat trotzdem Sachen, die man vielleicht auch positiv erwähnen soll. Man sollte die natürlich nicht in den Mittelpunkt rücken, Augenwischerei nicht, aber schon, dass man, ja eine Würdigung der Leistung der Mühe, die sich einfach jeder Referendar für so eine Stunde macht. Wir kennen das alle, wie lang man daran sitzt und wie viel Arbeit und Zeit man investiert. Ich glaube, das ist auch dann das, das ist jetzt natürlich wenig was so greifbar ist, es ist eher vielleicht auch eine persönliche Geschichte, aber das ist, glaube ich, auch ganz wichtig, dass da so eine gewisse Wertschätzung erfolgt, dass man nicht aus dieser Nachbesprechung rausgeht und denkt: «Ich kann nie ein guter Lehrer werden. Das ist ja unglaublich. Ich habe ja gerade so viel falsch gemacht.»[...] Ich glaube, das ist sehr demotivierend (R3: Ja) und sorgt dafür, ja dass die Kollegen und Kolleginnen dann auch wirklich (R3: weinten), ja weinten/[...]“

Die Unterrichtsnachbesprechungen und die Prüfungssituationen erscheinen für die Referendare als Situationen mit einem hochgradigen Stresspotential. Dies kann zu einem guten Stück mit dem Wesen von Prüfungssituationen erklärt werden, aber die Berichte der Auszubildenden offenbaren Gefühle und negative Emotionen, die teilweise als kleine Traumata verstanden werden können und über das normale Maß von Prüfungsanspannung deutlich hinaus gehen. Exemplarisch hier für steht der Bericht von R1, die ansonsten überwiegend in einer

positiven Art von ihrer Ausbildung berichtet, aber bis „auch heute noch nicht gerne“ an ihre Prüfung denkt, einfach „weil die Atmosphäre so mies war“:

R1: „Ja, vor allem passiert das ja auch nicht immer mit Worten, weil bei mir war ja das Schlimme meine Prüfung und da wird ja gar nicht kritisiert, da wird ja nur gefragt und das war viel schlimmer als jede Lehrprobe, die ich hatte, also ich denke auch heute noch nicht gerne daran und ich komme nicht, immer noch nicht so richtig damit klar, einfach weil die Atmosphäre so mies war und da ist es ja wirklich, wenn man so darüber nachdenkt, da hat ja keiner gesagt: «Das war schlecht, das war schlecht», sondern einfach nur so die Art, wie die sich so verhalten haben.“

Hierbei muss gesagt werden, dass das subjektiv empfundene Ausmaß des Stresses zwischen den Beteiligten der Gruppendiskussionen stark schwankt. Während R3, wie es im weiteren Verlauf der Auswertung deutlich werden sollte, sehr mit Stress, Unsicherheiten und Ängsten im Kontext der Unterrichtsnachbesprechungen zu kämpfen hatte, berichten die anderen Teilnehmer der Diskussion eher von einzelnen Situationen, die sie emotional gefordert oder auch in Einzelfällen an die die Grenzen ihrer Belastbarkeit gebracht haben. Diese Situationen werden als unangenehm erlebt, aber als Teil der Ausbildungsrealität anerkannt. R2 scheint aus ihrer Ausbildung keine derart negativen Erfahrungen zu kennen. Eine große Rolle für das Entstehen dieser negativen Emotionen, dies wird aus den Einschätzungen *aller Auszubildenden* deutlich, sind die hierarchischen ‚*Beziehungen zwischen Ausbildern und Auszubildenden*‘, die aus Sicht der Auszubildenden kaum dem Ideal entsprechen, das weiter oben herausgearbeitet wurde. Vielmehr ist da „immer dieser Rangunterschied, der da vor einem sitzt“, wie R4 ausführt, der dazu beitrage, dass die Auszubildenden so eingeschüchtert sind, dass sie in eine passive Schülerrolle fallen, anstatt dem von den Ausbildern gewünschten Ideal des selbstbewussten Novizen gerecht zu werden:

R4: „[...] Das ist glaube ich auch, was man so immer unterschwellig spürt, diese, ja dieser Unterschied, dieser Rangunterschied, einfach diese Autorität, die da vor einem sitzt und das ist das, worüber wir schon öfters im Seminar oder in der Referendarsgruppe auch geredet haben, dass man

sich auf einmal wieder wie ein Schüler fühlt (R2: nicken), dass man da auf einmal wieder sitzt und sich dabei ertappt, wie man sich rechtfertigt, wie man ja nervös wird, obwohl man eigentlich tatsächlich, ja man ist examiniert, man hat schon einen Hochschulabschluss, also man sollte da eigentlich tatsächlich mit mehr Selbstbewusstsein sitzen, aber man gerät unheimlich schnell durch diese, also dieses Gefälle einfach auch, dadurch dass man von den Leuten, die da sitzen, bewertet wird, gerät man unheimlich schnell in diese Schülerrolle, sag ich mal, wo man da sitzt und sich nur noch rechtfertigt und völlig unsicher wird.“

Diese Problematik der hierarchischen Beziehungen und der damit verbundenen Hemmung der Auszubildenden, sich mit ihren eigenen Ansichten und Überzeugungen zu Wort zu melden, zeigt auch die folgende Diskussionspassage, in der die Teilnehmer über die Frage diskutieren, wie offen verschiedene Qualitätseinschätzungen zum gesehenen Unterricht in der Nachbesprechung mit Hilfe von Bewertungsinstrumenten festgehalten werden könnten:

R3: „[...] Wenn beide jetzt ihr Kreuz machen sollen, weil ehrlich gesagt, sind wir an der Stelle wieder angekommen, wo wir ganz am Anfang mal gesagt haben, man kuscht, ehrlich gesagt, viel, muss man jetzt einfach mal so darstellen und wenn ich mein Kreuz nach meinem Fachleiter machen sollte, setze ich das bestimmt nicht gerade großartig woanders hin [...] Von der Grundidee her verstehe ich es, ist es wahrscheinlich auch schön und gut, wenn beide ihr Kreuz machen müssten, von der Umsetzung her finde ich es vielleicht dann irgendwie, wird es irgendwie unangenehm wieder mit diesem hierarchischen Gefälle.

R4: Aber habe ich das jetzt richtig verstanden, dass das wirklich so laufen soll, wie es R3 gerade beschrieben hat, dass quasi beide sich/

I: Es war eine Idee zum diskutieren.

R4: Das beide ihr Kreuzchen, also weil das finde ich, da würde ich dir (R3) zustimmen, das fände ich auch wieder.

R5: Am Tisch? Och ne (kopfschütteln).[...]

I: R1, du nickst.

R1: Nee, ich glaube auch, dass man das nicht so in der Mitte machen sollte. Also v.a. weil ich auch jemand bin, ich passe mich dann auch gerne (R3, R4: «mhm» (bejahend)) an. Und dann wäre es also besser, wenn der eine das Blatt hat und der eine Folie und dann legt man die übereinander.

R2: Ja und dann guckt man, was sich getroffen hat und was macht man, wenn man es nicht getroffen hat, also wenn es sich nicht trifft, dann redet man drüber oder dann sagt der Fachleiter: «Da haben sie sich leider falsch eingeschätzt.» Was macht man denn dann (R3: «mmh» (bejahend))?»

Von der „Grundidee“ ist die Schaffung von Transparenz „schön und gut“, aber durch die hierarchischen Beziehungen, da sind sich *alle Auszubildenden* einig, werden die Auszubildenden sich dem Urteil des Fachleiters unterordnen und eben nicht die eigene Position vertreten. Sie halten es also für sehr unwahrscheinlich, dass die Auszubildenden sich im Sinne einer – von beiden Akteursgruppen als Idealbild eingeschätzten – selbstbewussten Rolle verhalten. R2 geht in diesem Kontext sogar so weit zu hinterfragen, inwieweit es „bei jedem Ausbilder“ ein tatsächliches Interesse gibt, mit den Auszubildenden in einen offenen Diskurs zu treten oder ob diese nicht vielmehr denken: „das ist mir doch jetzt egal“:

R2: „Und ich frage mich, ehrlich gesagt, auch, inwieweit die Fachleiter das annehmen könnten, wenn ich das jetzt ganz anders einschätzen würde. Können die ja auch irgendwie so denken: Das ist mir doch jetzt egal (R4: «mhm» (bejahend)), was der da für ein Kreuz gemacht hat. Es ist jetzt das Kreuz von Belang, was ich hier gesetzt habe. Das ist jetzt auch wieder überspitzt vielleicht oder fies formuliert aber (...) ist halt die Frage. Also würde ich in Frage stellen, ob das bei jedem Fachleiter von Interesse ist, wie der Referendar sein Kreuz setzt (R1: «mhm» (bejahend)).“

Diese Ängste und Erfahrungen prägen die ‚Strategien der Akteure‘, mit denen sie in die Unterrichtsbesuche und -nachbesprechungen eintreten, und gleichzeitig sorgen die in diesem Rahmen sich entfaltenden dysfunktionalen Interaktions- und Kommunikationsmuster für das Entstehen von Interaktionssystemen, die Frustrationen, Ängste und Erschöpfung nähren. Ein zentraler Punkt in diesem Zusammenhang scheint die nicht zur Zufriedenheit der Auszubildenden wahrgenommene Beratungsfunktion (‚Beraten vs. Bewerten‘) der Ausbilder zu sein, wenn diese zwar im Rahmen von Bewertungen misslungene Unterrichtspraktiken feststellen, aber keine Alternativen für diese anbieten (können). R3 berichtet in diesem Zusammenhang von ihren Problemen, Kompetenzfortschritte zu erlangen, weil sie „zwar wusste, was falsch war“, aber nicht, wie sie es „hätte anders machen können“:

R3: „Was mir zum Beispiel sehr gefehlt hat insgesamt, war konstruktive Kritik. Dass, wenn man jetzt also über etwas gesprochen hat und es war klar, das und das ist nicht gut gelungen, dann fehlte immer dieser Schritt weiter, jetzt können wir mal gucken, sozusagen, was macht das denn aus, was wäre jetzt ein guter Weg gewesen und dann wieder zurück zu finden in meine Stunde und zu sagen: «Das wären Ideen, wie sie es hätten umsetzen können.» [...] Ich wusste am Schluss, was falsch war, das stand mit Kriterien auch irgendwie vielleicht fest, aber nicht, wie ich die jetzt dann wieder, wie die so formuliert sind oder ich da so was mit rausnehmen kann, dass ich sage, hätte ich auch verstanden, wie ich es hätte anders machen können logischerweise, und dann wäre ich bei dem Kriterium wahrscheinlich auf einen grünen Zweig gekommen.“

Weitergehend berichtet R3 von den sich immer weiter zuspitzenden emotionalen Problemen, die durch die – gefühlt – ausbleibende Beratung entstehen und die, nach ihren Erfahrungen, offenbar relativ weit verbreitet sind. Diese führen schlussendlich zu „Verzweiflung“ und dem Zweifel an der eigenen Eignung für den Lehrerberuf:

R3: „[...] Vielleicht hätte ich noch nicht mal unbedingt Lob gebraucht, aber wirklich diese Verzweiflung, jetzt nicht daraus ablesen zu können, was

soll ich denn beim nächsten Mal besser machen? Und wenn es mir nicht klar ist, und ich keine Möglichkeit sehe und keine Wege sehe, herrscht totale Verzweiflung und Panik. Und das war echt, fand ich, ein Tenor sollte gesagt haben: «Oh Gott, ja: einmal, klar, alles schrecklich und ich wüsste jetzt nicht mal, wie ich es anders machen sollte das nächste Mal.» Und das ist natürlich ein Gefühl, mit dem man, wie gesagt, in eine Nachbesprechung mit unfassbar viel Angst geht, in den nächsten UB mit unfassbar viel Angst geht und im schlimmsten Fall, dieses Thema: Manchmal überlege ich, ob ich das hier überhaupt zu Ende bringe, nicht relativ selten, irgendwie dann so in unseren Seminaren im Raum rumschwebte.“

Mit dieser dramatischen Einschätzung ist R3 in der Diskussionsrunde alleine, R2 relativiert diese dahingehend, dass sie diese „ganz schlimmen Erfahrungen“ mit Unterrichtsbesuchen nicht habe. Die anderen Auszubildenden pflichten R3 in abgeschwächter Form bei, sehen aber keinen derart dramatischen Stressor in der ausbleibenden oder unbefriedigenden Beratung durch die Ausbilder, wie z.B. R4, der darüber berichtet, dass er in einem seiner Fachseminare keine hilfreichen Beratungsangebote für seine Professionalisierung erhalten hat:

R4: „[...D]a kamen dann zwar auch Vorschläge, die aber für mich weniger nachvollziehbar waren.“

Betrachtet man die quantitative Verteilung der hier als negativ eingeschätzten ‚strategischen, verdeckten und konfliktären Kommunikation und Interaktion‘ zwischen Ausbildern und Auszubildenden, dann ist die unterschiedliche Codierhäufigkeit zwischen Ausbildern und Auszubildenden frappierend. Während in der Diskussion der Ausbilder lediglich zwei Passagen mit dieser Kategorie codiert wurden, stellt sie die am häufigsten genutzte Kategorie in der Diskussion der Auszubildenden dar, in der 20(!) Passagen in diese eingeordnet werden konnten. Somit deuten die Ausführungen *aller Auszubildenden* darauf hin, dass das prägende Interaktions- und Kommunikationsmuster von ihnen nicht dem oben umrissenen Ideal der ‚dialogischen und reflexiven Kommunikation und Interaktion‘ entspricht, sondern es darum geht, „schnellstmöglich herauszufinden, was der Fachleiter von einem möchte“, wie es R3 ausführt. „Auch wenn dies eigentlich nicht Sinn der Sache sein kann“, sieht R4

diese „Suche nach dem Treffer bei dem Fachseminarleiter“ ebenfalls als, de facto, die primäre Aufgabe der Auszubildenden:

R4: „[...] Weil diese Suche nach diesem Treffer bei dem Fachseminarleiter, also das ist was, was ich nicht nur selber auch gehört habe (R5: nicken), sondern auch was von anderen Kollegen kam und wo viele den Eindruck hatten, dass es tatsächlich darum geht, eher den Geschmack zu treffen eines Fachleiters. [...]“

Diese Kommunikationsstrategie wird an die nachfolgenden Referendarsjahrgänge quasi vererbt, wie R1 es ausführt:

R1: „Das ist ja auch das was man jetzt macht, wenn die neuen Referendare kommen (R5: nicken, «mhm» (bejahend)). Die sagen ich habe den und den Fachleiter und dann sagt man: «Also der mag dies, das, jenes» (R2, R3, R4, R5: allgemeine Zustimmung und nicken).“

Die Nutzung dieser Interaktionsstrategie hat weitreichende Folgen für die Logik, mit der sich die Auszubildenden und Ausbilder in den Nachbesprechungen begegnen. Es deutet sich an, dass sie sogar verhindert, dass die Ausbilder ihrer Rolle als Berater nachkommen können, da die Auszubildenden sich nicht trauen, die ausbleibende Befriedigung ihrer Beratungsbedürfnisse zu offenbaren und stattdessen ‚abnicken‘, was die Ausbilder ihnen anbieten, auch wenn sie dies für „Quatsch“ halten:

R4: „[...] W]eil ich glaube manchmal hört man sich auch Sachen an vom Fachleiter, nickt die auch vielleicht ab, aber denkt sich die ganze Zeit, wie gesagt, das ist mir auch passiert, wo ich gedacht habe: «Nein, das ist ja jetzt eigentlich Quatsch!», also zumindest für mein Empfinden. Das sagt man dann aber nicht und diskutiert darüber, weil eben dieses Gefälle da ist und mhm, ja. Das ist glaube ich auch was, was ja, was ich ändern will. Was gut wäre, wenn es sich ändert.“

Die Dominanz dieses Interaktionsmusters der Auszubildenden kann die Ursache dafür sein, dass diese deutlich seltener explizit die Problematik einer ‚nicht

fokussierten und überfrachteten Perspektivierung‘ thematisieren, als es die Ausbilder tun (zwei gegenüber zwölf codierten Passagen). Wie oben ausgeführt, wünschen sich die Referendare einen ‚idealen Ausbilder‘, der ihnen Orientierung für ihre Professionalisierung gibt und nicht „23 Aspekte auf einmal“ thematisiert, in diesem Sinne also eine ‚fokussierte, situationsangepasste Perspektivierung‘ anbietet. Da die Auszubildenden sich aber primär darauf konzentrieren „den Geschmack“ ihres Ausbilders zu treffen, ist es für sie vordergründig ein größeres Problem, wenn sie die Erwartungen des Ausbilders nicht durchschauen können, als wenn diese Erwartungen wenig hilfreich für die individuelle Professionalisierung und nicht situativ angepasst sind, sie aber klar zu erkennen sind und man so „den Geschmack“ des Ausbilders treffen kann. Nichtsdestotrotz finden sich in einigen Aussagen der Auszubildenden auch explizite Hinweise, dass eine fehlende Fokussierung zu Problemen bei der eigenen Professionalisierung führt, wie z.B. bei R5, die zu dem Sinn von verbindlichen, schriftlichen Bilanzierungen der Unterrichtsnachbesprechungen folgendes ausführt:

R5: „Die Ursprungsfrage war ja, vor langer, langer Zeit, ob uns dieser Bogen beim Nachgespräch in der Struktur helfen würde? Und da muss ich nämlich sagen: Auf jeden Fall. Egal, welcher jetzt und in welcher Länge oder Kürze, weil das ist eben das, was ich auch häufiger erlebt habe, dass man zwar ein unheimlich angeregtes Nachgespräch hatte, aber das Ganze so hin und her gesprungen ist (R4: «mhm» (bejahend)), dass da die Struktur fehlte und man deswegen am Ende so da saß und dachte: «Ja, ok. Worum ging es jetzt eigentlich? Was waren die wesentlichen Punkte?»“

Trotz einer „angeregten Gesprächsatmosphäre“, also Interaktionslogiken, die eher in Richtung der ‚Lernsituation‘ tendierten, fehlte R5 „häufiger“ ein roter Faden, eine Fokussierung der Nachbesprechung auf die aus den konkreten Situationen geborenen Professionalisierungsansätze. R4 bekräftigt diese Erfahrung.

In enger Verbindung mit den problematischen Interaktions- und kontraproduktiven Beziehungsmustern zwischen den Ausbildern und den Auszubildenden stehen die ‚unproduktiven Bewertungsprozesse‘, zumindest erleben alle Auszubildenden diese ganz oder teilweise als solche. Die Bewertungsprozesse sind aus der

Perspektive der Auszubildenden hochgradig von ‚*Intransparenz und Subjektivität*‘ geprägt, dies scheint für *mehr als die Hälfte der Auszubildenden* eine der bedeutendsten Erfahrungen im Kontext der Bewertungsprozesse gewesen zu sein, wie der Beginn der Gruppendiskussion zeigt:

I: „Wenn ihr an eure Erfahrungen mit Bewertungen im Rahmen eurer eigenen Lehrerausbildung zurückdenkt, was fällt euch da spontan ein? Bewertungen von Lehrproben!

R4: Wenig Transparenz, glaube ich, bei mir sehr häufig festgestellt, also nicht nur zum Negativen, auch selbst, wenn Lehrproben gut gelungen waren, auch gut bewertet wurden, selbst dann war es teilweise so, nicht in beiden Fächern, aber in einem Fach v.a. auch vorwiegend, dass nicht ganz klar war, warum sie so gut bewertet wurde. Genauso kam es vor, dass Lehrproben auch sich angehört haben in der Nachbesprechung als wären sie nicht gut, aber die Note war gleich (R3, R5: nicken) im Vergleich zu einer vorherigen Lehrprobe, die zum Beispiel besser war von der Note her. Erst, wirklich nur erster Eindruck.

R5: Ja, kann ich zustimmen.

R1: Ich auch.

R5: Außerdem war es auch noch so, dass manchmal, also ich das Gefühl hatte, dass mir gesagt wurde, ich soll irgendetwas in eine Richtung ändern, dann habe ich das meiner Ansicht nach getan und dann war es hinterher, war es wieder ganz anders. Also es war manchmal von der Kommunikation scheinbar schon etwas schwierig zu verstehen, was jetzt eigentlich von mir genau verlangt wird und ich hatte manchmal wirklich das Gefühl: Einmal rechts, einmal links und beides ist verkehrt.“

Intransparenz ist demnach nicht nur ein Problem schlecht bewerteter Lehrproben, sondern auch guter Bewertungen und stellt offensichtlich ein deutliches Hemmnis für klare Bilanzierungen über Kompetenzfortschritte auf Seiten der Auszubildenden dar. Die Problematik unterschiedlicher Bewertungsgrundlagen

unter den Ausbildern, wie sie auch die Ausbilder in ihrer Diskussion als Problem thematisiert haben, kennen die Auszubildenden aus eigener Erfahrung. Scheinbar spielen hier auch die Beziehungen *unter den Ausbildern* eine Rolle für die Intransparenz. R4 berichtet in diesem Zusammenhang nicht nur über die Unstimmigkeiten zwischen verschiedenen Ausbildern, sondern über unterschiedliche Bewertungsmaßstäbe ein- und desselben Ausbilders, je nachdem ob er alleine oder gemeinsam mit dem Hauptseminarausbilder bewertete, „auf einmal zählten andere Kriterien“:

R4: „Auch sehr interessant fand ich den Unterschied teilweise Fachleiter – Hauptseminarleiter oder Hauptseminarleiterin. Also, dass da auch scheinbar unterschiedliche Vorstellungen waren, die dann natürlich, die kamen in diesem Gespräch nicht unbedingt zum Ausdruck, aber wenn man gehört hat, was der Fachleiter vorher erzählt hat teilweise und wie er dann in dem Gespräch war, wenn der Hauptseminarleiter dabei war, das war was Anderes. Und dann auf einmal zählten auch andere Kriterien oder zumindest eine andere Gewichtung, und da hatte man auch das Gefühl, dass da es auf keinen Fall immer stimmig ist zwischen Fachseminar und Hauptseminar, dass dadurch also auch andere Kriterien gelten, aber zumindest andere Kriterien im Mittelpunkt stehen auf einmal. Und das, also das war dann durchaus interessant im negativen Sinne, weil man davon sich natürlich auch, man ist verwirrt natürlich. Man weiß auch in dem Gespräch nicht genau, in welche Richtung es dann geht und dementsprechend also, das war, was mir aufgefallen ist.“

Die Referendare sind die Leidtragenden von Intransparenz und fehlender Stringenz bei den Bewertungskategorien. Nicht nur, dass die Bewertungskategorien oftmals intransparent erscheinen, die Ausbilder verändern ihre eigenen Bewertungsmaßstäbe von Situation zu Situation und die Auszubildenden sind diejenigen, die dieser Unberechenbarkeit ausgeliefert sind, wobei erwähnt werden muss, dass R2 anmerkt, nie mit diesen Problemen konfrontiert worden zu sein. Gleichwohl zeigt sich hier eine strukturelle Eigenschaft der Bewertungsprozesse, die Hinweise darauf liefert, wie es zu den oben umschriebenen starken emotionalen Reaktionen, wie von R3 kommen kann, die auch R4 kennt, wie er weiter ausführt:

R4: „[...] Ich glaube, dass ist sehr demotivierend (R3: Ja) und sorgt dafür, ja dass die Kollegen und Kolleginnen dann auch wirklich (R3: weinten), ja weinten/Vollkommen, auch einfach fürs nächsten Mal verzweifelter waren, also, dass es immer schlimmer wurde, sozusagen, auch wenn man sich nicht immer unsicherer wurde, also jetzt habe ich das versucht, ok, hat nicht so ganz geklappt, aber da auch wieder diese fehlende Transparenz, dass man eben nicht genau weiß: Aha, ok, hier habe ich Mist gebaut, das muss ich anders machen oder das sollte ich besser machen, sondern auch da wieder dieses Gefühl, des in der Schwebe sein, gar nicht genau wissen, was so schlimm war oder was so schlecht war an einer Stunde.“

Obwohl sich in der Diskussion der Ausbilder keine Klarheit über die konkrete Ausgestaltung der Progression in den Bewertungsprozessen herstellen ließ, sahen diese darin scheinbar keine große Problematik. Aus der Diskussion der Auszubildenden lässt sich nun allerdings ableiten, dass neben der Unklarheit der Bewertungskriterien und der daraus resultierenden Intransparenz die Unklarheit darüber, wie sich die ‚Progression der Ausbildung‘ in der Bewertung niederschlägt, eine entscheidende Rolle bei den als unproduktiv erlebten Bewertungsprozessen spielt, wie die Einschätzung von R3 und R5 zeigen, die von *allen Auszubildenden* geteilt wird:

R3: „[...] Ich habe es nicht irgendwie nachvollziehen können, an welcher Stelle irgendwie was sich enger zieht oder schlimmer bewertet wird (R4: «mhm» (bejahend), R5: nicken) und die Noten fand ich jetzt auch irgendwie, konnte man jetzt irgendwie interpretieren: War das jetzt wirklich schon UPP- Standard? Haben die das jetzt schon angewendet und ich kann mir sicher sein, mit denen Kriterien würde ich auch eine UPP, mit derselben Lehrprobe, eine ähnliche Bewertung bekommen oder haben die jetzt noch gesagt: «Nein Frau R3 ist ja noch, hat ja noch vier Wochen oder so, da können wir noch ein bisschen warten.» Also auch da fand ich, konnte man nicht nachvollziehen, was da einfach mal am Anfang gesagt wurde, dass unterschiedliche Kriterien angesetzt werden würden, aber welche Art von Unterschieden es da gab, wusste man auch nicht.

R5: Oder auch welche wann dazukommen (R3: Ja., R4: «mhm» (bejahend)). Am Anfang hieß es wirklich: „Ja wir brauchen eine

Phasierung. „Aber irgendwann hat es wirklich aufgehört, dass uns gesagt wurde, was wir eigentlich können müssen (R2, R4: «mhm» (bejahend)).“

Im Zuge der Ausbildung wurde zwar offenbar kommuniziert, dass die Ausbildungsprogression eine Rolle in den Bewertungsprozessen spielt, aber für die Auszubildenden bleibt es eine Art von Kaffeesatzleserei, worin sich diese Progression denn nun ausdrückt, wie auch R2 bekräftigt:

R2: „[...]Was ich aber auch sehr erlebt habe ist (unv.), das muss jetzt UPP Standard sein und (...) hat uns aber auch nie jemand gesagt, welche Sachen da jetzt besonders in den Vordergrund rücken sollen (R4: «mhm» (bejahend)) oder was besonders wichtig ist.“

Eine Möglichkeit für produktivere und transparentere Bewertungsprozesse könnte der Einsatz von einheitlichen Bewertungsinstrumenten oder Dokumentationsbögen sein, wie sie auch das vorliegende Projekt anstrebt. Berücksichtigt man die Erfahrungen der Auszubildenden, muss dieser Hoffnung allerdings ein deutlicher Dämpfer erteilt werden. *Alle Auszubildenden* berichten von einer ausbleibenden oder fragwürdigen Nutzung der existierenden Bewertungsinstrumente des ZfsL Bonn, insofern decken sich ihre diesbezüglichen Erfahrungen mit den Einschätzungen der Ausbilder, wie sie weiter oben dargelegt wurden. So lagen mit dem bereits erwähnten Qualitätstableau und den daraus abgeleiteten ‚Bubbles‘ zwar Instrumente vor, aber eingesetzt wurden diese Instrumente scheinbar nicht adäquat, wie es die folgende Diskussionspassage dokumentiert:

R1: „Erinnern kann ich mich, aber so genau durchgesprochen haben wir das, glaub ich nicht. Also vielleicht ganz am Anfang?“

R5: In der Nachbesprechung kam das bei mir auch so, dass ich es benutzt habe, auch nicht zum Tragen.

R4: Das ist auch das Interessante.

R5: D.h., es hat sich nie jemand wirklich drauf bezogen (R4: Ja) und mir

dann mal gesagt: «Jetzt hier der Punkt», sondern es kam dann am Rande so ein bisschen mit rein (R4: Ja), aber das mal wirklich jemand gesagt hat: «Dieses Kriterium habe ich gut erfüllt oder nicht», der hat immer nur/[...]“

Irgendwann wurden die Instrumente offensichtlich thematisiert, aber „genau durchgesprochen“ oder „zum Tragen“ kamen sie nicht, höchstens mal „am Rande“.

Auch ein Instrument zur Dokumentation der Nachbesprechung¹⁰⁶ wurde nicht entsprechend seiner Intention genutzt, sondern, sofern es überhaupt genutzt wurde, eher pro forma ausgefüllt. Einen Sinn konnten die Auszubildenden bei der Nutzung dieses Instruments nicht entdecken, wie R3, R4 und R5 berichten:

R5: „Bei mir wurden meistens nur zwei Felder ausgefüllt (R4: «mhmm» (bejahend)), überhaupt.

R4: Nur die Daten eigentlich und!

R5 (R3, R4, R5 durcheinander): Ja, also der Bogen wurde zu fünfzig Prozent schon mal gar nicht genutzt (R4: «mhmm» (bejahend)). Ich weiß dann mal, da waren unten diese Vereinbarungsfelder mit: «Was soll die Schule tun?»

R3 (R3, R4, R5 durcheinander): Mit der Schule und mit dem Seminar.

R4 (R3, R4, R5 durcheinander): Links war das, was man gesprochen hat und rechts waren die Vereinbarungen, also untere Spalte.

R3: Und eigentlich dann links, wie die Schule noch helfen kann und rechts war, was man im Seminar nochmal aufgreifen sollte, die unteren hat man nie ausgefüllt

R5: Genau. Und der Rest war auch immer so, oh jetzt, nachdem die ganze Besprechung gelaufen ist, jetzt überlegen wir mal noch schnell, was

¹⁰⁶ Dieser Dokumentationsbogen ist im Anhang dokumentiert.

wir da reinschreiben (R4: Ja) und dann wurde irgendwas an den Haaren herbeigezogen, also das war immer sehr, die Dokumentation war da sehr merkwürdig.“

R1 kennt diese Instrumente noch nicht aus ihrer eigenen Ausbildung, die zwölf Monate länger zurückliegt als jene von R3, R4 und R5, aber auch R2, die ihr Examen gerade erst abgelegt hat, berichtet über einen ausbleibenden Einsatz:

R2: „Aber diese Bögen waren eigentlich genau dazu gedacht, dass theoretisch dann bei allen halt dann auch mit auf dem Tisch liegen (R4: Ja, oder?, R5: Ja), aber ich hab es auch tatsächlich überhaupt nie erlebt.“

Offensichtlich scheint bei den Auszubildenden auch keine vollkommene Klarheit darüber zu herrschen, welche Rolle diese Dokumentationsbögen denn nun spielen sollten bzw. wie verbindlich sie genutzt wurden, obwohl aus der Diskussion hervorgeht, dass diese Instrumente seit mindestens 12 Monaten Grundlage der Bewertungsprozesse sein sollten.

Rückblickend wird deutlich, dass die Interaktions- und Kommunikationsmuster, mit denen die Auszubildenden in die Unterrichtsbesuche und -nachbesprechungen eintreten, entscheidenden Anteil an dem Zustandekommen der ‚*Stresssituation*‘ haben. Durch den offensichtlich durch die hierarchischen Beziehungen befeuerten Bruch mit dem Ideal eines selbstbewusst auftretenden Auszubildenden positionieren sich die Auszubildenden nicht mit ihren eigenen Überzeugungen oder offenbaren sich nicht in Beratungssituationen. Dies geschieht auch dann, wenn sie die dort entwickelte Handlungsoptionen für „Quatsch“ halten, sodass die Ausbilder eben nicht „nah“ an die Auszubildenden und den „Motor der Entwicklung“ herankommen. Vielmehr lässt sich ein Bild zeichnen, dem zufolge die Ausbilder in dem Interaktionssystem der ‚*Stresssituation*‘ nur vordergründig mit den Auszubildenden interagieren. Diese wiederum spielen eine Rolle, indem sie versuchen den Ausbildern gerecht zu werden und fatalerweise durch genau dieses ‚Verstellen‘ nicht das tun, was der eigentlich Wunsch der Ausbilder ist. Es ist wenig verwunderlich, dass Interaktions- und Kommunikationssysteme mit solch verdeckten Motiven und Interaktionslogiken, v.a. auf Seite der Auszubildenden,

zu Stress führen, dessen Ausmaß aber offensichtlich individuell sehr unterschiedlich ist.

Die wesentlichen Erkenntnisse hinsichtlich der Logik der Ausbildungssituation, wie sie sich aus den Gruppendiskussionen rekonstruieren lässt, werden im Folgenden zusammenfassend dargestellt.

4.2.3 Zusammenfassende Interpretation ‚Logik der Ausbildungssituation/Rahmenbedingungen von Unterrichtsbesuchen‘

Auszubildende und Ausbilder teilen Vorstellungen darüber, wie die ‚idealen Rahmenbedingungen‘ der Ausbildung aussehen sollen! Selbstbewusste und zur Selbstreflexion fähige und willige Auszubildende treffen auf kompetente Ausbilder, die dazu in der Lage sind, situativ zu entscheiden, ob ihre Bewertungs- zu Gunsten ihrer Beraterfunktion in den Hintergrund treten muss und die dialogisch mit ihren Auszubildenden aus den Ausbildungssituationen heraus fokussieren können, welche Kompetenzaspekte für die weitere Professionalisierung entscheidend sind. Mehrere der Ausbilder und Auszubildenden haben erfahren, dass diese Vorstellungen, zumindest zeitweise, im Sinne einer ‚Lernsituation‘ realisiert werden können. Aber v.a. auf Seite der Auszubildenden zeigten sich deutliche Diskrepanzen zwischen den überwiegend realisierten Ausbildungsbedingungen und diesen Idealvorstellungen. Die Folgen, die diese Diskrepanzen zwischen der Ausbildungsrealität und dem -ideal haben, äußern sich in *Ausbildungssituationen, die Stress als prägendes Gefühl* hervorrufen und die Professionalisierung nur unzureichend fördern. Hierbei scheint das Ausmaß des Stresses zwischen den Akteursgruppen hochgradig unterschiedlich zu sein: Während die Ausbilder Stresssituationen ‚kennen‘, ‚fühlen‘ die Auszubildenden diese, wobei die individuellen Unterschiede der Stresswahrnehmung – darauf deuten die Ausführungen der Auszubildenden hin – sehr groß sind. Die *dominante Strategie der Auszubildenden*, „den Geschmack des Ausbilders zu treffen“, führt zu einem zentralen Konflikt mit den idealen Rollenerwartungen, die an die Adresse der Referendare von beiden Gruppen konsensuell formuliert wurde. Die Auszubildenden verhalten sich also wider besseres Wissen konträr zu ihrem eigenen Rollenideal.

Der Grund für diese paradox anmutende Interaktionslogik kann in den hierarchischen Beziehungen zwischen den Akteursgruppen ausgemacht werden, welche den Handlungsspielraum der Auszubildenden als prägende Strukturbedingung offenbar de facto deutlich einengen. Aus dieser Strategie folgen *Probleme für die gegenseitigen Rollenerwartungen*. So wird es für die Ausbilder schwer bis unmöglich eine Rolle als beratender Experte einzunehmen, wie die ideale Rollenzuschreibung an die Adresse der Ausbilder durch die Auszubildenden aussieht, wenn die Auszubildenden davor zurückschrecken, sich mit eigenen professionellen beliefs zu positionieren und auch – aus ihrer Sicht so zu betitelnden – „Quatsch“ schlucken, den die Ausbilder ihnen anbieten. Beratung kann nicht gelingen, wenn die Auszubildenden vor einer Selbstoffenbarung zurückschrecken, denn erfolgreiche „Ausbildung kann nur im Dialog gelingen“, wie es einer der Ausbilder in der Gruppendiskussion formulierte.

Inwiefern *Bewertungs- oder Beratungsinstrumente* dabei helfen können, die idealen Vorstellungen der Ausbildungssituationen häufiger zu realisieren, muss zu diesem Zeitpunkt als fragwürdig angesehen werden, denn wenn diese zwar existieren, aber nicht eingesetzt werden, ist wenig gewonnen. Weiterhin scheinen die Bewertungsinstrumente keine große Hilfe dabei zu sein, die Macht der handlungsbestimmenden Beziehungsprobleme zu entschärfen; zumindest deuten die Einschätzungen der Auszubildenden in diese Richtung.

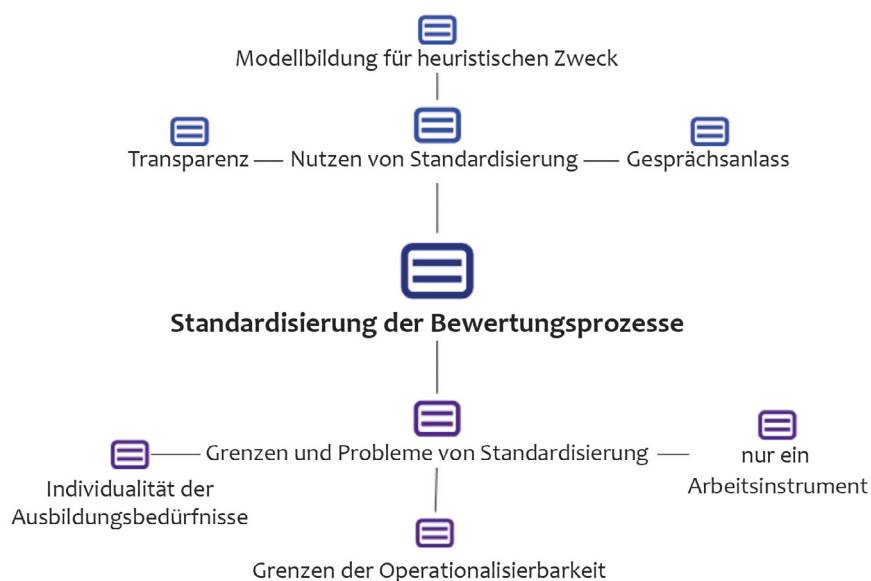
Abschließend muss gesagt werden, dass sich *gruppenspezifisch besonders fokussierte Probleme* finden, die von der jeweils anderen Gruppe nicht oder nur ansatzweise wahrgenommen werden. So betonen die Ausbilder beispielsweise die besondere Rolle der Ausbildungsprogression für die Bewertungsprozesse. Sie scheinen aber nur ein eingeschränktes Bewusstsein für die Problematik mangelnder Transparenz zu haben, wie sich diese Progression in den Bewertungen niederschlägt, während diese Intransparenz für die Auszubildenden ein zentrales Problem ist. Auf der anderen Seite erwarten die Auszubildenden von den Ausbildern, dass diese eine den konkreten Ausbildungsbedingungen und -bedürfnissen angepasste Fokussierung der Unterrichtsnachbesprechung leisten können. Dabei

fragen sie nicht, in welchen Grenzen dies überhaupt möglich ist und welchen aktiven Anteil sie selbst an dieser Perspektivierung haben.

4.3 Standardisierung der Bewertungsprozesse

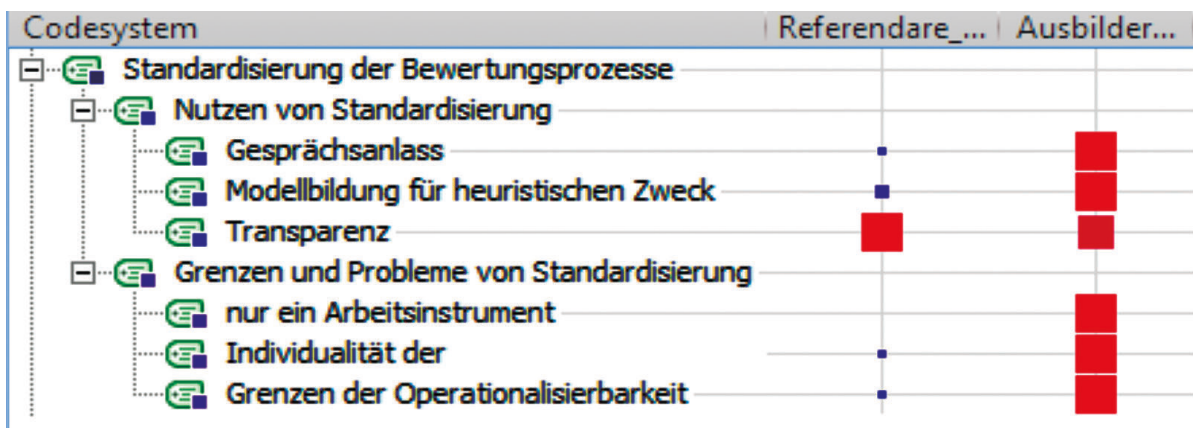
Die Abbildung 4.4 zeigt die Kategorie mit ihren Subkategorien, die hinsichtlich der grundlegenden qualitativen Einschätzung unterschieden werden, ob die Standardisierung von Bewertungsprozessen einen ‚Nutzen‘ für die Bewertungssituationen in der Ausbildung generiert oder ob die Standardisierung eher durch ‚Grenzen und Probleme‘ gekennzeichnet ist. Wenn im Folgenden von Standardisierung der Bewertungsprozesse gesprochen wird, dann ist eine Standardisierung gemeint, wie sie das hier entwickelte Bewertungskonzept vertritt. Diese zeichnet sich erstens durch eine *Standardisierung der Bewertungskategorien*, im Sinne der oben in Kapitel zwei geleisteten Operationalisierung eines guten Unterrichts und zweitens durch eine *Standardisierung des Bewertungsvorgangs* aus, die durch einheitlichen Bewertungsinstrumente, die die Vergleichbarkeit der Beobachtungen, sowie eine möglichst einheitliche Struktur der Nachbesprechung ermöglichen sollen.

Abbildung 4.4 : Subkategorien der Kategorie ‚Standardisierung der Bewertungsprozesse‘



Wie in der vorangegangenen Auswertung werden die Sichtweisen der Auszubildenden und Ausbilder mit Hilfe Code-Matrix-Browser von MaxQDA zunächst in ihrem rein quantitativen Auftreten deskriptiv gegenübergestellt, um dann getrennt voneinander interpretativ ausgewertet zu werden. Ein abschließendes Fazit führt beide Perspektiven wiederum hinsichtlich ihrer Gemeinsamkeiten und Unterschiede zusammen. Die Abbildung 4.5 stellt die Codierhäufigkeiten der einzelnen Subkategorien in beiden Gruppendiskussionen gegenüber, sodass wiederum erste Erkenntnisse über die unterschiedliche Rezeption der diskutierten Fragen in den Gruppen abgeleitet werden können. Die unterschiedliche Häufigkeit, mit der diese Kategorie in den beiden Gruppendiskussionen codiert worden ist, zeigt einen deutlichen Ausschlag in Richtung der Ausbilder. Während in der Gruppendiskussion der Auszubildenden 19 Diskussionspassagen mit den Subkategorien der Kategorie ‚Standardisierung der Bewertungsprozesse‘ codiert wurden, konnten in der Diskussion der Ausbilder mehr als drei mal so viele Passagen dieser Kategorie zugeordnet werden, nämlich 59. Diese höhere Zahl an Codings auf Seiten der Ausbilder lässt sich nicht durch die um ca. 25% längere Diskussionsdauer erklären, sodass ein im Vergleich zu den Auszubildenden schärferer Fokus für die Frage des Nutzens von Standardisierung abgeleitet werden kann. Mögliche Ausschläge der einzelnen Subkategorien in der Code-Matrix zu Gunsten der Ausbilder können also zu einem Teil durch die mehr als doppelt so häufig erfolgte Thematisierung erklärt werden, gleichzeitig sind vor diesem Hintergrund Ausschläge zu Gunsten der Auszubildenden als besonders auffällig zu bewerten.

Abbildung 4.5: Code-Matrix der Kategorie ‚Standardisierung der Bewertungsprozesse‘



Entsprechend dem unterschiedlichen Fokus der beiden Diskussionen sind die Schwerpunkte der zeilenweisen Verteilung der Codings auf Seiten der Ausbilder nicht verwunderlich. In fünf von sechs Subkategorien thematisieren die Ausbilder die Aspekte der Standardisierung deutlich häufiger als die Auszubildenden, einzig die Frage, inwiefern eine Standardisierung der Bewertungsprozesse zu mehr ‚*Transparenz*‘ von diesen beitragen kann, wird von den Auszubildenden etwas häufiger thematisiert (Auszubildende 8 Codings vs. Ausbilder 7 Codings). Dies deutet die wesentliche Funktion an, welche standardisierte Bewertungsprozesse für die Auszubildenden haben, nämlich eine Erhöhung der Transparenz, während die anderen möglichen positiven und negativen Aspekte der Standardisierung in der Diskussion der Auszubildenden keine große Rolle spielen (zwischen 0 und 4 Codings). Die Ausbilder hingegen diskutieren ausführlich einen möglichen Nutzen der Standardisierung, wobei hier die Frage, welchen Beitrag die Standardisierung für eine ‚*Modellbildung*‘ des Unterrichts in einem heuristischen Sinne, der der Professionalisierung dient, mit 14 Codings am häufigsten thematisiert wird. Noch breiter werden auf der negativen Seite der Kategorie die ‚*Grenzen der Operationalisierbarkeit*‘ diskutiert (21 Codings auf Ausbilderseite), sowie weniger häufig, aber immer noch relativ ausführlich, die Konflikte zwischen standardisierten Bewertungsprozessen und der ‚*Individualität der Ausbildungsbedürfnisse*‘ der Auszubildenden. Als Fazit aus den reinen Codierhäufigkeiten kann also festgehalten werden, dass die Ausbilder ein wesentlich differenzierteres Bild der ‚*Standardisierung von Bewertungsprozessen*‘ zeichnen als die Auszubildenden, sodass, wie schon in der zuvor thematisierten Kategorie der ‚*Logik der Ausbildungssituation*‘, offensichtlich gruppenspezifische Sensibilisierungen vorherrschend sind, die im Folgenden genauer rekonstruiert werden sollen.

4.3.1 Die Sichtweise der Ausbilder

Alle Ausbilder, bis auf einen, teilen die Auffassung, dass Standardisierung das Potential hat, Bewertungsprozesse in der zweiten Phase der Lehrerbildung positiv zu beeinflussen. Hierbei berichten die Ausbilder einerseits von ihren Erfahrungen mit Standardisierungsansätzen, andererseits schätzen sie ein, inwiefern sie die Standardisierungsabsichten des Bewertungskonzepts als sinnvoll erachten. Eine

zentrale Rolle bei diesen positiven Potentialen der Standardisierung spielt die ‚Modellbildung für einen heuristischen Zweck‘, heuristisch in dem Sinne, dass alle Maßnahmen, welche die Auszubildenden bei der zunehmenden Professionalisierung unterstützen, als sinnvoll angesehen werden. In diesem Kontext bieten standardisierte Bewertungsinstrumente aus der Perspektive der Ausbilder den Auszubildenden Modelle guten Unterrichts, an denen sie sich orientieren können, um nach und nach zunehmenden Durchblick in der zu Beginn der Ausbildung nahezu unüberschaubaren Komplexität des Unterrichts zu erhalten. A2 beschreibt diesen Beitrag der Standardisierung für eine gewinnbringende Ausbildung mit der Metapher eines Schachspiels:

A2: „Ich vergleiche einen Referendar im Anfangsstadium gerne mit einem Schachspieler. Der Schachspieler/Das Schachspielen hat ziemlich komplizierte Regeln, auch sehr komplex, die Regelstruktur. Der Referendar weiß, wie ein Pferdchen springen muss, wie man einen Bauern vorzieht, der rückt auch die Figuren, aber er überschaut nicht das ganze System, keine Taktik und keine Strategie im Grunde genommen (A1: «mhm» (bejahend)). [...] Im Schach lernen die erst mal Taktik und das sind unsere Kriterien hier (zeigt auf die Vorlage) (A4: nicken), sozusagen taktische Elemente (A3: «mhm» (bejahend)), das Regelwerk erweitert über taktische Aspekte, und so wird die Komplexität zunehmend in der Ausbildung größer. Und da hilft die Standardisierung, indem man eben bestimmte taktische Bereiche, die am Anfang erst mal leicht sind, mit hineinnimmt und die Komplexität schrittweise erhöht, in dem man z.B. die Heterogenität auch noch mit hinzunimmt oder vielleicht an erster Stelle, hängt auch von der Klasse ab, der Lerngruppenbezug ist ganz wichtig (A6: «mhm» (bejahend)).“

Die Standardisierung der Bewertungskriterien über die Operationalisierung guten Unterrichts liefert laut A2 also eine Hilfestellung, „taktische Elemente“ des Unterrichtens einzuüben und die Komplexität schrittweise zu erhöhen, damit die Auszubildenden nach und nach erlernen „das System zu überschauen“ und eben nicht nur einzelne Regeln kennen. A1 pflichtet den Ausführungen von A2 bei und sieht ebenfalls den Nutzen standardisierter Bewertungskategorien als „hilfreiche Stütze“ oder „Krücke“ für die Auszubildenden:

A1: „[...] Grundsätzlich ist natürlich, finde ich, diese Standardisierung, die Sie auch versuchen, erst mal, egal jetzt zu welchem Modell, eine hilfreiche Stütze oder eine Krücke, die einem erst mal einen Zugang eröffnen kann, wie möchte ich meinen Unterricht planen, worauf schaut eigentlich auch jemand, der Unterricht beurteilt. Das einfach mal transparent zu haben, das hilft erst mal.“

Neben der Herstellung von Transparenz, um die es weiter unten gehen wird, ermöglichen schriftlich fixierte Bewertungskriterien „erst mal einen Zugang“ zu der Frage, wie Unterricht geplant oder nach welchen Gesichtspunkten der Unterricht beobachtet wird. Die Standardisierung ist, in den Worten von A4 „[...] also wirklich eine Hilfestellung, um eine komplexe Wirklichkeit wahrzunehmen“. Standardisierte Bewertungskriterien sind im Sinne von A3 als Teil eines Klärungsprozesses zu verstehen, der für eine erfolgreiche Ausbildung und auch für jede Ausbildungsinstitution zwingend notwendig ist, damit sie erfolgreich sein kann. Wie weiter oben ausgeführt wurde, erfolgte der Einsatz der existierenden Bewertungsinstrumente am ZfsL Bonn bisher nicht stringent, aber die Entwicklung eines gemeinsamen Qualitätsverständnisses im Sinne einer Standardisierung von Bewertungskategorien dient der „Verständigung über theoretische Grundlagen“ der Bewertung, und „das ist ganz wichtig in einem Seminar“:

A3: „[...] Aber das heißt nicht das (Anm.: Gemeint ist die Entwicklung des ‚Qualitätstableaus‘ der Kriterien guten Unterrichts) war schlecht, sondern das ist, und so würde ich das auch einschätzen, dass diese Matrix, in der Form, in welcher Form auch immer, dass sie im Sinne einer Verständigung über theoretische Grundlagen und das Verständnis dieser theoretischen Grundlagen ganz wichtig ist (A6: nicken) in einem Seminar.“

Standardisierte Bewertungskategorien können also nicht nur den Auszubildenden helfen, einen ersten Zugang zur und zunehmende Orientierung in der Planung und Durchführung von Unterricht zu erhalten, sondern auch für die Ausbilder sorgen diese für Klärungsprozesse, die Grundlage produktiver Bewertungsprozesse sind, wie auch A2 zusammenfasst, für den es „keine Alternative“ zur Standardisierung von Bewertungskategorien gibt:

A2: „Also ich glaube auch, es gibt keine Alternative dazu. Denn Unterricht ist so komplex gestaltet!“

I: Keine Alternative zur Standardisierung?

A2: Ja zur Standardisierung, genau. Weil die Komplexität gar, anders gar nicht erfasst werden kann, dass ist ja der Begriff der Komplexität. Es muss fokussiert werden auf einzelne Kriterien. [...] Und die Standardisierung hilft uns selber auch zu reduzieren, die Komplexität auf wenige Begrifflichkeiten zu reduzieren, die für diese Stunde eine besondere Bedeutung hatten []. Und dazu brauchen wir die Reduktion, weil die Komplexität sonst uns selbst und auch die Referendare erschlägt.“

Eng mit dem positiven Nutzen im Sinne einer ‚heuristischen Modellbildung‘ verknüpft ist die Sicherstellung der ‚Transparenz‘ von Bewertungsprozessen, die durch standardisierte Bewertungskategorien ermöglicht wird, wie es z.B. A1 in dem oben zitierten Beitrag formuliert. An Kriterien orientierte Modellierungen guten Unterrichts sorgen dafür, dass die Auszubildenden „das“ (die Frage nach dem Planungs- und Auswertungsfokus von Unterricht) „einfach mal transparent haben“. Neben der Klärung von Bewertungskategorien sehen *mehrere Ausbilder* die Vorteile von standardisierten Bewertungsvorgängen, wenn es um die Frage transparenter Bewertungsprozesse geht. So sieht A3 den Vorteil standardisierter Beobachtungsbögen für Unterricht, auch wenn er diesen „grundsätzlich skeptisch gegenübersteht“:

A3: „[...] Also insofern stehe ich grundsätzlich solchen Listen ganz skeptisch gegenüber, aber sie haben auch einen Vorteil und das erlebe ich auch, dass wenn ich das nicht mache, ich natürlich die Vergleichbarkeit der Ergebnisse nicht so ohne weiteres gewährleisten kann.“

Standardisierte Bewertungsinstrumente können demnach zur Vergleichbarkeit der Bewertungsprozesse in den verschiedenen Unterrichtsbesuchen beitragen und somit Transparenz über diese herstellen. In diesem Kontext führt A3 weiter aus:

A3: „Und es [Anm: standardisierte Beobachtungsbögen, D. L.] schafft auch, ja die Möglichkeit Differenzierungen im Erreichungsgrad von Kompetenzstufen zu formulieren (A1: «mhm» (bejahend)). Das hilft ja auch den eigenen Fortschritt, den eigenen Lernfortschritt der Referendare und Referendarinnen zu beschreiben und zu benennen.

A1: Deswegen meinte ich auch Stütze.

A3: Das ist ein großer Vorteil finde ich und das ist auch, finde ich, die wesentliche Leistung von Standardisierung, aber auch nicht mehr (A6: «mhm» (bejahend)).“

Die Gewährleistung von Vergleichbarkeit ist in den Augen von A3 und A1 eine „wesentliche Leistung“ von Bewertungsvorgängen mit Hilfe standardisierter Instrumente, wobei diese Vergleichbarkeit hilft, ‚*Transparenz*‘ über den Kompetenzfortschritt eines einzelnen Auszubildenden herzustellen. Auf Basis dieser Zitate ist anzunehmen, dass beide Ausbilder auch eine Steigerung der Vergleichbarkeit der Bewertungsprozesse verschiedener Auszubildender durch standardisierte Bewertungsinstrumente sehen würden, wobei dies zunächst als These stehen bleiben muss, da die Zitate diesen Schluss nur nahe legen, er sich aber nicht direkt aus diesen ablesen lässt. Damit Bewertungsprozesse als produktiv wahrgenommen werden, dies wurde schon in der Auswertung der ersten Kategorie der ‚*Logik der Ausbildungssituation*‘ deutlich, muss es zu einer ‚*dialogischen und reflexiven Kommunikation und Interaktion*‘ zwischen den Ausbildern und den Auszubildenden kommen. Insofern ist die Frage wichtig, inwiefern standardisierte Bewertungsprozesse das Entstehen solcher Kommunikations- und Interaktionsmuster unterstützen. Aus den Ausführungen von *mehr als der Hälfte der Ausbilder* wird deutlich, dass standardisierte Bewertungsinstrumente hierbei einen ‚*Gesprächsanlass*‘ liefern können, der den Ausgangspunkt für produktive Gespräche über die Bewertung des gesehenen Unterrichts bieten kann und somit einerseits zu transparenten Bewertungsprozessen führt und darüber hinaus den Auszubildenden hilft, die individuelle Professionalisierung voranzutreiben. Die Standardisierung wirkt in diesem Sinne nicht dadurch positiv

auf die Bewertungsprozesse, dass sie valide und reliable Bewertungsvorgänge ermöglicht – ein, wie es bereits deutlich wurde, ohnehin vermessenenes Ziel – sondern dass sich Ausbilder und Auszubildende im Sinne der kommunikativen Validierung über den Unterricht und dessen Bewertung austauschen, wie z.B. A4 ausführt:

A4: „Sie (Anm.: Gemeint ist die Standardisierung, D. L.) hilft, weil wir uns über einen Punkt verständigen können, wie wir ihn jeweils deuten. Also jetzt: Was heißt klares Strukturieren? Das ist so billig gesagt (A2, A3: «mhm» (bejahend)). Ja, klare Strukturierung, aber was heißt für mich klare Strukturierung, was heißt das für Sie, was heißt das für meine Kinder, nicht meine Kinder, entschuldigung für die Schülerinnen und Schüler?“

Das bewertende Sprechen über Unterricht „ist ein Prozess“, wie es A3 ausführt und im Sinne der Professionalisierung gilt es die Auszubildenden dazu zu befähigen, über den eigenen Unterricht im Sinne einer kriterialen Analyse sprechen zu können. Hierbei können standardisierte Bewertungskriterien eine „Sprachhilfe“ sein, wie es mehrere Ausbilder einander ergänzend thematisieren:

A1 (A1, A3, A4 durcheinander): „Den Referendarinnen und Referendaren hilft das (Anm.: Gemeint sind standardisierte Bewertungskriterien, D. L.) manchmal weiter. Denjenigen zumindest, die erst mal nicht so recht wissen, worauf sie gucken sollen und wo sie (anschle!) können. Ich finde, denen hilft das auch weiter, nee also das, kann Orientierung geben genau (A6: nicken). Oder eine Sprachhilfe, absolut.

A3 (A1, A3, A4 durcheinander): Ja, wo sie (anschla?) sollen. Ja genau, genau, kann Orientierung geben. Ja genau, ganz genau.

A4 (A1, A3, A4 durcheinander): Ja oder eine Sprachhilfe sein, also wir haben ja gesagt, dass sie lernen müssen, ja zu reflektieren, dass sie vielleicht intuitiv was fühlen (A1, A3: «mhm» (bejahend)), dass sie es auch vielleicht beschreiben können im Hinblick auf eine konkrete Stunde und hier bekommen sie Vokabeln geliefert, Begriffe geliefert, Zusammenhänge, Konstrukte, mit denen sie dann argumentieren können.“

Bewertungsinstrumente, wie auch immer sie konkret gestaltet sein mögen, die im Sinne einer Standardisierung der Bewertungskriterien eine schriftliche Fixierung von Kriterien guten Unterrichts anbieten, können den Auszubildenden helfen, ihre je individuelle Professionalisierung voranzutreiben, indem diese ihre situativ gemachten Erfahrungen in den größeren Kontext eines Qualitätsverständnisses guten Unterrichts stellen können und so „Begriffe, Vokabeln und Zusammenhänge“ geliefert bekommen, mit denen sie, im oben zitierten Bild von A2, von der bloßen Kenntnis der Regeln des ‚(Schach-)Spiels Unterricht‘ zum Einsatz von zielführenden Strategien gelangen können.

Neben diesen positiven Aspekten von Standardisierung berichten die Ausbilder ebenso über die ‚*Grenzen und Probleme*‘, die diese für die Bewertungsprozesse darstellen kann. Der am breitesten thematisierte Problemkomplex der Standardisierung stellt die ‚*Grenzen der Operationalisierbarkeit*‘ guten Unterrichts generell in Frage, ohne die eine Standardisierung von Bewertungskriterien natürlich nicht funktionieren kann. Bedenken äußert z.B. A6 hinsichtlich der Trennschärfe der einzelnen Bewertungskriterien, die sowohl in dem bisher vom ZfsL Bonn genutzten ‚Qualitätstableau‘ als auch in dem im Kontext des Promotionsprojektes entwickelten Bewertungsinstrumentarium auftauchen:

A6: „[...] Und das fand ich auch einen großen Punkt hier, bei diesem Bogen und auch einen ähnlichen Punkt bei unserem Qualitätstableau, dass die Aspekte sich häufig überlappen, z.B. so was hoher Anteil echter Lernzeit überlappt sich ganz stark mit einem Punkt wie Binnendifferenzierung oder Klassenmanagement oder so was. Und das ist dann schwierig zu sagen, was ist hier der entscheidende Fokus, wie wird man bei einer Überlappung, wie nehme ich selber wahr in der Priorisierung und wie nimmt es der Referendar wahr? Und das muss nicht immer dasselbe sein (A3: nicken, «mhm»(bejahend)).“

Die teils unscharfen Abgrenzungen der einzelnen Bewertungskriterien können also zu einem unterschiedlichen Fokus in der Unterrichtsbewertung bei Ausbildern und Auszubildenden führen, was nicht unbedingt positiv ist, aber im Sinne des oben thematisierten ‚*Gesprächsanlasses*‘ kein großes Problem darstellen sollte,

da ja auch gerade eine unterschiedliche kategoriale Einordnung eines bestimmten Unterrichtsaspektes einen ‚*Gesprächsanlass*‘ zwischen den Akteuren bieten kann. Grundlegendere Bedenken lassen sich da schon aus weiteren Einwänden von A6 ableiten, wenn er hinterfragt, inwiefern zentrale Kriterien guten Unterrichts an Hand von Indikatoren beobachtbar sind:

A6: [...] „Wir müssen uns auch fragen: Was ist durch Indikatoren beobachtbar und wo sind die Grenzen von Indikatoren? Da gibt es z.B. einen Punkt, der mir aufgefallen war: Gemeinsame Verantwortung für den Unterricht. Das halte ich auch für ganz tragend für Unterricht, aber mir ist ganz unklar, wie ich das beobachten kann und ob man das überhaupt beobachten kann, und ob man das nicht nur im Gespräch zwischen Referendar und Fachausbilder darüber, wie Unterricht langfristig angelegt ist, klären kann oder ob man das tatsächlich in einer einzelnen Stunde beobachten kann. [...]“

Ähnliche Bedenken hinsichtlich der Operationalisierbarkeit äußert A4, der in Bezug auf das Kriterium eines lernförderlichen Klimas ausführt, dass die situativen Bedingungen darüber entscheiden, ob der Indikator eines geringen Umfangs an Störungen, der im vorliegenden Bewertungskonzept – und auch bei Hilbert Meyer – u.a. dazu genutzt wird, dieses Kriterium zu operationalisieren, geeignet ist, diese zu erfassen oder ob nicht in bestimmten Situationen viel mehr das Vorliegen von ‚Störungen‘ erst ein lernförderliches Klima aufzeigt:

A4: „[...] Also «Umfang der Störung ist gering», alleine was mit dem Begriff Störung verbunden ist (A2, A3: «mhm» (bejahend)), ist das für mich jetzt nur was Negatives oder ist das auch was Positives (A3: mhm, genau)? Und eine Störung, um jetzt ganz mal meine große Verehrung, also ich als Verehrer von Ruth Cohn: Ist Störung was ganz Wertvolles (A3: «mhm» (bejahend)). Nicht in jedem Fall, ich lege wert drauf, dass Ruth Cohn selber gesagt hat: Nicht Störungen haben per se Vorrang, aber sie nehmen sich Vorrang, wenn sie eben halt bedeutsam sind (A3: «mhm» (bejahend)), wenn sie meine Schülerinnen und Schüler sozusagen aus der Bahn werfen, weil sie etwas nicht mitmachen wollen. Wenn sie das artikulieren ist das super, weil sie dann aktiv sind, weil sie mitdenken und

sich weigern, meinem vorgedachten Lehrweg einfach zu folgen. Also nur um das anzudeuten, dass heißt: «Umfang der Störungen ist gering» ist einerseits sicherlich was ganz Wahres, aber ich muss auch dann wiederum daran denken, im Sinne der Kompetenz: Ja manche Schüler müssen erst mal sozialisiert werden. Ich muss, ich habe jetzt auch das leidvolle Erleben, dass ich meine Schüler erst mal erziehen muss, damit ich mit denen vernünftigen Unterricht machen kann.“

Berücksichtigt man diese Bedenken, dann stoßen standardisierte Bewertungsprozesse auf ihre Grenzen, wenn es um die Frage der Indikatorbildung geht. Kriterien guten Unterrichts können im Rahmen von Aushandlungsprozessen und durch Rückgriff auf empirische Forschung gefunden werden, aber einzelne dieser Kriterien werden im Rahmen der Beobachtungsmöglichkeiten der zweiten Phase der Lehrerbildung nicht beobachtbar sein und in diesem Sinne auch nur sehr eingeschränkt Zielkriterium von standardisierten Bewertungsinstrumenten oder gar Beobachtungsvorgängen sein können. A4 thematisiert die Problematik der Indikatorbildung weiter und führt aus, dass diese als Prozess begriffen werden müsse, der nie abgeschlossen ist, sondern immer wieder neu beschrritten werden muss.

A4: „[...] Und wir versuchen die Indikatoren zu finden, mit denen wir so was messen können (A6: «völlig richtig», A3: «mhm» (bejahend)). Und es wichtig zu wissen: Es sind immer nur Idealisierungen, Reduktionen, die nicht wirklich alles erfassen, aber die helfen vielleicht etwas besser manches zu verstehen und auch wieder zu verwerfen.“

Standardisierte Bewertungsinstrumente und -prozesse erscheinen demnach selber als prozesshaft und immer nur temporär gültig:

A4: [...] Manchmal scheint alles, was ich hier gemacht habe, Lug und Trug, aber es funktioniert bestens. Und trotzdem hat es sich bewahrheitet. Und so sind ja auch hier unsere Koryphäen, Meyer und Helmke, herangegangen. Sie haben ja letztendlich keine/Sie berufen sich auf Beobachtungen, in denen man sozusagen Bewährtes versucht hat zusammenzufassen und das ist der Weg, den wir jetzt erst mal voranschreiten. Und ich finde es jetzt wichtig, dass wir uns jetzt manchmal wagen, weiter wagen, aber dann

auch die Bereitschaft haben müssen: Nee, das ist es doch nicht so richtig, dieser Indikator ist unscharf, ist nicht präzise genug, ist untauglich.“

Vor dem Hintergrund der unausgereiften, wenn man so will, in einem frühen Stadium der wissenschaftlichen Evolution stehenden Operationalisierungsmöglichkeiten erscheint für A4 die Fixierung von Qualitätskriterien manchmal als „Lug und Trug“, weil sie es eben nicht schaffen, guten Unterricht generell zu operationalisieren, sondern es immer Ausnahmen geben kann, die die fixierten Kriterien nicht erfüllen, der Unterricht aber trotzdem gut in einem – wenn man so will – eigenen Sinn ist. Eine Erfahrung, die auch A3 aus seiner eigenen Ausbildertätigkeit kennt und standardisierten Bewertungskriterien deswegen eben nur eine eingeschränkte Gültigkeit zugesteht:

A3: „Zusammenfassend würde ich sagen, heißt das doch auch: Das ist nicht mehr, aber immerhin, es ist ein Modell oder es sind Modelle die dahinter stehen, aber es nicht eine Abbildung dessen, was wir beobachten können, sondern es sind mögliche Modelle, Verstehensmodelle. Und jeder wahrscheinlich von uns kennt Beispiele, wo jemand all das nicht erfüllt und trotzdem guten Unterricht macht (A6: «Ja», A4: «mhm» (bejahend)). Wo Schülerinnen und Schüler viel lernen, begeistert sind, also wo Wesentliches geschieht und trotzdem erfüllt der- oder diejenige nicht diese Kriterien, das gibt es auch (A4: «Ja!»). Es sind eben, ja Modelle der Wahrnehmung die helfen, aber eben begrenzt sind in ihrer Aussagekraft, das würde ich schon so sehen.“

Standardisierte Bewertungskriterien und darauf aufbauende Modelle können also nur eine prozesshafte, fluide Gültigkeit guten Unterricht zu operationalisieren beanspruchen, wie mit A6 ausgeführt werden kann:

A6: „Ein heuristisches Modell mit Grenzen, die die Wirklichkeit hergibt (A3, A4: «mhm» (bejahend)) und die man immer wieder neu verhandeln muss, weil das Modell sonst, ja obsolet wird. Also neu verhandeln heißt, innerhalb des einzelnen Unterrichtsbesuches, d.h. was den einzelnen, die Entwicklung des einzelnen Referendars angeht, d.h. auch was neue Erkenntnisse der Hirnforschung angeht, also es kann immer nur ganz

punktuell messen und dass ist anders als in der Physik (A1, A3: «mhm» (bejahend)).“

In diesem Sinne kollidiert der Wunsch, eine generell gültige Operationalisierung guten Unterrichts zu erarbeiten mit der ‚*Individualität der Ausbildungsbedürfnisse*‘, die sich in den unterschiedlichen Ausbilder- und Auszubildendenpersönlichkeiten äußert, aber auch in den vielfältigen Gestaltungsmöglichkeiten guten Unterrichts, jenseits der temporär fixierten Qualitätskriterien. In diesem Kontext berichtet A3 von seiner individuellen beruflichen Entwicklung als Ausbilder, in der er zunehmend Abstand von einer Standardisierung der Bewertungsprozesse genommen hat, eben wegen der Bedeutung der individuellen Kontexte:

A3: „So ich denke, angesichts auch der Lektüre Ihrer Vorschläge, dass ich anfangs in der Ausbildung mit ganz umfangreichen Listen ähnlicher Art gearbeitet habe und dachte, je mehr ich die ausarbeite, desto leichter kann ich Qualität von Unterricht erfassen (A6: nicken, «mhm» (bejahend)), aber ich bin damit nie zufrieden gewesen (A1: nicken), und ich habe inzwischen eigentlich es fast aufgegeben, überhaupt mit solchen Listen zu arbeiten. Das hat auch Nachteile, ganz klar, aber der Gewinn dabei ist, dass man eben Individualität und individuelle Lernwege sehr gut, man ist, also begleiten kann und das ist ein großes Plus nach meiner Wahrnehmung, wenn man eben nur eine kurze Zeit hat jemanden zu begleiten und ein Stück zu unterstützen. Diese GROßEN Unterschiede im individuellen Lernprozess. Also insofern stehe ich grundsätzlich solchen Listen ganz skeptisch gegenüber, [...]“

Durch die „großen Unterschiede im individuellen Lernprozess“, die A3 betont, können die Ausbilder keine allgemeingültige Operationalisierung und damit auch keine starr fixierten oder gar gewichteten Bewertungskriterien guten Unterrichts als Schablone der Bewertungsprozesse nutzen. *Alle Ausbilder* sind sich darin einig, dass die situative Betrachtung, vor dem Hintergrund der jeweiligen Bedürfnisse der Auszubildenden, den Ausschlag geben muss für die Bewertung:

A4: „Wir müssen selber der Individualität unserer Referendarinnen und Referendare gerecht werden (A1, A3: Ja genau), d.h. bei dem einen ist es das, bei dem anderen ist es das. Und das muss ich!

A2: Diskutieren auch in der Arbeit dann (A3: «mhm» (bejahend)).

A4: Das muss ich mit ihm, mit ihm gemeinsam herausfinden (A3: «mhm» (bejahend)). Also ich habe eine These und er muss draufantworten, stimmt oder stimmt nicht oder wie auch immer (A3: «mhm» (bejahend)). Also in diesem Reflexionsprozess und das, darum geht es da (A1: Ja).“

Diese situative Betrachtung kann nur im Dialog zwischen Ausbildern und Auszubildenden gelingen, was ein weiteres Mal die zentrale Bedeutung aufzeigt, die funktionierende und auf gegenseitiges Verständnis angelegte Kommunikations- und Interaktionsmuster, aus der Perspektive der Diskutanten für eine erfolgreiche Ausbildung haben. Neben den individuellen Professionalisierungsbedürfnissen der Auszubildenden spielen aus der Perspektive der Ausbilder auch die je situativ zu begründenden Bedürfnisse der Schüler an den Unterricht eine Rolle, die gegen eine schablonenhafte Interpretation der Standardisierung sprechen:

A1: „Ich wollte gerade sagen, es geht ja hier noch mal um, auch Qualitätsmerkmale. Also der kann ja durchaus Dinge da anlegen, die in der Planung vielleicht gut und richtig erscheinen, aber dann in seiner Durchführung qualitativ einfach nicht gelungen sind und die für ihn, vielleicht auch sogar für ihn persönlich, nicht die richtigen sind. Eine tolle Methode, die man vielleicht im Fachseminar besprochen, sogar vielleicht mit demselben Inhalt besprochen hat, aber für den einzelnen Menschen in dem Moment nicht greift. Die kann für mich gut sein, aber für ihn nicht.

A4: Ja oder für die Schüler nicht.

A2: Ja oder für die Schüler nicht (A4: «mhm» (bejahend))!“

Theoretisch kann ein bestimmter Unterricht extern festgelegten Qualitätskriterien voll entsprechen, aber trotzdem an dem Lernbedarf der konkret Beteiligten

vorbei gehen, sowohl auf Seite der in der Ausbildung befindlichen Lehrer, als auch auf Seite der Schüler, so das Credo, dass sich aus der Perspektive von *mehr als der Hälfte der Ausbilder* rekonstruieren lässt.

In diesem Sinne lässt sich für die Perspektive der Ausbilder zusammenfassend sagen, dass die Standardisierung bzw. Bewertungsinstrumente und -konzepte, die auf standardisierten Bewertungskriterien beruhen, auch ‚*nur ein Arbeitsinstrument*‘ sind, d.h. dass operationalisierte Qualitätskriterien, Bewertungsinstrumente, die die Bewertungsabläufe standardisieren sollen oder Skalen, die einer möglichst vergleichbaren Erfassung von Unterrichtsqualität dienen, keinen Wert an sich haben, sondern ihr Nutzen immer nur vor dem Hintergrund der jeweiligen Ausbildungssituation erkannt werden kann, die Standardisierung nicht „verabsolutiert werden darf“, wie es A4 ausführt:

A4: „[...] Das andere würde ich sagen, die Standardisierung ist in dem Sinne wirklich eine Hilfestellung, um eine komplexe Wirklichkeit wahrzunehmen. Wir müssen uns davor hüten sie zu verabsolutieren, das ist für mich was ganz Entscheidendes.“

Alle Ausbilder pflichten A4 in dieser Einschätzung bei, wie die folgende Diskussionspassage zeigt, in der A3 den Begriff des „Arbeitsinstruments“ prägt:

A3: „Vielleicht das Stichwort Arbeitsinstrument. Dass Standardisierung, das ist auch nicht mehr als Arbeitsinstrument und nicht mehr (A1: Ja). Darüber stehen andere Dinge. Es ist wirklich, es ist hilfreich, ja sehr hilfreich in mancher Hinsicht, aber auch nicht mehr, es ist keine Wahrheit.“

A1: Das muss auch den Auszubildenden klar sein. Das sieht oft so aus, als wenn das wirklich das Maß der Objektivität ist und das ist es nicht (A5: «mhm» (bejahend)).

A3: Genau, ist furchtbar.

A6: Gibt es in unserer Professur nicht (lachen in der Runde). Das stimmt (A3, A4: Ja).“

4.3.2 Die Sichtweise der Auszubildenden

Alle Auszubildenden berichten in einer oder mehrerer der Subkategorien über den ‚Nutzen von Standardisierung‘ für ihre eigene Professionalisierung. Bei *mehr als der Hälfte* der Auszubildenden haben standardisierte Bewertungskategorien, die schriftlich vorlagen, dazu geführt, dass diese im Sinne der ‚Modellbildung für einen heuristischen Zweck‘ einen Beitrag zur individuellen Kompetenzentwicklung geleistet haben. R4 berichtet in diesem Zusammenhang von der besonderen Herausforderung, den richtigen Fokus für die Planung und anschließende Auswertung einer Stunde zu finden, die sich ihm v.a. am Anfang der Ausbildung stellte:

R4: „[...] Das ist auch teilweise gerade am Anfang, also das finde ich unheimlich schwer, die zu (R3: Alle) beachten. A) alle sowieso, geht natürlich nicht, und dann, dass man auch wissen muss, was ist für meine Stunde das entscheidende Kriterium, sozusagen? Natürlich will ich wissen, dass Lernprogression zum Beispiel immer ein besonderer Faktor ist. Klar, aber, dass das je nach Stunde auch natürlich dann variiert, dass man dann sein Augenmerk tatsächlich auf ein zwei andere Kriterien legen müsste, das kriegt man gerade am Anfang auch noch nicht so hin, und dann versucht man das, glaub ich auch, genauso in der Nachbesprechung. Da/ Ich habe mich auch immer an diesen Kriterien versucht zu orientieren, aber das ist auch, ja es fehlt einem trotzdem so der Halt, sag ich mal, oder die, das wo man sagt: «So ok, das ist wirklich transparent und auch nachvollziehbar und pointiert» (R3: Ja). Also dass ich sagen kann: «Ok, das ist es jetzt worauf ich zu achten habe.»“

Offenbar waren die fixierten Kriterien im Rahmen des ‚Qualitätstableaus‘, das das ZfsL Bonn als Handreichung anbieten konnte, zumindest teilweise eine erste Orientierungshilfe, die R4 herangezogen hat. Aber einen wirklich „transparenten und auch nachvollziehbaren Halt“ haben die standardisierten Bewertungskategorien R4 nicht gegeben. R3 hat die Kurzform der standardisierten Bewertungskategorien des ZfsL Bonn, die sog. ‚Bubbles‘, pragmatisch für die Vorbereitung der anschließenden Unterrichtsnachbesprechungen genutzt und konnte diesen Instrumenten durchaus einen Nutzen abgewinnen:

R3: „Fürs Statement war es wirklich in dem Sinne allein schon mal, weil man nach der Stunde oftmals vollkommen erschöpft aus dem UB rauskam, also hatte man endlich hinter sich und da war auch wieder diese Übersichtlichkeit: Ein DIN A 4 Blatt mit ein Paar Unterpunkten und ich konnte nochmal echt sagen: «Boah, in dieser Stunde ist mir das richtig gut gelungen, aber das richtig schwer.» Weil allein nur so vier Unterpunkte, und du wusstest, aha genau Strukturierung, das ist wichtig (R2, R4: «mhm» (bejahend)), hat das heute bei mir gut geklappt, auf jeden Fall, dass kann ich gut ins Statement einbringen oder das hat halt überhaupt nicht geklappt (R5: «mhm» (bejahend), nicken), das muss ich wahrscheinlich erwähnen. Das war unfassbar hilfreich, fand ich, immer nach diesem UBs, nachdem man auch echt platt und fertig war (R5: «mhm» (bejahend)), nicken), was ganz Übersichtliches und Einfaches zu haben und da die Schwerpunkte zu suchen, zu sagen: «Hier die drei suche ich mir jetzt raus.»“

Der Überblick über die relevanten Qualitätskriterien, die – zumindest theoretisch – Grundlage der Bewertungsprozesse sind, war „unfassbar hilfreich“, v.a. auch vor dem Hintergrund der starken Erschöpfung, die R3 im Anschluss an ihre Unterrichtsbesuche gespürt hat. Insofern scheint sie „Halt“ in den standardisierten Bewertungskriterien gefunden zu haben, während dies bei R4 nur teilweise der Fall war. Ähnlich positiv berichtet auch R5 über die Bedeutung der vorliegenden Qualitätskriterien, die ihr eine Hilfestellung sowohl für die Planung von als auch die Reflexion über ihren Unterricht waren:

R5: „[...A]ber vorher war es ganz angenehm, wenn man seine Stunde geplant hat, als Vorbereitung so. Man konnte daran auch nochmal seine Stunde checken und schauen, welche Punkte habe ich denn hier in der Stunde von vornherein versucht zu berücksichtigen (R4: «mhm» (bejahend)), konnte sich die schon mal markieren und dann später im Statement gucken: Ok sind jetzt das auch die, die ich im Statement wirklich ansprechen will oder schmeiße ich einen raus, nehme ich einen anderen rein, konnte also anhand dieser Punkte auch nochmal seinen Plan durchgehen und das später nochmal für sich selbst überprüfen. Also das war schon wirklich sinnvoll. (...)“

Vorliegende Qualitätskriterien guten Unterrichts sind „wirklich sinnvoll“, damit die Auszubildenden mit den Herausforderungen der Unterrichtsbesuche umgehen können. Sie nutzen diese offensichtlich zur Vorbereitung der Besuche und überprüfen, ob ihr Unterricht zentralen Qualitätskriterien entspricht. Insofern scheint die Standardisierung von Bewertungskategorien durchaus eine Modellbildung im heuristischen Sinne für die Auszubildenden anzubieten.

Ein weiterer Nutzen, den *alle Auszubildenden* in standardisierten Bewertungsprozessen sehen, ist eine Steigerung der ‚*Transparenz*‘, wobei sowohl standardisierten *Bewertungskategorien*, als auch einer Standardisierung der *Bewertungsvorgänge* die Möglichkeit zugesprochen wird, eine Steigerung der ‚*Transparenz*‘ herbeizuführen. R3 erhofft sich durch eine Standardisierung der Bewertungskategorien eine verlässlichere Bewertung durch verschiedene Ausbilder und eine damit einhergehende Reduktion der Ängste, und R4 pflichtet bei, dass dies eine Vergleichbarkeit der Bewertungen gewährleisten könnte:

R3: „Ich glaube, dass es einem viel Ängste nehmen kann, weil man halt, ja einmal irgendwie weiß, was einen erwartet, das haben wir gleich am Anfang schon gesagt. [...] Glaube ich, dass es einen ja auch dahingehend beruhigen würde, dass ich während eineinhalb Jahren Ausbildung mir sicher sein kann, dass nicht plötzlich in der UPP mich fünf Personen erschrecken, die unfassbar böse gucken und ich das Gefühl habe: Oh Gott, vielleicht haben die jetzt gerade einen ganz anderen Maßstab (R1, R4: Ja.) und meine Stunde ist die ganze Zeit/Also wenn ich jetzt weiß, ich wurde eineinhalb Jahre so bewertet und ich bin meine UPP-Stunde hundertmal nach diesen Kriterien durchgegangen und frage Kolleginnen und Kollegen und wir haben die umgewälzt und alle sagen: Komm da kann nicht irgendwie alles komplett daneben gehen, dann muss einem das ja auf jeden Fall Sicherheit geben können. [...]

R4: Ja, auch da wieder Vergleichbarkeit einfach (R3: «mhm» (bejahend)), in jeglicher Hinsicht. Also seminarintern, aber auch seminarübergreifend, also das wäre /“

Eine Standardisierung der Bewertungsvorgänge durch Instrumente, die zur schriftlichen Bilanzierung des Unterrichts zwingen werden von *mehr als der Hälfte* der Auszubildenden ebenfalls positiv gesehen. R5 hebt die Bedeutung schriftlicher Bilanzierungen für die Transparenz der Bewertungsprozesse hervor, die aus ihrer Perspektive sowohl für eher selbstkritische als auch für jene Auszubildenden, die sich selbst in einem zu positiven Licht sehen, förderlich wären:

R5: „[...] Wo also einfach auch das, was die Fachleiter sagen, bei dem jeweiligen Referendar ganz anders ankommt. Bei dem einen kommt nur die Kritik an, obwohl die Fachleiter vielleicht denken, ich habe ihn auch gelobt, (R4: «mhm» (bejahend)) aber man hört es einfach auch nicht, man hört mit den falschen Ohr zu (R3: «mhm» (bejahend)). Und dass man da, wenn man da irgendwo eine Möglichkeit hätte am Ende, ja, irgendwie sich einzuschätzen und aber auch wirklich das schwarz auf weiß mal zu sehen, wo stehe ich denn, dann kann man vielleicht einerseits nicht mehr leugnen, wie schlecht man ist, aber auf der anderen Seite vielleicht auch gerade, wenn man sehr stark kritisiert wurde, trotzdem sagen, aber hier hatte ich ja auch gute Punkte, auch wenn die das jetzt nicht gesagt haben und sich dann ein bisschen wieder emotional fangen (R3: «mhm» (bejahend)), also für beide Referendarstypen fände ich es da wichtig.“

Eine verbindliche, an Kriterien orientierte schriftlich fixierte Bewertung kann, so erhoffen es sich die Auszubildenden, die Transparenz der Bewertungen für beide Seiten erhöhen und somit auch zu einer Reduzierung des subjektiven Faktors in den Bewertungsprozessen beitragen, wie R1 ausführt, die die Problematik auch aus ihrer eigenen Bewertungstätigkeit als Lehrerin kennt. Schriftlich fixierte, kriteriale Bewertungen können die Ausbilder zu mehr „Transparenz sich selbst gegenüber“ zwingen, so die mit der Standardisierung verbundene Hoffnung von R4, die er in der folgenden Passage äußert:

R1: „Vor allem auch, weil dann ja auch wirklich Schwarz auf Weiß, auch vor dem Fachleiter liegt, was er jetzt zu dem Referendar sagen kann (R4: «mhm» (bejahend)). Weil es ist ja, wie bei uns, wenn wir sonstige Mitarbeitsnoten geben, dann erinnert man sich an das eine und an das andere, aber an ganz viele Sachen gar nicht und deshalb habe ich mir auch

gedacht, dass das wahrscheinlich, wenn man es gut führt, schon gerechter werden kann, weil die dann einfach eher sehen: Ok, da habe ich jetzt aber auch mal ein schlechtes Kreuz gemacht, obwohl ich doch so einen guten Eindruck von der habe und mir gedacht habe, vielleicht/

R4: Also dass man sich zwingt selber, also Transparenz sich selber gegenüber (R1: Ja) aus der Sicht der Fachleiter, dass die dann auch mal wirklich eben tatsächlich weniger nach Gefühl gehen und tatsächlich dann auch mehr nach Kriterien, das ist/

R1: Oder bei jemandem, der mir dann nicht so sympathisch ist, weil es spielt ja schon rein auch (R4: Natürlich), dass ich dann sage, ok, da habe ich den aber gut bewertet (R2: Ja).

R4: Ja, genau so was.“

Aus der eigenen Ausbildung kennt R3 standardisierte Bewertungsinstrumente einer Ausbilderin, durch deren Einsatz die ‚Transparenz‘ der Bewertungsprozesse gesteigert werden konnte, weil man dadurch wusste „was die Fachleiterin“ von einem möchte:

R3: „[...] Und da waren dann verschiedene Kriterien aufgelistet, wie aktive Bewegungszeit, wie was für den Sportunterricht wichtig ist, kognitive Phasen. Also Kriterien waren links aufgezählt, dann hatte sie, glaube ich, so drei Kategorien, inwiefern das sehr erfüllt ist oder nicht so sehr erfüllt und sich selber dann rechts noch eine Spalte gelassen, in der sie dann immer schnell einträgt, was ihr dann noch halt an Notizen noch einfällt, was man in der Stunde besonders gesehen hat oder was verbesserungswürdig wäre. [...]

I: Fandest Du das gewinnbringend?

R3: Das vorher gesehen zu haben? Das sie das so bewertet?

I: Unter anderem, die ganze Art der Bewertung, so mit so einem standardisierten oder halb-standardisierten Bogen vorzunehmen.

R3: Ja doch. Das wäre wieder das – ehrlich gesagt – was wir eben gesagt haben, schnellstmöglich herauszufinden, was der Fachleiter von einem möchte, worauf es ankommt und das vor Augen zu haben, auch in einer übersichtlichen Art und Weise, hilft auf jeden Fall. [...]“

Der Nutzen von Standardisierung lässt sich auch aus der Perspektive der Auszubildenden in der Herstellung eines ‚Gesprächsanlasses‘ zusammenfassen, zu dem standardisierte Bewertungsprozesse, z.B. durch den Einsatz von fixierten Bewertungskategorien, einen Beitrag leisten können und es durch den auf Kriterien gestützten Austausch zwischen Ausbildern und Auszubildenden zu einem Klärungsprozess über die Bewertung und Möglichkeiten der Professionalisierung kommt, wie R3 zusammenfassend erläutert:

R3: „In Anbetracht vielleicht auch, was die Konstruktivität der Kritik angeht, was ich am Anfang mal gesagt habe. Dass ich es schlimm fand rauszugehen und man hatte nur, sozusagen wie eine Bewertung gesagt gekriegt: «Es war schlecht», aber man hat es nicht so richtig umgewalzt. Ich glaube auch mit, um so mehr Kriterien man hat irgendwie, so ein festes Raster, umso mehr kann man wirklich über konkretere Situationen oder Möglichkeiten eben halt reden, die einem dann hoffentlich irgendwie das transparenter machen, verdeutlichen und das Gefühl insgesamt auch dann verbessern und nachvollziehen zu können, wonach gerade bewertet wurde, warum, wie und dass ich das hoffentlich verändern und verbessern kann.“

Auffällig an der Diskussion der Auszubildenden ist, dass die negative Seite der Kategorie der ‚Standardisierung‘, also die Thematisierung von ‚Grenzen und Problemen‘ der Standardisierung, kaum eine Rolle spielt. Während zur positiv besetzten Seite der Kategorie immerhin 14 Diskussionspassagen codiert werden konnten, so waren es auf der negativen Seite lediglich fünf. Mit drei Codierungen noch am häufigsten vertreten ist der Konflikt der ‚Standardisierung‘ mit der ‚Individualität der Ausbildungsbedürfnisse‘, den *alle bis auf eine Auszubildende* sehen. Es dürfe durch die Standardisierung nicht dazu kommen, dass die Auszubildenden „in ein Raster“ gepresst werden, wie es R5 in der folgenden Diskussionspassage

formuliert. Ein „zu viel an Strukturierung“, die nicht flexibel gehandhabt wird, ist aus Sicht von R4 schädlich:

R4: „[...] Andererseits ist natürlich auch die Frage, inwiefern zu viel Strukturierung bei so einer Nachbesprechung vielleicht auch wieder schädlich ist (R1: «mhm» (bejahend)). Wenn man dann nur so sagt: «Ok Kriterium eins oder Kategorie eins, Kategorie zwei», das ist vielleicht auch nicht das Richtige. Das ist, eine, das ist, glaub ich, auch die Schwierigkeit bei der ganzen Geschichte.

R5: Ich denke auch, dass man grundsätzlich beides braucht (R4: «mhm» (bejahend)), einmal so Kriterien, wo man wirklich weiß, wo stehe ich denn und dann aber auch eine individuelle Bewertung, von dem, was man jetzt eigentlich getan hat, weil man kann so eine Stunde jetzt ja auch nicht zu hundert Prozent ins Raster reinpressen (R1, R2: «mhm» (bejahend), R4: Absolut nicht), sondern muss das ja auch immer nochmals Einzelerlebnis /“

Die problematischen Auswirkung eines unflexiblen „Rasters“ sieht auch R2, die darauf drängt, dass auch für „normale Unterhaltungen“ zwischen Ausbildern und Auszubildenden noch Platz sein muss und sich die Bewertung durch standardisierte Kriterien und Instrumente nicht „auf das Setzen von Kreuzchen“ reduziert werden darf:

R2: „Ich würde mich dem im Prinzip anschließen, aber echt auch noch mal das unterstreichen, dass vor lauter Kreuzchen und jetzt haben wir, Zack, hier das richtige Kästchen gefunden, dann dieses ganz normale, was heißt normale, dass man sich darüber unterhält und vielleicht auch mal ein bisschen irgendwo anders hinkommt. Dass das dann auch bitte noch sein darf (R4, R5: «mhm» (bejahend), nicken), sonst würde ich auch denken, ich muss hier immer in diesem Raster mich unterhalten und will ich aber gerade gar nicht (R4: «mhm» (bejahend)), oder was anderes würde jetzt gerade besser passen. [...]“

Insofern stecken in diesen Beiträgen der Auszubildenden durchaus Hinweise, dass sie die Sichtweise der Ausbilder teilen, dass die Standardisierung „nur ein

Arbeitsinstrument‘ ist und sie nicht als verdinglichter Zweck an sich zu betrachten ist, aber so explizit wie die Ausbilder äußern die Referendare diesen Aspekt nicht. Gleichmaßen selten thematisieren die Auszubildenden die ‚*Grenzen der Operationalisierbarkeit*‘, die *lediglich von R4* in zwei Passagen kritisch in den Blick genommen werden. So verweist er u.a. in Bezug auf das Kriterium der ‚Klarheit und Strukturiertheit‘ darauf, dass die Indikatoren eine „relativ schwammige Geschichte“ seien und eben nicht einfach für sich ständen:

R4: „[...] Wie gesagt, ich finde mit drei: «Unterrichtsziele, Inhalte und eingesetzte Methoden stehen in einem sinnvollen Verhältnis zueinander.» Was ist denn sinnvolles Verhältnis zueinander? Da sind wir doch wieder bei so einer so relativ schwammigen Geschichte und das ist, finde ich, deswegen ein Punkt, den ich, wo ich gesagt habe: Weiß ich nicht, das ist/Wo fängt der Sinn an, wo hört der Sinn auf?“

Die Frage, „wo der Sinn anfängt und wo er aufhört“, kann nur im Rahmen von Gesprächen zwischen den Ausbildern und den Auszubildenden beantwortet werden, also im Sinne kommunikativer Validierung, für die die ‚*Standardisierung*‘ ‚*Gesprächsanlässe*‘ bieten kann.

4.3.3 Zusammenfassende Interpretation ‚*Standardisierung der Bewertungsprozesse*‘

Oben wurde ausgeführt, dass die Auszubildenden Fragen des Nutzens und der Grenzen und Probleme von Standardisierung deutlich seltener thematisierten als die Ausbilder und sich hierbei v.a. auf die Fragen der ‚*Grenzen und Probleme von Standardisierung*‘ beziehen. Hier zeigt sich sehr deutlich der gruppenspezifische Fokus der beiden Diskussionen und eine unterschiedliche Sensibilisierung für die Herausforderungen, die für die jeweils andere Akteursseite erwachsen, wenn die Auszubildenden die Hoffnungen und positiven Potentiale der Standardisierung sehen, aber nicht die Probleme, die diese mit sich bringen kann und die sich v.a. für die Ausbilder stellen. Gleichwohl lassen sich auch Gemeinsamkeiten aus der Einschätzung der beiden Gruppen hinsichtlich der Rolle von Standardisierung für konstruktive Bewertungsprozesse ableiten. Beide Seiten betonen den

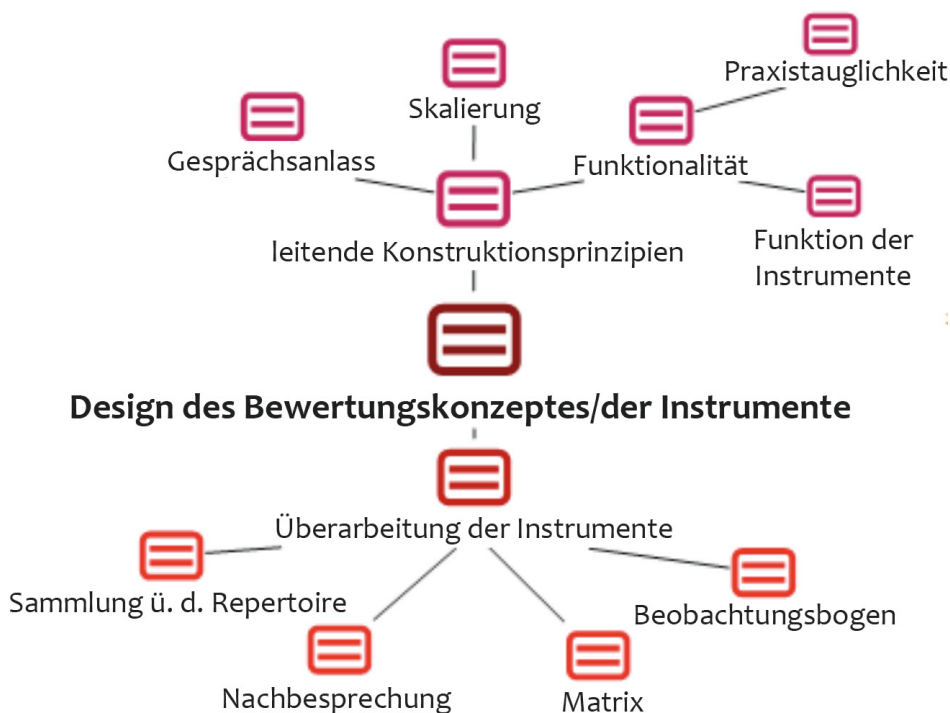
diskursiven Charakter der Bewertungsprozesse und lehnen eine Verdinglichung der Standardisierung und standardisierender Bewertungsinstrumente als Zweck an sich ab. Standardisierung scheint aus der Perspektive beider Gruppen ihren Sinn in der Herstellung von ‚*Gesprächsanlässen*‘ über den gesehenen Unterricht zu erhalten, wobei die Ausbilder diesen Aspekt wesentlich deutlicher verbalisieren, als es die Auszubildenden tun. Offenbar teilen beide Gruppen die Einschätzung, dass die Standardisierung und hierbei v.a. schriftlich fixierte Bewertungskriterien, die auf Modellen guten Unterrichts beruhen, eine wichtige Hilfestellung bei der Professionalisierung der Auszubildenden im Sinne einer ‚*heuristischen Modellbildung*‘ bieten. In diesem Kontext muss aber die Frage gestellt werden, inwiefern die Berichte der Auszubildenden, wie sie standardisierte Bewertungskriterien zur Vorbereitung ihres Unterrichts bzw. zur Reflexion über diesen genutzt haben, den Erwartungen der Ausbilder an diese Modellbildung entsprechen? Oder anders formuliert, ob die beiden Gruppen jeweils eigene Heuristiken der Modellbildung haben. Die Verwendung der Qualitätskriterien als „Checkliste“ vor einem Unterrichtsbesuch oder in der -nachbesprechung verhindert sicher nicht die gewünschte Entwicklung der eigenen Lehrerpersönlichkeit der Auszubildenden, wie sie die Ausbilder wünschen. Allerdings ist es fraglich, inwieweit diese pragmatische, doch eher auf die Erfüllung der an die Auszubildenden gestellten Erwartungen, ausgerichtete Nutzung der Qualitätskriterien den Wünschen der Ausbilder entspricht.

4.4 Design des Bewertungskonzepts/der Instrumente

Die Kategorie des ‚*Designs des Bewertungskonzepts bzw. der Instrumente*‘ nimmt, anders als die beiden zuvor diskutierten Kategorien, keine dichotome qualitative Einschätzung der Diskussionsaspekte vor, sondern rekonstruiert die Einschätzungen der beteiligten Diskutanten, inwieweit die pre-Alpha-Version des Bewertungskonzepts die in sie gesetzten Erwartungen erfüllen kann, wie die Abbildung 4.6 zeigt. Aus diesem Grund werden die Perspektiven der beiden Gruppen auch nicht getrennt ausgewertet, sondern es wird nach Designvorschlägen gesucht,

die aus der Perspektive beider Gruppen Sinn machen. Hierbei ließen sich erstens ‚leitende Konstruktionsprinzipien‘ von Bewertungsinstrumenten finden, die es aus Sicht der Diskutanten zu berücksichtigen gilt, und zweitens formulierten diese konkrete ‚Vorschläge zur Überarbeitung der Instrumente‘.

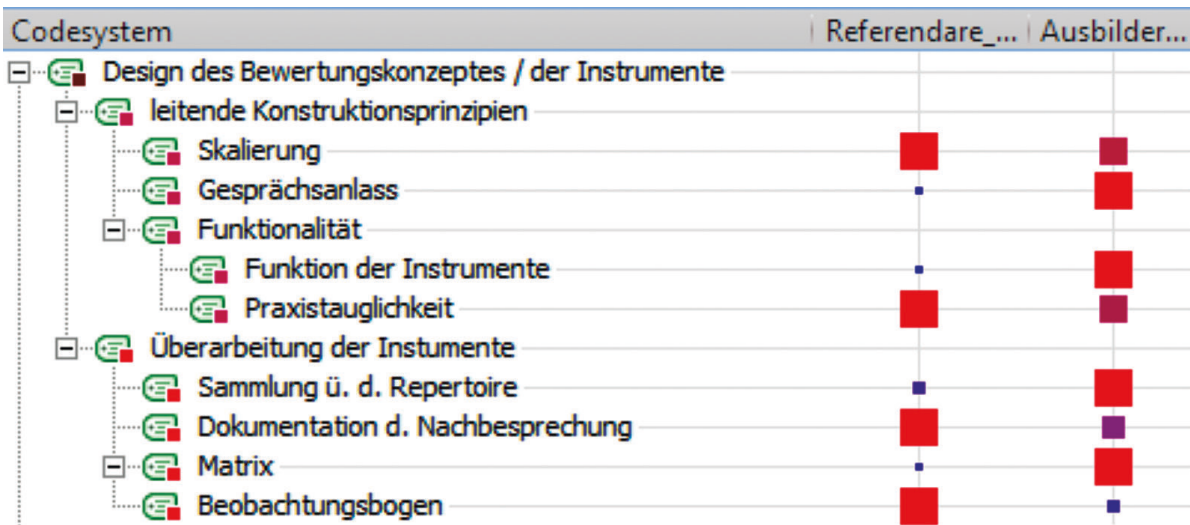
Abbildung 4.6: Subkategorien der Kategorie ‚Design des Bewertungskonzeptes/ der Instrumente‘



Die Sichtweisen der Auszubildenden und Ausbilder werden, wie gehabt, zunächst mit Hilfe des Code-Matrix-Browser von MaxQDA in ihrem rein quantitativen Auftreten deskriptiv gegenübergestellt, um dann gemeinsam interpretativ ausgewertet zu werden. Aus dieser Auswertung werden Schlüsse für die Überarbeitung der pre-Alpha-Version des Bewertungskonzeptes abgeleitet, sodass dieses zur Alpha-Version fortentwickelt werden kann. Die Ausbilder diskutierten das ‚Design des Bewertungskonzeptes‘ in insgesamt 103 Diskussionspassagen, während die Auszubildenden dies in lediglich 76 Passagen taten. Dieser Unterschied kann, zumindest zu einem gewissen Teil, durch die um 25% längere Dauer der Gruppendiskussion erklärt werden, da die Zahl von 27 zusätzlichen Codings in der Diskussion der Ausbilder im Vergleich zur Diskussion der Auszubildenden pro-

zentual nahezu genau der längeren Dauer der Diskussion entspricht. In der Verteilung der Codings in der Code-Matrix, die Abbildung 4.7 zeigt diese, zeigen sich allerdings deutliche Unterschiede zwischen den beiden Gruppen:

Abbildung 4.7 : Code-Matrix der Kategorie ‚Design des Bewertungskonzeptes/ der Instrumente‘



Bei den zeilenweisen Häufigkeitsverteilungen der einzelnen Subkategorien fällt auf, dass jene, die häufiger von den Ausbildern thematisiert werden, auf Seite der Auszubildenden deutlich seltener in den Fokus gelangen, wohingegen jene Subkategorien, die in der Diskussion der Auszubildenden eine größere Rolle gespielt haben, von den Ausbildern nur unwesentlich seltener thematisiert worden sind. Hier zu nennen ist v.a. die Bedeutung der Instrumente für die Herstellung eines ‚Gesprächsanlasses‘ mit 17 (Ausbilder) gegenüber lediglich drei (Auszubildende) codierten Passagen auf der einen Seite, sowie die Frage der ‚Skalierung‘ der Instrumente mit 8 codierten Passagen in der Ausbilder- und 10 codierten Passagen in der Auszubildendendiskussion, auf der anderen Seite. Hier zeigen sich teilweise die bereits in den vorangegangenen Kapiteln thematisierten spezifischen Blickwinkel mit denen die Ausbilder und Auszubildenden auf die Ausbildungsbedingungen und ihre Herausforderungen blicken, die je gruppenspezifische Herausforderungen und entsprechende Fokusse bedingen. In der konkreten Frage des Konzeptdesigns scheinen die Ausbilder eine umfassendere Perspektive

einzunehmen, als es die Auszubildenden tun. Die einzige Ausnahme von dieser Regelmäßigkeit bildet die Frage, wie der Beobachtungsbogen überarbeitet werden sollte, die von den Auszubildenden mehr als drei mal so häufig diskutiert wurde wie von den Ausbildern.

Die Auswertung folgt den einzelnen Subkategorien der Kategorie für beide Gruppen gemeinsam und sucht nach Designvorschlägen, welche beide Gruppen thematisiert haben oder welche von der jeweils anderen Gruppe zumindest nicht abgelehnt wurden, um daraus Leitlinien für die Überarbeitung der pre-Alpha-Version des Bewertungskonzepts ableiten zu können.

4.4.1 Leitende Konstruktionsprinzipien

Aus den Diskussionen beider Gruppen ließen sich *leitende Konstruktionsprinzipien*, welche bei dem Design eines Bewertungskonzepts, das den Anforderungen und Wünschen der beteiligten Akteure entgegenkommt, rekonstruieren. Hierbei geht es v.a. um die Frage der *Funktionalität* der Instrumente, in deren Kontext die Diskutanten einerseits die Frage thematisierten, wie es um die *Praxistauglichkeit* eines standardisierten Bewertungskonzepts bestellt ist und andererseits klar machten, dass die *Funktion der Instrumente* bei ihrem Einsatz in der Ausbildung geklärt werden muss. Der Sinn und Zweck der standardisierten Bewertungsinstrumente und -prozesse ergebe sich eben nicht aus sich selbst, sondern könne nur vor dem Hintergrund der besonderen Ausbildungsbedürfnisse in der zweiten Phase der Lehrerbildung geklärt werden. In der Diskussion der Auszubildenden wird die Frage *nur von einer Auszubildenden* angerissen: Sie fragt, ob die Instrumente bei der Reflexion von Unterricht helfen sollen oder der Dokumentation der Nachbesprechung dienen. Sie verweist insofern nur implizit auf die Bedeutung einer Klärung der Intention des Einsatzes der Instrumente. *Die Hälfte der Ausbilder* thematisiert die Bedeutung der Klärung dieser Aspekte wesentlich expliziter, hierbei v.a. A3, der mehrfach darauf verweist, dass ein und dasselbe Instrument in der Ausbildung unterschiedliche Funktionen haben könne und eine Klärung der Funktion eines Instruments innerhalb der Logik eines umfassenderen Bewertungskonzepts deswegen notwendig sei:

A3: „Deshalb muss auch eine Dokumentation, es ist ähnlich wie mit dem Repertoire, muss, da muss die Funktion stehen, die dieser Bogen hat. Sonst kann man ihn auch gar nicht verstehen. Also eine Dokumentation um einer Dokumentation willen macht überhaupt keinen Sinn für mich, sondern ich muss wissen, (A4: «mhm» (bejahend)) was ich mit der Dokumentation will. Ist es Grundlage für den Selbstlernprozess? Ist es Grundlage für die Rechtfertigung meiner Bewertung am Ende? Das sind ja zwei ganz unterschiedliche Funktionen (A6: Genau) und ich denke man muss klären, welche Funktion hat, die kann auch wechseln!“

Aus der Perspektive der Ausbilder erhalten die Bewertungsinstrumente ihren Sinn nicht aus sich selbst heraus, ebenso wenig wie dies für die Standardisierung gilt, also muss eine ausdrückliche Klärung hinsichtlich der Frage geschehen, welche Rolle die Bewertungsinstrumente einnehmen. A6 pflichtet A3 in der obigen Passage bei und führt im weiteren Verlauf der Diskussion selber aus, dass die Klärung der „Funktionalität“, wie er es bezeichnet, zentral sei, damit ein Bewertungskonzept seinem heuristischen Zweck innerhalb der Ausbildung, wie er in den vorangegangenen Kapiteln rekonstruiert worden ist, gerecht werden könne:

A6: „[...] Und die Funktionalität der unterschiedlichen Instrumente im Blick zu behalten. Die haben nicht alle dieselbe Funktion und das von Anfang an auch immer klar zu machen. Richtet sich in erster Linie an und dient dem Ziele zu (A1, A3: «mhm» (bejahend)). Dann wird dieser heuristische Prozess von Anfang an immer mitgedacht. Und (...) ja!“

Neben der Klärung der ‚Funktion der Instrumente‘ in den Ausbildungssituationen spielt die Frage der ‚Praxistauglichkeit‘ für beide Gruppen eine entscheidende Rolle, hierbei scheint der Umfang des Konzepts als Ganzes ein neuralgischer Punkt zu sein, wie die Auszubildenden bei der Frage nach dem ersten Eindruck des Bewertungskonzepts ausführen:

I: „[...] Dieses Konzept in Gänze zu bewerten. Wenn ihr euch das/Ihr habt euch das ja zu Hause angeguckt, was ist der erste Eindruck, den ihr davon gekriegt habt?“

R5: Umfangreich (allgemeines Lachen)

R3: Komplex, ja habe ich auch gedacht, das war auch gerade das Adjektiv.

R5: Da hat man verdammt viel Arbeit hinterher, war so mein erster Gedanke. [...]

R3: Also erst mal schon verdammt viel und natürlich benötigt es einfach viel Zeit bis man die ganzen Listen, auch wenn selbst die meisten auch relativ einfach mit ankreuzen oder so und so gemacht sind, wirklich aber viel Zeit, bis man es ausfüllt (R1, R4: «mh» (bejahend)).“

Auf Seiten der Ausbilder sieht zumindest A6 den Umfang des Konzepts kritisch, auch wenn die Entwicklung eines fachübergreifend einsetzbare Konzepts „wahnsinnsschwierig“ sei, müsse die Frage nach der „Handhabbarkeit“ gestellt werden:

A6: „[...] Ich verstehe schon, dass das wahnsinnsschwierig ist, aber ich frage mich, wie handhabbar ein Teil ist, was aus 19 Seiten besteht, oder 27 Seiten besteht.“

Die weiteren Diskussionen um die ‚Praxistauglichkeit‘ drehen sich v.a. um die Frage des Einsatzes der Beobachtungsbögen, die auf beiden Seiten der Akteure eine überwiegend negative Würdigung erfahren. Interessant sind hier v.a. die Berichte *von der Hälfte der Ausbilder*, dass diese im Zuge ihrer Ausbildung versucht haben, mit Hilfe standardisierter Beobachtungsbögen eine möglichst nachvollziehbare Bewertung der Unterrichtsbesuche zu gewährleisten, aber alle von der Nutzung standardisierter Bögen abgerückt sind, weil diese in der Praxis nicht gewinnbringend eingesetzt werden konnten. Als Gründe führt z.B. A5 an, dass die Beobachtungsbögen ihn von seiner eigentlichen Aufgabe, den Unterricht zu beobachten, abgehalten hätten.

A5: „Ja ich mach jetzt, bin jetzt auch seit drei Jahren Fachleiter und hatte zu Beginn, ähnlich wie du A1, ganz viele Beobachtungsbögen, einmal kriterienorientiert, einmal kompetenzorientiert und bin dann wirklich, hab mich dann einmal bei einem Unterrichtsbesuch ertappt, dass ich

wirklich primär nur auf diese Bögen geguckt habe und gar nicht mehr auf den Unterricht und auf den Lehrer und auf die Schüler geschaut habe. [] Aber ich habe für mich mir wirklich gemerkt, dass im oder während des Unterrichts mich diese Bögen ein bisschen behindern und dass ich da lieber genau (A3: «mhm» (bejahend), nicken)) /“

Ähnliche Erfahrungen haben auch A1 und A3 im Zuge ihrer Tätigkeit als Ausbilder gemacht und die gleichen Schlüsse wie A5 gezogen. Die anderen Ausbilder berichten nicht explizit über ihre Erfahrungen mit standardisierten Beobachtungsbögen, aber teilen scheinbar die Einschätzungen von A1, A3 und A5. Letzten Endes zeigen sich hier wieder die strukturell bedingten Grenzen der Beobachtungsfähigkeit der Ausbilder, wie sie bereits in der Auswertung der vorherigen Kategorien deutlich wurde, die grundsätzlich in Frage stellen, welchen Sinn der Einsatz von standardisierten, mit zahlreichen Bewertungskategorien versehenen Beobachtungsbögen im Kontext formativer Evaluation von Unterricht hat. A4 fasst die grundsätzliche Problematik in Bezug auf die Nutzung von Beobachtungsbögen ein weiteres Mal zusammen:

A4: „Praktikabel für wen, unter welchen Bedingungen? Wenn ich mich jetzt an einen Beobachtungsbogen hier setzen muss und ich muss das alles im Kopf haben, wird es eben sehr schwierig so viele Items wahrzunehmen. Und ich muss, habe 45 Minuten jetzt im Moment vor Augen, demnächst ja vielleicht auch, jetzt manchmal auch schon 60 und wenn ich die, muss ich/Manche sind vielleicht querschnittsmäßig für eine Stunde zu benennen, manche müssen aber eigentlich mehrfach in der Stunde benannt werden, macht es eben halt Aufwand, sozusagen mich immer zu verorten.“

Diese Frage der begrenzten Beobachtungsfähigkeit der Ausbilder wird von den Auszubildenden nicht explizit thematisiert, was insofern nicht verwunderlich ist, als dass diese Herausforderung die Ausbilder und nicht sie betrifft und die Problemsensibilisierung der Gruppen ja offensichtlich von den gruppenspezifischen Herausforderungen geprägt wird. Hingegen werden von beiden Gruppen Probleme gesehen, die Beobachtungsbögen unter den gegebenen Umständen der Ausbildung zufriedenstellend auszuwerten, was R2 in der folgenden plakativen Aussage zusammenfasst:

R2: „Also, ich glaube ich, würde auch schon wieder denken, wenn ich das jetzt hier alles besprechen muss, dann gehe ich lieber vorher (Allgemeines Lachen). [...]“

Alle Auszubildenden und Ausbilder scheinen sich einig zu sein, dass eine Auswertung der Beobachtungsbögen nicht in Gänze, sondern nur hinsichtlich ausgewählter Kriterien geleistet werden sollte, da eine umfassende Auswertung aller Kategorien den Rahmen der Unterrichtsnachbesprechungen sprengen würde, wie R2 im Anschluss an das obige Zitat weiter ausführt:

R2: „[...] Also ich würde mir, also ich glaube ich könnte, würde das als gewinnbringend dann empfinden, wenn man bestimmte Sachen raus sucht (R4: Genau, R5 «mhm» (bejahend)), also wenn man irgendwie nicht alle 23 Punkte bespricht (R5: nicken.), sondern irgendwie, weiß ich nicht, im Seminar dann sagt: «Am Anfang achten Sie besonders auf die Punkte ddddmmmmm» und da gucken wir dann mal nach, und wenn das in Richtung UPP-Standard oder was auch immer wo drüber gesprochen wird gehen könnte, dann könnte man mehr Instrumente oder mehr Punkte noch genauer dazunehmen. Dann würde es mir, glaube ich, helfen.“

Abgesehen davon, dass die Auszubildenden hier ihre Hoffnungen in eine Klärung der Beobachtungsschwerpunkte über eine transparente Ausbildungsprogression setzen, was vor dem Hintergrund der bisherigen Ausführungen wenig wahrscheinlich erscheint, sind sie sich doch mit den Ausbildern darüber einig, dass es im Sinne der ‚Praxistauglichkeit‘ nur Sinn macht, ausgewählte Aspekte eines Beobachtungsbogens zu besprechen und diesen nicht in seiner „ganzen Komplexität“, wie es A1 formuliert:

A1: Aber das müsste vielleicht klar gestellt werden, dass man nicht einen derartigen ganzen Bogen in seiner ganzen Komplexität bespricht (A3: Nee, genau), sondern dass man tatsächlich (A4: Auswählt), wenn ich Unterricht beobachte, gezielt auf die entsprechende Stunde, die ich betrachte, auswähle (A3: «mhm» (bejahend)) oder auswählen lasse vom Referendar oder der Referendarin (A4: Ja was ich vorziehe). [...]“

Die Probleme für einen Einsatz in der Ausbildungspraxis, die relativ umfangreiche und differenzierte Beobachtungsbögen mit sich bringen werden auch deutlich bei den abschließenden Empfehlungen, welche die Diskutanten zum Ende der Diskussion abgeben, bei denen der Appell „sich kürzer zu fassen“ in beiden Gruppen immer wieder auftaucht:

R4: „Immer: Weniger ist mehr irgendwie (R5: Genau). [...]“

R5: „Also ich denke auch eben, wenn wir es wirklich kürzer fassen, [...]“

R3: „[...]F]ände ich es persönlich, um es nachzuvollziehen als Referendar angenehmer, wenn es kürzer und prägnanter wäre.[...]“

A3: „Wählen Sie mehr aus! Wählen Sie mehr aus! (A1, A4, A6: Ja) Seien Sie mutiger, was exemplarisches Vorgehen betrifft (A6: Ja).“

A4: Also nicht zu komplex (A6: Ja).

A3: Nicht so komplex, auswählen. [...]“

Aus den Diskussionen beider Gruppen wird deutlich, dass die Frage, mit welchen Skalen der Kompetenzaufbau erfasst werden soll, durchaus eine Rolle für den heuristischen Zweck des Bewertungskonzepts spielt. Allerdings weniger die Frage welche Art der Skala genutzt wird, als vielmehr die Tatsache, dass durch *irgendeine* Art von Skala *Teilaspekte* einer allgemeinen Unterrichtskompetenz bewertet werden. In der Diskussion der Auszubildenden wird zunächst deutlich, dass es offensichtlich einen Konflikt zwischen der ‚*Praxistauglichkeit*‘ einer ‚*Skalierung*‘ und der Genauigkeit, mit der sich Unterrichtskompetenzen erfassen lassen, gibt. Diese werden besonders in den Ausführungen von R4 deutlich, der zunächst klar für die Verwendung der Ratingskala für den Beobachtungsbogen votiert:

R4: „Da sehe ich auch das Problem, dass man gerade am Anfang, wäre man, glaube ich, völlig erschlagen. Wenn ich das vorgesetzt bekomme, würde ich als Schüler, als Schüler – sage ich jetzt auch schon – (allgemeines Lachen) als Referendar tatsächlich erst mal denken: «Woah, wie soll ich das

denn alles umsetzen?» und hätte dann auch Panik (R2: Kann ich nicht!), dass ich erst mal versuche alles zu machen, also alle 22 oder zwölf oder wie viel (R5: 23), 23 Punkte hier angucke und denke: «Oh Gott, vielleicht musst du die versuchen umzusetzen, also irgendwann am Ende musst du alle in eine Stunde packen.» [...] Und deswegen, aus diesem Grunde auch würde ich, wenn ich mich entscheiden müsste, definitiv für diesen Ratingskalaentwurf entscheiden (R5: nicken). Auch da wieder Gründe anhand der Übersichtlichkeit wirklich, also das kriegen die auf zwei DINA 4 Seiten, während wir hier vier oder fünf oder so was haben, ich weiß es gar nicht oder sechs sogar.“

Im weiteren Verlauf der Diskussion gibt R2 zu bedenken, dass eine Ratingskala im Vergleich zu einer ausformulierten Kompetenzskala eine weniger eindeutige Erfassung von Kompetenzen ermögliche und deswegen für sich genommen keine zufriedenstellende Grundlage für eine Unterrichtsnachbesprechung biete, was R4 in Abweichung seiner ersten Positionierung durchaus als Vorteil einer ausformulierten Kompetenzskala anerkennt:

R2: „Ich würde nochmal das von dir eben (R5) unterstreichen, dass man tatsächlich das oder das hast du im Prinzip auch gesagt. Diese ausformulierten Sachen, also Besprechungsgrundlage für so ein ankreuzen in der Tabelle, sonst bleibt das sehr schwammig, sonst kann man, kann ich mir selber ausdenken, was jetzt gut ist oder was jetzt nicht gut ist (R4, R5: nicken, «mhm» (bejahend)). Aber in dem Moment, wo man das in einem Text mit so quasi Indikatoren hat, dann hat man auch etwas Griffiges.“

R4: „Auch wieder wenn man nur stehen hat: „Die Schüler und Schülerinnen verstehen die Aufgabenstellung, können erläutern, was sie gerade tun“, ist ja sehr kurz und knapp gehalten (R1, R2: «mhm» (bejahend)). Wenn man aber daneben stehen hat: „Die Aufgabenstellungen sind nicht klar, die Schüler müssen mehrfach neu instruiert werden, bevor sie mit dem Arbeiten beginnen“. Das ist, definiert genau, aha, was dann eben nicht stark ausgeprägt ist und ich glaube, das ist dann schon/da lässt sich wenig drüber diskutieren einfach, das ist relativ klar vorgegeben, da kann man ziemlich genau sagen, wie man die Stunde sieht, ob das erfüllt ist.“

Die Operationalisierung von Kompetenzstufen über einer Kompetenzskala, auch wenn sie nur in einem pragmatischen Sinne erfolgt und nicht Teil einer komplexen Kompetenzmodellierung ist, scheint also durchaus einen Zugewinn an Transparenz für die Erfassung von Unterrichtsqualität und -kompetenz zu ermöglichen, so die Sichtweise von der *Hälfte der Auszubildenden*. Bei der weitergehenden Thematisierung wird der zusätzliche Nutzen, den die Kompetenzskala im Vergleich zur Ratingskala bringt, kritisch hinterfragt. R3 und R4 ziehen in Zweifel, um wie viel genauer eine Kompetenzskala denn nun eine Teilkompetenz erfassen kann und wägen diesen – geringen? – Zugewinn an Genauigkeit gegen die reduzierte Praxistauglichkeit ab:

R4: „Es ist halt auch, weil, wenn ich den Text hier mal so lese, ist tatsächlich eigentlich ja nur, es geht ja darum, wenn wir jetzt zum Beispiel die Aufgabenstellung nehmen, also weiter diesen Punkt zwei: Aufgabenstellung. Es gibt ja die Variante Aufgabenstellung nicht klar und dann eben die Schüler müssen nachfragen. So und das Ganze gibt es auf der Stufe: niedrig, mittel, ist teilweise klar, teilweise unklar, teilweise müssen sie nachfragen, teilweise nicht und hoch ist dann, sie müssen überhaupt nicht nachfragen, alles ist klar. Also, das ist ja immer/Und es geht ja eigentlich (R3: Ja, eben) bei allen Punkten (R3: Genau) also das ist halt, man liest sehr viel für relativ wenig (R3: Ja) Information, sag ich mal, was ja im Prinzip dahintersteckt (R3: Ja). Weil der Text ist ja auch immer gleich und wird tatsächlich immer nur in Nuancen verändert.“

R3: Das meinte ich eben auch, genau. Das im Grunde genommen da viel steht, obwohl da nur ganz kleine Unterschiede sind zwischen niedrig, mittel und hoch. Und das wird einfach nur in netter Weise nochmal eloquent bezogen auf, wie gesagt, bei z.B. zwei da Aufgabenstellung und das andere.“

Im weiteren Verlauf der Diskussion beziehen die Auszubildenden auch die dritte Skalenvariante der Bewertungsinstrumente ein, die Kompetenzeinschätzungen in Form einer Einordnung auf einer offenen Ratingskala zwischen einem positiven und negativen Extrempunkt vornimmt, die im Kontext der Diskussion mit

dem Begriff des „Kontinuums“ belegt wurde. Während mehrere Auszubildende sich zunächst sehr angeregt über die Vor- und Nachteile einer mehrstufigen Ratingskala im Vergleich zu dem „Kontinuum“ austauschen, endet die Diskussion in Einmütigkeit, als R1 ausführt, dass sie sich frage, ob sie „es relevant finde“, ob denn nun mit der einen oder anderen Variante versucht würde die Kompetenz zu erfassen:

I: „Was denkst du als Referendar? Welche Skala, glaubst du, fändest du besser?“

R4: Ich persönlich bin, ich bin ein Fan der Ratingskala, weil die im Prinzip das gleiche macht. Man sollte vielleicht dann, muss halt überlegen, ob eins, zwei, vier Stufen oder null bis vier, wären dann ja fünf Stufen, also Null, eins, zwei, drei, vier, zur Auswahl. Würde mir persönlich reichen, sind wir natürlich schon fast wieder im Notenbereich, aber trotzdem, dass ich sage: So, wir haben diese Stufen, und da ordne ich das ein. So, und das würde mir an Flexibilität auch auf dieser Skala durchaus reichen, weil/ Klar, hier hat man wesentlich mehr Flexibilität auch/

I: Hier ist jetzt dieses Kontinuum?

R5: Also hier ist, genau hier ist das Kontinuum, da hat man natürlich noch mehr Flexibilität, aber da finde ich, wird man dann, kommt man dann auch schnell in den Bereich der Schwammigkeit. Also man sollte sich dann schon, wenn man fünf Stufen hat, finde ich, sollte man sich schon zwischen diesen entscheiden können. Und da würde mir die Ratingskala tatsächlich eher zusagen.

R1: Finde ich auch (Nicken).

R3: Aber, wenn wir jetzt nachher davon ausgehen, dass wir im Referendariat, wie wir eben noch gesagt haben, als Lernprogression bewertet werden sollen. Finde ich das der Riesenvorteil/

I: Das ist das Kontinuum?

R3: Kontinuum, Entschuldigung, richtig. Dass man auf dem Kontinuum ja eintragen kann, ob der, ob ich in einem gewissen Bereich, wenn ich jetzt das Datum vom letzten UB habe und in das Datum von diesem UB eintragen muss, muss ich automatisch eintragen, ob ich mein Kreuzchen jetzt weiter links, also ist es jetzt noch mal schlechter geworden oder besser geworden, was wiederum heißen soll!

R4: Aber da hindert dich auch eine Ratingskala nicht dran. Ne Ratingskala ist genau auch in der Waagerechten, da hast du dann nur vorne eins, zwei, drei, vier, fünf drüber stehen und kannst aber auch genau dein, kannst theoretisch ja auch sagen: „Hier bin ich im Einserbereich.“ Wie gesagt, es ging ja dann nur darum im Einserbereich nochmal zwischen 1,3 und 1,9 zu differenzieren. Und ich finde da hört es dann wirklich auf.

R3: Nee, nee, ich meine genau das eben nicht, zwischen 1,3 und 1,9 oder so, ich meine dass man den jetzigen UB im Verhältnis zum letzten UB stellen muss, bei diesem Kontinuum.

R4: Aber genau das kannst du doch auch, ja aber das kannst bei der Ratingskala genau so tun.

R5: Aber es ist nicht so, so einfach. Weil wenn du jetzt z.B. sagst: «Ist immer noch schlecht. ABER EIN BISSCHEN BESSER!»

R3 (R3, R4, R5 durcheinander): Als das Letzte, als der letzte UB, richtig.

R5 (R3, R4, R5 durcheinander): Also ich würde, ich würde z.B./

R4 (R3, R4, R5 durcheinander): Dann hast du beim ersten Mal null gegeben und beim zweiten gibst du einen Punkt. [...]

I: R1 du hast gerade geschmunzelt bei der Diskussion?

R1: Ja, weil ich mich gefragt habe, ob das jetzt so relevant finde, ob das ein oder das andere.

R4: Ja, das geht mir genauso.

I: Kannst du das ein bisschen ausführen?

R1: Ja, ich habe mir gedacht, eigentlich reicht mir auch diese Skala, also ob ich jetzt/

I: Diese Skala ist? Dieses Kontinuum oder die Ratingskala?

R1: Äh die, nee die Ratingskala. Weil, ja ich weiß nicht, weil ich habe/ Gut, wenn man das jetzt wieder aus Ausbilersicht sieht, vielleicht ist, hilft dem das eher, wenn er diese Pfeile hat, aber bei mir ist das ja so, ich weiß ja eigentlich, wie das bewertet und ob die mir gesagt haben: «Ja das war jetzt schon etwas besser, als wir das beim letzten mal gesehen haben» oder so. Also weiß ich nicht.“

Bei einem Blick auf die Diskussion der Ausbilder erhält dieses, aus der Diskussion der Auszubildenden ableitbare Fazit, dass es erstens einen Konflikt zwischen der Erfassungsgenauigkeit einer Skala und der ‚Praxistauglichkeit‘ einen Konflikt gibt und zweitens die Art der genutzten Skala nur eine untergeordnete Rolle für den praktischen Wert eines Bewertungsinstrumente hat, weitere Bestätigung. Auch die Ausbilder scheinen die Erfassungsgenauigkeit der Kompetenzskala höher einzuschätzen als jene der Ratingskala, aber *mehr als die Hälfte der Ausbilder schlägt vor, die Kompetenzskala als Handreichung zur Erläuterung der Rating-skala zu nutzen und nicht als eigenes Erfassungsinstrument, da dies den Rahmen der Ausbildungskontexte sprengt, wie es z.B. A5 und A4 ausführen, die auf die Bedeutung einer transparenten Klärung der einzelnen Bewertungsstufen verweisen, aber zeitgleich auf die „Praktikabilität“ der Instrumente:*

A5: „Auf Ihre Frage noch mal zurückkommen, empfinde ich diesen kompetenzorientierten Bogen transparenter und objektiver. Sowohl für mich, als auch für die Referendare hinterher in der Nachbesprechung. Diese Skalierung (Anm.: Gemeint ist die Ratingskala, D. L.), ist dann ja doch wieder sehr individuell, sehr subjektiv. Wo mache ich jetzt mein

Kreuzchen? Dann müsste ich dann ja eigentlich auch noch mal ganz genau definieren (A4: Sich verständigen, A1, A3: nicken), wo ist ein Minus und wo ist ein Plus, wie definiere ich das Minus, wie definiere ich das Plus, wenn man das ganz genau gemeinsam mit den Referendaren definiert hat, dann ist das natürlich praktikabler, nur Minus und Plus zu machen.“

A4: „[...] Also insofern ist das hier für mich (verweist auf Kompetenzskala), der mit den ausformulierten dreistufigen Kompetenzstufen, da liegt, ist ein erster Schritt dazu zu kommen. In der Praktikabilität scheint mir das aber hoch komplex zu sein, dann ist eher so etwas, wie vorher schon einige (verweist auf Vorschläge von A6), worauf soll ich insbesondere achten, zu welchen Phasen wesentlich/[...]“

Aus der Diskussion der Ausbilder wird im weiteren Verlauf deutlich, dass der Zugewinn standardisierter und skalierteter Bewertungsinstrumente für die Ausbildungspraxis in erster Linie nicht eine möglichst große Steigerung der Objektivität und Transparenz der Erfassungsprozesse ist, sondern das Schaffen von ‚Gesprächsanlässen‘ zwischen den Akteuren, um im gemeinsamen Austausch eine Einschätzung des Unterrichts und der relevanten Qualitätskriterien zu erarbeiten. Skalierte Instrumente sind nicht „die ganze Wahrheit“, sondern eine „Hilfestellung“, wie es A4 formuliert:

A4: „[...] Und deshalb ist dann eine Reduktion und damit eine etwas klarere Verständigung, wohl wissend, dass das bitte nicht die gesamte Wahrheit ist, aber eine Hilfestellung ist, um klar zu kommen. So eine Reduktion in diese drei Stufen der Kompetenzskala, ein pragmatisch sinnvoller Weg.“

Ein weiteres Mal wird deutlich, welche große Rolle eine offene Kommunikation zwischen Ausbildern und Auszubildenden spielt, damit es zu gelingenden Bewertungssituationen kommen kann. Die Instrumente und damit auch die ihnen zu Grunde liegende Skalierung spielen dabei nur eine untergeordnete Rolle, sie sind nur ein ‚Arbeitsinstrument‘ wie es im Kontext des Nutzen der Standardisierung weiter oben formuliert worden ist und wie es *alle Ausbilder sehen*. Auf Grund der großen Komplexität und der unendlichen Vielfalt der Unterrichtskontexte

erscheint der Anspruch einer möglichst objektiven Messung der Unterrichtskompetenz ohnehin ebenso vergeblich wie deplatziert, denn für gelingende Bewertung gehe es gar nicht darum, „wie gut ich es messen kann“, sondern darum, „Kriterien gemeinsam“ zu entwickeln, wie A3 ausführt:

A3: „Ich würde mich ganz eindeutig für diese Kompetenzskala entscheiden und auf gar keinen Fall für das Andere. WEIL die Ratingskala für mich suggeriert: Wir könnten es messen. [...] Und es geht ja gar nicht darum, wie gut ich es kann, sondern es geht um Kriterien gemeinsam zu entwickeln für eine Kompetenz und da finde ich, ist das sehr hilfreich, die drei, diesen Unterschied zwischen niedrig, mittel, hoch, den kann auch ein Anfänger relativ gut leisten, das ist ja auch ein bisschen, was Sie (A6) sagen. Und deshalb würde ich immer sagen: Das reicht absolut aus, genau das ist praktikabel und eine Differenzierung mit zusätzlichen Anmerkungen, das kann dann kommen, aber von daher finde ich ist das absolut ausreichend, so was zu nehmen. Das hat, finde ich, viele Vorzüge. Ich käme nie auf die Idee das (Anm.: Gemeint ist die Ratingskala, D. L.) zu machen.“

Prägendes Konstruktionsprinzip von Bewertungskonzepten und -instrumenten muss es also sein, den Akteuren diese so wichtigen ‚Gesprächsanlässe‘ zu liefern, um ihre subjektiven Vorstellungen über guten Unterricht in eine tendenzielle Deckung zu überführen oder zumindest auf dem Weg dorthin eine gemeinsame Sprache für die Beobachtungen zu finden:

A4: „[...] Sprich, aber das Eigentliche ist dann, dass wir uns da verständigen, dass wir versuchen das zu deuten in Hinblick auf die Beobachtungen. Können wir uns auf gemeinsame Beobachtungen stützen und die damit verknüpfen und damit ein Gespür für diese Kompetenz, für dieses Beobachtungsmerkmal entwickeln, das wir an dieser Stelle eventuell bestimmen könnten?“

Hierbei gilt es aus der Perspektive *aller Ausbilder*, die Auszubildenden, im Sinne der ‚Modellbildung für einen heuristischen Zweck‘, sprachfähig zu machen, ihnen Fachvokabular und theoretische Modelle für das anzubieten, was sie zwar ‚fühlen‘, aber noch nicht verbalisieren können:

A1 (A1, A3, A4 durcheinander): „Den Referendarinnen und Referendaren hilft das manchmal weiter. Denjenigen zumindest, die erst mal nicht so recht wissen, worauf sie gucken sollen und wo sie (anschle!) können. Ich finde, denen hilft das auch weiter, nee also das, kann Orientierung geben genau (A6: nicken). Oder eine Sprachhilfe, absolut.

A3 (A1, A3, A4 durcheinander): Ja, wo sie (anschla?) sollen. Ja genau, genau, kann Orientierung geben. Ja genau, ganz genau.

A4 (A1, A3, A4 durcheinander): Ja oder eine Sprachhilfe sein, also wir haben ja gesagt, dass sie lernen müssen, ja zu reflektieren, dass sie vielleicht intuitiv was fühlen (A1, A3: «mhm» (bejahend)), dass sie es auch vielleicht beschreiben können im Hinblick auf eine konkrete Stunde und hier bekommen sie Vokabeln geliefert, Begriffe geliefert, Zusammenhänge, Konstrukte mit denen sie dann argumentieren können.“

Diese Bedeutung ‚Gesprächsanlässe‘ zu generieren erkennen auch die Auszubildenden in ihrer Diskussion, wenn z.B. R5 den Begriff des ‚Gesprächsanlasses‘ von sich aus als zentrales Kriterium für das Schaffen von Transparenz hochhält und ihr *mehr als die Hälfte* der anderen Auszubildenden beipflichten:

I: „Ich würde gerne noch mal als Frage in den Raum rein werfen: Glaubt ihr, das kann zu mehr Transparenz führen, wenn (...) eine Bewertung so abläuft, dass am Ende eines Unterrichtsbesuches sowohl ihr, als auch der Fachleiter auf so einer Skala ein Kreuz machen müsst.

R5: Ja.

R4: Ja, definitiv ja.

R5: Auf jeden Fall (nicken).

I: Warum?

R5: Weil man dann auch, also es schult ja auch die Reflexion nochmal.

Man sieht dann: Was habe ich gesehen, was hat der Fachleiter gesehen? Wo sind die Unterschiede, ohne dass man vorher jetzt vielleicht darüber groß diskutiert hat, und damit schafft man überhaupt erst die Gesprächsanlässe, wenn beide sagen, es war schlecht oder es war gut, dann muss nicht erst mal was ausloten, sondern man hat ganz schnell, ist man auf dem Punkt und sagt: Ok, hier sehen wir das unterschiedlich, woran liegt es oder hier sind wir beide schlecht, sagen wir beide, es war schlecht? Kann man direkt ins Problem einsteigen.“

Aus diesen Ausführungen lassen sich zusammenfassend die leitenden Konstruktionsprinzipien rekonstruieren, welche bei der Überarbeitung des Bewertungskonzepts von der pre-Alpha- zur Alpha-Version berücksichtigt werden sollten. Erstens muss eine *Klärung erfolgen, welche Funktion die Instrumente im Rahmen der Ausbildung haben*, um Klarheit über ihren spezifischen Beitrag für den Prozess der Professionalisierung zu generieren. Die Instrumente haben nämlich, wie es die Ausbilder formulieren, keinen Zweck in sich, sondern sind immer nur als ‚*Arbeitsinstrumente*‘ zu begreifen, deren Sinn sich in den konkreten Ausbildungssituationen entfaltet. Zweitens können *Skalen* aus der Perspektive beider Gruppierungen zu einer *Steigerung der Transparenz und Klärung der Bewertungsmaßstäbe* beitragen. Sie erhalten jedoch ihren maßgeblichen Sinn nicht in einer möglichst großen Objektivierung des Bewertungsprozesses, da dieser Anspruch vor dem Hintergrund der Ausführungen der Ausbilder ohnehin fragwürdig erscheint, sondern in der *Generierung von Gesprächsanlässen*. Das wesentliche Konstruktionsprinzip eines Bewertungskonzepts als Ganzem und der einzelnen Bewertungsinstrumenten, aus denen es sich zusammensetzt, muss diesen Dreh- und Angelpunkt gelingender Bewertungsprozesse im Blick haben, die sich maßgeblich dadurch auszeichnen, dass Ausbilder und Auszubildende ‚eine gemeinsame Sprache‘ finden, um sich über den gesehenen Unterricht auszutauschen und wechselseitiges Verständnis für die Interpretationen des jeweils anderen zu entwickeln.

4.4.2 Überarbeitung der Instrumente

Eine zentrale Rolle in der pre-Alpha-Version des Bewertungskonzepts nimmt der ‚*Beobachtungsbogen*‘ ein. Er sollte der Beobachtung des Unterrichts nach standardisierten Kriterien dienen und weitergehend die Nachbesprechung strukturieren, sowie im Rahmen der Sammlung über das Repertoire eine transparente Dokumentation des Kompetenzfortschritts ermöglichen. *Alle Auszubildenden* formulieren in ihrer Diskussion durchaus die Hoffnung, dass durch dieses Instrument einzelne dieser Ziele erreicht werden, sind aber vor dem Hintergrund der Komplexität des Instruments skeptisch, inwiefern es in der Praxis eingesetzt werden kann. Der bereits oben ausgeführte Konflikt zwischen der ‚*Praxistauglichkeit*‘ einerseits, die ein möglichst reduziertes und intuitiv verständliches Instrument benötigt, und dem Anspruch einer möglichst genauen Erfassung einzelner Teilaspekte guten Unterricht andererseits, der ein differenziertes und umfassendes Instrument erfordert, sorgt für eine große Spannung zwischen den zentralen Konstruktionsprinzipien, der ein gewinnbringender aber dennoch praxistauglicher Beobachtungsbogen begegnen muss. Zusammen mit den wenig ermutigenden Berichten der Ausbilder über den Einsatz von standardisierten Beobachtungsbögen, wie sie ebenfalls oben ausgeführt wurden, ergibt sich ein eher ernüchterndes Fazit bezüglich des Potentials standardisierter Beobachtungsbögen, die Bewertungsprozesse positiv zu beeinflussen. Zentrale Schlüsse hinsichtlich der Einsetzbarkeit von Beobachtungsbögen bietet auch die Rezeption der ‚*Matrix*‘. Diese hatte in der pre-Alpha-Version nicht den Stellenwert eines eigenen Bewertungsinstrumentes, sondern sollte vielmehr nur Grundlage eines Klärungsprozesses sein, welche der Qualitätsbereiche guten Unterrichts relevant für die Bewertungsprozesse in der Zweiten Phase der Lehrerbildung sind. Dieser anvisierte Klärungsprozess konnte in den Gruppendiskussionen nicht geleistet werden, weil v.a. die Ausbilder dieser Idee sehr skeptisch gegenüberstanden und die Frage, inwiefern die Matrix eine umfassende und vollständige Operationalisierung der Merkmale guten Unterrichts anbiete, auf deren Basis eine Auswahl der für die Ausbildung relevanten Kriterien erfolgen könne, negierten. Für A3 ist die Frage nach der Vollständigkeit der Operationalisierung von Grund auf falsch gestellt:

A3: „[...] Insofern meine ich, wäre die Frage gar nicht so sehr zu stellen: Ist die vollständig oder was ist wichtiger oder nicht? Das finde ich, spielt eigentlich eine völlig untergeordnete Rolle, sondern die Funktion zu benennen. Also Verständigung über Theoriehintergrund (A1: «mhm» (bejahend)) und Hilfe zur Orientierung und sprachlich und so weiter (A1, A6: «mhm» (bejahend)). Also das wäre entscheidend und dann ist eigentlich sekundär, was da steht. Hauptsache ist, es ist als Quelle benannt.“

Eine umfassende und möglichst vollständige Operationalisierung guten Unterrichts spielt aus der Perspektive *aller Ausbilder* keine große Rolle für gewinnbringende Bewertungsprozesse, und zwar deshalb, weil diese auf Grund der Eigenschaften von Unterricht sowieso von niemanden zu leisten sei, wie es auch schon im Rahmen der ‚*Grenzen der Operationalisierbarkeit*‘ weiter oben rekonstruiert wurde. Entsprechend ist aus der Perspektive der Ausbilder viel entscheidender, welche „Funktion“ die Matrix im Rahmen der Heuristik der Ausbildung hat, als dass die Frage nach ihrer Vollständigkeit entscheidend wäre, wie auch A6 ausführt:

A6: „Erschöpfend kann das nicht sein. Das ist von Natur aus, wie wir das alle schon beschrieben haben, mit diesem heuristischen Instrument nicht gegeben (A3: nicken) [...]“.

Entsprechend deuten die Ausführungen der Ausbilder darauf hin, dass die ursprünglich intendierte Funktion der Matrix zu verwerfen ist. Wenn schon in ihrem Rahmen keine erschöpfende Operationalisierung der Merkmale guten Unterrichts geleistet werden kann, dann macht es keinen Sinn, aus dieser Matrix die für die Bewertungsprozesse in der zweiten Phase relevanten Kriterien ableiten zu wollen. Vielmehr erhält sie aus der Perspektive der Ausbilder einen anderen Sinn. Sie dient der „Verständigung über einen Theoriehintergrund“, wie es A3 formuliert, bzw. der Darstellung der „Komplexität von Unterricht“, wie A 2 ergänzt:

A2: „Auch die Darstellung von Komplexität von Unterricht (A3: Ja richtig, genau) wird dadurch sichtbar und der Respekt vor qualitativ gutem Unterricht (A3: Ja genau). Wenn das der Anspruch, wenn das alles

Unterricht ausmacht, dann hat man doch einen ganz anderen Respekt vor Unterricht (A3, A4: Ja), weil da ja alles möglicherweise irgendwann mal eine Rolle spielen kann in der Stunde.

A6: Das ist ein guter Satz!“

Die ‚*Dokumentation der Nachbesprechung*‘ wird von beiden Gruppen als grundlegend sinnvoll begrüßt, wobei die Funktion, die diesem Instrument zugewiesen wird, zwischen den beiden Gruppen teilweise differiert. Für *alle Auszubildenden* ist die wesentliche Funktion einer Dokumentation der Nachbesprechung, die Transparenz der Bewertungsprozesse zu steigern, v.a. durch Bilanzierungen, die „schwarz auf weiß“ vorliegen und einzelne Lehrproben verbindlich und dauerhaft nachprüfbar bewerten, wie es R1 stellvertretend für *alle Auszubildenden* auf den Punkt bringt:

R1: „Vor allem auch, weil dann ja auch wirklich schwarz auf weiß, auch vor dem Fachleiter liegt, was er jetzt zu dem Referendar sagen kann (R4: «mhm» (bejahend)). Weil es ist ja, wie bei uns, wenn wir sonstige Mitarbeitsnoten geben, dann erinnert man sich an das eine und an das andere, aber an ganz viele Sachen gar nicht und deshalb habe ich mir auch gedacht, dass das wahrscheinlich, wenn man es gut führt, schon gerechter werden kann, weil die dann einfach eher sehen: Ok, da habe ich jetzt aber auch mal ein schlechtes Kreuz gemacht, obwohl ich doch so einen guten Eindruck von der habe und mir gedacht habe, vielleicht!“

Die Ausbilder heben in ihrer Diskussion wiederum hervor, dass die Klärung der Funktion eines Dokumentationsbogens entscheidend für seinen Nutzen sei, denn sonst „kann man ihn gar nicht verstehen“, wie es A3 formuliert:

A3: „Deshalb muss auch eine Dokumentation, es ist ähnlich wie bei dem Repertoire, muss, da muss die Funktion stehen, die dieser Bogen hat. Sonst kann man ihn auch gar nicht verstehen. Also eine Dokumentation um einer Dokumentation willen macht überhaupt keinen Sinn für mich, sondern ich muss wissen, (A4: «mhm» (bejahend)) was ich mit der Dokumentation will. Ist es Grundlage für den Selbstlernprozess? Ist es Grundlage für die Rechtfertigung meiner Bewertung am Ende? Das sind

ja zwei ganz unterschiedliche Funktionen (A6: Genau) und ich denke, man muss klären, welche Funktion sie hat, die kann auch wechseln! Aber Dokumentation in dem Sinne, was sind wesentliche Aspekte der Beratung gewesen und was machen wir damit, also das wäre auf jeden Fall sinnvoll, mit Funktion dieser Dokumentation.“

Laut A3, dem *mehr als die Hälfte der Ausbilder* beipflichten, kann eine ‚*Dokumentation der Nachbesprechung*‘ demnach verschiedene Funktionen erfüllen, wobei es den Anschein hat, dass sich diese Funktionen nicht gegenseitig ausschließen müssen, sondern dass die Funktion von ein- und demselben Instrument „auch wechseln“ kann. Hinsichtlich der Frage der „Rechtfertigung der Bewertung“, wie A3 es formuliert, zeigen sich deutliche Parallelen zu den Hoffnungen, welche die Auszubildenden an das Instrument richten. Für *alle Ausbilder* scheint die wesentliche Funktion die einer ‚*Dokumentation der Nachbesprechung*‘ zu sein. Ihr Wert liegt aber weniger in einer „Rechtfertigung der Bewertung“, als vielmehr in ihrer Rolle als „Grundlage für den Selbstlernprozess“ der Auszubildenden, damit der Prozess der Professionalisierung produktiv gestaltet werden kann, wie es neben A3 z.B. auch A4 explizit ausführt:

A4: „Wir haben ja jetzt bei uns im Seminar auch, versuchen wir auch die Nachbesprechung zu dokumentieren, also haben einen eigenen Bogen entwickelt. Und die Chance, die mit so einer Dokumentation verknüpft ist, ist dass ein Referendar an seiner Entwicklung weiterarbeiten kann.“

Hier zeigen sich wiederum die gruppenspezifischen Sensibilisierungen und Motive der einzelnen Akteure: Während es für die Ausbilder essentiell ist, den „Selbstlernprozess“ der Auszubildenden anzuregen, ist es für die Auszubildenden entscheidend, eine transparente und verbindliche Bewertung ihres Ausbilders zu erhalten, damit dieser sie möglichst „gerecht“ bewerten kann. Der Ansatz, die Kompetenz als Expertise über die Reflexionsfähigkeit zu erfassen, erfährt auf beiden Seiten ein überwiegend positives Echo. *Alle Auszubildenden, bis auf eine Person*, halten die Bewertung der Reflexionsfähigkeit auf Basis der gewählten offenen Ratingskalen sowohl für gewinnbringend für ihre eigene Professionalisierung, als auch praktisch gut umsetzbar:

R4: „[...] Aber grundsätzlich finde ich das, auch was die Kriterien angeht, also diese Selbstreflexion, also die Fähigkeit zur Selbstreflexion, was die Stunde angeht, als auch die Art und Weise, wie man das Statement hält und dann auch dieser Aspekt mit den Alternativen, finde ich, sind drei wichtige Punkte, also finde ich sehr gut gewählt, also finde ich, könnte ich mir vorstellen, dass das auch für Fachleiter schnell auszufüllen ist, relativ einfach und trotzdem eine ganz gute Übersicht bietet, um einen Überblick zu bekommen einfach (R5: Ja).“

R5: „[...] Ich persönlich finde das eben absolut wichtig (R4: «mhm» (bejahend)), weil ich denke, dass sich genau über diese drei Punkte definiert, egal unabhängig von meiner Einzelstunde, wie die gelaufen ist, ob ich eben in der Lage bin, ja selbstkritisch mit meinem Unterricht umzugehen und mich deswegen auch selbst, eigenständig zu verbessern. Und da finde ich auch gerade die Alternativenentwicklung absolut wichtig (R1: «mhm» (bejahend)) [...]“

Neben der Tatsache, dass die Art der Erfassung laut R4 hilft, „einen Überblick zu bekommen“, sieht R5 in der Erfassung der Kompetenz als Expertise über die Reflexionsfähigkeit und der Kompetenz als Repertoire über die Frage der Fähigkeit, Alternativen zu entwickeln, als entscheidende Kriterien, um zu bewerten, ob ein Auszubildender dazu in der Lage ist, „selbstkritisch mit seinem Unterricht“ umzugehen. In diesem Sinne schreibt R5 diesem Ausschnitt des Dokumentationsbogens die Fähigkeit zu, genau jenen Aspekt zu erfassen, der für die Ausbilder der Dreh- und Angelpunkt gelingender Ausbildung ist, nämlich die Frage nach dem Willen und der Fähigkeit zur Selbstreflexion auf Seite der Auszubildenden. R3 steht der Erfassung der Kompetenz als Repertoire skeptisch gegenüber, weil sie sich durch diese unter Druck gesetzt fühlt:

R3: „Aber das finde ich so, also das ist halt eher wieder, finde ich so, leistungsorientiert, abfragend halt. Kommt sie auf Alternativen? Dass ich da jetzt erst mal logischerweise aus meinen Fehlern heraus versuche, Alternativen zu finden kann ja auch gerne in der Tradition so bleiben, aber ich finde, es sollte halt nicht so krass in der Dokumentation als

beurteilendes Kriterium dastehen (R4: Ja aber I), damit es nicht wieder dahin geht, dass diese Nachbesprechung nicht so beratend ist, wie wir es uns oft im Referendariat gewünscht hätten, dass man viel lernt.“

Die anderen Auszubildenden teilen diese Kritik nicht, vielmehr betonen sie, dass es ab einem gewissen Zeitpunkt der Ausbildung auch die Pflicht der Referendare sein müsse, Alternativen für nicht gelungene Aspekte des Unterrichts aufzuzeigen. *Alle Ausbilder* teilen die positive Einschätzung der Mehrzahl der Referendare, was die Erfassung der Kompetenz als Expertise und als Repertoire angeht. So führt z.B. A3 in ähnlicher Art und Weise wie zuvor R5 die übergeordnete Bedeutung der Kompetenz als Expertise und als Repertoire für eine gelingende Ausbildung an:

„A3: „Ich glaube schon, dass es eine sinnvolle Idee ist (A5: Ja), weil es einfach den Unterschied deutlich macht. Also wenn ich mir darüber Gedanken mache: Kompetenz als Expertise oder Kompetenz als Repertoire, lerne ich ja relativ viel über Unterricht, also in so fern würde ich schon denken es ist, es macht Sinn (A4: mhm, gut). Das hat, finde ich schon, kann man drüber nachdenken.“

Hinsichtlich des Einsatzes in der Praxis betonen *alle Auszubildenden* die Problematik der ausbleibenden Nutzung der Instrumente und appellieren deswegen für Instrumente mit einem möglichst geringen Umfang. Die Dokumentation der einzelnen Aspekte, der in der Nachbesprechung thematisierten Kompetenzbereiche in Form einer Tabelle, wie es die pre-Alpha-Version vorschlägt, halten sie für „too much“, wie R4 es formuliert und R5 ihm beipflichtet, dass dies „in der Praxis nicht gemacht wird“:

R4: „Nee wirklich, du hast keine Lust mehr. Also ganz pragmatisch, du sitzt da und hast schon so viel geschrieben und gemacht und getan und du schreibst auch immer mit. Hier trägst du dann nichts mehr ein, glaube ich (R5: nicken). Also ich glaub hier ein, maximal eine Spalte, Rest sagst du: «Komm, ich weiß im Prinzip (wo es dran lag?), ich habe es jetzt hier auch drauf stehen, ich habe es mir auch angekreuzt». Ich glaube, das ist too much.“

R5: „Also ich finde es grundsätzlich, von der Hilfe her, fände ich es super, wenn diese Tabelle ausgefüllt würde (R4: «mhm» (bejahend)) und ich hinterher wirklich mehr als die Vereinbarungen, die jetzt getroffen werden sollen, sagen kann: «Aha, in dem Bereich Umgang mit Heterogenität, da wurde dies und das besprochen». Aber wirklich in der Praxis, denke ich auch, wird es nicht gemacht werden!“

Die ‚Sammlung über das Repertoire‘ wird von allen Auszubildenden grundlegend positiv bewertet. Sie erfährt aus der Auszubildendenperspektive ihren primären Sinn darin, einen Überblick über den Ausbildungsverlauf zu ermöglichen:

I: „So ein Dokument, das die Entwicklung im Zeitverlauf über die Unterrichtsbesuche dokumentiert, wie auch immer das jetzt aussehen soll, würde als gewinnbringend betrachtet werden?“

R3: Ja, auf jeden Fall (allgemeine Zustimmung).

R4: Definitiv, also!

R3: Und gab es bisher auf jeden Fall nicht. Also das genau, das war halt mein!

R4: Habe ich auch nie so was gesehen (R5: nicken), also das ist/Ich habe auch da wieder, man hatte auch noch nicht das Gefühl, dass die Seminarleiter immer so diese Entwicklung vor Augen haben, dass die genau wissen, aha, ok, das war letztes Mal, damals bei UB2 war es aber da und das Problem, jetzt ist sie da wirklich sicher, jetzt hat er aber dafür!“

Die Auszubildenden erhoffen sich durch das Repertoire eine verbindlichere und nachvollziehbarere Form der Bewertung, als sie es erlebt haben. Sie gehen davon aus, dass die Ausbilder nicht immer vor Augen hatten, welche Kompetenzfortschritte sie in ihrer Ausbildung erzielen konnten, weil diesen der Überblick über ihre Unterrichtsbesuche vermeintlich fehlte, und eine ‚Sammlung über das Repertoire‘ den Fachleitern einen besseren Überblick ermöglichen würde, wie es auch R2 und R5 ausführen:

R5: „[...] Ich finde das sehr schön, dass man einfach auch mal sieht, was wurde denn gemacht, was wurde eingesetzt (R1, R4: «mhm» (bejahend)), um darüber reden zu können.

R2: Also ich hätte jetzt gesagt, das brauche ich nicht, aber das weiß ich ja noch.

R5: Ja?

R2 und R5: Aber der Fachleiter nicht.“

Die Ausbilder sind in ihren Erwartungen an die ‚Sammlung über das Repertoire‘ anspruchsvoller. Allein einen Überblick über die verschiedenen Unterrichtsbesuche herzustellen ist für sie kein sinnvoller Zweck; „eine Dokumentation um einer Dokumentation willen macht überhaupt keinen Sinn“, wie A3 oben bereits zitiert wurde. Folglich sind sich *alle Ausbilder* darin einig, dass eine rein beschreibende Darstellung der Vielfalt des gesehenen Unterrichts keinen Sinn mache, schlechter Unterricht könne auch vielfältig schlecht sein, sodass „ein Repertoire als solches noch nichts aussagt“, wie A3 zusammenfasst:

A3: „Also ein funktionales, zielorientiertes Repertoire ist sinnvoll (A1: Ist sinnvoll, A4: Ja, A5: «mhm» (bejahend)), aber nicht ohne das (A1, A5: «mhm» (bejahend)). Ein Repertoire als solches sagt noch nichts aus.“

Ausgehend von der Annahme, dass es im Sinne der Professionalisierung ist, ein breites Repertoire an Inszenierungstechniken von Unterricht zu kennen und anwenden zu können, macht das Instrument aus der Perspektive der Ausbilder Sinn, aber nur vor dem Hintergrund der jeweils spezifischen Anforderungskontexte, die sich aus den konkreten Ausbildungssituationen heraus ergeben, wie die folgende Diskussionspassage zeigt:

A3: „Ok, das ist die Funktion die das da hat. Dann müsste das da hinter in meinen Augen: Grundlegende Einordnung des Unterrichts, im Sinne eines zu ERWERBENDEN Repertoires oder so was. Dann hat das für mich schon einen ganz anderen Sinn.“

A1: Dann wäre meine Frage aber auch gleichzeitig: Ist das dann gleichzeitig ein Qualitätsmerkmal?

A3: Ja, das ist die Frage (A6: «mhm» (bejahend))!

A1 (am Ende durcheinander mit A4 und A6): Also ich hab so einige Referendare vor Augen, die dann gerne auch experimentieren und dann das Angewandte auch zeigen. Haben sie (gehört?), aber eigentlich stehen sie noch auf einem ganz anderen Punkt, wo sie mit diesen vielfältigen Methoden, mit vielfältigen Medien meinetwegen in dem eigentlichen Unterricht dann selbst überfordert sind, und es einfach nicht gewinnbringend ist für Schülerinnen und Schüler (A6 nicken). Und ich auf diesem Repertoirebogen dann aber wunderbar stehen habe: «Hat dieses und jenes gemacht.»

A4 (A1, A4, A6 durcheinander): Hat alles gemacht.

A6 (A1, A4, A6 durcheinander): Gruppenpuzzle im Doppeldecker.

A1: Aber was sagt mir das über die Qualität?

A3: Nichts

[...]

A4: Ich kann mir beides vorstellen (A3: nicken), ich kann also, aber das ist dann wieder abhängig wirklich vom Subjekt der Ausbildung, sage ich mal. Also ich kann mir wunderbar/Also ich habe jetzt zwei vor Augen wo ich sage: «Bei der Einen könnte ich mir diesen Bogen, super! Alles ausgefüllt, alles da, perfekt!» Bei dem Anderen auch alles da, aber leider mangelhaft.»

Die Rezeption der ‚Sammlung über das Repertoire‘ auf Seiten der Ausbilder ist zwiespältig. Einerseits sprechen sie der Idee, Unterrichtskompetenz an Hand der Frage zu erfassen, ob ein Auszubildender über ein breites und, wie A3 es formuliert, „funktionales, zielorientiertes“ Repertoire verfügt, durchaus Sinn zu, andererseits werfen sie ein, dass einzelne Auszubildende „noch auf einem ganz anderen

Punkt stehen“, wie es A1 formuliert, und offensichtlich erst andere Professionalisierungsschritte geleistet werden müssen, bevor es Sinn macht, ein breites Repertoire an Inszenierungstechniken anzustreben. Entsprechend ist es Konsens unter *allen Ausbildern*, dass eine ‚*Sammlung über das Repertoire*‘, die eine sinnvolle Aussage über den Kompetenzerwerb von Auszubildenden abliefern will, zwingend eine qualitative Einschätzung vornehmen müsse. A4 verweist in diesem Kontext auf den Begriff der „Methodentiefe“ und stellt ihn jenem der „Methodenvielfalt“ gegenüber:

A4: „Ja ich möchte noch mal an Hilbert Meyer erinnern, der selber mit seinem Begriff der Methodenvielfalt eine Korrektur später verbunden hat, der Methodentiefe (A3: «mhm» (bejahend), A2: Ja, nicken). Ich kann wirklich Partnerarbeit so unterschiedlich gestalten wie nur irgendetwas, (A3, A1: nicken) und da das Repertoire zu beherrschen würde für mich eine hohe Methodenkompetenz bedeuten. D.h. ich könnte fünf Stunden sehen, Partnerarbeit und könnte trotzdem eine hohe Variabilität wahrnehmen. [...]“

In der Diskussion der Auszubildenden thematisiert lediglich R5 diese Problematik, stößt dabei aber in dasselbe Horn wie die Ausbilder und pocht auf die Berücksichtigung qualitativer Einschätzungen in einer ‚*Sammlung über das Repertoire*‘:

I: „[...] Also auch ein qualitatives Element?“

R5: Ja, genau, um halt einfach zu sehen auch, ok da wiederholt sich zwar eine Methode oder da wiederholt sich der Einsatz (R3: «mhm» (bejahend)) des Computers oder sonst was (R3: «mhm» (bejahend)), aber dazwischen ist ein Lernzuwachs passiert.“

Kritik üben die Ausbilder an den Einordnungskriterien, welche die pre-Alpha-Version der Sammlung über das Repertoire anbietet (Sozialformen, Medien, Unterrichtsdramatik), die aus der Perspektive der Ausbilder keinen Sinn machen. Größere Bedeutung für eine sinnvolle Bewertung des Erwerbs eines Repertoires im Kontext gelingender Ausbildung haben nach Meinung von *mehr als der Hälfte*

der Ausbilder die Frage nach der Vielfalt der Studententypen, die ein Auszubildender inszenieren kann, wie es etwa A6 ausführt:

A6: „[...] Und zwar, was die sog. Unterrichtstypen angeht, obwohl das ja was ist, was wir am Seminar immer mal wieder diskutieren, aber auch nicht abschließend diskutiert haben, gibt es ja Dinge wie eine Erarbeitungsstunde, denke ich mal (A3: «mhm» (bejahend)) und eine Einstiegsstunde. Das ist was, was ein Referendar im ersten Halbjahr sicher beherrscht (A1: Eine Einstiegsstunde?). Aber so was wie eine Abschluss- oder eine Evaluationsstunde empfinde ich schon als wirklich deutlich komplexer, aber auch eine Stunde im offenen Unterricht (A3: «mhm» (bejahend)). [...] Also für mich wäre ganz entscheidend, auch wenn wir derzeit noch keine abschließende Skala haben von Unterrichtstypen, wäre für mich das auf jeden Fall hier eine wichtige Ergänzung: Hat der/Welche Unterrichtstypen sind vorgekommen? Hat der z.B. eine Übungsstunde jemals gezeigt (A3: «mhm» (bejahend)) oder sind das alles Einstiegsstunden in neue Reihen?“

Darüber hinaus betonen *alle Ausbilder* die Bedeutung fachlicher Qualitätskriterien, die als Maßstab für eine qualitative Einschätzung einer Unterrichtsstunde herangezogen werden müssen, ohne dass diese festgelegt werden könnten. Vielmehr müsse eine ‚Sammlung über das Repertoire‘ als Bewertungsinstrument offen für fachspezifische Einschätzungen sein:

I: „[...] Ich entnehme jetzt den Äußerungen, wenn ich die zusammenfasse, dass die, wenn eine Sammlung über das Repertoire ein Instrument werden soll, das wesentlich offener gestaltet werden sollte (A3: Völlig richtig), als das, wie es hier vorgeschlagen ist. (A3, A5, A6; «mhm» (bejahend), nicken) und fachspezifischer?

A6: Fachspezifisch füllbarer (A3: Füllbarer, nicken).

I: Fachspezifisch füllbarer

A6: Komisches Wort, aber!

A3: Ja, ist so, genau

A1: Diese Spalte!

A2 (A2, A5 durcheinander): Man könnte auch zwei Spalten machen: allgemeine Aspekte in Beispielen und dann wieder (A3: fachspezifisch) speziell fachspezifische. Sind ja zwei verschiedene Fächer dann, damit man das noch mal trennt voneinander.“

Als weiteren Überarbeitungsvorschlag wirft A3 ein, dass es sinnvoll sei mit den Auszubildenden zu bilanzieren, welche Aspekte des Kompetenzfortschrittes aus einer konkreten Stunde herausgearbeitet werden können und diese als erhaltenswerte Professionalisierung zu benennen:

A3: „[...] Es ist eine besondere Leistung, wenn ich im Rahmen von Reflexion nach einer Unterrichtsstunde mit dem Referendar oder der Referendarin überlege: Was soll vom Repertoire, Ihrem Repertoire, was hat diese Stunde heute an Repertoireerweiterung gebracht? Und das kann sehr unterschiedlich sein (A4: Ja). Im fachlichen Bereich, also dass ich das benennen lasse, das ist ja auch eine Qualität. Das fände ich hilfreich.“

Zusammen mit den allgemeinen Konstruktionsprinzipien eines Bewertungskonzepts, wie sie im vorangegangenen Unterkapitel rekonstruiert wurden, bildet dieses Feedback beider Akteursgruppen das Fundament für die Überarbeitung des Bewertungskonzepts zur Alpha-Version, wie sie im folgenden Unterkapitel dokumentiert ist.

4.5 Von der pre-Alpha- zur Alpha-Version des Bewertungskonzepts

Im Zuge der Gruppendiskussionen kristallisierte sich heraus, dass einzelne Konstruktionsmerkmale und Bestandteile der pre-Alpha-Version des Bewertungskonzepts die erfolgreiche Erprobung in der Ausbildungspraxis behindern würden. Die Weiterentwicklung zur Alpha-Version dient dem Zweck, das Bewertungskonzept so zu überarbeiten, dass die absehbaren Probleme, die aus Unzulänglichkeiten im

Design der Instrumente resultieren, minimiert werden. Die Alpha-Version des Bewertungskonzepts ist im Anhang dokumentiert.

Entsprechend der wiederholt geäußerten Appelle der Ausbilder, die Funktion der einzelnen Instrumente im Kontext der Bewertungsprozesse, aber auch in der größeren Ausbildungslogik zu klären, wurde an den Beginn eines jeden Instruments eine *Erklärung zu der intendierten Funktion* von diesem gestellt.

Die für das Konzept einschneidendste Entscheidung ist der *Verzicht auf standardisierte Beobachtungsbögen*, da sich aus den Diskussionen beider Gruppen rekonstruieren ließ, dass ein erfolgreicher Einsatz von diesen mehr als fraglich ist. Ein wesentlicher Grund für diese Entscheidung ist der Konflikt zwischen der Notwendigkeit eines möglichst einfachen und intuitiv verständlichen Instruments, damit dieses im Sinne der *Praxistauglichkeit* auch tatsächlich eingesetzt wird einerseits, und der Notwendigkeit, ein komplexes und differenziertes Instrument, das den Ansprüchen eines standardisierten Erfassungsprozesses gerecht wird, andererseits. Die Berichte der Ausbilder weisen sehr deutlich darauf hin, dass vor dem Hintergrund der nur beschränkten menschlichen Ressourcen zur Beobachtung von Unterricht und der gleichzeitigen Dokumentation dieser Beobachtungen standardisierte Beobachtungsbögen keinen Mehrwert hinsichtlich der Genauigkeit der Erfassungsprozesse im Kontext formativer Evaluation bringen. Kurzum lautete das Fazit aus den Gruppendiskussionen: Ein Bogen, der einen differenzierten Erfassungsprozess ermöglichen würde, wäre nicht praxistauglich, ein praxistauglicher Bogen wäre derart unterkomplex, dass der Erfassungsprozess an sich mehr als fragwürdig erscheinen müsste.

An die Stelle des Beobachtungsbogens tritt die *Matrix der Qualitätsbereiche guten Unterrichts*, die in der Alpha-Version in den Rang eines Bewertungsinstruments gehoben wird, während ihr vorher nur eine Hilfestellung bei der Auswahl der für die Beobachtungsbögen relevanten Qualitätsbereiche zukam. Da im Rahmen der Diskussionen, v.a. in jener der Ausbilder, deutlich wurde, dass der Anspruch, eine umfassende und vollständige Operationalisierung guten Unterrichts zu erarbeiten, von den Akteuren zurückgewiesen wurde, ist dieses ursprüngliche Vorgehen fallengelassen worden. Stattdessen erhält die Matrix nun die Funktion,

Transparenz über das dem Bewertungskonzept zu Grunde liegende Verständnis guten Unterrichts herzustellen und weiterhin als loser Beobachtungs- und Planungsleitfaden zu fungieren, der den Ausbildern und Auszubildenden als gemeinsame Grundlage für die Beobachtung von und das Sprechen über Unterricht dienen soll. Weiterhin wurde die Sichtweise der Ausbilder aufgegriffen, dass es sich bei den definierten Qualitätsbereichen und operationalisierten Merkmalen guten Unterrichts um „Beispiele“ handelt und nicht etwa um eine erschöpfende Darstellung der Merkmale guten Unterrichts. Die Ausführungen in den jeweiligen Spalten der Matrix wurden passend umformuliert.

Der *Dokumentationsbogen für die Nachbesprechung* erfuhr weiterreichende Änderungen. An erster Stelle wurden die Einordnungskriterien des Unterrichts den Impulsen der Ausbilder angepasst und mit dem Vermerk versehen, dass die Einordnung hinsichtlich eines zu erwerbenden Repertoires geschehen soll. Diese Einordnungskriterien setzen sich nun zusammen aus der Zuordnung zu einem Stundentyp und den berücksichtigten fachlichen Lern- und Kompetenzbereichen. Die zuvor dort genannten Einordnungskriterien (Sozialformen, Medien, Unterrichtsdramatik) finden in einer offenen Kategorie der „sonstigen relevanten Aspekte“ Platz. Ergänzt wird die Einordnung im Sinne des zu erwerbenden Repertoires durch eine qualitative Einschätzung der Stunde, die an prägenden Merkmalen des Unterrichts zu begründen ist, wie es v.a. die Ausbilder als unerlässlich für eine sinnvolle Erhebung der Kompetenz als Repertoire gefordert haben. Da sich trotz ausgiebiger Literaturrecherche keine fächerübergreifend nutzbare Taxonomie unterschiedlicher Stundentypen finden ließ, wurde dies auf Basis bereits vorliegender Arbeiten des ZfsL formuliert und an das Bewertungskonzept angehängt. An die Stelle einer tabellarischen Übersicht der in der Nachbesprechung thematisierten Kompetenzbereiche treten in der Alpha-Version des Dokumentationsbogens drei Abschnitte, um die wichtigsten Bewertungs- und Beratungsaspekte der Nachbesprechung, getrennt nach denen, für die konkrete Stunde, „wesentlichen“ Qualitätsbereiche zu dokumentieren. Diese Anpassung entspricht dem Appell der Ausbilder und der Auszubildenden, eine situationsangepasste Fokussierung der Nachbesprechung auf die „wesentlichen Aspekte der Stunde“ vorzunehmen und das Design des Instruments auf diese Fokussierung

auszurichten. Die Dokumentation jedes relevanten Qualitätsbereiches schließt mit einer offenen Ratingskala, auf der im Sinne möglichst hoher Transparenz und Verbindlichkeit festgehalten werden soll, inwieweit der Auszubildende die Ansprüche des Qualitätsbereiches umgesetzt hat. Im Sinne der von den Ausbildern oft thematisierten Bedeutung der Ausbildungsprogression für die Bewertung, die für die Auszubildenden aber nur unklar zu erkennen war, sollen die Ausbilder die Bewertung auf der Skala einmal vor dem Hintergrund des aktuellen Ausbildungsstandes einschätzen („Ausbildungsstand“), aber darüber hinaus auch angeben, welche Bewertung sie im Sinne eines zu erreichenden „Examensstandards“ abgeben würden. Die Erfassung der Kompetenz als Expertise und der Kompetenz als Repertoire wird nach dem positiven Feedback aus den Gruppendiskussionen in die Alpha-Version übernommen. Die Skalen werden lediglich entsprechend der Idee, die Ausbildungsprogression in den Bewertungsskalen abzubilden, ebenfalls um den Hinweis in den Bewertungen einen „Ausbildungsstand“ und einen „Examensstandard“ zu berücksichtigen ergänzt. Ebenfalls in das Instrument eingefügt wurde die letzte Seite, die einer abschließenden Bilanzierung der Bewertung dient und einerseits jene Aspekte schriftlich fixieren soll, die im Sinne der Professionalisierung in ein dauerhaftes Handlungsrepertoire überführt werden sollten und andererseits jene Aspekte benennt, die einer weiteren Professionalisierung bedürfen.

Das Design der *Sammlung über das Repertoire* wurde den Änderungen des Dokumentationsbogens angepasst, sodass auch hier die geänderten Kriterien zur Einordnung des Unterrichts ihren Niederschlag fanden. Da der Beobachtungsbogen nicht länger Teil des Bewertungskonzepts war, fiel er entsprechend auch aus der Sammlung über das Repertoire heraus. Die Dokumentation der in den Unterrichtsbesuchen relevanten Qualitätsbereiche sowie deren realisierte Qualitätsausprägungen wird in der Alpha-Version nun über offene Ratingskalen erfasst, wie auch schon in dem Dokumentationsbogen für die Nachbesprechung. Die jeweils realisierten Bewertungen des Examensstandes sind aus den Dokumentationsbögen der Nachbesprechung in die jeweiligen Skalen der Sammlung über das Repertoire zu übertragen, sodass im Verlauf der Ausbildung ein differenziertes Bild entsteht, wann welche Qualitätsbereiche in welcher Qualitätsausprägung

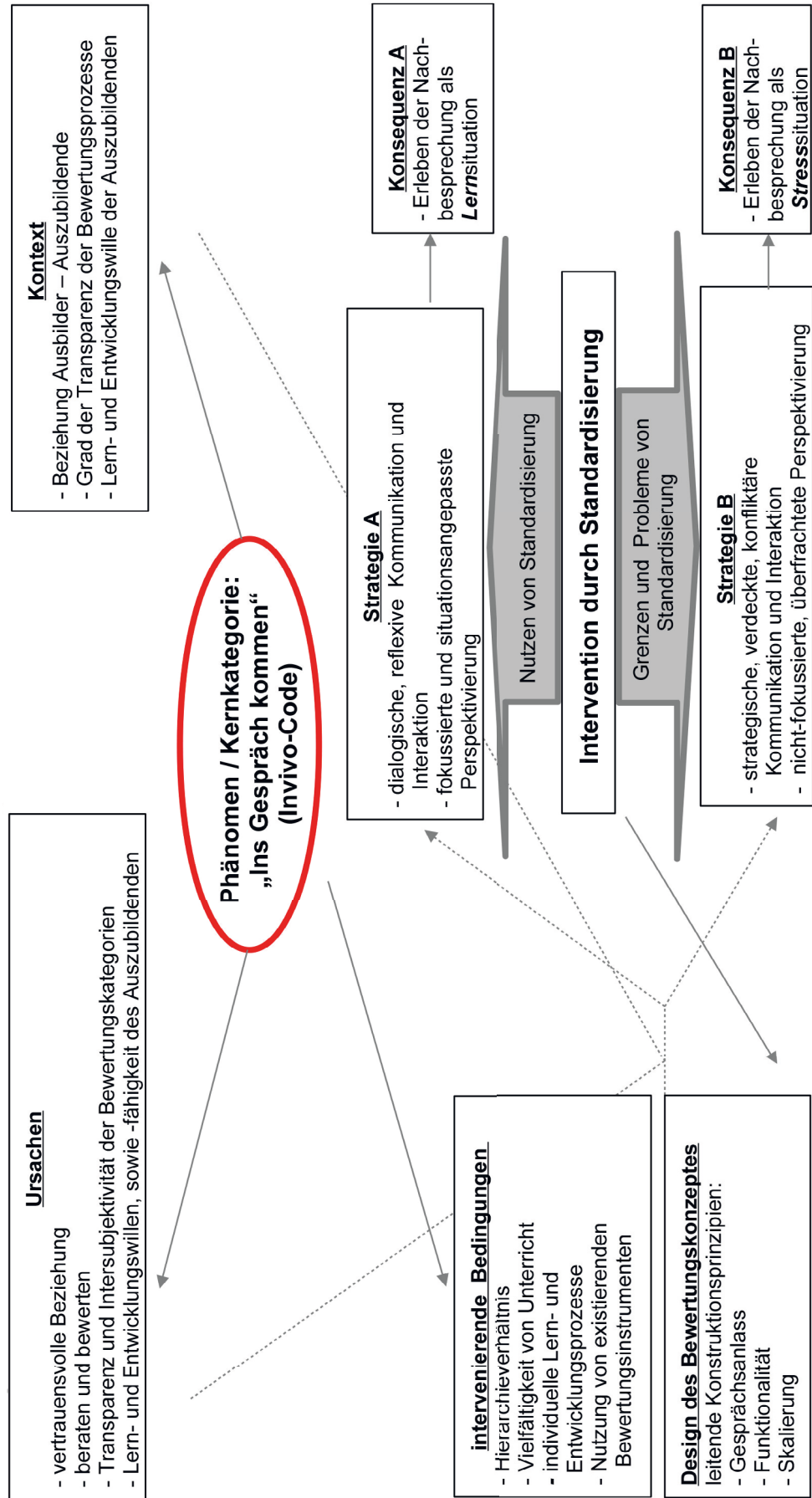
realisiert werden konnten. Die Auflistung der gezeigten positiven Aspekte einer allgemeinen Unterrichtskompetenz dient der schriftlichen Fixierung eines Kerns der professionellen Handlungskompetenz der Auszubildenden.

Ausgehend von diesen Änderungen des Bewertungskonzepts ist die Frage zu stellen, inwiefern überhaupt noch von einem standardisierten Bewertungskonzept die Rede sein kann, wenn zentrale Aspekte der Standardisierung der Entwicklung von der pre-Alpha- zur Alpha-Version zum Opfer gefallen sind. Weiter oben wurde der Standardisierungsanspruch des Bewertungskonzepts erstens an einer *Standardisierung der Bewertungskategorien* festgemacht, die durch die in Kapitel zwei geleistete Operationalisierung eines guten Unterrichts gesichert werden sollte. Da diese Operationalisierung in Form der Matrix nach wie vor Teil des Bewertungskonzepts ist, kann dem Konzept nach wie vor eine Standardisierung der Bewertungskategorien zugesprochen werden. Der weitergehende Anspruch, eine *Standardisierung des Bewertungsvorgangs* zu ermöglichen, muss deutlich zurückgenommen werden. Durch den Wegfall der standardisierten Beobachtungsbögen ist die Vergleichbarkeit der Beobachtungs- und Erhebungsprozesse nicht mehr gewährleistet. Erhalten bleibt allerdings die Absicht, eine möglichst einheitliche Struktur der Nachbesprechung zu gewährleisten dadurch, dass der Dokumentationsbogen bestimmte zentrale Aspekte in der Nachbesprechung vorsieht (Einordnung des Unterrichts, Festlegung auf relevante Qualitätsbereiche, Einschätzung des Reflexionsfähigkeit). Letztendlich kann demnach durchaus noch von einem standardisierten Bewertungskonzept gesprochen werden, da es bestimmte Standards eines guten Unterrichts definiert und versucht zentrale Teilkompetenzen einer allgemeinen Unterrichtskompetenz (Kompetenz als Expertise/als Repertoire) auf einheitliche Art und Weise zu erfassen.

4.6 Abschließende Interpretation der Gruppendiskussionen – axiales Codieren

Neben der eher pragmatisch orientierten Überarbeitung der pre-Alpha-Version des Bewertungskonzepts sollen die Erkenntnisse bezüglich des Einflusses der Standardisierung auf die Güte der Bewertungsprozesse, die sich aus der Auswertung der Gruppendiskussion generieren ließen, hier abschließend zusammenfassend dargestellt werden. Dem *axialen Codieren* im Sinne der Grounded Theory folgend (vgl. Strübing 2008: 26–20) werden hier Querverbindungen zwischen den verschiedenen Auswertungskategorien hergestellt. Hierbei werden diese in einen Ursache-Wirkungszusammenhang eingeordnet und, wie es die Grounded Theory nennt, auf eine *Kernkategorie* zugespitzt, die als zentraler Aspekt des hier untersuchten sozialen Phänomenbereichs verstanden werden kann. Die Kernkategorie der Bewertungsprozesse im Rahmen der zweiten Phase der Lehrerbildung bezieht sich auf Frage, ob es den Akteuren im Rahmen der Unterrichtsnachbesprechung gelingt, über den gesehenen Unterricht „*ins Gespräch zu kommen*“. Die Bedeutung des offenen, kommunikativen Austausches zwischen den einzelnen Akteuren für gewinnbringende Bewertungs- und Beratungsprozesse tauchte im Verlauf der Gruppendiskussionen beider Akteursgruppen immer wieder auf. Das Erkenntnispotential dieser zunächst vollkommen banal anmutenden Feststellung zeigt sich durch das axiale Codieren, in dessen Rahmen die besonderen Ursachen herausgearbeitet werden können, die ein „ins Gespräch kommen“ erst ermöglichen. Die Abbildung 4.8, die sich in der Darstellung an den Vorschlag von Strübing anlehnt (vgl. ders. 2008: 26–30), zeigt das axiale Codieren der Kernkategorie:

Abbildung 4.8: Axiales Codieren der Gruppendiskussionen



Im Zentrum des axialen Codierens steht die *Kernkategorie*, das *zentrale soziale Phänomen* der Bewertungsprozesse in der zweiten Phase der Lehrerbildung, das „ins Gespräch kommen“ *zwischen Ausbildern und Auszubildenden*. Diese Kernkategorie ist ein in-vivo Code, d.h. die Bezeichnung dieser Kategorie ist wörtlich aus den Gruppendiskussionen entnommen worden, und in diesem Wortlaut in beiden Diskussionen an mehreren Stellen genutzt worden.

Der im rechten oberen Kasten dargestellte *Kontext* umreißt die sozialen Rahmenbedingungen, innerhalb derer sich das Phänomen des „ins Gespräch Kommens“ entfaltet. Wie in der Auswertung der Gruppendiskussionen deutlich geworden sein sollte, spielen in diesem *Kontext* die *Beziehungen zwischen den Ausbildern und Auszubildenden* eine entscheidende Rolle, ebenso wie der *Grad an Transparenz der Bewertungsprozesse*. Weiterhin ist auf Seite der Auszubildenden ein *Lern- und Entwicklungswillen* notwendig, damit Ausbilder und Auszubildende tatsächlich „ins Gespräch kommen“. Dieser *Kontext* ist zunächst nicht wertend zu verstehen. Er schließt sowohl jene Ausprägungen der Interaktionssysteme ein, welche das „ins Gespräch Kommen“ verhindern, etwa fehlendes Interesse an professioneller Entwicklung auf Seiten der Auszubildenden, als auch jene welche dieses ermöglichen. Insofern umfasst der *Kontext* des „ins Gesprächs Kommens“ eine Vielfalt konkreter Ausprägungsgrade, die im Sinne gelingender Ausbildungsprozesse sowohl positiv als auch negativ bewertet werden können.

Damit die Akteure in diesem abstrakteren *Kontext* tatsächlich „ins Gespräch kommen“, müssen konkrete *Ursachen* gegeben sein, d.h. die prinzipielle Offenheit des *Kontextes* engt sich hinsichtlich jener Ausprägungsgrade ein, die erfolgreiche Bewertungsprozesse charakterisieren. Die Eigenschaften dieser *Ursachen* sind im Kasten links oben aufgeführt. Die Beziehungen zwischen Ausbildern und Auszubildenden müssen einen gewissen Grad an *Vertrauen* zwischen beiden Seiten ermöglichen. Der Ausbilder muss es schaffen ein Mindestmaß an *Transparenz und intersubjektiver Nachvollziehbarkeit seiner* Bewertungen zu gewährleisten und sich seiner Doppelrolle als *Berater und Bewerter* bewusst sein, sowie diese Rollen situationsangepasst ausfüllen. Ebenso muss der konkrete Auszubildende die *Bereitschaft und Fähigkeit* haben, sich im Rahmen eines *Professionalisierungsprozesses* zu entwickeln.

Intervenierende Bedingungen, die im Kasten mittig links dargestellt sind, verhindern, dass diese *Ursachen* ihre Wirksamkeit entfalten können. Dominiert der Einfluss dieser *intervenierenden Bedingungen*, werden Ausprägungsgrade des *Kontextes* realisiert, die ein „ins Gespräch Kommen“ verhindern. Die *Hierarchieverhältnisse* zwischen den Akteuren können hierbei offene Kommunikation verhindern und für kontraproduktive Interaktionsmuster sorgen. Die *unendliche Vielfalt des Unterrichts* stellt den Ausbilder vor die große Aufgabe, den Unterricht nicht nur vor dem Hintergrund festgelegter Qualitätskriterien zu bewerten, sondern in seinen Bewertungen auch die individuellen Kontextbedingungen, in denen sich der Unterricht entfaltet, sowie den *individuellen Lern- und Entwicklungswegen* des Auszubildenden gerecht zu werden. Die *ausbleibende Nutzung von bestehenden Bewertungsinstrumenten*, aus welchen Gründen auch immer, sorgt für Irritation und Verunsicherung auf Seiten des Auszubildenden und verhindert, dass dieser mit dem Ausbilder tatsächlich „ins Gespräch kommt“.

Aus dem Zusammenwirken der *Ursachen* und *intervenierenden Bedingungen* des „ins Gesprächs Kommens“, entstehen die verschiedenen *Strategien der Akteure*, die sich idealtypisch in der Dichotomie von zwei Strategien darstellen lassen. *Dominieren die Ursachen* der Kernkategorie ein Interaktionssystem, ist die Folge die *Strategie A*, die sich durch eine dialogische, reflexive Kommunikation und eine fokussierte und situationsangepasste Perspektivierung der Bewertungs- und Beratungsgespräche auszeichnet. Die Akteure kommen „ins Gespräch“, sodass als *Konsequenz* ein soziales Interaktionssystem entsteht, das sich im Sinne der oben geleisteten Charakterisierung als ‚*Lernsituation*‘ beschreiben lässt. Auf der anderen Seite sorgt eine *Dominanz der intervenierenden Bedingungen* für eine Realisierung der Strategie B, in deren Rahmen die Akteure einer strategischen, verdeckten und konfliktären Form der Kommunikation und Interaktion folgen. Die Perspektivierung auf den Unterricht ist nicht fokussiert und tendenziell überfrachtet, sodass keine Professionalisierungsschritte zurückgelegt werden können, und als *Konsequenz* ein soziales Interaktionssystem entsteht, das als ‚*Stressituation*‘ zu bezeichnen ist.

An dieser Stelle interveniert die *Standardisierung* in die Interaktionssysteme: Sie kann die soziale Situation der Unterrichtsnachbesprechung sowohl in Richtung einer Realisierung der Strategie A, als auch in Richtung der Strategie B beeinflussen. Wird die *Standardisierung als ein Arbeitsinstrument* verstanden und an die Bedürfnisse der Ausbildungssituationen angepasst, dann hilft sie, Lernsituationen zu erzeugen, in dem sie die Transparenz der Bewertungsprozesse erhöht, Orientierung gibt und eine gemeinsame Gesprächsgrundlage für Ausbilder und Auszubildende liefert. Eine *Verdinglichung der Standardisierung zum Selbstzweck* sorgt für gegenteilige Effekte; sie wird den konkreten Ausbildungssituationen und -bedürfnissen nicht gerecht. Sie ordnet singuläre, so nie wieder herstellbare unterrichtliche Kontexte einem abstrakten Qualitätsverständnis unter, welches für die konkrete Situation unangemessen sein kann und tendiert dazu, die Beobachtungsfähigkeit von Ausbildern zu überfordern. Folglich muss ein standardisiertes Bewertungskonzept, das die Akteure darin unterstützen will, im Sinne der Strategie A zu interagieren, diese Aspekte in Form der *leitenden Konstruktionsprinzipien* berücksichtigen. Diese besagen, dass die Funktion bzw. zumindest die *Intention von Bewertungsinstrumenten* innerhalb eines Bewertungskonzepts geklärt sein muss. Zentrales Konstruktionskriterium von Bewertungsinstrumenten ist die Frage, ob sie die Akteure dazu animieren, über den gesehenen Unterricht miteinander „ins Gespräch zu kommen“ und sich über die jeweiligen subjektiv gefärbten Deutungen auf Basis einer gemeinsamen (Fach-)Sprache auszutauschen. Schlussendlich müssen Bewertungskonzepte berücksichtigen, dass *Skalierungen* gefunden werden müssen, welche die im Rahmen des Gesprächs getroffenen Aussagen verbindlich schriftlich fixieren. Die Überarbeitung des Bewertungskonzepts zur Alpha-Version versucht, diesen Ansprüchen gerecht zu werden und so die Realisierung der Kernkategorie zu unterstützen. Die Erprobung des Bewertungskonzepts in der Ausbildungspraxis und die anschließenden Interviews mit den beteiligten Akteuren dienen der Überprüfung dieser These, wie sie im folgenden Kapitel dokumentiert wird.

5 | Auswertung der Interviews

Die Interviews haben im Kontext des Forschungsprojektes v.a. die Funktion, den Einsatz des Bewertungskonzepts zu evaluieren und damit ein Fazit bilden zu können, inwiefern das Bewertungskonzept die in Kapitel eins formulierten Zielsetzungen erfüllen kann und an welchen Stellen es ggf. Anpassungen geben sollte. Die Analyse der Interviews strukturiert sich dabei nach den Unterrichtsfächern, in denen die Erprobung stattgefunden hat. Hierbei werden für jeden Erprobungskontext die Einschätzungen der drei Auszubildenden, die sich an der Erprobung beteiligt haben, den Einschätzungen des jeweiligen Ausbilders gegenübergestellt. Die Auswertung der Interviews schließt mit einer Darstellung der aus den Interviews rekonstruierbaren Überarbeitungsvorschlägen des Konzepts und der damit in Verbindung stehenden Weiterentwicklung des Konzepts zur Beta-Version ab. Die Dokumentation der Ergebnisse beschränkt sich auf die dritte und damit letzte Version des hierarchischen Kategoriensystems (Gamma-Version).

5.1 Das hierarchische Kategoriensystem der Interviews

Abbildung 5.1 zeigt die dritte Version des hierarchischen Kategoriensystems, das sich aus der Analyse der Interviews ergeben hat, ohne die Subkategorien der beiden Hauptkategorien.

Abbildung 5.1: Hierarchisches Kategoriensystem des Interviews (Gamma-Version)



Konzept Einsatz Ausbilder



Konzeptbewertung

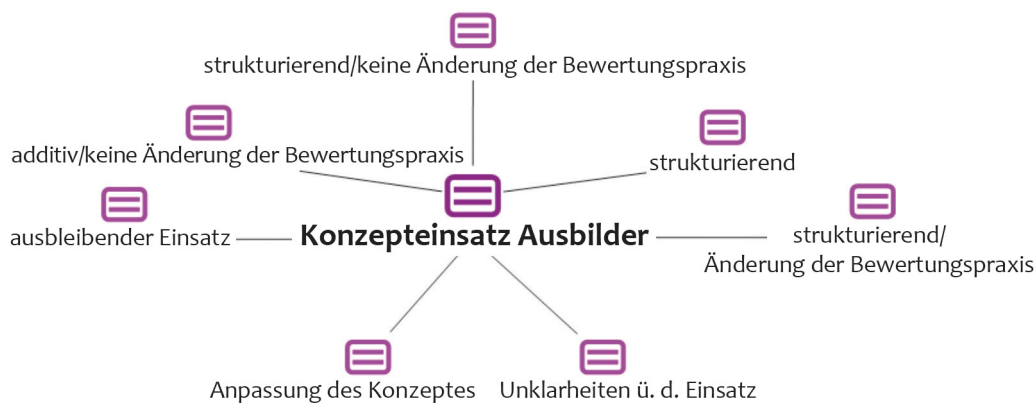
Die Kategorie des ‚*Konzept Einsatz der Ausbilder*‘ erfasst diese verschiedenen Muster der Konzepterprobung aus der Perspektive der beteiligten Akteure. Diese verschiedenen Einsatzmuster des Konzepts haben, wie die Auswertung zeigen

wird, maßgeblichen Einfluss auf die Wirkungen, die das Bewertungskonzept entfaltet. Die bedeutendste Kategorie für den Zweck der Interviews ist jene der ‚*Konzeptbewertung*‘, in der die wertenden Einschätzungen der Akteure hinsichtlich der Zielsetzungen des Bewertungskonzept strukturiert zusammengefasst sind. Innerhalb der Auswertung dieser Kategorien werden immer wieder Bezüge zur Auswertung der Gruppendiskussionen, sowie der dort erarbeiteten Erkenntnisse hergestellt. Die Auswertung der Interviews beginnt mit der Darstellung der verschiedenen Muster des ‚*Konzeptesinsatzes der Ausbilder*‘.

5.2 Konzeptesinsatz der Ausbilder

Die Kategorie, die in Abbildung 5.2 mit ihren Subkategorien dargestellt ist, nimmt zunächst keine qualitative Einschätzung der unterschiedlichen Einsatzvarianten des Bewertungskonzept vor, sondern unterscheidet nur die unterschiedlichen Muster, mit denen die Ausbilder das Konzept in der Praxis erprobt haben. Hierbei beschränkt sich die Auswertung auf eine deskriptive Darstellung der Berichte der Nutzungsmuster und ordnet diese nur in jene Subkategorien ein, in die sie nach den Berichten sicher einzuordnen sind.

Abbildung 5.2: Subkategorien der Kategorie ‚Konzeptesinsatz Ausbilder‘



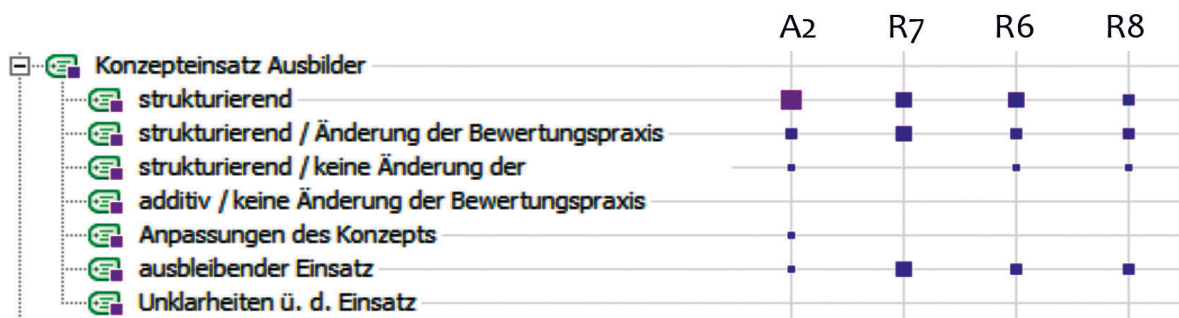
So wurde die Subkategorie eines Einsatzes der ‚*strukturiierend mit einer Änderung der Bewertungspraxis*‘ ist, nur dann codiert, wenn die Akteure explizit von

einer Änderung der Bewertungspraxis berichtet haben, die durch den Einsatz des Bewertungskonzepts ausgelöst wurde. Hat das Konzept zwar die Unterrichtsnachbesprechung oder die Bewertungsprozesse strukturiert, es wird aber nicht deutlich, ob diese Bewertungspraxis auch ohne das Konzept schon so durchgeführt wurde, dann wurde der Bericht lediglich als ‚*strukturierend*‘ eingestuft. Gleichermaßen ließen sich Berichte darüber rekonstruieren, dass das Konzept zwar ‚*strukturierend*‘ eingesetzt wurde, aber dies explizit ‚*keine Änderung der Bewertungspraxis beinhaltet*‘, da die Bewertungspraxis auch vor Einsatz des Konzepts auf die gleiche Art und Weise ablief. Weiterhin berichteten die Akteure über eine ‚*additive*‘ Nutzung des Konzepts, die ebenfalls ‚*keine Änderung der Bewertungspraxis*‘ mit sich brachte, sondern sich dadurch auszeichnete, dass die Ausbilder das Konzept zwar einsetzen, aber erst nachdem sie ihre Nachbesprechung wie gewohnt durchgeführt hatten und auf diese Weise bereits zu einer Bewertung gelangten, bevor das Konzept eingesetzt wurde. Über einen gänzlich ‚*ausbleibenden Einsatz*‘ des Bewertungskonzepts wurde ebenfalls an zahlreichen Stellen berichtet. Ebenso bedeutsam für die Entwicklung des Konzepts sind die Berichte über ‚*Anpassungen des Konzepts*‘, die von den Ausbildern vorgenommen wurden, sowie ‚*Unklarheiten über den Einsatz*‘, die Hindernisse für eine erfolgreiche Erprobung darstellten. Auffällig ist, dass sich die Berichte über die Erprobung des Konzepts bei *allen Ausbildern* in verschiedene Subkategorien des Konzepteinsatzes einordnen ließen, was damit zusammenhängt, dass sie das Konzept z.B. nur in einzelnen Unterrichtsbesuchen eingesetzt haben, über die dann Berichte hinsichtlich eines ‚*strukturierenden*‘ oder ‚*additiven*‘ Einsatzes generieren ließen, wohingegen der ‚*ausbleibende Einsatz*‘ in anderen Unterrichtsbesuchen auch als solcher codiert wurde. Die folgende Auswertung begibt sich demnach auf die Suche nach dem *dominanten Muster des Einsatzes* für jeden Ausbilder, um daraus allgemein Schlüsse hinsichtlich der Entwicklung und Implementation eines Bewertungskonzepts abzuleiten. Hierbei werden die Perspektiven der Ausbilder und der Auszubildenden zunächst mit Hilfe der Code-Matrix nach den Häufigkeitsverteilungen der verschiedenen Subkategorien gegenübergestellt, um darauf aufbauend interpretativ ausgewertet zu werden.

5.2.1 Einsatz bei A2 (Ausbilder 2)

Abbildung 5.3 zeigt die Code-Matrix des Konzepteinsatzes von A2, wobei die Häufigkeitsverteilungen *spaltenweise* dargestellt sind, d.h. je größer ein Symbol ist, desto häufiger wurde diese Subkategorie in der Auswertung der einzelnen Interviews codiert, ergo dominiert aus der Perspektive des jeweiligen Interviewten das entsprechende Nutzungsmuster den Einsatz. Aus der Code-Matrix wird deutlich, dass sowohl Ausbilder als auch Auszubildende den Einsatz des Bewertungskonzepts ähnlich einschätzen. Alle Akteure sehen den ‚strukturierenden‘ Einsatz als dominantes Muster und konstatieren A2 in der Erprobung des Konzepts durchaus eine ‚Änderung der Bewertungspraxis‘.

Abbildung 5.3: Code-Matrix der Konzepteinsatzes bei A2



Ebenso berichten alle Akteure über einen teils ‚ausbleibenden Einsatz‘ des Bewertungskonzepts, bei R7 und R8 nehmen diese Berichte einen durchaus erheblichen Teil ein, da die Codierhäufigkeit dieser Subkategorie zumindest 50% der Häufigkeiten eines ‚strukturierenden‘ Einsatzes ausmacht, wenn man den rein ‚strukturierenden‘ Einsatz und einen ‚strukturierenden Einsatz mit Änderungen der Bewertungspraxis‘ gemeinsam betrachtet.

Aus den einzelnen codierten Interviewpassagen dieser Kategorie wird deutlich, dass A2 das Bewertungskonzept gemäß der formulierten Intention eingesetzt hat und die Auszubildenden diesen Einsatz so erkennen konnten. In der folgenden Passage berichtet der Ausbilder auf Nachfrage über den Einsatz des Dokumentationsbogens für die Nachbesprechung und bejaht explizit die Frage nach einer Veränderung der Bewertungspraxis, da das Konzept ihm bei der „Strukturierung des Nachgesprächs“ geholfen habe:

I: „[...] Ob die Art wie Sie die Nachbesprechung geführt haben, ob die sich dadurch verändert hat?

A2: Ja schon, ich bin stärker verwiesen wirklich eine klare Struktur einzuhalten. Durch die Reduktion auf die Qualitätsbereiche wird man selber, was wir ja von den Referendaren im Unterricht erwarten, eine klare Struktur, wird selber noch mal stärker noch mal dran erinnert, wirklich/ Man hat ja nur eine Stunde Zeit oder anderthalb Stunden maximal, wirklich sich zu konzentrieren auf die wesentlichen Punkte. Und auch innerhalb der Operationalisierung, der Beispiele in diesen Bereichen drin zu bleiben. Also es hilft einem selber auch noch mal durch die Unterpunkte und durch die Aufschlüsselung selber stärker nochmal, dass man sich nicht verliert in verschiedenen Qualitätsbereichen, sondern dass man wirklich die Konzentration auch dabei lässt dann bei drei oder vier Qualitätsbereichen. Es hilft einem selber bei der Strukturierung des Nachgesprächs.“

Auch aus den codierten Textstellen bei den Auszubildenden zeigt sich der strukturierende Einsatz des Bewertungskonzepts, für den A2 gesorgt hat. R8 berichtet hierbei von den ohnehin schon produktiven Bewertungsprozessen und Unterrichtsnachbesprechungen von A2, die durch den Einsatz des Bewertungskonzepts aber „vielleicht noch ein bisschen nachhaltiger“ wurden, insofern sich also durchaus eine, wenn auch nur kleine, ‚Änderung der Bewertungspraxis‘ feststellen ließ:

I: „[...] Konnten Sie Unterschiede feststellen in den Lehrproben, in denen der Ausbilder dieses Konzept angewandt hat und in denen der gleiche Ausbilder dieses Konzept nicht angewandt hat?

R8: Also ich muss sagen, dass der Ausbilder an sich generell immer sehr intensiv diese Nachbesprechungen auch wahrgenommen hat, um uns wirklich auch zu helfen, uns weiter auch zu fördern in bestimmten Bereichen. Und dass ihm das immer sehr wichtig war, da auch sehr genau drauf einzugehen, jetzt auch unabhängig davon, ob das Konzept dabei war oder nicht, aber im Nachhinein ist halt dieses Konzept, wenn man es dann nachher noch mal bekommt und sich dann noch mal durchliest,

was er wirklich alles aufgeschrieben hat, nochmal vielleicht ein bisschen nachhaltiger, weil es einfach länger in Erinnerung bleibt und auch verschriftlicht ist. Sicher macht man sich auch selbst Notizen, auch zu dem was an Anregungen kommt, wie man sich weiterentwickeln kann, aber dann bekommt man es auch noch mal konkret, schwarz auf weiß, und es ist halt ein Unterschied, was man raus hört und was man vielleicht nachher auch dann liest, was er aufgeschrieben hat.“

Auch R7 schließt sich dieser Einschätzung an und meint während der Erprobung des Bewertungskonzepts auf Seiten des Ausbilders eine zunehmende Sicherheit bei dem Einsatz des Bewertungskonzepts entdeckt zu haben:

I: „Hatten Sie den Eindruck, dass dieser Dokumentationsbogen die Nachbesprechung strukturiert hat oder hatten Sie den Eindruck, der Ausbilder fährt sein Programm und macht dann additiv am Ende nochmal dieses Dokument?“

R7: Nee, der Fachleiter hat sich schon sehr stark an diesem Bogen orientiert, was ihm anfangs auch, glaube ich, schwer fiel, aber mit der gewonnenen Erfahrung hat er sich besser einarbeiten können und es ging dann auch, die Nachbesprechung vollzog sich schneller und er hat sich dann schon doch sehr an den Ablauf des Bogens gehalten.“

Offensichtlich hat v.a. der Dokumentationsbogen für die Nachbesprechung die Bewertungspraxis von A2 verändert, wohingegen andere Instrumente des Bewertungskonzepts nicht oder nur teilweise eingesetzt wurden, wodurch sich die codierten Interviewpassagen eines ‚ausbleibenden Einsatzes‘ erklären lassen. Alle Auszubildenden haben vom Ausbilder am Ende der Ausbildung nur einen Teil der Sammlung über das Repertoire erhalten, nämlich die Übersichtstabelle von Seite zwei des Instruments. Die Abbildung der Kompetenzentwicklung in den verschiedenen Qualitätsbereichen des Unterrichts, sowie die Entwicklung der Kompetenz als Expertise und der Kompetenz als Repertoire, gemessen an der Reflexionsfähigkeit, die auf verschiedenen Skalen abgebildet werden sollte, haben die Referendare dahingegen nicht erhalten. Insofern lässt sich interpretieren, dass ‚Unklarheit über den Einsatz‘ der Sammlung über das Repertoire vorlag, obwohl

A2, wie alle anderen Ausbilder der Erprobung auch, an der Entwicklung des Bewertungskonzepts beteiligt war und insofern gut über die Einsatzbereiche der verschiedenen Instrumente informiert war. Weiterhin berichten *alle Auszubildenden* über eine nur marginale oder ausbleibende Nutzung der Matrix, wie z.B. R6, die sich für die Schaffung von noch mehr Transparenz über die Bewertungskriterien einen stärkeren Einsatz der Matrix gewünscht hätte:

R6: „[...]Also der einzige oder auch Hauptkritikpunkt würde ich sagen bei der Umsetzung: Einfach, dass die Matrix irgendwie so ein bisschen/ Ja die war halt da, aber es war irgendwie so ein Schatten, so ein bisschen. Also die hätte, meiner Meinung nach hätte die noch mehr einbezogen werden können. Weil sie ja nun mal da ist und weil ich sie, wie gesagt, auch sinnvoll finde und ich glaube, dass die neben dem Bogen einfach noch mehr die Nachbesprechung hätte strukturieren können, wenn man irgendwie so ein bisschen an der lang gehandelt hätte.“

Für R6 war die Matrix „irgendwie so ein Schatten“, da sie nicht in dem Maße eingesetzt wurde, wie es das Konzept vorgesehen hatte. Über den ausbleibenden Einsatz der Matrix berichtet auch R8, die darin aber kein Problem sah, da sie die Einsatzmöglichkeiten dieses Instruments gar nicht gesehen hatte, obwohl sie auf Seite eins der Matrix explizit formuliert waren:

R8: „[...] Ich bin auch nicht auf die Idee gekommen, das (Anm.: Gemeint ist die Matrix der Qualitätsbereiche, D. L.) zu benutzen, das wäre vielleicht noch besser gewesen, weil das ja noch ausführlicher ist. Aber ich muss ehrlich sagen, jetzt wo sie es sagen, aber das habe ich auch gar nicht im Blick gehabt, dass ich auch das quasi für die Vorbereitung des Statements auch mal mir anschauen könnte.“

Interessant und für die Erprobung des Konzepts wichtig ist der Hinweis von R8, dass sie auf andere, bereits vor dem Beginn der Konzepterprobung eingesetzte Instrumente des ZfsL Bonn ausgewichen ist. Dies sind die bereits mehrfach angesprochenen Bubbles, die von R8 an Stelle der Matrix zur Vorbereitung ihres Statements herangezogen wurden. Die Frage, inwieweit diese Konkurrenzsituation des in der Erprobung befindlichen Bewertungskonzepts einerseits und der

bestehenden Instrumente des ZfsL andererseits die Ergebnisse der Erprobung verfälschen, wird weiter unten, im Fazit dieses Unterkapitels geführt. Betrachtet man die Einschätzung von A2 hinsichtlich des Einsatzes der Matrix, fällt ein bereits aus der Auswertung der Gruppendiskussionen bekanntes Muster auf: Die beiden Akteursgruppen hegen wechselseitige Erwartungen, die dem jeweiligen Gegenüber nicht klar sind, sodass es zu Verwirrungen und Unzufriedenheit kommt. In diesem Sinne spielt A2, in der Frage des Einsatzes der Matrix, den Ball an die Auszubildenden zurück, wenn er den Grund für den Einsatz dieses Instruments in der „Gewissenhaftigkeit“ der Auszubildenden festmacht:

I: „[...] Sie haben schon angesprochen, dass die Referendare auch in, ich sage mal den Verhandlungen der Nachbesprechungen die Qualitätsbereiche herangezogen haben. Haben die sich explizit auch auf die Matrix bezogen, haben die die Matrix herangezogen?“

A2: Unterschiedlich, auch wieder die, die sehr gewissenhaft sind, die haben tatsächlich konkret auch die Matrix vorliegen gehabt, die hatten sie auch dabei und die waren auch inhaltlich vorbereitet und die haben auch die Matrix wirklich auch benutzt und die Aspekte sich das angeschaut und die auch mit eingesetzt in ihr Formular. Es gibt natürlich immer welche, die das sehr locker gestalten dann und nicht so konsequent dann nutzen, das Angebot nicht so konsequent nutzen dann.“

Für A2 war der Einsatz der Matrix ein „Angebot“, dass er offensichtlich nicht verbindlich einforderte, wenn die Auszubildenden nicht von selbst darauf kamen. Die Auszubildenden wiederum waren von dem ausbleibenden Einsatz, wie es oben gezeigt wurde, verunsichert und teils auch unzufrieden und forderten diese – die hierarchische Beziehung schimmert durch – auch nicht offensiv ein oder nutzten sie nicht von sich aus, weil sie von A2 klarere Ansagen hinsichtlich des Instrumenteneinsatzes erwartet hätten.

Schlussendlich zeigt sich, dass in der Erprobung bei A2 v.a. der Dokumentationsbogen die Nachbesprechung strukturiert und hier die Bewertungspraxis zumindest in Ansätzen durchaus geändert hat. Die Matrix wurde hingegen nur

marginal in die Unterrichtsnachbesprechungen einbezogen und die Sammlung über das Repertoire nur teilweise ausgefüllt.

5.2.2 Einsatz bei A5

Aus der Perspektive des Ausbilders lässt sich durch die reine Häufigkeitsverteilung der Subkategorien, wie sie die Code-Matrix in Abbildung 5.4 darstellt, kein dominantes Nutzungsmuster ableiten. Dies, das wird im weiteren Verlauf der Auswertung deutlich, hängt mit den durchaus umfangreichen Berichten des Ausbilders über die Erfahrungen mit dem Einsatz des Bewertungskonzepts ab, die aber nur sporadisch erfolgten, wobei dieser Umstand vom Ausbilder nur an wenigen Stellen erwähnt wurde.

Abbildung 5.4: Code-Matrix der Konzepteinsatzes bei A5



Für zwei der drei Auszubildenden dieses Faches dominieren die Berichte des ausbleibenden Einsatzes die Erprobung sehr deutlich, sie berichten weiterhin lediglich über ‚Unklarheiten im Einsatz‘ des Bewertungskonzepts, aber nicht über andere Formen des Konzepteinsatzes, wohingegen R13 am häufigsten von einem ‚additiven Einsatz‘ des Konzepts berichtet, und sich der ‚ausbleibende Einsatz‘ erst an zweiter Stelle der Häufigkeitsverteilung einordnet.

Wirft man einen Blick auf die einzelnen Berichte der Akteure, dann kristallisiert sich der ‚ausbleibende Einsatz‘ sehr deutlich als das dominante Nutzungsmuster des Konzepts heraus. So antwortet A5 schon auf die einleitende Frage des Interviews, welche Erfahrung aus der Erprobung des Bewertungskonzepts prägend war, wie folgt:

A5: „Dass es mir leid tut, dass ich jetzt Ihnen vielleicht nicht so viel weiter helfen konnte, weil die Erprobung jetzt nicht so vielfältig war, dass ich/Drei Referendare hatten sich ja bereit erklärt dieses Konzept zu erproben, aber das auf mehrere Umstände bedingt, ich glaube, ich bei jedem Referendar das maximal zweimal eingesetzt habe.“

Die Erprobung ist bei A5 offenbar „nicht so vielfältig“ gelaufen, wie es aus seiner Perspektive für eine sinnvolle Erprobung wünschenswert gewesen wäre, den Einsatz des Bewertungskonzepts taxiert er auf „maximal“ zwei Einsätze bei jedem Referendar, was einer Nutzung von ca. 50%¹⁰⁷ bei den durchgeführten Unterrichtsbesuchen entsprechen würde. Erweitert man die Auswertung um die Perspektive der Auszubildenden, dann muss diese Nutzungsquote weiter reduziert werden. *Alle Auszubildenden* berichten über lediglich einen einzigen Einsatz des Bewertungskonzepts, der im Zuge des Einstiegs der Erprobung im Beisein des Autors dieser Arbeit stattgefunden hat. An weitere Einsätze in der Verantwortung von A5 kann sich keiner der Auszubildenden erinnern. Frappierend ist, dass *alle Auszubildenden* diesen ausbleibenden Einsatz als prägende Erfahrung der Konzepterprobung auf die offene Einleitungsfrage des Interviews angeben, wie etwa R12:

I: „Ich würde gerne von Ihnen wissen: Wenn Sie an die Erprobung dieses Bewertungskonzepts denken, den Einsatz dieses Bewertungskonzepts, was fällt Ihnen da spontan als Erstes ein?“

R12: Dass nicht viel funktioniert hat und dass es eigentlich gar nicht zum Einsatz gekommen ist, so richtig.“

Fraglich ist, warum das Bewertungskonzept in der Erprobung kaum zum Einsatz gekommen ist, obwohl sich A5 freiwillig an dem Projekt beteiligt hat und an der Entwicklung des Bewertungskonzepts im Rahmen der Gruppendiskussionen

¹⁰⁷ Die Erprobung des Konzepts startete nach dem ersten Ausbildungshalbjahr der Referendare, in dem diese schon zwei Unterrichtsbesuche durchgeführt hatten und die meisten noch vier Besuche ableisten mussten. In Einzelfällen fehlten den beteiligten Auszubildenden bei dem Beginn der Erprobung lediglich drei Unterrichtsbesuche, andere mussten noch fünf durchführen.

beteiligt war, also von einer relativ hohen intrinsischen Motivation für die Teilnahme an der Erprobung auszugehen ist. Auf die Nachfrage nach den Gründen für den ausbleibenden Einsatz nennt A5 drei Bereiche, in denen die Gründe zu verorten sind, zeitlicher Druck, organisatorische Rahmenbedingungen und Konflikte mit den Auszubildenden:

I: „Das ist ja vielleicht schon ein interessanter Punkt, wie kam das, dass Sie das nur zweimal eingesetzt haben, welche Umstände haben dazu geführt, dass Sie das nicht getan haben?“

A5: Also Umstände/Einmal war es der zeitliche Faktor, weil ich finde, wenn man mit diesem Konzept arbeitet, man wirklich Zeit braucht für die Nachbesprechung und was in diesem Durchgang sowieso mein Problem war/Viele sagen ich war zu lieb, vielleicht für die Referendare, ich habe eben auch meine Reisekostenabrechnung noch gemacht, da habe ich das gesehen, gerade bei der R12, da hatte ich einmal Zeit, als Sie dabei waren und die anderen beiden Male, da hatte ich drei Unterrichtsbesuche an einem Tag, und da hat sich das wirklich reduziert auf das Wichtigste. 45 Minuten waren eingeplant, und da sind wir dann nicht mehr dazu gekommen. Aber da war es mir bedeutsamer, dann wirklich die Termine bei den anderen Referendaren wahrzunehmen, weil die eben in Zeitdruck standen, jetzt durch neue OVPs, 18 Monate und so. Das war ein Grund, der zeitliche Faktor. Dann war einmal, persönlich bedingt, mit einem Teilnehmer oder einer Probandin sind/ Also ich habe das für angebracht gehalten, dass wir spazieren gegangen sind und dabei die Nachbesprechung gemacht haben, aus mehreren Gründen, und beim dritten Probanden war das so, dass einmal der Kernseminarleiter mit war, der hat sich die Gesprächsmoderation nicht aus den Fingern nehmen lassen, und bei den letzten Unterrichtsbesuchen wir persönliche Probleme hatten und ich deshalb auf mein altbewährtes Konzept zurückgegriffen habe, weil da habe ich mich sicherer gefühlt. Weil ich immer die Wahrnehmung hatte, dass er versucht, mich immer irgendwo anzugreifen, PING PENG! Und da bin ich dann auf mein altes Konzept umgestiegen. Also das war mir sicherer.“

Der zeitliche Druck, dem A5 sich auf Grund der verkürzten Ausbildungsdauer des Vorbereitungsdienstes ausgesetzt sah, kann nicht der ursächliche Grund für die ‚ausbleibende Nutzung‘ des Konzepts sein, schließlich haben auch A2 und A6 unter den gleichen Rahmenbedingungen eine Erprobung des Konzepts gewährleisten können. Spannender erscheinen schon die Hinweise auf die Hemmnisse zu sein, dies sich aus der organisatorischen Logik der Unterrichtsnachbesprechungen ergeben, wie z.B. sein Einwurf, dass sich „der Kernseminarleiter die Moderation nicht hat aus der Hand nehmen lassen“. Hier taucht das bereits aus der Auswertung der Gruppendiskussionen bekannte Problem auf, dass offensichtlich auch hierarchische Beziehungen unter den Ausbildern maßgeblichen Einfluss auf die Bewertungsprozesse haben und generell dazu führen, dass sich die Bewertungsprozesse bei ein und demselben Ausbilder unterscheiden können, je nachdem, ob er alleine oder gemeinsam mit einer anderen Person bewertet, und dass dies im konkreten Fall zu einer ‚ausbleibenden Nutzung‘ des Bewertungskonzepts geführt hat. A5 berichtet daneben offensichtlich von Konflikten in den Ausbildungsbeziehungen, die dazu geführt haben, dass er das Konzept nicht eingesetzt hat. Bei einer Probandin hat er es „für angebrachter gehalten, spazieren zu gehen und dabei die Nachbesprechung zu machen“, als das Konzept einzusetzen, bei einem weiteren Probanden fühlte er sich dadurch unter Druck gesetzt, da dieser „versucht hat mich immer irgendwo anzugreifen“, sodass er auf sein „altbewährtes Konzept zurückgegriffen“ hat, weil er sich dort „sicherer gefühlt“ habe. Weiterhin gibt A5 die Problematik, das Bewertungskonzept mit den fachlichen Anforderungen des konkreten Unterrichtsfachs in Einklang zu bringen, als Grund für den ausbleibenden Einsatz an¹⁰⁸:

I: „Und wenn ich es richtig verstanden habe vorhin, lag das, dass das so geholpert hat, in dem Einsatz der Instrumente primär daran, dass, ich sage mal die Logik, um die es in der Ausbildung ging, auf anderen Ebenen sich abspielte. Da ging es um Beziehungsaspekte, zwischen Ihnen/

¹⁰⁸ Die Konflikte des Bewertungskonzepts mit den fachdidaktischen Anforderungen der Unterrichtsfächer werden in einem eigenen Unterkapitel (siehe Kapitel 5.3.4) abgehandelt. Dies geschieht aus Gründen der Anonymisierung, damit auf Grund der Zugehörigkeit zu Fächern der Ausbilder keine Rückschlüsse auf konkrete Personen möglich sind.

A5: Zeitliche Aspekte, Beziehungsaspekte und, ja und glaube ich auch die Schwierigkeit, eben das so auf den Fachunterricht zu übertragen

Die Auszubildenden haben die ausbleibende Erprobung des Konzepts registriert und zumindest *zwei Auszubildende* haben explizit nachgefragt, warum das Konzept nicht zum Einsatz kommt und darum gebeten, es für die Nachbesprechung des Unterrichts zu nutzen, aber darauf keine klare Reaktion erhalten. R14 berichtet dazu:

I: „Ok, haben Sie irgendwie das Konzept für sich mitgebracht und sind noch mal drauf zu sprechen gekommen?“

R14: Dabei hatte ich es nicht. Ich bin noch einmal in einer Nachbesprechung, in einer Nach-Nachbesprechung darauf zu sprechen gekommen. Da wurde gesagt, ja es wird zugeschickt und weil das nicht geschehen ist, habe ich auch nicht mehr nachgefragt, weil die Beziehung ohnehin etwas belastet war.“

Nach Abschluss der letzten Unterrichtsbesuche haben *alle drei Auszubildenden* mit der Sammlung über das Repertoire zumindest eines der drei Instrumente erhalten, die Teil des Bewertungskonzepts waren, aber alle drei berichten über eine Nutzung des Instruments, die nicht der Intention entsprach, sondern ‚*additiv*‘ geschehen ist, wie R13 ausführt:

I: „Sie haben ja vorhin schon gesagt, Sie haben noch dieses dritte Instrument, die Sammlung über das Repertoire gekriegt, aber haben da ja schon, wenn ich es richtig in Erinnerung habe, einschränkend gesagt: Das passte nicht, was da drin steht.“

R13: Es war einfach nur sehr kurz. Also es war, das passte irgendwie schon, aber es war so ein bisschen, ja im Nachhinein halt ausgefüllt. Also man merkte, das musste man jetzt machen, weil das jetzt hier so ein Projekt war, aber es war jetzt nicht so, dass das irgendwie gemacht wurde um mir damit zu helfen. [...] Und ich habe es mir durchgelesen, und es war schon eher so eine grobe Erinnerung an das, was da in diesen UBs passiert ist. Ich glaube nicht, dass das wirklich aus vielen Aufzeichnungen entstanden ist.“

R14 schließt sich der Einschätzung von R13 an und ist der Auffassung, dass man ganz deutlich sehe, dass das Instrument nicht entsprechend seiner Intention eingesetzt wurde:

R14: „[...] Also man liest ganz deutlich, dass dieses Gutachten nicht prozedural angefertigt wurde, sondern rückblickend zu einem Zeitpunkt. Hier ist ein Meinungsmoment abgefasst.“

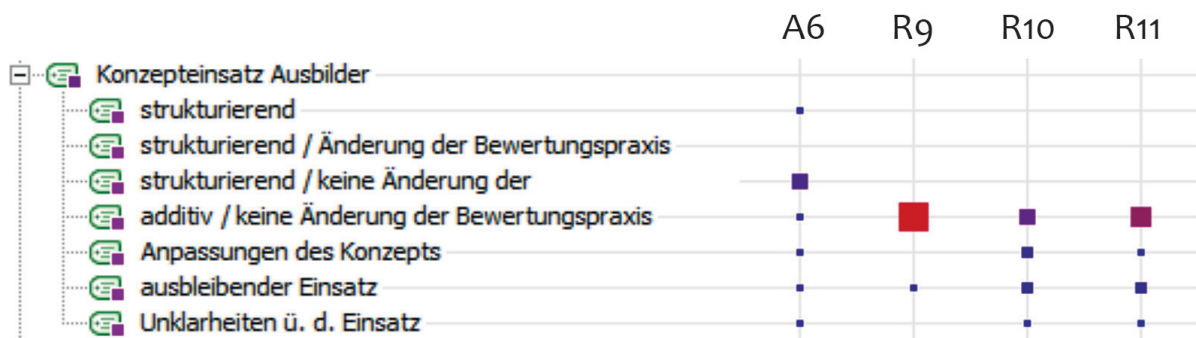
Im Fall von A5 scheint allerdings, abweichend von A2, weniger die ‚Unklarheit über den Einsatz‘ des Repertoires entscheidend für den nur mangelhaften Einbezug des Instruments gewesen zu sein, als vielmehr die Tatsache, dass A5 auf Grund der ausbleibenden Nutzung des Dokumentationsbogens schlichtweg keine Datengrundlage hatte, um die Sammlung über das Repertoire auszufüllen.

Zusammenfassend lässt sich also festhalten, dass im Fall von A5 die Erprobung des Bewertungskonzepts in weiten Teilen nicht stattgefunden hat und dort, wo die Instrumente genutzt wurden – im Fall der Sammlung über das Repertoire – die Nutzung lediglich als additiv zu beschreiben ist. Die Gründe für diese ausbleibende Nutzung lassen sich, A5 folgend, in organisatorischen Problemen des Ausbildungsalltags, in Beziehungsproblemen zu den Auszubildenden und in einem Konflikt mit der fachlichen Passung des Bewertungskonzepts ausmachen.

5.2.3 Einsatz bei A6

Die Unterschiede in der Wahrnehmung zwischen dem Ausbilder und den Auszubildenden fallen bei einer Betrachtung der Code-Matrix für den Einsatz bei A6 (vgl. Abb. 5.5) auf den ersten Blick ins Auge. Die am häufigsten codierten Subkategorien von A6 befinden sich im Bereich eines ‚strukturierenden‘ Einsatzes des Bewertungskonzepts, auch wenn A6 durch den Einsatz des Bewertungskonzepts scheinbar keine großen Veränderungen an seiner Bewertungspraxis festgestellt hat. Die Referendare hingegen sind sich einig, dass das dominante Erprobungsmuster des Konzepts ein ‚additiver‘ Einsatz war, eine Strukturierung der Bewertungsprozesse durch das Konzept hat keiner der Auszubildenden wahrgenommen.

Abbildung 5.5: Code-Matrix des Konzepteinsatzes bei A6



Vertieft man die Analyse um einen genaueren Blick auf die qualitativen Daten, dann zeigen sich Gründe, warum sich aus der Code-Matrix eine unterschiedliche Wahrnehmung der beiden Akteursgruppen herauslesen lässt. A6 berichtet mehrfach darüber, wie er die Instrumente des Bewertungskonzepts genutzt hat und hierbei das Konzept in seinem Sinne ‚angepasst‘ hat. Insofern kam das Konzept durchaus ‚strukturierend‘ zum Einsatz, aber erstens ließ A6 sich dabei nicht von seinem ursprünglichen Ablauf der Nachbesprechung abbringen, sondern der vom Bewertungskonzept intendierte Ablauf ähnelte dem ohnehin praktizierten, und zweitens wurden die Instrumente des Konzepts so an die individuelle Ausbildungslogik angepasst, dass der Einsatz eher ‚additiv‘ erschien, als dass er die Bewertungspraxis tatsächlich strukturiert hätte. Auf die direkte Frage, inwiefern die Instrumente seine Bewertungspraxis strukturiert haben, antwortet A6 wie folgt:

I: „Haben diese Instrumente die Nachbesprechung dann strukturiert oder sind Sie eher Ihren normalen Gang gegangen und haben die dann so letzten Endes so ein bisschen additiv am Ende dran gemacht oder haben Sie sich strukturieren lassen?“

A6: Also ich habe es so gemacht, wie ich es auch sonst mache, habe gesagt, welche Punkte haben Sie zur Nachbesprechung und die Referendarinnen und Referendare kennen ja dieses Bonner Qualitätstableau, kannten in diesem Falle auch dieses hier, hatten also alle Möglichkeiten, hierauf zurückzugreifen. Und ich habe dann wo es nötig war, weitere Punkte hinzugefügt. Die hätte man teilweise auf dieser Ebene formulieren können,

aber es gab z.B. Punkte, die sich hier nicht abbilden lassen, wie z.B. vertieftes Textverstehen, kommt hier unter dem Oberbegriff vor inhaltliche Klarheit oder so was. Aber, oder auch Lernaufgabenkonstruktion, das sind Dinge, die im Fachdidaktikseminar, wenn Sie sich dran erinnern, riesige Rollen spielen, die aber hier nur unter dem großen Hut der inhaltlichen Klarheit sich wiederfinden, aber innerhalb der Fachdidaktik wird eben inhaltliche Klarheit, zerfällt in ganz viele Einzelteile und spielt eben dann einen entscheidende Raum und dann habe ich eben auch neue Faktoren angefügt.“

A6 hat die Nachbesprechung also „so gemacht, wie er sie auch sonst macht“ und in den Instrumenten jene „Faktoren angefügt“, die aus einer fachdidaktischen Perspektive entscheidend waren. Insofern hatte das Konzept keine große Bedeutung für die Strukturierung der Nachbesprechung, eher lief es in dem ohnehin stattfindenden Nachbesprechungen nebenher mit und – salopp formuliert – störte diese nicht groß, da es seiner Logik nach dem ohnehin praktizierten Ablauf der Nachbesprechung von A6 entsprach, wie auch die folgende Interviewpassage zeigt, in der A6 dem erprobten Konzept keine großen Unterschied zum tradierten Ablauf attestiert:

A6: „Ich verstehe jetzt nicht ganz, wo der Gewinn ist im Vergleich zu einer herkömmlichen Nachbesprechung, wenn eine Agenda aufgemacht wird.

I: Was ist eine Agenda, können Sie das umreißen?

A6: Also der Unterrichtsbesuch endet damit, dass oder die Nachbesprechung fängt an, dass der Referendar, die Referendarin ein Statement abgibt, dann geben die Ausbilder ein Statement ab und danach steht dann die Frage: Zu welchen Punkten melden Sie Beratungsbedarf an. [...]

I: Und das konnte das Konzept (Anm.: Gemeint ist das in der Erprobung befindliche Bewertungskonzept, D. L.) auch leisten?

A6: Das konnte das Konzept auf jeden Fall leisten.

I: Aber es wäre jetzt kein großer Unterschied gewesen zu dem ohnehin vorhandenen Qualitätstableau?

A6: Genau, und ich vermute mal, dass andere Seminare, ohne dass ich mich damit täglich beschäftige, ähnliche Hilfen haben.“

Entsprechend macht es aus einer Ausbilderperspektive durchaus Sinn, von einem ‚strukturierenden‘ Einsatz zu sprechen, aber diese Strukturierung rührte nicht aus dem erprobten Bewertungskonzept her, sondern daher, dass dessen intendierte Strukturierung in weiten Teilen mit der ohnehin praktizierten Struktur der Bewertungsprozesse von A6 übereinstimmte, sodass für die Auszubildenden der Einsatz des Konzepts v.a. an jenen Stellen klar wurde, an denen es für sich stehend eingesetzt wurde und diese Art des Einsatzes schloss sich dann eben ‚additiv‘ an die eigentliche Nachbesprechung, in der die wesentlichen Bewertungsvorgänge bereits gelaufen waren an, wie z.B. R 10 berichtet, der eine „Transformation“ des eigentlichen Bewertungsvorgangs in das in Erprobung befindliche Bewertungskonzept, genauer den Dokumentationsbogen für die Nachbesprechung, wahrgenommen hat:

I: „Konnten Sie bei dem Ausbilder einen Unterschied feststellen zwischen den Lehrproben, die anfangs ohne das Konzept bewertet worden sind im Vergleich zu den Lehrproben, die dann mit dem Konzept bewertet worden sind?“

R10: Nein, aus dem Grund, weil er parallel bewertet hat. Also er hat tatsächlich seinen klassischen Auswertungsbogen immer dabei gehabt in der Lehrprobe, hat diesen auch zuerst ausgefüllt und hat dann überlegt, wie er von seinem Konzept ausgehend, transformierend in Ihr Konzept das Ganze umschreiben kann.“

Auch R9 und R11 berichten von ähnlichen Erfahrungen, in denen der Ausbilder die abgeschlossenen Bewertungsprozesse auf den in der Erprobung befindlichen Dokumentationsbogen „übertragen“ hat, wie es R11 ausdrückt:

I: „Konnten Sie einen Unterschied feststellen, hinsichtlich der Nachbesprechung, im Vergleich zu den Lehrproben, die von dem Ausbilder ohne das Konzept bewertet worden sind und den Lehrproben, die mit dem Konzept bewertet worden sind?“

R11: Nee, das lag aber glaube ich, das lag aber daran, dass er sich eher an seinem Konzept orientiert hat und das nachher auf Ihr Konzept übertragen hat.

I: Was meinen Sie damit „auf Ihr Konzept übertragen“?

R11: Ja, also im Prinzip, dass er halt, dass wir die Besprechung und alles andere, haben wir nach seinem Konzept gemacht, in dem er geguckt hat, was er da hin geschrieben hat und danach haben wir dann, quasi gemeinsam, die Kreuze dahin gemacht, [...]“.

Die Matrix der Qualitätsbereiche wurde aus der Perspektive der Auszubildenden auf die gleiche Art und Weise eingesetzt. Abgeschlossene Bewertungsaspekte wurden nachträglich auf das Instrument bezogen, sie wurden den Qualitätsbereichen „zugeordnet“ wie R11 es nennt:

I: „Klang jetzt für mich so: Die Matrix mit diesen Qualitätsbereichen, die wurden, wurde die genutzt? Wurde die herangezogen? [...]“

R11: Ja, um es abzugleichen, um dann nachher zu sagen: Wo kann man das denn jetzt hier hin packen, was könnten wir zu klare Strukturierung packen? Das ungefähr.

I: Also d.h. auch im Nachhinein, nachdem schon die Bewertung geschehen war und die Erläuterung wurde/Wo könnte man das denn hinpacken und nicht etwa, dass man guckt: Ok ich habe mir/Folgender Punkt, Klarheit und Strukturiertheit hat es bei Ihnen gehapert, ich glaube, das lag da und daran, lassen Sie uns da mal hingucken?

R11: Genau richtig und das ist halt das, wenn man wahrscheinlich mehrere UBs nacheinander und auch wirklich an Hand dieses Konzepts

und nicht erst in die Übertragung nimmt, dass man dann sagen kann: Ok, das hier sind die Punkte, die gucken wir uns an, daran müssen wir dann arbeiten.“

Die Matrix wurde also nicht dazu genutzt, um über den Unterricht ins Gespräch zu kommen, wie es als Zielsetzung der Bewertungskonzepts formuliert wurde, sie generierte keine Gesprächsanlässe, sondern sie wurden nach Abschluss eines Bewertungsaspektes „abgehakt“, wie es R9 formuliert:

I: „Wurde in der Nachbesprechung explizit Bezug auf die Matrix genommen, auf die dort festgelegten Qualitätsbereiche, von Ihnen oder von dem Ausbilder, wurde die herangezogen?“

R9: Explizit nicht, es wurde herangezogen um es, ich nenne es mal abzuhaken.“

Gleichermaßen wurde das dritte Instrument, die *Sammlung über das Repertoire*, nicht in dem Sinne genutzt, wie es die Intention des Bewertungskonzepts vorsah. A6 berichtet darüber, dass er zum Abschluss der Ausbildung durchaus versucht hat das Instrument zu nutzen, aber mit der Nutzung sehr unzufrieden war, da es dem „komplexen Systemanspruch“, der seiner Ausbildung zu Grunde liegt, nicht gerecht werden könne, obwohl er am Beginn der Erprobung große Hoffnungen in das Instrument gesetzt habe:

A6: „[...] Habe mich auch sehr schwer getan mit diesem Repertoire hinterher, was ich mir vorher als großartig vorgestellt habe, weil ich eben denke, dass es diesem komplexen Systemanspruch, den ich zumindest habe, wenn ich Lehrerinnen und Lehrer ausbilde, nicht gerecht werden kann. Weil es zu linear angelegt ist.“

Diese Äußerungen von A6 lassen darauf schließen, dass das Instrument, trotz der negativen Rezeption durch A6, die weiter unten ausführlicher dargestellt wird, genutzt wurde. Wirft man einen Blick auf die Berichte der Auszubildenden, muss dies durchaus in Frage gestellt werden, denn deren Berichte lassen auf nur eine unvollständige Nutzung schließen. R11 berichtet, dass ihr am Ende der Aus-

bildung kein Repertoire ausgehändigt wurde, wohingegen R9 und R10 durchaus ein ausgefülltes Exemplar zugegangen ist, dieses aber nur unvollständig ausgefüllt worden sei, wie beide berichten:

I: „Sie haben ja auch von xy dieses dritte Instrument gekriegt, diese Sammlung über das Repertoire.

R9: Ja, wobei er da ein oder zwei Lehrproben vergessen hat. Deswegen, ja/Das habe ich bekommen. [...]

I: Hat Ihnen diese Sammlung über das Repertoire geholfen oder haben Sie den Eindruck, dass diese Sammlung über das Repertoire helfen kann, zentrale Stärken Ihrer Lehrerpersönlichkeit zu erkennen aber auch Entwicklungsfelder? Wird das daraus deutlich aus diesem Instrument?

R9: Könnte, ja, wenn es denn vollständig ausgefüllt worden wäre, was hier ja auch einfach nicht der Fall ist. Könnte es sicherlich hilfreich sein, kann ich jetzt nicht beantworten, weil das hier ja nicht vollständig ausgefüllt ist, von daher eingeschränkt. Ich könnte mir vorstellen, das hilft/[...]“

R9 „kann sich vorstellen“, dass das Instrument hilft eine Bilanzierung der eigenen Professionalisierung vorzunehmen, aber verbleibt bei dieser Einschätzung im Konjunktiv, denn für eine Beantwortung der Frage wäre es notwendig gewesen, dass es „vollständig ausgefüllt worden wäre“: Auch R10 berichtet über die gleichen Einschränkungen der Instrumentennutzung und teilt die Einschätzung von R9 eines potentiell positiven Nutzens des Instruments:

I: „Zeigt das Repertoire, zeigt diese Sammlung über das Repertoire, nichtsdestotrotz vielleicht zentrale Kompetenzen Ihrer Lehrerpersönlichkeit oder Entwicklungsfelder und, oder Entwicklungsfelder?“

R10: Nein, weil es nicht vollständig ausgefüllt ist. Es sind von – wie viele Kriterien haben Sie, eins, zwei drei, vier, fünf, sechs, sieben, acht, neun, zehn, elf, zwölf–sind vier ausgefüllt, und das auch nur mit zwei Unterrichtsbesuchen von vieren. Und deswegen ist eben auch die Entwicklung so mit diesem

Instrument, wie er es ausgefüllt hat, nicht nachvollziehbar. Interessant wäre es für mich tatsächlich dieses Repertoire so zu haben, dass er mir das mal mit all meinen Unterrichtsbesuchen ausfüllt und eben auch entsprechend, ja in Hinblick auf eine Leistungsbewertung nach Examensstandard einordnen würde. Und ich glaube, dann könnte man da durchaus gut erkennen, dass da so eine Entwicklung stattgefunden hat.“

Schlussendlich kann man das Fazit ziehen, dass der strukturierende Einfluss des Bewertungskonzepts im Fall von A6 nicht vorgelegen hat, sondern die Struktur der Bewertungsprozesse, wie sie das Bewertungskonzept intendiert, mit den von A6 praktizierten Prozessen in weiten Teilen übereinstimmte. A6 erlebte den Einsatz des Konzepts entsprechend durchaus ‚strukturierend‘, aber eben nicht im Sinne einer Änderung der Bewertungspraxis, sondern in dem Sinne, dass er die durch das Bewertungskonzept intendierte Struktur der Bewertungsprozesse ohnehin schon umsetzte, das Konzept also „keinen Gewinn“ darstellte. Für die Auszubildenden waren lediglich die Verschriftlichung im Sinne des Dokumentationsbogens und die Einordnung von bereits geklärten Aspekten des Unterrichts in die Matrix der Qualitätsbereiche auffällig, die im Anschluss an die eigentlichen Bewertungsprozesse erfolgte, sodass sie daraus den Schluss zogen, dass der Einsatz des Bewertungskonzepts ‚additiv‘ erfolgte. Auch hier, das zeigen die Zitate von A6, kam es zu einer Konkurrenzsituation zwischen dem in der Erprobung befindlichen Bewertungskonzept und den im ersten Ausbildungshalbjahr genutzten Bewertungsinstrumenten des ZfsL Bonn, dem Qualitätstableau und der Zusammenfassung von diesem, den sog. Bubbles, wie es bereits bei einzelnen Auszubildenden im Erprobungskontext von A2 festgestellt werden konnte. Die Auswirkung dieser Konkurrenzsituation auf die Erprobung wird weiter unten diskutiert.

5.2.4 Zusammenfassende Interpretation ‚Konzept Einsatz Ausbilder‘

Drei Ausbilder, drei verschiedene Nutzungsmuster – so lässt sich das Fazit der Auswertung der Kategorie des ‚Konzept Einsatzes der Ausbilder‘ pointiert formulieren. Die dominanten Nutzungsmuster decken dabei eine große Spannbreite möglicher

Einsatzformen ab, von einer de facto nicht erfolgten Erprobung des Konzepts bei A5, über eine additive Ergänzung der eigenen Bewertungspraxis bei A6 bis zu einer erfolgten Umsetzung des Konzepts mit zumindest – partiellen – Veränderungen der Bewertungspraxis bei A2. Einem ersten Impuls folgend müsste dieses Fazit der Kategorie als verheerend im Sinne der Konzepterprobung gebrandmarkt werden. Verschiebt man allerdings den Blickwinkel etwas, so lassen sich die erlebten Probleme der Erprobung als *Implementationsprobleme* deuten, aus denen sich wichtige Erkenntnisse für die Veränderung der Bewertungspraxis in der zweiten Phase der Lehrerbildung ableiten lassen. Offensichtlich wiesen die im Zuge der beruflichen Sozialisation entwickelten und individuell geprägten Bewertungsmuster der Ausbilder A5 und A6 eine derart hohe Stabilität auf, dass sie, trotz des für sich genommen sehr günstigen Settings, für eine Veränderung der Bewertungsprozesse¹⁰⁹ überwiegend stabil blieben. Folglich muss als erster wesentlicher Schluss aus der Erprobung des Bewertungskonzepts abgeleitet werden, dass eine Veränderung der Bewertungsprozesse in der Zweiten Phase der Lehrerbildung allein durch die Entwicklung und das Angebot von Bewertungskonzepten nicht zu erwarten ist, die den hier formulierten Ansprüchen genügt. Die hochgradige Autonomie der Ausbilder, gepaart mit langjährig gewachsenen, individuell geformten Praxen der Bewertung, verhindern die Wirksamkeit von neuen Bewertungskonzepten, so kann aus der Erprobung geschlussfolgert werden. Der im Kapitel zwei aufgestellten These, dass sich die Bewertungspraxen der Ausbilder nur sehr eingeschränkt beeinflussen lassen, kann damit zunächst die größere empirische Bedeutsamkeit zugesprochen werden, als der alternativen These. Diese besagte, dass ein standardisiertes Bewertungskonzept eine Steuerungswirkung zu Gunsten transparenter Bewertungsprozesse entfalten würde. Auf diesen Schluss deuteten auch schon die Ergebnisse der Gruppendiskussionen hin, in denen beide Akteursgruppen davon berichten, dass existierende Bewertungsinstrumente in der Ausbildungspraxis oftmals keinen befriedigenden Einsatz fanden. Vor dem Hintergrund der oben ausgeführten Theorien institutionellen Wandels¹¹⁰ erscheint dieser ‚Widerstand‘ in

¹⁰⁹ Siehe Kapitel 3.3.

¹¹⁰ Siehe Kapitel 3.3.

den Veränderungsprozessen zunächst nicht als unerwartet. Allerdings überrascht das Ausmaß dieser Hindernisse, die sich einer erfolgreichen Erprobung in den Weg stellen durchaus. Wenn schon im Rahmen der *Erprobung* eines standardisierten Bewertungskonzepts, in einem für Veränderungsprozesse als günstig einzuschätzenden Kontext derartige Widerstände gegen den Einsatz auftauchen, welches Ausmaß werden diese bei einem tatsächlichen Implementationsversuch in der Ausbildungspraxis annehmen? Gerade das Beispiel von A5 hat gezeigt, dass es eine starke Handlungslogik gibt, im Falle von Konfliktsituationen mit den Auszubildenden auf erlernte und – vermeintlich? – bewährte Muster der Bewertungen auszuweichen, weil dies sicheres Terrain verspricht. Die Funktion des Widerstands scheint hier in der Tat der Abwehr von Ängsten, v.a. von Kontrollverlust zu dienen (Reh 2007: 292). Vor dem Hintergrund der konfliktär geprägten Beziehungen zwischen den Akteuren, wie sie in der Forschung zur zweiten Phase der Lehrerbildung immer wieder rekonstruiert werden¹¹¹, erscheint eine „Abwehrreaktion“ gegen die versuchte Implementation, wie Oelkers es nennt, auch im Bereich der Studienseminare naheliegend.

„Als ein Kernproblem erweist sich dabei, wie neue Instrumente und Verfahren in der Schule Akzeptanz finden oder wie sie implementiert werden können, ohne immer gleich abgestoßen zu werden. Man kann sich das Innenleben von Schulen wie einen Immunkörper vorstellen, der abwehrt, was nur die Belastungen steigert. Wer etwas verändern will, darf also nicht wie eine drohende Krankheit erscheinen. (Oelkers 2006: 106).

Ob Prozessbegleitung der Implementation, professionelle Institutionsentwicklung oder Change Management (vgl. z.B.: Buhren & Rolff 2012, Rolff 2007, Rolff et al. 2011) diesen Implementationshindernissen erfolgreich begegnen können, kann in dieser Arbeit nicht systematisch überprüft werden, aber die Erfahrungen des Konzepteinsatzes lassen einen Erfolg eher als unwahrscheinlich erscheinen. Wenn noch nicht einmal intrinsisch motivierte Ausbilder in einem von der Leitung abgesicherten, experimentellen Rahmen, durch dieses Konzept zur

¹¹¹ Siehe Kapitel 1.2.

Änderung ihrer Bewertungspraxis zu bewegen sind, wann werden es die Akteure dann im Alltag tun?

Nun war es nicht das Ziel dieses Forschungsprojektes, die Implementation eines Bewertungskonzepts wissenschaftlich zu begleiten, sondern jenes zu entwickeln und die Erprobung desselben mit qualitativen Methoden zu evaluieren. Man könnte also argumentieren, dass die gescheiterte Veränderung der Bewertungspraxis schlichtweg in einem schlecht designten Bewertungskonzept zu suchen ist, das so einfach nicht in der Ausbildungspraxis einsetzbar war, die gescheiterte Erprobung also kein *Implementationsproblem*, sondern ein *Designproblem* war. Wie die Auswertung des folgenden Unterkapitels zeigen wird, ist diese Interpretation allerdings zu kurz gegriffen. Das Konzept weist an zentralen Stellen durchaus Schwächen auf, aber es erfährt in vielen Aspekten, v.a. durch die Auszubildenden und teilweise auch durch die Ausbilder, eine positive Würdigung. An den Stellen, an denen sich Probleme des Konzeptdesigns offenbaren, sind diese allerdings oft grundsätzlicher Natur, sodass es fraglich erscheint, ob es überhaupt ein Bewertungskonzept geben kann, was diese Probleme löst und die in Kapitel eins formulierten Ansprüche erfüllt.

Bevor die Arbeit zu diesem Auswertungsschritt fortschreitet, muss allerdings ein bereits mehrfach angesprochener Aspekt der Konzepterprobung in seiner Auswirkung auf die Aussagekraft eben dieser diskutiert werden. Die Erprobung des Bewertungskonzepts begann nach dem Ende des ersten Ausbildungshalbjahres der Auszubildenden, sodass diese schon mindestens einen, in den meisten Fällen jedoch zwei Unterrichtsbesuche absolviert hatten. Grundlage der Bewertung dieser Unterrichtsbesuche war – zumindest theoretisch – das vom ZfsL Bonn genutzte Qualitätstableau, sowie als Hilfestellung in Kurzform die sog. Bubbles. Im Zuge der Auswertung der Kategorie des ‚*Konzeptesinsatzes*‘ wurde mehrfach deutlich, dass diese Bewertungsgrundlagen in Konkurrenz zur Erprobung des Bewertungskonzepts standen, die Auszubildenden also nicht die in der Erprobung befindlichen Instrumente, sondern das Tableau oder die Bubbles für die Vorbereitung des Unterrichts, bzw. der Nachbesprechung genutzt haben. Insofern ließe sich auf den ersten Blick vermuten, dass die Konzepterprobung hinsichtlich des

Nutzens zentraler Instrumente, wie z.B. der Matrix, keine Schlüsse zulässt, weil die Instrumente nicht gemäß ihrer Intention genutzt worden sind. Streng genommen kann man dieser Perspektive folgen, aber im Zuge der Auswertung im folgenden Unterkapitel wird deutlich, dass die Matrix und das Qualitätstableau prinzipiell eine sehr ähnliche Struktur und eine nahezu identische Funktion aufweisen. Beide definieren zentrale Qualitätsbereiche guten Unterrichts, die Matrix des Bewertungskonzepts in Bezug auf Meyer und Helmke. Das Qualitätstableau nutzt Meyers Kriterien und erweitert sie um Aspekte, die den Ausbildern der ZfsL Bonn wichtig waren. Insofern kann der positive Nutzen, den das Qualitätstableau stiftet, zumindest teilweise auch auf die Matrix übertragen werden. Interessant ist die Frage, warum die Auszubildenden in Einzelfällen nicht die Matrix, sondern weiterhin das Qualitätstableau oder sogar nur die sehr knapp gehaltenen Bubbles genutzt haben. Die Konkurrenzsituation der Instrumente wird durch diese Perspektive nicht negiert, sondern in die Evaluation des Einsatzes einbezogen, mit dem Ziel herauszufinden, welches Instrumentendesign den größten Beitrag für produktive Bewertungsprozesse liefert.

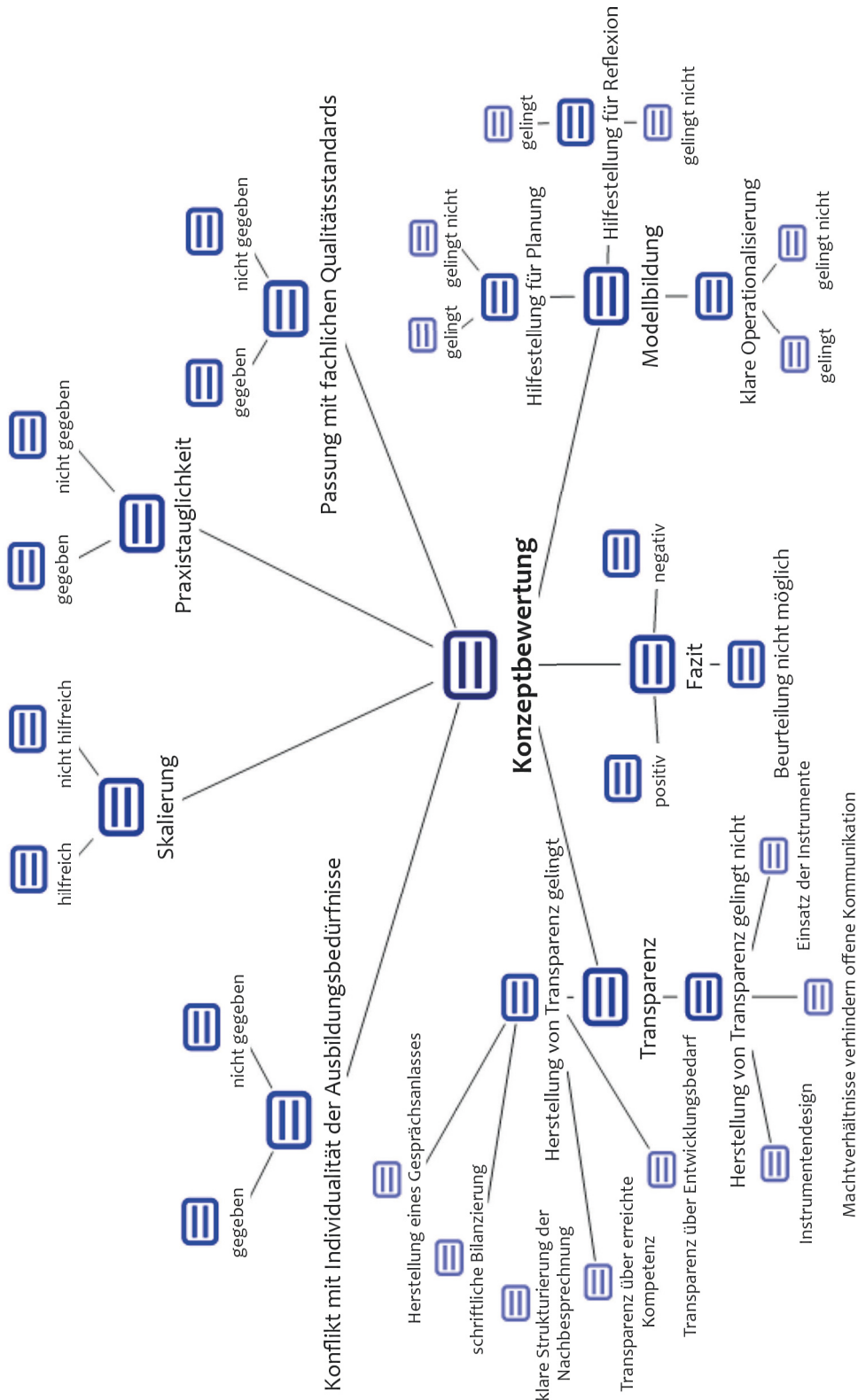
5.3 Konzeptbewertung

Die Kategorie der ‚*Konzeptbewertung*‘, die mit ihren Subkategorien in Abbildung 5.6 dargestellt ist, entspricht dem Herzstück der Evaluation der Konzepterprobung. Mit ihren Subkategorien erfasst sie die Einschätzungen der Akteure, inwieweit das Bewertungskonzept, die Erwartungen die in es gesetzt wurden, erfüllt. Hierbei beziehen sich die Subkategorien einerseits auf die in Kapitel eins formulierten Ansprüche des Bewertungskonzepts und andererseits auf jene ‚*leitenden Konstruktionsprinzipien*‘, die aus den Gruppendiskussionen abgeleitet werden konnten. Die Subkategorie eines möglichen ‚*Konflikts mit der Individualität der Ausbildungsbedürfnisse*‘ trägt dem Anspruch Rechnung, dass das Konzept mit der Vielfalt möglicher Professionalisierungswege, sowie verschiedenen Lehrerpersönlichkeiten vereinbar sein soll, wie es in Kapitel 2 ausgeführt wurde. Diese Subkategorie nimmt eine dichotome Unterteilung des Feedbacks der Akteure vor

und unterscheidet dieses hinsichtlich der Frage, ob ein solcher Konflikt durch das Konzept gegeben war oder eben nicht. Die Bedeutung einer erhöhten ‚*Transparenz*‘ der Bewertungsprozesse wurde im Rückgriff auf den Forschungsstand zur zweiten Phase der Lehrerbildung herausgearbeitet, er zeigte sich aber auch im Zuge der Auswertung der Gruppendiskussionen als zentrales Qualitätskriterium produktiver und akzeptierter Bewertungsprozesse. Die Rekonstruktion der Sichtweisen, die sich in den Interviews offenbarten, werden für diese Subkategorie zunächst einer dichotomen Unterscheidung unterworfen, welche die Beiträge danach gliedert, ob die Akteure davon berichten, dass die ‚*Herstellung von Transparenz gelingt*‘, bzw. durch das Konzept die ‚*Herstellung von Transparenz nicht gelingt*‘. Diese beiden Subkategorien unterscheiden ihrerseits nach verschiedenen Gründen, warum die Herstellung von Transparenz gelingt oder eben nicht. Die Subkategorie der ‚*Skalierung*‘ greift den in Kapitel zwei formulierten Anspruch auf, dass die Erfassung der Kompetenz durch das Bewertungskonzept wissenschaftlichen Ansprüchen genügen soll, aber ebenso die aus den Gruppendiskussionen abgeleitete These, dass eine klare Skalierung der Bewertungen zu einer Klärung von Bewertungskriterien führen kann und damit einerseits die Transparenz der Bewertungsprozesse steigert, andererseits aber auch mehr Verständigungsmöglichkeiten zwischen den Akteuren eröffnet. Auch diese Subkategorie nimmt eine dichotome Unterscheidung der Einschätzungen der Akteure vor und fragt, inwiefern die in dem Bewertungskonzept genutzten Skalierungen ‚*hilfreich*‘ oder ‚*nicht hilfreich*‘ für produktive Bewertungsprozesse waren. Aus den zuvor formulierten Ansprüchen des Bewertungskonzepts als auch aus den Ergebnissen der Gruppendiskussionen abgeleitet ist die Subkategorie der ‚*Praxistauglichkeit*‘, welche ebenso eine dichotome Qualitätseinschätzung für die Einschätzung des Feedbacks nutzt (‚*gegeben*‘ vs. ‚*nicht gegeben*‘). Ein zentraler Anspruch an das Bewertungskonzept, der sich aus den Gruppendiskussionen ableiten ließ, war die Frage, ob das Bewertungskonzept den Auszubildenden eine Hilfestellung für die ‚*Modellbildung*‘ von Unterricht bietet. Diese Subkategorie unterscheidet drei Unter Aspekte einer allgemeinen Modellbildung guten Unterrichts, nämlich die Fragen, inwiefern das Bewertungskonzept eine ‚*Hilfestellung für die Planung*‘, bzw. eine ‚*Hilfestellung für die Reflexion*‘ von Unterricht anbietet, sowie eine ‚*klare*

Operationalisierung‘ desselben bereitstellt, wie es auch dem in Kapitel zwei formulierten Anspruch an das Bewertungskonzept entspricht. Alle drei Subkategorien unterscheiden das Feedback der Akteure dahingehend, ob sie berichten, dass diese Hilfestellungen ‚*gelingen*‘, bzw. ‚*nicht gelingen*‘. Schlussendlich fasst eine Subkategorie die abschließenden ‚*Fazits*‘ der Akteure zur Erprobung des Bewertungskonzepts zusammen und fragt danach, ob diese ‚*positiv*‘ oder ‚*negativ*‘ ausfallen oder ggf. eine ‚*Beurteilung nicht möglich*‘ ist.

Abbildung 5.6: Subkategorien der Kategorie ‚Konzeptbewertung‘



Die Subkategorie der ‚*Passung mit fachlichen Standards*‘ wird in einem eigenen Unterkapitel ausgewertet. Dies geschieht, weil die den Ausbildern zugesicherte Anonymität bei der Auswertung des Fachbezuges nur noch eingeschränkt gegeben ist, sodass die Auswertung von Entwicklungsaspekten des Konzepts mit Fachbezug von anderen Aspekten, die teils auf die persönlichen Eigenschaften der Ausbilder zurückweisen, getrennt werden sollte. Diese Subkategorie greift den Anspruch auf, ein Bewertungskonzept zu entwickeln, das fachübergreifend in einer gesamten Ausbildungsinstitution eingesetzt werden kann und nicht nur von einzelnen Ausbildern in isolierten Fächern und trennt die Rückmeldungen der Akteure ebenfalls in eine dichotome Unterscheidung auf (‚*gegeben*‘ vs. ‚*nicht gegeben*‘).

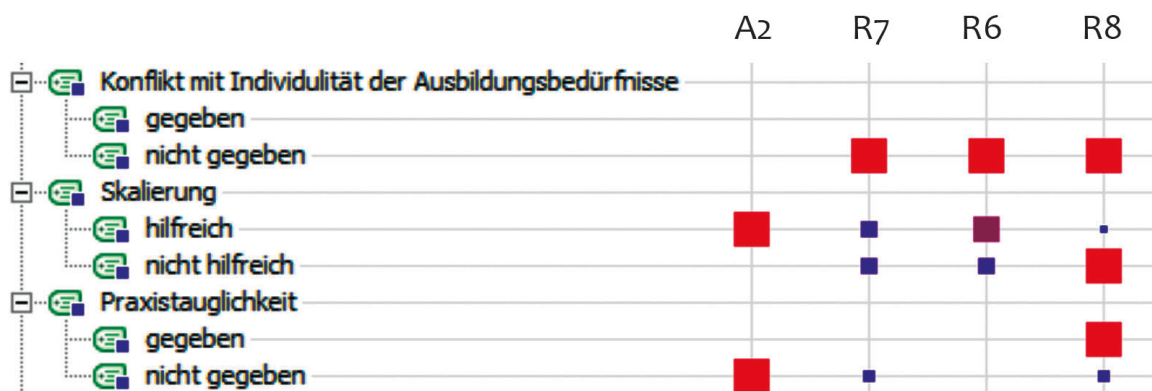
Die folgende Auswertung wird wieder getrennt nach den Ausbildern erfolgen, bei denen das Bewertungskonzept erprobt worden ist. Hierbei werden die Perspektiven der Ausbilder den Perspektiven der jeweiligen Auszubildenden gegenübergestellt. In einem abschließenden Fazit werden die Ergebnisse der drei verschiedenen Kontexte, in denen das Bewertungskonzept erprobt worden ist, zusammengeführt. Die Auswertung beginnt jeweils mit der gemeinsamen Betrachtung der drei dichotom codierten Subkategorien ‚*Praxistauglichkeit*‘, ‚*Skalierung*‘ und des ‚*Konflikts mit der Individualität der Ausbildungsbedürfnisse*‘. Es schließt sich die Analyse der Kategorien der ‚*Transparenz*‘ und der ‚*Modellbildung*‘ an. Die Auswertung für jeden Ausbildungskontext endet jeweils mit dem ‚*Fazit*‘ der Akteure. In allen Fällen dient eine Darstellung der Code-Matrix der jeweiligen Subkategorie(n) einem ersten Überblick über die verschiedenen Perspektiven des jeweiligen Ausbilders und der Auszubildenden. Die Subkategorie der ‚*Passung mit fachlichen Qualitätsstandards*‘ wird, wie oben begründet, in einem eigenen Unterkapitel erfolgen.

5.3.1 Konzeptbewertung bei A2

Die Abbildung 5.7 zeigt die Code-Matrix der drei dichotom codierten Subkategorien. Die Darstellung der Code-Matrix geschieht, abweichend vom Unterkapitel zur Konzeptnutzung, nun wieder zeilenweise, um die Unterschiede in der Konzeptrezeption zwischen den Akteuren herauszustellen. Die Code-Matrix

zeigt in den drei Subkategorien eine sehr fragmentierte Rezeption des Bewertungskonzepts, sowohl hinsichtlich der Unterschiede zwischen dem Ausbilder und den Auszubildenden, als auch auf der Ebene der Auszubildenden und in den intrapersonalen Bewertungen der einzelner Akteure.

Abbildung 5.7: Code-Matrix der Subkategorien ‚Skalierung‘, ‚Praxistauglichkeit‘, ‚Konflikt mit Individualität der Ausbildungsbedürfnisse‘



Von Ausbilderseite werden keine Angaben zu der Frage des ‚Konflikts mit der Individualität der Ausbildungsbedürfnisse‘ gemacht, wohingegen alle Auszubildenden hier angeben, keinen Konflikt erlebt zu haben. Die Rolle der ‚Skalierung‘ wird, zumindest der Code-Matrix folgend, relativ heterogen bewertet. Der Ausbilder betont in vier Passagen ihren Nutzen, zwei Auszubildende berichten ebenfalls in mehreren Passagen positiv über deren Einfluss, nennen aber auch negative Aspekte, und bei R8 überwiegt die negative Rezeption deutlich. Markantester Aspekt dieser drei Subkategorien ist aus der Perspektive des Ausbilder eine nicht gegebene ‚Praxistauglichkeit‘ des Konzepts, die er an insgesamt sechs Textstellen thematisiert. Die Einschätzungen auf Auszubildendenseite könnte heterogener nicht ausfallen: R6 äußert sich nicht zu dieser Thematik, R7 äußert in einer Interviewpassage, dass keine ‚Praxistauglichkeit‘ gegeben sei. R8 bewertet das Konzept als überwiegend praxistauglich, aber erwähnt zeitlich Einschränkungen von dieser.

Die deutliche Einschätzung der Auszubildenden bezüglich der Subkategorie des ‚Konflikts mit der Individualität der Ausbildungsbedürfnisse‘ zeigen auch die

die codierten Interviewpassagen *aller Auszubildenden* an, wie z.B. R6, die davon spricht, dass sie das Bewertungskonzept am Beginn der Ausbildung „zwar nicht so ganz durchschaut“ habe, es im weiteren Verlauf aber eine „Unterstützung“ gewesen sei:

I: „Hat Sie das Konzept eingeengt oder haben Sie den Eindruck gehabt, das gibt Ihnen die Möglichkeit, sich in Ihrer Lehrerpersönlichkeit zu entfalten?“

R6: Ja auf jeden Fall, also auf jeden Fall eher Entfaltung. Also eingeengt gefühlt habe ich mich gar nicht dadurch. Anfangs, als ich es nicht so ganz durchschaut habe, habe ich mir schon so ein bisschen Sorgen gemacht, aber dann nachher gar nicht mehr. Also ich habe es eher als Unterstützung gesehen, so als Hilfe: Da kann ich nochmal nachgucken, auf was ich achten muss und so was. Und ich fand es wirklich hilfreich, also eingeengt fand ich nee, fand ich überhaupt nicht.“

Ein Blick auf die codierten Passagen der Subkategorie ‚Skalierung‘ spiegelt eine überwiegend positive Bewertung der eingesetzten offenen Ratingskala wieder, sowohl im Bereich der ‚Reflexionskompetenz‘, als auch zur Bilanzierung der erreichten Kompetenzstufe bei der Realisierung der relevanten Qualitätsbereiche des Unterrichts. A2, R6 und R7 teilen diesbezüglich eine sehr ähnliche Einschätzung. A2 führt etwa aus, dass die Nutzung der Skalen für die Bilanzierung der Reflexionskompetenz für die Auszubildenden Transparenz hinsichtlich der bereits erreichten Kompetenzstufe herstellen konnten:

I: „Wie haben Sie diese Seite neun, diese Skalen zur Reflexionskompetenz, wie haben Sie das erlebt? Hatten Sie den Eindruck gehabt, das war gut, das hat mir auch geholfen um noch mal ein Fazit zu ziehen?“

A2: Das war gut, auf jeden Fall, glaube ich schon. Schafft auch Transparenz wiederum für den Referendar, damit er weiß: „A“ Ausbildungsstand, „E“ Examensstandard. Beides habe ich auch natürlich unterschiedlich dargestellt. Und, ja da kamen, da waren Referendare dabei, die waren schon relativ weit, und da ist der Ausbildungsstand sozusagen schon sehr

positiv zu bewerten und schon nah am Examensstandard dran. [...] Und insofern finde ich das hier, dass es inhaltlich hier in dieser Form, auch wie es hier gemacht wird, finde ich das ganz gut, dass der Referendar praktisch darin auch ausgebildet wird und Transparenz zurückgemeldet bekommt, wo er jetzt steht, weil das genau im Examen jetzt bedeutungsvoll auf ihn zukommt.“

Auffällig bei der Rückmeldung von R6 und R7 ist, dass beide die positive Würdigung der genutzten ‚Skalierung‘ schon als Antwort auf die offene Impulsfrage am Interviewanfang formulieren, deren positiven Nutzen für beide also eine offensichtlich prägende Erfahrung der Konzepterprobung war, die R7 z.B. wie folgt formuliert:

I: „Wenn Sie an die Bewertung mit den Instrumenten, mit dem Bewertungskonzept zurückdenken, was fällt Ihnen da spontan ein, was kommt Ihnen als erstes in den Kopf?“

R7: Das Erste wäre, dass es zeitlich aufwendig war, das ist mir aufgefallen zu erst, aber auch v.a. wie ich finde, eine Transparenz und eine bessere Übersichtlichkeit, besonders auf Grund des, ich nenne es mal Zeitstrahl, also diese Leiste, wo der Fachleiter dann eintragen konnte: Ausbildungsstand und Examensstand. Das fand ich sehr anschaulich und auch wirklich hilfreich um zu sehen, inwiefern man sich verbessert hat und wo man, wo der Weg hinführen muss um eine gute Note auch im Examen zu bekommen und was man gerade schon leisten kann im Ausbildungszustand.“

Hinsichtlich der ‚Skalierung‘ betont R7, dass diese „anschaulich und hilfreich“ sei. Ein Urteil, dass er sowohl auf die Bilanzierung der Reflexionsfähigkeit der Referendare bezieht, als auch auf die Bedeutung der Skalen für eine klare Diagnostik der realisierten Standards in den Qualitätsbereichen des Unterrichts. R6 teilt diese Einschätzung, da für sie die Skalen „auf einen Blick“ eine Antwort auf die Frage geben:

„R6: [...] Wo stehe ich, wo kann ich noch hin oder wo sollte ich vielleicht noch hin? Also das fand ich so zur Visualisierung eigentlich sehr hilfreich.“

R8 ist hinsichtlich der Rezeption der Skalen skeptischer. Sie ist ihnen nicht zwingend negativ gegenüber eingestellt, aber einen wirklichen Mehrwert in der Bilanzierung von Lehrproben stellen diese für sie auch nicht dar, sie „sollen es einfach noch mal visualisieren“, aber eine Hilfestellung ist dies für sie nicht:

R8: „[...] Ja es soll ja einfach noch mal visualisieren, quasi was darüber genannt wird, wie man das sieht. Ich denke, wie gesagt, das ist eigentlich, dass wenn irgendwas kritisiert wird, dass das halt noch nicht auf Examensstandard ist, ist es eigentlich klar. Deswegen, ja/Dieses „setzt nicht um“ und „voll und ganz“ ist ja auch ein bisschen schwammig, ist halt so wenig greifbar und da kann man nicht so richtig konkrete Handlungs /, keine konkreten Maßnahmen von ableiten sozusagen.“

Auch R7 äußert sich teils kritisch zu den Skalen, dies aber hinsichtlich ihrer Nutzung, bei der der Ausbilder die Bewertung vorgegeben und nicht dialogisch mit ihm entwickelt hat. Auf diesen Sachverhalt wird weiter unten bei der Weiterentwicklung des Konzepts zur Beta-Version eingegangen.

Die ‚Praxistauglichkeit‘ des Konzepts wird von A2 in vielen Interviewpassagen kritisch hinterfragt. Die grundsätzliche Struktur des Konzepts, wie die Bewertung von Unterricht ablaufen soll, hat nach Einschätzung von A2 eine hohe Relevanz für die Ausbildungspraxis, aber der Umsetzung in der Alpha-Version des Konzepts steht er skeptisch gegenüber:

A2: „[...] In der Durchführung der Nachbesprechung glaube ich, dass die Referendare damit stellenweise überfordert sind, mit der Komplexität überfordert sind. Und dass da, wenn sie selber ein Statement abgeben sollen, bezogen auf dieses Konzept und auch später in der Reflexion, die da anschließend drauf folgt in der Nachbesprechung, glaube ich schon, dass das möglicherweise wiederum zu differenziert ist, um da die Übersicht zu behalten. Das fängt schon mit formalen Bereichen an, das sind ganz viele Blätter, die man hier in der Hand hat.[...] Und ich glaube, dass die, dass es auch für Referendare dann in der Nachbesprechung an dieser Stelle unübersichtlich wird, also da finde, müsste reduziert werden an einigen Stellen, zumindest für die Referendare, es müssen ja nicht beide

das gleiche Konzept haben. [...] Und da ist es mir, glaube ich schon, zu unübersichtlich, aber das könnten die Referendare, ist eine Vermutung die ich habe, auch auf Grund der Erfahrung, dass das Konzept zwar auf dem Tisch gelegen hat, die Referendare hatten es auch dabei, aber die kamen überhaupt nicht dazu, auch nicht während der Besprechung, dann irgendetwas einzutragen.“

A2 kritisiert v.a. den Umfang des Konzepts, der aus seiner Perspektive die Auszubildenden überfordert sieht, wenn sie das Konzept tatsächlich gemäß seiner Intention nutzen sollten. Hierbei scheint A2 durchaus auch die inhaltliche Komplexität der im Konzept angebotenen Modellierung von gutem Unterricht zu meinen, wie die folgende Interviewpassage noch mal verdeutlicht:

A2: „[...] Auf der anderen Seite habe ich selbst bemerkt in den Gesprächen, die ich geführt habe, dass die Komplexität auch wiederum sehr groß ist. Wenn man also in einer konkreten Analysesituation steht, muss ja die Entscheidung getroffen werden, sowohl von Referendaren als auch von mir, es soll ja auf dialogischer Ebene stattfinden: Welche Qualitätsbereiche sind in dieser Stunde jetzt maßgeblich, gut oder möglicherweise auch kritisch zu betrachten? Und da finde ich, sind Referendare, glaube ich schon, zumindest am Anfang der Ausbildung überfordert, in dieser Komplexität jetzt selber zu erkennen, auf Grund ihrer geringen Erfahrung auch, was war jetzt das Wesentliche im Unterricht? D.h. ich als Experte bin da erheblich stärker gefordert, insofern wird das doch am Anfang der Ausbildung zumindest sehr stark auf mich darauf ankommen, was die Analyse der Stunde selbst betrifft. [...]“

„Zumindest am Anfang der Ausbildung“ ist der Anspruch, das Bewertungskonzept im Sinne einer Gesprächsgrundlage über den gesehenen Unterricht, für einen Dialog mit dem Auszubildenden zu nutzen, zu hoch gegriffen, so die Einschätzung von A2, die sich durchaus mit dem Feedback der Auszubildenden deckt. Es sei an das Zitat von R6 weiter oben erinnert, in dem diese sagt, dass ihr die Komplexität des Bewertungskonzepts am Anfang der Erprobung durchaus „etwas Sorgen gemacht“ habe. Eine Erfahrung von der auch R8 berichtet, bei der die Vielfalt des Qualitätsverständnisses guten Unterrichts und die umfangreiche Operationalisierung von diesem in der Matrix sogar „ein bisschen Angst“ ausgelöst hat:

R8: „Also ich finde halt diese Liste, die man halt davor hat, mit diesen ganzen Qualitätsbereichen (Anm: Die Matrix der Qualitätsbereiche), die finde ich halt sehr, sehr ausführlich. Sicherlich hilfreich und ich wüsste auch nicht was man da alles draus streichen könnte, aber es macht ein bisschen Angst, wenn man denkt, was alles man leisten soll, sozusagen in einer 45-Minuten Stunde die man zeigt, was da alles eine Rolle spielen kann, weil man kann ja einfach diese ganzen Punkte nicht, in einer Stunde sowieso nicht, auch wirklich denen gerecht werden, sage ich, mal und das macht einem so ein bisschen mulmiges Gefühl, wenn man denkt: Ok jetzt habe ich hier drei, vier Punkte, wo ich denke, die sind jetzt wichtig in der Stunde, aber da sind noch ganz viele, die fehlen da irgendwo. [...]“.

Spannend ist dieser Bericht von R8, da sie eine der Auszubildenden ist, die angab, trotz der Erprobung des Bewertungskonzepts, weiterhin das Qualitätstableau des ZfsL, sowie v.a. Bubbles genutzt zu haben, weil diese Instrumente, auf Grund ihrer reduzierten Komplexität, ihrer Meinung nach „einfach funktionaler“ gewesen sind:

I: Wäre jetzt eine Vermutung, aber haben Sie eine Vermutung warum sie intuitiv auf die Bubbles ausgeschwenkt sind? Warum Sie intuitiv dahin gegangen sind und nicht die Matrix genutzt haben?

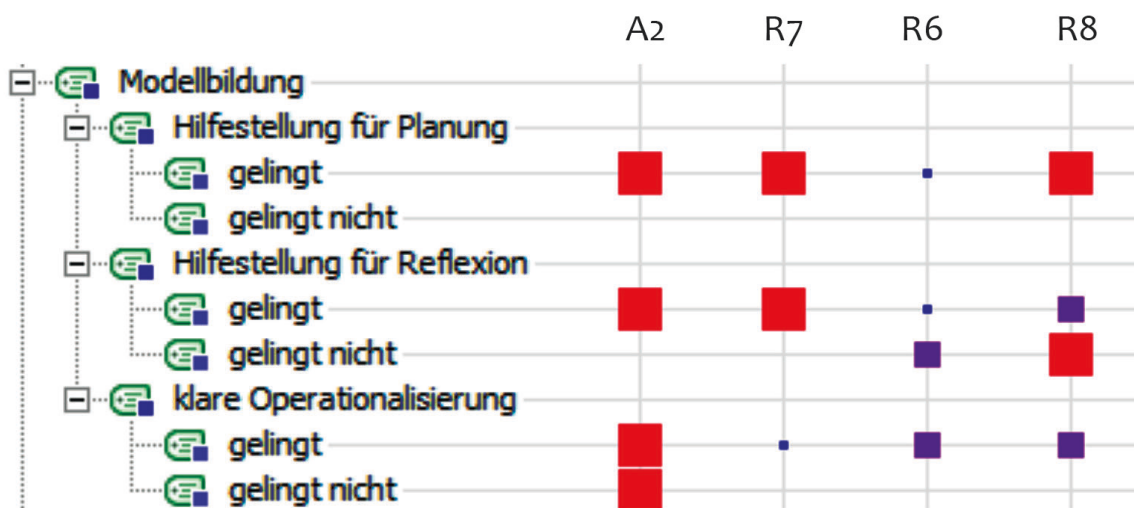
A8: Ich denke, weil dann vielleicht auch die Matrix zu ausführlich ist und dann doch der Überblick bei diesem anderen Blatt halt, in der Kürze der Zeit, die man nur hat, man kann sich ja dann nicht alle Bereiche auch durchlesen, wirklich einfach funktionaler gewesen ist, in dem Moment.

Fasst man die Auswertung dieser drei Subkategorien für die Konzepterprobung bei A2 zusammen, können folgende Schlüsse gezogen werden: Das Bewertungskonzept lässt offensichtlich einen ausreichend großen Rahmen für die ‚Individualität der Ausbildungsbedürfnisse‘ der Referendare, es engt sie nicht unzulässig ein. Die gewählte ‚Skalierung‘ der Instrumente scheint durchaus gewinnbringend zu sein um die Transparenz der Bewertungsprozesse zu erhöhen und die Professionalisierungsprozesse der Auszubildenden durch eine verbindliche Bilanzierung einzelner Kompetenzbereiche zu unterstützen. Denjenigen, denen die Skalierung

keine große Hilfe ist, fügt sie keinen Schaden zu. Im Sinne der ‚*Praxistauglichkeit*‘ ist eine Reduzierung des Umfangs des Konzepts anzustreben, ebenso wie es durchaus sinnvoll sein kann, Instrumente verschiedener Komplexitätsgrade zu haben, die den jeweils individuellen Bedürfnissen der Auszubildenden in den verschiedenen Phasen der Ausbildung entgegenkommen. Hierauf deutet die Nutzung der zwar strukturgleich aufgebauten, aber wesentlich weniger umfangreichen Instrumente des ZfsL Bonn, durch R8 hin.

Abbildung 5.8 zeigt die Code-Matrix für die Subkategorie ‚*Modellbildung*‘, die im Folgenden ausgewertet wird. Die Code-Matrix zeigt hier zunächst eine deutliche weniger fragmentierte Rezeption der Subkategorie, als es bei den drei zuvor diskutierten Subkategorien der Fall war. *Alle Akteure* sind sich einig, dass das Bewertungskonzept eine ‚*Hilfestellung für die Planung*‘ von Unterricht bietet, ebenso geben sie an, dass dem Konzept eine ‚*klare Operationalisierung*‘ guten Unterrichts gelinge. Lediglich A2 merkt an einer einzelnen Interviewpassage hierzu kritisch etwas an. Etwas uneindeutiger ist die Zahl der codierten Interviewpassagen im Fall der Subkategorie einer ‚*Hilfestellung für die Reflexion*‘, die das Konzept bieten soll. Neben positiven Anmerkungen aller Akteure finden sich bei R6 und R8 ebenso häufig negativ codierte Textpassagen, wie positive. Ein Blick in die qualitativen Daten kann diese Verwirrung, welche der rein deskriptive Blick auf die Codierhäufigkeiten mit sich bringt, klären.

Abbildung 5.8: Code-Matrix der Subkategorie ‚*Modellbildung*‘



Die Frage, inwieweit das Bewertungskonzept eine Hilfestellung für die Planung von Unterricht bietet, wurde von den Akteuren einhellig positiv beschieden. Insofern zeigen auch die codierten Interviewpassagen eine einhellig positive Einschätzung dieses Aspektes an. Von Ausbilderseite wird die diesbezügliche Leistung des Konzepts als „sehr positiv“ gelobt:

A2: „[...] Also insofern, wenn ich mir diese Matrix der Qualitätsbereiche angucke, dann ist das erst mal aus meiner Sicht sehr positiv, insbesondere auch für die Referendare in der Vorbereitung ihres eigenen Unterrichts, so wissen sie noch mal genau: Das können Beobachtungsmerkmale sein, das sind die Schwerpunkte, auf die es ankommt, und danach wird Unterricht möglicherweise beurteilt.“

Dieser Einschätzung schließen sich *alle Auszubildenden* an, wobei R7 hierbei sogar erwähnt, dass er insb. die Matrix der Qualitätsbereiche für die Vorbereitung des Unterrichts genutzt hat und diese anderen Instrumente, die ihm vom ZfsL ausgehändigt wurden, explizit vorgezogen hat, weil er wusste, „da sind die wichtigen Informationen drin“:

I: „[...] Haben Sie die im Vorfeld der Unterrichtsplanung herangezogen?“

R7: Wenn ich unsicher war dann habe ich dieses Blatt genommen und habe wirklich mal reingeschaut. Was macht auch guten Unterricht aus, um die Qualität des Unterrichts zu verbessern. Ich hatte so ähnliche Formulare oder Informationen auch von den anderen Fachleitern bekommen, aber durch die Menge an Materialien fehlt der Überblick und da wusste ich ganz genau, da sind die wichtigen Informationen drauf.“

Ebenso berichtet R8 über den hilfreichen Einfluss des Konzepts, insb. auch der Matrix, die ihr „Formulierungshilfen“ geboten und dabei geholfen habe, „bestimmte Punkte noch mal im Blick zu haben“:

I: „Ok, Sie haben jetzt gesagt: Es ist auch hilfreich für die Vorbereitung der Unterrichtsbesuche. Haben Sie da die Matrix oder andere Bereiche des Konzepts genutzt?“

R8: Ich habe da mal drauf geschaut, sicher. Ich habe auch bei der Unterrichtsvorbereitung, auch wenn es darum geht bei dem didaktischen Kommentar von dem Verlaufsplan, dass man da einfach Formulierungshilfen wirklich zum Teil hat, wenn man denkt: Ok stimmt, das ist ja eigentlich das, was ich meine. Dann hat man hier nochmal, wie gesagt, das ist vielleicht ein bisschen schöner formuliert, wie man es dann auch übernehmen kann, vielleicht auch in der Planung. Das fand ich ganz hilfreich, auf jeden Fall. Und dass man auch verschiedene Punkte noch mal im Blick hat und auch bedenkt bei der Planung.“

Hier zeigt sich eine frappierende Ähnlichkeit des Berichts von R8 mit dem Anspruch des Bewertungskonzepts, der aus den Gruppendiskussionen abgeleitet wurde, dass das Konzept im Sinne der Modellbildung guten Unterrichts den Auszubildenden – mit den Worten von A3 – eine „Sprachhilfe“ sein soll und dass es dazu dienen soll, einen „fokussierten Blick“ auf „die wesentlichen Aspekte“ von Unterricht zu erhalten.

Eine weniger einhellige Positionierung als die vorherigen Subkategorie suggeriert die Code-Matrix für die Subkategorie der ‚Hilfestellung für die Reflexion‘. A2 interpretiert die Wirkung des Bewertungskonzepts in diesem Bereich rein positiv und hebt dabei v.a. die Bedeutung der Matrix für eine fokussierte Nachbesprechung hervor:

I: „Aber wenn man das zusammenfasst, ein Fazit zieht: Durch die Matrix, durch das Instrument wird das Gespräch schon fokussiert, auf die wesentlichen Bereiche gelenkt?“

B2: Die Matrix ist aus meiner Sicht hervorragend geeignet, wenn es einem gelingt eine sinnvolle Reduktion bezogen auf die konkrete Stunde vorzunehmen. [...]“

Eine Einschätzung, der sich auch R7 anschließt:

R7: „Genau und durch die Matrix habe ich dann auch immer meine Auswahl getroffen hinsichtlich des Beratungsbedarfs und auch die

Übersicht vorher mir zu Herzen genommen um dann zu entscheiden, welche Bereiche, welche Qualitätsbereiche ich besprechen möchte mit dem Fachleiter, und das fand ich schon immer sehr gut und Probleme sind mir nicht aufgefallen.“

Während R6 die Einschätzung von A2 und R7 über den Nutzen der Matrix für eine fokussierte Nachbesprechung teilt und auch hier wieder die Bedeutung als Formulierungshilfe hervorhebt, wurde sie durch den Dokumentationsbogen eher verwirrt, was ihrer Ansicht nach aber v.a. mit dessen Einsatz zu tun hatte. Dieser Aspekt des Konzepteinsatzes wird weiter unten thematisiert. Für R8 war v.a. der Dokumentationsbogen für die Nachbesprechung eine Hilfestellung, da durch diesen „die Möglichkeit da ist, mal einen Vergleich herzustellen“ zwischen der Selbst- und der Fremdwahrnehmung durch den Ausbilder:

R8: „[...] Was halt sehr schön war hier, dass man sich selbst auch einschätzen konnte, dass die Möglichkeit da ist, mal einen Vergleich herzustellen: Wie schätzt denn der Prüfer einen ein, wie schätzt man sich selbst ein? Auch wenn das immer auch ein bisschen kritisch zu sehen ist, denke ich meinerseits, aber das war auf jeden Fall ein Unterschied. An sich war das sehr übersichtlich und gut strukturiert, eben auch dadurch, dass es wirklich auch konkret auf drei Punkte jetzt auch begrenzt werden sollte, die dann ganz gezielt fokussiert wurden und nicht dieser ganze, die große Liste von vielen Anforderungen, die man ja hat an guten Unterricht, ganz gut runter gebrochen denke ich, durch diesen Bogen.“

Der Hinweis von R8, dass es für sie eine Hilfestellung war, dass der Dokumentationsbogen für die Nachbesprechung die „vielen Anforderungen“ an guten Unterricht auf drei Qualitätsbereiche „heruntergebrochen“ hat, verweist auf ihre eher kritische Rezeption der Matrix als Reflexionshilfe. Wie bereits weiter oben ausgeführt, griff sie für die Vorbereitung der Unterrichtsnachbesprechung auf die Instrumente des ZfsL zurück, weil diese weniger umfangreich und für sie deswegen „funktionaler“ waren.

Ein einhelligeres Bild der Konzeptwahrnehmung zeigt sich wiederum bei der Subkategorie einer ‚klaren Operationalisierung‘ guten Unterrichts, die das Kon-

zept laut *allen Akteuren* leisten konnte und somit die „Komplexität des Unterrichts“ abbilden, wie A2 es formuliert:

A2: „[...] Also zuerst mal habe ich das als Kommunikationsgrundlage verstanden, dieses Konzept. Und da ist erst mal positiv zu bewerten, aus meiner Sicht, dass hier klare Qualitätsbereiche vorgegeben werden. Das finde ich, ist sehr gut gelungen, an Hand der Auflistung der einzelnen Qualitätsbereiche. Die Bereiche, es sind ja glaube ich insg. 14, decken sehr gut die Komplexität des Unterrichts ab.[...] So, und damit ist auch für die Referendare, die das ja auch in die Hand bekommen, eine große Transparenz gegeben, was wird von ihnen im Unterricht erwartet? Was wird erwartet, wenn guter Unterricht durchgeführt werden soll? Das ist das Positive. Auch die Operationalisierung ist ja hier wiedergegeben, und das schafft auch wiederum Klarheit, wonach es gemessen wird.“

Alle Auszubildenden teilen diese Einschätzung, wie aus den codierten Interviewpassagen hervorgeht. R6 führt aus, welche Bedeutung diese ‚klare Operationalisierung‘ für den eigenen Professionalisierungsprozess hatte:

„R6: „[...] Ich fand halt gut, dass dies in diesen einzelnen Tabellen, also mit diesem Bewertungsraster da, mit diesen Qualitätsbereichen, die Matrix, das fand ich eigentlich sehr hilfreich, weil das nochmal genauer aufgeschlüsselt war. [...] Weil das mir dann leichter gefallen ist bei manchen Wörtern, wo ich natürlich im Prinzip weiß, was es ist aber vielleicht jetzt auf meine Stunde speziell gedacht hätte, das habe ich da gar nicht so untergebracht, aber an Hand der Erklärung dann gesehen habe: Ach so stimmt, das spielt da ja noch rein und das auch. Und von daher fand ich das sehr nützlich, auf jeden Fall, doch.

I: Also hat Ihnen geholfen nochmal auch zu sehen, was Sie schon können, wenn ich das richtig verstehe, was Sie eh schon berücksichtigt haben?

R6: Genau, das war das. Also, dass ich einfach gesehen habe, so eine Vernetzung mehr, fand ich da auch noch mal deutlich. Dass man gesehen hat, womit ein bestimmter Begriff einfach, also mit wie vielen Dingen der verbunden wird einmal und dann aber auch, manchmal war es dann

tatsächlich so, dass man sich gedacht hat: Ach so, das mache ich ja schon und das mache ich auch schon. Das war dann so ein bisschen die Beruhigung: Ich muss jetzt nicht noch zehn neue Dinge irgendwie integrieren, sondern vielleicht nur zwei und der Rest ist irgendwie schon, ja so ein bisschen ritualisiert, also es läuft schon so halt automatisch ab.“

Durch die Operationalisierung guten Unterrichts im Rahmen der Matrix hat R6 „so eine Vernetzung“ der verschiedenen Qualitätsbereiche erkannt, also genau jene Perspektive eingenommen, die von den Ausbildern im Rahmen der Gruppendiskussion immer wieder als zentraler Aspekt gelingender Professionalisierung genannt wurde.

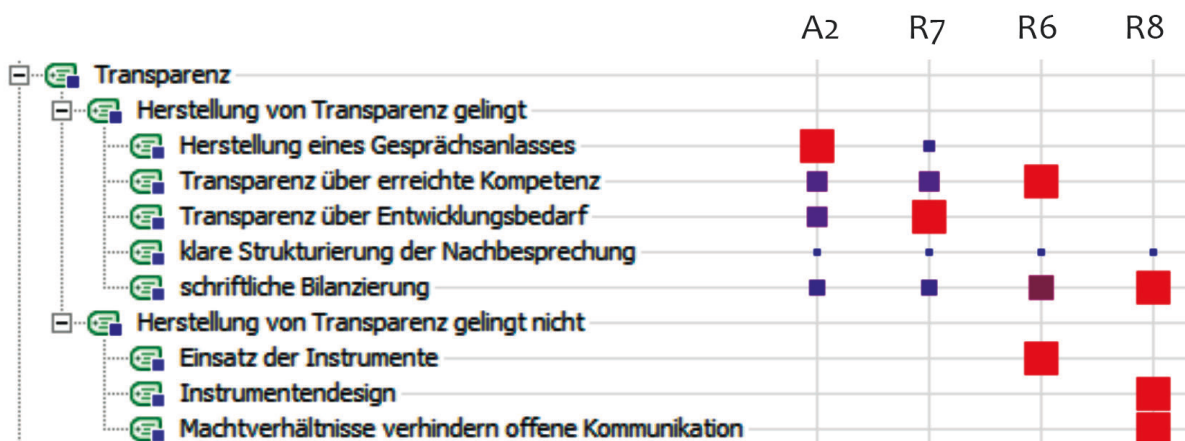
Kritisch kommt A2 noch mal auf den Aspekt der Systematik, bzw. Hierarchie der verschiedenen Qualitätsbereiche zu sprechen, ein Diskussionsaspekt, über den schon in der Gruppendiskussion der Ausbilder leidenschaftlich diskutiert wurde und über den kein Konsens erzielt werden konnte. A2 vertritt, wie in der Gruppendiskussion, die Auffassung, dass der Qualitätsbereich der ‚Klarheit und Strukturiertheit‘ hierarchisch über die anderen zu stellen sei, da ohne die Gewährleistung von diesem, kaum Lernerfolge zu erzielen seien.

Zieht man für diese Subkategorie im Rahmen der Erprobung des Konzepts bei A2 ein Fazit, dann fällt dieses relativ positiv im Sinne des Bewertungskonzepts aus. Das Konzept scheint aus der Perspektive aller Akteure dazu geeignet zu sein, eine ‚Modellbildung‘ guten Unterrichts im Sinne der drei Subkategorien zu leisten, welche die Auszubildenden bei dem Erwerb von Unterrichtskompetenz unterstützt. Es fallen Erfahrungsberichte der Auszubildenden über den Nutzen der Instrumente auf, die beinahe den Wortlaut der Forderungen der Ausbilder aus den Gruppendiskussionen treffen. Einzelne kritische Punkte betreffen wiederum den Umfang der Instrumente, insbesondere der Matrix, was wiederum auf den bereits im Fazit des vorangegangenen Unterkapitels formulierten Punkt verweist, dass es Instrumente mit differenzierendem Umfang geben sollte, um den unterschiedlichen Bedürfnissen der Auszubildenden entgegen zu kommen.

Auf den zweiten Blick etwas verwirrend ist die positive Würdigung der Matrix, die laut A2, R6 und R7 zwar hilfreich für die Vorbereitung des Statements war, in der Nachbesprechung aber selten herangezogen wurde wie, in der Analyse der Kategorie des ‚Konzepteinsatzes der Ausbilder‘ deutlich herausgearbeitet wurde.

Die Code-Matrix für die Subkategorie der ‚Transparenz‘ (vgl. Abb. 5.9) zeigt für den Erprobungskontext von A2 eine deutliche Mehrzahl von Codings im positiven Bereich der Subkategorie, sowie eine relativ gleichwertige Sichtweise des Ausbilder und der Auszubildenden. Die Interviews von A2 und R7 weisen keine Codings auf der negativen Code-Seite auf, dafür aber Passagen in allen positiven Subkategorien. Jene von R6 und R8 zeigen zumindest teilweise auch eine negative Rezeption des Bewertungskonzepts, hinsichtlich seiner Leistung die ‚Transparenz‘ der Bewertungsprozesse zu erhöhen. Auf der positiven Seite der Subkategorie scheint die Kernleistung des Konzepts in einer ‚klaren Strukturierung der Nachbesprechung‘ und einer ‚schriftlichen Bilanzierung‘ der erfolgten Bewertung zu liegen, über die jeweils alle Akteure berichten. Die ‚Herstellung eines Gesprächsanlasses‘ und die Sicherstellung von ‚Transparenz über den Entwicklungsbedarf‘ finden sich als codierte Passagen jeweils nur in den Berichten von A2 und R7 wieder; diese Leistung des Konzepts scheint also weniger augenscheinlich zu sein als die zuvor aufgeführten. Die Herstellung von ‚Transparenz über die erreichte Kompetenz‘ scheint wiederum aus der Perspektive von drei der vier Akteure zu gelingen.

Abbildung 5.9: Code-Matrix der Subkategorie ‚Transparenz‘



Die ‚schriftliche Bilanzierung‘ der Bewertung sehen alle Akteure als eine wesentliche Leistung des Konzepts an. Hierbei spielt, wie es zu erwarten war, der Dokumentationsbogen für die Nachbesprechung eine Hauptrolle. Er „schafft Transparenz und reduziert auf die wesentlichen Aspekte“, so A2:

I: „Diese Dokumentation des Gesprächs im Nachhinein an die Referendare aber zu geben, würden Sie das als gewinnbringend, als sinnvoll erachten?“

A2: Auf jeden Fall, das schafft wiederum Transparenz, es reduziert auf die wesentlichen Aspekte. Das ist auf jeden Fall nochmal eine große Hilfe. Und gerade am Ende ist ja noch mal so eine Art Zusammenfassung, hier auf der Seite 10: «Aus dieser Stunde sollen folgende Aspekte usw., dauerhaftes Handlungsrepertoire», da werden ja auch die positiven Aspekte aufgelistet und unten drunter die kritischen Aspekte. Das sind genau die Punkte, die i. W. das Gespräch bestimmt haben. Und insofern schaffen sie wiederum Transparenz und auch noch mal Klarheit.“

Dies ist eine Einschätzung, der sich *alle Auszubildenden* anschließen und einhellig betonen, dass der Dokumentationsbogen des Bewertungskonzepts einen deutlichen Mehrwert gegenüber den bisher praktizierten Formen der Dokumentation des ZfsL bot. Für R6 scheint dieser Gewinn durch das Bewertungskonzept der prägende Eindruck der Konzepterprobung gewesen zu sein, da sie dies als Antwort auf die offene Impulsfrage des Interviews formuliert:

I: „Wenn Sie auf den Einsatz dieses Bewertungskonzepts zurückblicken, was fällt Ihnen da spontan als erstes ein, was kommt Ihnen da in den Kopf als aller Erstes?“

R6: [...] Ich fand es einfach gut, dass ich einfach mal schwarz auf weiß was hatte, wie es abgelaufen ist und ich das auch noch mal nachlesen konnte. Weil ich fand immer/Es war vorher in den Nachbesprechungen so, dass man, man hat halt viele Informationen bekommen, die Nachbesprechung war auch irgendwie produktiv, aber man ist danach eh nicht so ganz aufnahmefähig, man hat halt vieles mitgeschrieben. Aber ich fand es so, man hatte es einfach noch mal ganz klar auch strukturiert nach einzelnen

Fragen, man konnte noch mal nachgucken, wie war das jetzt mit dem Methodeneinsatz, so spezielle Aspekte.“

Für R8 ist der Dokumentationsbogen des Bewertungskonzepts „irgendwie nachhaltiger angelegt“ als jener des ZfsL, da er eine klarere Bilanzierung der Bewertung vornehme. R6 verweist darüber hinaus auf den positiven Nutzen der Sammlung über das Repertoire, auch wenn ihr nur die Übersichtstabelle des Instruments ausgehändigt wurde, so konnte sie durch diese sehen, dass „Sachen zwischendurch auch gut waren“ und „nicht alles schlecht“ war, diese also eine durchaus bedeutsame Rolle bei der Aufrechterhaltung ihrer Motivation gespielt hat:

I: „Hat Ihnen diese Sammlung über das Repertoire nochmal geholfen mehr Transparenz zu kriegen über Ihre Entwicklung?“

R6: [...] Und das fand ich eigentlich noch mal gut zu sehen auch, weil da ja auch dann durch verschiedene Methoden oder was weiß ich, was man ausprobiert hat, einfach noch mal, ja deutlich gemacht wird, dass man doch schon ein paar verschiedene Sachen probiert hat und die dann auch zwischendurch auch mal gut waren und nicht alles schlecht war. Und von daher fand ich das schon ganz gut, ja.“

R7 verweist, als Reaktion auf die Frage nach Änderungen der Bewertungspraxis durch die Konzepterprobung einerseits ebenfalls auf die Bedeutung der schriftlichen Dokumentation der Bewertung, gleichzeitig aber auch darauf, dass der Ablauf der Nachbesprechung für ihn „einfach transparenter“ war:

I: „Ich meine beim Ausbilder, im Vergleich der Lehrproben, die vorher waren.“

R7: Vorher hatte ich nur eine, die frei lief mit dem xy zusammen, aber was ich jetzt nachträglich finde, ist einfach diese Übersichtlichkeit, dass man ein Formular bekommt, wo man ganz genau weiß, wo steht man, was ist zu verbessern, was lief gut. Ist einfach übersichtlicher und auch der Ablauf des Gesprächs war für mich transparenter, wann ist was von

mir gefordert, was läuft in der Nachbesprechung ab? Das war für mich insgesamt übersichtlicher.“

Diese ‚klare Strukturierung der Nachbesprechung‘ ist ebenfalls ein Aspekt, den alle Akteure durch die Erprobung des Konzepts erlebt haben, wobei R8 auch bei diesem Aspekt auf positive Veränderungen der Bewertungspraxis durch die Konzepterprobung verweist, denn ein „ausführliches Feedback“ bekomme man „halt ansonsten nicht in dieser Form“:

I: „Ich würde gerne von Ihnen einfach wissen, wenn Sie an den Einsatz, die Erprobung dieses Bewertungskonzepts zurückdenken, was fällt Ihnen da spontan als Erstes ein?“

R8: Ja, eine sehr gute Strukturiertheit in der Besprechung, der Nachbesprechung an Hand dieses Dokumentationsbogens. Also man konnte wirklich ganz gezielt einige Aspekte sich anschauen aus diesem Wust von Anforderungen an guten Unterricht, sage ich mal. [...]“

Angesichts der Tatsache, dass R8 diesen Aspekt als Antwort auf die offene Impulsfrage des Interviews formuliert, scheint der Zugewinn an Struktur in der Nachbesprechung für sie eine der prägenden Erfahrung der Konzepterprobung zu sein.

Eine besondere Rolle bei der Gewährleistung von ‚Transparenz über die erreichte Kompetenz‘ der Referendare spielt nach Aussage von A2, R6 und R7 die Auflistung jener Handlungsaspekte der gezeigten Unterrichtsstunde, die von den Auszubildenden in eine allgemeine Handlungskompetenz übernommen werden sollten, und die auf Seite fünf des Dokumentationsbogens zu vermerken waren. Durch diese schriftliche Fixierung würden, wie es A2 in der bereits oben zitierten Interviewpassage ausdrückt, die „wesentlichen Aspekte“ des Unterrichtsbesuches hervorgehoben und damit der Professionalisierung der Auszubildenden zugeführt. Für R6 hat die Auflistung von Aspekten einer bereits ‚erreichten Kompetenz‘ wiederum eine große Bedeutung für das Aufrechterhalten der Motivation, denn die positiven Aspekte des eigenen Kompetenzfortschrittes könne man als Referendar „nicht oft genug hören“:

R6: „[...] Aber ich glaube, für uns als Referendare, das kann man nicht oft genug hören, weil gerade, wenn man aus der Nachbesprechung rausgeht und sich eben so inhaltliches oder so was eher so im negativen Bereich, zumindest gefühlt, anordnet ist es glaube ich ganz wichtig, dass man aber dann auch mal noch mal sagt: Ja aber das und das läuft halt auch schon gut. Und dafür fand ich halt speziell diese Strukturierung auf diesem Bogen sehr sinnvoll, dass man einfach gesehen hat: Das ist aber auch alles noch da, das ist nicht weg, auch wenn ich inhaltlich vielleicht noch einiges irgendwie verbessern kann.“

Für A2 hilft die Auflistung des Entwicklungsbedarfs auf der gleichen Seite des Dokumentationsbogens bei der Herstellung von ‚*Transparenz über den Entwicklungsbedarf*‘, eben weil so die „wesentlichen Aspekte“ der Stunde auch auf der negativen Seite der Professionalisierung noch einmal fixiert würden. R6 und R8 äußern sich nicht zu dieser Subkategorie, aber R7 schließt sich der Einschätzung von A2 an, da ihm die Verschriftlichung „Möglichkeiten der Entwicklung“ gegeben haben:

I: „Haben Sie den Eindruck gehabt, dass die Instrumente dazu beigetragen haben die Bewertungen besser nachvollziehen zu können?“

R7: Ja, auf jeden Fall. Das fand ich schon, gerade auch auf Grund der letzten Seiten, wo aufgeschrieben wurde, genau Verbesserungsbedarf oder Möglichkeiten der Entwicklung, die haben mir immer schon auch geholfen und ja, positiv.“

Zumindest zwei der Befragten im Erprobungsbereich von A2 äußern explizit, dass das Bewertungskonzept dabei helfen kann, ‚*Gesprächsanlässe*‘ zwischen dem Ausbilder und den Auszubildenden zu generieren und damit einen Beitrag zur Realisierung des Dreh- und Angelpunktes gelingender Bewertung zu leisten. A2 hat das Bewertungskonzept als „Kommunikationsgrundlage“ verstanden, wie er es von sich aus am Beginn des Interviews formuliert, durch die ein gemeinsames Verständnis von gutem Unterricht entwickelt werden kann, wobei offensichtlich die Matrix eine bedeutsame Rolle erfüllt:

I: „[...] Haben Sie erlebt, dass durch diese Qualitätskriterien ein gemeinsames Verständnis von guten Unterricht entwickelt werden konnte und auch für beide klar war?“

A2: Das glaube ich schon. Wie gesagt, das Problem ist es manchmal, die Auswahl zu treffen, was war für die Stunde jetzt relevant, aber wenn man da jetzt ein Agreement gefunden hat in der Auswahl der Qualitätsbereiche, dann kann man wunderbar hier die Struktur hinlegen. Die Operatoren sind ja da und dann kann man auch markieren, mit Farbstiften z.B. und das habe ich ja teilweise auch gemacht, ganz klar sagen: Das sind die einzelnen Punkte innerhalb dieses Qualitätsbereiches, die Operatoren und an den Beispielen ist es dann klar gemacht worden. Also das ist auf jeden Fall positiv zu bewerten.“

Auch die Skalierung des Dokumentationsbogens hat scheinbar eine Rolle dabei gespielt, ‚Gesprächsanlässe‘ zu generieren. Wenn die Einordnung des Unterrichts aus der Perspektive von Ausbilder und Auszubildenden in unterschiedlichen Bereichen der Skala verortet wurde, „dann haben wir eben noch mal darüber gesprochen“:

A2: „[...] Also bei dieser Einschätzung gab es im Gespräch mit den Referendaren in der Regel nie ein Problem. Da waren wir in der Regel auf gemeinsamer Linie. Und wenn es da einen Unterschied gab, dann haben wir eben da noch mal drüber gesprochen, und ich habe dann noch mal versucht zu erklären, warum ich das doch ein bisschen weiter nach links oder recht ansetze. In der Regel geht es ja darum, wenn der Abstand größer ist, da ist ja eher Diskussionsgrundlage, wenn es positiver ist, da ist ja in der Regel keine Diskussionsgrundlage.“

Auch R7 berichtet davon, dass durch das Instrumentendesign Klärungsprozesse angestoßen wurden:

I: „Sie haben jetzt gerade schon gesagt: «Gemeinsam den Bogen ausgefüllt.» Haben Sie das wirklich dialogisch gemacht oder hat Ihnen der Ausbilder gesagt, was er da rein schreibt und das erklärt?“

A7: Also, ich kann mich noch an einige Situationen erinnern, dass ich auf jeden Fall meine Meinung zu diesen Aspekten nennen konnte und der Fachleiter das dann auch berücksichtigt hat, dass ihm das vorher nicht so klar war, weil es in meinem Entwurf nicht stand und dass er dann das nochmal berücksichtigen konnte und das dann auch verändert hat. Das fand ich auch gut.“

Neben diesen überwiegend positiven Berichten der Akteure über den Einfluss des Bewertungskonzepts, sprechen R6 und R8 auch über Aspekte, die verhindert haben, dass das Konzept zur Herstellung von ‚Transparenz‘ beitragen konnte. R8 bewertet den Beitrag des ‚Instrumentendesigns‘ für die Herstellung von Transparenz skeptischer als die anderen Akteure, v.a. als A2 und R7. Die Nutzung der ‚Skalierung‘, das wurde bei der Auswertung dieser Subkategorie schon deutlich, sorgt aus ihrer Perspektive für keine Steigerung der ‚Transparenz‘. Während R7, wie oben dokumentiert, darüber berichtete, dass durch die Skalen Klärungsprozesse angestoßen wurden, ist die Aussage dieser für R8 „noch ein bisschen schwach“. V.a. die Unterscheidung zwischen einem Ausbildungs- und Examensstand, die von R6 und R8 lobend hervorgehoben wurde, erscheint ihr „ein bisschen schwammig“:

I: Ich sehe ja auch, dass dieser Unterschied zwischen Ausbildungs- und Examensstandard da vermerkt wurde. Hatten Sie den Eindruck, dass das schlüssig erläutert werden konnte, warum es da einen Unterschied gibt oder hatten Sie den Eindruck das wird eher willkürlich gezogen?

R8: Diese Sache ist, dass man nicht so richtig genau weiß, was ist denn jetzt der Examensstandard? Ein Ausbildungsstandard, da kann man sich jetzt besser drauf verorten, aber ein Examensstandard ist, glaube ich immer, von wegen dass es besser werden muss noch. Es ist irgendwie auch eigentlich auch klar, wenn man nicht gerade „voll und ganz“ schon eingeschätzt wird, deswegen finde ich das auch ein bisschen, ja schwammig einfach, so von der Distanz aus. Macht/Hat sicher/Visualisiert das ein bisschen, das, was er vielleicht meint, aber es ist/ Von der Aussagekraft finde ich es doch ein bisschen, vielleicht noch ein bisschen schwach.“

R8 spricht die bereits aus der Auswertung der Gruppendiskussionen bekannte Problematik an, keine Transparenz über die Ausbildungsprogression zu haben, welche die Ausbilder in die Bewertungsprozesse einbeziehen. Während sich diese Problematik für R6 und R7 durch die Klärungsprozesse, die durch den Dokumentationsbogen und insb. die Skalierung ausgelöst werden, entschärft, bietet das Konzept für R8 keine Hilfestellung.

Weitergehend berichtet R6 darüber, dass es trotz des Einsatzes des Dokumentationsbogens offensichtlich zu Unklarheiten über den Schwerpunkt der Nachbesprechungen kam. Hierbei verweist sie darauf, dass ein anderer ‚Einsatz der Instrumente‘ diese Problematik zumindest hätte deutlich machen können, wenn Ausbilder und Auszubildende sich über die ausgefüllten Dokumentationsbögen ausgetauscht hätten, das alleinige Ausfüllen aber offensichtlich keine ausreichenden Abgleich der Perspektiven ermöglichte:

R6: „[...] Manchmal war es so, dass ich es deswegen nicht nachvollziehen konnte, weil in der Nachbesprechung halt auf ganz viel erst mal eingegangen wurde, also es war manchmal nicht so ganz schwerpunktmäßig, und dann war im Dokumentationsbogen, waren da natürlich jetzt keine Punkte, die nicht genannt wurden, aber manchmal war es so, dass man sich gewundert hat, dass das jetzt so, sage ich mal, der Schwerpunkt ist. Also, ich habe es halt einmal auch selbst ausgefüllt um das mal zu vergleichen, wir hatten das eigentlich auch so abgesprochen, dass wir das mal machen. Im Endeffekt/Ich habe es ein-, zweimal gemacht, aber wir haben nicht drüber gesprochen. Das war halt nicht so, das wurde auch nicht verglichen. War so angedacht und ich weiß, einmal habe ich halt ausgefüllt und da habe ich mich echt gewundert, wie weit das auseinander ging, also so von den Schwerpunkten.“

Ein ebenfalls aus der Auswertung der Gruppendiskussionen bekannter Aspekt ist derjenige, dass die hierarchischen Beziehungen zwischen Ausbildern und Auszubildenden eine offene Kommunikation und damit die Herstellung von Transparenz behindern. R8 berichtet über ihre Skepsis, auf den Skalen des Dokumentationsbogens eine offene Einschätzung ihrer gezeigten Kompetenz abzugeben,

weil sie befürchtet, dass dies negativ in die Bewertung einfließt, demnach aus ihrer Perspektive die ‚Machtverhältnisse offene Kommunikation verhindern‘:

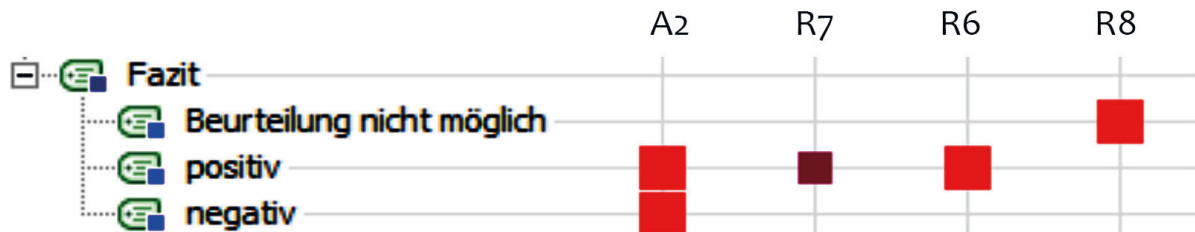
R8: „[...] Das ist auch der Punkt, den ich sehr kritisch finde, weil man weiß nicht so richtig/Man ist in einer Bewertungssituation, ob diese Einschätzung wenn man sich selbst komplett falsch einschätzt beispielsweise, ob das dann irgendwie auch negativ mit in die Bewertung dann einfließt. Das finde ich halt das Schwierige, dass man selbst auch da irgendwo da eintragen soll. Ist zwar schön, dass man sich einbringen kann, aber durch diese Bewertungssituation finde ich das ein bisschen schwierig.“

Wirft man einen zusammenfassenden Blick auf die Auswertung der Subkategorie der ‚Transparenz‘ im Rahmen der Erprobung von A2, dann erscheint die Konzepterprobung ein relativ positives Bild des Einflusses des Konzepts zu liefern. Die Instrumente des Konzepts, allen voran der Dokumentationsbogen und auch die von ihm genutzte Skalierung, tragen zu einer Steigerung der Transparenz bei. Hierbei erwähnen alle Auszubildenden Vorteile des erprobten Konzepts gegenüber den bisher praktizierten Formen der Dokumentation, da es dem Konzept gelänge, Klärungsprozesse im Sinne eines ‚Gesprächsanlasses‘ zu generieren. Die Betonung bereits erreichter Kompetenzfortschritte ist für R6 eine wichtige Quelle der Motivation, die durch den Einsatz des Konzepts sichergestellt wird – eine vor dem Hintergrund des Stresspotentials der zweiten Phase sicherlich wichtige Erkenntnis. Vereinzelt Kritikpunkte finden sich in der bereits bekannten Auffassung von R8, dass die Skalen keine große Hilfe bei der Sicherstellung von ‚Transparenz‘ leisten, sondern bestenfalls „schwammige“ Auskünfte erteilen, sowie der erneuten Problematisierung, dass die Herstellung von ‚Transparenz‘ in Bewertungsprozessen nur bis zu einem bestimmten Grad möglich sei, wenn die Interaktionssysteme, in denen diese ablaufen, derart hierarchisch strukturiert seien.

Wie die Code-Matrix (vgl. Abb. 5.10) zeigt, fällt das ‚Fazit‘ der Akteure im Erprobungsbereich von A2 überwiegend positiv aus. Lediglich R8 gibt, an eine abschließende und umfassende Beurteilung des Konzepts nicht leisten zu können, weil sie es dafür „nicht intensiv genug“ für die Vorbereitung ihrer Unterrichtsbesuche genutzt habe. Die anderen beiden Auszubildenden geben abschließend

eine rein positive Bewertung ab und attestieren dem Konzept, eine Hilfestellung im Hinblick auf erfolgreiche Unterrichtsbesuche geleistet zu haben.

Abbildung 5.10: Code-Matrix der Subkategorie ‚Fazit‘



R7 hebt hierbei die Unterstützung des Konzepts sowohl für die Planung von, als auch der Reflexion über Unterricht hervor:

I: „Haben Sie den Eindruck, dass das Konzept Ihnen insgesamt eine Hilfestellung geliefert hat um erfolgreiche Unterrichtsbesuche zu machen?“

A7: Ja, zum Einen fand ich es hilfreich im Vorfeld, weil ich ganz genau wusste, wo ich die Informationen herbekomme für qualitativ hochwertigen Unterricht und zum anderen habe ich das für die Nachbesprechung gut nutzen können um mir dann nochmal über die Qualitätsbereiche, mir drei herauszusuchen um sie dann, mich darauf im Statement vorzubereiten.“

Eine ähnliche Bewertung gibt auch R6 ab, die aber noch hinzufügt, dass die schriftliche Dokumentation nach einem einheitlichen Muster eine wichtige Hilfestellung für sie war:

I: „Wenn Sie ein Fazit ziehen: Haben Sie den Eindruck, dass das Konzept dazu beigetragen hat die Ausbildungskontexte, in denen es eine Rolle gespielt hat, positiv zu beeinflussen?“

R6: Ja, insgesamt würde ich es auf jeden Fall positiv bewerten. Wie gesagt, u.a. für mich, für die Planung hat es unheimlich geholfen und auch nachher einfach wirklich noch mal zusammengefasst so eine Information zu haben über die einzelnen Lehrproben. Weil, das fand ich vorher, das hat auf jeden Fall gefehlt. Man ist rausgegangen, man hatte mündlich irgendwas gehört, aber man hatte einfach nichts, nichts wo man nochmal nachgucken konnte.“

A2 bewertet das Konzept insgesamt ambivalenter als die Auszubildenden und hebt einzelne positive Aspekte, als auch negative Aspekte des Konzepts hervor. Besonders positive Erfahrungen hat A1 mit der Matrix der Qualitätsbereiche, sowie dem Ansatz, die Kompetenz als Expertise über die Reflexionsfähigkeit zu erfassen, gemacht. Aus seiner Perspektive haben sie entscheidenden Einfluss auf eine strukturierten Nachbesprechung gehabt:

I: „Haben Sie das (Anm.: Gemeint ist der Einsatz des Bewertungskonzepts, D. L.) als positiv erlebt?“

A2: Positiv, ja auf jeden Fall positiv. Allerdings bin ich jemand, der vorher auch schon versucht hat, sehr strukturiert die Nachbesprechung an Hand dieser Qualitätsbereiche, die dann eben anders genannt werden, Qualitätskriterien oder wie auch immer, schon durchzuführen. Aber es war trotzdem eine Hilfe, insbesondere auch die Reflexionskompetenz, dass die nochmal ganz deutlich in den Blick genommen wird, weil das ja auch, das kommt ja noch dazu, das haben wir vorher vielleicht nicht so bewusst getan oder ich nicht so bewusst getan, weil das ja auch ein Aspekt ist, der im Examen eine ganz wichtige Rolle spielt.“

Kritisch sieht A1 die Komplexität und den Umfang des Konzepts, die einen seiner Meinung nach zu hohen Aufwand bedingen, sich in das Konzept einzuarbeiten, was ihn zum dem Schluss leitet, „dass die Komplexität, die Vielfalt dieser Zettel, rein formal einfach reduziert wird.“ Ebenso würdigt A1 die Sammlung über das Repertoire kritisch, die seiner Meinung nach zu oberflächlich ist, um eine Abbildung der Kompetenzentwicklung der Auszubildenden leisten zu können:

I: „Hatten Sie den Eindruck, dass Sie durch, dadurch, dass Sie diese Sammlung über das Repertoire noch mal zusammengefasst haben, dass für Sie nochmal die Kompetenzentwicklung der Referendare, Ihnen noch mal klarer geworden ist?“

A2: (Gleich?) Ich glaube, dafür ist es wieder zu grob im Grunde genommen. Man hat eine Übersicht, welche Kompetenz war in welcher Stunde fokussiert, wo lagen die Probleme, in welchem Bereich oder was war

besonders gut? Das kommt ja in den kurzen Stichworten zum Ausdruck, aber es ist eben doch sehr stark komprimiert und zusammengefasst, sodass man die Progression, die im Laufe der Ausbildung stattfindet, dafür ist es, glaube ich, wiederum auch zu undifferenziert. Da guckt man besser hier in diese Einzelauflistung hinein, die dann differenzierter auch!/[...]“

Während A2 hier den Einsatz der Sammlung über das Repertoire korrekt beschreibt, also auch die Skalen der einzelnen Qualitätsbereiche erwähnt, auf denen die Kompetenzentwicklung im Zeitverlauf dokumentiert werden sollte, berichten die Auszubildenden davon, nur die Übersichtstabelle des Repertoires erhalten zu haben. Das ist ein Widerspruch, der so wohl stehen bleiben muss.

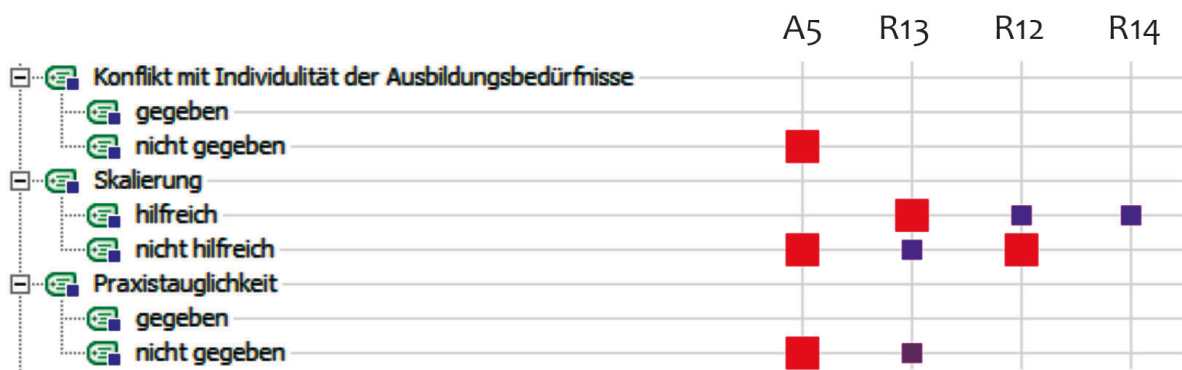
5.3.2 Konzeptbewertung bei A5

Wie im Kapitel zum ‚*Konzept Einsatz der Ausbilder*‘ ausgeführt wurde, setzte A5 das Bewertungskonzept bei jedem Auszubildenden seines Seminar in lediglich einem einzigen Unterrichtsbesuch ein, sodass Bewertungen über den Konzept Einsatz in diesem Ausbildungskontext entsprechend kaum generiert werden können. Gleichwohl sollen in diesem Unterkapitel zumindest die spärlichen Erfahrungen der Akteure aufgearbeitet werden, um sie in das abschließende Gesamtfazit der Arbeit einbauen zu können. Der für die Konzeptentwicklung interessantere Aspekt aus dem Erprobungskontext von A5 sind allerdings die Gründe für den ausbleibenden Konzept Einsatz, wie sie im vorangegangenen Kapitel rekonstruiert wurden, so wie die Konflikte zwischen dem Ausbilder und den Auszubildenden und deren Bedeutung für die Bewertungsprozesse, weshalb auf diese im Rahmen der folgenden Auswertung noch einmal etwas ausführlicher eingegangen wird.

Die Code-Matrix der drei dichotom codierten Subkategorien (vgl. Abb. 5.11) ist, wie es die in weiten Teilen ausbleibende Erprobung vermuten lässt, durch relativ viele Leerstellen ausgezeichnet. Lediglich in der Subkategorie der ‚*Skalierung*‘ geben alle Akteure ihre Einschätzung wieder, inwieweit diese hilfreich oder eben nicht hilfreich für produktive Bewertungsprozesse war, wobei das Feedback der Akteure hier sehr heterogen ist. Hinsichtlich eines ‚*Konflikts mit der Individualität der Ausbildungsbedürfnisse*‘ äußert sich in diesem Fall, abweichend von

den anderen beiden Erprobungskontexten, lediglich der Ausbilder, aber nicht die Auszubildenden, wobei dieser Konflikt auch hier nicht gesehen wird. Die ‚Praxistauglichkeit‘ des Bewertungskonzepts wird auch hier offensichtlich eher kritisch eingeschätzt, wobei lediglich A5 und R13 eine Einschätzung zu dieser Subkategorie abgeben. A5 ist der Auffassung, dass das Bewertungskonzept die Auszubildenden in der Analyse und Planung von Unterricht nicht unsachgemäß eingeschränkt hat, insofern also kein ‚Konflikt mit der Individualität der Ausbildungsbedürfnisse‘ vorlag. Die Auszubildenden äußern sich zu dieser Subkategorie nicht, was vor dem Hintergrund des ausbleibenden Einsatzes nicht weiter verwunderlich ist.

Abbildung 5.11: Code-Matrix der Subkategorien ‚Skalierung‘, ‚Praxistauglichkeit‘, ‚Konflikt mit Individualität der Ausbildungsbedürfnisse‘



In der Subkategorie der ‚Skalierung‘ finden sich zahlreiche Einschätzungen der Akteure hinsichtlich ihres Nutzens für produktive Bewertungsprozesse. A5 äußert grundlegende Kritik an der Idee, Unterrichtsqualität durch Bewertungen von einzelnen Qualitätsbereichen mit festgelegten Skalierungen erfassen zu wollen. In Bezug auf die Nutzung der Skalen für die einzelnen Qualitätsbereiche, als auch derjenigen der Reflexionskompetenz, führt er aus, dass er deren Nutzung „schwierig finde“:

A5: „[...] Wo ich da, das dann auch manchmal schwierig finde. Dann hatten wir einen Qualitätsbereich und dann immer Bezug auf den Ausbildungsstand und den Examenstand. Jetzt ist aber auch gerade in [meinem Fach] so, dass so viele unterschiedliche Stunden oder die Inhalte der Unterrichtsstunden waren immer anders und da kam manchmal, was weiß

ich, die inhaltliche Klarheit kann bei einem [didaktischen Schwerpunkt] gegeben sein, aber dann bei Anwendung [eines andern] dann eben nicht mehr, und da kann man jetzt nicht so sagen, man muss das so machen. Es gibt in [meinem Fach] ganz, ganz viele Möglichkeiten. Und deshalb hatte ich da auch manchmal Schwierigkeiten, habe ich auch mitgenommen. Was ist denn Examensstandard und was/Also das war schwierig¹¹².

I: Weil diese unterstellte Eindimensionalität so eines Qualitätsbereiches, der ist mal klar, der ist mal weniger klar, nicht so ohne weiteres auf die Vielfalt des Unterrichts zu übertragen ist?

A5: Ja, genau. Super zusammengefasst.

Für A5 macht die Erfassung an Hand isolierter Qualitätsbereiche also keinen großen Sinn, da diese Art suggeriere, sie bilde den Kompetenzfortschritt ab, sie aber die vielfältigen Kontexte, in denen Unterricht ablaufen kann, gar nicht zufriedenstellend berücksichtigen könne. Demnach könnten die Skalen auch widersprüchliche Ergebnissen liefern, wenn die Anforderungen für die Realisierung ein- und desselben Qualitätsbereichs, je nach Komplexität der Gestaltung der konkreten Unterrichtsstunde, differieren würden:

A5: „[...] Und was ich ja eben auch schon meinte, ist eben die Sache: Ja dann ist er vielleicht in dem Bereich, hat er dann eben da jetzt eine Leistung gezeigt, die vielleicht auch schon voll und ganz dem Examensstandard entspricht, aber einen Unterrichtsbuch später hätte es wieder ganz anders sein können.“

Die Skalen würden in diesem Fall einen Kompetenzrückschritt abbilden, obwohl vielleicht einfach das Anspruchsniveau für die Realisierung eines Qualitätsbereiches gestiegen ist. Eine Problematik, die so auch von anderen Ausbildern gesehen wird, wie weiter unten in dem Kapitel zu Konflikten des Konzepts mit fachdidaktischen

¹¹² Aus Gründen der Anonymisierung wurden die konkreten fachlichen Bezüge, die A5 ausführt, verallgemeinert. Die eckigen Klammern geben an, wo dies geschehen ist. Eine genauere Analyse der Konflikte des Bewertungskonzepts mit den fachdidaktischen Anforderungen werden in einem eigenen Unterkapitel (siehe Kapitel 5.3.4) aufgearbeitet.

Ansprüchen ausgearbeitet wird. *Alle Auszubildenden* bewerten die ‚Skalierung‘ des Bewertungskonzepts grundlegend positiv. Sie sprechen ihr einerseits auf Basis ihrer wenigen Erfahrungen, die sie mit dem Bewertungskonzept gemacht haben, Sinn zu, bzw. befürworten die Idee des Konzeptdesigns, wie im Fall von R14, obwohl die ‚Skalierung‘ in seinem Fall nicht Unterrichtsnachbesprechung genutzt wurde:

R14: „[...] Ich würde es gerne im Umgang mit anderen Referendaren noch mal erleben, wie es (Anm.: Die Skalierung, D. L.) funktioniert. Ich halte es nämlich tatsächlich, aber das ist natürlich rein hypothetisch, für sinnvoll, v.a. auf Grund der axiologischen Skaleneinteilung. Dieses freie Setzen des X in einem sehr freien Raum des Kreuzes, ist glaube, ich sehr gut nachvollziehbar, v.a. weil man über kleine Nuancen diskutieren kann. Und wahrscheinlich sind das genau diese kleinen Differenzen, die Fachleitung und Referendar sehen können, die ein ganz fruchtbares Gespräch ausmachen können. Ich könnte mir sogar vorstellen, dass der Referendar sich ein Kreuz setzt, die Fachleitung ein Kreuz setzt, die liegen wahrscheinlich gar nicht oft weit auseinander, wenn sie weit auseinander liegen, kann es ein sehr fruchtbares Gespräch werden, wenn sie minimal auseinander liegen, kann es auch ein fruchtbares Gespräch werden, ich kann über die Dauer meine Entwicklung beobachten, ich bin nicht an irgendwelche sehr abstrakten Noten gebunden, die versuchen 13 Unterrichtsqualitätskriterien abzudecken und es ist vollkommen unmöglich. Schade, dass ich es nicht erleben durfte (lacht).

R14 formuliert hier sehr genau die Intention der Skalierung, Gesprächsanlässe zu liefern und damit Klärungsprozesse anzustoßen, falls es zu Missverständnissen zwischen Ausbildern und Auszubildenden kommt. Allerdings merkt er selber an, dass diese positive Einschätzung „jetzt rein hypothetisch“ sei, da das Konzept eben nicht gemäß seiner Intention eingesetzt wurde. R12 und R13 können zumindest von Erfahrungen aus einem Unterrichtsbesuch berichten, in dem die ‚Skalierung‘ des Konzepts für die Bewertung herangezogen wurde. Sie schätzen deren Nutzen positiv ein. R13 beschreibt sich selbst „eher so als Negativdenker“ und führt aus, dass ihr erst durch Fixierung der Bewertung auf den Skalen der Qualitätsbereiche

klar wurde, wie gut die Bewertung ausgefallen war, weil sie „immer eher so höre: Das war nicht so gut“:

I: Ok, also dass diese Verortung auf Skalen für bestimmte Qualitätsbereiche, wie gut die jetzt erfüllt waren, dass die noch mal eine Klärung herbei geführt haben.

R13: Genau, also eine Klärung und manchmal auch dann ein bisschen verwirrender waren. Also man hatte vorher den Eindruck, das war eigentlich alles so weit gut oder auch teilweise sehr gut, bei mir eher: Oder das war nicht so gut. Also ich bin eher der Negativdenker, also bei mir, ich höre dann eher immer so: Das war nicht so gut, und dann wurde es auf einmal in einer Skala eingetragen und die Skala war gar nicht so schlecht. Und das fand/dadurch relativiert sich einiges noch mal oder wird noch mal deutlicher, wenn man dann sich mal einen genauen Skalenwert anguckt.

Dies ist eine Einschätzung, die R13 auch auf die ‚Skalierung‘ der Reflexionskompetenz ausdehnt und der sich auch R12 anschließen kann, weil für sie durch die ‚Skalierung‘ eine klare Bilanzierung der Kompetenzfortschritte, aber auch der noch zu leistenden Professionalisierung, gelingen kann:

I: „[...] Sie haben ja schon gesagt, diese Skalen, die die Qualitätsbereiche einschätzen hinsichtlich Ihrer qualitativen Realisierung, die haben Ihnen geholfen, die fanden Sie gut? Können Sie das noch mal erläutern, warum?“

R12: Weil ich dann weiß, dass ich zu diesem Zeitpunkt und mit dem, was man bisher gelernt hat, es halt sehr gut umgesetzt habe, aber ich weiß halt gleichzeitig, dass ich mich darauf nicht ausruhen darf und dass das im Hinblick auf das Examen eben noch nicht ausreichend ist.“

Gleichzeitig merkt R12 an, dass durch die Art des Instrumenteneinsatzes seitens A5 die positiven Auswirkungen der ‚Skalierung‘ nicht voll zum Tragen kommen konnten, weil dieser erst zum Ende eine Unterscheidung zwischen einem Ausbildungs- und einem Examenstand „mit Verstand“ treffen konnte:

I: „Also, Sie hatten den Eindruck, dass diese Unterscheidung zwischen Ausbildungs- und Examenstand irgendwie willkürlich gezogen worden ist?“

R12: Nicht richtig willkürlich, aber dass da vorher halt nicht wirklich drüber nachgedacht worden ist und erst in dem Moment, und dass es erst im Laufe des Prozesses bei dieser Auswertung von Skala zu Skala dann irgendwann klarer wurde, aber die ersten mit Sicherheit willkürlich und dann zu Ende dann eher, ja mit Verstand ausgefüllt.“

Darüber hinaus merkt R13 kritisch an, dass es für einen Ausbilder auch nicht immer leicht sei, sich auf einen konkreten Skalenwert in der Bewertung festzulegen, da man sich damit in Gefahr begeben, „sich festnageln“ zu lassen. Eine Problematik, die sie auch aus ihrer eigenen Tätigkeit als Lehrerin kennt:

R13: „[...] Ich glaube, es ist auch für einen Fachleiter manchmal schwierig und das weiß ich, weil ich eben jetzt auch Lehrerin bin, sich so explizit auf einen bestimmten Wert, also damit könnte man einen schon so ein bisschen festnageln. Und das ist, glaube ich, manchmal schwierig, weswegen da ganz viel Streuung ja auch immer ist.“

Hinsichtlich der ‚Praxistauglichkeit‘ erwähnen A5 und R13, wie schon die Akteure zuvor, den zu großen Umfang des Konzepts und die damit in Verbindung stehenden Fragen der Komplexität des Konzepts und des Zeitaufwands für seinen Einsatz. A5 berichtet zwar, dass die drei Probanden, die an der Erprobung des Konzepts beteiligt gewesen sind, mit diesem umgehen konnten, dies aber nur, „weil sie sehr fit waren“ und „die anderen Referendare“ aus seinem Seminar „das gar nicht verstanden hätten“:

I: Wie war Ihr Eindruck von dem Konzept generell, Schrägstrich wenn Sie es eingesetzt haben, konnte man das so in der Ausbildungspraxis einsetzen, wie es bisher designt war oder eher nicht?

R5: Also bei den drei Probanden, ich nenne sie jetzt mal alle Probanden. Da war es möglich, da war das aber möglich, weil die alle drei sehr fit waren, das alles kognitiviert haben. Ich habe mir heute noch überlegt, wie

es gewesen wäre, hätte ich es bei anderen Referendarinnen und Referendaren bei mir im Seminar eingesetzt, die hätten das gar nicht verstanden. [...] Und die wären wirklich absolut davon überfordert gewesen.

A5 stellt diese Überforderung hier als These auf, die sich nicht aus der Erprobung ergeben hat, aber seine Bedenken decken sich hier mit den bereits zuvor von A2 vorgebrachten Einwänden. Weiterhin benannte A5 den zeitlichen Aufwand für den Konzepteinsatz als ein Hindernis, wie es schon im Unterkapitel zum ‚Konzepteinsatz‘ deutlich herausgearbeitet wurde und wie es auch R13 bei der Erprobung wahrgenommen hat:

R13: „[...] Das Einzige, was halt das Schwierige noch daran war, dass es halt so für die Fachleiter zu lange gedauert hat. Also da hatte ich auch das Gefühl, dass es bei A5 so ein bisschen der Fall war, dass es so ihm zu lange gedauert hat. Ich hatte den Eindruck, dass war der Kritikpunkt daran. [] Vielleicht müsste man den auch reduzieren ein Stück, das kann ich nicht einschätzen, weil ich mich da nicht so reingedacht habe in das Konzept. Aber ich glaube Fachleiter mögen das an manchen Stellen/[...]“

Aus der Auswertung der drei Subkategorien der ‚Skalierung‘, ‚Praxistauglichkeit‘ und des ‚Konflikt mit der Individualität der Ausbildungsbedürfnisse‘, können trotz des nur marginal erfolgten Einsatzes gewisse Schlüsse abgeleitet werden. Die ‚Skalierung‘ scheint aus der Perspektive der Auszubildenden durchaus einen Beitrag für produktive Bewertungsprozesse zu leisten, wenn sie Bilanzierungen anbietet, die offensichtlich eine höhere Klarheit und Verbindlichkeit haben, als die reine Verbalisierung durch den Ausbilder. Allerdings kritisiert A5, dass diese ‚Skalierung‘ der Komplexität des Unterrichts nicht gerecht werde, da die Realisierung einzelner Qualitätsbereiche eben unterschiedlich komplex sein könne, weswegen Veränderungen auf einer Skala nicht notwendigerweise Kompetenzfort- oder -rückschritte abbilden müssten. Hinsichtlich der ‚Praxistauglichkeit‘ des Instrumente wiederholen sich die Forderungen nach einer Reduzierung der Komplexität und des Umfangs, wie sie zuvor schon im Eprobungskontext von A2 formuliert wurden. Der ausbleibenden Nutzung des Bewertungskonzepts ist es zuzuschreiben, dass die Auszubildenden keine Einschätzung darüber abgeben

können, ob durch das Bewertungskonzept ein ‚*Konflikt mit der Individualität der Ausbildungsbedürfnisse*‘ entsteht.

Auch bei der Subkategorie der ‚*Modellbildung*‘ findet sich nur sehr lückenhafte Rückmeldung der Akteure, die durch den in weiten Teilen ausbleibenden Einsatz des Bewertungskonzepts zu klären ist. Die Code-Matrix in Abbildung 5.12 zeigt noch weniger codierte Interviewpassagen als jene der zuvor ausgewerteten Subkategorien. Für die Subkategorie einer ‚*Hilfestellung für die Planung*‘ von Unterricht seitens des Konzepts gibt nur R13 eine Rückmeldung. In den anderen beiden Subkategorien ließen sich zwar immerhin von zwei Akteuren Interviewpassagen codieren, aber diese konnten den jeweils positiv und negativ besetzten Seiten der Subkategorien zugeordnet werden, sodass die rein deskriptive Darstellung der Codierhäufigkeiten mehr Fragen aufwirft, als dass sie diese beantwortet.

Abbildung 5.12: Code-Matrix der Subkategorie ‚*Modellbildung*‘

	A5	R13	R12	R14
Hilfestellung für Planung				
gelingt				
gelingt nicht		■		
Hilfestellung für Reflexion				
gelingt		■		
gelingt nicht	■			
klare Operationalisierung				
gelingt				■
gelingt nicht	■			

R13 ist die einzige Auszubildende, die darüber berichtet, das Bewertungskonzept für die Vorbereitung der Unterrichtsnachbesprechung eingesetzt zu haben, wobei der Nutzen des Konzepts als ‚*Hilfestellung für die Reflexion*‘ für sie darin liegt, eine „Abhakliste“ zu haben, um ihr Statement in der Nachbesprechung vorzubereiten:

R13: „[...]ich hatte das dann genutzt um das, meinen zweiten UB vorzubereiten, allerdings eher so als Abhakliste. Ich habe meine Stunde ganz normal gemacht wie immer und danach habe ich mir angeguckt, weil ich mich darauf vorbereiten wollte, wie das hinterher in der Nachbesprechung läuft und habe dann auch mir angeguckt, was ich im Statement/Also mir

hat das im Statement ein bisschen geholfen, weil ich im Statement auf diese Bereiche eingehen wollte, weil ich glaube!

I: Aus der Matrix, diese Qualitätsbereiche.

R13: Genau, weil ich/Es macht ja Sinn, dass die Nachbesprechung genau die Punkte gleich anspricht, dass ich die Punkte auch im Statement benutze, um meine Stunde gut oder an manchen Stellen vielleicht die Kritik vorwegzunehmen. [...]"

I: Würden Sie sagen, dass war gewinnbringend für Sie?

R13: Ja, allein weil es auch ein bisschen schneller geht. Also man hat/ Wenn man sich, das habe ich jetzt leider nicht so konsequent gemacht, weil es hinterher nicht mehr weiter benutzt wurde, aber ich glaube, es ist ganz gewinnbringend auch nach einer Stunde, wo man ja noch eine Viertelstunde Zeit hat sich auf die Nachbesprechung vorzubereiten, wenn man die noch mal nimmt und abhakt. [...]"

Den Beitrag des Bewertungskonzepts für die erfolgreiche Gestaltung der Unterrichtsnachbesprechung und hierbei offensichtlich insb. der Matrix, beurteilt R13 als „gewinnbringend“, wohingegen ihr das Bewertungskonzept keine ‚Hilfestellung für die Planung‘ von Unterricht gewesen sei, da sie diese auch nicht gebraucht habe, da sie „schon immer perfektionistisch und strukturiert“ gewesen sei:

R13: „[...] Also, ich habe das nicht genommen und habe dann meine Stunden geplant, ich glaube aber weiterhin, dass es Leute in meinem Seminar geben würde, die das bräuchten. Ich glaube, das hat was mit einer Struktur zu tun. Also für mich, die schon immer perfektionistisch und strukturiert ist, habe die Struktur für mich. Ich glaube aber, es gibt genug Leute, die alleine diese Struktur schon für die Planung in der Stunde überhaupt bräuchten.“

Obwohl R13 das Bewertungskonzept nicht als Unterstützung für die Planung von Unterricht herangezogen hat, ist sie der Meinung, dass es „genug Leute“ gäbe, die diese Hilfestellung benötigen würden.

A5 hat nicht erlebt, dass das Bewertungskonzept eine ‚Hilfestellung zur Reflexion‘ leisten konnte, weil es nicht dabei geholfen habe, Struktur in die gemeinsame Besprechung des Unterrichts zu bringen, sondern eher hinderlich dabei war, Klarheit über Entwicklungsfelder zu schaffen:

I: “[...] Konnte man mit diesen Qualitätsbereichen, mit diesen Instrumenten, dem Dokumentationsbogen über die Nachbesprechung und der Matrix, haben die geholfen einen Beratungsbedarf klar zu formulieren und auch Entwicklungsaufgaben zu fokussieren, die die Referendare hatten?

A5: Ja mit dem Beratungsbedarf war ja eher schwierig. Die haben ja dann eigentlich ihre Probleme, haben sie wahrgenommen und geschildert und dann haben wir ja gemeinsam geguckt, zu welchem Qualitätsbereich gehört das? Aber da war ja eben das große Problem, weil ich immer noch nicht sagen kann, ob es jetzt da zu dem Bereich gehört oder da zu diesem Bereich [...].“

Offensichtlich hat das Bewertungskonzept, in den wenigen Versuchen, in denen es eingesetzt wurde, die Unterrichtsnachbesprechung nicht strukturiert. Es wurde vielmehr ‚additiv‘ an die wesentlichen Beratungs- und Bewertungsgespräche angefügt, sodass es eher Probleme bereitete, die besprochenen Aspekte bestimmten Qualitätsbereichen zuzuordnen, als dass die angebotene Struktur Klärungsprozesse angestoßen hätte.

R14 spricht dem Konzept zu, dass es die ‚klare Operationalisierung‘ guten Unterrichts, die es leisten wollte, durchaus leisten konnte. Allerdings bleibt auch seine Einschätzung primär eine oberflächliche, da das Bewertungskonzept nicht ausreichend eingesetzt wurde, um eine fundierte Wertung zu dieser Subkategorie abgeben zu können:

I: „Ok, dann erübrigt sich jetzt quasi der gesamte Frageblock. Gut haben Sie nichtsdestotrotz dieses Konzept für sich genutzt um Ihre Unterrichtsbesuche vorzubereiten, haben Sie da mal reingeguckt?“

R14: Ich habe da reingeschaut am Anfang, ja, und hatte auch das Gefühl, das entspricht dem was ich gerne können möchte, was ich wichtig finde für Unterricht und ich glaube, dass ich auch damit Unterricht beobachten könnte.“

A5 kritisiert entgegen der Positionierung von R14 eine mangelnde Passung der Operationalisierung guten Unterrichts mit fachlichen Qualitätsstandards, die er an das Verständnis von gutem Unterricht anlegt. Diesem Aspekt wird in dem eigenen Unterkapitel der ‚*Passung mit fachlichen Qualitätsstandards*‘ Rechnung getragen.

Das abschließende Fazit der Auswertung dieser Subkategorie fällt, wie zu erwarten, relativ dürftig aus. Der ausbleibende Einsatz des Bewertungskonzepts führt dazu, dass wenig verwertbares Feedback hinsichtlich der Leistung des Konzepts, eine ‚*Modellbildung*‘ guten Unterrichts auf Seiten der Auszubildenden zu unterstützen, generiert werden kann. Das Nutzungsmuster des Konzepts, das zeigt auch der Beitrag von A5 zur Frage, inwieweit das Konzept eine Hilfestellung für die Reflexion bietet, schränkte dessen Möglichkeiten, überhaupt wirksam zu werden, massiv ein. Lediglich R13 berichtet von einer ‚*Hilfestellung zur Reflexion*‘, die das Konzept für sie anbot, die sie aber für die ‚*Planung*‘ von Unterricht so nicht wahrnahm, wobei diese ausbleibende Nutzung des Konzepts zur Planung von Unterricht zumindest teilweise mit ihrer Selbsteinschätzung, sich selbst „gut strukturieren“ zu können, zusammenhängt und nicht ursächlich auf das Konzeptdesign zurückzuführen ist.

Die Code-Matrix der Subkategorie der ‚*Transparenz*‘ (siehe Abbildung 5.13) zeigt als auffälligstes Merkmal, dass keine einzige Interviewpassage das Ausbilders zu der Frage, inwiefern das Bewertungskonzept die ‚*Transparenz*‘ der Bewertungsprozesse steigern kann, codiert wurde. Dieser Aspekt hat in dem Interview mit A5 entsprechend kaum eine Rolle gespielt, bzw. keine Rolle spielen können, da

sich die Frage i.W. um die Gründe des ausbleibenden Einsatzes und der (vermeintlichen?) Mängel des Bewertungskonzepts gedreht hat. Weiterhin sind natürlich die vielen leeren Plätze in der Code-Matrix auffällig, die dokumentieren, dass durch die ausbleibende Erprobung des Bewertungskonzepts weder positive noch negative Erfahrungen hinsichtlich des Beitrags des Konzepts zur Sicherstellung transparenter Bewertungsprozesse gemacht werden konnten. Etwas aus dem Rahmen fällt dabei die breit aufgestellte und durchweg positive Rezeption des Konzepts in dieser Subkategorie durch R13, wohingegen R12 und R14 nur eine singuläre Einschätzung zur Leistung des Bewertungskonzepts abgeben.

Abbildung 5.13: Code-Matrix der Subkategorie ‚Transparenz‘



Wie bereits erwähnt, scheint das Bewertungskonzept bei R13, trotz des i.W. ausbleibenden Einsatzes den Eindruck erweckt zu haben, dass es einen wesentlichen Beitrag zur Sicherstellung von ‚Transparenz‘ in den Bewertungsprozessen leisten kann. Sie verweist hierbei explizit auf Unterschiede in der ‚klaren Strukturierung der Nachbesprechung‘ des einen Unterrichtsbesuchs, in dem das Bewertungskonzept erprobt wurde und den anderen Unterrichtsnachbesprechungen, die sie in ihrer Ausbildung erlebt hat:

R13: „[...] Es macht ja Sinn, dass die Nachbesprechung genau die Punkte gleich anspricht, dass ich die Punkte auch im Statement benutze um meine Stunde gut oder an manchen Stellen vielleicht die Kritik vorwegzunehmen. Das fand ich halt gut, und das ist bisher im Seminar ja nicht so. Also, wir haben zwar dieses Bonner Qualitätstableau mit den zwölf Feldern oder

wie sie heißen und das Problem ist, auf die geht man ja nicht in der Fülle ein. Also ich kann/Ich saß da dann immer vor, hatte zwölf Felder und habe mir dann willkürlich dann drei genommen, weil, warum auch nicht, drei musste ich ja nehmen.[...] Und das finde ich ein bisschen klarer, man weiß genau, was kommt gleich im Gespräch. Ich kann jetzt nicht behaupten, dass ich das im Vorfeld/[...]“

Durch den festgelegten Ablauf der Unterrichtsnachbesprechung und -dokumentation, ist die Struktur der Nachbesprechung für R13 „ein bisschen klarer“, ein Umstand, von dem sie sich wünscht, dass er in der gesamten Ausbildung eine weitere Verbreitung finden würde, um der erlebten „Willkür“ etwas entgegenzusetzen:

R13: „[...] Ich habe immer das Gefühl, es ist alles sehr willkürlich oder sehr von Personen abhängig und auch wie versteht man sich mit seinem Fachleiter oder wie gut ist man? Und ich glaube, so ein striktes Konzept so, das muss jetzt passieren, das muss passieren, das muss passieren, und hier halten wir es auch noch fest, würde helfen, das Ganze ein bisschen einheitlicher zu gestalten und dann im Sinne der Referendare mehr, sehr viel Transparenz, was von uns ja immer gefordert wird.“

Weiterhin berichtet R13 darüber, dass durch den Einsatz des Bewertungskonzepts und hierbei v.a. des Dokumentationsbogens für die Nachbesprechung ‚Gesprächsanlässe‘ geschaffen wurden, die zu einer höheren ‚Transparenz‘ der Bewertungsprozesse beigetragen haben:

I: „Die Instrumente sind ja letzten Endes gar nicht herangezogen worden. Bei dem einen Mal, wo sie herangezogen worden sind, hatten Sie den Eindruck, dass eine bessere Klärung, eine höhere Transparenz über die Bewertung hergestellt werden konnte?“

R13: Ja, ich glaube schon, weil man eben ganzhaft über irgendwelche Kreuze und Notizen, die man sich gemacht hat, ja reden musste. Also es gab ja immer wieder einen Gesprächsimpuls oder zumindest für mich nochmal die Gewissheit: Ok, das hat ihn anscheinend wirklich gestört. [...]

*Also, da hätten Sie noch und da hätten Sie noch und da hätten sie noch.
Das fand ich schon, hat der Klarheit geholfen.“*

Das Potential des Bewertungskonzepts, diese ‚Gesprächsanlässe‘ zu generieren, hat auch R14 gesehen, wobei er selbst anmerkt, dass dies natürlich nur eine „hypothetische“ Einschätzung sei, da er die Erprobung des Dokumentationsbogens nicht erlebt hat:

R14: „[...] Dieses freie Setzen des X in einem sehr freien Raum des Kreuzes ist, glaube ich, sehr gut nachvollziehbar, v.a. weil man über kleine Nuancen diskutieren kann. Und wahrscheinlich sind das genau diese kleinen Differenzen, die Fachleitung und Referendar sehen können, die ein ganz fruchtbares Gespräch ausmachen können.[...]“

Durch die ‚schriftliche Bilanzierung‘ des Unterrichtsbesuches, die durch den Dokumentationsbogen geleistet werden sollte, konnte für R13 außerdem eine „klare Dokumentation“ der Unterrichtsbewertung gelingen. Hierbei attestiert sie dem Bewertungskonzept anscheinend eine „klarere Dokumentation“, als es die bisher in ihrer Ausbildung praktizierten Verfahren gewährleisten konnten:

I: Können Sie sich vorstellen, ich weiß nicht, ob Sie den Dokumentationsbogen so noch vor Augen haben, Sie haben ja gerade schon die Skalen angesprochen, dass diese Art der Dokumentation gewinnbringender gewesen wäre? Oder, Sie haben ihn ja auch einmal gekriegt und haben vorhin da schon Bezug genommen, dass die Skalen Ihnen geholfen haben.

R13: Ja, ich glaube, gewinnbringend ist der insofern, als dass er sehr viel klarer ist oder klarer dokumentiert wird, was wirklich in dieser Stunde passiert ist oder wie ich als Lehrperson agiert habe. Also ich glaube, es ist ganz gut um Klarheit zu schaffen.

R14 berichtet darüber, dass der Einsatz des Bewertungskonzepts die problematischen Interaktionsmuster, die aus den hierarchischen Beziehungen zwischen den Akteuren resultieren, nicht egalisieren kann, sondern der Einsatz „eine zu-

sätzliche Belastung“ darstellt, wenn der Ausbilder den mit dem Konzept in Verbindung stehenden Abmachungen nicht nachkommt:

R14: „Eine wichtige Erfahrung ist sicherlich, dass nach der, dem Unterrichtsnachgespräch in dem das Ganze benutzt wurde und ich keine Rückmeldung bekommen habe, der Moment des Einforderns kein einfacher ist. Also wenn man dann vor der Fachleitung steht und fragt: «Wie sieht es eigentlich damit aus?» Und dann gesagt bekommt: «Ja kommt bald», dann fragt man doch eher kein zweites Mal nach, zumindest ich hatte keine Lust, ein zweites Mal nachzufragen und das erschwert den Umgang damit. Also wenn die Fachleitung damit nicht souverän umgeht und nicht diszipliniert umgeht, dann ist es für den Referendar eine zusätzliche Belastung, weil er auf etwas wartet, womit er gerne arbeiten würde, es aber nicht kommt.“

Auch hier ‚*verhindern die Machtverhältnisse offene Kommunikation*‘ und die Herstellung von ‚*Transparenz*‘ über den Einsatz des Konzepts, denn wenn R14 sich nicht traut den Einsatz einzufordern, kann das Konzept auch keine Wirkung zeigen. Ebenso zeigen die bereits oben, in der Auswertung der Subkategorie der ‚*Skalierung*‘, zitierten Berichte von R12, dass ein unklarer und missverständlicher ‚*Einsatz der Instrumente*‘ keinen Beitrag zu transparenten Bewertungsprozessen leisten kann.

Die Berichte von R14 führen darüber hinaus drastisch vor Augen, welche dysfunktionalen Mechanismen dadurch entstehen, dass die ‚*Machtverhältnisse offene Kommunikation verhindern*‘. In seinem Fall haben diese maßgeblichen Anteil an einem, wie er es nennt, „Regress der Beziehung“ zwischen sich und A5 gehabt, an dessen Ende R14 die Ausbildungssituationen nur noch ertrug, aber keinerlei Hoffnungen auf einen Kompetenzfortschritt mit ihnen verband:

I: Können Sie das ein bisschen näher einfach erzählen. Was für ein Regress, was für Spiele das waren?

R14: Wenn man ein gewisser Typ Mensch ist und einen gewissen Anspruch an sich hat und das Gefühl hat, von der anderen Seite wird das nicht gespiegelt und wird nicht eingefordert, und es gibt keine klare Kriterien,

an die man sich halten kann, dann neige zumindest ich dazu, zu sagen: Ok, gut dann spiegele ich das jetzt meinem Gegenüber, nachdem ich ein Feedback gegeben habe, in dem ich mich auf eine gewisse Art und Weise verhalte und da ich ja ein Produkt meines Ausbilders bin, er auch an mir sehen kann: Ok, da und da hapert es. Ich habe so ein bisschen die Hoffnung gehabt, dass auf Grund meines Verhaltens ein Verhalten der Fachleitung folgt, dass sichtlich macht, dass gewisse Klarheiten und Strukturen fehlen und zwar in der Ansage, also in den Erwartungen an mich selbst. Das ist nicht geschehen und dadurch ist schlichtweg die Lust flöten gegangen. Ich hatte keine Lust mehr, diese Unterrichtsbesuche bei dieser Fachleitung so vorzubereiten, dass sie ihm genügen, weil ich nicht ihm als etwas, ihm nicht genügen wollte, wo ich doch der Meinung bin, dass er mir doch überhaupt nichts anbietet an Struktur.

I: Also so ein sehr starker Rollenkonflikt, der dann auch zum Tragen kommt.

R14: Ja, ja ganz klar. [...]

I: Ok, d.h./Die Bewertungssituation und Beratungssituation waren, wenn ich das richtig raus höre, für Sie eigentlich dann so, dass Sie das akzeptiert haben, dass Sie das jetzt noch machen müssen, aber dass Sie für sich eigentlich nicht mehr mit Lernerfolgen da raus gegangen sind?

R14: Ja, das ist richtig.

I: Weil Sie einfach auch Ihren Ausbilder in seiner Rolle nicht mehr so ernst nehmen konnten?

R14: Ja, das ist richtig. Und, das kommt nach dazu, weil ich mir auch den Notendruck genommen haben. Ich habe mir gesagt: Es ist mir jetzt egal, was dabei herauspringt. Ich habe mir einfach ausgerechnet, wie viel Prozent diese Note anteilig ist und habe mir dann gesagt, gut, das ist in Ordnung.

Die enttäuschten Erwartungen an die Adresse des Ausbilders haben bei R14 dazu geführt, dass er seine Handlungsstrategien nicht mehr an der eigenen Professionalisierung ausgerichtet hat, sondern eine extreme Variante der ‚*strategischen, verdeckten und konfliktiven Interaktion*‘ an den Tag gelegt hat, die oben als maßgeblicher Hemmschuh erfolgreicher Ausbildungsverläufe ausgemacht wurde. Entsprechend kann die Bedeutung der ‚*Beziehung zwischen Ausbildern und Auszubildenden*‘ für eine gelingende Ausbildung, wie sie aus den Gruppendiskussionen herausgearbeitet wurde, bestätigt werden. Sie hat entscheidenden Einfluss auf die Wahl der Handlungsstrategien der Akteure, dadurch auf die zustande kommenden Interaktions- und Kommunikationsmuster und damit letztendlich auf die Frage, ob die Ausbildungssituationen Merkmale im Sinne der oben umrissenen ‚*Stresssituation*‘ oder eher einer ‚*Lernsituation*‘ aufweisen.

Schlussendlich zeigt die Auswertung der Subkategorie der ‚*Transparenz*‘, dass ein Bewertungskonzept die ‚*Machtverhältnisse*‘ zwischen den Akteuren nur sehr eingeschränkt beeinflussen kann. Ignoriert ein Ausbilder das Bewertungskonzept, ist es für die Auszubildenden eine große Herausforderung dessen Einsatz einzufordern. R14 spricht in diesem Zusammenhang von einer zusätzlichen Belastung. Trotz des singulären Einsatzes berichtet R13 über positive Auswirkungen des Bewertungskonzepts, die so aus der Erprobung von A2 bereits bekannt sind. Die ‚*schriftliche Bilanzierung*‘ durch den Dokumentationsbogen sorgt für eine Verbindlichkeit der Bewertung, die R13 ebenso schätzt wie die ‚*klare Strukturierung der Nachbesprechung*‘ und der ‚*Gesprächsanlässe*‘, die durch das Konzept generiert werden.

Vor dem Hintergrund der Ausführungen zum ‚*Konzepteinsatz*‘ bei A5, sowie den nur lückenhaften Bewertungen in der Kategorie ‚*Konzeptbewertung*‘, ist es wenig verwunderlich, dass das ‚*Fazit*‘ der Erprobung bei A5 von allen Akteuren damit abschließt, dass das Bewertungskonzept umfassend nicht beurteilt werden kann. Die Code-Matrix in Abbildung 5.14 zeigt die entsprechend einhellige Positionierung der Akteure.

Abbildung 5.14: Code-Matrix der Subkategorie ‚Fazit‘

	A5	R13	R12	R14
Fazit				
Beurteilung nicht möglich	■	■	■	■
positiv				
negativ				

A5 bescheinigt der Erprobung keine Veränderung seiner Bewertungsprozesse, da „die Struktur der Nachbesprechung „ja auch schon vorher so war“ und meint, „dass das Konzept sehr sinnvoll ist, wenn man es konsequent einsetzt“:

I: „Hat das Konzept nachhaltig Nachbesprechungen verändert, in dem es eingesetzt worden ist?“

A5: Nein. Und das ist ja auch wahrscheinlich. Weil ich habe die Struktur der Nachbesprechung, die war ja vorher schon so. Statement, Feedback, Festlegung von Beratungsaspekten, Beratungsgespräche und dann hinterher nochmal eben eine Metareflexion. Das ist ja so geblieben.

I: Haben Sie noch wichtige Erfahrungen aus der Erprobung dieses Bewertungskonzepts, die ich bisher nicht angesprochen haben, die Ihnen wichtig sind?

(...)

A5: Nein, das habe ich ja eben schon gesagt: Ich glaube, dass das wirklich, wenn man das konsequent einsetzt und wenn man das vielleicht noch ein bisschen mehr auf den Fachunterricht modifiziert, dass das sehr sinnvoll ist. [...]“

Die abschließende Bewertung von A5 verbleibt nebulös, da einerseits positive Untertöne der Konzeptrezeption durchklingen („dass das sehr sinnvoll ist“), aber andererseits eben genau der konsequente Einsatz, den A5 als Bedingung einer sinnvollen Konzeptnutzung formuliert, bei ihm nicht zu realisieren war.

Aus den Interviews *aller Auszubildenden* ließ sich die Einschätzung rekonstruieren, dass eine abschließende und umfassende Konzeptbewertung nicht möglich

war, sondern bestenfalls eine Bewertung einzelner singulärer Aspekte des Konzepts, die darüber hinaus größtenteils bloß „hypothetischen Charakter“ haben, wie R14 es formuliert, da die Konzepterprobung ja de facto nicht stattgefunden habe. So zu verstehen ist auch seine Reaktion auf die Impulsfrage des Interviews:

I: „Ja, wenn Sie an den Einsatz des Bewertungskonzepts zurück denken, diese Erprobungsphase, was fällt Ihnen da spontan als Erstes ein?“

R14: „Spontan fällt mir ein: Schade, dass ich es nicht hundertprozentig ausprobieren konnte, weil es schlichtweg nicht durchgeführt wurde. [...]“

Eine Einschätzung, die auch für R12 die prägende Erfahrung der Konzepterprobung geblieben ist, wie sie ebenfalls auf die offene Impulsfrage angibt:

R12: „Dass nicht viel funktioniert hat und dass es eigentlich gar nicht zum Einsatz gekommen ist, so richtig.“

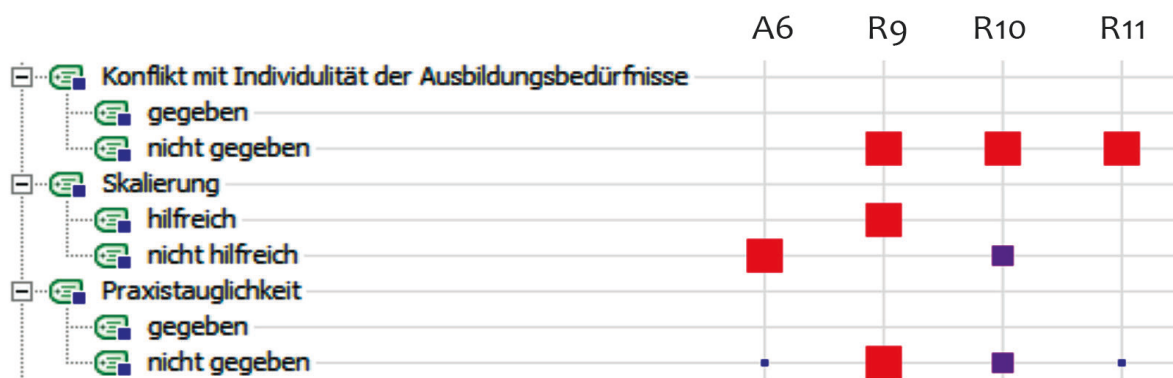
Entsprechend bleiben als Ergebnis der Konzepterprobung v.a. bruchstückhafte Einschätzungen der Akteure zu den einzelnen Subkategorien zurück, die auf Grund der mangelnden Stringenz der Erprobung nicht zu einem eindeutigen und zufriedenstellenden ‚Fazit‘ aggregiert werden können

5.3.3 Konzeptbewertung bei A6

Die deskriptive Auswertung der Code-Matrix (vgl. Abb. 5.15) der dichotom codierten Subkategorien zeigt eine relativ eindeutige Rezeption des Bewertungskonzepts, die lediglich hinsichtlich der Subkategorie der ‚Skalierung‘ Unterschiede zwischen den Positionen der Akteure wiedergibt. Ansonsten zeigen Ausbilder und Auszubildende bei der Frage der ‚Praxistauglichkeit‘ des Konzepts eine einhellig negative Positionierung, die Auszubildenden bei der Beurteilung eines möglichen ‚Konflikts mit der Individualität der Ausbildungsbedürfnisse‘ des Konzepts eine einstimmig positive Beurteilung, zu der sich der Ausbilder seinerseits nicht äußert. Die Bewertungen der Akteure fallen in dieser Erprobungsgruppe wesentlich eindeutiger aus als in der Gruppe von A2, in der der Ausbilder und die

Auszubildenden in einzelnen Subkategorien sowohl Stärken als auch Schwächen des Konzepts benannt haben, hier nun aber ausschließlich eindeutige Positionierungen der Akteure zu finden sind. Wie schon in der Erprobungsgruppe von A2 berichtet keiner der Auszubildenden aus der Gruppe von A6, dass es durch den Einsatz des Bewertungskonzepts zu einem wahrgenommenen ‚Konflikt mit der Individualität der Ausbildungsbedürfnisse‘ gekommen sei, das Konzept demnach genügend Freiräume für die Entfaltung und Entwicklung der eigenen Lehrerpersönlichkeit gelassen habe.

Abbildung 5.15 : Code-Matrix der Subkategorien ‚Skalierung‘, ‚Praxistauglichkeit‘, ‚Konflikt mit Individualität der Ausbildungsbedürfnisse‘



R10 berichtet darüber, dass er das Konzept als „theoretische Optimalvorgabe“ verstanden hat, welches es nicht in Gänze umzusetzen gelte, sondern für ihn nur als „Orientierungsrahmen“ fungiere:

I: „Hatte das Konzept Ihnen genug Freiräume für die Planung und Durchführung Ihres Unterrichts gegeben, oder haben Sie sich dadurch eingeengt gefühlt?“

R10: Nein, es war keine Einengung, es ist/Natürlich muss in letzter Konsequenz ich das verantworten, was ich mache. Das ist ein Orientierungsrahmen, aber es gibt sicherlich auch mal gute Begründungen, warum man jetzt einen Aspekt nicht unbedingt so definitiv wahrnimmt. Es sind ja Optimalvorgaben für ein theoretisches Modell und dahingehend schaut man sich das mal an, guckt dann, inwiefern die Umsetzung gegeben ist, wenn was fehlt, brauche ich es unbedingt in dem Unterricht?“

Auch R9 und R11 berichten über keine Einengung durch Konzept, wobei R9 dies damit begründet, dass sie das Konzept „wenig benutzt“ habe und auch bei R11 der Einsatz des Konzepts nicht stringent erfolgte. Ein Umstand, der weiter unten ausführlicher zu thematisieren ist.

Die ‚Skalierung‘ ist die in der Erprobung von A6 einzig wirklich kontrovers rezipierte Subkategorie. A6 und R10 sind der Auffassung, dass die Skalierung keine Hilfestellung für die Gestaltung erfolgreicher Unterrichtsnachbesprechungen anbot, wohingegen R9 durchaus einen positiven Nutzen dieser gesehen hat. Der Ausbilder hat generelle Vorbehalte gegen die Idee, einzelne Qualitätsbereiche aus der Gesamtbewertung herauszugreifen und diese für sich genommen hinsichtlich der Frage zu bewerten, ob sie zufriedenstellend umgesetzt worden sind oder nicht. Eine solche „eindimensionale“ Sichtweise, sei „per se ungeeignet“, um Unterricht zufriedenstellend zu bewerten:

A6: „[...] Ich glaube aber, nachdem ich jetzt auch viel darüber nachgedacht habe, v.a. als ich diese Repertoirebögen ausgefüllt habe, dass das für Unterricht nicht geeignet ist. Dass wir immer in so einer Art Gummibandsituation sind, in der, wo verschiedenen Nägel eingeschlagen sind und dass mal mehr dieser Punkt betrachtet wird und mal mehr jener Punkt betrachtet wird, aber dass das ganze Wirkungsgefüge trotzdem immer noch erhalten bleibt. Und deswegen glaube ich im Nachhinein, was ich vorher völlig anders beurteilt habe, dass diese Sicht mehrere eindimensionale Dinge aufeinanderzulegen, per se ungeeignet ist.“

Im Rahmen der Erprobung des Konzepts musste A6 einen Wandel in der eigenen Vorstellung über die Bewertung von Unterricht erleben. Die, wie er es nennt, „eindimensionale Bewertung“ des Unterrichts über „ein lineares Abbildungssystem“ habe sich als inadäquat erwiesen, um Unterricht zufriedenstellend zu bewerten, eine Sichtweise, die er vor der Erprobung „völlig anders beurteilt“ habe:

A6: „[...] Und der andere Punkt, eben diese Frage, dass Unterricht als komplexes System andere Arten von Abbildung als ein lineares Abbildungssystem braucht.

I: Sondern eines, dass die Überkomplexität dem gerecht wird und das kann eigentlich kein standardisiertes sein, wenn ich das richtig verstehe?

A6: Ja. [...]

I: Aber so eine formalisierte Beobachtung mit Skalen oder auch Einzelbereichen, die man sich herauspicks und formalisiert dann abhandelt hinsichtlich einem Fortschritt auf einer Skalierung, das erscheint!

A6: Hat sich mir jetzt in der wenigen Anzahl, also ich habe es ja vielleicht zehn mal gemacht, als nicht so hilfreich erwiesen, wie ich es gedacht habe.“

Diese grundlegend negative Bewertung des Konzeptdesigns bezieht A6 auch auf die Skalen der Reflexionskompetenz und teilt damit die Rezeption der ‚Skalierung‘, wie sie auch A5 bereits an den Tag legte. R10 erschien sowohl die Skalierung des Dokumentationsbogens als auch des Repertoires wenig hilfreich, um eine klare Bilanzierung seiner erreichten Kompetenz leisten zu können. Die Einordnungen auf den Skalen verblieben für ihn eine „vage Ahnung“, aber sie konnten keinen Klärungsprozess initiieren:

I: „Konnte dieser Unterschied/Sie sagten gerade, das war ein bisschen enttäuschend, konnte das für Sie schlüssig erklärt werden, woher die Unterschiede kommen, von einem Ausbildungsstand und einem Examensstand?

R10: Nee, also da war lediglich so diese vage Ahnung einer steigenden Anspruchsprogression, aber jetzt keine qualitativen Kriterien, die noch zusätzlich erfüllt werden müssten um zu sagen: Ok, das ist jetzt Ausbildungsstand vollständig, aber Examensstand bräuchte ich noch dieses oder jenes, das fehlte.

I: Sondern es war: Da ist ein Unterschied, aber wie Sie genau dahin kommen, wo, man an welchen Kriterien das festzustellen ist!

R10: Genau, es wird schwerer und anspruchsvoller.

I: Also es blieb nebulös.

R10: Genau.“

R9 hat die Nutzung der Skalen grundlegend anders erlebt, als „gewinnbringend“, wie sie es formuliert:

R9: „ [...] Ja und nichtsdestotrotz, im Laufe der Ausbildung habe ich das schon als positiv wahrgenommen. Das muss ich sagen, weil ich auch schauen konnte, wo stehe ich jetzt und das für mich sehr gewinnbringend war und ich das auch positiv konnotiert hab, ja wunderbar, hier kannst du was mit anfangen.“

Auch die Unterscheidung eines Ausbildungs- und Examenstandes konnte mit Hilfe der Skalen für R9 schlüssig erläutert werden:

R9: „[...] Ja er hat halt immer darauf hingewiesen, zum jetzigen Stand ist es so und so und zum jetzigen Ausbildungsstand ist das gefordert, das haben sie erfüllt, aber im Examen muss das Kriterium xy noch mit eingeführt werden. Also er hat das schon transparent gemacht. Und ich konnte auch gut was damit anfangen.“

R11 berichtet hingegen, dass sie keine Einschätzung über den Nutzen der Skalen abgeben konnte, da sie während der Erprobung gar nicht ausgefüllt worden seien, was wiederum auf den nur lückenhaften Einsatz des Bewertungskonzepts bei A6 hinweist.

Die Frage der ‚Praxistauglichkeit‘ des Bewertungskonzepts wird von den Ausbildern und Auszubildenden wieder einstimmig beurteilt und zwar negativ. Alle Akteure heben die fehlende ‚Praxistauglichkeit‘ des Bewertungskonzepts hervor. Für A6 liegt diese Sichtweise in der bereits ausgeführten Logik begründet, mit der das Bewertungskonzept versucht Unterricht zu bewerten und die nach seiner Meinung nicht dazu geeignet ist, die „systemischen Zusammenhänge“ guten Unterrichts zu erfassen. Vielmehr habe er durch die Erprobung des Konzepts erst erfahren, „was ich daran habe, mich verstehend in Stunden hineinzuschreiben“:

A6: „[...] Ich habe, weiß jetzt überhaupt erst, was ich daran habe, wenn ich mich in Stunden verstehend hineinschreibe und nicht gleich messen muss auf welcher Ebene ich was, wo, wie beurteile und wenn ich die Beurteilung weitestmöglich zurückschieben kann und auch noch mal in, nach dem Punkt herausschieben kann, wenn das Gespräch auch mit der Referendarin, dem Referendar abgeschlossen ist.“

R9 attestiert dem Konzept aus einem anderen Grund eine mangelnde ‚Praxistauglichkeit‘, denn es sei viel zu umfangreich und unübersichtlich, so ihr prägender Eindruck, den sie als Reaktion auf die offene Impulsfrage der Intervieweröffnung formuliert:

I: Wenn Sie an den Einsatz dieses von mir entwickelten Bewertungskonzepts zurückblicken, was fällt Ihnen da spontan als Erstes ein?

R9: Viel, viel, wie soll man das sagen, Blätter. Viel, viel, ja wie nennt man das, Blätter/Ja viele Blätter, die man ausfüllen muss, so das fällt mir spontan ein.

Durch den Umfang und die Komplexität des Konzepts fühle man sich v.a. am Beginn der Ausbildung „erschlagen“:

R9: „[...] Ja, man wurde am Anfang von diesem Konzept schlichtweg erschlagen, das habe ich auch, ja ganz am Anfang, ja vergessen zu sagen. Da diese Papierflut/Man wurde von einer Papierflut überschll erschlagen, also diese formalen Sachen. Und dann auch, gerade was das Fachliche angeht, diese ganzen Fachbegriffe, die man einfach noch nicht zuordnen konnte. Vieles habe ich schon intuitiv gemacht, das maße ich mir an zu sagen oder mit meinem Verstand geregelt, aber das, didaktisch, inhaltlich und fachlich so auf den Punkt zu bringen, wie das hier ist, dazu bedurfte es einige Monate um das auch so verknüpfen zu können. Und das hat gedauert einfach. Also das, wie gesagt von Anfang an finde ich das sehr, sehr schwer.“

Das ist ein Kritikpunkt, den auch R11 und R10 so unterstützen, wobei letzter formuliert, dass die Matrix auf Grund ihrer Komplexität, gerade für die Vorbereitung der Nachbesprechung „zu sperrig“ sei und er deswegen auf das Qualitätstableau des ZfsL ausgewichen sei. Als weiteren Kritikpunkt merkt er an, dass das Konzept den Anspruch, die Kompetenzentwicklung über einen längeren Zeitraum abzubilden, nur eingeschränkt einlösen konnte, da die Unterrichtsnachbesprechungen „in verschiedenen Phasen der Ausbildung ganz verschiedene Schwerpunkte hatten“ und ein Vergleich der erreichten Kompetenzstufen innerhalb eines Schwerpunktes deswegen über die Ausbildung hinweg nicht geleistet werden konnte:

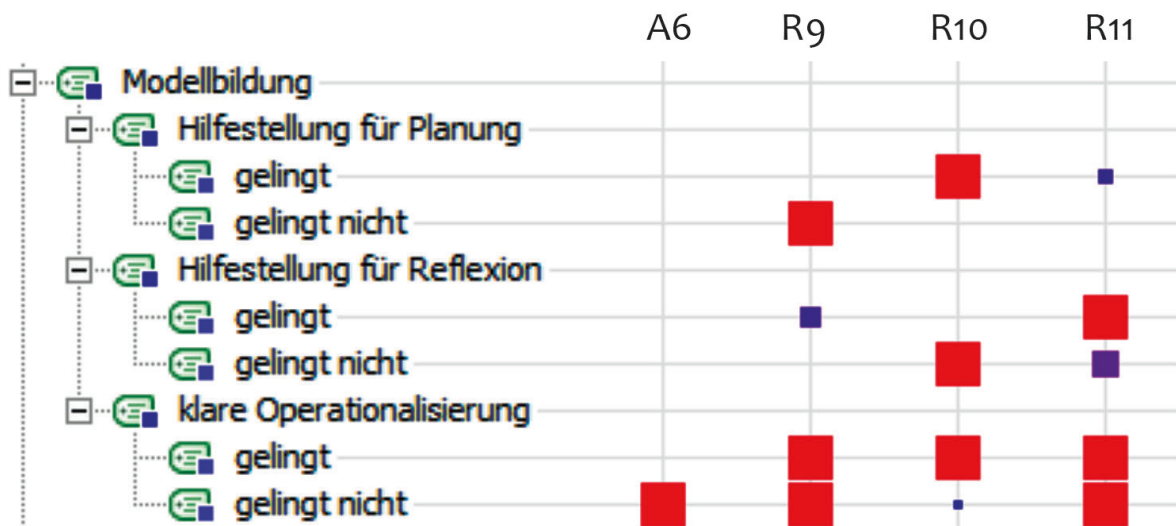
R10: „[...] Zugleich war das Problem sicherlich, dass die Unterrichtsnachbesprechungen in verschiedenen Phasen der Ausbildung ganz verschiedene Schwerpunkte jedes Mal hatten, und tatsächlich dieser Bogen überwiegend nur schwerpunktmäßig ausgefüllt wurde und sich eben auch mit der Verschiebung der, für die Ausbildung relevanten, zu entwickelnden Kompetenzen, eben auch diese durchgängige Transparenz nicht herstellen konnte. Also beispielsweise zu Beginn der Ausbildung ging es um eine sinnvolle Phasierung des eigenen Unterrichts, um die Herstellung von Transparenz für die Lernenden, während dann am Ende der Ausbildung ein Punkt wie sinnstiftendes Kommunizieren und nachhaltige Ergebnissicherung mehr in den Fokus gerückt sind.

Die Auswertung dieser drei Subkategorien für die Konzepterprobung bei A6 zeigen teils Übereinstimmungen, teils Abweichungen von den Ergebnissen der anderen Erprobungskontexte. Einen ‚Konflikt mit der Individualität der Ausbildungsbedürfnisse‘ sieht auch in diesem Fall keiner der Akteure, wie es schon in den zuvor diskutierten Kontexten der Fall war. Hinsichtlich der Rolle der ‚Skalierung‘, mit der das Konzept versucht Kompetenzfortschritte zu erfassen, formuliert A6 weitreichende Kritik an der generellen Logik, auf diese Art und Weise Unterrichtsqualität und Handlungskompetenzen erfassen zu wollen. Die systemische Qualität von Unterricht kann seiner Meinung nach nicht durch dieser Art der Skalierung erfasst werden. Zwar teilt R10 diese tiefgreifende Kritik von A6 nicht, aber eine wirkliche Hilfestellung für produktive Bewertungsprozesse waren die

Skalen für ihn nicht, wohingegen R9, wie bereits zuvor andere Auszubildende, von dem positiven Einfluss der Skalierung auf die Bewertungsprozesse berichtete. Die bereits von mehreren Akteuren geforderte Reduzierung des Konzeptumfangs, zur Gewährleistung der ‚*Praxistauglichkeit*‘ tauchte auch in der Erprobung bei A6 wieder auf, v.a. vor dem Hintergrund der zu hohen Komplexität für den Beginn der Ausbildung, wie sie auch schon A2 beklagte. Wie schon zuvor R8 berichtet auch R10, dass er als Reaktion auf den zu großen Umfang der Matrix auf die weniger umfangreichen Instrumente des ZfsL auswich.

Abbildung 5.16 zeigt die Code-Matrix der Subkategorie ‚*Modellbildung*‘ für den Erprobungskontext von A6. Die Matrix zeigt eine sehr heterogene Rezeption des Bewertungskonzepts hinsichtlich der Subkategorie der ‚*Modellbildung*‘. Lediglich die Positionierung von A6 ist eindeutig, er äußert sich nämlich nur zur Frage der ‚*klaren Operationalisierung*‘ guten Unterrichts durch das Konzept – ausschließlich negativ.

Abbildung 5.16: Code-Matrix der Subkategorie ‚*Modellbildung*‘



Zu den weiteren Subkategorien finden sich keine Interviewpassagen des Ausbilders. Das Feedback der Auszubildenden weist keine klare Linie auf, in jeder der Subkategorien finden sich Bewertungen sowohl auf den positiven, als auch auf den negativen Codeseiten. Hierbei sticht sicherlich die Subkategorie der ‚*klaren Operationalisierung*‘ heraus, für die sich bei *allen Auszubildenden* sowohl positive, als auch negative Codings finden lassen. Hinsichtlich der Hilfestellungen, die das

Konzept für die ‚*Reflexion*‘ und die ‚*Planung*‘ von Unterricht bietet, sind die Rückmeldungen zwar durchaus heterogen, aber bis auf jene von R11 hinsichtlich der ‚*Hilfestellung zur Reflexion*‘, die das Konzept bietet, sind sie für die einzelnen Personen eindeutig auf einer Seite der Subkategorien verortet.

Zumindest für zwei der drei Auszubildenden aus dem Seminar von A6 war das Bewertungskonzept eine Hilfestellung für die Planung von Unterricht. Hierbei heben sowohl R10 als auch R11 hervor, dass insbesondere die genauen Operationalisierungen der Qualitätsbereiche guten Unterrichts in der Matrix eine Hilfe bei der Planung und Vorbereitung von Unterrichtsbesuchen waren. R11 berichtet, dass die Matrix ihr gezeigt habe, „worauf muss ich achten“ und geholfen habe „Formulierungen zu finden“:

I: „Können Sie ein bisschen näher umreißen, warum Sie glauben, dass die Matrix helfen kann oder wo sie Ihnen geholfen hat, wo Sie da mal reingeguckt haben?“

R11: Ja, also ich finde tatsächlich, dass der Teil, also die beiden Erläuterungen helfen, v.a. wenn man nicht so genau weiß: Ok, was bedeutet denn jetzt, Aufgabenstellung soll verständlich und in sich stimmig sein oder wo ist denn das Ziel nachher? Und dass man dann das einfach noch mal, einfach tatsächlich auch um Formulierungen zu finden, also was ist denn der, worauf muss ich darauf achten? Dann hilft es wahrscheinlich. [...]“

Letzten Endes formuliert auch R11, wie schon zuvor R8, hier in eigenen Worten genau jenen Aspekt der Professionalisierung, den A3 in der Gruppendiskussion mit dem Ausdruck des „sprachfähig Werdens“ umriss. Die Matrix unterstützt R11 darin, ihre Überlegungen hinsichtlich der Planung von Unterricht mit dem Fachvokabular verbalisieren zu können. Sie sieht darin einen wesentlichen Schritt zur Erlangung von Unterrichtskompetenz im Sinne der Kompetenz als Expertise. Auch R10 berichtet in ähnlicher Art und Weise über den Beitrag, den die Matrix für seine Professionalisierung geleistet hat. Für ihn war sie „so eine Art Checkliste“, auf die man „immer zurückgreifen“ konnte, wenn man einen „konkreten Bereich des Unterrichtsbesuches“ vorbereitet hat.

I: „Sie hatten ja schon gesagt, die Matrix hatten Sie durchaus mal herangezogen in der Unterrichtsvorbereitung. Hat Ihnen die Matrix, wie hat die Ihnen da geholfen? Welche Bedeutung spielte die?“

R10: Na ja, die Matrix, das war im Wesentlichen so eine Art Checkliste. Wenn man da irgendeinen konkreten Bereich des Unterrichtsbesuchs vorbereitet, konnte man immer zur Matrix greifen und kurz überlegen: Für diese Phase der Stunde beispielsweise ist das drin, ist das drin, ist das drin, bzw. brauche ich das und ist es dann drin, wenn ich es konkret brauche? Also in dieser Übersichtsdarstellung ganz gut, eben auch weil es wesentlich prägnanter noch mal und an der Stelle differenzierter und differenzierter nützlich war bei der Vorbereitung, als das, was wir vom Seminar haben, weil das ist ja dann doch recht Vage, sinnstiftendes Kommunizieren, ja schön ne, kann alles und nichts sein.“

Explizit führt R10 hier aus, dass für ihn die Matrix auf Grund ihrer Differenziertheit für die Vorbereitung von Unterrichtsbesuchen besser geeignet gewesen ist als die Instrumente des ZfsL. Ein interessanter Punkt, wurde diese Komplexität doch z.B. von A2 als Hinderungsgrund für den erfolgreichen Einsatz benannt. Für R9 war das Bewertungskonzept und dabei auch die Matrix der Qualitätsbereiche nicht hilfreich, um ihre Unterrichtsbesuche vorzubereiten. Sie beklagt eine mangelnde Klarheit über den Zweck der Matrix; ihr „sei einfach nicht klar gewesen, was sie damit machen sollte“. Dies ist ein relativ verblüffendes Urteil, wenn man sich vor Augen führt, dass das Bewertungskonzept den Auszubildenden in einem 30-minütigen Vortrag vorgestellt wurde, der Einsatz in einer Mail ein weiteres Mal erläutert wurde mit der Bitte, Unklarheiten durch Rückfragen zu klären, und die Funktion, welche die Matrix haben sollte, explizit auf ihrer ersten Seite erklärt war. Für R9 fehlte eine „umfangreiche Instruktion“, um das Bewertungskonzept und die Matrix sinnvoll nutzen zu können. Allerdings boten die Instrumente des ZfsL R9 auch keine große Hilfestellung bei der Vorbereitung ihrer Unterrichtsbesuche, da sie die Planung „mehr intuitiv“ gemacht habe:

R9: „[...] Und wenn ich Unterricht geplant habe, was heißt die Dinge vom Seminar? Also ich habe da auch nicht gegessen und die Meyerkriterien gemacht, sondern: Welche Lerngruppe sitzt vor mir? Also ich habe das mehr intuitiv gemacht, natürlich darauf abgestimmt, auf die Fachseminare, aber eher für mich: Was gibt es, was habe ich für eine Lerngruppe, wie stehe ich zu der Lerngruppe, welche Methoden und welche Inhalte kann ich daran anschließen? Das hat jetzt weniger auch was mit diesem Konzept zu tun oder mit dem Konzept des Seminars.

I: Ok, das klingt für mich jetzt so, als hätten Sie für sich, als würden Sie bei der Unterrichtsplanung ohnehin sehr stark so vor gehen, auf Grund Ihrer, sage ich mal, Ihrer eigenen schon erwachsenden Expertise und dem was Sie empfinden. Sie nicken, also dass Sie eher, klingt für mich jetzt so, lernen und auch Planung, eine Professionalisierung auf Grund eigener Erfahrung haben.

R9: Oft. Also, ich glaube ich bin einfach, natürlich war ich nicht ausgebildet und man hat viel zu lernen, aber von Anfang an stand immer der Schüler, der Mensch und ich, Erziehung und Beziehung und daran schließt sich Methode, Didaktik und Inhalt an. Das ist auch so mein Leitbild für mich, und so habe ich von Anfang an gehandelt. Natürlich wurde ich da, Gott sei Dank, auch unterfüttert während meiner Ausbildung, sonst hätte ich sie ja nicht machen müssen. Aber das sind, ist einfach meine Einstellung die jetzt auch wirklich weniger mit diesem Entwurf zu tun hat, ja wo ich mich, daran habe ich mich weniger festgehalten.

R9 formuliert die für die Erprobung wesentliche Quintessenz selber: Der mangelnde Einsatz des Bewertungskonzepts und v.a. der Matrix bei der Unterrichtsvorbereitung hat in ihrem Fall aller Wahrscheinlichkeit nach „weniger mit diesem Entwurf“ zu tun, als vielmehr etwas mit ihrer etwas eigensinnigen Professionalisierungsstrategie, ihrem „Leitbild“, wie sie es nennt: ihrer eigenen „Intuition“ mehr zu vertrauen als von außen kommenden Ausbildungskonzepten. Inwiefern eine „umfangreiche Instruktion“ über den Einsatz des Bewertungskonzepts diese Strategie hätte ändern können, sei dahingestellt.

Gleichzeitig berichtet R9 allerdings darüber, dass die Matrix ihr durchaus eine ‚Hilfestellung für die Reflexion‘ im Rahmen der Unterrichtsnachbesprechung geboten habe. „Nachdem ich diese verstanden habe“, konnten auch in ihrem Fall die Matrix und dabei insb. die Indikatoren dabei helfen, die Reflexion im Rahmen der Nachbesprechung vorzubereiten:

I: Ok, wo haben Sie genau rein geguckt? Eher in die Matrix mit den Qualitätsbereichen oder eher in den Dokumentationsbogen nochmal geguckt?

R9: In die Matrix. Nachdem ich sie verstanden, habe ich dann schon in diese Matrix geschaut und gerade hier bei den Beispielen von Indikatoren vieles entnommen um dann in meinem Statement unterzubringen.

I: War das eine Hilfestellung, die wesentlichen Qualitätsbereiche in den Blick zu nehmen, der Stunde? Die wesentlichen Qualitätsbereiche, die die Stunde haben erfolgreich werden lassen oder auch vielleicht haben nicht so erfolgreich werden lassen?

R9: Ja, also in meinem Fall schon. Also nachdem, wobei ich sage, nachdem ich es dann verstanden habe, habe ich mir das hier schon vorgenommen, ja.

R11 berichtet zwar darüber, dass sie das Bewertungskonzept nicht für die Vorbereitung der Unterrichtsnachbesprechung genutzt habe, aber Gleichwohl lässt sich aus ihrem Bericht schlussfolgern, dass insb. die Matrix eine Rolle als Hilfestellung bei der zunehmenden Ausbildung der Reflexionsfähigkeit spielen kann. R11 nutzte das Qualitätstableau des ZfsL, sowie die Bubbles für die Vorbereitung der Unterrichtsnachbesprechung, aber hätte an statt des Qualitätstableaus „genauso gut die Matrix ran holen können“, wie sie selbst formuliert:

I: „Haben Sie das Konzept, die Matrix oder andere Instrumente des Konzepts herangezogen um Ihr Statement zur Stunde abzugeben und produktiv über Ihre Stunde reflektieren zu können?“

R11: Nee, dafür habe ich dann tatsächlich immer die Bubbles genommen, weil das auf einem DIN A4 Blatt relativ schnell übersichtlich war und daran habe ich es dann aufgebaut. Ich habe nachher, als es Richtung UPP ging und Richtung Examen ging, da habe ich mir damals aber tatsächlich das Qualitätstableau ran gezogen, das ja, wie gesagt, fast genauso aufgebaut ist um meine Nachbesprechung im Vorhinein, also ich habe die Nachbesprechung, habe ich immer schon, bevor, also habe ich am Tag vorher schon vorbereitet, weil ich ja ungefähr wusste, was ich da erzählen möchte. Und dafür habe ich nachher dieses Tableau herangeholt, aber da hätte ich wahrscheinlich auch genauso gut das hier ran holen können, das Tableau war für mich einfach nur schneller zugänglich.“

Auch für R11 waren die auf einer Seite zusammengefassten Bubbles das sinnvollste Instrument, um unmittelbar nach dem Unterrichtsbesuch die Nachbesprechung vorzubereiten. Das ist ein weiterer Hinweis, dass ein komplexes Instrument in dieser Ausbildungssituation wahrscheinlich wenig Einsatz finden wird. Im Vorfeld von Unterrichtsbesuchen und in späten Ausbildungsphasen, also etwa in der Vorbereitung auf das Examen, hat das ausdifferenziertere Instrument des Qualitätstableaus für R11 aber eine zunehmend wichtigere Rolle gespielt, womit deutlich wird, dass die Instrumente je nach Komplexitätsgrad individuelle didaktische Funktionen wahrnehmen. Der Einschätzung von R11, dass das Qualitätstableau und auch die Matrix zu umfangreich waren, um während der Unterrichtsnachbesprechung sinnvoll eingesetzt zu werden, schließt sich auch R10 an:

R10: „[...] Wir haben natürlich, wir haben auch so eine Form „Meyer plus“, also wir haben Meyers Kriterien und dann noch mal vier vom Bonner Seminar spezifisch und natürlich man muss sagen, es ist in dieser massiven Ausdifferenzierung, entspricht sicherlich fachwissenschaftlichen Standards, ist gleichzeitig aber praktisch im Rahmen einer Unterrichtsbesprechung so nicht nutzbar.“

Interessant an seinem Bericht ist weiterhin, dass auch er auf Grund seines recht individuellen Umgangs mit den Ausbildungsanforderungen angebotene Instrumente wahrscheinlich kaum für die Unterstützung der Reflexion der eigenen Unterrichtsstunde nutzen würde, denn er bezeichnet sich selbst als „ad-hoc-Statement-Typ“:

I: „Ok, d.h. für die Vorbereitung Ihres Statements, für die Reflexion war das zu viel, zu ausdifferenziert?“

R10: Das war zu viel. Ich muss dazu aber auch noch sagen, ich bin eher ein ziemlicher ad hoc Statement-Typ. Also ich habe mir selten länger als fünf Minuten für die Statements Zeit genommen.“

Der Bereich der ‚klaren Operationalisierung‘ ist die erste Subkategorie zu der sich A6 äußert. Hier bringt er erneut seine relativ fundamentalen Einwände gegen das Bewertungskonzept vor, und stellt sie seinem „hermeneutischen Vorgehen“ gegenüber:

A6: „[...]Wir versuchen wirklich, eigentlich mit einem hermeneutischen Vorgehen da ranzugehen. Und wir versuchen dann im Blick kleine Ausbildungssituationen zu fokussieren, die wir hinterher im Gespräch zum Thema machen. [...] Und es gab auch einen Fall, in dem ich das Instrument eingesetzt habe, in dem ich hinterher gar nicht zu einer Beurteilung kommen konnte, weil die Referendarin eine Situation gewählt hatte, die sich als Bewertungssituation nicht eignet, als Leistungssituation nicht eignet. Ich konnte die trotzdem gut beraten in diesem Punkt, also es war eine Stunde zur Klausurvorbereitung, und es war eine Stunde, wo die die ganze Stunde vorne an der Tafel gestanden hat und in einem eng gelenkten Unterrichtsgespräch genau das Tafelbild hinbekommen hat, was ihr wichtig war, dass es die Schülerinnen und Schüler hatten, um am nächsten Tag die Klausurfragen zu können. Verstehe ich sofort, dass solche Stunden nötig sind, sind aber als Leistungssituation in dem Sinne, wie wir Referendarausbildung betreiben, überhaupt nicht geeignet.

I: D.h., wenn ich da schon mal kurz einhaken darf, die Stunde, wie sie gesehen wurde, hatte ihr Recht an diesem didaktischen Ort, sie war auch so für guten Unterricht begründbar in einem größeren Sinne, dass das sinnvoll war, das zu machen, aber es war mit diesem Schema, das hinter dem Bewertungskonzept lag, was guter Unterricht ist, nicht deckungsgleich, nicht passend?

A6: Richtig, richtig. Und solche Missverhältnisse können eben auch sonst vorkommen, dass eine Stunde sich sehr an den Bedürfnissen der Lernenden orientiert und nur ganz einzelne Dinge in den Blick nimmt und trotzdem guter Unterricht ist.“

Aus der Perspektive von A6 hätte die konkrete Ausbildungssituation auf Basis der Qualitätskriterien guten Unterrichts, die das Bewertungskonzept formuliert, als schlecht bewertet werden müssen, obwohl „solche Stunden nötig sind“. Im Sinne von A6 muss also geschlussfolgert werden, dass die Operationalisierung guten Unterrichts, die das Konzept leisten will, eben doch nur situativ zu interpretieren ist und keine allgemeine Erfassungsgrundlage bieten kann. A6 erblickt deutliche Parallelen zwischen der Art, wie das Bewertungskonzept Unterrichtsqualität erfassen will, und dem Vorgehen der Schulinspektion in Nordrhein Westfalen, der Qualitätsanalyse, durch die nur eine oberflächliche Bewertung von „äußeren Abläufen“ geschehen könne, aber keine Bewertung von „Verstehenstiefe und solchen Dingen mehr“:

A6: „[...] Die Qualitätsanalyse kümmert sich ja auch keinen Deut um so was wie Verstehenstiefe, Anwendung von Kompetenzmodellen, etc., die gucken, ob die Tische so stehen, ob die Kinder auf den Plätzen sitzen, ob die Computer arbeiten, also eine ganz von außen gerichtete Unterrichtswahrnehmung.

I: Das würde für mich jetzt spannend sein, das wäre für mich spannend: Als Fachausbilder, meinen Sie, dass die Qualitätsanalyse auf diese Art und Weise beurteilen kann, ob ein Unterricht gut ist?

A6: Nein. Unterricht gut ist im Sinne von äußeren Abläufen, aber sie kann darüber nichts sagen über Verstehenstiefe, solche Dinge mehr. [...]“

R11 als die Auszubildende, um deren Unterricht es bei der oben geschilderten Ausbildungssituation ging, zieht fundamental andere Schlüsse. Für sie hat gerade die aus ihrer Perspektive misslungene Unterrichtsnachbesprechung gezeigt, dass das Konzept durchaus in der Lage ist, guten Unterricht zu operationalisieren, es vielmehr die fehlende Flexibilität des Ausbilders zeige, sich von den individuellen Qualitätsvorstellungen zu lösen:

R11: „[...] Also der UB, der mir da besonders im Gedächtnis bleibt, war der, als er halt beschlossen hat, dass es keine Leistungssituation ist, weil vorher zwei Stunden ausgefallen sind und in der nächsten Stunde die Klausur geschrieben wurde. Die Stunde lief genauso, wie ich das geplant hatte und ich war eigentlich relativ zufrieden mit der Stunde, und er war aber der Meinung: Ja nee, man könnte es ja nicht so offen gestalten wie das, also man kann nicht diese Zauberstunde gestalten, wie es für einen UB ja verlangt würde und deswegen könnte man den nicht bewerten, weil es keine Leistungssituation wäre. Das Problem ist, dass man auf der Matrix, also auf diesen verschiedenen Skalen, alles positiv ankreuzen konnte, aber halt für ihn das im Prinzip schlecht war und es keine Note gab, weil es nicht der Situation entsprach. [...] Wie kann es sein, dass wir dann auf den Skalen alles positiv ankreuzen, es aber dann trotzdem negativ ist oder darin endet, dass er mir keine Note geben möchte für diesen UB?

I: Klingt jetzt erst mal sehr suspekt, so auf den ersten Blick. Ich höre da auch raus, Ihnen ist es auch bis jetzt nicht so ganz transparent?

R11: Nein, immer noch nicht.

I: Ok, meinen Sie, das hatte was mit diesem Bewertungskonzept zu tun oder ist die Intransparenz, wäre die auch ohne Konzept?

R11: Die wäre auch ohne dieses Bewertungskonzept, aber dadurch ist es erst recht nochmal aufgefallen. [...] Das Gute ist, dass eigentlich durch dieses Konzept klar wurde, was mir klar war, nämlich, dass die Stunde an sich gut war. Also, dass halt alle Kriterien ausreichend erfüllt wurden und es dadurch quasi noch komischer war zu sagen: Ja, aber es war halt keine Leistungssituation und deshalb kriegen Sie keine Note. Also, so gesehen könnte man sagen, der Transparenz wegen hat dieses Konzept durchaus seinen Sinn erfüllt, weil ich dadurch gesehen habe, was alles gut gelaufen ist, aber es hat halt gegen seine Äußerungen dann gesprochen und die Begründungen war dann irgendwann falsch.

Während A6 auf Nachfrage angibt, die Stunde hätte auf Grund der Operationalisierung durch das Bewertungskonzept schlecht bewertet werden müssen, interpretiert R11 die Quintessenz der Nachbesprechung vollkommen anders: die Bewertung im Kontext des Qualitätsrahmens des Bewertungskonzepts sei positiv ausgefallen; der Ausbilder habe auf Grund seiner individuellen Meinung über sinnvolle und nicht sinnvolle Ausbildungskontexte die Bewertung verweigert. Insofern ist die Frage, inwiefern in den Anwendungssituationen von R11 eine ‚klare Operationalisierung‘ guten Unterrichts gelungen ist, sehr zwiespältig zu beantworten. Einerseits konnte aus der Perspektive von R11 mit dem erprobten Bewertungskonzept sehr wohl belegt werden, dass ihr Unterricht gelungen war, eben dadurch, dass auf den Skalen der einzelner Qualitätsbereiche hohe Bewertungen abgegeben wurden. Andererseits verneint A6 diese Logik und interpretiert die Bewertung des entsprechenden Unterrichtsbesuches in zweifacher Hinsicht vollkommen anders: Erstens sei dieser auch auf Basis des vom Bewertungskonzept vertretenen Qualitätsverständnisses als schlecht zu bewerten. Wodurch offensichtlich wird, dass zwischen R11 und A6 massive Missverständnisse hinsichtlich der Endbewertung des Unterrichtsbesuches im Raum stehen blieben. Zweitens negierte A6, dass auf die vom Bewertungskonzept angestrebte Art und Weise Unterrichtsqualität überhaupt zufriedenstellend erfassbar sei; allenfalls könne es höchstens eine Bewertung von „äußeren Abläufen“ gewährleisten. Somit wird auf jeden Fall deutlich, dass das Bewertungskonzept seinen zentralen Anspruch, in Frage von Konflikten und Missverständnissen zwischen den Akteuren eine Klärungsgrundlage zu bieten, verfehlte.

Auch bei R9 und R10 wird die Frage, inwieweit das Bewertungskonzept eine ‚klare Operationalisierung‘ guten Unterrichts anbietet, nicht eindeutig beantwortet. Zwar berichtet R9 darüber, dass die Skalierung der einzelnen Qualitätsbereiche für sie „total sinnvoll“ war, aber hierbei scheinbar v.a. der prozessuale Charakter der Kompetenzerfassung positiv wahrgenommen wurde, wie bereits in der Subkategorie der ‚Skalierung‘ für R9 erörtert wurde und weniger die Tatsache der Operationalisierung der einzelnen Qualitätsbereiche hilfreich war. In Bezug auf diese Operationalisierung wiederholt R9 vielmehr ihre bereits zuvor geäußerte Kritik, dass die Matrix in dieser Operationalisierung viel zu umfangreich war, sodass „man einfach nur am blättern war“:

„I: Dass es hier diese Matrix gab, wo in verschiedenen, wo die Unterrichtsqualität in mehrere Qualitätsbereiche aufgegliedert war. Hat das geholfen, die Transparenz zu erhöhen warum Unterricht gut ist oder warum Unterricht als nicht so gut angesehen worden ist?“

R9: (...) Würde ich erst mal verneinen wollen, weil es einfach viel war. Also, ja weil man einfach nur am blättern war und/Also ich sage im ersten Teil der Ausbildung überhaupt nicht, nee, nein. Weil man damit, ja, nee, noch nicht so was anfangen konnte, um es selber einzuordnen, für sich, nee.“

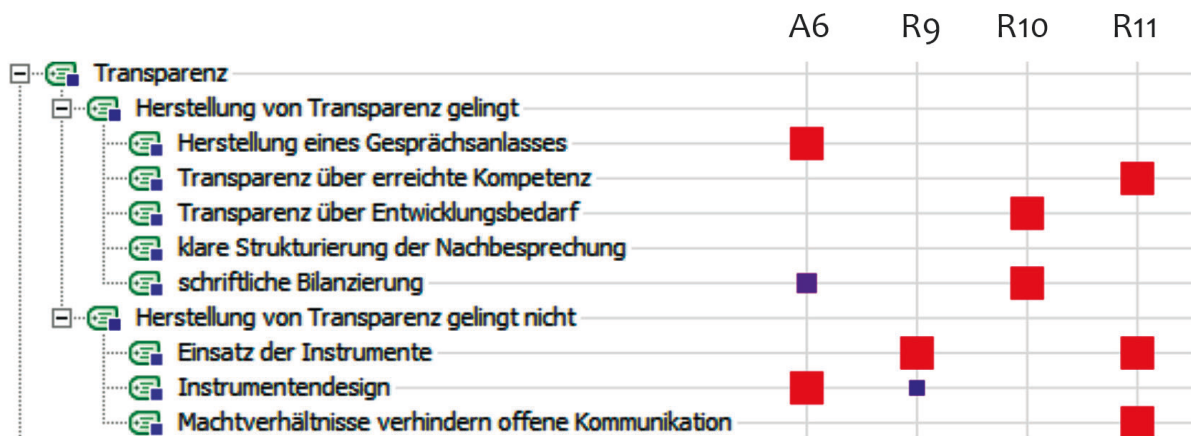
Hinsichtlich der Operationalisierung guten Unterrichts über klar voneinander abgegrenzte Kriterien gibt auch R10 ein widersprüchliches Feedback. Einerseits führt er aus, dass die Operationalisierung guten Unterrichts im Rahmen der Matrix, „sicherlich fachwissenschaftlichen Standards“ entspricht, andererseits sieht er einen Konflikt mit der „realen unterrichtlichen Praxis“, sodass auch im Fall von R10 offen bleibt, inwiefern das Bewertungskonzept im Ausbildungskontext von A6 eine ‚klare Operationalisierung‘ guten Unterrichts anbieten konnte, welche die Professionalisierung der Auszubildenden unterstützte.

Betrachtet man die Rekonstruktion der Akteurseinschätzungen zur Leistung des Bewertungskonzepts hinsichtlich einer ‚Modellbildung‘ zusammenfassend, bestätigen sich bereits herausgearbeitete Erkenntnisse. Offensichtlich macht es Sinn, in ein Bewertungskonzept niveaudifferente Instrumente einzubeziehen, um einerseits den unterschiedlichen Ansprüchen und Fähigkeiten der Auszubildenden während der verschiedenen Phasen der Ausbildung entgegen zu kommen und außerdem die Einsetzbarkeit des Konzepts in verschiedenen Ausbildungskontexten sicherzustellen. Eine umfangreiche Matrix kann bei der Vorbereitung von Unterrichtsbesuchen helfen, ist aber auf Grund ihrer Komplexität nur eingeschränkt geeignet, um nach einem Unterrichtsbesuch die Nachbesprechung vorzubereiten. Das Feedback zeigt außerdem, dass das Konzept sich einem Konflikt stellen muss, der sich nicht ohne weiteres auflösen lässt: Einerseits eine allgemeingültige Operationalisierung guten Unterrichts anbieten zu wollen, andererseits

aber die je besondere Logik konkreter Unterrichtssituationen immer in Rechnung stellen zu müssen.

Abbildung 5.17 zeigt die Code-Matrix der Subkategorie ‚Transparenz‘ für die Erprobung im Fachseminar von A6. Wie schon die zuvor ausgewertete Subkategorie, so zeigt auch die Subkategorie der ‚Transparenz‘ ein sehr heterogenes Bild der Rezeption. Auffällig ist v.a., dass in vier von sieben Unterkategorien jeweils in dem Interview von lediglich einem Akteur entsprechende Passagen codiert werden konnten. Drei von diesen vier Subkategorien liegen im positiven Bereich hinsichtlich der Frage, ob das Bewertungskonzept dabei geholfen hat ‚Transparenz‘ über die Bewertungsprozesse herzustellen. Dies deutet darauf hin, dass im Erprobungskontext von A6 das Konzept seine Ansprüche, für transparente Bewertungsprozesse zu sorgen, nicht umsetzen konnte. Immerhin für zwei der negativ bewerteten Subkategorien finden sich Interviewpassagen von zwei Akteuren, aber auch dies bedeutet, dass lediglich die Hälfte der Teilnehmer der Erprobung darüber berichten, dass die Herstellung von ‚Transparenz‘ am ‚Einsatz der Instrumente‘ oder durch die ‚Machtverhältnisse‘ in den Beziehungen zwischen dem Ausbilder und den Auszubildenden scheiterte.

Abbildung 5.17: Code-Matrix der Subkategorie ‚Transparenz‘



Wie schon zuvor A2 bewertet es auch A6 positiv, dass die Unterrichtsnachbesprechung schriftlich dokumentiert wurde und sieht in dieser ‚schriftlichen Bilanzierung‘ gleichzeitig auch die ‚Herstellung eines Gesprächsanlasses‘, da sie die „Verpflichtung zur Verständigung formalisiert“:

I: „Haben Sie den Eindruck, dass diese Instrumente auf irgendeine Art und Weise dazu beigetragen haben Transparenz herzustellen?“

A6: Ja schon, in dem Sinne, dass es noch einmal in der einen oder anderen Weise diese Verpflichtung sich zu verständigen formalisiert. Das finde ich schon einen guten Punkt.

I: Also die schriftliche Fixierung, die Bilanzierung der Nachbesprechung?

A6: Ja, was ich sonst, also ich schreibe sonst eben so ein paar Sätze dazu.“

Allerdings wird nicht deutlich, ob A6 einen Mehrwert der Dokumentation, im Sinne des erprobten Ausbildungskonzepts, im Vergleich zu seiner bisher praktizierten Form der Dokumentation sieht. Offensichtlich ist diese Dokumentation durchaus komplexer als „so ein paar Sätze“, die A6 aufschreibt. Dies wird in seinen weiteren Ausführungen und einem Zitat einer Auszubildenden deutlich, welche die Gleichwertigkeit der beiden Dokumentationsformen hervorhob:

A6: „[...] Das fand ich auch einen ganz interessanten Aspekt, die letzte Referendarin [...] die sagte dann: Ja ich verstehe das, dass Sie, wenn Sie das in Ihrer Form machen, dass sie das als gleichwertig ansehen. Weil ich für mich eben eine Form gefunden habe, in der ich das immer in einer relativ, auch für mich formalisierten Form mache, das hinterher den Referendarinnen und Referendaren gebe.

I: Also das haben Sie sowieso, ohnehin immer schon gemacht?

A6. Ja schon seit zehn Jahren. Und habe diesen Bogen übrigens auch immer wieder überarbeitet, je nachdem, wenn eine neue OVP kam oder wenn irgendwas/Inzwischen habe ich einen für die ersten drei Besuche und für die letzten drei Besuche.“

Auch R10 sieht in der schriftlichen Dokumentation der Nachbesprechung eine Steigerung der ‚Transparenz‘ der Bewertung, wobei er dem erprobten Bewer-

tungskonzept explizit einen größeren Nutzen zuschreibt als der tradierten Form der Dokumentation von A6, da diese eben nicht „sinnhaft gewichtet“ habe:

I: „Sie haben jetzt selber das Wort Bilanzierung gerade genutzt, zumindest wie ich gehört habe. War das für Sie wichtig, nochmal diese schriftlich fixierte Dokumentation zu kriegen der Nachbesprechung?“

R10: Ja natürlich, also wenn man zwei Monate teilweise zwischen den UBs im gleichen Fach liegen hat, erinnert man sich natürlich kaum noch an diese Gespräche. Auch die Gesprächsnotizen sind natürlich, so wie man so Unterrichtsnachbesprechung macht, teilweise nicht immer ganz sinnvoll geordnet. Man nimmt Impulse auf, die hereinkommen. Also das war mehr so eine, ja eine Sammlung, die ich dann an persönlichen Notizen hatte oder ein Gesprächsprotokoll von ihm. Da war nichts sinnhaft gewichtet und dadurch, da er mir das eben noch über dieses Instrument sinnhaft gewichtet hat, konnte ich daraus ganz sicherlich noch was rausziehen.“

Für R10 ist dies die wesentliche Leistung des Konzepts, wie er in seinem abschließenden Fazit, das weiter unten dokumentiert ist, noch mal hervorhebt. R9 äußert sich nur diffus zu dieser Subkategorie mit dem bereits mehrfach zitierten Hinweis, dass die Skalierung der Qualitätsbereiche hilfreich war, aber weiteres Feedback zum Nutzen der schriftlichen Dokumentation formuliert sie nicht. Ebenso wenig R11, die, wie bekannt, darüber berichtete, dass ihr keine Dokumentationsbögen ausgehändigt wurden, insofern auch keine Steigerung der Transparenz stattfinden konnte.

Darüber, dass das Bewertungskonzept zu einer ‚klaren Strukturierung‘ der Nachbesprechung und damit zu ‚Transparenz‘ über die Erwartungen in dieser beitragen konnte, berichtet keiner der Akteure, was vor dem Hintergrund der ‚additiven‘ Nutzung des Bewertungskonzepts wenig verwunderlich ist. Entsprechend verhinderte u.a. der ‚Einsatz der Instrumente‘ in diesem Kontext der Erprobung, dass sie dabei helfen konnten, die ‚Transparenz‘ der Bewertungsprozesse zu erhöhen, wie z.B. bei R9, bei der der Einsatz des Bewertungskonzepts, „einfach eine begleitende Erscheinung“ war:

I: Ok und der Dokumentationsbogen für die Nachbesprechung, klang jetzt für mich so, der hat das jetzt, also dieser Dokumentationsbogen, der hat die Transparenz nicht erhöht, weil ohnehin schon ein Bogen da war, der das!

R9: Ja, also bei diesem Fachleiter schon. Es war ein Protokollbogen des Fachleiters da, der ausgefüllt wurde und deswegen war das einfach eine begleitende Erscheinung.

Ebenfalls zeigt das oben ausführlich diskutierte Beispiel der missverständlichen Unterrichtsnachbesprechung bei R11, dass das Instrumentarium durch den ‚*additiven*‘ Einsatz eher Missverständnisse auslöst, als dass es als Gesprächsgrundlage für die Klärung von Missverständnissen dienen kann. Für R11 zeigte die Bewertung ihres Unterrichtsbesuches an den einzelnen Skalen der Qualitätsbereiche, dass er als gelungen zu bewerten ist, für A6 sollte die Bewertung das Gegenteil aufzeigen.

In diesem Kontext berichtet R11 auch darüber, dass die ‚*Machtverhältnisse eine offene Kommunikation*‘ zwischen dem Ausbilder und ihr verhindert hätten, der Einsatz von Instrumenten zur Steigerung der Transparenz entsprechend Augenwischerei sei:

R11: „[...] Ich glaube, es liegt einfach am Fachleiter, ich bin mit xy einfach nicht gut ausgekommen, das weiß xy inzwischen auch, glaube ich, das weiß ich auch, das ist jetzt kein großes Geheimnis und ich finde, dass er jemand ist, der sehr schwer anzuzweifeln ist, weil er das direkt persönlich nimmt und das direkt in seinem Verhalten nachher widerspiegelt. Und von daher war das, hätte er sein Kreuz überall machen können und ich hätte gesagt: Ja, ist ok. Weil es also mir nicht geholfen hätte. Und dadurch war das Ausbildungsinstrument dann quasi hinlänglich. Also er hat es dann auch so, dann gemacht wie er das für richtig hält.

I: D.h., wenn ich das jetzt richtig verstehe: Steigerung der Transparenz, schön und gut, aber auf Grund der besonderen Beziehung zwischen Ausbilder und Auszubildender ist die Transparenz hinfällig, weil es eine Machtbeziehung ist?

R11: Ja, richtig, genau.

I: Der Sie sich auch einfach unterordnen mussten.

R11: Ja, also es blieb mir nichts anderes übrig. Ich habe am Anfang noch revoltiert, wie sich das auf xy Laune auswirkt. Meine Mitreferendarin hat das noch stärker gemacht, da hat es sich extrem nachher ausgewirkt, und von daher habe ich irgendwann angefangen abzunicken und habe versucht das Beste irgendwie rauszuholen und gut war es.

Insofern berichtet R11 hier von sich aus darüber, dass sie sich in Unterrichtsnachbesprechungen mit ihren persönlichen Überzeugungen und Einstellungen nicht offenbart, sondern sich auf Grund der Unterordnung und Bewertungssituation nach den (vermeintlichen) Vorstellungen ihres Ausbilders richtet, also in genau jenes dysfunktionale Interaktionsmuster gerät, das oben als wesentlicher Grund misslingender Ausbildung anzusehen ist. Die anderen beiden Auszubildenden berichten über keine derartigen Erfahrungen hinsichtlich problematischer Beziehungen zwischen dem Ausbilder und sich.

Mit R10 erläutert einer der Auszubildenden, dass das Bewertungskonzept einen Beitrag geleistet hat, *Transparenz über den Entwicklungsbedarf*^f herzustellen, da es „konkret die Baustellen“ benannt hat, an denen er arbeiten musste:

I: Haben Sie den Eindruck gehabt, dass Ihnen das Konzept hier eine Hilfestellung geliefert hat, um erfolgreiche Unterrichtsbesuche zu leisten?

A10: Es hat zumindest konkret die Baustellen, an denen ich arbeiten muss, ganz genau lokalisiert, es hat sich mir verdeutlicht, ich wusste genau, bei der Planung meiner nächsten Stunde muss ich v.a. auf diese Bereiche achten. Insofern war es eine Hilfe.

Auch in diesem Fall thematisieren weder A6, noch R9 oder R11 in ihren Interviews Aspekte, die sich in diese Subkategorie einordnen ließen.

Abschließend lässt sich für diese Subkategorie festhalten, dass R9 wiederum den Umfang des Konzepts und damit das ‚Instrumentendesign‘ dafür verantwortlich macht, dass es keinen Beitrag zur Steigerung von transparenten Bewertungsprozessen liefern kann, weil es eben durch seinen Umfang selbst nicht transparent sei. A6 wiederholt in diesem Kontext ihre Kritik an der generellen Logik, mit der das Konzept versucht, Unterrichtsqualität zu bewerten, denn die Auswahl einer geringen Zahl von Qualitätsbereichen sei nicht dazu geeignet, „ein komplexes Bild des Unterrichts“ wiederzugeben:

I: „Und die Instrumente, wenn man das mal gegenüberstellt, die Instrumente haben Sie nicht darin unterstützt diesen Prozess auch leisten zu können?“

A6: Dieses Instrument, ist, fand ich deswegen dann schwierig, weil es mir immer auch so das Gefühl gab, wenn ich jetzt nur zwei oder drei ausgewählt habe, dass ich zu wenig gesehen habe. Aber de facto glaube ich, so wie wir von Referendaren möglichst vertieftes Verstehen angeleitet wissen wollen, ist auch meine Aufgabe zu vertieftem Verstehen zu kommen und nicht zu einem Oberflächenverstehen auf vielen Ebenen. Und im Glücksfalle sind Referendarinnen und Referendare und ich, sehen wir denselben Fall. Es ist auch sicherlich in 80% der Fälle so, dass die Referendare hinterher sagen: Wo sehen sie Beratungsbedarf? Dass sie genau den Punkt nennen, den ich auch gesehen habe. Aber hier, das suggeriert mir eigentlich, ich soll auf ganz viele Punkte gucken und kann die nur noch oberflächlich wahrnehmen.

I: D.h. auch eine Auswahl, ich habe das jetzt rausgehört, eine Auswahl von nur zwei, drei Punkten aus dieser Matrix der Qualitätsbereiche führt dazu, dass Sie den Eindruck haben, die Stunde wird dann zwar an diesen zwei, drei Bereichen besprochen, aber das Ganze ist oberflächlich, das ist keine tiefgehende Analyse der Stunde?

A6: Wenn ich nur zwei, drei auswähle ist das zum einen häufig dann in Ordnung, aber es ergibt sich niemals hinterher ein komplexes Bild, also das sehe ich ja hier daran, wie viel leer hier ist bei meinen Repertoirebögen. [...]“

Weiterhin kritisiert A6, dass das Bewertungskonzept einerseits „ein zu grobes Raster“ für die Bewertung von wirklicher Unterrichtsqualität vorgegeben habe, aber zugleich ein zu „enges Raster“. Diese zunächst paradox anmutende Einschätzung klärt A6 durch den fehlenden fachdidaktischen Bezug des Konzepts auf:

A6: „[...]Wenn ich als Fachdidaktikerin ausbilde, dann habe ich selbstverständlich fachdidaktische Inhalte deutlich stärker im Blick als diese Allgemeindidaktischen. Und d.h., dieser wesentliche fachdidaktische Punkt, ich erinnere mich noch gut an die erste Diskussion, wo A2 immer gesagt hat: «Inhaltliche Klarheit steht über Allem» und inhaltliche Klarheit ist da so das Pasprototo für, wie ich inzwischen verstanden habe, damals habe ich noch gedacht, man kann es ja auch ein bisschen einseitig sehen. Inzwischen denke ich: Ja inhaltliche Klarheit heißt hier fachdidaktische Vermittlung. Und das ist so das Pasprototo dann für alles was Fachdidaktik bedeutet.

In diesem Sinne sei das Konzept einerseits zu weit gefasst, weil es fachdidaktische Beurteilungen nur im viel zu grob differenzierenden Bereich der „inhaltlichen Klarheit“ zulasse, andererseits viel zu eng, weil es allgemeindidaktische Aspekte in den Blick nehme, für die A6 sich als Fachausbilder „nur unter Fußnote 27 klein b“ interessiere.

Hinsichtlich des Beitrags des Konzepts, für ‚*Transparenz*‘ in den Bewertungsprozessen zu sorgen, können mehrere Aspekte als Fazit aus der Erprobung bei A6 benannt werden. Auch wenn nur zwei Akteure sich klar zu dieser Frage positioniert haben, wird deutlich, dass eine ‚*schriftliche Dokumentation*‘ und v.a. eine solche die eine standardisierte Art der Dokumentation gewährleistet, die ‚*Transparenz*‘ von Bewertungsprozessen, auch über einen längeren Zeitraum, sicherstellt. Die Erprobung hat auch gezeigt, dass das Nutzungsmuster des Konzepteinsatzes entscheidend über seine Leistung bestimmt. In einem ‚*additiven*‘ Setting kann das Konzept Stärken, die es in einem ‚*strukturierenden*‘ Einsatz ausspielte, nicht zeigen. Eng in Verbindung mit dem Einsatzmuster steht die Akzeptanz, die das Konzept auf Seite des Ausbilders genießt. Fehlt diese, wie im Fall von A6, der die generelle Logik der Bewertung

über isolierte Qualitätsbereiche ablehnt, so kann ein Instrumentarium auch nicht als Grundlage von Klärungsprozessen dienen, es sorgt nur für neue Missverständnisse. Schlussendlich zeigen sich aus der Perspektive von A6 starke Konflikte eines fachübergreifend angelegten Konzepts mit fachdidaktischen Bedürfnissen.

Das ‚Fazit‘, was sich aus Erprobung bei A6, R9, R10 und R11 ziehen lässt, ist, es war nicht anders zu erwarten, sehr heterogen (vgl. Abb. 5.18).

Abbildung 5.18: Code-Matrix der Subkategorie ‚Fazit‘



In jeder der drei Bewertungsstufen der Code-Matrix finden sich Codings von mindestens zwei Akteuren, wobei A6, R9 und R10 eine nach der rein deskriptiven Verteilung eindeutige Bewertung des Konzepts abgeben. In dem Interview von R11 finden sich hingegen Passagen, die sich in jede der drei Einschätzungen der Subkategorie einordnen lassen.

Die eindeutig negative Bewertung des Konzepts durch A6 wurde im Verlauf der Auswertung der ‚Konzeptbewertung‘ mehr als deutlich und beruht i.W. auf der Erfahrung, dass die Art, wie das Bewertungskonzept Kompetenzfortschritte erfassen will, der Art Unterricht zu bewerten, wie A6 sie für richtig hält, entgegenläuft, wie er auch den prägenden Eindruck der Erprobung als Antwort auf die Impulsfrage des Interviews formuliert:

I: „Wenn Sie an den Einsatz dieses Bewertungskonzepts, dieser drei verschiedenen Instrumente zurück denken, was fällt Ihnen das spontan ein?“

A6: Dass ich vorher sehr viel Positives davon erwartet habe, dass es zu einer Objektivierung von Unterrichtsbeobachtung führt. Dass ich inzwischen mir aber der Sache gar nicht mehr so sicher bin, ob es in diesem Maße dafür geeignet ist, weil ich glaube, dass die Herkunft von der Qualitätsanalyse solcher Instrumente her für einen ganz anderen Zugriff geeignet ist, als das

was, wir in der Ausbildung betreiben. Unsere Aufgabe ist nicht in erster Linie Unterricht ganz schnell zu bewerten, sondern Unterricht zu verstehen und das will die Qualitätsanalyse ja eigentlich gar nicht, sondern wir versuchen wirklich, eigentlich mit einem hermeneutischen Vorgehen da ran zu gehen. Und wir versuchen dann im Blick kleine Ausbildungssituation zu fokussieren, die wir hinterher im Gespräch zum Thema machen.“

Entsprechend kommt A6 zu dem Schluss, dass „Unterricht als komplexes System andere Arten von Abbildung als ein lineares Abbildungssystem braucht.“

Des weiteren zeigt die Erprobung bei A6, dass das Konzept seine Ansprüche nicht einlösen kann, wenn die grundlegende Konstruktionslogik vom Ausbilder abgelehnt wird. In diesem Fall führt es eher zu Verwirrung, als dass es Klärungsprozesse anstößt. Ein additiver Einsatz richtet mehr Schaden an, als dass er nützt, was z.B. das Fazit von R11 zeigt, die aus dem Einsatz des Konzepts durchaus geschlussfolgert hat, dass es die Transparenz steigern kann, aber die von ihr verstandene Bewertung von der von A6 Intendierten massiv abwich:

R11: „[...] Das Gute ist, dass eigentlich durch dieses Konzept klar wurde, was mir klar war, nämlich, dass die Stunde an sich gut war. Also, dass halt alle Kriterien ausreichend erfüllt wurden und es dadurch quasi noch komischer war zu sagen: Ja aber, es war halt keine Leistungssituation und deshalb kriegen Sie keine Note. Also so gesehen könnte man sagen, der Transparenz wegen hat dieses Konzept durchaus seinen Sinn erfüllt, weil ich dadurch gesehen habe, was alles gut gelaufen ist, aber es hat halt gegen seine Äußerungen dann gesprochen und die Begründungen war dann irgendwann falsch.“

Vordergründig kommt R11 also zu einer positiven Bewertung des Konzeptes, durch die aber im Sinne der Schaffung von Kommunikationsanlässen zwischen Ausbilder und Auszubildenden, als der zentralen Zielsetzung des Konzepts, nichts gewonnen ist. Ein weiteres Mal zeigt dieser Aspekt, welche Bedeutung die ‚Beziehung zwischen Ausbildern und Auszubildenden‘ für gelingende Bewertung hat und dass kein Instrument diese Probleme beseitigen kann.

Das Beispiel von R11 zeigt ebenso wie das von R9, dass Bewertungskonzepte nur dann einen Nutzen haben können, wenn auch die Auszubildenden diese als Richtschnur ihrer Professionalisierung anerkennen. Insofern treiben diese beiden Auszubildenden die, zurückhaltend gesagt, autonome Ausbildungssteuerung, wie sie in deutlich abgemilderter Form auch bei R13 zu finden war, auf die Spitze. Mit einer Einstellung, wie sie etwa R9 an den Tag legt, nämlich dass sie ihren Unterricht eher „intuitiv“ plane, wird jedes halbwegs standardisierte Bewertungskonzept hinfällig. Gleiches gilt, wenn man wie R11 als Auszubildende zum Fazit der Konzepterprobung kommt, dass man seine Unterrichtsentwürfe „nie an diesen Konzepten“ ausgerichtet habe, weil man selbst schon „eine umfassende Ausbildung als Dozentin an der Uni“ habe und sich nach „den eigenen Maßstäben“ ausrichte:

I: Haben Sie den Eindruck gehabt, dass das Konzept nichtsdestotrotz Ihnen auf irgendeine Art und Weise geholfen hat erfolgreiche Unterrichtsbesuche abzuliefern?

A11: Also ich richte meine/Ich richte/Ich habe, glaube ich, nie meine Unterrichtsentwürfe nach diesen Konzepten ausgerichtet. Also, es liegt aber auch daran, dass ich eine relativ umfassende didaktische Ausbildung hatte und selber Dozentin an der Uni bin, d.h. ich richte mich tatsächlich nach meinen Maßstäben und die haben auch immer ganz gut funktioniert, und ich kenne auch Helmke und Meyer in- und auswendig.

Aus dem Zusammenspiel dieser Ablehnung von standardisierten Bewertungskonzepten und dem ‚additiven‘ Einsatz durch A6 wird auch das ‚Fazit‘ von R9 mehr als nachvollziehbar, dass erfolgreiche Unterrichtsnachbesprechungen „auf die Persönlichkeit und die Kompetenz des Ausbilders“ und nicht auf das Bewertungskonzept zurückzuführen seien:

I: „Abschließend: Hat das Konzept dazu beigetragen, dieses Bewertungskonzept, dass Sie und Ihre Ausbilderin die Nachbesprechung produktiv gestalten konnten, dass Sie mit einem gefühlten Lernerfolg draus gegangen sind, dass Sie gesagt haben: Heute habe ich was gelernt hier?“

R9: Also, ich bin eigentlich aus jeder Sitzung oder aus jeder Nachbesprechung, ja mit gewinnbringend und mit Lösungsmöglichkeiten und Weiterentwicklungsmöglichkeiten rausgegangen. Ich würde es jetzt einschränkend auf das Konzept zurückführen wollen, als vielmehr auf die Persönlichkeit und die Kompetenz des Ausbilders.“

Hier wird entsprechend geschlussfolgert, dass für R9 eine Beurteilung des Bewertungskonzepts nicht möglich ist, da es der Bewertungspraxis von A6 untergeordnet wurde.

Für R10 hat der Einsatz des Konzepts „einen geringen Unterschied gemacht“, eben diese „gewichtende Zusammenfassung“ der Unterrichtsbesuche sei ein Mehrwert, aber darüber hinaus sieht auch er keinen positiven Beitrag des erprobten Bewertungskonzepts.

5.3.4 Passung mit fachlichen Qualitätsstandards

Ein zentrales Anliegen des hier zu entwickelnden Bewertungskonzepts war es, eine Bewertungsgrundlage anbieten zu können, die innerhalb einer gesamten Ausbildungsinstitution, also fachübergreifend eingesetzt werden kann. Dieser Anspruch, das wird im Verlauf der Auswertung der Subkategorie der ‚*Passung mit fachlichen Qualitätsstandards*‘ deutlich, wird nur schwer, wahrscheinlich sogar gar nicht umzusetzen sein. Diese Subkategorie ordnet sich der Kategorie der ‚*Konzeptbewertung*‘ unter, wird aber isoliert von den anderen Subkategorien ausgewertet. Dies geschieht weil durch den Bezug auf die Überlegungen aus der Perspektive eines bestimmten Unterrichtsfaches Angehörige des ZfsL dazu in der Lage sind, die Ausführungen der jeweiligen Studienteilnehmer konkreten Ausbildern des ZfsL zuzuordnen. Damit die Anonymität der Studienteilnehmer, v.a. hinsichtlich sensibler Fragen wie jener zu den Konflikten in den Ausbildungsbeziehungen, gewährleistet ist, wird die folgende Auswertung zwar Bezug auf das Ausbildungsfach nehmen, aber nicht auf die Zuordnung der Nummerierung der Probanden.

In den Interviews haben lediglich zwei der Ausbilder mit den Fächern Deutsch und Sport sich zu dieser Subkategorie geäußert. Hierbei äußerten sie ausnahmslos Bedenken, ob das Bewertungskonzept mit den fachlichen bzw. fachdidaktischen Bewertungskriterien in Einklang gebracht werden kann. Der Ausbilder des Faches Deutsch führt an, dass die wesentlichen fachdidaktischen Aspekte, die für die Bewertung einer Unterrichtsstunde sowie die Professionalisierung in dem konkreten Unterrichtsfach, entscheidend seien, nicht zufriedenstellend durch das Bewertungskonzept erfasst werden könnten. Vielmehr würden alle fachdidaktischen Aspekte „unter dem großen Hut“ der inhaltlichen Klarheit verhandelt, der dann in „ganz viele Einzelteile“ zerfällt, die mit Inhalten zu füllen eben seine Aufgaben gewesen sei:

Deutsch: „[...] Und ich habe dann, wo es nötig war, weitere Punkte hinzugefügt. Die hätte man teilweise auf dieser Ebene formulieren können, aber es gab z.B. Punkte, die sich hier nicht abbilden lassen, wie z.B. vertieftes Textverstehen, kommt hier unter dem Oberbegriff vor inhaltliche Klarheit oder so was. Aber, oder auch Lernaufgabenkonstruktion, das sind Dinge die im Fachdidaktikseminar, wenn Sie sich dran erinnern, riesige Rollen spielen, die aber hier nur unter dem großen Hut der inhaltlichen Klarheit sich wiederfinden, aber innerhalb der Fachdidaktik wird eben inhaltliche Klarheit, zerfällt in ganz viele Einzelteile und spielt eben dann einen entscheidende Raum und dann habe ich eben auch neue Faktoren angefügt. Aber das war einerseits ein zu genaues Raster, weil hier zu viele Faktoren vorkamen und andererseits ein zu grobes Raster, weil inhaltliche Klarheit hier!

I: Klingt ja erst mal paradox. Können Sie diese Paradoxie vielleicht doch noch mal näher umschreiben?

Deutsch: Ja, wenn ich, ich bin ja auch Kernseminarleiter, wenn ich als Kernseminarleiter ausgebildet hätte, hätte ich mich, glaube ich, hier dran eher abarbeiten können, weil ich dann ja den fachlichen Ertrag einer Stunde und wie dieser fachliche Ertrag durch die Lehrprozesse zu Stande kommt, nur peripher in den Blick nehme. Als Fachdidaktikerin ist das

selbstverständlich mein Hauptanliegen und damit war in diesem Punkt, dieser Punkt inhaltliche Klarheit zu umfassend und ich habe daraus einzelne Punkte herausgelöst. Und die zum Schwerpunkt gemacht, weil eben auch der inhaltliche Klarheitsbereich zu grob war um das zu erfassen, z.B. vertieftes Verstehen, richtiges Anwenden von Lesekompetenzmodellen, nehmen wir das mal als Beispiel. Aber andererseits, weil das eben der Knackpunkt der Stunde war! [...]

I: D.h., das allgemein didaktische, allgemein pädagogische, die Begrifflichkeiten des Konzepts fassen nicht die eigentlich wesentlichen fachlichen, fachdidaktischen Aspekte, die für wirklich fruchtbare Lernprozesse notwendig sind?

Deutsch: Genau. Und da muss man eben in dieser großen Kiste kleinere Kisten rausholen und sagen: Die steht jetzt auf dem Tisch.

Den Anspruch, den der Ausbilder hier formuliert, erscheint mehr als sinnvoll, aber gleichzeitig scheint es kaum möglich, dass ein fachübergreifend einsetzbares Bewertungskonzept es leisten kann, „diese kleineren Kisten“ mit den fachdidaktischen Bewertungskriterien systematisch zu erfassen. Das Konzept bot zwar mit dem Bereich der „Klarheit und Strukturiertheit“ bzw. der inhaltlichen Klarheit eine „große Kiste“, in die sich die fachdidaktischen Aspekte einordnen ließen, aber das Konzept kann auf diese Art und Weise seinen eigenen Ansprüchen nicht mehr gerecht werden. Die wesentlichen Kommunikationsprozesse, die als zentraler Mechanismus gelingender Bewertungsprozesse ausgemacht worden sind, können nicht mehr mit Hilfe der Struktur und des Vokabulars des Konzept gestaltet werden, sondern richten sich nach fachdidaktischen Aspekten. Sie spielen sich also in einem Resonanzbereich ab, auf den das Konzept nicht zugreifen kann. Sehr ähnliche Einwände formuliert der Sportausbilder, der ebenso Probleme sieht, die wesentlichen fachdidaktischen Fragen im Rahmen des Bewertungskonzepts zufriedenstellend zu thematisieren:

Sport: Ich habe ja dann versucht, das ein bisschen zu modifizieren auf den Sportunterricht. (...) Dieses klassische Merkmal hoher Anteil echter

Lernzeit, da gilt im Sportunterricht ja wirklich, dass das motorische Erleben im Mittelpunkt steht, dass die Bewegungszeit gegeben ist. [...] Die Schülerinnen und Schüler sollen ja durch eine Praxis, ja ein Problem reflektieren und die Lösung finden. Und das ist unheimlich schwierig, diese reflektierte Praxis zu inszenieren oder wie kann man mit Hilfe der Bewegung wirklich Wissen vermitteln? Und da haben die Referendare, aber da habe ich auch immer noch große Probleme wie das realisiert werden kann. Und das ist ja dann/Und das hat bestimmt dieser/Bestimmt 70% immer der Nachbesprechung in Zeit in Angriff genommen.

I: Also ein so sehr spezifisches, sehr spezifischer Aspekt, nenne ich es jetzt mal, des Sportunterrichts?

Sport: Ja genau.

I: Der dann auch dazu führt, dass dieser Qualitätsbereich eigentlich immer eine Rolle spielt, weil der in der Professionalisierung der Sportlehrer eine sehr, sehr große Baustelle ist.

Sport: Immer, ja also. [...] Und dann die inhaltliche Klarheit auch, der didaktische Schwerpunkt, da haben die natürlich auch unheimlich große Probleme. Wie reduziere ich jetzt, das ich wirklich mit Hilfe der Praxis auch ein theoretisches Wissen vermitteln kann und so? Das ist auch unheimlich schwierig und das ist ja/Auch der thematische Gang, finde ich ganz, ganz wichtig, aber das ist, das sind inhaltliche Klarheit und eigentlich die Bewegungslernzeit, bzw. der hohe Anteil echter Lernzeit, das ist ganz wichtig und das wurde primär thematisiert, weil viele dann eben auch irgendwie keine optimale Passung hatten und nicht reduziert haben und dementsprechend vielleicht dreißig Minuten die Schülerinnen und Schüler sich nicht bewegt haben.

I: Und das hat wozu geführt, diese Fokussierung auf diese drei sehr zentralen Qualitätsbereiche?

Sport: Dass es auf einem, für die Probanden glaube ich schwer war, weil ich hatte ja damals, ich glaube auch auf Basis Ihrer Vorlagen die didaktische Schwerpunktsetzung in die inhaltliche Klarheit gesetzt. Und

das war schwierig für die zu verstehen. Warum ist das das Qualitätsmerkmal inhaltlicher Klarheit? Die haben immer unter inhaltlicher Klarheit/Ok, ich habe die Transparenz, die Transparenz muss für die Schülerinnen und Schüler gegeben sein, das ist die inhaltliche Klarheit.

I: D.h., ich verstehe das jetzt so: Die fachlichen, fachdidaktischen Qualitätsbereiche eines guten Sportunterrichts lassen sich eben nicht einfach hier so ohne Weiteres in diese Qualitätsbereiche unterordnen?

Sport: Genau, genau.

I: Und es gab viele Reibungsverluste bei der Übersetzung der fachdidaktischen Konzepte in dieses primär allgemeindidaktische Konzept?

Sport: Ja, genau, genau.

Auch im Fach Sport wurden also die zentralen fachdidaktischen Fragen zwar einem Qualitätsbereich eingeordnet, aber dies führte nicht zu Klärungsprozessen, sondern, wie der Ausbilder berichtet, eher zu Verständnisschwierigkeiten auf Seite der Auszubildenden. Entsprechend scheint das Bewertungskonzept hier keine Hilfestellung für die Bewertung von Unterricht geboten, sondern eher für Verwirrung gesorgt zu haben. Eine Forderung an eine wirklich hilfreiche Beta-Version des Bewertungskonzepts ist es entsprechend, die fachliche Passung bzw. Offenheit für fachdidaktische Bewertungskategorien zu erhöhen.

Beide Fachausbilder sehen aus der fachdidaktischen Perspektive ein Problem darin, Kompetenzfortschritte der Auszubildenden über die vorgeschlagene Skalierung zu messen, da die Realisierung einzelner Qualitätsbereiche sich je nach Unterrichtsgegenstand unterscheidet und man eben sprichwörtlich Äpfel mit Birnen vergleicht, wenn Unterrichtsbesuche mit verschiedenen Unterrichtsgegenständen auf der gleichen Skala bewertet werden sollen:

Sport: „[...] Jetzt ist es aber auch gerade im Sport so, dass so viele unterschiedliche Stunden oder die Inhalte der Unterrichtsstunden waren immer anders und da kam manchmal, was weiß ich, die inhaltliche

Klarheit kann bei motorischem Lernen gegeben sein, aber dann bei Anwendung eines Spielvermittlungskonzepts dann eben nicht mehr und da kann man jetzt nicht so sagen, man muss das so machen. Es gibt im Sport ganz, ganz viele Möglichkeiten. Und deshalb hatte ich da auch manchmal Schwierigkeiten, habe ich auch mitgenommen.“

Der Ausbilder in Deutsch geht bei dieser Kritik sogar noch weiter und formuliert, dass die Skalierung wesentliche Aspekte der Qualitätsbereiche gar nicht erfassen könne, denn die Realisierung der Qualitätsbereiche sei bei unterschiedlichen Unterrichtsgegenständen unterschiedlich komplex. Folglich führten Erfassungen mit derselben Skala zu missverständlichen Aussagen:

I: Wir hatten ja für die Qualitätsbereiche und auch für die Reflexionskompetenz diese Skalen. Konnten Sie da für sich und auch für die Referendare schlüssig Unterscheidungen vornehmen zwischen Ausbildungsstand und Examensstand?

Deutsch: Theoretisch ja, aber praktisch fand ich das dann nicht so hilfreich wie ich vorher gedacht habe.

I: Warum, können Sie das erörtern?

Deutsch: Ja das hat genau mit dieser Frage der komplexen Systeme zu tun. Also wenn ein Referendar am, erst mal hat das mit dem weiten Feld der inhaltlichen Klarheit zu tun, dass man da eben so unterschiedliche Bereiche herausnehmen kann. Und wenn ein Referendar am Anfang ein Märchen liest, dann ist das eben eine einfach klare Sache, dazu gibt es 27.000 gute veröffentlichte Beispiele und kein Referendar kann da jetzt irgendwas glanzvoll glorioses Neues machen. Dann macht ein Referendar ein Poetry Slam, dazu gibt es so gut wie noch keine Literatur dazu und erprobt sich in einem ganz neuen Bereich. Dann ist das vielleicht auf der inhaltlichen Klarheit, ist das Märchen ganz oben und der Poetry Slam ist ganz unten, trotzdem hat der Referendar für sich und seine Ausbildungssteuerung, dass er gesagt hat: Einmal in meiner Ausbildung probiere ich mal was Neues, Konzeptionelles, mache selber was, hat der trotzdem unglaublich

viel gelernt, aber wenn ich so einen fairen Strich machen sollte, müsste ich vielleicht sagen: Also inhaltliche Klarheit ihr Lieben, die Schüler haben doch überhaupt nicht mitgekriegt, wo es lang ging oder was. Trotzdem war das eine spannende Ausbildungssituation.

I: Und für die Professionalisierung des jeweiligen sehr wichtig, wenn ich das richtig verstehe.

Deutsch: Ja, genau!

I: Aber formalisiert nach diesem Konzept bewertet wäre es schlecht?

Deutsch: [...] Ja, man muss auch Unterrichtsentwicklung in irgendeiner Form mal betreiben und wenn man was, wir hatten gerade mit dem Praxissemester zu tun, da ging es auch um Scheitern von forschendem Lernen bei den Studierenden. Wenn man sein Scheitern gut reflektiert, dann ist das eine sehr viel höhere Leistung, als wenn man ein konventionelles Unterrichtsdesign, was tausend mal durchgeführt ist, zum tausend und ersten mal durchführt.

Aus den Ausführungen des Ausbilders geht hervor, dass aus seiner Perspektive die Skalierung wesentliche Professionalisierungsschritte, die ein Auszubildender leisten müsse, nicht erfassen kann, denn „ein gut reflektiertes Scheitern“ sei eine sehr viel höhere Leistung, als wenn man ein „konventionelles Unterrichtsdesign zum tausend und ersten Mal durchführt“.

Zusammenfassend kann das Fazit gezogen werden, dass es sehr schwierig sein wird, das Bewertungskonzept in der Beta-Version so fortzuentwickeln, dass es einerseits fachübergreifend eingesetzt werden kann und ein funktionales Maß an Standardisierung der Bewertungsprozesse anbietet und andererseits genug Offenheit für fachdidaktische Bewertungskriterien bildet. Die Tragweite dieses Konflikts ist auch dem Deutschausbilder erst während der Konzepterprobung klar geworden:

I: D.h., ich versuche da jetzt zwei Sachen rauszuziehen. Die erste Frage, die ich habe, das klingt für mich ja so ein bisschen so, als wenn der Anspruch,

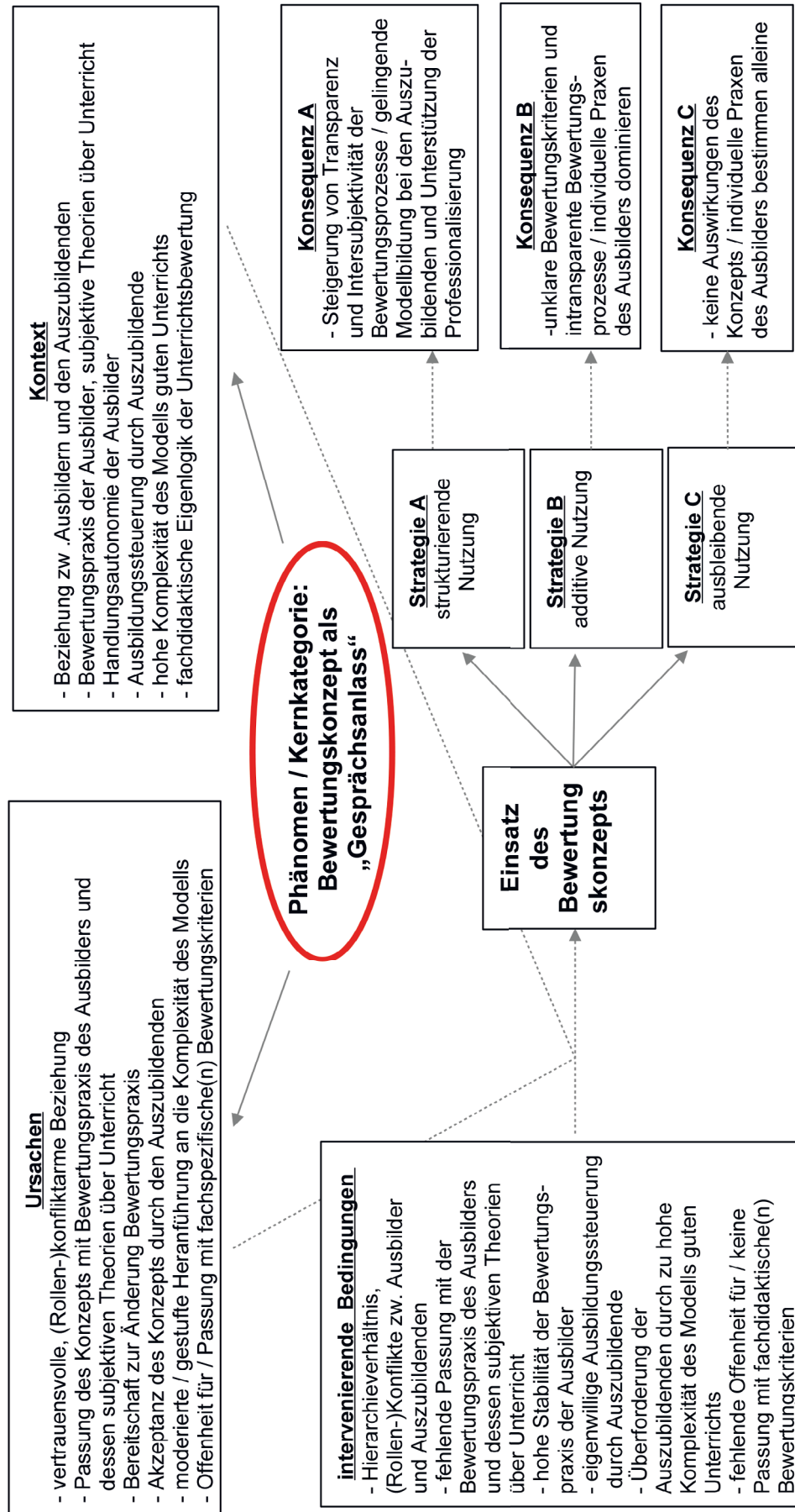
den ich versuche mit dem Bewertungskonzept umzusetzen, wirklich ein Bewertungsinstrument zu entwickeln, was für alle Fächer funktioniert, schwierig wird.

Deutsch: Ja. Sehe ich jetzt so, habe ich vorher überhaupt nicht so gesehen.

5.4 Abschließende Interpretation der Interviews – axiales Codieren

Bevor in dem abschließenden Unterkapitel 5.5 das Bewertungskonzept zur Beta-Version fortentwickelt wird, sollen die wesentlichen Ergebnisse der Interviews zusammenfassend betrachtet werden. Dies geschieht wie in der abschließenden Interpretation der Gruppendiskussionen in Form des axialen Codierens, in dessen Rahmen die Ursache-Wirkungszusammenhänge eines erfolgreichen Einsatzes des Bewertungskonzepts in ihrer Systematik dargestellt werden. Die *Kernkategorie* eines erfolgreichen Konzepteinsatzes, in dessen Kontext es zu einer positiven Veränderung der Bewertungsprozesse kommt, ist die Frage, ob das Bewertungskonzept als „*Gesprächsanlass*“ dient. Gelingt dies, generiert das Bewertungskonzept Impulse, durch die Ausbilder und Auszubildende in einen Klärungsprozess eintreten. Sie entwickeln gegenseitiges Verständnis für die subjektiven Wertungen bezüglich des gesehenen Unterrichts und können die Qualitätseinschätzungen des jeweils anderen begründend nachvollziehen. Kurzum: Ausbilder und Auszubildende *kommen ins Gespräch* über ihre verschiedenen Sichtweisen auf den konkreten Unterricht, wie es die Kernkategorie des axialen Codierens der Gruppendiskussionen formulierte. So wird es den Auszubildenden ermöglicht, berufliche Handlungskompetenz im Bereich des Unterrichtens – eben Unterrichtskompetenz – zu entwickeln. Sie verstehen, warum ihr Ausbilder einzelne ihrer Entscheidungen als sinnvoll erachtet und andere als weniger geeignet guten Unterricht anzuleiten. Sie können die so gewonnene Expertise in anderen Kontexten anwenden. Abbildung 5.19 zeigt die Kontextbedingungen in ihren systemischen Zusammenhängen, die gegeben sein müssen, damit das Bewertungskonzept ‚*Gesprächsanlässe*‘ generieren kann.

Abbildung 5.19: Axiales Codieren der Interviews



Im Zentrum des axialen Codierens steht die *Kernkategorie*, die Frage, ob das Bewertungskonzept *Gesprächsanlässe* generieren kann, wobei die Kernkategorie auch in diesem Fall ein In-vivo-Code ist. Der im oberen rechten Kasten dargestellte *Kontext* beschreibt die abstrakteren sozialen Rahmenbedingungen, in denen sich durch den Einsatz des Bewertungskonzepts *potentiell* Gesprächsanlässe generieren lassen. Auch hier spielen die *Beziehungen zwischen den Ausbildern und den Auszubildenden* eine große Rolle, ebenso wie die *Bewertungspraxen der Ausbilder und ihre subjektiven Theorien über die Eigenschaften guten Unterrichts*. Die faktisch existente *große Handlungsautonomie der Ausbilder* stellt eine wesentliche Rahmenbedingung dar, in der ein Bewertungskonzept seine Wirkung entfalten muss. Allerdings beeinflussen die Auszubildenden durch ihre *individuelle Ausbildungssteuerung* den erfolgreichen Einsatz eines Konzepts ebenso. Die *hohe Komplexität der Modellierung guten Unterrichts* fordert einen bewussten Umgang mit dieser Komplexität, damit sie nicht überfordernd wirkt. Schlussendlich hat die Erprobung gezeigt, dass die *fachdidaktische Eigenlogik der Unterrichtsbewertung* massiven Einfluss auf den Einsatz eines fachübergreifend eingesetzten Bewertungskonzepts hat.

Für den erfolgreichen Einsatz eines Bewertungskonzepts muss der *abstrakte Kontext* Ausdruck in *konkreten Ursachen* finden. Hierbei ist zunächst eine *möglichst vertrauensvolle Beziehung zwischen dem Auszubildenden und dem Ausbilder* zu nennen, in der es möglichst wenige (Rollen-)Konflikte zwischen den beiden Akteuren gibt. Außerdem ist ein Mindestmaß von *Übereinstimmung zwischen der bestehenden Bewertungspraxis des Ausbilders* und der Logik, mit der das Bewertungskonzept Unterricht bewerten möchte, erforderlich. Damit das Konzept seine Wirksamkeit entfalten kann, muss der Ausbilder ebenfalls dazu *bereit sein, seine gewachsenen Bewertungspraxen zu ändern*. Hierbei muss das Maß der Änderungsbereitschaft um so größer sein, je stärker die bisher praktizierte Praxis von der durch das Konzept intendierten abweicht. Der Auszubildende muss das Bewertungskonzept bzw. das Verständnis guten Unterrichts, wie es das Konzept vertritt, als *Steuerungselement für seine Professionalisierung akzeptieren* und anwenden. Um einer Überforderung vorzubeugen kommt dem Ausbilder die Rolle zu,

durch seine professionelle Expertise eine *schrittweise Heranführung an die Komplexität des Modells* guten Unterricht anzubieten. Weiterhin muss das Konzept eine *Offenheit für die konkreten fachdidaktischen Bewertungskriterien* aufweisen. Diese müssen so in das Bewertungskonzept integriert werden können, dass eine fruchtbare fachdidaktische Ausbildung gelingen kann, ohne dass diese die Bewertungsprozesse so dominiert, dass die Struktur des eingesetzten, fachübergreifenden Bewertungskonzepts durch eine fachdidaktische Strukturierung abgelöst wird.

Prägen die *intervenierenden Bedingungen* den Kontext, dann können die Ursachen eines erfolgreichen Konzepteinsatzes keine Wirkung entfalten, das Bewertungskonzept wird kaum Gesprächsanlässe liefern können. Hierbei entfaltet das (unvermeidliche!) *Hierarchieverhältniss zwischen dem Ausbilder und dem Auszubildenden* wiederum seine unheilvolle Wirkung dahingehend, dass Auszubildende sich mit ihren tatsächlichen Wertungen über Unterricht zurückhalten und insofern nur oberflächliche Interaktionen zwischen Ausbilder und Auszubildendem zustande kommen. *Rollenkonflikte*, so zeigen zwei Beispiele aus den Interviews, können dafür sorgen, dass die Ausbilder von den Auszubildenden in ihrer Rolle als Experte nicht ernst genommen werden, wodurch auch die Wirkungen des Konzepteinsatzes betroffen sind. Die *Logik, mit der das Bewertungskonzept Unterricht bewerten möchte*, kann der tradierten Logik eines Ausbilders diametral entgegen stehen, sodass dies einen maßgeblichen Hinderungsgrund für den erfolgreichen Konzepteinsatz darstellt. Gleiches gilt, wenn das vom Konzept aufgestellte *Modell guten Unterrichts den subjektiven Theorien* eines Ausbilders, über das was guten Unterricht ausmacht, *zuwider läuft*. Gepaart mit der bei allen Ausbildern beobachteten *hohen Stabilität der Bewertungspraxis* bilden sich Systemzusammenhänge, die einen erfolgreichen Einsatz des Konzepts verhindern können. Ebenso muss der Auszubildende sowohl das Modell guten Unterrichts als auch die Kompetenzmodellierung des Bewertungskonzepts zumindest so weit anerkennen, dass er bereit ist, sich in seiner Ausbildungssteuerung auf beide einzulassen und die *Steuerung nicht primär eigenwillig* durch vom Konzept abweichende Logiken bestimmt. Ist ein Auszubildender von der Komplexität des Bewertungskonzepts und hierbei insb. *von dem Modell guten Unterrichts überfor-*

dert, sind die Bedingungen für einen erfolgreichen Konzepteinsatz ebenfalls nicht gegeben. Abschließend kann festgehalten werden, dass eine *fehlende Offenheit des Bewertungskonzepts für fachdidaktische Bewertungskategorien* zu einem Scheitern des Konzepteinsatzes führen kann.

Aus dem Zusammenwirken des abstrakten *Kontexts* und den konkreten *Ursachen* eines erfolgreichen Konzepteinsatzes bzw. den *intervenierenden Bedingungen*, die einen solchen Erfolg verhindern, lassen sich aus der Auswertung der Interviews drei idealtypische Strategien der Akteure ableiten, mit denen diese auf die Systemzusammenhänge reagieren. Hierbei ist darauf zu verweisen, dass die Ausbilder zwar die im Interaktionssystem dominierenden Akteure sind und demnach deren realisierte Handlungsstrategie entscheidenden Einfluss auf den Systemcharakter hat, aber die Handlungsstrategien der Auszubildenden ebenso Einfluss auf die Realisierung einer bestimmten Systemausprägung nehmen.

- Dominieren die Ursachen eines erfolgreichen Konzepteinsatzes einen konkreten Ausbildungskontext, kommt es zu einer *strukturierenden Nutzung des Konzepts (Strategie A)*, durch die Gesprächsanlässe generiert werden können. In der *Konsequenz* entsteht eine Steigerung der Transparenz und der intersubjektiven Nachvollziehbarkeit der Bewertungsprozesse. Dies führt auf Seiten der Auszubildenden wiederum dazu, dass diese ein für sich stimmiges Modell guten Unterrichts erhalten, welches sie in ihrer Professionalisierung unterstützt.
- Eine partielle Abweichung von den Ursachen, bei gleichzeitig stärkerem Einfluss der intervenierenden Bedingungen, führt zur *Strategie B*, die sich durch eine *additive Nutzung* des Konzepts auszeichnet und als *Konsequenz unklare Bewertungskriterien und intransparente Bewertungsprozesse* nach sich zieht. Auch wenn es dem Bewertungskonzept gelingt Gesprächsanlässe zu generieren, so stoßen diese sich typischerweise an individuellen und dominierenden Logiken des Ausbilders. Folglich erscheint es für die Auszubildenden unklar, welche Bewertungskriterien die entscheidenden sind, sodass das Konzept keinen produktiven Beitrag für deren Professionalisierungsprozess leisten kann.

- *Strategie C* entspringt einer Dominanz der intervenierenden Bedingungen und drückt sich in einer größtenteils *ausbleibenden Nutzung* des Bewertungskonzepts aus. Das Konzept zeigt in der *Konsequenz* keine nennenswerten Auswirkungen, da die individuellen Bewertungspraxen des Ausbilders die tatsächlich einzig relevanten Aspekte der Bewertung sind und die Professionalisierung der Auszubildenden strukturieren.

Bei jeder dieser Strategien gilt, dass ein Bündel verschiedener Faktoren dazu führt, dass die *Ursachen* einer erfolgreichen Umsetzung des Bewertungskonzepts oder eben die *intervenierenden Bedingungen* eine jeweils größere Wirksamkeit entfalten. Im Fall von A5 scheinen die Konflikte zwischen ihm und den Auszubildenden maßgeblichen Einfluss auf die Wahl der Strategie gehabt zu haben, während bei A6 offensichtlich die fehlende Passung mit der subjektiven Theorie, durch welche Bewertungsprozesse guter Unterricht überhaupt erfasst werden kann, ein maßgebliches Hindernis für die Realisierung eines strukturierenden Einsatzes war.

Die folgende Überarbeitung des Bewertungskonzepts zur Beta-Version greift diese Zusammenhänge auf und diskutiert, wie das Konzept überarbeitet werden kann, damit die Generierung von ‚*Gesprächsanlässen*‘ wahrscheinlicher wird.

5.5 Von der Alpha- zur Beta-Version

Aus der Auswertung der Interviews lassen sich weniger klare Hinweise zur Überarbeitung der Alpha-Version ziehen, als zuvor aus den Gruppendiskussionen für die Überarbeitung der pre-Alpha-Version. Das axiale Codieren der Interviews hat gezeigt, dass die von den Akteuren gewählten Strategien zum Einsatz des Bewertungskonzept maßgeblichen Einfluss auf die Wirkungen von diesem haben. Dieser Einsatz hängt seinerseits von einem Bündel von Faktoren ab, auf die das Design des Bewertungskonzepts keinen Einfluss nehmen kann. Widerspricht die Logik, mit der das Bewertungskonzept den Unterricht bewerten will, der tradierten Bewertungspraxis eines Ausbilders und seinen subjektiven Theorien, wie

sich Unterrichtsqualität erfassen lässt, dann wird ein Bewertungskonzept kaum produktive Impulse für die Gestaltung der Bewertungsprozesse geben können, wie das Beispiel von A6 zeigte. Eine *Änderung des Konzeptdesigns* wird auf diese Probleme aller Voraussicht nach *keine adäquate Reaktion* darstellen, da mit den Änderungen des Konzepts im Sinne eines Ausbilders die Passung mit den Bewertungspraxen und den subjektiven Theorien anderer Ausbilder wahrscheinlich abnimmt. Hierbei sei z.B. an die nicht zu klärende Diskussion um eine mögliche Hierarchie innerhalb der von der Matrix dargestellten Merkmale guten Unterrichts erinnert oder an die Frage, ob Unterrichtsqualität über die Erfassung und Beurteilung isolierter Qualitätsbereiche bewertet werden kann. Ein standardisiertes Bewertungskonzept wird diese unterschiedlichen Sichtweisen nicht vereinen oder auch nur zufriedenstellend versöhnen können. Gleichzeitig ist davon auszugehen, dass in jeder Ausbildungsinstitution der zweiten Phase der Lehrerbildung Akteure, die sowohl der einen als auch der anderen Perspektive anhängen, zu finden sein werden. Folglich muss noch nicht einmal die problematische Rezeption des Bewertungskonzept aus der Perspektive fachdidaktischer Überlegungen in den Blick genommen werden, um dem zentralen Anliegen dieser Dissertation, nämlich ein standardisiertes Bewertungskonzept zu entwickeln, das innerhalb einer gesamten Ausbildungsinstitution eingesetzt werden kann, eine Absage zu erteilen. Damit ist nicht gesagt, dass das Dissertationsprojekt als solches gescheitert ist. Wie im abschließenden Fazit der Arbeit gezeigt wird, lassen sich aus der Konzeptentwicklung und -erprobung wichtige Erkenntnisse über die Rahmenbedingungen gewinnen, denen sich produktive Bewertungsprozesse in der zweiten Phase der Lehrerbildung stellen müssen. Aber auf diese strukturellen Bedingungen und die daraus resultierenden Herausforderungen und Fragen können Überlegungen zum *Design* eines Bewertungskonzepts eben nur sehr eingeschränkt eine Antwort geben. Gleichwohl sollen in diesem Unterkapitel generalisierbare Erkenntnisse hinsichtlich der Frage des Konzeptdesigns, die sich aus der Konzepterprobung gewinnen lassen, dargestellt werden und – sofern dies möglich ist – für eine Überarbeitung der Alpha-Version des Bewertungskonzepts genutzt werden.

Die *Matrix der Qualitätsbereiche* bot für einen Großteil der Auszubildenden eine Hilfestellung für die erfolgreiche ‚*Modellbildung*‘ guten Unterrichts und unterstützte sie so in ihrer Professionalisierung. Auffällig war hierbei, dass mehrere Auszubildende in ihren Berichten darüber, wie die Matrix ihre Professionalisierungsprozesse unterstützte, jenen Prozess beschrieben, der von den Ausbildern in der Gruppendiskussion als zentrale Zielsetzung der Ausbildung angesehen wurde: Die Entwicklung der Fähigkeit, über Unterricht und dessen Planung mit Hilfe von Fachvokabular zu sprechen und eigene didaktische Entscheidungen kategorial begründen zu können. Dies deckt sich stark mit dem hier vertretenen *Kompetenzzusammenhang als Expertise*, dessen Ausbildung das Konzept unterstützen sollte. Gleichzeitig scheint die Komplexität guten Unterrichts, wie sie die Matrix darstellt, v.a. zu Beginn der Ausbildung eine Überforderung für einige Auszubildenden darzustellen. Diese Komplexität sorgte ebenfalls dafür, dass die Matrix die Vorbereitung der Unterrichtsnachbesprechung größtenteils nicht unterstützen konnte und mehrere Auszubildende auf Grund dieser Überforderung in den konkreten Ausbildungsphasen und -situationen auf die weniger komplexen Instrumente des ZfsL ausgewichen sind. Entsprechend wird das Bewertungskonzept der Beta-Version erweitert um eine *reduzierte Version der Matrix*, die auf die Auflistung von Indikatoren für die Beobachtung der Einzelmerkmale der Qualitätsbereiche verzichtet, und eine *Übersicht über die Qualitätsbereiche guten Unterrichts in Form einer Mindmap*. In dieser wird auf jede nähere Erläuterung der Qualitätsbereiche verzichtet, um den Umfang des Instruments möglichst gering zu halten. Der Umfang des Bewertungskonzepts nimmt dadurch jedoch zunächst einmal zu. Der Spielraum der Ausbilder und Auszubildenden weitet sich jedoch dahingehend aus, dass sie *Instrumente verschiedener Komplexitätsgrade* zur Verfügung haben, aus denen sie je nach den konkreten Anforderungen in den verschiedenen Ausbildungsphasen und -situationen auswählen können.

Die Operationalisierung guten Unterrichts und mit ihr die Matrix der Qualitätsbereiche wurden im praktischen Einsatz mit der Problematik konfrontiert, dass die wesentlichen Bewertungsprozesse in den konkreten Ausbildungskontexten oft von *fachdidaktischen Überlegungen beeinflusst wurden*, die so nicht in die Operationalisierung eingeordnet werden konnten. So lautet zumindest das Fazit

der Ausbilder in den Fächern Sport und Deutsch. Entsprechend müsste die Matrix als Instrument eine größere Offenheit für fachdidaktische Qualitätskriterien aufweisen, als sie es bisher tut. Damit wird eine Forderung, die schon aus den Gruppendiskussionen rekonstruiert wurde, hier ein zweites Mal gestellt. Die Erfahrungen der eben genannten Ausbilder deuten allerdings darauf hin, dass eine solche Passung zwischen der notwendigerweise allgemeindidaktisch ausgerichteten Matrix der Qualitätsbereiche und den fachdidaktischen Fokussierungen in den Unterrichtsnachbesprechungen nur sehr schwer zu realisieren ist. Dominiert eine Qualitätsauffassung wie jene der Matrix die Nachbesprechung, werden sich fachdidaktische Kriterien additiv in die Besprechungslogik einordnen müssen. Dies wird vermutlich die bekannten Folgen eines solchen Einsatzmuster für die Klarheit der Kompetenzdiagnostik und die Transparenz der Bewertungsprozesse haben. Das gilt ebenso andersherum, wenn die fachdidaktische Bewertung die allgemeindidaktische Perspektivierung dominiert und die zweite additiv an die Bewertung angefügt wird. Es entsteht demnach ein Dilemma der Bewertung, das sich wahrscheinlich nicht auflösen lässt. Auch der Vorschlag von Hilbert Meyer, ein fachdidaktisches Kriterium im Sinne eines „Jokers“ als weiteren Qualitätsbereich in das Bewertungskonzept einzubeziehen (vgl. Meyer 2010b: 26) wird die Problematik nicht lösen, da es um die *Frage eines dominanten Bewertungsmusters* in den Unterrichtsnachbesprechungen geht und dieses *nur allgemein- oder fachdidaktisch*, mit den jeweils spezifischen Vor- und Nachteilen sein kann. Daraus lässt sich schlussfolgern, dass die für produktive Bewertungsprozesse *notwendige fachdidaktische Perspektivierung ein strukturelles Hindernis* für die Entwicklung fachübergreifend nutzbarer Bewertungsinstrumente darstellt, welche für mehr Transparenz in den Bewertungsprozessen sorgen und für die Auszubildenden eine hilfreiche Modellbildung guten (Fach-)Unterrichts anbieten sollen.

Der *Dokumentationsbogen für die Nachbesprechung* rief in den Interviews ein geteiltes Echo hervor. Auf der einen Seite lobten mehrere Auszubildende und ein Ausbilder seine Leistung ‚*Transparenz*‘ herzustellen, indem er zu einer klar strukturierten Unterrichtsnachbesprechung sowie zu einer verbindlichen und verlässlichen schriftlichen Dokumentation des Besprochenen anleitete. Aus den

Interviews ging hervor, dass mehrere Auszubildende durch die Art der Dokumentation größere Klarheit bezüglich ihrer erreichten Kompetenzfortschritte, aber auch über ihre Entwicklungsaufgaben erhielten. Die für die Beurteilung der Qualitätsbereiche und der Reflexionskompetenz genutzte ‚Skalierung‘ rief ebenfalls größtenteils ein positives Echo hervor, sorgte sie doch für eine Bilanzierung einzelner Teilkompetenzen „schwarz auf weiß“, wie es mehrere Auszubildende formulierten. Zeitgleich zeigte sich in der Erprobung aber auch, dass ein Dokumentationsbogen nur dann gewinnbringend eingesetzt werden kann, wenn die von ihm unterstellte Logik, wie Unterricht bewertet werden kann, mit den subjektiven Theorien des Ausbilders und der Auszubildenden ein Mindestmaß an Übereinstimmung aufweist. Die Erprobung des Bewertungskonzepts im Seminar von A6 weist darauf hin, dass es sinnvoller sein kann, wenn ein Ausbilder einen Dokumentationsbogen nutzt, der seiner eigenen Logik, Unterricht zu bewerten entspricht, auch wenn dieser nicht seminarübergreifend eingesetzt wird. Ein additiver Einsatz sorgt nicht für eine Klärung erreichter Kompetenzfortschritte oder eine Steigerung der Transparenz von Bewertungsprozessen, sondern erzeugt neue Missverständnisse in den Unterrichtsnachbesprechungen. Insofern ist es fragwürdig, ob es Sinn macht, einen Dokumentationsbogen zu entwerfen, der seminarübergreifend eingesetzt werden soll. Der Zweck dieses Instruments, zu einer verbindlichen und transparenten Bilanzierung des erreichten Kompetenzfortschritts zu kommen, wird durch ein standardisiertes Instrument auf einer institutionellen Ebene wahrscheinlich nicht zu erreichen sein. Ein standardisierter Dokumentationsbogen, der fachübergreifend eingesetzt werden kann, wird hingegen nichts anderes als Banalitäten über den Unterricht erfassen. Dies zeigen auch die Erfahrungen der Auszubildenden, wie sie in der Gruppendiskussion schildern. Vielmehr erscheint es deshalb sinnvoll zu sein, alle Ausbilder dazu zu verpflichten, eine schriftliche Dokumentation anzufertigen, ohne deren Design vorzuschreiben, denn dass eine schriftliche Dokumentation die Transparenz von Bewertungsprozessen erhöht und dafür sorgt, dass die Ausbilder ihre Bewertung möglichst klar begründen, dies scheint Konsens unter den Teilnehmern der Erprobung zu sein. Dabei haben die Interviews auch gezeigt, dass eine schriftliche Bilanzierung, die der Bewertungslogik des Ausbilders entspricht, die Transparenz

der Bewertungsprozesse deutlich positiver beeinflusst als ein fachübergreifend, aber eben additiv eingesetzter Dokumentationsbogen. Folglich wird der Dokumentationsbogen in der Beta-Version des Bewertungskonzepts keiner Veränderung unterzogen, sondern verbleibt als *Vorschlag für individuelle Entwürfe von Dokumentationsbögen*.

In dem Kontext der nur im Ansatz gelungenen Erprobung des Dokumentationsbogens ist auch auf die *Gütekriterien* einzugehen, mit deren Hilfe die Qualität des Erfassungsprozesses von Unterrichtskompetenz sichergestellt werden sollte. Die Auswertungsinterviews deuten auch hier wieder darauf hin, dass die Instrumente nur dann einen Erfassungsprozess, der den in Kapitel zwei formulierten Gütekriterien gerecht wird, gewährleisten können, wenn das Bewertungskonzept keine zu starke Abweichung von der tradierten Form der Unterrichtsbewertung des Ausbilder aufweist, und von den Auszubildenden als Leitlinie für die eigene Professionalisierung anerkannt wird. Eine *kommunikative Validierung* des Urteils kann nur dann gelingen, wenn die vom Konzept genutzte Bewertungslogik vom Ausbilder akzeptiert wird. Ansonsten, auch hier sei wiederum auf die Erprobung bei A5 und A6 verwiesen, entstehen aus der Nutzung der Instrumente eher Missverständnisse als klare Bewertungen. Ebenso wenig kann der Einsatz der Instrumente in einem solchen Setting *stimmige Urteile* ermöglichen. Das Kriterium der *Stimmigkeit* setzt voraus, dass eine Bewertung u.a. keine inneren Widersprüche aufweist. Dies wird allein schon dadurch konterkariert, dass ein Ausbilder ein Bewertungsinstrument nutzen soll, welches seiner eigenen Auffassung widerspricht wie guter Unterricht in seinem Wesenskern erfasst werden kann. Entsprechend *stimmt* sein eigenes Urteil hier aller Voraussicht nach nicht mit jenem des Bewertungskonzepts überein. Die Folgen einer solchen *Unstimmigkeit* zeigen z.B. die Missverständnisse und Frustrationen, von denen A6 und R11 berichten, als die Bewertung eines Unterrichtsbesuchs mit dem Bewertungskonzept zu diametral unterschiedlichen Einschätzungen auf beiden Seiten führte. Entsprechend lässt sich auch aus diesen Überlegungen ein Plädoyer für individualisierte Dokumentationsbögen ableiten.

Die *Sammlung über das Repertoire* wurde im Rahmen der Erprobung offensichtlich von keinem Ausbilder entsprechend der Intention eingesetzt. Einzelne positive Einschätzungen der Auszubildenden beruhen i.W. auf Mutmaßungen, dass das Instrument eine transparente Bilanzierung ihrer Kompetenzfortschritte anbieten könnte. Tatsächlich realisiert werden konnte dieser Anspruch bei keinem der Auszubildenden. Inwieweit ein solches Instrument in der Praxis eingesetzt werden kann, dessen Zweck sich v.a. dann erfüllt, wenn alle Ausbilder eines Referendars sich auf die Bewertungslogik des Instruments einlassen und bereit sind, ihre Bewertungen gemeinsam in dieses einzutragen, muss vor dem Hintergrund dieser problematischen Erprobungsphase als sehr fragwürdig angesehen werden. Wie oben ausgeführt wurde, bot die Konzepterprobung günstige Rahmenbedingungen für die Veränderungen der Bewertungsprozesse¹¹³. Wenn nun die Sammlung über das Repertoire selbst unter diesen günstigen Erprobungsbedingungen nicht von einzelnen Ausbildern eingesetzt wurde, dann ist der Einsatz eines solchen Instruments in der kooperativen Anwendung von mehreren Ausbildern unter Praxisbedingungen höchst unwahrscheinlich. Zumal die Nutzung eines solchen Instruments nur dann funktionieren kann, wenn sich *alle* Ausbilder einer Institution auf den Einsatz einheitlicher Bewertungsinstrumente und zu Grunde liegender Kategorien guten Unterrichts einigen können. Eine Voraussetzung, die an sich schon eine große Herausforderung darstellt, bevor an die Implementation gemeinsam genutzter Bewertungsinstrumente gedacht werden kann. Entsprechend verbleibt die Sammlung über das Repertoire als Zukunftsvision bestehen, die im Sinne nachhaltiger und klar bilanzierter Kompetenzentwicklungsprozesse theoretisch ihre Berechtigung hat, die in die Praxis umzusetzen aber als relativ unwahrscheinlich angesehen werden muss.

Fasst man diese Ausführungen zur Überarbeitung des Bewertungskonzepts zur Beta-Version zusammen, bleibt letzten Endes kein in sich geschlossenes Bewertungskonzept bestehen. Vielmehr zeigt sich die *Beta-Version als ein Angebot von Impulsen für die Gestaltung von Bewertungsprozessen*, die Hinweise für die Bedingungen produktiver Bewertung im Sinne formativer Evaluation in der zwei-

¹¹³ Siehe Kapitel 3.3.

ten Phase der Lehrerbildung geben können. Diese erhalten ihre Berechtigung als Impulsgeber für die Weiterentwicklung der Bewertungspraxis daraus, dass sie das Feedback von Stakeholdern aus der konkreten Ausbildungspraxis heraus berücksichtigen. Die Beta-Version des Bewertungskonzept kann in diesem Sinne auch *nicht mehr als standardisiertes Bewertungskonzept* verstanden werden, da der individuellen Bewertungspraxis der Ausbilder ein größerer Stellenwert eingeräumt wird als der Standardisierung von Bewertungskategorien und Erfassungsprozessen. Zusammenfassend können aus der Fortentwicklung des Konzepts zur Beta-Version, *Standards guter Unterrichtsbewertung in der zweiten Phase der Lehrerbildung* abgeleitet werden. Aus der Erprobung lässt sich schlussfolgern, dass sich ‚gute Unterrichtsbewertung‘ durch folgende Aspekte auszeichnet:

- Eine schriftlich fixierte Qualitätsdefinition guten Unterrichts liegt vor.
 - Die Qualitätsdefinition klärt zentrale allgemein- und fachdidaktische Kriterien guten Fachunterrichts und operationalisiert sie durch Beispiele.
 - Die Auszubildenden nutzen die Qualitätsdefinition für die Vorbereitung von oder der Reflexion über Unterricht.
- Ein Dokumentationsbogen für die Nachbesprechung liegt vor.
 - Die Intention des Dokumentationsbogens ist geklärt (Hilfestellung für die Beratung oder Leitfaden für Bewertung)
 - Der Dokumentationsbogen gibt begründet an auf welche Art und Weise er Unterrichtsqualität bewertet (offen, standardisiert, mit Hilfe von Skalen, ggf. gar nicht (weil Beratungsinstrument), etc.).
- Die Instrumente werden eingesetzt.
 - Die ausgefüllten Instrumente werden den Auszubildenden im Anschluss an die Unterrichtsbesuche zeitnah ausgehändigt.

Die Beispielinstrumente und aus der Erprobung abgeleiteten Impulse zur Gestaltung der Bewertungsprozesse sind auf der Internetseite www.unterrichtskompetenz.de zusammengefasst, damit sie der Fachöffentlichkeit möglichst einfach zugänglich sind.

6 | Abschließendes Fazit – Was leistet ein standardisiertes Bewertungskonzept in der zweiten Phase der Lehrerbildung?

Die Einleitung sowie Kapitel eins formulierten zwei Ziele dieser Arbeit: *Erstens* sollte ein Bewertungskonzept für die zweite Phase der Lehrerbildung entwickelt werden, das auf die bekannten Problemfelder der Bewertungsprozesse reagiert und diese zumindest teilweise entschärft. *Zweitens* sollten aus der Erprobung des Bewertungskonzepts Erkenntnisse im Sinne gegenstandsbezogener Theoriebildung über die Rolle von Standardisierung für produktive Bewertungsprozesse in der zweiten Phase der Lehrerbildung generiert werden. Mit Blick auf diese beiden Ziele muss ein zwiespältiges Fazit gezogen werden: einerseits ist es nicht gelungen, ein in der Praxis einsetzbares Bewertungskonzept für die zweite Phase der Lehrerbildung zu entwickeln, aber andererseits konnten im Rahmen der Evaluation der Konzepterprobung ein erheblicher Teil jener Strukturen rekonstruiert werden, in denen dieser Misserfolg gründet. In diesem Sinne konnte das Ziel der Arbeit, begründete gegenstandsbezogene Thesen über die Rolle von Standardisierung für produktive Bewertungsprozesse in der zweiten Phase der Lehrerbildung zu generieren, erreicht werden. Die Arbeit leistet damit einen Beitrag zur Theoriebildung über die Bewertungsprozesse in der zweiten Phase der Lehrerbildung und bietet durch die Beta-Versionen der Bewertungsinstrumente, sowie die ‚Standards guter Unterrichtsbewertung in der zweiten Phase der Lehrerbildung‘ Impulse, zur Gestaltung von Bewertungsprozessen.

Die Evaluation des Konzepteinsatzes hat gezeigt, dass die *Beziehungen zwischen den Ausbildern und den Auszubildenden* die Bewertungsprozesse und auch den Einsatz von Bewertungsinstrumenten massiv prägen. Teilweise verhindern Konflikte zwischen den Akteuren den Einsatz von Instrumenten gemäß ihrer Intention (siehe Beispiel von A5). In anderen Fällen ruft die unterschiedliche Rezeption der Instrumente Konflikte hervor, die im Sinne der Rollenhierarchie nicht geklärt werden können. Aus diesen Ergebnissen lässt sich schlussfolgern, dass die

hierarchischen Beziehungen zwischen den Ausbildern und den Auszubildenden und die daraus resultierenden Machtverhältnisse *stabile Strukturbedingungen der Ausbildung* und der Bewertungsprozesse sind, mit denen sich ein Bewertungskonzept arrangieren muss. Folgt man den Erfahrungen der Konzepterprobung, so werden es Bewertungsinstrumente nicht leisten können, Konflikte über die Bewertung dadurch zu klären, dass eine schriftlich fixierte Bewertungsgrundlage vorliegt, wie es die formulierte Zielsetzung des Bewertungskonzepts war. Auf Grund des Abhängigkeitsverhältnisses fragte R13 „kein zweites Mal nach“, als die Instrumente nicht eingesetzt wurden, und R11 „hat am Anfang noch gegen die Bewertungspraxis revoltiert“, dieses Verhaltensmuster aber geändert, als sie merkte „wie sich das auf die Laune von A6 ausgewirkt hat“.

Neben den hierarchischen Beziehungen zwischen den Akteuren ist die *hohe Autonomie*, mit der die Ausbilder faktisch agieren, eine weitere zentrale Systembedingung in den Ausbildungskontexten der zweiten Phase der Lehrerbildung. Die Ausbilder entwickeln, basierend auf ihren subjektiven Theorien über die Eigenschaften guten Unterrichts sowie über die Möglichkeiten, seine Qualität zu erfassen, individuelle Bewertungspraxen. Absichten, die Bewertungsprozesse durch festgelegte Bewertungskategorien oder Skalierungen zu standardisieren, erzeugen teils paradoxe Ergebnisse, da die Ausbilder Instrumente nutzen sollen, die ihren subjektiven Theorien und fest verankerten Beliefs widersprechen. In der Konsequenz entstehen Bewertungspraktiken, die in sich unstimmig sind und neue Konflikte und Unklarheiten hervorrufen. Diese paradoxen und konfliktbehafteten Bewertungspraxen könnten im Sinne der Theorien institutionellen Wandels zunächst als Ausdruck des Widerstands gegen die Neuerung an sich verstanden werden, der sich durch ein systematisches Change Management, welches den möglichen Implementierungsprozess auf der Ebene einer Ausbildungsinstitution begleitet, auf Dauer reduzieren lässt. Allerdings sprechen zwei Aspekte gegen diese Perspektive: *Erstens* deutet die Erprobung darauf hin, dass die hochgradige Autonomie des Ausbilderhandelns die erfolgreiche Implementierung von fächerübergreifend eingesetzten, standardisierten Bewertungskonzepten torpedieren wird. Die Hoffnung ein standardisiertes Bewertungskonzept so zu implementieren,

dass es die Bewertungsprozesse im Sinne der oben formulierten Zielsetzungen beeinflusst, erscheint utopisch. Entsprechend kann aus der Evaluation der Konzepterprobung geschlussfolgert werden, dass es für produktive und transparente Bewertungsprozesse in der zweiten Phase der Lehrerbildung sinnvoller erscheint, wenn die einzelnen Ausbilder in einen *individuellen Entwicklungsprozess zur Professionalisierung ihrer Bewertungspraxis* einsteigen, als die Bewertungspraxis durch standardisierte Konzepte entwickeln zu wollen. Die Standards guter Bewertung, wie sie im Kapitel 5.5 entwickelt wurden, bieten Qualitätskriterien für diesen individuellen Entwicklungsprozess an. *Zweitens* deuten die Ergebnisse der Erprobung darauf hin, dass die von den Ausbildern vorgebrachten Einwände gegen ein fächerübergreifend eingesetztes, standardisiertes Bewertungskonzept, teils durchaus sachlich begründete Einwände sind und damit ‚mehr‘ als der ‚normale Widerstand‘ gegen institutionelle Entwicklungsprozesse. Die Ausführungen von A5 und A6, sowie die der Ausbilder der Fächer Deutsch und Sport, haben gezeigt, dass sie mitunter gute Gründe hatten, an ihrer tradierten Bewertungspraxis festzuhalten. Hier ist zum Einen etwa auf die mangelnde Passung des Konzepts mit fachdidaktischen Bewertungskategorien zu verweisen, welche eine entscheidende Rolle für die Qualitätseinschätzung der Stunde spielten. Zum Anderen gab es durchaus begründete Einwände gegen die Logik des Konzepts, mittels isolierter Qualitätsbereiche Unterrichtsqualität beurteilen zu wollen. Wurde das Konzept in diesen Fällen trotzdem eingesetzt, erzeugte es widersprüchliche Urteile und sorgte für Missverständnisse zwischen den Akteuren. Diesen rational begründeten Widerstand gegen das standardisierte Bewertungskonzept, ja die Standardisierung der Bewertungsprozesse an sich, kann man nur schwerlich mit Prozessbegleitung, bzw. Change Management begegnen. Vielmehr erscheint auch hier die an gemeinsamen Qualitätskriterien orientierte, aber primär individuell gesteuerte Entwicklung der Bewertungspraxis der erfolgsversprechendere Weg zu sein

Die stark ausgeprägte Autonomie der Ausbilder ermöglicht es demnach, dass die Ausbildungsinstitutionen, in den Worten von Oelkers, „listig“ auf die „Reformattacke“ der Standardisierung reagieren. Die autonomen Handlungsmöglichkeiten der Ausbilder stellen quasi den „Immunkörper“ der zweiten Phase

dar, „der abwehrt, was nur die Belastungen steigert“ (vgl. Oelkers 2006: 106). Insofern kann man aus diesen Erkenntnissen aber auch den mit der standardbasierten Bildungsreform einhergehenden Befürchtungen einer totalitären Vereinnahmung des Bildungssystems für die zweite Phase der Lehrerbildung eine Absage erteilen. Der „Immunkörper“ des Systems bzw. der Institutionen und der in diesen agierenden Akteure funktioniert sehr effektiv gegen Neuerungen und weist eine hohe Stabilität gewachsener Praktiken auf. Totalitäre bzw. autoritäre Systemlogiken entstehen in der zweiten Phase der Lehrerbildung vielmehr aus den Abhängigkeitsverhältnissen zwischen den Auszubildenden und den Ausbildern und nicht aus Standardisierungsbestrebungen.

Bezieht man dieses Fazit auf die in Kapitel 2.2.4 formulierten Thesen bezüglich der Wirkungen eines standardisierten Bewertungskonzepts, so weist die erste der beiden Thesen, die von einer Stabilität der Bewertungspraxen der Ausbilder ausgeht, deutlich höhere empirische Relevanz auf als die zweite, die von einer Steigerung der Transparenz und Klarheit der Bewertungsprozesse durch den Einsatz eines standardisierten Instruments ausgeht. Eine reflektierte und transparent gemachte individuelle Bewertungspraxis erscheint deswegen sinnvoller als ein fach- und ausbilderübergreifend eingesetztes, standardisiertes Konzept. Die durch die Evaluation der Konzepterprobung rekonstruierten Hindernisse eines erfolgreichen Konzeptesinsatzes können demnach auch nur partiell als ‚Kinderkrankheiten‘ eines neu implementierten Bewertungskonzepts verstanden werden, auf die mit Change Management angemessen reagiert werden kann. Vielmehr scheinen sie stabile Kontextbedingungen der Bewertungsprozesse zu sein, mit denen ein Bewertungskonzept dauerhaft umgehen muss. Abschließend kann gesagt werden, dass die in der Einleitung formulierte Einschätzung, wonach der „Flaschenhals“ der standardbasierten Bildungsreform für die zweite Phase der Lehrerbildung die Entwicklung von praxistauglichen Diagnostikinstrumenten ist, umformuliert werden muss. Die Arbeit hat gezeigt, dass für die erfolgreiche Umsetzung der standardbasierten Bildungsreform im Bereich der Lehrerbildung die Entwicklung von standardisierten Diagnostikinstrumenten keine hinreichende Bedingung ist. *Die Frage der Implementation scheint vielmehr der entscheidende „Flaschenhals“ zu sein, den die Reform überwinden muss, will sie Wirkung zeigen.* Die aus der

Konzepterprobung gewonnen Erkenntnisse bezüglich der Strukturen, in welchen die Diagnostikinstrumente eingesetzt werden, deuten allerdings darauf hin, dass eine erfolgreiche Implementation standardisierter Instrumente und Konzepte am „Immunkörper“ des Systems scheitern werden und der Einsatz standardisierter Instrumente für gewinnbringende Bewertungsprozesse, welche die Entwicklung der Auszubildenden unterstützen, auch nur bedingt förderlich wären. Folglich schließt die Arbeit mit einem Fazit, das der in der Einleitung geschilderten Intention des Projektes entgegenläuft. Frei nach Lipowskys Schlussfolgerung, dass es für erfolgreichen Unterricht „auf die Lehrer ankommt“ (vgl: Lipowsky 2006), kann hier konstatiert werden: Auf die Ausbilder kommt es an! Genauer kommt es auf ihre Bereitschaft an, die eigenen Bewertungspraktiken zu hinterfragen und diese für sich und ihre Auszubildenden begründend in transparente Instrumente zu übersetzen, deren Design sich an seminarübergreifend gültigen Qualitätsvorstellungen ‚guter Bewertungen‘ orientieren. Standardisierte Bewertungskonzepte sind in diesem Prozess „auch nur ein Arbeitsinstrument“, wie es A3 in der Gruppendiskussion der Ausbilder formulierte.

Literatur- und Quellenverzeichnis

- Abel, Jürgen & Faust, Gabriele* (Hrsg.) (2010): *Wirkt Lehrerbildung?... Antworten aus der empirischen Forschung*. Münster, New York, München, Berlin: Waxmann.
- Abs, Hermann Josef* (2007): Überlegungen zur Modellierung diagnostischer Kompetenz bei Lehrerinnen und Lehrern. In: Lüders, Manfred & Wissinger Jochen (Hrsg.): *Kompetenzentwicklung und Programmevaluation. Forschung zur Lehrerbildung*. Münster, New York, München, Berlin: Waxmann, S. 63–84.
- Abs, Hermann Josef & Anderson-Park, Eva* (2014): Programme zur Berufseinführung: die zweite Phase der Lehrerbildung. In: Terhart, Ewald/Bennewitz, Hedda/Rothland, Martin (Hrsg.): *Handbuch der Forschung zum Lehrerberuf*. Münster, New York, München, Berlin: Waxmann, S. 489–510.
- Allemann-Ghionda, Christina & Terhart, Ewald* (2006): Kompetenzen und Kompetenzentwicklungen von Lehrerinnen und Lehrern: Ausbildung und Beruf. In: *Zeitschrift für Pädagogik* (52), 51. Beiheft, S. 7-11.
- Altrichter, Herbert & Feindt, Andreas* (2011): Lehrerinnen und Lehrer erforschen ihren Unterricht: Aktionsforschung. In: Terhart, Ewald/Bennewitz, Hedda/Rothland, Martin (Hrsg.): *Handbuch der Forschung zum Lehrerberuf*. Münster, New York, Berlin, München: Waxmann, S. 214–231.
- American Education Research Association* (2004): *Teacher matters: Evidence from Value-Added Assessments*. Onlinedokument, online unter: <http://www.aera.net/Portals/38/docs/Publications/Teachers%20Matter.pdf>, abgerufen am 30.8.2012.
- Antoni, Conny* (2007): Partizipation. In: Schuller, Heinz & Sonntag, Karl Heinz (Hrsg.): *Handbuch der Arbeits- und Organisationspsychologie*. Göttingen: Hogrefe, S. 773–780.
- Arnold, Karl-Heinz* (2009): Lehr-Lernforschung ohne Allgemeine Didaktik? Über die Notwendigkeit einer integrierten Wissenschaft vom Unterricht. In: Arnold, Karl-Heinz/Blömeke, Sigrid/Messner, Rudolf/Schlömerkemper, Jörg (Hrsg.) (2009): *Allgemeine Didaktik und LehrLernforschung. Kontroversen und Entwicklungsperspektiven einer Wissenschaft vom Unterricht*. Bad Heilbrunn: Klinkhardt, S. 27–45.
- Arnold, Karl-Heinz/Blömeke, Sigrid/Messner, Rudolf/Schlömerkemper, Jörg* (Hrsg.) (2009): *Allgemeine Didaktik und Lehr-Lernforschung. Kontroversen und Entwicklungsperspektiven einer Wissenschaft vom Unterricht*. Bad Heilbrunn: Klinkhardt.
- Baecker, Dirk* (Hrsg.) (2004): *Niklas Luhmann. Einführung in die Systemtheorie*. Heidelberg: Carl-Auer-Systeme Verlag.
- Bellmann, Johannes* (2012): „The very Speedy solution“ – Neue Erziehung und Steuerung im Zeichen von Social Efficiency. In: *Zeitschrift für Pädagogik* (58), Heft 2, S. 143–158.
- Baumert, Jürgen & Köller, Olaf* (2000): Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In: Baumert, Jürgen/Bos, Wilfried/Lehmann, Rainer (Hrsg.): *TIMMSS/III, Band 2*, Opladen: Leske und Budrich, S. 271–315.

- Baumert, Jürgen & Kunter, Mareike* (2006): Stichwort: Professionelle Kompetenz von Lehrkräften. In: *Zeitschrift für Erziehungswissenschaft* (9), Heft 4, S. 469–520.
- Becker, Georg E.* (2008): *Unterricht durchführen. Handlungsorientierte Didaktik Teil II.* Weinheim und Basel: Beltz.
- Benner, D.* (2014): *Allgemeine Pädagogik. Eine systematisch-problemgeschichtliche Einführung in die Grundstrukturen pädagogischen Denkens und Handelns.* Weinheim und München: Beltz Juventa.
- Blömeke, Sigrid* (2004): Empirische Befunde zur Wirksamkeit der Lehrerbildung. In: *Blömeke, Sigrid/Reinhold, Peter/Tulodziecki, Gerhard/Wildt, Johannes* (2004): *Handbuch der Lehrerbildung.* Bad Heilbrunn: Klinkhardt, S. 59–91.
- Blömeke, Sigrid* (2007): Qualitativ – Quantitativ, induktiv – deduktiv, Prozess – Produkt, national – international. In: *Lüders, Manfred & Wissinger Jochen* (Hrsg.): *Kompetenzentwicklung und Programmevaluation. Forschung zur Lehrerbildung.* Münster, New York, München, Berlin: Waxmann, S. 13–36.
- Bohl, Thorsten/Helsper, Werner/Holtappels, Heinz Günter/Schelle, Carla* (Hrsg.): *Handbuch Schulentwicklung.* Bad Heilbrunn: Verlag Julius Klinkhardt.
- Bohmsack, Fritz* (1995): Widerstand von Lehrkräften gegen Innovationen in der Schule. In: *Die Deutsche Schule* (87), Heft 1, S. 21–35.
- Borko, Hilda/Whitcomb, Jennifer A./Byrnes, Kathryn* (2008): Genres of research in teacher education. In: *Cochran-Smith, Marilyn/Feiman-Nemser, Sharon/McIntyre, D. John/Demers, Kelly E.* (Hrsg.): *Handbook of research in teacher education. Enduring questions in changing contexts* (third edition). New York: Routledge, S. 1017–1049.
- Bortz, Jürgen & Döring, Nicola* (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Heidelberg: Springer.
- Bovet, Gislinde & Frommer, Helmut* (2009³): *Praxis Lehrerberatung – Lehrerbeurteilung.* Baltmannsweiler: Schneider Verlag.
- Brophy, Jere* (1999): *Teaching.* In: *Educational Practices Series, 1.* Onlinedokument, online unter: http://www.ibe.unesco.org/fileadmin/user_upload/archive/publications/EducationalPracticesSeriesPdf/prac01e.pdf, abgerufen am 22.1.2013.
- Bruder, Simone/Klug, Julia/Hertel, Silke/Schmitz, Bernhard* (2010): Modellierung der Beratungskompetenz von Lehrkräften. In: *Zeitschrift für Pädagogik* (56), 56. Beiheft, S. 274–285.
- Buhren, Claus G. & Rolff, Hans-Günter* (2012): *Handbuch Schulentwicklung und Schulentwicklungsberatung.* Weinheim und Basel: Beltz.
- Cochran-Smith, Marilyn* (2005): The new teacher education: For better or for worse? In: *Educational Researcher* (34), Heft 7, S. 3–17.
- Cochran-Smith, Marilyn/Feiman-Nemser, Sharon/McIntyre, D. John/Demers, Kelly E.* (Hrsg.) (2008³): *Handbook of research in teacher education. Enduring questions in changing contexts.* New York: Routledge.

- Dietrich, Fabian* (2014): Professionalisierungskrisen im Referendariat. Rekonstruktion zu Krisen und Krisenbearbeitungen in der zweiten Phase der Lehrerbildung. Wiesbaden: Springer VS.
- Ditton, Hartmut* (2006:) Unterrichtsqualität. In: Arnold, Karl-Heinz/Sandfuchs, Uwe/Wiechmann, Jürgen (Hrsg.): Handbuch Unterricht. Bad Heilbrunn: Klinkhardt, S. 235–243.
- Dollase, Rainer* (2007:) Mehrperspektivische Bewertung in der Lehrerbildung. In: Seminar, Lehrerbildung und Schule (13), Heft 1, S. 10–19.
- Dresing, Thorsten & Pehl, Thorsten* (2012⁴): Praxisbuch Interview und Transkription. Regelsysteme und Anleitungen für qualitative ForscherInnen. Onlinedokument, online unter: www.audiotranskription.de/praxisbuch, abgerufen am 5.12.2012.
- Elke, Gabriele* (2007): Veränderung von Organisationen–Organisationsentwicklung. In: Schuller, Heinz & Sonntag, Karl Heinz (Hrsg.): Handbuch der Arbeits- und Organisationspsychologie. Göttingen: Hogrefe, S. 752–759.
- Englert, Rudolf/Porzelt, Burkhard/Reese, Annegret/Stams, Elisa* (2006): Innenansichten des Referendariats. Wie erleben angehende Religionslehrer/innen an Grundschulen ihren Vorbereitungsdienst? Eine empirische Untersuchung zur Entwicklung (religions-)pädagogischer Handlungskompetenz. Berlin: Lit-Verlag.
- Fischer, Christian & Schilmöller, Reinhard* (2010)(Hrsg.): Was ist guter Unterricht? Qualitätskriterien auf dem Prüfstand. Münster: Aschendorf.
- Flick, Uwe* (2005³): Qualitative Sozialforschung–Eine Einführung. Reinbek bei Hamburg: Rowohlt.
- Flick, Uwe* (2006): Qualitative Evaluationsforschung zwischen Methodik und Pragmatik – Einleitung und Überblick. In: ders. (Hrsg.): Qualitative Evaluationsforschung Konzepte – Methoden – Umsetzungen. Reinbek bei Hamburg: Rowohlt, S. 9–29.
- Frey, Andreas & Jung, Claudia* (2011): Kompetenzmodelle und Standards in der Lehrerbildung. In: Terhart, Ewald/Bennewitz, Hedda/Rothland, Martin (Hrsg.) (2011): Handbuch der Forschung zum Lehrerberuf. Münster, New York, München, Berlin: Waxmann, S. 540–572.
- Frey, Andreas* (2006): Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften – eine erste Standortbestimmung zu bereits publizierten Instrumenten. In: Zeitschrift für Pädagogik (52), 51. Beiheft, S. 30–45.
- Forst, Ursula* (Hrsg.) (2006): Unternehmen Bildung. Die Frankfurter Einsprüche und kontroverse Positionen zur aktuellen Bildungsreform. Paderborn: Schöningh. (= Sonderheft der Vierteljahresschrift für wissenschaftliche Pädagogik)
- Gerstenberg, Frank* (2007): „Die schlimmste Zeit meines Lebens“. Süddeutsche Zeitung, 15.01.2007, online unter: <http://www.sueddeutsche.de/karriere/lehrer-im-referendariat-die-schlimmste-zeit-meines-lebens-1.592394>, abgerufen am 09.03.2012.
- Goodwin, A. Lin & Oyler, Celia* (2008): Teacher educators as gatekeepers: deciding who is ready to teach. In: Cochran-Smith, Marilyn/Feiman-Nemser, Sharon/McIntyre, D. John/Demers, Kelly E. (Hrsg.): Handbook of research in teacher education. Enduring questions in changing contexts (third edition). New York: Routledge, S. 468–490.

- Gräsel, Cornelia/Krolak-Schwerdt, Sabine/Nölle, Ines/Hörstermann, Thomas* (2010): Diagnostische Kompetenz von Grundschullehrkräften bei der Erstellung der Übergangsempfehlung: Eine Analyse aus der Perspektive der sozialen Urteilsbildung. In: Zeitschrift für Pädagogik (56), 56. Beiheft, S. 286–295.
- Grouws, Douglas A. & Cebulla Kristin J.* (2000): Improving student achievement in mathematics. In: Educational Practices Series, 4. Onlinedokument, online unter: http://www.ibe.unesco.org/fileadmin7/user_upload/archive/publicationEducationalPracticesSeriesPdf/prac04e.pdf, abgerufen am 22.01.2013.
- Gruschka, Andreas* (2006): Bildungsstandards oder das Versprechen, Bildungstheorie in empirischer Bildungsforschung aufzuheben. In: Frost, Ursula (Hrsg.): Unternehmen Bildung. Die Frankfurter Einsprüche und kontroverse Positionen zur aktuellen Bildungsreform. Paderborn: Schöningh, S. 140–158.
- Gruschka, Andreas* (2007): „Was ist guter Unterricht?“ In: Pädagogische Korrespondenz. Heft 36 (20), S. 10–36.
- Gruschka, Andreas* (2013): Bildung – Kompetenz. In: Asdonk, Jupp (Hrsg.): Von der Schule zur Hochschule. Analysen, Konzeptionen und Gestaltungsperspektiven des Übergangs. Münster, New York, München, Berlin: Waxmann, S. 77–86.
- Hattie, John* (2009): Visible Learning. A Synthesis of over 800 meta-analyses relating to achievement. London, New-York: Routledge.
- Helmke, Andreas & Weinert, Franz Emanuel* (Hrsg.) (1997): Entwicklung im Grundschulalter. Weinheim: Beltz.
- Helmke, Andreas* (2006): Was wissen wir über guten Unterricht? In: Pädagogik (58), Heft 2, S. 42–45.
- Helmke, Andreas/Helmke, Tuyet/Schrader, Friedrich-Wilhelm* (2007): Unterrichtsqualität: Brennpunkte und Perspektiven der Forschung. In: Arnold, Karl-Heinz (Hrsg.): Unterrichtsqualität und Fachdidaktik. Bad Heilbrunn: Klinkhardt, S. 51–72.
- Helmke, Andreas* (2010): Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts. Seelze-Velber: Klett/Kallmeyer.
- Herrmann, Ulrich & Hertrampf, Herbert* (2002): „Lehrer“ – eine Selbstdefinition. In: Herrmann, Ulrich (Hrsg.): Wie lernen Lehrer ihren Beruf? Empirische Befunde und praktische Vorschläge. Weinheim und Basel: Beltz, S. 200–220.
- Holtappels, Heinz Günter & Rolff, Hans-Günter* (2010): Einführung: Theorien der Schulentwicklung. In: Bohl, Thorsten/Helsper, Werner/Holtappels, Heinz Günter/Schelle, Carla (Hrsg.): Handbuch Schulentwicklung. Bad Heilbrunn: Verlag Julius Klinkhardt, S. 73–79.
- Jürgens, Eiko & Standop, Jutta* (Hrsg.) (2010): Was ist guter Unterricht? Namhafte Expertinnen und Experten Antworten. Bad Heilbrunn: Klinkhardt.
- Kelle, Udo & Kluge, Susanne* (2010²): Vom Einzelfall zum Typus – Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung. Wiesbaden: VS Verlag.

- Kiel, Ewald* (2010): Unterrichtsforschung. In: Tippelt, Rudolf & Schmidt, Bernhard (Hrsg.) (2010³) Handbuch Bildungsforschung. Wiesbaden: VS Verlag, S. 773–790.
- Klieme, Eckhard* (2002): Was ist guter Unterricht? Ergebnisse der TIMSS-Videostudie im Fach Mathematik. In: Bergsdorf, Wolfgang/Court, Jürgen/Eckert, Manfred/Hoffmeister, Hans (Hrsg.): Herausforderungen der Bildungsgesellschaft. Ringvorlesung im Sommersemester 2002. Weimar: Rhino Verlag, S. 89–113.
- Klieme, Eckhard u.a.* (2003): Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Hrsg. v. Bundesministerium für Bildung und Forschung (BMBF). Bonn.
- Klieme, Eckhard* (2004): Was sind Kompetenzen und wie lassen sie sich messen? In: Pädagogik (56), Heft 4, S. 10–13.
- Klieme, Eckhard* (2002): Bildungsstandards als Beitrag zur Qualitätsentwicklung im Schulsystem. In: DIPF informiert, Nr. 3, August 2002, S. 2–6. Onlinedokument, online unter: <http://www.dipf.de/de/publikationen/pdf/dipf-informiert/dipf-informiert-nr.-3>, abgerufen am 19.6.2013.
- Klieme, Eckhard & Rakoczy, Katrin* (2003): Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland. Opladen: Leske und Budrich, S. 333–360.
- Klieme, Eckhard & Hartig Johannes* (2007): Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In: Zeitschrift für Erziehungswissenschaft (10), Sonderheft 8/2007, S. 11–29.
- Kloepfer, Inge* (2012): Frontalunterricht macht klug. In: Frankfurter Allgemeine Zeitung, 15.12.2012. Onlinedokument, online unter: <http://www.faz.net/aktuell/wirtschaft/bildungswesen-frontalunterricht-macht-klug11994686.html>, abgerufen am 24.01.2013.
- Koch, Lutz* (2010): Kompetenz: Konstrukt zwischen Defizit und Anmaßung. In: Vierteljahresschrift für wissenschaftliche Pädagogik (86), Heft 3, S. 323–331.
- Koller, Hans-Christoph* (2012): Grenzsicherung durch Annäherung? Zum Spannungsverhältnis zwischen Bildungstheorie und empirischer Bildungsforschung. In: Zeitschrift für Pädagogik (58), Heft 1, S. 6–21.
- Köller, Olaf* (2010): Bildungsstandards. In: Tippelt, Rudolf & Schmidt, Bernhard (Hrsg.) (2010³) Handbuch Bildungsforschung. Wiesbaden: VS Verlag, S. 529–548.
- Kösel, Edmund* (2002⁴): Die Modellierung von Lernwelten, Band I – Die Theorie der Subjektiven Didaktik -. Bahlingen: SD-Verlag.
- Kromrey, Helmut* (2006¹¹): Empirische Sozialforschung. Stuttgart: Lucius und Lucius.
- Kuckartz, Udo* (2006): Quick and dirty? – Qualitative Methoden der drittmittelfinanzierten Evaluation in der Umweltforschung. In: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung Konzepte – Methoden – Umsetzungen. Reinbek bei Hamburg: Rowohlt, S. 267–283.
- Kuckartz, Udo/Dresing, Thorsten/Rädiker, Stefan/Stefer, Claus* (2008²): Qualitative Evaluation. Der Einstieg in die Praxis. Wiesbaden: VS Verlag.

- Kuckartz, Udo* (2010): Einführung in die computergestützte Analyse qualitativer Daten. Wiesbaden: VS Verlag.
- Kühn, Thomas & Koschel, Kay-Volker* (2011): Gruppendiskussionen – ein Praxishandbuch. Wiesbaden: VS Verlag.
- Kultusministerkonferenz (2004a)*: Standards für die Lehrerbildung: Bildungswissenschaften i.d.F. Vom 12.06.2014. Online unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf, abgerufen am 30.12.2010.
- Kultusministerkonferenz (2004b)*: Bildungsstandards der Kultusministerkonferenz. Online unter: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Konzeption-Entwicklung.pdf, abgerufen am 16.8.2012.
- Kultusministerkonferenz (2008)*: Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken. Online unter: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildungb.pdf, abgerufen am 30.12.2014.
- Kunter, Mareike* (2010): Modellierung von Lehrerkompetenzen. In: Zeitschrift für Pädagogik (56), 56. Beiheft, S. 307–312.
- Kunter, Marieke & Klusmann Uta* (2010): Kompetenzmessung bei Lehrkräften – Methodische Herausforderungen. In: Unterrichtswissenschaft (38), Heft 10, S. 68–86.
- Kunter, Marike und Voss Thamar* (2011): Das Modell der Unterrichtsqualität in COACTIV: eine multikriteriale Analyse. In: Kunter, Marike/Baumert, Jürgen/Blum, Werner/Klusmann, Uta/Krauss, Stefan/Neubrand, Michael (Hrsg.): Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV. Münster, New York, München, Berlin: Waxmann, S. 85–113.
- Kunter, Mareike/Kleickmann, Thilo/Klusmann, Uta/Richter, Dirk* (2011): Die Entwicklung professioneller Kompetenz von Lehrkräften. In: Kunter, Marike/Baumert, Jürgen /Blum, Wener/Klusmann, Uta/Krauss, Stefan/Neubrand, Michael (Hrsg.): Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV. Münster, New York, München, Berlin: Waxmann, S. 55–68.
- Lacher, Sabine und Oelkers, Jürgen* (2004): Deutsche Lehrerbildung im internationalen Vergleich. In: Blömeke, Sigrid/Reinhold, Peter/Tulodziecki, Gerhard/Wildt, Johannes (Hrsg.): Handbuch der Lehrerausbildung. Bad Heilbrun: Klinkhardt, S. 128–150.
- Ladenthin, Volker* (2005): Letzte Wahrheit: Bildungsstandards. In: PÄD Forum – unterrichten, erziehen. Hohengehren: Schneider Verlag, S. 11.
- Ladenthin, Volker* (2010): Kompetenzorientierung als Indiz pädagogischer Orientierungslosigkeit. In: Vierteljahresschrift für wissenschaftliche Pädagogik (86), Heft 3, S. 346–358.
- Lamnek, Siegfried* (2005^{2a}): Gruppendiskussionen, Theorie und Praxis. Weinheim und Basel: Beltz.
- Lamnek, Siegfried* (2005^b): Qualitative Sozialforschung. Weinheim und Basel: Beltz.
- Lehmann, Rena* (2012): Lehrer klagen: Erziehung mangelhaft. In: Rhein-Zeitung, 22.03.2012.

- Lehmeier, Heinz* (1995): Braucht die handlungswissenschaftlich begründete Pädagogik eine eigene Forschungsmethode? In: *Zeitschrift für Pädagogik* (41), Heft 3, S. 631–650.
- Lenhard, Hartmut* (2004): Zweite Phase an Studienseminaren und Schulen. In: Blömeke, Sigrid/Reinhold, Peter/Tulodziecki, Gerhard/Wildt, Johannes (2004): *Handbuch der Lehrerbildung*. Bad Heilbrunn: Klinkhardt, S. 275–290.
- Lennartz, Alice* (2010): Unterrichtsbeobachtungen – Ein Instrument zur Qualitätsentwicklung. In: Fischer, Christian & Schilmöller Reinhard (Hrsg.): *Was ist guter Unterricht? Qualitätskriterien auf dem Prüfstand*. Münster: Aschendorf, S. 160–168.
- Lipowsky, F.* (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In: *Zeitschrift für Pädagogik* (51), 51. Beiheft, S. 47-65.
- Lüders, Manfred & Rauin, Udo* (2004): Unterrichts- und Lehr-Lern-Forschung. In: Helsper, Werner & Böhme, Jeanette (Hrsg.): *Handbuch der Schulforschung*. VS Verlag, Opladen, S. 691-720.
- Luhmann, Niklas & Schorr, Karl Eberhard* (1982): Das Technologiedefizit der Erziehung und die Pädagogik. In: dies. (Hrsg.): *Zwischen Technologie und Selbstreferenz: Fragen an die Pädagogik*. Frankfurt a.M.: Suhrkamp, S. 11–40.
- Luhmann, Niklas* (1986): Systeme verstehen Systeme. In: Luhmann, Niklas & Schorr, Karl Eberhard (Hrsg.): *Zwischen Intransparenz und verstehen: Fragen an die Pädagogik*. Frankfurt a.M.: Suhrkamp, S. 72–118.
- Luhmann* (2002): *Das Erziehungssystem der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- Luhmann* (2004): Erziehender Unterricht als Interaktionssystem. In: Lenzen, Dieter (Hrsg.): *Niklas Luhmann – Schriften zur Pädagogik*. Frankfurt am Main: Suhrkamp, S. 11–22.
- Maag Merki, Katharina & Werner, Silke* (2011): Erfassung und Bewertung professioneller Kompetenz. In: Terhart, Ewald/Bennewitz, Hedda/Rothland, Martin (Hrsg.): *Handbuch der Forschung zum Lehrerberuf*. Münster, New York, München, Berlin: Waxmann, S. 573–591.
- Meyer, Hilbert* (2003): Zehn Merkmale guten Unterrichts. Empirische Befunde und didaktische Ratschläge. In: *Pädagogik* (55), Heft 10, S. 36–43.
- Meyer, Hilbert* (2007a⁴): *Was ist guter Unterricht?* Berlin: Cornelsen Scriptor.
- Meyer, Hilbert* (2007b): Zehn Merkmale guten Unterrichts. Anleitung zur Überarbeitung und zum Einsatz eines Beobachtungsbogens. In: Endres, Wolfgang (Hrsg.): *Lernen lernen – Wie stricken ohne Wolle? 13 Experten streiten über Konzepte und Modelle zur Lernmethodik*. Weinheim und Basel: Beltz, S. 167–187.
- Meyer, Hilbert* (2010a): Merkmale guten Unterrichts – Ein Kriterienmix. In: Jürgens, Eiko & Standop, Jutta (Hrsg.): *Was ist guter Unterricht? Namhafte Expertinnen und Experten Antworten*. Bad Heilbrunn: Klinkhardt, S. 159–174.
- Meyer, Hilbert* (2010b): Qualitätsmerkmale guten Unterrichts in der Diskussion. In: Fischer, Christian & Schilmöller Reinhard (Hrsg.): *Was ist guter Unterricht? Qualitätskriterien auf dem Prüfstand*. Münster: Aschendorf, S. 6–38.

- Oelkers, Jürgen* (2006): Qualitätsentwicklung in Schulen: Probleme und konkrete Erfahrungen aus der Schweiz. In: Frost, Ursula (Hrsg.): *Unternehmen Bildung. Die Frankfurter Einsprüche und kontroverse Positionen zur aktuellen Bildungsreform*. Paderborn: Schöningh, S. 105–125.
- Oelkers, Jürgen & Reusser, Kurt* (2008): *Qualität entwickeln – Standards sichern – mit Differenzen umgehen*. Bonn: Bundesministerium für Bildung und Forschung.
- Oser, Fritz & Renold, Ursula* (2005): Kompetenzen von Lehrpersonen – über das Auffinden von Standards und ihre Messung. In: *Zeitschrift für Erziehungswissenschaft* (8), Beiheft 4/2005, S. 119–140.
- Oser, Fritz/Düggeli, Albert/Heinzer, Sarah* (2010): Qualitätsmessung von Lehrpersonen-Kompetenzen. Ein neuer Ansatz. In: Abel, Jürgen & Faust, Gabriele (Hrsg.): *Wirkt Lehrerbildung? Antworten aus der empirischen Forschung*. Münster, New York, München, Berlin: Waxmann, S. 133–153.
- Paulsen, Nina* (2012): Guter Lehrer, guter Unterricht. In: *Hamburger Abendblatt*, 12.03.2012.
- Pille, Thomas* (2013): *Das Referendariat. Eine ethnographische Studie zu den Praktiken der Lehrerbildung*. Bielefeld: Transcript Verlag.
- Przyborski, Aglaja & Wohlrab-Sahr, Monika* (2009²): *Qualitative Sozialforschung – Ein Arbeitsbuch*. München: Oldenbourg.
- Radtke, Frank-Olaf* (2006): Das neue Erziehungsregime. Steuerungserwartungen, Kontrollphantasien und Rationalitätsmythen. In: Frost, Ursula (Hrsg.): *Unternehmen Bildung. Die Frankfurter Einsprüche und kontroverse Positionen zur aktuellen Bildungsreform*. Paderborn: Schöningh, S. 45–49.
- Reich, Kersten* (2008⁴). *Konstruktivistische Didaktik. Lehr- und Studienbuch mit Methodenpool*. Weinheim und Basel: Beltz.
- Reh, Sabine* (2005): Rezension von: Meyer, Hilbert: *Was ist guter Unterricht?*, Cornelsen Verlag Scriptor 2004. In: *Erziehungswissenschaftliche Revue* (4), Nr. 3 Onlinedokument, online unter: <http://www.klinkhardt.de/ewr/58922047.html>, abgerufen am 25.01.2013.
- Reh, Sabine* (2010): Widerstand in Schulentwicklungsprozessen. In: Bohl, Thorsten/Helsper, Werner/Holtappels, Heinz Günter/Schelle, Carla (Hrsg.): *Handbuch Schulentwicklung*. Bad Heilbrunn: Verlag Julius Klinkhardt, S. 292–294.
- Rolff, Hans-Günter* (2007): *Studien zu einer Theorie der Schulentwicklung*. Weinheim und Basel: Beltz.
- Rolff, Hans-Günter/Buhren, Claus G./Lindau-Bank, Detlev/Müller, Sabine* (2011⁴): *Manual Schulentwicklung. Handlungskonzepte zur pädagogischen Schulentwicklungsberatung*. Weinheim und Basel: Beltz.
- Saleh, Issa M. & Khine, Myint Swe* (Hrsg.) (2011): *Practitioner Research in Teacher Education. Theory and Best Practices*. Frankfurt a.M.: Peter Lang.

- Schiersmann, Christiane & Thiel, Heinz-Ulrich* (2011³): Organisationsentwicklung – Prinzipien und Strategien von Veränderungsprozessen. Wiesbaden: VS Verlag.
- Schilmöller, Reinhard* (2006): Guter Unterricht – eine Technik? In: Vierteljahresschrift für wissenschaftliche Pädagogik (82), Heft 1, S. 70–88.
- Schilmöller, Reinhard* (2010): Guter Unterricht – Gute Schule. Notwendige Kontexte gelingender Unterrichtspraxis. In: Fischer, Christian & Schilmöller Reinhard (Hrsg.): Was ist guter Unterricht? Qualitätskriterien auf dem Prüfstand. Münster: Aschendorf, S. 76–102.
- Schnell, Rainer/Hill, Paul B./Esser, Elke* (2011⁹): Methoden der empirischen Sozialforschung. München: Oldenbourg.
- Schorn, Ariane* (2000). Das „themenzentrierte Interview“. Ein Verfahren zur Entschlüsselung manifester und latenter Aspekte subjektiver Wirklichkeit. Forum Qualitative Sozialforschung (1), Heft 2, Art. 23, online unter: <http://nbnresolving.de/urn:nbn:de:0114-fqs0002236>, abgerufen am 06.08.2012
- Schubarth, Wilfried /Speck, Karsten/Große, Ulrike/Seidel, Andreas/Gemsa, Charlotte* (2006): Die zweite Phase der Lehrerbildung aus Sicht der Brandenburger Lehramtskandidatinnen und Lehramtskandidaten. Die Potsdamer LAK-Studie 2004/2005. In: Schubarth, Wilfried & Pohlenz, Philipp (Hrsg.): Qualitätsentwicklung und Evaluation in der Lehrerbildung. Die Zweite Phase: Das Referendariat. Potsdam: Universitätsverlag Potsdam, S. 13–175.
- Schubarth Wilfried/Speck, Karsten/Seidel, Andreas* (2007): Endlich Praxis! Die zweite Phase der Lehrerbildung. Potsdamer Studien zum Referendariat. Frankfurt am Main: Peter Lang.
- Schulte, Michaela* (2008): Das Studienseminar. Die Ausbildung im Studienseminar 8Gymnasium & Gesamtschule) aus der Perspektive der Referendare. Online unter: http://dokumentix.ub.uni-siegen.de/opus/volltexte/2008/365/pdf/Druckversion_Diss_Schulte.pdf, abgerufen am 15.03.2012.
- Schulte, Klaudia/Bögelholz, Susanne /Watermann, Rainer* (2008): Selbstwirksamkeitserwartungen und pädagogisches Professionswissen. In: Zeitschrift für Erziehungswissenschaft (11), Heft 2, S. 268–287.
- Seidel, Tina & Prenzel, Manfred* (2007): Wie Lehrpersonen Unterricht wahrnehmen und einschätzen – Erfassung pädagogisch-psychologischer Kompetenzen mit Videosequenzen. In: Zeitschrift für Erziehungswissenschaft (10), Sonderheft 8, S. 201–216.
- Seidel, Tina /Blomberg, Geraldine/Stürmer, Kathleen* (2010): „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt Observe. In: Zeitschrift für Pädagogik (56), 56. Beiheft, S. 296–306.
- Shirley, Dennis* (2009): American Perspectives on German Educational Theory and Research – A Closer Look at Both the American Educational Context and the German Didaktik-Tradition. In: Arnold, Karl-Heinz/Blömeke, Sigrid/Messner, Rudolf/Schlömerkemper (Hrsg.): Allgemeine Didaktik und Lehr-Lernforschung. Kontroversen und Entwicklungsperspektiven einer Wissenschaft vom Unterricht. Bad Heilbrunn: Klinkhardt, S. 195–209.

- Stegmaier, Ralf* (2007): Umgehen mit Veränderungen. In: Schuller, Heinz & Sonntag, Karl Heinz (Hrsg.): Handbuch der Arbeits- und Organisationspsychologie. Göttingen: Hogrefe, S. 314–319.
- Storr, Birgit* (2006): „In der Lehrprobe machst du ´ne Show“. Das Referendariat als Gegenstand rekonstruktiver Sozialforschung. Berlin: Logos Verlag.
- Strietholt, Rolf & Terhart, Ewald* (2009): Referendare beurteilen. Eine explorative Analyse von Beurteilungsinstrumenten in der Zweiten Phase der Lehrerbildung. In: Zeitschrift für Pädagogik (55), Heft 4, S. 622–645.
- Strübing, Jörg* (2008²): Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung des Verfahrens der empirisch begründeten Theoriebildung. Wiesbaden: VS Verlag.
- Teller, David* (2008): Gegenseitige Irritation – vom Umgang mit theoretischer Vielfalt. In: Arndt, Friedrich/Dege, Carmen/Ellermann, Christian/Mayer, Maximilian/Teller, David/Zimmermann, Lisbeth (Hrsg.): Ordnungen im Wandel. Globale und lokale Wirklichkeiten im Spiegel transdisziplinärer Analysen. Bielefeld: Transcript Verlag, S. 217–222.
- Teller, David* (2009): Politisch-moralische Urteilsbildung am Beispiel des Luftsicherheitsgesetzes – zwischen „Mündigkeit“ und „Herrschaftslegitimation“. In: Gesellschaft – Wirtschaft – Politik (58), Heft 4, S. 595–605.
- Tenorth, Heinz-Elmar* (2013): Viele Befunde, aber kein Handlungswissen. In: Frankfurter Allgemeine Zeitung, 01.03.2013, Seite 8.
- Terhart, Ewald* (1999): Konstruktivismus und Unterricht. In: Zeitschrift für Pädagogik (45), Heft 5, S. 629–647.
- Terhart, Ewald* (2002): Fremde Schwestern. Zum Verhältnis von Allgemeiner Didaktik und empirischer Lehr- Lernforschung. In: Zeitschrift für pädagogische Psychologie (16), Heft 2, S. 77–86.
- Terhart, Ewald* (2006): Standards und Kompetenzen in der Lehrerbildung. In: Hilligus, Annegret Helen & Rinkens, Hans-Dieter (Hrsg.) (2006): Standards und Kompetenzen – neue Qualität in der Lehrerbildung? Neue Ansätze und Erfahrungen in nationaler und internationaler Perspektive. Münster: LIT, S. 29–42.
- Terhart, Ewald* (2007): Erfassung und Beurteilung der beruflichen Kompetenz von Lehrkräften. In: Lüders, Manfred & Wissinger Jochen (Hrsg.) Kompetenzentwicklung und Programmevaluation. Forschung zur Lehrerbildung. Münster, New York, München, Berlin: Waxmann, S. 37–62.
- Terhart, Ewald/Bennewitz, Hedda/Rothland, Martin* (Hrsg.) (2011): Handbuch der Forschung zum Lehrerberuf. Münster, New York, München, Berlin: Waxmann.
- Terhart, Ewald* (2013): Widerstand von Lehrkräften in Schulreformprozessen: Zwischen Kooperation und Obstruktion. In: McElvany, Nele & Holtappels, Heinz Günter (Hrsg.): Empirische Bildungsforschung. Theorien, Methoden, Befunde und Perspektiven. Münster: Waxmann, S. 75–92.

- Tillmann, Klaus-Jürgen* (2011): Konzepte der Forschung zum Lehrerberuf. In: Terhart, Ewald/Bennewitz, Hedda/Rothland, Martin (Hrsg.): Handbuch der Forschung zum Lehrerberuf. Münster, New York, Berlin, München: Waxmann, S. 232–240.
- Waldow, Florian* (2012): Taylorismus im Klassenzimmer: John Franklin Bobbitts Vorschläge zur *standard-based reform*. In: Zeitschrift für Pädagogik (58), Heft 2, S. 159–175.
- Walke, Jutta* (2007): Die Zweite Phase der Lehrerbildung. Ein Überblick über Stand, Problemlagen und Reformtendenzen. Online unter: http://stifterverband.info/wissenschaft_und_hochschule/lehre/lehrerbildung/die_zweite_phase_der_lehrerbildung.pdf, abgerufen am 03.04.2012.
- Watzlawick, Paul/Beavin, Janet H. /Jackson, DonD.* (1982⁶): Menschliche Kommunikation, Formen, Störungen, Paradoxien. Bern, Stuttgart, Wien: Verlag Hans Huber.
- Werner-Bentke, Frank* (2010): Lehrerausbildung aus der Perspektive von GymnasialreferendarInnen. Eine Deutungsmusteranalyse vor dem Hintergrund der Professionalisierungsdebatte. Hamburg: Verlag Dr. Kovac.
- Witzel, Andreas* (2000). Das problemzentrierte Interview. Forum Qualitative Sozialforschung (1), Heft 1, Art. 22, online unter: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0001228>, abgerufen am 6.08.2012.
- Wolfram, Claudia* (2009): „Manchmal war ich kurz davor aufzugeben. Anspruch und Wirklichkeit. In: Pädagogik (61), Heft 9, S. 19–21.
- Zeichner, Kenneth M.* (2006): Konzepte von Lehrerexpertise und Lehrerausbildung in den Vereinigten Staaten. In: Zeitschrift für Pädagogik (52), 51. Beiheft, S. 97–113.
- Zumwalt, Karen & Craig, Elizabeth* (2005): Teacher's Characteristics: Research on the indicators of Quality. In: Cochran-Smith, Marilyn & Zeichner, Kenneth M. (Hrsg.): Studying Teacher Education. The Report on the AERA Panel on Research on Teacher Education. Mahwah/New Jersey, London: Lawrence Erlbaum Associates. S. 157–260.

Zitierte Verordnungen

Baden-Württemberg:

AprOGymn 2004: Verordnung des Kultusministeriums über den Vorbereitungsdienst und die Zweite Staatsprüfung für die Laufbahn des höheren Schuldienstes an Gymnasien vom 10. März 2004. Online unter: <http://www.landesrecht-bw.de/jportal/?quelle=purl&psml=bsbawueprod.psml&max=true&docId=jlr-GymLehrPrOBWrahmen&doc.part=X>, abgerufen am 9.03.2012.

Bayern:

ZALG 1992: Zulassungs- und Ausbildungsverordnung für das Lehramt an Gymnasien in der Fassung der Bekanntmachung vom 29.9.1992. Online unter: <http://www.gesetze-bayern.de/jportal/portal/page/bsbayprod.psml?doc.id=jlr-GymLehrZAOBY1992rahmen&showdoccase=1¶mfromHL=true#focuspoint>, abgerufen am 9.03.2012.

Bremen:

Ausbildungs- und Prüfungsordnung für Lehramtler 2008: Verordnung über die Ausbildung der Lehramtsreferendarinnen und -referendare im Vorbereitungsdienst und über die Zweite Staatsprüfung für die Lehramter an öffentlichen Schulen

Online unter: http://www.lis.bremen.de/sixcms/media.php/13/APO_fuer%20Lehraemter_Bremen.pdf, abgerufen am 9.03.2012.

Nordrhein-Westfalen:

OVP 2011: Ordnung des Vorbereitungsdienstes und der Staatsprüfung für Lehramter an Schulen. Online unter: http://www.schulministerium.nrw.de/BP/Schulrecht/Lehrerausbildung/OVP_vom_10__April_2011.pdf, abgerufen am 9.03.2012.

Kerncurriculum für die Ausbildung im Vorbereitungsdienst für Lehramter in den Zentren für schulpraktische Lehrerausbildung und in den Ausbildungsschulen: Online unter: <http://www.schulministerium.nrw.de/BP/Schulrecht/Lehrerausbildung/Kerncurriculum.pdf>, abgerufen am 6.05.2013

Kernlehrplan für das Gymnasium – Sekundarstufe I (G8) in Nordrhein-Westfalen Erdkunde : Online unter: http://www.standardsicherung.schulministerium.nrw.de/lehrplaene/upload/lehrplaene_download/gymnasium_g8/gym8_erdkunde.pdf, abgerufen am 5.6.2013

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2006): Richtlinien und Lehrpläne für das Gymnasium – Sekundarstufe I – in Nordrhein-Westfalen, Erdkunde. Frechen: Ritterbach Verlag. (Unveränderter Nachdruck von 1993)

Anhang

Matrix der Qualitätsbereiche guten Unterrichts in Anlehnung an Helmke und Meyer (pre-Alpha-Version)

Qualitätsbereiche Quellen der Kompetenzen und Indikatoren	Erklärung, Operationalisierung in Teilkompetenzen	Indikatoren (in einer Unterrichtsstunde) [teilweise sind einzelne Indikatoren verschiedenen Qualitätsbereichen zugeordnet]
Klarheit und Strukturiertheit <ul style="list-style-type: none"> ▪ klare Strukturierung (Meyer) ▪ inhaltliche Klarheit (Meyer) ▪ Klassenführung (Helmke) ▪ Klarheit und Strukturiertheit (Helmke) 	<ul style="list-style-type: none"> ▪ Stimmigkeit von Zielen, Inhalten und Methoden ▪ Folgerichtigkeit des methodischen Gangs (Ausmaß der Lenkung und der emotionalen Nähe, Komplexität und Abstraktion sind der Lernsituation angepasst) ▪ Folgerichtigkeit des thematischen Gangs (sinnvolle didaktische Inszenierung der Inhalte nach (fach-) didaktischen Prinzipien) ▪ methodischer Grundrhythmus ▪ Aufgaben, Regel- und Rollenklarheit ▪ Aufgabenstellung verständlich und in sich stimmig hinsichtlich der Ziel-, Inhalts- und Methodenwahl (Lernstrukturanalyse; welche Lernschritte muss ein Schüler vollziehen um zu einem Ergebnis zu kommen? Lernstandsanalyse; verfügen die SuS über die notwendigen Ressourcen zum Lösen der Aufgabe?) ▪ Klarheit der Ergebnissicherung (die Unterrichtsstunde bringt einen klar zu benennenden Erkenntnisfortschritt) ▪ Verbindlichkeit der Ergebnissicherung (Klarheit über definitiv zu fixierende Wissens- und Kompetenzbestände der Lerngruppe) 	<ul style="list-style-type: none"> ▪ SuS verstehen Aufgabenstellungen und können erläutern was sie gerade tun ▪ Regeln werden eingehalten ▪ Rituale werden sinnvoll eingesetzt ▪ einzelne Unterrichtsschritte werden klar markiert ▪ Umfang der Störungen ist gering ▪ wenige, aber sinnvolle, Planungskorrekturen ▪ sprachliche Klarheit ▪ Transparenz über aktuelles Thema und übergeordnete Fragestellungen ▪ Anknüpfen an die Vorstellungs- und Lebenswelt der SuS ▪ klare schriftliche Ergebnisfixierungen ▪ Wiederholung und Zusammenfassung wichtiger Ergebnisse, am besten durch SuS
hoher Anteil echter Lernzeit <ul style="list-style-type: none"> ▪ hoher Anteil echter Lernzeit (Meyer) ▪ Klassenführung (Helmke) 	<ul style="list-style-type: none"> ▪ „time on task“ (tatsächlich aufgewendete Lernzeit) ▪ Hohes Momentum (Vermeidung unnötiger Unterbrechungen des Unterrichtsflusses, sowie von Hektik, Herstellen eines „geschmeidigen“ Unterrichtsverlaufs) 	<p>die SuS sind nicht gelangweilt und unaufmerksam die SuS arbeiten aktiv mit und lassen sich nicht durch Kleinigkeiten ablenken es entstehen reichhaltige Arbeitsergebnisse der Lehrer gibt gezielte Lernhilfen der Lehrer lässt den nach Fragen Zeit, zum überlegen</p>

Qualitätsbereiche Quellen der Kompetenzen und Indikatoren	Erklärung, Operationalisierung in Teilkompetenzen	Indikatoren (in einer Unterrichtsstunde) [teilweise sind einzelne Indikatoren verschiedenen Qualitätsbereichen zugeordnet]
lernförderliches Klima <ul style="list-style-type: none"> ▪ lernförderliches Klima (Meyer) ▪ lernförderliches Klima (Helmke) ▪ Schülerorientierung (Helmke) 	<ul style="list-style-type: none"> ▪ Gegenseitiger Respekt SuS/SuS-Lehrer ▪ verlässliche Regeln ▪ gemeinsame Verantwortung von Lehrer und SuS für erfolgreichen Lernprozess ▪ Gerechtigkeit und Fürsorge des Lehrers gegenüber den SuS ▪ konstruktiver Umgang mit Fehlern, auf Lehrer wie auf SuS-Seite ▪ Balance zwischen Lern- und Leistungssituationen ▪ Balance zwischen „Lockerheit“ und „Anspruch“ ▪ Schülerfeedback 	<ul style="list-style-type: none"> ▪ Respektvoller Umgang der SuS untereinander und in der SuS-Lehrer-Beziehung (kein aggressives Verhalten, Beleidigungen und Demütigungen, keine Diskriminierung einzelner SuS, keine Rivalitäten zwischen Gruppen) ▪ humorvolle Atmosphäre bei gleichzeitiger Leistungsorientierung ▪ Ermutigungen durch den Lehrer
sinnstiftendes Kommunizieren <ul style="list-style-type: none"> ▪ (Meyer) 	<ul style="list-style-type: none"> ▪ Kommunikationsprozess der „persönlichen qualitativen Sinngabe“ des Gelehrten für die SuS 	<ul style="list-style-type: none"> ▪ Metakommunikation über die Unterrichtsinhalte, bzw. über den Unterricht ▪ Anregungen des Lehrers zur Einordnung der Inhalte in den eigenen „Sinnhorizont“ der SuS ▪ Verknüpfungen zu anderen Unterrichtsthemen oder Alltagswissen ▪ persönliche Stellungnahmen der SuS zu Inhalten ▪ SuS-Feedback
Methodenvielfalt <ul style="list-style-type: none"> ▪ Methodenvielfalt (Meyer) ▪ Angebotsvielfalt (Helmke) 	<p>der Unterricht wird auf vielfältige Art und Weise inszeniert (Aktions- und Sozialformen, Medieneinsatz, Handlungsformen, methodische Großformen, Lernorte)</p> <ul style="list-style-type: none"> ▪ Unterricht den verschiedenen Lernvoraussetzungen der SuS anpassen (Adaptivität) ▪ auch leistungsschwachen SuS wertschätzend begegnen ▪ Angebote zum Erwerb von Lernstrategien 	<p>der Unterricht zeigt ein vielfältiges Repertoire an den zuvor erwähnten Inszenierungstechniken</p>
Umgang mit Heterogenität <ul style="list-style-type: none"> ▪ individuelles Fördern (Meyer) ▪ Umgang mit Heterogenität (Helmke) 	<ul style="list-style-type: none"> ▪ Innere Differenzierung des Unterrichts (binnendifferenzierende Aufgabenstellungen, unterschiedliche Lernmaterialien werden eingesetzt, für SuS mit Lernschwierigkeiten stehen zusätzliche Lernhilfen bereit, Möglichkeit für leistungsstarke SuS Routineaufgaben nicht zu erledigen und nach eigenen Schwerpunkten zu arbeiten) 	<ul style="list-style-type: none"> ▪ Innere Differenzierung des Unterrichts (binnendifferenzierende Aufgabenstellungen, unterschiedliche Lernmaterialien werden eingesetzt, für SuS mit Lernschwierigkeiten stehen zusätzliche Lernhilfen bereit, Möglichkeit für leistungsstarke SuS Routineaufgaben nicht zu erledigen und nach eigenen Schwerpunkten zu arbeiten)

Qualitätsbereiche Quellen der Kompetenzen und Indikatoren	Erklärung, Operationalisierung in Teilkompetenzen	Indikatoren (in einer Unterrichtsstunde) [teilweise sind einzelne Indikatoren verschiedenen Qualitätsbereichen zugeordnet]
Konsolidierung/Sicherung <ul style="list-style-type: none"> ▪ intelligentes Üben (Meyer) ▪ Konsolidierung/Sicherung (Helmke) 	<ul style="list-style-type: none"> ▪ in den Unterricht werden intelligente Übungsphasen eingebaut (lernstandsadäquate Entwicklung von Lernstrategien, Rhythmisierung der Übung) 	<ul style="list-style-type: none"> ▪ in den Übungsphasen werden Hilfestellungen bereitgestellt ▪ Übungen leiten zum Transfer, der Qualitätssteigerung oder der Automatisierung von Wissen gezielt an ▪ eine Selbstkontrolle des Lernerfolgs ist möglich ▪ auf Phasen der Stoffvermittlung folgen Phasen der kommunikativen Anwendung
transparente Leistungserwartungen <ul style="list-style-type: none"> ▪ transparente Leistungserwartungen (Meyer) ▪ lernförderliches Klima (Helmke) 	<ul style="list-style-type: none"> ▪ Lernangebote bereitzustellen die den Ansprüchen der Richtlinien, den Bildungsstandards und den Fähigkeiten der SuS entsprechen ▪ Transparenz über Lernziele ▪ Rückmeldungen über den Lernfortschritt geben 	<ul style="list-style-type: none"> ▪ die Lernziele der Unterrichtsstunde/-einheit sind transparent ▪ die Grundlagen der Leistungsbewertung werden offen gelegt ▪ deutliches Lehrer-Feedback, ob Beiträge „richtig“ waren ▪ SuS-Feedback zu Beiträgen ▪ es findet regelmäßig eine Rückmeldung über den Leistungsstand statt
vorbereitete Umgebung <ul style="list-style-type: none"> ▪ (Meyer) 	<ul style="list-style-type: none"> ▪ ordentliche, mit sinnvollem Lernmaterial ausgestattete Räume, zu denen Lehrer und SuS ein Gefühl von „Ownership“ haben ▪ Belüftung, Beleuchtung und Akustik sind stimmig ▪ ausreichend Platz für verschiedene Inszenierungstechniken des Unterrichts und den Bewegungsdrang der SuS 	<ul style="list-style-type: none"> ▪ ökologisch und ergonomisch sinnvolle Raumausstattung ▪ sauberer und gepflegter Eindruck der Räume ▪ pfleglicher Umgang mit den Materialien ▪ Es gibt „Funktionsecken“ in den Räumen
Aktivierung <ul style="list-style-type: none"> ▪ (Helmke) 	<ul style="list-style-type: none"> ▪ kognitive Aktivierung im Sinne der Selbststeuerung des Lernens, es erfolgt eine „tiefe“ Verarbeitung des Stoffes (Lernstrategien) ▪ soziale Aktivierung (kooperative Arbeitsformen) ▪ aktive Teilhabe 	<ul style="list-style-type: none"> ▪ sinnvoller Einsatz kooperativer Arbeitsformen ▪ Einsatz von Hilfestellungen zum Erwerb von Lernstrategien

Qualitätsbereiche Quellen der Kompetenzen und Indikatoren	Erklärung, Operationalisierung in Teilkompetenzen	Indikatoren (in einer Unterrichtsstunde) [teilweise sind einzelne Indikatoren verschiedenen Qualitätsbereichen zugeordnet]
Motivierung ■ (Helmke)	<ul style="list-style-type: none"> ■ extrinsische und intrinsische Motivationsgründe bewegen die SuS zum Mitarbeiten ■ anspruchsvolle Lehrererwartungen 	<ul style="list-style-type: none"> ■ Motivierende, überraschende Einstiege ■ „fantasiereiche“ Unterrichtsgestaltung ■ Herstellung von Situationen, in denen SuS Selbstwirksamkeit erleben ■ leistungsbezogene, transparente Bewertung ■ klare, entwicklungsbezogene Perspektive
Kompetenzorientierung ■ (Helmke)	<ul style="list-style-type: none"> ■ die Gestaltung des Unterrichts orientiert sich an den am Ende des Lernprozesses zu erreichenden Standards 	<ul style="list-style-type: none"> ■ klare Kompetenzorientierung in den Aufgabenstellungen ■ klar erkennbare Lernprogression über die AFBs
Lernzuwachs ■ (QA NRW)	<ul style="list-style-type: none"> ■ Die SuS erreichen einen intendierten Lernzuwachs 	<ul style="list-style-type: none"> ■ SuS können die zum Beginn der Stunde aufgeworfene (Leit-)Fragen beantworten ■ SuS formulieren selbstständig Lösungen für Aufgabenstellungen und stellen diese vor ■ SuS Antworten auf Transferfragen

pre-Alpha-Version des Bewertungskonzeptes Beobachtungsbogen ‚A‘ mit Kompetenzskala (pre-Alpha-Version)

Qualitätsbereich	Indikatoren	Indikator relevant	nicht angezeigt	
Klarheit und Strukturiertheit	1) Es ist eine transparente und stimmige Phasierung des Unterrichts erkennbar, einzelne Unterrichtsschritte werden klar markiert			
	2) Die SuS verstehen die Aufgabenstellungen und können erläutern was sie gerade tun			
	3) Unterrichtsziele, Inhalte und eingesetzte Methoden stehen in einem sinnvollen Verhältnis zueinander			
	4) Die Inhalte der Stunde werden didaktisch sinnvoll inszeniert und knüpfen an zentrale fachdidaktische Prinzipien an			
	5) Die Äußerungen des Lehrers zeichnen sich durch akustische Verständlichkeit und sprachliche Klarheit aus			

	Kompetenzstufen		
	niedrig	mittel	hoch
	Eine Phasierung des Unterrichts ist in Ansätzen erkennbar, der Wechsel zwischen den verschiedenen Phasen ist nicht flüssig, es wird nur ansatzweise Transparenz über den Unterrichtschwerpunkt hergestellt	Eine Phasierung des Unterrichts ist deutlich erkennbar, der Wechsel zwischen den verschiedenen Phasen ist in Ansätzen sinnvoll angeleitet, einzelne Unterrichtsschritte deuten auf den Schwerpunkt des Unterrichts hin	Eine Phasierung des Unterrichts ist deutlich erkennbar, der Wechsel zwischen den verschiedenen Phasen wird sinnvoll angeleitet, bzw. geschieht durch im Vorfeld angelegte Lernarrangements, einzelne Unterrichtsschritte sorgen für Transparenz hinsichtlich des Stundenschwerpunktes
	Die Aufgabenstellungen sind nicht klar, die SuS müssen mehrfach neu instruiert werden, bevor sie mit dem Arbeiten beginnen	Die Aufgabenstellung sind i.W. klar, es erfolgen lediglich kleinere, zusätzliche Erklärungen	Die Aufgabenstellungen sind klar formuliert, die SuS können erläutern, welche Aufgabe sie gerade lösen und welche Rolle sie im Lernprozess einnimmt
	Die eingesetzten Methoden und Inhalte leiten nicht systematisch zum Unterrichtsziel hin, stattdessen ist die Methoden-, Ziel- und Inhaltswahl teils widersprüchlich	Die eingesetzten Methoden und Inhalte leiten zum Unterrichtsziel hin, wobei die Methoden keinen systematischen Beitrag zum Verständnis der Inhalte leisten, sondern relativ unverbunden neben den Inhalten stehen	Die eingesetzten Methoden und Inhalte leiten systematisch auf das zu erreichende Unterrichtsziel hin, die eingesetzten Methoden helfen systematisch, die Inhalte zu verstehen
	fachdidaktische Prinzipien strukturieren die Stunde kaum, es ist kein inhaltlicher Spannungsbogen feststellbar	fachdidaktische Prinzipien sind in der Strukturierung erkennbar und zeigen in Ansätzen einen inhaltlichen Spannungsbogen auf	fachdidaktische Prinzipien strukturieren die Stunde sinnvoll und führen zu einem klar ausgeprägten inhaltlichen Spannungsbogen
	Die Lehrersprache ist akustisch teils schwer zu verstehen, Unsicherheits- und Vagheitsausdrücke erzeugen missverständliche Botschaften, Abschweifungen machen es schwer, den inhaltlichen Kern des Gedankenganges zu erkennen	Die Lehrersprache ist akustisch meistens gut verständlich, Unsicherheits- und Vagheitsausdrücke verwässern die Aussagen selten, der inhaltliche Kern der Ausführungen ist trotz einiger Abschweifungen auszumachen	Die Lehrersprache ist akustisch verständlich, Unsicherheits- und Vagheitsausdrücke werden kaum benutzt, es finden sich stringente Ausführungen ohne Abschweifungen

Qualitätsbereich	Indikatoren	Indikator relevant	nicht angezeigt	
hoher Anteil echter Lernzeit	6) Die meisten SuS sind nicht gelangweilt oder überfordert			
	7) Die Stunde weist einen hohen Anteil ‚echter Lernzeit‘ auf			
Kompetenzorientierung	8) Die Aufgaben zeigen eine klare Berücksichtigung der Kompetenzorientierung			
	9) Im Kontext der Unterrichtsstunde/-reihe ist eine klare Progression über die Anforderungsbereiche zu erkennen			
Aktivierung	10) Kooperative oder selbstgesteuerte Arbeitsformen oder werden gewinnbringend eingesetzt			

	Kompetenzstufen		
	niedrig	mittel	hoch
	Ein erheblicher Anteil der SuS konzentriert sich nicht auf die Aufgabenstellungen, weil sie überfordert oder gelangweilt sind	Die meisten SuS folgen den Arbeitsaufträgen, lediglich wenige SuS sind überfordert oder gelangweilt	Nahezu alle SuS arbeiten konzentriert und zielgerichtet an den Aufgabenstellungen
	Ein erheblicher Teil der Stunde wird für Planungskorrekturen, nachträgliche Erklärungen, Disziplinierungen o.ä. des Lehrer verwendet	Planungskorrekturen, nachträgliche Klärungen von Aufgabenstellungen, Inhalten o.ä. nehmen nur einen geringen Teil der Stunde in Anspruch	Ein Großteil der Stunde wird für lernwirksame Kommunikation oder Arbeitsprozesse genutzt. Die Kommunikation ist zielführend und inhaltlich reichhaltig, Planungskorrekturen, nachträgliche Klärungen, etc. sind kaum notwendig
	Die Aufgaben weisen nur einen sehr losen Bezug zu den Kompetenzbereichen der KLP auf, eine klare Progression der Aufgabenstellungen über die Kompetenzbereiche lässt sich kaum festmachen	Die Aufgaben zeigen Bezüge zu den Kompetenzbereichen der KLP auf, eine klare Progression der Aufgabenstellungen über die Kompetenzbereiche ist in Ansätzen zu erkennen	Die Aufgaben zeigen klare Bezüge zu den Kompetenzbereichen der KLP auf, es zeigt sich eine deutliche Progression der Aufgabenstellungen über die Kompetenzbereiche
	Die einzelnen Stunden der Unterrichtsreihe zeigen nur losen Bezug zueinander, ein roter Faden, dessen Ziel ein Kompetenzerwerb in allen 3 AFBs ist, lässt sich kaum erkennen	Die einzelnen Unterrichtsstunden der Unterrichtsreihe bauen mehr oder weniger logisch aufeinander auf und leiten zum Kompetenzerwerb in verschiedenen AFB an	Die einzelnen Unterrichtsstunden der Unterrichtsreihe bauen logisch aufeinander auf und sorgen durch eine klare Progression über die AFBs für einen systematischen Kompetenzaufbau
	Kooperative oder selbstgesteuerte Arbeitsformen werden in den Unterricht eingebaut, aber sie führen kaum zur aktiven Einbindung der SuS in den Lernprozess	Kooperative oder selbstgesteuerte Arbeitsformen werden in den Unterricht eingebaut, sie führen im eingeschränkten Rahmen zur Aktivierung der SuS	Kooperative oder selbstgesteuerte Arbeitsformen werden in den Unterricht eingebaut, sie versetzen die SuS in eine aktive Rolle im Lernprozess

Qualitätsbereich	Indikatoren	Indikator relevant	nicht angezeigt	
Umgang mit Heterogenität	11) Eine Binnendifferenzierung reagiert sowohl auf die Bedürfnisse schwächerer als auch leistungsstärkerer SuS			
lernförderliches Klima	12) Der Lehrer reagiert angemessen auf Unterrichtsstörungen			
	13) Ein großer Teil der Lerngruppe bringt sich aktiv in den Unterricht ein			
sinnstiftendes Kommunizieren	14) Die SuS werden dazu angeregt die Unterrichtsinhalte in ihren eigenen Sinnhorizont einzuordnen			
	15) Das Gelernte wird mit anderen Unterrichtsinhalten verknüpft			
	16) Die SuS geben sich untereinander Rückmeldung über Inhalte oder Lernfortschritte			

	Kompetenzstufen		
	niedrig	mittel	hoch
	Lernhilfen stehen nicht bereit, lediglich spontane Reaktionen des Lehrers zeigen eine Beachtung unterschiedlicher Leistungsstärke	Es finden sich Ansätze binnendifferenzierenden Arbeitens, z.B. durch Zusatzaufgaben, etc., wobei diese Differenzierung eher quantitativ angelegt ist	Es findet eine qualitative Differenzierung des Unterrichts statt, die sich z.B. in unterschiedlichen Schwierigkeitsgraden von Aufgabenstellungen, Lernhilfen für schwächere SuS oder Transferaufgaben für stärkere SuS zeigt
	Unterrichtsstörungen führen zu Brüchen im Unterrichtsablauf, der Fokus der Kommunikation bewegt sich stark vom Unterrichtsgeschehen weg	Unterrichtsstörungen beeinträchtigen den eigentlichen Unterrichtsablauf nur mäßig, aber der Fokus der Kommunikation bewegt sich zeitweilig vom Unterrichtsgeschehen weg	Unterrichtsstörungen werden schnell geklärt und beruhigt, ohne dass sich der Fokus der Kommunikation vom eigentlichen Unterrichtsthema wegbewegt
	Einige wenige SuS bringen sich in den Unterricht ein	Ein wesentlicher Teil der Lerngruppe bringt sich in den Unterricht ein, wobei ein nicht unerheblicher Teil der SuS passiv bleibt	Der Großteil der Lerngruppe bringt sich in den Unterricht ein, nur einzelne SuS ziehen sich aus den Arbeits- und Kommunikationsprozessen heraus
	Der Unterricht ist abstrakt und konkret kaum fassbar, die Verbindung von allgemeinen zu lernenden Inhalten und Beispielen bleibt unklar	Der Unterricht ermöglicht es den SuS, z.B. durch konkrete Beispiele, im Ansatz die lebensweltliche Bedeutsamkeit der im Unterricht gewonnenen Erkenntnisse zu erfassen	Der Unterricht bindet allgemein gewonnene Erkenntnisse systematisch und geplant an konkrete Beispiele zurück und zeigt den „Bildungsgehalt“ der Inhalte
	Das Gelernte bleibt isoliert im Stundenkontext stehen	Die Verknüpfung des Gelernten erfolgt spontan, z.B. durch Lehrerfragen oder Impulse	Die Verknüpfung des Gelernten erfolgt planvoll und systematisch in eigenen Arbeitsphasen, z.B. durch vorbereitete Aufgaben und Materialien
	SuS-Feedback ist abseits ungeplanter und spontaner SuS-Kommunikation nicht zu beobachten	SuS-Feedback ist in einzelnen Phasen des Unterrichts vorgesehen, wird aber unstrukturiert und ohne klare Vorgaben eingesetzt	SuS-Feedback ist in einzelnen Phasen des Unterrichts vorgesehen und wird durch systematische Anleitungen und transparente Leistungserwartungen unterstützt

Qualitätsbereich	Indikatoren	Indikator relevant	nicht angezeigt	
transparente Leistungserwartungen	17) Der Lehrer gibt deutliches Feedback, ob die Beiträge „richtig“ waren			
	18) Der Lehrer unterstützt die SuS Fehler selbstständig zu beheben			
	19) Dem Lehrer gelingt es Klarheit über die Lernziele herzustellen			
vorbereitete Umgebung	20) Die Arbeitsmittel und Medien werden zielführend eingesetzt und sorgen für Verständnis			
	21) Die Sitzordnung und die Raumgestaltung stehen in einem sinnvollen Verhältnis zu Zielen, Inhalten und Methoden des Unterrichts			

	Kompetenzstufen		
	niedrig	mittel	hoch
	Der Lehrer gibt Feedback zu Schülerbeiträgen, wobei die dem Feedback zu Grunde gelegten Leistungserwartungen unklar bleiben	Der Lehrer gibt Feedback zu Schülerbeiträgen, er muss die der Bewertung zu Grunde liegenden Leistungserwartungen allerdings zuerst erläutern	Der Lehrer gibt ein gezieltes und stimmiges Feedback zu Schülerbeiträgen, er bezieht sich auf transparente und den SuS bekannte Leistungserwartungen
	Der Lehrer weist SuS auf Fehler hin und korrigiert sie für die SuS	Der Lehrer weist SuS auf Fehler hin und hilft ihnen durch Fragen und Hinweise den Fehler zu beheben	Der Lehrer weist SuS auf Fehler hin und gibt gezielte Hilfestellungen oder verweist auf Lernhilfen oder festgelegte Standards, damit die SuS den Fehler selbstständig beheben können
	Der Lehrer nennt die Lernziele der Unterrichtsstunde/-einheit	Der Lehrer leitet die SuS fragend-entwickelnd dazu an, die Lernziele der Unterrichtsstunde/-einheit selbstständig zu benennen	Vorbereitete Materialien oder Strukturierungen führen die SuS dahin, die Lernziele der Unterrichtsstunde selbstständig zu benennen
	Die eingesetzten Medien und Arbeitsmittel führen zu einer i.W. zielgerichteten, aber wenig motivierenden und nicht immer schülergerechten Art der Auseinandersetzung	Die eingesetzten Medien und Arbeitsmittel sind i.W. schülergerecht, sie zeigen den zu erarbeitenden Erkenntnisfortschritt auf, wobei dieser mitunter zu trivial auf der Hand liegt oder zu komplex für das Verständnis der SuS ist	Die eingesetzten Medien und Arbeitsmittel sind motivierend und schülergerecht, sie zeigen den zu erarbeitenden Erkenntnisfortschritt klar, aber nicht trivial auf
	Die Sitzordnung und Raumgestaltung sind nicht oder wenig durchdacht auf die Arbeits- und Kommunikationsprozesse abgestimmt, sie unterstützen den reibungslosen Ablauf der Unterrichtsphasen nicht	Sitzordnung und Raumgestaltung sind i.W. auf die in der Stunde intendierten Arbeits- und Kommunikationsprozesse abgestimmt, aber unterstützen den reibungslosen Ablauf der Unterrichtsphasen nur bedingt	Sitzordnung und Raumgestaltung sind sinnvoll auf die in der Stunde intendierten Arbeits- und Kommunikationsprozesse abgestimmt, sie unterstützen den reibungslosen Ablauf der Unterrichtsphasen

Qualitätsbereich	Indikatoren	Indikator relevant	nicht angezeigt	
Lernzuwachs	22) Am Ende der Stunde steht ein klar zu benennender Erkenntnisfortschritt der verbindlich fixiert wird oder die SuS formulieren selbstständig Lösungen für Aufgabenstellungen und stellen diese vor			
	23) Die SuS können den erzielten Erkenntnisfortschritt für weitere Lernprozesse nutzen			

	Kompetenzstufen		
	niedrig	mittel	hoch
	Der Erkenntnisfortschritt wird schriftlich fixiert, wobei sich diese Phase durch eine stark lenkende Rolle des Lehrers auszeichnet	Der Erkenntnisfortschritt wird von SuS selbstständig in gestalteten Produkten oder schriftlichen Lösungen vorgestellt, wobei der Lehrer Korrekturen vornehmen muss, ggf. werden alternative Lösungen gezeigt oder der Lösungsweg verdeutlicht	Der Erkenntnisfortschritt wird von SuS selbstständig in gestalteten Produkten oder schriftlichen Lösungen vorgestellt, ggf. werden alternative Lösungen gezeigt oder der Lösungsweg verdeutlicht
	Der in der Stunde erzielte Erkenntnisfortschritt bleibt i.W. isoliert stehen und wird nicht für weitere Lernprozesse genutzt	Der in der Stunde erzielte Erkenntnisfortschritt wird zur Beantwortung von weitergehenden Fragen genutzt, ohne dass die SuS diese in einer eigenen systematisch angeleiteten Arbeitsphase bearbeiten	Der in der Stunde erzielte Erkenntnisfortschritt wird für weiterführende, systematisch angelegte Lernarrangements genutzt (gestalterische oder kommunikative Aufgaben, Transferaufgaben, etc.)

Beobachtungsbogen 'B' mit Ratingskala (pre-Alpha-Version)

Qualitätsbereich	Indikatoren	Indikator relevant?	Unterrichtsmerkmale 1 = Kompetenz kaum ausgeprägt/4 Kompetenz stark ausgeprägt			
Klarheit und Strukturiertheit	1) Es ist eine transparente und stimmige Phasierung des Unterrichts erkennbar, einzelne Unterrichtsschritte werden klar markiert					
	2) Die SuS verstehen die Aufgabenstellungen und können erläutern was sie gerade tun					
	3) Unterrichtsziele, Inhalte und eingesetzte Methoden stehen in einem sinnvollen Verhältnis zueinander					
	4) Die Inhalte der Stunde werden didaktisch sinnvoll inszeniert und knüpfen an zentrale fachdidaktische Prinzipien an					
	5) Die Äußerungen des Lehrers zeichnen sich durch akustische Verständlichkeit und sprachliche Klarheit aus.					
Hoher Anteil echter Lernzeit	6) Die meisten SuS sind nicht gelangweilt oder überfordert					
	7) Die Stunde weist einen hohen Anteil ‚echter Lernzeit‘ auf					
Kompetenzorientierung	8) Die Aufgaben zeigen eine klare Berücksichtigung der Kompetenzorientierung					
	9) Im Kontext der Unterrichtsstunde/-reihe ist eine klare Progression über die Anforderungsbereiche zu erkennen					
Aktivierung	10) Kooperative oder selbstgesteuerte Arbeitsformen oder werden gewinnbringend eingesetzt					
Umgang mit Heterogenität	11) Eine Binnendifferenzierung reagiert sowohl auf die Bedürfnisse schwächerer als auch leistungsstärkerer SuS					

Qualitätsbereich	Indikatoren	Indikator relevant?	Unterrichtsmerkmale 1 = Kompetenz kaum ausgeprägt/4 Kompetenz stark ausgeprägt			
lernförderliches Klima	12) Der Lehrer reagiert angemessen auf Unterrichtsstörungen					
	13) Ein großer Teil der Lerngruppe bringt sich aktiv in den Unterricht ein					
sinnstiftendes Kommunizieren	14) Die SuS werden dazu angeregt die Unterrichtsinhalte in ihren eigenen Sinnhorizont einzuordnen					
	15) Das Gelernte wird mit anderen Unterrichtsinhalten verknüpft					
	16) Die SuS geben sich untereinander Rückmeldung über Inhalte oder Lernfortschritte					
transparente Leistungserwartungen	17) Der Lehrer gibt deutliches Feedback, ob die Beiträge „richtig“ waren					
	18) Der Lehrer unterstützt die SuS Fehler selbstständig zu beheben					
	19) Dem Lehrer gelingt es Klarheit über die Lernziele herzustellen					
vorbereitete Umgebung	20) Die Arbeitsmittel und Medien werden zielführend eingesetzt und sorgen für Verständnis					
	21) Die Sitzordnung und die Raumgestaltung stehen in einem sinnvollen Verhältnis zu Zielen, Inhalten und Methoden des Unterrichts					
Lernzuwachs	22) Am Ende der Stunde steht ein klar zu benennender Erkenntnisfortschritt der verbindlich fixiert wird/Die SuS formulieren selbstständig Lösungen für Aufgabenstellungen und stellen diese vor					
	23) Die SuS können den erzielten Erkenntnisfortschritt für weitere Lernprozesse nutzen					

Dokumentationsbogen für die Nachbesprechung (pre-Alpha-Version)

ReferendarIn: XY

Grundlegende Einordnung des Unterrichts

Datum des UBs/ Signatur Ausbilder	eingesetzte Sozialformen (kursiv = dominante Sozialform)	Eingesetzte Medien	„Unterrichtsdramatik auf den Punkt gebracht“
04.11.2011/ Tel	Partnerarbeit Unterrichtsgespräch	Computer (WebQuest)	computergestützter erarbeitender Unterricht

Kompetenz als Expertise

erkennt Probleme
im Stundenablauf nicht

erkennt Probleme
ansatzweise

erkennt Probleme
vollständig



Notizen:

beschreibt additiv
und chronologisch

analysiert
kriteriengeleitet



Notizen:

Kompetenz als Repertoire:

entwickelt keine
sinnvollen Alternativen

entwickelt mit Hilfe
sinnvolle Alternativen

entwickelt selbstständig
sinnvolle Alternativen



Notizen:

Angesprochene Merkmalsbereiche

- Klarheit und Strukturiertheit
- hoher Anteil echter Lernzeit
- Umgang mit Heterogenität
- Kompetenzorientierung
- Aktivierung
- lernförderliches Klima
- sinnstiftendes Kommunizieren
- transparente Leistungserwartungen
- vorbereitete Umgebung
- Lernzuwachs

Kompetenzbereich	Indikatoren/Aspekte	Thematisierung

Kompetenzbereiche und Kompetenzausprägungen im Zeitverlauf Beobachtungsbogen

Kompetenzbereich	Indikatoren	Indikator relevant	nicht angezeigt	
Klarheit und Strukturiertheit	1) Es ist eine transparente und stimmige Phasierung des Unterrichts erkennbar, einzelne Unterrichtsschritte werden klar markiert			
	Beobachtet/Signatur, Datum			
	2) Die SuS verstehen die Aufgabenstellungen und können erläutern was sie gerade tun			
	Beobachtet/Signatur, Datum			
	3) Unterrichtsziele, Inhalte und eingesetzte Methoden stehen in einem sinnvollen Verhältnis zueinander			
	Beobachtet/Signatur, Datum			
	4) Die Inhalte der Stunde werden didaktisch sinnvoll inszeniert und knüpfen an zentrale fachdidaktische Prinzipien an			
	Beobachtet/Signatur, Datum			
	5) Die Äußerungen des Lehrers zeichnen sich durch akustische Verständlichkeit und sprachliche Klarheit aus			
	Beobachtet/Signatur, Datum			

	Kompetenzstufen		
	niedrig	mittel	hoch
	Eine Phasierung des Unterrichts ist in Ansätzen erkennbar, der Wechsel zwischen den verschiedenen Phasen ist nicht flüssig, es wird nur ansatzweise Transparenz über den Unterrichtschwerpunkt hergestellt	Eine Phasierung des Unterrichts ist deutlich erkennbar, der Wechsel zwischen den verschiedenen Phasen ist in Ansätzen sinnvoll angeleitet, einzelne Unterrichtsschritte deuten auf den Schwerpunkt des Unterrichts hin	Eine Phasierung des Unterrichts ist deutlich erkennbar, der Wechsel zwischen den verschiedenen Phasen wird sinnvoll angeleitet, bzw. geschieht durch im Vorfeld angelegte Lernarrangements, einzelne Unterrichtsschritte sorgen für Transparenz hinsichtlich des Stundenschwerpunktes
Tel/23.02.2012			
	Die Aufgabenstellungen sind nicht klar, die SuS müssen mehrfach neu instruiert werden, bevor sie mit dem Arbeiten beginnen	Die Aufgabenstellungen sind i.W. klar, es erfolgen lediglich kleinere zusätzliche Erklärungen	Die Aufgabenstellungen sind klar formuliert, die SuS können erläutern, welche Aufgabe sie gerade lösen und welche Rolle sie im Lernprozess einnimmt
	Die eingesetzten Methoden und Inhalte leiten nicht systematisch zum Unterrichtsziel hin, stattdessen ist die Methoden-, Ziel- und Inhaltswahl teils widersprüchlich	Die eingesetzten Methoden und Inhalte leiten zum Unterrichtsziel hin, wobei die Methoden keinen systematischen Beitrag zum Verständnis der Inhalte leisten, sondern relativ unverbunden neben den Inhalten stehen	Die eingesetzten Methoden und Inhalte leiten systematisch auf das zu erreichende Unterrichtsziel hin, die eingesetzten Methoden helfen systematisch, die Inhalte zu verstehen
	fachdidaktische Prinzipien strukturieren die Stunde kaum, es ist kein inhaltlicher Spannungsbogen feststellbar	fachdidaktische Prinzipien sind in der Strukturierung erkennbar und zeigen in Ansätzen einen inhaltlichen Spannungsbogen auf	fachdidaktische Prinzipien strukturieren die Stunde sinnvoll und führen zu einem klar ausgeprägten inhaltlichen Spannungsbogen
	Die Lehrersprache ist akustisch teils schwer zu verstehen, Unsicherheits- und Vagheitsausdrücke erzeugen missverständliche Botschaften, Abschweifungen machen es schwer, den inhaltlichen Kern des Gedankenganges zu erkennen	Die Lehrersprache ist akustisch meistens gut verständlich, Unsicherheits- und Vagheitsausdrücke verwässern die Aussagen selten, der inhaltliche Kern der Ausführungen ist trotz einiger Abschweifungen auszumachen	Die Lehrersprache ist akustisch verständlich, Unsicherheits- und Vagheitsausdrücke werden kaum benutzt, es finden sich stringente Ausführungen ohne Abschweifungen

Kompe-tenz- bereich	Indikatoren	Indikator relevant	nicht angezeigt	
hoher Anteil echter Lernzeit	6) Die meisten SuS sind nicht gelangweilt oder überfordert			
	Beobachtet/Signatur, Datum			
	7) Die Stunde weist einen hohen Anteil ‚echter Lernzeit‘ auf			
	Beobachtet/Signatur, Datum			
Kompetenzorientierung	8) Die Aufgaben zeigen eine klare Berücksichtigung der Kompetenzorientierung			
	Beobachtet/Signatur, Datum			
	9) Im Kontext der Unterrichtsstunde/-reihe ist eine klare Progression über die Anforderungsbereiche zu erkennen			
Beobachtet/Signatur, Datum				
Aktivierung	10) Kooperative oder selbstgesteuerte Arbeitsformen oder werden gewinnbringend eingesetzt			
	Beobachtet/Signatur, Datum			

	Kompetenzstufen		
	niedrig	mittel	hoch
	Ein erheblicher Anteil der SuS konzentriert sich nicht auf die Aufgabenstellungen, weil sie überfordert oder gelangweilt sind	Die meisten SuS folgen den Arbeitsaufträgen, lediglich wenige SuS sind überfordert oder gelangweilt	Nahezu alle SuS arbeiten konzentriert und zielgerichtet an den Aufgabenstellungen
	Ein erheblicher Teil der Stunde wird für Planungskorrekturen, nachträgliche Erklärungen, Disziplinierungen o.ä. des Lehrer verwendet	Planungskorrekturen, nachträgliche Klärungen von Aufgabenstellungen, Inhalten o.ä. nehmen nur einen geringen Teil der Stunde in Anspruch	Ein Großteil der Stunde wird für lernwirksame Kommunikation oder Arbeitsprozesse genutzt. Die Kommunikation ist zielführend und inhaltlich reichhaltig, Planungskorrekturen, nachträgliche Klärungen, etc. sind kaum notwendig
	Die Aufgaben weisen nur einen sehr losen Bezug zu den Kompetenzbereichen der KLP auf, eine klare Progression der Aufgabenstellungen über die Kompetenzbereiche lässt sich kaum festmachen	Die Aufgaben zeigen Bezüge zu den Kompetenzbereichen der KLP auf, eine klare Progression der Aufgabenstellungen über die Kompetenzbereiche ist in Ansätzen zu erkennen	Die Aufgaben zeigen klare Bezüge zu den Kompetenzbereichen der KLP auf, es zeigt sich eine deutliche Progression der Aufgabenstellungen über die Kompetenzbereiche
	Die einzelnen Stunden der Unterrichtsreihe zeigen nur losen Bezug zueinander, ein roter Faden dessen Ziel ein, Kompetenzerwerb in allen 3 AFBs ist, lässt sich kaum erkennen	Die einzelnen Unterrichtsstunden der Unterrichtsreihe bauen mehr oder weniger logisch aufeinander auf und leiten zum Kompetenzerwerb in verschiedenen AFB an	Die einzelnen Unterrichtsstunden der Unterrichtsreihe bauen logisch aufeinander auf und sorgen durch eine klare Progression über die AFBs für einen systematischen Kompetenzaufbau
	Kooperative oder selbstgesteuerte Arbeitsformen werden in den Unterricht eingebaut, aber sie führen kaum zur aktiven Einbindung der SuS in den Lernprozess	Kooperative oder selbstgesteuerte Arbeitsformen werden in den Unterricht eingebaut, sie führen im eingeschränkten Rahmen zur Aktivierung der SuS	Kooperative oder selbstgesteuerte Arbeitsformen werden in den Unterricht eingebaut, sie versetzen die SuS in eine aktive Rolle im Lernprozess

Kompe-tenz- bereich	Indikatoren	Indikator relevant	nicht angezeigt	
Umgang mit Heterogenität	11) Eine Binnendifferenzierung reagiert sowohl auf die Bedürfnisse schwächerer als auch leistungsstärkerer SuS			
	Beobachtet/Signatur, Datum			
lernförderliches Klima	12) Der Lehrer reagiert angemessen auf Unterrichtsstörungen			
	Beobachtet/Signatur, Datum			
	13) Ein großer Teil der Lerngruppe bringt sich aktiv in den Unterricht ein			
sinnstiftendes Kommunizieren	14) Die SuS werden dazu angeregt die Unterrichtsinhalte in ihren eigenen Sinnhorizont einzuordnen			
	Beobachtet/Signatur, Datum			
	15) Das Gelernte wird mit anderen Unterrichtsinhalten verknüpft			
	Beobachtet/Signatur, Datum			
	16) Die SuS geben sich untereinander Rückmeldung über Inhalte oder Lernfortschritte			
	Beobachtet/Signatur, Datum			

				Kompetenzstufen		
				niedrig	mittel	hoch
	Lernhilfen stehen nicht bereit, lediglich spontane Reaktionen des Lehrers zeigen eine Beachtung unterschiedlicher Leistungsstärke	Es finden sich Ansätze binnendifferenzierenden Arbeitens, z.B. durch Zusatzaufgaben, etc., wobei diese Differenzierung eher quantitativ angelegt ist	Es findet eine qualitative Differenzierung des Unterrichts statt, die sich z.B. in unterschiedlichen Schwierigkeitsgraden von Aufgabenstellungen, Lernhilfen für schwächere SuS oder Transferaufgaben für stärkere SuS zeigt			
	Unterrichtsstörungen führen zu Brüchen im Unterrichtsablauf, der Fokus der Kommunikation bewegt sich stark vom Unterrichtsgeschehen weg	Unterrichtsstörungen beeinträchtigen den eigentlichen Unterrichtsablauf nur mäßig, aber der Fokus der Kommunikation bewegt sich zeitweilig vom Unterrichtsgeschehen weg	Unterrichtsstörungen werden schnell geklärt und beruhigt, ohne dass sich der Fokus der Kommunikation vom eigentlichen Unterrichtsthema wegbewegt			
	Einige wenige SuS bringen sich in den Unterricht ein	Ein wesentlicher Teil der Lerngruppe bringt sich in den Unterricht ein, wobei ein nicht unerheblicher Teil der SuS passiv bleibt	Der Großteil der Lerngruppe bringt sich in den Unterricht ein, nur einzelne SuS ziehen sich aus den Arbeits- und Kommunikationsprozessen heraus			
	Der Unterricht ist abstrakt und konkret kaum fassbar, die Verbindung von allgemein zu lernenden Inhalten und Beispielen bleibt unklar	Der Unterricht ermöglicht es den SuS, z.B. durch konkrete Beispiele, im Ansatz die lebensweltliche Bedeutsamkeit der im Unterricht gewonnenen Erkenntnisse zu erfassen	Der Unterricht bindet allgemein gewonnene Erkenntnisse systematisch und geplant an konkrete Beispiele zurück und zeigt den „Bildungsgehalt“ der Inhalte			
	Das Gelernte bleibt isoliert im Stundenkontext stehen	Die Verknüpfung des Gelernten erfolgt spontan, z.B. durch Lehrerfragen oder Impulse	Die Verknüpfung des Gelernten erfolgt planvoll und systematisch in eigenen Arbeitsphasen, z.B. durch vorbereitete Aufgaben und Materialien			
	SuS-Feedback ist abseits ungeplanter und spontaner SuS-Kommunikation nicht zu beobachten	SuS-Feedback ist in einzelnen Phasen des Unterrichts vorgesehen, wird aber unstrukturiert und ohne klare Vorgaben eingesetzt	SuS-Feedback ist in einzelnen Phasen des Unterrichts vorgesehen und wird durch systematische Anleitungen und transparente Leistungserwartungen unterstützt			

Kompe-tenz- bereich	Indikatoren	Indikator relevant	nicht angezeigt	
transparente Leistungserwartungen	17) Der Lehrer gibt deutliches Feedback, ob die Beiträge „richtig“ waren			
	Beobachtet/Signatur, Datum			
	18) Der Lehrer unterstützt die SuS Fehler selbstständig zu beheben			
	Beobachtet/Signatur, Datum			
vorbereitete Umgebung	20) Die Arbeitsmittel und Medien werden zielführend eingesetzt und sorgen für Verständnis			
	Beobachtet/Signatur, Datum			
	21) Die Sitzordnung und die Raumgestaltung stehen in einem sinnvollen Verhältnis zu Zielen, Inhalten und Methoden des Unterrichts			
	Beobachtet/Signatur, Datum			

Kompetenzstufen		
niedrig	mittel	hoch
Der Lehrer gibt Feedback zu Schülerbeiträgen, wobei die dem Feedback zu Grunde gelegten Leistungserwartungen unklar bleiben	Der Lehrer gibt Feedback zu Schülerbeiträgen, er muss die der Bewertung zu Grunde liegenden Leistungserwartungen allerdings zuerst erläutern	Der Lehrer gibt ein gezieltes und stimmiges Feedback zu Schülerbeiträgen, er bezieht sich auf transparente und den SuS bekannte Leistungserwartungen
Der Lehrer weist SuS auf Fehler hin und korrigiert sie für die SuS	Der Lehrer weist SuS auf Fehler hin und hilft ihnen durch Fragen und Hinweise den Fehler zu beheben	Der Lehrer weist SuS auf Fehler hin und gibt gezielte Hilfestellungen oder verweist auf Lernhilfen oder festgelegte Standards, damit die SuS den Fehler selbstständig beheben können
Der Lehrer nennt die Lernziele der Unterrichtsstunde/-einheit	Der Lehrer leitet die SuS fragend-entwickelnd dazu an, die Lernziele der Unterrichtsstunde/-einheit selbstständig zu benennen	Vorbereitete Materialien oder Strukturierungen führen die SuS dahin, die Lernziele der Unterrichtsstunde selbstständig zu benennen
Die eingesetzten Medien und Arbeitsmittel führen zu einer i.W. zielgerichteten, aber wenig motivierenden und nicht immer schülergerechten Art der Auseinandersetzung	Die eingesetzten Medien und Arbeitsmittel sind i.W. schülergerecht, sie zeigen den zu erarbeitenden Erkenntnisfortschritt auf, wobei dieser mitunter zu trivial auf der Hand liegt oder zu komplex für das Verständnis der SuS ist	Die eingesetzten Medien und Arbeitsmittel sind motivierend und schülergerecht, sie zeigen den zu erarbeitenden Erkenntnisfortschritt klar, aber nicht trivial auf
Die Sitzordnung und Raumgestaltung sind nicht oder wenig durchdacht auf die Arbeits- und Kommunikationsprozesse abgestimmt, sie unterstützen den reibungslosen Ablauf der Unterrichtsphasen nicht	Sitzordnung und Raumgestaltung sind i.W. auf die in der Stunde intendierten Arbeits- und Kommunikationsprozesse abgestimmt, aber unterstützen den reibungslosen Ablauf der Unterrichtsphasen nur bedingt	Sitzordnung und Raumgestaltung sind sinnvoll auf die in der Stunde intendierten Arbeits- und Kommunikationsprozesse abgestimmt, sie unterstützen den reibungslosen Ablauf der Unterrichtsphasen

Kompe-tenz- bereich	Indikatoren	Indikator relevant	nicht angezeigt	
Lernzuwachs	22) Am Ende der Stunde steht ein klar zu benennender Erkenntnisfortschritt der verbindlich fixiert wird oder die SuS formulieren selbstständig Lösungen für Aufgabenstellungen und stellen diese vor			
	Beobachtet/Signatur, Datum			
	23) Die SuS können den erzielten Erkenntnisfortschritt für weitere Lernprozesse nutzen			
	Beobachtet/Signatur, Datum			

Kompetenzstufen		
niedrig	mittel	hoch
Der Erkenntnisfortschritt wird schriftlich fixiert, wobei sich diese Phase durch eine stark lenkende Rolle des Lehrers auszeichnet	Der Erkenntnisfortschritt wird von SuS selbstständig in gestalteten Produkten oder schriftlichen Lösungen vorgestellt, wobei der Lehrer Korrekturen vornehmen muss, ggf. werden alternative Lösungen gezeigt oder der Lösungsweg verdeutlicht	Der Erkenntnisfortschritt wird von SuS selbstständig in gestalteten Produkten oder schriftlichen Lösungen vorgestellt, ggf. werden alternative Lösungen gezeigt oder der Lösungsweg verdeutlicht
Der in der Stunde erzielte Erkenntnisfortschritt bleibt i.W. isoliert stehen und wird nicht für weitere Lernprozesse genutzt	Der in der Stunde erzielte Erkenntnisfortschritt wird zur Beantwortung von weitergehenden Fragen genutzt, ohne dass die SuS diese in einer eigenen systematisch angeleiteten Arbeitsphase bearbeiten	Der in der Stunde erzielte Erkenntnisfortschritt wird für weiterführende systematisch angelegte Lernarrangements genutzt (gestalterische oder kommunikative Aufgaben, Transferaufgaben, etc.)

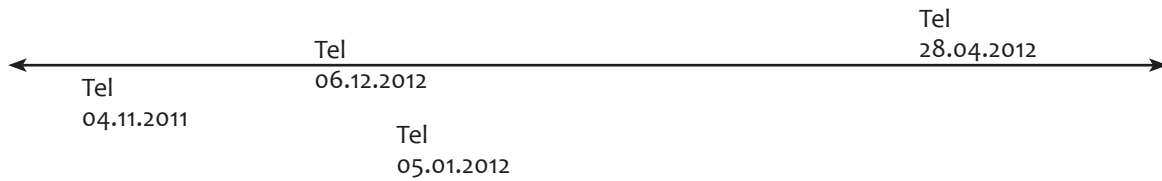
Entwicklung der Reflexionsfähigkeit von XY

Kompetenz als Expertise

erkennt Probleme
im Stundenablauf nicht

erkennt Probleme
ansatzweise

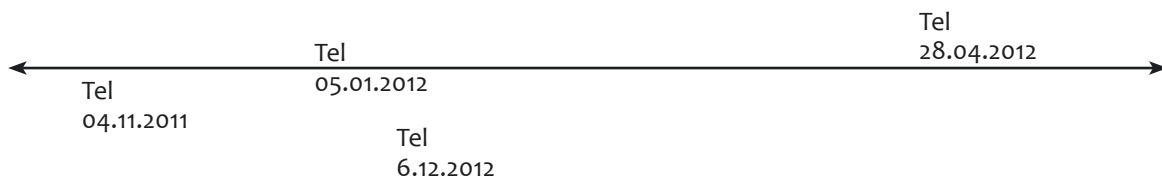
erkennt Probleme
vollständig



Anmerkungen:

beschreibt additiv
und chronologisch

analysiert
kriteriengeleitet



Anmerkungen:

Kompetenz als Repertoire

entwickelt keine
sinnvollen Alternativen

entwickelt mit Hilfe
sinnvolle Alternativen

entwickelt selbstständig
sinnvolle Alternativen



Anmerkungen:

Alpha-Version des Bewertungskonzeptes

Operationalisierung der Studentypen

(nach HS Dahmen und FS Pabst, Studienseminar Bonn, April 2008, ergänzt und erweitert durch Teller, Februar 2013)

Die folgende Operationalisierung von Studentypen basiert auf folgenden Überlegungen:

- Die Studentypen sind idealtypisch aufbereitet, in der Unterrichtspraxis sind sie nicht immer trennscharf voneinander abzugrenzen.
- Die Abgrenzung der Stunden folgt dem Zweck, Auszubildenden einen Überblick über die Vielfalt der verschiedenen Studentypen zu geben und eine Orientierungshilfe für die Unterrichtsplanung, -beobachtung und -auswertung bereitzustellen (Heuristik).
- Jeder Unterrichtsstunde sollte ein bestimmter Schwerpunkt zugeordnet werden können.

Studentypen:

- Einstiegsstunde, Einführungsstunde in ein Unterrichtsvorhaben (Erarbeitung/Präsentation der Schwerpunkte für die zukünftige Arbeit zu einem Unterrichtsthema/-inhalt)
- Erarbeitungsstunde (Erarbeitung und Sicherung von Ergebnissen bezogen auf eine bestimmte Fragestellung)
- Übungs- und Anwendungsstunde (Arbeit mit Lernaufgaben zur Sicherung erlernter Kompetenzen, Übertragen von Erlerntem auf neue Lerninhalte oder Entwicklung metakognitiver Kompetenzen)
- Auswertungsstunde (problemorientierte Auseinandersetzung mit einem Thema, Entwicklung eines Urteils und kritische Würdigung der bisherigen Arbeitsergebnisse)
- thematisch orientierte Abschlussstunde (Zusammenführen der Ergebnisse eines Unterrichtsvorhabens, kriteriengeleitete Auseinandersetzung mit den Ergebnissen und offenen Fragen einer Unterrichtsreihe)
- Evaluationsstunde (Auswertung der Unterrichtsreihe/des Unterrichts im Hinblick auf inhaltliche, methodische oder soziale Aspekte)

- offener Unterricht (eigenständige Erarbeitung eines Themas durch die Lernenden in verschiedenen Sozialformen)
- Experimentalstunde
(die Lernenden führen selbstständig Experimente zu einer festgelegten Thematik durch)
- Exkursion
(der Unterricht wird an einen außerschulischen Lernort verlegt)

Matrix der Qualitätsbereiche guten Unterrichts (in Anlehnung Helmke und Meyer¹¹⁴) (AlphaVersion) Die Matrix der Qualitätsbereiche guten Unterrichts erfüllt folgende Funktion:

- eine theoretische Fundierung, ein „idealtypisches“ Modell für das hier vertretene Verständnis von „gutem Unterricht“ zu schaffen, ohne unter den angegebenen Qualitätsbereichen eine Hierarchie zu unterstellen
- eine *beispielhafte* und *heuristische* Operationalisierung der Qualitätsbereiche anzubieten
- einen Orientierungsrahmen sowohl für die Planung von Unterricht als auch für die Beobachtung und reflektierende Auswertungsgespräche über Unterricht anzubieten

Qualitätsbereich	Erklärung, beispielhafte(!) Operationalisierung in Einzelmerkmalen	Beispiele(!) von Indikatoren (in einer Unterrichtsstunde) [teilweise sind einzelne Indikatoren verschiedenen Qualitätsbereichen zugeordnet]
Klarheit und Strukturiertheit <ul style="list-style-type: none"> ▪ klare Strukturierung (Meyer) ▪ inhaltliche Klarheit (Meyer) ▪ Klassenführung (Helmke) ▪ Klarheit und Strukturiertheit (Helmke) 	<ul style="list-style-type: none"> ▪ Stimmigkeit von Zielen, Inhalten und Methoden Folgerichtigkeit des methodischen Gangs (Ausmaß der Lenkung und der emotionalen Nähe, Komplexität und Abstraktion sind der Lernsituation angepasst) ▪ Folgerichtigkeit des thematischen Gangs (sinnvolle didaktische Inszenierung der Inhalte nach (fach-)didaktischen Prinzipien) ▪ methodischer Grundrhythmus ▪ Aufgaben, Regel- und Rollenklarheit ▪ Aufgabenstellung verständlich und in sich stimmig hinsichtlich der Ziel-, Inhalts- und Methodenwahl (Lernstrukturanalyse; welche Lernschritte muss ein Schüler vollziehen um zu einem Ergebnis zu kommen? Lernstandanalyse; verfügen die Lernenden über die notwendigen Ressourcen zum Lösen der Aufgabe?) ▪ Klarheit der Ergebnissicherung (die Unterrichtsstunde bringt einen klar zu benennenden Erkenntnisfortschritt) ▪ Verbindlichkeit der Ergebnissicherung (Klarheit über definitiv zu fixierende Wissens- und Kompetenzbestände der Lerngruppe) 	<ul style="list-style-type: none"> ▪ die Lernenden verstehen Aufgabenstellungen und können erläutern was sie gerade tun ▪ Regeln werden eingehalten ▪ Rituale werden sinnvoll eingesetzt ▪ einzelne Unterrichtsschritte werden klar markiert ▪ Umfang der Störungen ist gering ▪ wenige, aber sinnvolle, Planungskorrekturen ▪ sprachliche Klarheit ▪ Transparenz über aktuelles Thema und übergeordnete Fragestellungen ▪ Anknüpfen an die Vorstellungs- und Lebenswelt der Lernenden ▪ klare schriftliche Ergebnisfixierungen ▪ Wiederholung und Zusammenfassung wichtiger Ergebnisse, am besten durch die Lernenden

¹¹⁴ Vgl. Helmke, Andreas (2010): Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts. Seelze-Velbert: Klett/Kallmeyer und Meyer, Hilbert (20074): Was ist guter Unterricht? Berlin: Cornelsen Scriptor.

Qualitätsbereich <ul style="list-style-type: none"> ▪ Quellen der Merkmale und Indikatoren 	Erklärung, beispielhafte(!) Operationalisierung in Einzelmerkmalen	Beispiele(!) von Indikatoren (in einer Unterrichtsstunde) [teilweise sind einzelne Indikatoren verschiedenen Qualitätsbereichen zugeordnet]
hoher Anteil echter Lernzeit <ul style="list-style-type: none"> ▪ hoher Anteil echter Lernzeit (Meyer) ▪ Klassenführung (Helmke) 	<ul style="list-style-type: none"> ▪ „time on task“ (tatsächlich aufgewendete Lernzeit) ▪ Hohes Momentum (Vermeidung unnötiger Unterbrechungen des Unterrichtsflusses, sowie von Hektik, Herstellen eines „geschmeidigen“ Unterrichtsverlaufs) 	<ul style="list-style-type: none"> ▪ die Lernenden sind nicht unaufmerksam ▪ die Lernenden arbeiten aktiv mit und lassen sich nicht durch Kleinigkeiten ablenken ▪ es entstehen reichhaltige Arbeitsergebnisse ▪ der Lehrer gibt gezielte Lernhilfen ▪ der Lehrer lässt den Lernenden nach Fragen Zeit, Antworten zu überlegen
lernförderliches Klima <ul style="list-style-type: none"> ▪ lernförderliches Klima (Meyer) ▪ lernförderliches Klima (Helmke) 	<ul style="list-style-type: none"> ▪ gegenseitiger Respekt der Lernenden /zw. Lernenden und Lehrer ▪ verlässliche Regeln ▪ gemeinsame Verantwortung von Lehrer und Lernenden für erfolgreichen Lernprozess (Arbeitsvertrag) ▪ Gerechtigkeit und Fürsorge des Lehrers gegenüber den Lernenden ▪ konstruktiver Umgang mit Fehlern, auf Lehrer wie auf Lernenden-Seite ▪ Balance zwischen Lern- und Leistungssituationen ▪ Balance zwischen „Lockerheit“ und „Anspruch“ ▪ Schülerfeedback 	<ul style="list-style-type: none"> ▪ Respektvoller Umgang der Lernenden untereinander und in der Lernenden-Lehrer-Beziehung (kein aggressives Verhalten, Beleidigungen und Demütigungen, keine Diskriminierung einzelner Lernender , keine Rivalitäten zwischen Gruppen) ▪ humorvolle, entspannte Atmosphäre bei gleichzeitiger Leistungsorientierung ▪ Ermutigungen durch den Lehrer
Methodenvielfalt und -tiefe <ul style="list-style-type: none"> ▪ Methodenvielfalt (Meyer) ▪ Angebotsvielfalt (Helmke) 	<ul style="list-style-type: none"> ▪ der Unterricht wird auf vielfältige Art und Weise inszeniert (Aktions- und Sozialformen, Medieneinsatz, Handlungsformen, methodische Großformen, Lernorte) ▪ die eingesetzten Unterrichtsmethoden weisen eine Passung auf die Lerngruppe und die intendierten Lernziele auf ▪ die eingesetzten Methoden leiten zur tiefen fachlichen Durchdringung an 	<ul style="list-style-type: none"> ▪ der Unterricht zeigt ein vielfältige Repertoire an den zuvor erwähnten Inszenierungstechniken ▪ die Lernenden nutzen die Unterrichtsmethoden im Sinne der Planung und können die für sie vorgesehenen Rollen sicher einnehmen ▪ Metakommunikation über den Methodeneinsatz

Qualitätsbereich <ul style="list-style-type: none"> ▪ Quellen der Merkmale und Indikatoren 	Erklärung, beispielhafte(!) Operationalisierung in Einzelmerkmalen	Beispiele(!) von Indikatoren (in einer Unterrichtsstunde) [teilweise sind einzelne Indikatoren verschiedenen Qualitätsbereichen zugeordnet]
Umgang mit Heterogenität <ul style="list-style-type: none"> ▪ individuelles Fördern (Meyer) ▪ Umgang mit Heterogenität (Helmke) 	<ul style="list-style-type: none"> ▪ Unterricht den verschiedenen Lernvoraussetzungen der Lernenden anpassen (Adaptivität/Passung) ▪ Lernenden verschiedener Leistungsbereiche, sowohl leistungsschwachen als auch leistungsstarken, wertschätzend begegnen ▪ Angebote zum Erwerb von Lernstrategien 	<ul style="list-style-type: none"> ▪ innere Differenzierung des Unterrichts (binnendifferenzierende Aufgabenstellungen, unterschiedliche Lernmaterialien werden eingesetzt, für Lernende mit Lernschwierigkeiten stehen zusätzliche Lernhilfen bereit, Möglichkeit für leistungsstarke Lernende, Routineaufgaben nicht zu erledigen und nach eigenen Schwerpunkten zu arbeiten)
<ul style="list-style-type: none"> ▪ Konsolidierung/Sicherung ▪ intelligentes Üben (Meyer) ▪ Konsolidierung/Sicherung (Helmke) 	<ul style="list-style-type: none"> ▪ in den Unterricht werden intelligente Übungsphasen eingebaut (lernstandsadäquate Entwicklung von Lernstrategien, Rhythmisierung) 	<ul style="list-style-type: none"> ▪ in den Übungsphasen werden Hilfestellungen bereitgestellt ▪ Übungen leiten zum Transfer, der Qualitätssteigerung oder der Automatisierung von Wissen gezielt an ▪ eine Selbstkontrolle des Lernerfolgs ist möglich ▪ auf Phasen der Stoffvermittlung folgen Phasen der kommunikativen Anwendung
transparente Leistungserwartungen <ul style="list-style-type: none"> ▪ transparente Leistungserwartungen (Meyer) ▪ lernförderliches Klima (Helmke) 	<ul style="list-style-type: none"> ▪ Lernangebote bereitzustellen die den Ansprüchen der Richtlinien, den Bildungsstandards und den Fähigkeiten der Lernenden entsprechen ▪ Transparenz über Lernziele ▪ Rückmeldungen über den Lernfortschritt geben 	<ul style="list-style-type: none"> ▪ die Lernziele der Unterrichtsstunde/-einheit sind transparent ▪ die Grundlagen der Leistungsbewertung werden offen gelegt ▪ deutliches Lehrer-Feedback, ob Beiträge „richtig“ waren ▪ Lernenden-Feedback zu Beiträgen

Qualitätsbereich <ul style="list-style-type: none"> ▪ Quellen der Merkmale und Indikatoren 	Erklärung, beispielhafte(!) Operationalisierung in Einzelmerkmalen	Beispiele(!) von Indikatoren (in einer Unterrichtsstunde) [teilweise sind einzelne Indikatoren verschiedenen Qualitätsbereichen zugeordnet]
sinnstiftendes Kommunizieren (Meyer)	<ul style="list-style-type: none"> ▪ Kommunikationsprozess der „persönlichen qualitativen Umgebung“ des Gelernten für die Lernenden 	<ul style="list-style-type: none"> ▪ Metakommunikation über die Unterrichtsinhalte, bzw. über den Unterricht ▪ Anregungen des Lehrers zur Einordnung der Inhalte in den eigenen „Sinnhorizont“ der Lernenden ▪ Verknüpfungen zu anderen Unterrichtsthemen oder Alltagswissen ▪ persönliche Stellungnahmen der Lernenden zu Inhalten ▪ Lernenden-Feedback
vorbereitete Umgebung (Meyer)	<ul style="list-style-type: none"> ▪ ordentliche, mit sinnvollem Lernmaterial ausgestattete Räume, zu denen Lehrer und Lernende ein Gefühl von „Ownership“ haben ▪ Belüftung, Beleuchtung und Akustik sind stimmig ▪ ausreichend Platz für verschiedene Inszenierungstechniken des Unterrichts und den Bewegungsdrang der Lernenden 	<ul style="list-style-type: none"> ▪ ökologisch und ergonomisch sinnvolle Raumausstattung ▪ sauberer und gepflegter Eindruck der Räume ▪ pfleglicher Umgang mit den Materialien ▪ es gibt „Funktionsecken“ in den Räumen
Aktivierung (Helmke)	<ul style="list-style-type: none"> ▪ kognitive Aktivierung im Sinne der Selbststeuerung des Lernens, es erfolgt eine „tiefe“, anstatt bloß oberflächlicher „mechanischer“ Verarbeitung des Stoffes (Lernstrategien) ▪ soziale Aktivierung (kooperative Arbeitsformen) ▪ aktive Teilhabe 	<ul style="list-style-type: none"> ▪ sinnvoller Einsatz kooperativer Arbeitsformen ▪ Einsatz von Hilfestellungen zum Erwerb von Lernstrategien ▪ Lernenden wird (Teil-)Verantwortung für den Lernerfolg zugemutet
Motivierung (Helmke)	<ul style="list-style-type: none"> ▪ extrinsische und intrinsische Motivationsgründe bewegen die Lernenden zum mitarbeiten ▪ anspruchsvolle Lehrererwartungen 	<ul style="list-style-type: none"> ▪ motivierende, überraschende Einstiege ▪ „fantasiereiche“ Unterrichtsgestaltung ▪ Herstellung von Situationen, in denen Lernende Selbstwirksamkeit erleben ▪ leistungsbezogene, transparente Bewertung ▪ klare entwicklungsbezogene Perspektive

Qualitätsbereich <ul style="list-style-type: none"> ▪ Quellen der Merkmale und Indikatoren 	Erklärung, beispielhafte(!) Operationalisierung in Einzelmerkmalen	Beispiele(!) von Indikatoren (in einer Unterrichtsstunde) [teilweise sind einzelne Indikatoren verschiedenen Qualitätsbereichen zugeordnet]
Kompetenzorientierung (Helmke)	<ul style="list-style-type: none"> ▪ die Gestaltung des Unterrichts orientiert sich an den am Ende des Lernprozesses zu erreichenden Kompetenzfortschritten der Lernenden ▪ der erfolgreiche Kompetenzerwerb wird evaluiert 	<ul style="list-style-type: none"> ▪ klare Kompetenzorientierung in den Aufgabenstellungen und Leistungsüberprüfungen ▪ fachdidaktische Prinzipien leiten den Kompetenzerwerb systematisch an ▪ klar erkennbare Lernprogression über die Anforderungsbereiche der Aufgabenstellungen
Lernzuwachs (QA NRW)	<ul style="list-style-type: none"> ▪ ein Großteil der Lernenden erreicht einen intendierten Lernzuwachs inhaltlicher, methodischer oder sozialer Art ▪ der Lernzuwachs wird dokumentiert, von den Lernenden gezeigt 	<ul style="list-style-type: none"> ▪ Lernende können die zum Beginn der Stunde/Reihe aufgeworfene (Leit-)Fragen beantworten ▪ Lernende formulieren selbstständig Lösungen für Aufgabenstellungen und stellen diese vor ▪ Lernende Antworten auf Transferfragen ▪ die Kompetenzfortschritte werden für die Formulierung weiterer Arbeitsschritte oder Fragestellungen genutzt ▪ die Lernenden stellen Ergebnisse/Produkte eines Arbeitsprozesse vor und bewerten die Aussagekraft ihrer Ergebnisse ▪ die Lernenden reflektieren ihren Arbeitsprozess und formulieren Alternativen für Probleme

Dokumentationsbogen für die Nachbesprechung (Alpha-Version)

Die Dokumentation der Nachbesprechung erfüllt folgende Funktionen:

- eine transparente und verbindliche Bilanzierung des gesehenen Unterrichts vorzunehmen, inklusive einer Fixierung von beispielhaft gezeigten Kompetenzen des/der Auszubildenden, die in ein dauerhaftes Handlungsrepertoire überführt werden sollten
- eine grundlegende Einordnung des gezeigten Unterrichts im Hinblick auf ein zu erwerbendes Repertoire von Unterricht vorzunehmen (vgl. das Instrument „Sammlung über das Repertoire“) und in Beziehung zu zentralen allgemein- und fachdidaktischen Prinzipien und Fragestellungen zu setzen
- eine für die/den AuszubildendeN transparente Einschätzung über ihre/seine Fähigkeiten zur professionellen Selbstreflektion vorzunehmen
- einen Abgleich der Selbstwahrnehmung des/der Auszubildenden mit der professionellen Fremdwahrnehmung durch den/die AusbilderIn systematisch und transparent anzuleiten
- die Nachbesprechung mit einer transparenten und schriftlich fixierten Einschätzung über den derzeitigen Leistungsstand des/der Auszubildenden zu beenden, die im Anschluss an den Unterrichtsbesuch zeitnah ausgehändigt wird
- einen Abgleich des erreichten Ausbildungsstandes mit einem zu erreichenden Examenstandard vorzunehmen

AuszubildendeR: XY

AusbilderIn: XY

Grundlegende Einordnung des Unterrichts im Sinne eines zu erwerbenden Repertoires

Datum des Ubs

Studentyp: (siehe Operationalisierung der Studententypen)

Abgedeckte fachspezifische Lern- und Kompetenzbereiche: (siehe fachspezifische Operationalisierung)

Sonstige relevante Aspekte: (z.B. Medien, Unterrichtsmethoden, Sozialformen, Grad der LehrerInnenzentrierung, Besonderheiten der Lerngruppe,...)

Qualitative Einschätzung der Unterrichtsstunde (ggf. Bewertung) erfasst an prägenden Merkmalen der jeweiligen Stunde: (z.B.: Stimmigkeit von Methoden, Inhalten und Zielen, lerngruppenadäquate Auswahl von Inhalten, Methoden und Sozialformen,...)

Thematisierung einer Auswahl der die Unterrichtsstunde prägenden Qualitätsbereiche (in Bezug zum Studentyp, Lernbereich, angestrebter Kompetenzerweiterung, sowie der Lerngruppe)

(Einzuschätzen und in die Skala einzutragen nach den Anforderungen des derzeitigen Ausbildungsstandes („A“) und des zu erreichenden Examenstandards („E“))

Qualitätsbereich: _____

Operationalisierung/Indikatoren in der Stunde:

Thematisierung (positiv wie negativ):

Qualitätseinschätzung (Ausbildungsstand (A) und Examenstandard (E)):

Der/die Auszubildende setzt die Anforderungen des Qualitätsbereiches...

nicht um

teilweise um

voll und ganz um



Qualitätsbereich: _____

Operationalisierung/Indikatoren in der Stunde:

Thematisierung (positiv wie negativ):

Qualitätseinschätzung (Ausbildungsstand (A) und Examenstandard (E)):

Der/die Auszubildende setzt die Anforderungen des Qualitätsbereiches...

nicht um

teilweise um

voll und ganz um



Qualitätsbereich: _____

Operationalisierung/Indikatoren in der Stunde:

Thematisierung (positiv wie negativ):

Qualitätseinschätzung (Ausbildungsstand (A) und Examenstandard (E)):

Der/die Auszubildende setzt die Anforderungen des Qualitätsbereiches...

nicht um

teilweise um

voll und ganz um




Reflexionskompetenz von XY

(Einzuschätzen und in die Skala einzutragen nach den Anforderungen des derzeitigen Ausbildungsstandes („A“) und des zu erreichenden Examensstandards („E“))

Kompetenz als Expertise

Der/die Auszubildende beschreibt bestimmende Aspekte der Unterrichtsstunde...

nicht ansatzweise vollständig



ggf. Notizen:

Der/die Auszubildende beschreibt...

additiv und chronologisch analysiert kriteriengeleitet



ggf. Notizen:

Kompetenz als Repertoire

Der/die Auszubildende entwickelt für Probleme im Stundenablauf...

keine mit Hilfe selbstständig
 sinnvollen Alternativen sinnvolle Alternativen sinnvolle Alternativen



ggf. Notizen:

Aus dieser Stunde sollten folgende Aspekte einer professionellen Unterrichtskompetenz in ein dauerhaftes Handlungsrepertoire überführt werden:

Für die Entwicklung der eigenen Unterrichtskompetenz sollten als Aspekte für eine weitere Professionalisierung berücksichtigt werden:

Sammlung über das Repertoire (Alpha-Version)

Die Sammlung über das Repertoire erfüllt folgende Funktionen:

- eine Übersicht über die Vielfalt verschiedener vom/von der Auszubildenden gezeigten Unterrichtsinszenierungen bereitzustellen, die dokumentiert, dass der/die Auszubildende Unterricht anleiten kann, welcher den vielfältigen Ansprüchen seines/ihres Faches (Studentypen, Lern- und Anforderungsbereiche) als auch den Bedürfnissen der jeweiligen Lerngruppen gerecht wird
- eine Bilanzierung von gezeigten Einzelaspekten einer allgemeinen Unterrichtskompetenz, die in ein dauerhaftes Handlungsrepertoire des/der Auszubildenden übernommen werden sollten, vorzunehmen
- die Entwicklung der professionellen Reflexionsfähigkeit des/der Auszubildenden im Zeitverlauf zu dokumentieren

AuszubildendeR: XY

„Repertoire“ des Unterrichts von XY

Lfd. Nr.	Datum des UBs/Signatur AusbilderIn, Fach	Studententyp	abgedeckte fachspezifische Lern- und Kompetenzbereiche	sonstige relevante Aspekte (z.B. Medien, Unterrichtsmethoden, Sozialformen, Grad der LehrerInnenzentrierung, Besonderheiten der Lerngruppe,...)	Qualitative Einschätzung der Unterrichtsstunde (ggf. Bewertung nach Ausbildungsstand) erfasst an prägenden Merkmalen der jeweiligen Stunde (z.B.: Stimmigkeit von Methoden, Inhalten und Zielen, lerngruppenadäquate Auswahl von Inhalten, Methoden und Sozialformen,...)
1	04.11.2011/ Tel, Sozialwissenschaften	Erarbeitungsstunde	Ökonomie, Urteils- und Handlungskompetenz	JG. 10, Pro- Contra Diskussion	Zielgerichteter Einsatz der Großmethode der Pro-Contra Diskussion für die Hinführung zu Urteilskompetenz Bewertung: 1,3
	...				


Entwicklung der Reflexionskompetenz von XY

(die Einschätzung über die Erreichung des „Examensstandards“ ist aus den Dokumentationsbögen der Nachbesprechung zu übertragen)

Kompetenz als Expertise

Der /die Auszubildende beschreibt bestimmende Aspekte der Unterrichtsstunde...

nicht ansatzweise vollständig



Der/die Auszubildende...

beschreibt *additiv und chronologisch* *analysiert kriteriengeleitet*



Kompetenz als Repertoire

Der/die Auszubildende entwickelt für Probleme im Stundenablauf...

keine mit Hilfe selbstständig
 sinnvollen Alternativen sinnvolle Alternativen sinnvolle Alternativen



Kompetenzbereich Motivierung:

Der/die Auszubildende setzt die Anforderungen des Qualitätsbereiches...

nicht um

teilweise um

voll und ganz um



Kompetenzbereich Kompetenzorientierung:

Der/die Auszubildende setzt die Anforderungen des Qualitätsbereiches...

nicht um

teilweise um

voll und ganz um



Kompetenzbereich Lernzuwachs:

Der/die Auszubildende setzt die Anforderungen des Qualitätsbereiches...

nicht um

teilweise um

voll und ganz um



Leitfaden für die *problemzentrierte Gruppendiskussion* zur Auswahl zentraler Teilkompetenzen und der Rezeption der Instrumente mit *Fachleitern*

Problemstellung:

- *Wie werden die Bewertungsinstrumente wahrgenommen, besonders der Ratingbogen für die Unterrichtsbesuche?*
- *Welche Teilkompetenzen einer allgemeinen Unterrichtskompetenz sollten Eingang in den Bewertungsbogen finden?*

Phasen der Diskussion:

- Einführung (Zusammenfassung der Rahmenbedingungen, des Erkenntnisinteresses und der Grundregeln der Diskussion)
- Warm-Up (allg. Aspekte des Themas, Herstellung einer vertrauensvollen und intimen Gesprächsatmosphäre)
- Hauptteil (Bearbeitung der verschiedenen Themenblöcke)
- Abschlussteil (Zusammenfassung der Ergebnisse, stimmiger Ausstieg)

Grundregeln der Gruppendiskussion:

- Im Zentrum steht die Interaktion der Teilnehmer untereinander, der Moderator gibt für diese nur (problemzentrierte) Impulse oder greift ein, wenn Regeln der Diskussion missachtet werden
- Trichterfunktion: zunehmende thematische Verengung auf das Erkenntnisinteresse und die Themenblöcke nach offenem Beginn
- Anregung zur wechselseitigen Diskussion
- Schlüsselfragen in verschiedenen Themenblöcken mehrfach beantworten lassen
- Rückspiegelung der „gelernten“ Aspekte, damit die Diskutanten vertiefen und ergänzen können
- möglichst offenes Design des Leitfadens und offene Formulierung der Fragen, um den Befragten die Möglichkeit zu geben, die Zusammenhänge in eigenen Worten zu entfalten

Einführung

- *Erkenntnisinteresse heute: Welche Wahrnehmung haben Sie von den Beurteilungsinstrumenten?*
- *Im Mittelpunkt heute stehen die verschiedenen Perspektiven auf das Instrument, ausgehend von verschiedenen Lehrerpersönlichkeiten und verschiedenen Fächern, es gibt kein „richtig“ oder „falsch“, sondern nur verschiedene Perspektiven auf die vorgestellte Idee, Unterricht im Referendariat zu bewerten.*
- *Es geht nicht um die Identifizierung von Mehrheiten oder Minderheiten!*
- *Rückfragen von mir, ggf. auch überspitzte, dienen dazu, den Inhalt ihrer Aussagen möglichst klar herauszuarbeiten*
- *Aufzeichnung lediglich zur Transkription des Interviews, Wiedergabe der Inhalte nur anonymisiert, Rückbezug auf Fächer kann ausgeführt werden*

Warm-Up

Schilderung von Erfahrungen über die Bewertung im Referendariat oder Unterrichtsnachbesprechungen mit Referendaren (Probleme, best-practice, Herausforderungen,...)

- *Wenn Sie an ihre Erfahrungen mit Bewertungen von Unterrichtsbesuchen im Rahmen ihrer Tätigkeit als Fach- oder Hauptseminarleiter zurückdenken, was fällt ihnen da spontan ein?*

Hauptteil (questioning route)

[Frageblock 1: *Wahrnehmung des Instrumentes, der „Logik“ mit der Unterricht bewertet werden soll*] **nicht vorlesen!**

- **Einstiegsfrage:**
 - *Was ist der erste Eindruck, den sie von den Bewertungsinstrumenten haben?*

▪ Überleitung:

- *Inwiefern ergibt sich dies für sie?/Wie kommen Sie zu dieser Auffassung?*

▪ Schlüsselfragen:

- *Was halten Sie von der standardisierten Form der Unterrichtsbewertung?*
 - ◆ *positive Aspekte*
 - ◆ *negative Aspekte*
 - ◆ *Vor dem Hintergrund Ihrer eigenen Erfahrung*
- *Beobachtungsbögen*
 - ◆ *Wenn Sie den Beobachtungsbogen mit den Ratingskalen und jenen mit den Kompetenzstufen vergleichen, welche Form erscheint Ihnen sinnvoller um Unterricht zu bewerten?*
 - ▲ *positive Aspekte der beiden Varianten*
 - ▲ *negative Aspekte der beiden Varianten*
 - ◆ *Lässt sich das den Bögen zu Grunde liegende Verständnis von „gutem Unterricht“ sinnvoll auf Ihre Unterrichtsfächer übertragen?*
- *Was halten Sie von dem Dokumentationsbogen für die Nachbesprechung?*
- *Was halten Sie von der „Sammlung über das Repertoire“?*

▪ Schlussfragen:

- *Glauben Sie, dass die vorliegenden Instrumente zu einer Verbesserung der Bewertungsprozesse von Lehrproben beitragen können?*

▪ Möglichkeiten zum Durchbrechen von Gruppenkonformität:

- *Widersprechen sich die Vielfalt möglicher Lehrerpersönlichkeiten und eine standardisierte Bewertung nicht?*
- *Bedingt eine gerechte und transparente Bewertung der Auszubildenden nicht zwangsläufig eine Standardisierung der Bewertungsinstrumente?*

[Frageblock 2: Auswahl der Teilkompetenzen], nicht vorlesen

▪ **Einstiegsfrage:**

- Die Beobachtungsbögen umfassen nicht alle Merkmale „guten Unterrichts“ von Helmke und Meyer, wie sie in der Matrix mit den Qualitätsbereichen guten Unterrichts zusammengefasst sind. Wenn Sie sich die Auswahl der Qualitätsbereiche und Merkmale in den Beobachtungsbögen anschauen, halten Sie diese Auswahl für sinnvoll und erschöpfend, um guten Unterricht zu operationalisieren?

▪ **Überleitung:**

- Inwiefern ergibt sich dies für sie?/Wie kommen Sie zu dieser Auffassung?

▪ **Schlüsselfragen:**

- Sollten weitere Merkmale aus der Matrix in den Beobachtungsbögen übernommen werden?
 - ♦ Welche?
 - ♦ Woran machen Sie dies fest?
- Sollten einzelne Merkmale aus dem Beobachtungsbogen gestrichen werden?
 - ♦ Woran machen sie fest, dass diese gestrichen werden sollen?
- Fehlen zentrale Merkmale einer allgemeinen „Unterrichtskompetenz“ in der Matrix?
 - ♦ Können Sie diese fehlenden Merkmale umreißen oder formulieren?

▪ **Schlussfragen:**

- Halten Sie die vorliegende Auswahl von Merkmalen guten Unterrichts, bzw. eine um Anregungen ergänzte Auswahl, für hilfreich um die Referendare bei der Entwicklung von „Unterrichtskompetenz“ anzuleiten?
 - ♦ (Fokus auf Auswahl und umfassenden Geltungsanspruch der ausgewählten Kompetenzen)

▪ **Möglichkeiten zum Durchbrechen von Gruppenkonformität:**

- Halt Sie es für vermessen, für die Ausbildung eine „erschöpfende“ Darstellung zentraler Unterrichtskompetenzen erarbeiten zu wollen?
- Führt eine Formulierung von Teilkompetenzen zu einer unzulässigen und theoretisierenden „Zerstückelung“ einer kreativen Tätigkeit wie Unterrichten?
- Brauchen Referendare einen ausformulierten Orientierungsrahmen, der ihnen möglichst konkret anzeigt, was sie lernen sollen?

[Frageblock 3: Was macht eine „gute“ Bewertung von Unterrichtsbesuchen ihrer Meinung nach aus?/Wie kann das vorliegende Konzept dazu beitragen?], nicht vorlesen

Optional, wenn noch Zeit ist!!!

- **Einstiegsfrage:**
 - Schildern Sie kurz, wie Sie sich den „optimalen“ Ablauf eines Bewertungsgesprächs im Anschluss an eine Lehrprobe vorstellen.
- **Überleitungsfrage:**
 - Wie oft erleben Sie Bewertungsgespräche, die sich diesem Ideal annähern?
- **Schlüsselfragen:**
 - Kann eine Standardisierung der Bewertung dabei helfen diesen „idealen Ablauf des Bewertungsgesprächs“ zu realisieren?
- **Schlussfragen:**
 - Glauben Sie, dass die vorliegenden Bewertungsinstrumente dazu beitragen können, die Bewertungsrealität im Referendariat dem von Ihnen formulierten Ideal anzunähern?
- **Möglichkeiten zum Durchbrechen von Gruppenkonformität:**
 - Konfliktfeld zwischen Standardisierung und Offenheit ansprechen.

Leitfaden für die problemzentrierte *Gruppendiskussion* zur Rezeption der Bewertungsinstrumente mit *Referendaren*

Problemstellung:

- *Wie wurden die Bewertungsprozesse im Vorbereitungsdienst wahrgenommen?*
- *Wie werden die entwickelten Bewertungsinstrumente wahrgenommen?*
- *Können die vorgestellten Instrumente helfen, die erlebten Missstände der Bewertungsprozesse zu verbessern?*

Phasen der Diskussion:

- Einführung (Zusammenfassung der Rahmenbedingungen, des Erkenntnisinteresses und der Grundregeln der Diskussion)
- Warum-Up (allg. Aspekte des Themas, Herstellung einer vertrauensvollen und intimen Gesprächsatmosphäre)
- Hauptteil (Bearbeitung der verschiedenen Themenblöcke)
- Abschlussteil (Zusammenfassung der Ergebnisse, stimmiger Ausstieg)

Grundregeln der Gruppendiskussion:

- Im Zentrum steht die Interaktion der Teilnehmer untereinander, der Moderator gibt für diese nur (problemzentrierte) Impulse oder greift ein, wenn Regeln der Diskussion missachtet werden
- Trichterfunktion: zunehmende thematische Verengung auf das Erkenntnisinteresse und die Themenblöcke nach offenem Beginn
- Anregung zur wechselseitigen Diskussion
- Schlüsselfragen in verschiedenen Themenblöcken mehrfach beantworten lassen
- Rückspiegelung der „gelernten“ Aspekte, damit die Diskutanten vertiefen und ergänzen können
- möglichst offenes Design des Leitfadens und offene Formulierung der Fragen, um den Befragten die Möglichkeit zu geben, die Zusammenhänge in eigenen Worten zu entfalten

Einführung

- Erkenntnisinteresse heute: Welche Wahrnehmung haben Sie von den Beurteilungsinstrumenten, der Art, wie mit den Instrumenten Unterrichtsbesuche bewertet werden sollen?
- Im Mittelpunkt heute stehen die verschiedenen Perspektiven auf das Instrument, ausgehend von verschiedenen Lehrerpersönlichkeiten und verschiedenen Fächern, es gibt kein „richtig“ oder „falsch“, sondern nur verschiedene Perspektiven auf die vorgestellte Idee Unterricht im Referendariat zu bewerten.
- Es geht nicht um die Identifizierung von Mehrheiten oder Minderheiten!
- Rückfragen von mir, ggf. auch überspitzte, dienen dazu den Inhalt ihrer Aussagen möglichst klar herauszuarbeiten
- Aufzeichnung lediglich zur Transkription des Interviews, Wiedergabe der Inhalte nur anonymisiert, Rückbezug auf Fächer kann ausgeführt werden

Warm-Up

Schilderung von Erfahrungen über die eigene Bewertung im Referendariat (Probleme, best-practice, Herausforderungen,.....)

- Wenn Sie an ihre Erfahrungen mit Bewertungen im Rahmen ihrer eigenen Lehrerausbildung zurückdenken, was fällt Ihnen da spontan ein?
- Können Sie Aspekte der Bewertung umreißen die....
 - ◆ ... sie als negativ wahrgenommen haben?
 - ◆ ... sie als positiv wahrgenommen haben?

Hauptteil (questioning route)

[Frageblock 1: *Wahrnehmung des Instrumentes, der „Logik“, mit der Unterricht bewertet werden soll*] nicht vorlesen!

▪ Einstiegsfrage:

- *Was ist der erste Eindruck, den sie von den Sie von den Bewertungsinstrumenten haben?*

▪ Überleitung:

- *Inwiefern ergibt sich dies für sie?/Wie kommen Sie zu dieser Auffassung?*

▪ Schlüsselfragen:

- *Was halten Sie von der standardisierten Form der Unterrichtsbewertung?*
 - ♦ *positive Aspekte*
 - ♦ *negative Aspekte*
 - ♦ *Vor dem Hintergrund Ihrer eigenen Erfahrung*
- *Halten Sie den Beobachtungsbogen für sinnvoll, um Unterrichtsbesuche zu bewerten?*
- *Wenn Sie sich den Dokumentationsbogen für die Nachbesprechung vor Augen führen, glauben Sie, dass dieser zu einer produktiven Gestaltung der Unterrichtsnachbesprechung beitragen kann?*
- *Wenn Sie sich die Sammlung über das Repertoire vor Augen führen, glauben Sie, dass diese Ihnen bei der Entwicklung Ihrer individuellen Lehrerpersönlichkeit geholfen hätte?*

▪ Schlussfragen:

- *Glauben Sie, dass die vorliegenden Instrumente zu einer Verbesserung der Bewertungsprozesse der Lehrproben beitragen könnten?*

▪ Möglichkeiten zum Durchbrechen von Gruppenkonformität:

- *Widersprechen sich die Vielfalt möglicher Lehrerpersönlichkeiten und eine standardisierte Bewertung nicht?*
- *Bedingt eine gerechte und transparente Bewertung der Auszubildenden nicht zwangsläufig eine Standardisierung der Bewertungsinstrumente?*

[Frageblock 2: Was macht eine „gute“ Bewertung von Unterrichtsbesuchen ihrer Meinung nach aus?/Wie kann das vorliegende Konzept dazu beitragen?], nicht vorlesen

▪ **Einstiegsfrage:**

- *Schildern Sie kurz, wie Sie sich den „optimalen“ Ablauf eines Bewertungsgespräches im Anschluss an eine Lehrprobe vorstellen.*

▪ **Überleitungsfrage:**

- *Wie oft haben sie Bewertungsgespräche erlebt, die sich diesem Ideal annähern?*

▪ **Schlüsselfragen:**

- *Kann eine Standardisierung der Bewertung dabei helfen, diesen „idealen Ablauf des Bewertungsgespräches“ zu realisieren?*

▪ **Schlussfragen:**

- *Glauben Sie, dass die vorliegenden Bewertungsinstrumente dazu beitragen können, die Bewertungsrealität im Referendariat dem von Ihnen formulierten Ideal anzunähern?*

▪ **Möglichkeiten zum Durchbrechen von Gruppenkonformität:**

- *Konfliktfeld zwischen Standardisierung und Offenheit ansprechen.*

Leitfaden für das *problemzentrierte Interview* zur Erprobung der Bewertungsinstrumente mit *Fachleitern*

Problemstellung:

- *Ist das Konzept in der Ausbildungspraxis einsetzbar, weist es eine Adaptivität mit der Praxis auf?*
- *Bietet das Konzept eine Hilfestellung für die Planung, Durchführung und Analyse von Unterricht, bietet es Gesprächsanlässe?*
- *Kann das Konzept die Transparenz der Bewertungsprozesse sicherstellen?*

Phasen des Interviews:

- Einführung (Zusammenfassung der Rahmenbedingungen und des Erkenntnisinteresses)
- Warm-Up (allg. Aspekte des Themas, Herstellung einer vertrauensvollen und intimen Gesprächsatmosphäre)
- Hauptteil (Bearbeitung der verschiedenen Themenblöcke)
- Abschlussteil (Zusammenfassung der Ergebnisse, stimmiger Ausstieg)

Grundregeln des Interviews:

- Im Zentrum steht die Wahrnehmung der Interviewten, der Interviewer gibt durch Fragen (problemzentrierte) Impulse
- Trichterfunktion: zunehmende thematische Verengung auf das Erkenntnisinteresse und die Themenblöcke nach offenem Beginn
- Schlüsselfragen in verschiedenen Themenblöcken mehrfach beantworten lassen
- Rückspiegelung der „gelernten“ Aspekte, damit die Interviewten vertiefen und ergänzen können
- möglichst offenes Design des Leitfadens und offene Formulierung der Fragen, um den Befragten die Möglichkeit zu geben, die Zusammenhänge in eigenen Worten zu entfalten

Einführung

- **Erkenntnisinteresse heute:** *Wie haben Sie den Praxiseinsatz der Bewertungsinstrumente erlebt? Unterstützen die Instrumente die Referendare darin ihren eigenen Unterrichtsstil zu finden?*
- *Im Mittelpunkt heute steht Ihre Perspektive auf die Instrumente, ausgehend von Ihrer Ausbilderpersönlichkeit und Ihrem Ausbildungsfach, es gibt kein „richtig“ oder „falsch“, sondern Ihre Perspektive auf die vorgestellte Idee, Unterricht im Referendariat zu bewerten, zählt.*
- *Im Mittelpunkt stehen die Ausbildungssituationen, in denen des Bewertungskonzept eingesetzt wurde. Wird der Kontext auf andere Referendare und Ausbildungskontexte ausgedehnt, so wird dies erwähnt.*
- *Rückfragen von mir, ggf. auch überspitzte, dienen dazu den Inhalt ihrer Aussagen möglichst klar herauszuarbeiten*
- *Aufzeichnung lediglich zur Transkription des Interviews, Wiedergabe der Inhalte nur anonymisiert, Rückbezug auf Fächer kann ausgeführt werden*

Warm-Up

- *Schilderung von Erfahrungen über die Bewertung mit den Instrumenten (Probleme, Verbesserungen im Gegensatz zu ansonsten praktiziertem Vorgehen,....)*
 - *Wenn Sie auf den Einsatz der drei verschiedenen Instrumente zurückblicken, was fällt ihnen da spontan ein?*

Hauptteil (questioning route)

[Frageblock 1: Ist das Konzept in der Ausbildungspraxis einsetzbar, weist es eine Adaptivität mit der Praxis auf?] nicht vorlesen!

▪ Einstiegsfrage:

- Wie war Ihr Eindruck: Konnte das Bewertungskonzept, so wie es war, in der Ausbildungspraxis eingesetzt werden oder gab es Probleme beim Einsatz?

▪ Überleitung:

- Inwiefern ergibt sich dies für Sie?/Wie kommen Sie zu dieser Auffassung?
- War der zeitliche Aufwand, der für den Einsatz des Konzeptes nötig war, angemessen?

▪ Schlüsselfragen:

- Wie haben Sie die Instrumente des Bewertungskonzeptes in der Nachbesprechung genutzt?
 - ◆ Haben Sie die Nachbesprechung mit Hilfe des Konzeptes strukturiert oder die Instrumente im Anschluss an die eigentliche Nachbesprechung thematisiert und ausgefüllt?
- Haben Sie alle Instrumente genutzt?
 - ◆ Matrix
 - ◆ Dokumentation über die Nachbesprechung
 - ◆ Sammlung über das Repertoire
- Gab es für Sie wichtige Erkenntnisse, zentrale Erfahrungen beim Einsatz der Instrumente?
- Genauere Nachfragen zur Dokumentation über die Nachbesprechung:
 - ◆ Wer hat den Bogen ausgefüllt? War dieses Vorgehen sinnvoll/gewinnbringend?
 - ◆ Wann wurde er ausgefüllt? War dieses Vorgehen sinnvoll/gewinnbringend?
 - ◆ Wie haben Sie die Skalen zur Reflexionskompetenz und zu den Qualitätsbereichen des Unterrichts ausgefüllt?
 - ▲ Alleine oder dialogisch mit dem Auszubildenden? Haben Sie dieses Vorgehen als sinnvoll/gewinnbringend erlebt?
 - ▲ Konnten Sie eine Unterscheidung zwischen dem Ausbildungsstand und dem Examenstandard auf der Skala verorten? Konnte Sie diesen Unterschied schlüssig erklären? Nach welcher Logik haben Sie diese die Unterscheidung begründet?

▪ Schlussfragen:

- Würden Sie sagen, dass das Bewertungskonzept passend für den Einsatz in der Ausbildungspraxis ist?

[Frageblock 2: Bietet das Konzept eine Hilfestellung für die Planung und Analyse von Unterricht, bietet es Gesprächsanlässe?] nicht vorlesen!

▪ **Einstiegsfrage:**

- *Haben Sie den Eindruck, dass das Konzept den Referendaren eine Hilfestellung für erfolgreiche Unterrichtsbesuche geliefert hat?*

▪ **Überleitung:**

- *Inwiefern ergibt sich dies für sie?/Wie kommen Sie zu dieser Auffassung?*

▪ **Schlüsselfragen:**

- *Hatten Sie den Eindruck, dass sich die Referendare bei der Planung und der Analyse von Unterricht auf die Qualitätsbereiche des Unterrichts bezogen haben?*
- *Hat das Konzept genügend Freiräume für die Analyse und Bewertung des Unterrichts gelassen oder hat das Konzept den Blick eingeeengt?*
- *Hat Ihnen das Konzept dabei geholfen in der Nachbesprechung ein auf wesentliche Punkte fokussiertes Auswertungsgespräch zu führen?*
- *Wie hat es dabei geholfen?*
- *Hat es geholfen „die wesentlichen“ Qualitätsbereiche der Stunde, jene, denen im Rahmen der gezeigten Stunde ein besondere Bedeutung zukam, zu fokussieren?*
- *Wurde von Ihnen oder vom Referendar Bezug auf die in der Matrix festgehaltenen Qualitätsbereiche genommen?*

▪ **Schlussfragen:**

- *Würden Sie sagen, dass der Einsatz des Bewertungskonzeptes sinnvoll ist?*
- *Wie begründen Sie dies?*

[Frageblock 3: Kann das Konzept die Transparenz der Bewertungsprozesse sicherstellen?] nicht vorlesen!

▪ **Einstiegsfrage:**

- Konnten die Referendare Ihre Bewertungen nachvollziehen? Haben die Instrumente dabei geholfen, die Bewertung nachvollziehen zu können?

▪ **Überleitungsfrage:**

- Inwiefern ergibt sich dies für Sie?

▪ **Schlüsselfragen:**

- War die Matrix, war die Unterscheidung der Qualitätsbereiche hilfreich um Klarheit über Ihre jeweilige Bedeutung für die Qualität des gezeigten Unterrichts zu bekommen?
- Konnten Sie an Hand der Qualitätsbereiche schlüssig begründen, warum der Unterrichtsbesuch gut oder schlecht bewertet wurde?
- Haben die Instrumente, haben die unterschiedlichen Qualitätsbereiche geholfen einen Beratungsbedarf zu formulieren und Entwicklungsaufgaben zu fokussieren?
- Konnten Sie die Nachbesprechungen mit einer Bilanzierung abschließen?
- Welche Rolle hat dabei der Dokumentationsbogen für die Nachbesprechung gespielt?
- Hat sich Ihre Bewertungspraxis durch den Einsatz des Bewertungskonzeptes verändert?
 - ◆ Wie sah dieser Unterschied aus?
 - ◆ Haben Sie ihn als positiv erlebt?
- Hat Ihnen die Sammlung über das Repertoire dabei geholfen zu einer umfassenden Einschätzung über die Kompetenzentwicklung des Referendars zu kommen?

▪ **Schlussfragen:**

- Hat das Konzept dazu beigetragen, dass Sie und Ihre Referendare die Nachbesprechungen so gestalten konnten, dass die Referendare am Ende mit einem gefühlten Lernerfolg aus den Nachbesprechungen gegangen sind?
- Würden Sie anderen Seminaren bzw. ZfsL den Einsatz dieser Instrumente empfehlen?

Abschlussteil

Fragen:

- *Haben Sie noch wichtige Erfahrungen aus dem Einsatz der Instrumente, die im Laufe des Interviews nicht angesprochen wurden?*
- *Was ist Ihnen hinsichtlich Ihres Feedbacks besonders wichtig?*
- *Welchen Ratschlag würden Sie mir für die weitere Entwicklung der Instrumente mit auf den Weg geben?*

Leitfaden für das *problemzentrierte Interview* zur Erprobung der Bewertungsinstrumente mit Referendaren

Problemstellung:

- *Ist das Konzept in der Ausbildungspraxis einsetzbar, weist es eine Adaptivität mit der Praxis auf?*
- *Bietet das Konzept eine Hilfestellung für die Planung, Durchführung und Analyse von Unterricht, bietet es Gesprächsanlässe?*
- *Kann das Konzept die Transparenz der Bewertungsprozesse sicherstellen?*

Phasen des Interviews:

- Einführung (Zusammenfassung der Rahmenbedingungen und des Erkenntnisinteresses)
- Warm-Up (allg. Aspekte des Themas, Herstellung einer vertrauensvollen und intimen Gesprächsatmosphäre)
- Hauptteil (Bearbeitung der verschiedenen Themenblöcke)
- Abschlussteil (Zusammenfassung der Ergebnisse, stimmiger Ausstieg)

Grundregeln des Interviews:

- Im Zentrum steht die Wahrnehmung der Interviewten, der Interviewer gibt durch Fragen (problemzentrierte) Impulse
- Trichterfunktion: zunehmende thematische Verengung auf das Erkenntnisinteresse und die Themenblöcke nach offenem Beginn
- Schlüsselfragen in verschiedenen Themenblöcken mehrfach beantworten lassen
- Rückspiegelung der „gelernten“ Aspekte, damit die Interviewten vertiefen und ergänzen können
- möglichst offenes Design des Leitfadens und offene Formulierung der Fragen, um den Befragten die Möglichkeit zu geben, die Zusammenhänge in eigenen Worten zu entfalten

Einführung

- **Erkenntnisinteresse heute:** *Wie haben Sie den Praxiseinsatz der Bewertungsinstrumente erlebt, ist das Bewertungskonzept geeignet die Nachbesprechungen der Lehrproben positiv zu beeinflussen?*
- *Im Mittelpunkt heute steht ihre Perspektive auf die Instrumente, ausgehend von Ihrer Auszubildendenpersönlichkeit und Ihrem Ausbildungsfach, es gibt kein „richtig“ oder „falsch“, sondern Ihre Perspektive auf die vorgestellte Idee, Unterricht im Referendariat zu bewerten zählt.*
- *Im Mittelpunkt stehen die Ausbildungssituationen, in denen des Bewertungskonzept eingesetzt wurde, wird der Kontext auf andere Ausbilder und Ausbildungskontexte ausgedehnt, so wird dies erwähnt.*
- *Rückfragen von mir, ggf. auch überspitzte, dienen dazu den Inhalt ihrer Aussagen möglichst klar herauszuarbeiten*
- *Aufzeichnung lediglich zur Transkription des Interviews, Wiedergabe der Inhalte nur anonymisiert, Rückbezug auf Fächer kann ausgeführt werden*

Warm-Up

- *Schilderung von Erfahrungen über die Bewertung mit den Instrumenten (Probleme, Verbesserungen im Gegensatz zu ansonsten praktiziertem Vorgehen,....)*
 - *Wenn Sie auf den Einsatz des Bewertungskonzeptes zurückblicken, was fällt ihnen da spontan ein?*
 - *...ggf. im Unterschied zu den Lehrproben, die ohne das Konzept bewertet wurden, sowie die Nachbesprechungen in Ihrem anderen Fach.*

Hauptteil (questioning route)

[Frageblock 1: Ist das Konzept in der Ausbildungspraxis einsetzbar, weist es eine Adaptivität mit der Praxis auf?] nicht vorlesen!

▪ Einstiegsfrage:

- Wie war Ihr Eindruck: Konnte das Bewertungskonzept so wie es war in der Ausbildungspraxis eingesetzt werden oder gab es Probleme beim Einsatz?

▪ Überleitung:

- Inwiefern ergibt sich dies für sie?/Wie kommen Sie zu dieser Auffassung?
- War der zeitliche Aufwand, der für den Einsatz des Konzeptes nötig war, angemessen?

▪ Schlüsselfragen:

- Wurden die Instrumente des Bewertungskonzeptes in der Nachbesprechung genutzt?
- Wie wurden sie genutzt?
 - ◆ Leiteten sie die Strukturierung der Nachbesprechung oder wurden sie additiv im Anschluss an die eigentliche Nachbesprechung thematisiert?
- Wurden alle Instrumente genutzt
 - ◆ Matrix
 - ◆ Dokumentation über die Nachbesprechung
 - ◆ Sammlung über das Repertoire
- Genauere Nachfragen zur Dokumentation über die Nachbesprechung:
 - ◆ Wer hat den Bogen ausgefüllt? War dieses Vorgehen sinnvoll/gewinnbringend?
 - ◆ Wann wurde er ausgefüllt? War dieses Vorgehen sinnvoll/gewinnbringend?
 - ◆ Wie wurden die Skalen zur Reflexionskompetenz und zu den Qualitätsbereichen des Unterrichts ausgefüllt?
 - ▲ I.W. vom Ausbilder oder dialogisch mit dem Auszubildenden? War dieses Vorgehen sinnvoll/gewinnbringend?
 - ▲ Wurde die Unterscheidung zwischen dem Ausbildungsstand und dem Examenstandard schlüssig erklärt oder eher willkürlich gezogen? Nach welcher Logik wurde die Unterscheidung begründet?

▪ Schlussfragen:

- Würden Sie sagen, dass das Bewertungskonzept passend für den Einsatz in der Ausbildungspraxis ist?

[Frageblock 2: Bietet das Konzept eine Hilfestellung für die Planung und Analyse von Unterricht, bietet es Gesprächsanlässe?] nicht vorlesen!

▪ **Einstiegsfrage:**

- Haben Sie den Eindruck, dass das Konzept Ihnen eine Hilfestellung für erfolgreiche Unterrichtsbesuche geliefert hat?

▪ **Überleitung:**

- Inwiefern ergibt sich dies für sie?/Wie kommen Sie zu dieser Auffassung?

▪ **Schlüsselfragen:**

- Hat Ihnen das Konzept bei der Planung von Unterricht geholfen?
 - ♦ Wie hat es geholfen?
 - ♦ Hat es geholfen, „die wesentlichen“ Qualitätsbereiche der Stunde, jenen denen im Rahmen der gezeigten Stunde ein besondere Bedeutung zukam, zu fokussieren?
 - ♦ Welche Rolle hat dabei die Matrix der Qualitätsbereiche gespielt?
- Hat Ihnen das Konzept genügend Freiräume für die Planung und Durchführung Ihres Unterrichts gegeben oder haben Sie sich durch das Konzept eingeengt gefühlt?
- Hat Ihnen das Konzept dabei geholfen ein eigenes Statement zur Stunde abzugeben und in der Nachbesprechung produktiv über Ihre Stunde reflektieren zu können?
 - ♦ Wie hat es dabei geholfen?
 - ♦ Hat es geholfen „die wesentlichen“ Qualitätsbereiche der Stunde, jene denen im Rahmen der gezeigten Stunde ein besondere Bedeutung zukam zu fokussieren?
 - ♦ Wurde von Ihnen oder vom Ausbilder Bezug auf die in der Matrix festgelegten Qualitätsbereiche genommen?
 - ▲ Hat die Matrix Ihnen dabei geholfen eine gewinnbringend Reflexion der Unterrichtsstunde zu leisten? Inwiefern?

▪ **Schlussfragen:**

- Würden Sie sagen, dass der Einsatz des Bewertungskonzeptes sinnvoll ist?
 - ♦ Wie begründen Sie dies?

[Frageblock 3: Kann das Konzept die Transparenz der Bewertungsprozesse sicherstellen?] nicht vorlesen!

▪ **Einstiegsfrage:**

- Waren die Bewertungen Ihres Ausbilders für sie nachvollziehbar? Haben die Instrumente dabei geholfen, die Bewertung nachvollziehen zu können?

▪ **Überleitungsfrage:**

- Inwiefern ergibt sich dies für Sie?

▪ **Schlüsselfragen:**

- War die Matrix, war die Unterscheidung der Qualitätsbereiche hilfreich, um Klarheit über Ihre jeweilige Bedeutung für die Qualität des gezeigten Unterrichts zu bekommen?
- Konnte an Hand der Qualitätsbereiche schlüssig begründet werden, warum der Unterrichtsbesuch gut oder schlecht bewertet wurde?
- Haben die Instrumente, haben die unterschiedlichen Qualitätsbereiche geholfen einen Beratungsbedarf zu formulieren und Entwicklungsaufgaben zu fokussieren?
- Schloss eine Nachbesprechung mit einer für Sie zufriedenstellenden Bilanzierung, die sie nachvollziehen konnten?
 - ◆ Welche Rolle hat dabei der Dokumentationsbogen für die Nachbesprechung gespielt?
- Konnten Sie einen Unterschied zu den Bewertungen feststellen, in denen Ihr Ausbilder noch ohne das Konzept bewertet hat?
 - ◆ Wie sah dieser Unterschied aus?
- Gab es Unterschiede zur Bewertung in dem Fach, wo ohne das Bewertungskonzept bewertet wurde?
 - ◆ Würden Sie diesen Unterschied dem Bewertungskonzept zuschreiben?
- Konnten Sie die Bewertung in Ihrem Abschlussgutachten nachvollziehen?
 - ◆ Deckte sich die Bewertung mit der Einschätzung im Sinne der „Sammlung über das Repertoire“?
 - ◆ Hat die „Sammlung über das Repertoire“ Ihnen geholfen, zentrale Kompetenzen ihrer Lehrerpersönlichkeit zu erkennen, aber auch Entwicklungsfelder?
 - ◆ Zeigt sich in der Sammlung über das Repertoire ihre Kompetenzentwicklung?

▪ **Schlussfragen:**

- Hat das Konzept dazu beigetragen, dass Sie und Ihre Ausbilder die Nachbesprechungen so gestalten konnten, dass Sie am Ende mit einem gefühlten Lernerfolg aus den Nachbesprechungen gegangen sind?

Abschlussteil

Fragen:

- *Haben Sie noch wichtige Erfahrungen aus dem Einsatz der Instrumente, die im Laufe des Interviews nicht angesprochen wurden?*
- *Was ist Ihnen hinsichtlich Ihres Feedbacks besonders wichtig?*
- *Welchen Ratschlag würden Sie mir für die weitere Entwicklung der Instrumente mit auf den Weg geben?*

Transkriptionsregeln

- Es wird wörtlich transkribiert, also nicht lautsprachlich oder zusammenfassend. Vorhandene Dialekte werden möglichst wortgenau ins Hochdeutsche übersetzt. Wenn keine eindeutige Übersetzung möglich ist, wird der Dialekt beibehalten, zum Beispiel: Ich gehe heuer auf das Oktoberfest.
- Wortverschleifungen werden nicht transkribiert, sondern an das Schriftdeutsch angenähert. Beispielsweise wird aus „Er hatte noch so’n Buch genannt“ wird zu „Er hatte noch so ein Buch genannt“ und „hamma“ wird zu „haben wir“. Die Satzform wird beibehalten, auch wenn sie syntaktische Fehler beinhaltet, beispielsweise: „bin ich nach Kaufhaus gegangen.“
- Alle Angaben, die einen Rückschluss auf Personen zulassen, werden anonymisiert
- Wort- und Satzabbrüche sowie Stottern werden geglättet, bzw. ausgelassen, Wortdoppelungen nur erfasst, wenn sie als Stilmittel zur Betonung genutzt werden: „Das ist mir sehr, sehr wichtig.“. „Ganze“ Halbsätze, denen nur die Vollendung fehlt, werden jedoch erfasst und mit dem Abbruchzeichen/gekennzeichnet.
- Interpunktion wird zu Gunsten der Lesbarkeit geglättet, das heißt bei kurzem Senken der Stimme oder uneindeutiger Betonung, wird eher ein Punkt als ein Komma gesetzt. Dabei sollen Sinneinheiten beibehalten werden.
- Pausen werden durch drei Auslassungspunkte in Klammern (...) markiert.
- Besonders betonte Wörter oder Äußerungen werden durch GROSSSCHREIBUNG gekennzeichnet.
- Jeder Sprecherbeitrag erhält eigene Absätze. Zwischen den Sprechern gibt es eine freie, leere Zeile. Auch kurze Einwürfe, sofern sie aus mehr als zwei Wörtern bestehen, werden in einem separaten Absatz transkribiert. Sprechen zwei Personen durcheinander, so werden die Ausführungen in getrennten Absätzen transkribiert und es wird vermerkt, welche Personen durcheinander gesprochen haben. Mindestens am Ende eines Absatzes werden Zeitmarken eingefügt.
- Reine Verständnissignale der gerade nicht Sprechenden werden nicht transkribiert (Beispiel: Sprechende Person schaut in die Runde und sucht nach Zustimmung, andere Person nickt daraufhin).
- Nonverbale und verbale Äußerungen der gerade nicht Sprechenden, welche auf Unterstützung oder Ablehnung einer Aussage hindeuten (etwa wie lachen, seufzen, kopfschütteln, nicken, bejahende/verneinende Lautäußerungen) werden beim Einsatz in Klammern notiert, z.B. R3: „mhm“ (bejahend), nicken.

- Unverständliche Wörter werden mit (unv.) gekennzeichnet. Vermutete Wortlaute, bzw. Satzteile werden mit einem Fragezeichen in Klammern gesetzt, z.B. (Xylomethanolin?). Generell werden alle unverständlichen Stellen mit einer Zeitmarke versehen, wenn innerhalb von einer Minute keine Zeitmarke gesetzt ist.
- Die interviewende Person wird durch ein „I:“, die befragte Person durch ein „A/R:“ gekennzeichnet. Bei mehreren Interviewpartnern in Gruppendiskussionen wird dem Kürzel „A/R“ eine entsprechende Kennnummer zugeordnet (z.B. A1 für Ausbilder, R1 für Referendare))

„Bonner Qualitätstableau“

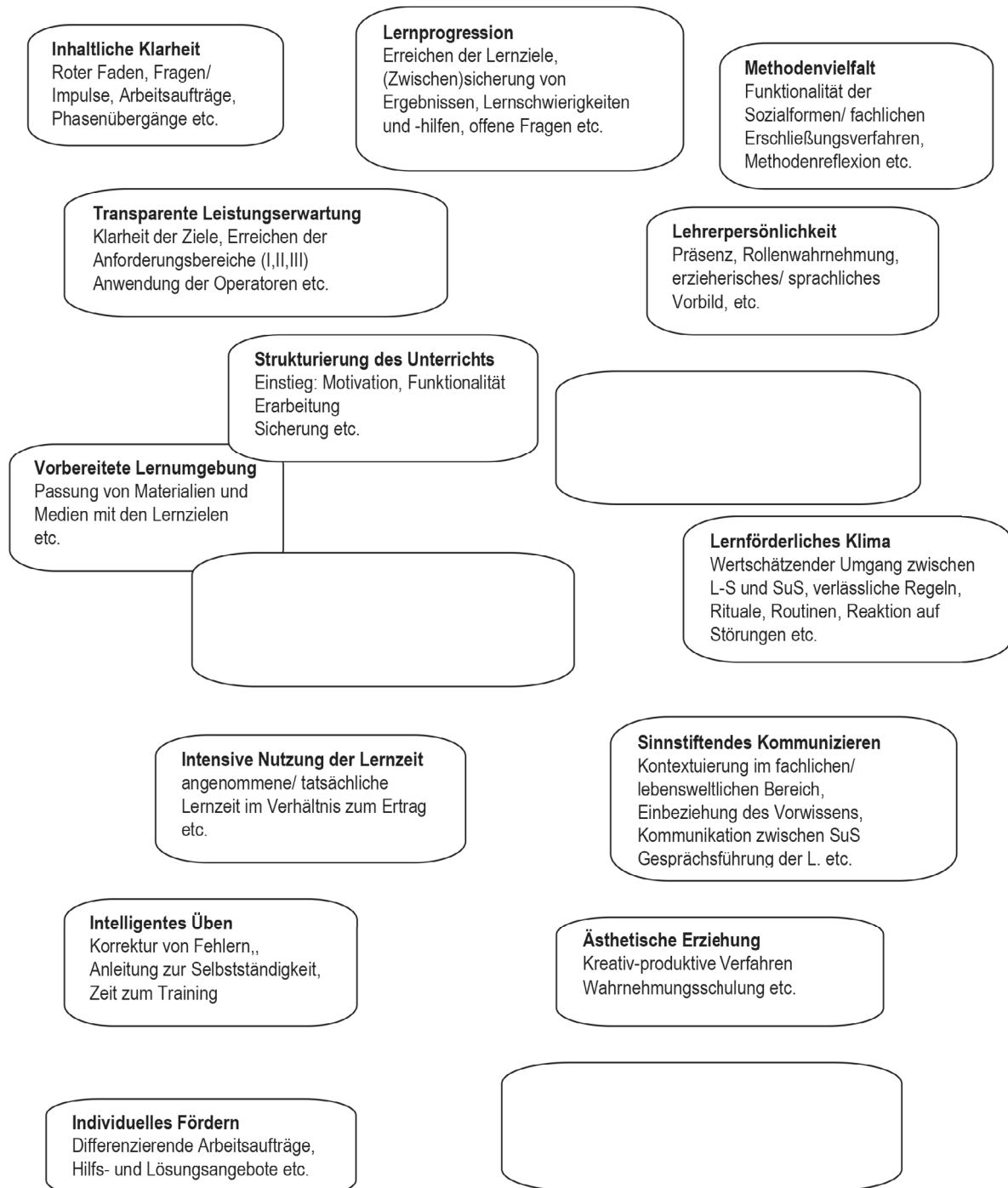
Guter Unterricht im Seminar Bonn wird besonders gut sichtbar:

Kriterien guten Unterrichts nach H. Meyer	... bei der Planung	Durchführung von Unterricht (Bildungs- u. Erziehungsauftrag von Richtlinien, Lehrplänen u. Schulcurricula kennen und Thema sachangemessen u. schülerorientiert kommunizieren)	Rahmenvorgabe (Bezogen auf Unterricht)
1. klare Strukturierung		<ul style="list-style-type: none"> • Klarheit und Akzentuierung von Thema, Ziel und Prozess • Transparente und stimmige Phasierung (= Artikulation) • Vernetzung der Unterrichtsphasen 	
2. Inhaltliche Klarheit und Fachlichkeit		<ul style="list-style-type: none"> • Verständliche Themenbezüge herstellen • Orientierenden Arbeitsrahmen festlegen • Aufgabenstellung verständlich und klar vermitteln • Aufmerksamkeit für Thema und Problemstellung erzielen 	
3. Sinnstiftendes Kommunizieren		<ul style="list-style-type: none"> • Verknüpfung mit vergangener Themenstellung und Ausblick auf neue Themenstellung • Vorwissen anschließen • Metakommunikation gewinnbringend in den Prozess einbringen • Erhellung der Bedeutung des Themas für das Selbst- bzw. Weltverständnis 	<ol style="list-style-type: none"> 1. Unterrichten 2. Erziehen 3.
4. Anteil echter Lernzeit/Selbsttätigkeit der Lernenden	<ol style="list-style-type: none"> 1. Diagnose-, Entscheidungs- u. Begründungskompetenz (didaktisch-methodisch, päd.-psycholog) 2. Fachkompetenz (Inhalte, Methoden) 3. Reflexionskompetenz 1. Empathiefähigkeit <p>→ fachspezifisch auszufordern insbesondere bezüglich didaktischer Reduktion und sachlogischer Materialauswahl</p>	<ul style="list-style-type: none"> • Eigenaktive und zielbezogene Lernzeiten ermöglichen • Individuelle Lernwege ermöglichen und unterstützen • Zu eigenen Denkprozessen und Lösungswegen ermutigen • Lernzuwachs fördern 	<ol style="list-style-type: none"> 4. Diagnostizieren und Fördern 5. Beraten 6. Leistung messen und beurteilen 7.
5. Lernförderliches Klima		<ul style="list-style-type: none"> • Anregende Lernumgebungen gestalten • Regeln und Rituale verankern • Wertschätzung aller Lernenden • Angemessene Reaktionen auf Unterrichtsstörungen • Klima der Fehlerfreundlichkeit • Reibungslose Organisation gewährleisten • Methodenreflexion in den Unterricht einbringen • Balance von Variabilität und Konstanz in der Mediatorschließung • Vielfalt von Inszenierungstechniken funktional einsetzen • Vielfalt von Handlungsmustern funktional einsetzen, u. a. Unterrichtsgespräche zielgerichtet und interaktionsfördernd führen 	<ol style="list-style-type: none"> 6. Organisieren und verwalten 7. Evaluieren, Innovieren und Kooperieren
6. Methodenvielfalt			

doc 14,5

<p>7. Individuelles Fördern</p>		<ul style="list-style-type: none"> • verschiedene Lernkanäle, Lernwege, Sozialformen, Präsentationsformen, Kontrollformen, Materialien, Aufgaben auf Grundlage diagnostizierter Heterogenität ermöglichen bzw. nutzen (z.B. Schwierigkeiten wahrnehmen und angemessen darauf reagieren) • Differenzierte Hilfs- u. Lösungsangebote (Anspruchsniveau), Berücksichtigung des individuellen Lern- u. Arbeitstempos • Differenziertes Angebot eigenverantwortlicher/kooperativer Lernformen innerhalb/außerhalb von Unterricht (Referat, Portfolio, Lerntagebuch/Lerntteams, Lerntandems etc.) 	<p>Rahmenvorgabe (bezogen auf Unterricht)</p>
<p>8. Intelligentes Üben</p>	<p>bei der Planung</p>	<ul style="list-style-type: none"> • Interesse wecken, Vielfalt ermöglichen (Binnendifferenzierung in Darstellung/Verbalisierung, Lernprogression) • Transparente Unterscheidung von Lern-, Förder- und Leistungssituationen • ausreichend Zeit für das Üben; Anleiten zur Selbstständigkeit • thematische und/oder methodische Vernetzung der Übung(sphase) • Vermittlung von Lernstrategien, Rückmeldung von Übungsleistungen • Konstruktiver Umgang mit Fehlern (Fehlerdiagnose als Übungsanlass) 	<p>1. Unterrichten</p> <p>2. Erziehen</p> <p>3. Diagnostizieren und Fördern</p> <p>4. Beraten</p> <p>5. Leistung messen und beurteilen</p> <p>6. Organisierer und verwalten</p> <p>7. Evaluieren, Innovieren und Kooperieren</p>
<p>9. Transparente Leistungserwartungen</p>		<ul style="list-style-type: none"> • Offenlegung zentraler LZ u. Leistungserwartungen; ggf. Modifikation (Anspruchsniveau) • Offenlegung von Bewertungskriterien, Beurteilungsbereiche/-gegenstände, (verschiedene) Formen d. Leistungsermittlung/ -dokumentation • klare \wedge (Operatoren, sichere Verwendung) u. angemessene Zeitr Vorgaben • Feedback (L-SS; SS-SS): situativ, regelmäßig, institutionalisiert 	
<p>10. Vorbereitete Umgebung</p>	<p>Diagnose-, Entscheidungs- u. Begründungskompetenz</p>	<ul style="list-style-type: none"> • Materialien/strukturierte Ausstattung, Sitzordnung • Visualisierung von Regeln und Ritualen (Sicherheit) • Klare, störungsfrei wahrnehmbare u. lernprozessbezogene Visualisierungen 	
<p>11. Lernergebnis/Progression</p>	<p>Fachkompetenz (Inhalte, Methoden)</p> <p>Reflexionskompetenz u. Empathiefähigkeit</p>	<ul style="list-style-type: none"> • deutliche Markierung/Strukturierung/Gewichtung von (Teil-)Ergebnissen im Lernprozess • Herstellung und Sicherung weiterführender (verbaler, medialer, ...) Lernergebnisse durch fachlich, fachmethodisch, terminologisch angemessene Verfahren (Präsentation, Visualisierung, reflexive Rückschau) • angemessene schriftliche Fixierungen 	
<p>12. Kreative Gestaltung ästhetische Erziehung</p>	<p>→ fachspezifisch ausdifferenzieren insbesondere bezüglich didaktischer Reduktion und sachlogischer Materialauswahl</p>	<ul style="list-style-type: none"> • Kreativ-produktive Verfahren u. Materialienverwendung • Differenzierte Wahrnehmungsschulung und Handlungsorientierung 	
<p>13. LehrerInnen-Persönlichkeit / Wahrnehmung der Lehrerrolle</p>		<ul style="list-style-type: none"> • Wertschätzendes Kommunizieren bzw. Haltung (verbal u. non-verbal; aktiv u. rezeptiv) • Begeisterungsfähigkeit, Humor u. Selbstironie • sprachliches Vorbild • Präsenz u. Durchsetzungsfähigkeit 	
<p>... bei der Reflexion</p>			
<p>Erkennt nicht – erkennt in Ansätzen – erkennt selbstständig und ordnet ein – entwickelt (mit Hilfe) Alternativen</p>			

„Bubbles“ des ZfsL Bonn



Dokumentationsbogen für die Nachbesprechung des ZfsL Bonn

Unterrichtsnachbesprechung – Statement nach der Stunde

Sie haben nach jeder Lehrprobe und später auch nach jeder UPP ca. 10' Zeit, ein Statement vorzubereiten. Das Statement sollte nicht länger als 5-10' min dauern. Dabei zeigen Sie Ihre Fähigkeit, Ihre geplante und durchgeführte Stunde kritisch zu reflektieren. Das Statement sollte gezielt eingehen auf

- a) Planungsentscheidungen, die im schriftlichen Unterrichtskonzept nicht thematisiert wurden, und
- b) Gesichtspunkte, die die Qualität der Stunde beeinflusst haben. Die Darstellung muss - statt additiv beschreibend - nach Aspekten gegliedert, argumentativ und bewertend erfolgen.

1. Überlegen Sie, welche Aspekte in der Stunde von zentraler Bedeutung waren (Anregungen¹ dazu S.2). **Sie können auch eigene Gesichtspunkte wählen** (freie Felder)

2. Gliedern Sie Ihr Statement!

Statement:

Was ich noch zu einem späteren Zeitpunkt besprechen möchte:

Beratungsbedarf habe ich zu folgenden Gesichtspunkten:

Matrix der Qualitätsbereiche guten Unterrichts, reduzierte Version (Beta-Version)

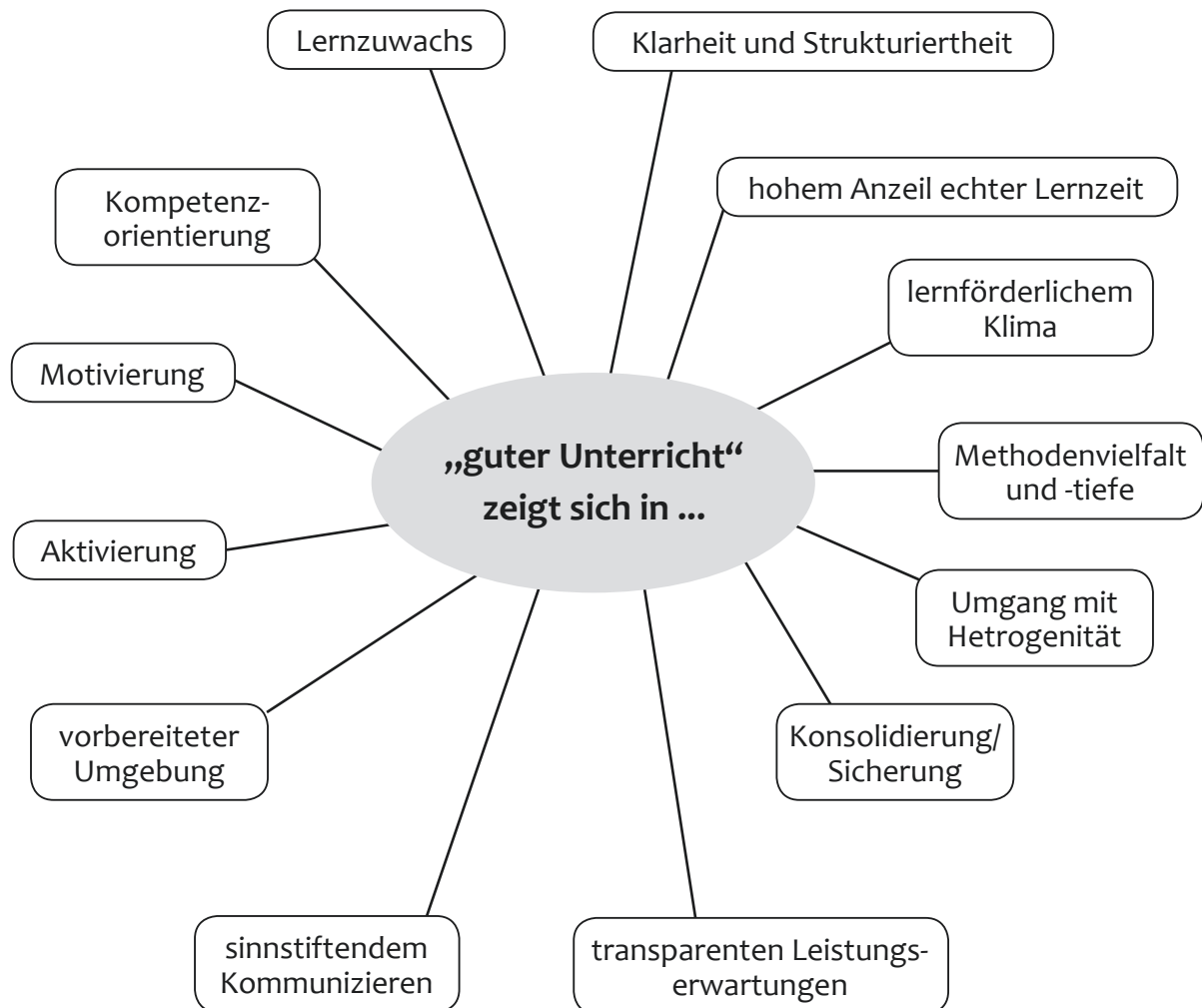
- Die Matrix der Qualitätsbereiche guten Unterrichts erfüllt folgende Funktion:
- eine theoretische Fundierung, ein „idealtypisches“ Modell für das hier vertretene Verständnis von „gutem Unterricht“ zu schaffen, ohne unter den angegebenen Qualitätsbereichen eine Hierarchie zu unterstellen
- einen Orientierungsrahmen sowohl für die Planung von Unterricht als auch für die Beobachtung und reflektierende Auswertungsgespräche über Unterricht anzubieten

Qualitätsbereich ▪ Quellen der Merkmale und Indikatoren	Erklärung, beispielhafte(!) Operationalisierung in Einzelmerkmalen
Klarheit und Strukturiertheit ▪ klare Strukturierung (Meyer) ▪ inhaltliche Klarheit (Meyer) ▪ Klassenführung (Helmke) ▪ Klarheit und Strukturiertheit (Helmke)	▪ Stimmigkeit von Zielen, Inhalten und Methoden ▪ Folgerichtigkeit des methodischen Gangs (Ausmaß der Lenkung und der emotionalen Nähe, Komplexität und Abstraktion sind der Lernsituation angepasst) ▪ Folgerichtigkeit des thematischen Gangs (sinnvolle didaktische Inszenierung der Inhalte nach (fach-)didaktischen Prinzipien) ▪ methodischer Grundrhythmus ▪ Aufgaben, Regel- und Rollenklarheit ▪ Aufgabenstellung verständlich und in sich stimmig hinsichtlich der Ziel-, Inhalts- und Methodenwahl (Lernstrukturanalyse; welche Lernschritte muss ein Schüler vollziehen um zu einem Ergebnis zu kommen? Lernstandanalyse; verfügen die Lernenden über die notwendigen Ressourcen zum Lösen der Aufgabe?) ▪ Klarheit der Ergebnissicherung (die Unterrichtsstunde bringt einen klar zu benennenden Erkenntnisfortschritt) ▪ Verbindlichkeit der Ergebnissicherung (Klarheit über definitiv zu fixierende Wissens- und Kompetenzbestände der Lerngruppe)
hoher Anteil echter Lernzeit ▪ hoher Anteil echter Lernzeit (Meyer) ▪ Klassenführung (Helmke)	„time on task“ (tatsächlich aufgewendete Lernzeit) Hohes Momentum (Vermeidung unnötiger Unterbrechungen des Unterrichtsflusses, sowie von Hektik, Herstellen eines „geschmeidigen“ Unterrichtsverlaufs)
lernförderliches Klima ▪ lernförderliches Klima (Meyer) ▪ lernförderliches Klima (Helmke)	▪ gegenseitiger Respekt der Lernenden /zw. Lernenden und Lehrer ▪ verlässliche Regeln ▪ gemeinsame Verantwortung von Lehrer und Lernenden für erfolgreichen Lernprozess (Arbeitsvertrag) ▪ Gerechtigkeit und Fürsorge des Lehrers gegenüber den Lernenden ▪ konstruktiver Umgang mit Fehlern, auf Lehrer wie auf Lernenden-Seite ▪ Balance zwischen Lern- und Leistungssituationen ▪ Balance zwischen „Lockerheit“ und „Anspruch“ ▪ Schülerfeedback

Qualitätsbereich ■ Quellen der Merkmale und Indikatoren	Erklärung, beispielhafte(!) Operationalisierung in Einzelmerkmalen
Methodenvielfalt und -tiefe ■ Methodenvielfalt (Meyer) ■ Angebotsvielfalt (Helmke)	<ul style="list-style-type: none"> ■ der Unterricht wird auf vielfältige Art und Weise inszeniert (Aktions- und Sozialformen, Medieneinsatz, Handlungsformen, methodische Großformen, Lernorte) ■ die eingesetzten Unterrichtsmethoden weisen eine Passung auf die Lerngruppe und die intendierten Lernziele auf ■ die eingesetzten Methoden leiten zur tiefen fachlichen Durchdringung an
Umgang mit Heterogenität ■ individuelles Fördern (Meyer) ■ Umgang mit Heterogenität (Helmke)	<ul style="list-style-type: none"> ■ Unterricht den verschiedenen Lernvoraussetzungen der Lernenden anpassen (Adaptivität/Passung) ■ Lernenden verschiedener Leistungsbereiche, sowohl leistungsschwachen als auch leistungsstarken, wertschätzend begegnen ■ Angebote zum Erwerb von Lernstrategien
Konsolidierung/Sicherung ■ intelligentes Üben (Meyer) ■ Konsolidierung/Sicherung (Helmke)	<ul style="list-style-type: none"> ■ in den Unterricht werden intelligente Übungsphasen eingebaut (lernstandsadäquate Entwicklung von Lernstrategien, Rhythmisierung der Übung)
transparente Leistungserwartungen ■ transparente Leistungserwartungen (Meyer) lernförderliches Klima (Helmke)	<ul style="list-style-type: none"> ■ Lernangebote bereitzustellen die den Ansprüchen der Richtlinien, den Bildungsstandards und den Fähigkeiten der Lernenden entsprechen ■ Transparenz über Lernziele ■ Rückmeldungen über den Lernfortschritt geben
sinnstiftendes Kommunizieren (Meyer)	<ul style="list-style-type: none"> ■ Kommunikationsprozess der „persönlichen qualitativen Sinngebung“ des Gelernten für die Lernenden
vorbereitete Umgebung (Meyer)	<ul style="list-style-type: none"> ■ ordentliche, mit sinnvollem Lernmaterial ausgestattete Räume, zu denen Lehrer und Lernende ein Gefühl von „Ownership“ haben ■ Belüftung, Beleuchtung und Akustik sind stimmig ■ ausreichend Platz für verschiedene Inszenierungstechniken des Unterrichts und den Bewegungsdrang der Lernenden
Aktivierung (Helmke)	<ul style="list-style-type: none"> ■ kognitive Aktivierung im Sinne der Selbststeuerung des Lernens, es erfolgt eine „tiefe“, anstatt bloß oberflächlicher „mechanischer“ Verarbeitung des Stoffes (Lernstrategien) ■ soziale Aktivierung (kooperative Arbeitsformen) ■ aktive Teilhabe
Motivierung (Helmke)	<ul style="list-style-type: none"> ■ extrinsische und intrinsische Motivationsgründe bewegen die Lernenden zum mitarbeiten ■ anspruchsvolle Lehrererwartungen

Qualitätsbereich ▪ Quellen der Merkmale und Indikatoren	Erklärung, <i>beispielhafte(!)</i> Operationalisierung in Einzelmerkmalen
Kompetenzorientierung (Helmke)	<ul style="list-style-type: none">▪ die Gestaltung des Unterrichts orientiert sich an den am Ende des Lernprozesses zu erreichenden Kompetenzfortschritten der Lernenden▪ der erfolgreiche Kompetenzerwerb wird evaluiert
Lernzuwachs (QA NRW)	<ul style="list-style-type: none">▪ ein Großteil der Lernenden erreicht einen intendierten Lernzuwachs inhaltlicher, methodischer oder sozialer Art▪ der Lernzuwachs wird dokumentiert, von den Lernenden gezeigt

Qualitätsbereiche guten Unterrichts



Unterrichtskompetenz erfassen und entwickeln

David Luhr

Das Forschungsprojekt *Unterrichtskompetenz erfassen und entwickeln* nimmt die beklagten Missstände der Bewertung in der zweiten Phase der Lehrerbildung zum Anlass, um im Rahmen von Aktionsforschung ein Bewertungskonzept für Unterrichtsbesuche zu entwickeln. Ausgehend vom Forschungsstand bezüglich der Probleme von Bewertungsprozessen in der zweiten Phase nutzt das Projekt Methoden der qualitativen Sozialforschung, um gemeinsam mit den betroffenen Akteuren die Merkmale eines „guten Bewertungskonzeptes“ zu erarbeiten. Aufbauend auf den Konzepten guten Unterrichts von Meyer und Helmke wurde ein standardisiertes Bewertungskonzept mit drei Bewertungsinstrumenten entwickelt, die einem Praxistest unterzogen wurden. In anschließenden Interviews berichteten Ausbilder und Auszubildende über die Veränderungen, welche die Bewertungsprozesse durch den Einsatz des Konzeptes erfahren haben. Das zwiespältige Fazit: „Die Standardisierung ist auch nicht mehr als ein Arbeitsinstrument!“

ISBN 978-3-8405-0134-0

EUR 28,80



9 783840 501340