



Westfälische Wilhelms-Universität Münster
Fachbereich Psychologie, Institut IV

Prof. Dr. Heinz Holling

Reliabilität und Validität von Verfahren der Präferenzmessung

**Ein meta-analytischer Vergleich verschiedener Verfahren
der Conjoint-Analyse**

**Inaugural-Dissertation
zur Erlangung des Doktorgrades der
Philosophischen Fakultät
der Westfälischen Wilhelms-Universität Münster**

**vorgelegt von
Diplom-Psychologe Marcus Heidbrink
aus Münster**

August 2006

Tag der mündlichen Prüfung: 19.09.2006 und 23.11.2006

Dekan: Prof. Dr. Dr. h.c. Wichard Woyke

Referent: Prof. Dr. Heinz Holling

Korreferent: Prof. Dr. Dr. h.c. Bernd Schäfer

Inhaltsverzeichnis

Abbildungsverzeichnis	6
Tabellenverzeichnis	7
1. Einleitung	9
2. Konzept und Verfahren der Präferenzstrukturmessung	13
2.1 Multiattribute Präferenzstrukturmessung	13
2.2 Zentrale Begrifflichkeiten	16
2.3 Modellierung multiattributer Präferenzen	20
2.3.1 Das Präferenzstrukturmodell.....	20
2.3.2 Nichtkompensatorische Entscheidungsregeln	21
2.3.3 Kompensatorische Entscheidungsregeln	22
2.4 Verfahren der Präferenzstrukturmessung	23
2.4.1 Kompositionelle Nutzenmessung	23
2.4.2 Dekompositionelle Nutzenmessung	26
2.4.3 Hybride Nutzenmessung	33
2.5 Die Weiterverarbeitung der individuellen Präferenzstrukturen	37
2.5.1 Die Segmentierungsverfahren	38
2.5.2 Die Wahlanteilssimulatoren.....	39
3. Die Reliabilität und Validität der Präferenzstrukturmessung	42
3.1 Die Generalisierbarkeitstheorie als Rahmenkonzept zur Reliabilitätsbestimmung	42
3.2 Übliche Vorgehensweisen zur Reliabilitätsbestimmung in der Präferenzmessung	46
3.3 Übliche Vorgehensweisen zur Validitätsbestimmung in der Präferenzmessung	51
3.3.1 Validitätsarten	51
3.3.2 Kategorisierung der in der Präferenzmessung verwendeten Validitätsarten	52
3.3.3 Die Maßzahlen der in der Präferenzmessung verwendeten Validitätsarten	55

4. Die Meta-Analyse als systematisches Verfahren zur Integration von Forschungsergebnissen	62
4.1 Die bisherige, zusammenfassende Forschung	62
4.2 Die Zielsetzung und die Vorteile einer Meta-Analyse	62
4.3 Das Vorgehen bei einer Meta-Analyse	64
4.4 Die statistischen Vorgehensweisen bei einer Meta-Analyse	66
4.4.1 Korrelationskoeffizienten als Effektgrößen	66
4.4.2 Der Ansatz nach Hedges und Olkin	68
4.4.3 Der Ansatz nach Hunter und Schmidt	70
4.4.4 Integration der unterschiedlichen Effektgrößen der Messgütebestimmung der Präferenzmessverfahren.....	75
5. Aufbau und Durchführung der Meta-Analyse	78
5.1 Spezifikation der Forschungsfrage und Definition der Charakteristika der relevanten Forschungsliteratur	78
5.2 Literatursuche	78
5.3 Kodierung der Studienmerkmale	80
5.4 Analyse der Datenbasis	81
5.5 Vorgehen bei der Datenanalyse und bei der Moderatorenanalyse	86
6. Ergebnisse der Meta-Analyse	87
6.1 Ergebnisse der Meta-Analyse für Effektgrößen der Validität	87
6.1.1 Ergebnisse der Meta-Analyse für Validitätsmaße auf der individuellen Ebene	87
6.1.2 Einflussfaktoren auf die Höhe der Validität	93
6.1.3 Ergebnisse der Meta-Analyse für Validitätsmaße auf der aggregierten Ebene	99
6.2 Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität.....	102

7. Zusammenfassende Diskussion und Ausblick.....	106
7.1 Zusammenfassung der Ergebnisse der Meta-Analyse	106
7.1.1 Zusammenfassung der Ergebnisse der Meta-Analyse zur Reliabilität	106
7.1.2 Zusammenfassung der Ergebnisse der Meta-Analyse zur Validität	107
7.2 Diskussion der Ergebnisse der Meta-Analyse	111
 Anhang.....	 115
 Literaturverzeichnis.....	 137

Abbildungsverzeichnis

Abb. 3.1 Modellanpassung und Reliabilität bei der Präferenzstrukturmessung 49

Abb. 3.2 Kategorisierung der in der Präferenzstrukturmessung verwendeten Arten der
Validitätsbestimmung 53

Abb. 3.3 Modellanpassung, Reliabilität und Validität bei der Präferenzstrukturmessung 54

Abb. 3.4 Grafische Übersicht über die in der Meta-Analyse verwendeten Arten und
Maßzahlen der Kriteriumsvalidität der Präferenzmessung 60

Abb. 5.1 Übersicht über die Größe der Untersuchungsdesigns der in der Meta-Analyse
berücksichtigten Effektgrößen 84

Tabellenverzeichnis

Tab. 3.1 Tabellarische Übersicht über die Häufigkeiten der in der Meta-Analyse verwendeten Effektgrößen der Reliabilitätsbestimmung der Präferenzmessung 50

Tab. 3.2 Tabellarische Übersicht über die in der Meta-Analyse verwendeten Arten und Maßzahlen der Kriteriumsvalidität der Präferenzmessung 60

Tab. 3.3 Tabellarische Übersicht über die Häufigkeiten der in der Meta-Analyse verwendeten Effektgrößen der Kriteriumsvalidität der Präferenzmessung 61

Tab. 5.1 Übersicht über die Herkunft der in der Meta-Analyse verwendeten Effektgrößen . 79

Tab. 5.2 Übersicht über das Erscheinungsjahr der in der Meta-Analyse verwendeten Effektgrößen..... 81

Tab. 5.3 Übersicht über die untersuchten Objektbereiche der in der Meta-Analyse verwendeten Effektgrößen 82

Tab. 5.4 Übersicht über die Merkmalsgewinnung bei den Untersuchungen, die in die Meta-Analyse eingegangen sind 83

Tab. 5.5 Übersicht über die Anzahl der Wahlalternativen in den Hold-out-Szenarien für die in der Meta-Analyse verwendeten Effektgrößen 85

Tab. 6.1 Ergebnisse der Meta-Analyse für Effektgrößen der Validität 87

Tab. 6.2 Ergebnisse der Meta-Analyse auf der alleinigen Basis der Hitraten-Effektgrößen . 89

Tab. 6.3 Ergebnisse der Meta-Analyse auf der alleinigen Basis des Korrelationskoeffizienten Pearson's r 89

Tab. 6.4 Ergebnisse der Meta-Analyse auf der alleinigen Basis von Spearman's rho..... 90

Tab. 6.5 Ergebnisse der Meta-Analyse auf der alleinigen Basis von Kendall's tau 90

Tab. 6.6 Ergebnisse der Meta-Analyse auf der alleinigen Basis der Korrelationskoeffizienten Pearson's r, Spearman's rho und Kendall's tau 91

Tab. 6.7 Ergebnisse der Meta-Analyse auf der alleinigen Basis der Hitraten-Effektgrößen korrigiert für Hold-out-Wahlsets mit maximal zehn Alternativen 92

Tab. 6.8 Korrelation ausgewählter Faktoren mit den Effektgrößen der Validität..... 94

Tab. 6.9 Durchschnittliche Validität in Abhängigkeit der Publikationsart 94

Tab. 6.10 Durchschnittliche Validität in Abhängigkeit der Probandenart 95

Tab. 6.11 Durchschnittliche Validität in Abhängigkeit des Objektbereichs der Präferenzmessung 96

Tab. 6.12	Durchschnittliche Validität in Abhängigkeit der Art der Individualisierung der Teilnutzenwerte bei der wahlbasierten Conjoint-Analyse	97
Tab. 6.13	Durchschnittliche Validität in Abhängigkeit der Verwendung eines Segmentierungsverfahrens	98
Tab. 6.14	Ergebnisse der Meta-Analyse für die Effektgröße der mittleren absoluten Abweichung (MAE) zwischen vorhergesagten und beobachteten Wahlanteilen	99
Tab. 6.15	Ergebnisse der Meta-Analyse für die Effektgröße der Wurzel der mittleren quadrierten Abweichung (RMSE) zwischen vorhergesagten und beobachteten Wahlanteilen.....	100
Tab. 6.16	Ergebnisse der Meta-Analyse für die Effektgröße der mittleren absoluten Abweichung (MAE) in Abhängigkeit des Verfahrens zur Wahlanteilssimulation...	101
Tab. 6.17	Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität	102
Tab. 6.18	Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität auf der Input-Ebene.....	103
Tab. 6.19	Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität auf der Output-Ebene.....	104
Tab. 6.20	Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität in Abhängigkeit der Art der Reliabilitätsmessung.....	105

1. Einleitung

Der Messung von Präferenzen kommt eine wichtige Bedeutung bei der Untersuchung und Prognose von Entscheidungsverhalten von Individuen und Gruppen zu. Umfragen in Europa und den USA mit Anwendern in Wissenschaft und Praxis lassen darauf schließen, dass weltweit jährlich mehrere tausend Studien zur Präferenzmessung durchgeführt werden (Wittink & Cattin, 1989; Wittink, Vriens & Burhenne, 1994; Melles & Holling, 1998). Es gibt mehr als 1.300 wissenschaftliche Publikationen zu diesem Thema (http://www.geocities.com/melles99/melles_e.html).

Die Popularität von Präferenzmessungen resultiert aus der Annahme, dass sich mit der Kenntnis der Präferenz eines Menschen gute Vorhersagen über dessen Auswahl- und Entscheidungsverhalten treffen lassen. Im bedeutendsten Anwendungsgebiet der Präferenzmessmethoden, dem Marketing, werden die Präferenzstrukturen von Antwortpersonen genutzt, um Prognosen über das Kaufverhalten von Konsumenten zu treffen, Produkte zu entwickeln, Preise festzulegen und Marktsegmentierungen oder Marktanteilsprognosen vorzunehmen.

Darüber hinaus finden Methoden der Präferenzmessung unter anderem Verbreitung in den Themengebieten des Personalmanagements, beispielsweise im Personalmarketing (Wiltinger, 1997), in der Personalauswahl (Niebergall & Schulz, 1996), in der Vergütungsberatung (Leptien, 1995), bei der Generierung von Anforderungsprofilen (Willms, 2001) oder zur Identifikation von High-Potentials (Franke, 1999).

Wie bei jedem anderen Messinstrument stellt sich auch im Rahmen der Präferenzmessung die Frage, wie reliabel und valide das jeweilige Erhebungsverfahren misst. Angesichts der bedeutenden Ableitungen, welche in den Anwendungsgebieten aus den erhobenen Präferenzen resultieren, erscheint es nicht verwunderlich, dass sich zahlreiche Forscher mit der Bestimmung der Reliabilität und der Validität von Verfahren der Präferenzmessung befasst haben. In der dieser Arbeit zugrunde liegenden Literaturrecherche konnten mehrere hundert Studien von mehr als einhundertundfünfzig unterschiedlichen Forschern identifiziert werden, welche Angaben über die Reliabilität oder die Validität von Präferenzmessverfahren enthielten.

Um den großen Umfang an Publikationen und Ergebnissen über die Güte der Ergebnisse von Präferenzmessverfahren zu strukturieren und anderen Forschern den jeweiligen Status quo der wissenschaftlichen Erkenntnis rasch zugänglich zu machen, haben einzelne Forscher ein Review der vorhandenen Literatur vorgenommen. Die größte Beachtung fanden die Arbeiten von Green und Srinivasan (1978, 1990), die wesentlich zur Verbreitung und Popularität der Methoden der Präferenzmessung, und hier insbesondere der Conjoint-Analyse, beigetragen haben. Im Bereich der Reliabilitätsbestimmung kommt der Arbeit von Bateson, Reibstein und Boulding (1987) zentrale Bedeutung zu. Ein neueres Review zur Validität einiger Verfahren der Präferenzmessung stammt von Sattler und Hensel-Börner (2001).

Die wissenschaftliche Methode des narrativen Reviews bietet interessierten Forschern einen guten Einstieg in ein neues Thema und informiert überblicksartig über die bisherigen Entwicklungen. Sie stellt eine übliche Form der Zusammenfassung des gegenwärtigen Wissenstands in einem Forschungsgebiet dar.

Andererseits lässt sich an der Methode des Reviews kritisieren, dass sie nicht in einem unwesentlichen Maß den subjektiven Einflüssen des Forschers unterliegt und dem Forscher bei der Auswahl und der Gewichtung der zu berücksichtigenden Studien zu viele Freiheiten lässt.

Zahlreiche Forscher empfehlen aus theoretischen und empirischen Erkenntnissen heraus, kein narratives Review, sondern eine Meta-Analyse als wissenschaftliche Methode zur Zusammenfassung und Integration von Forschungsergebnissen eines Themengebiets zu verwenden (z. B. Beaman, 1991; Schulze, Holling & Böhning, 2003).

In der vorliegenden Arbeit wird erstmalig eine Meta-Analyse durchgeführt, um die bisherigen Erkenntnisse zur Prognosegüte und zur Zuverlässigkeit von Verfahren der Präferenzmessung zusammenzufassen und die Reliabilitäts- bzw. die Validitätsdaten der zahlreichen Einzelstudien einer gemeinsamen, statistischen Auswertung zu unterziehen.

Die Meta-Analyse ist ein vor allem in der Psychologie und der Erziehungswissenschaften vor rund dreißig Jahren entwickeltes Verfahren zur systematischen Zusammenfassung von inhaltlich verwandten Studien eines Themengebiets. Die Meta-Analyse hat seit seiner Entwicklung in den 1970er Jahren eine weite Verbreitung in den Sozialwissenschaften erreicht und kann gerade in der Psychologie als Standardverfahren zur Integration von Forschungsergebnissen gelten.

Schulze (2004) berichtet beispielsweise alleine für das Jahr 1998 von knapp 350 psychologischen Publikationen mit meta-analytischem Bezug in der deutsch- und englischsprachigen Literatur. Zunehmendes Interesse findet die Methode der Meta-Analyse auch in anderen Disziplinen wie der Medizin. Bislang weniger üblich scheint der Einsatz der Meta-Analyse im Marketing zu sein. Bijmolt und Pieters (2001) erwähnen in einer Übersichtsarbeit zur Anwendung der Meta-Analyse im Marketing elf meta-analytische Untersuchungen.

Die Methode der Meta-Analyse bietet die Möglichkeit, die Stärke eines statistischen Effekts über die berücksichtigten Einzelstudien hinweg gemeinsam abzuschätzen. Es wird angestrebt, die Einzelergebnisse in ein einheitliches Effektmaß zu überführen, welches dann aufgrund der gepoolten Stichprobe eine größere statistische Aussagekraft aufweist als die Ergebnisse einer empirischen Einzelstudie. Es lassen sich mit einer Meta-Analyse zudem Faktoren, welche einen systematischen Einfluss auf das Ergebnis haben, als sogenannte Moderatorvariablen identifizieren.

Im Forschungsbereich der Präferenzmessung interessiert zentral, ob sich die unterschiedlichen Gruppen an Messverfahren in ihrer Prognosegüte unterscheiden. In der Vergangenheit wurden daher in mehreren Studien diesbezüglich zwei oder mehr Verfahrensvarianten direkt miteinander verglichen. Im Vergleich der Ergebnisse konnte bislang keine einstimmige Rangordnung der Verfahren in Bezug auf die Vorhersagegüte abgeleitet werden.

In dieser Arbeit werden 135 Studien mit Hilfe der Meta-Analyse zu einer mittleren Effektgröße für die Validität der Präferenzmessverfahren zusammengeführt. Neben anderem wird überprüft, ob die Art des verwendeten Messverfahrens einen signifikanten Einfluss auf die Höhe der Validität aufweist.

Als Verfahrensgruppen werden in der Meta-Analyse die traditionelle Conjoint-Analyse, die adaptive Conjoint-Analyse (ACA), andere hybride Verfahren und die wahlbasierte Conjoint-Analyse miteinander verglichen. Zudem wird untersucht, ob sich kompositionelle Formen der direkten Nutzenmessung in ihrer Messgüte signifikant von den dekompositionellen Verfahren der Conjoint-Familie unterscheiden.

Die Ergebnisse der durchgeführten Meta-Analyse zeigen, dass sich die Verfahrensgruppen der Präferenzmessung signifikant in ihrer Prognosegüte unterscheiden. In der Klarheit der Ergebnisse hebt sich die durchgeführte Meta-Analyse deutlich von der bisherigen Forschung ab.

Die vorliegende Arbeit besteht aus zwei Teilen. Im ersten Teil der Arbeit werden die in der Meta-Analyse berücksichtigten Methoden der Präferenzstrukturmessung und die praktisch üblichen Vorgehensweisen bei der Reliabilitäts- und Validitätsbestimmung zusammenfassend besprochen. Im zweiten Teil werden die Vorgehensweise und die Ergebnisse der durchgeführten Meta-Analyse zur Messgüte der unterschiedlichen Verfahrensgruppen der Präferenzmessung dargestellt.

Der erste Teil untergliedert sich in drei Kapitel, von denen das erste neben dem Konzept der Präferenzstrukturmessung die alternativen, in der Meta-Analyse berücksichtigten Verfahrensgruppen der kompositionellen und dekompositionellen Nutzenmessung darstellt. Im zweiten Kapitel des Theorieteils wird näher auf die Kriterien der Messgüte eingegangen. Hier soll die Reliabilität und Validität sowie deren Operationalisierungen im Zusammenhang mit den Instrumenten der Präferenzmessung erörtert werden. Im abschließenden Kapitel des Grundlagenteils wird die Methode der Meta-Analyse in konzeptioneller Hinsicht zusammengefasst.

Im zweiten Teil der vorliegenden Arbeit wird zusammenfassend von der durchgeführten Meta-Analyse berichtet. Dazu werden in Kapitel fünf zunächst der Aufbau und die einzelnen Durchführungsschritte der Meta-Analyse beschrieben. Im Kapitel sechs finden sich schließlich die Ergebnisse der Meta-Analyse zu den unterschiedlichen Kriterien der Reliabilität und der Validität der Präferenzmessverfahren. Im abschließenden Kapitel sieben findet sich eine zusammenfassende Diskussion der wesentlichen Ergebnisse der durchgeführten Meta-Analyse.

2. Konzept und Verfahren der Präferenzstrukturmessung

Dieses Kapitel greift das Konzept der Präferenzstrukturmessung zusammenfassend auf und beschreibt im weiteren Verlauf die wesentlichen Verfahrensgruppen der Präferenzstrukturmessung in Abgrenzung zueinander. Anhand der traditionellen Conjoint-Analyse (CA) werden die Verfahrensschritte einer dekompositionellen Präferenzstrukturmessung prototypisch beschrieben. Die Verfahrensalternativen der hybriden Conjoint-Analyse, der adaptiven Conjoint-Analyse und der wahlbasierten Conjoint-Analyse werden in ihren Unterschieden und Erweiterungen zur traditionellen CA diskutiert. Weiterhin wird in diesem einführenden Kapitel die Konzeption und übliche Vorgehensweise der direkten Nutzenmessung beschrieben.

Die Bedeutung dieses Kapitels für die Meta-Analyse liegt in einer Klassifizierung der unterschiedlichen Verfahren der Präferenzmessung. Die in diesem Kapitel vorgestellten Verfahren werden in der Meta-Analyse berücksichtigt. Die Klassifizierung dient in diesem Sinne als Entscheidungskriterium für die Aufnahme einer Untersuchung in die Meta-Analyse.

2.1 Multiattribute Präferenzstrukturmessung

Die Präferenz einer Person bezieht sich immer auf ein Beurteilungsobjekt. Es handelt sich bei einer Präferenz um eine zeitlich überdauernde, affektiv-bewertende Einstellung gegenüber einem Objekt. In ihrer Wortbedeutung geht die Präferenz allerdings über das umfangreich beforschte Konstrukt der Einstellung (z. B. Fishbein & Ajzen, 1975) hinaus. Von der Präferenz einer Person wird dann gesprochen, wenn eine Person ein Beurteilungsobjekt gegenüber einem oder mehreren anderen Objekten vorzieht. Die Präferenz beinhaltet also immer eine relationale Beurteilung zwischen verschiedenen Objekten und kann somit nicht isoliert von anderen gegenüber einem Objekt alleine bestehen.

Beurteilungsobjekte, gegenüber denen Menschen Präferenzen ausbilden können, sind materielle Produkte wie Autos, Computer und Wohnungen, aber auch Dienstleistungen, Menschen, Ideen und Konzepte, Organisationen etc. Eine Grundidee der Präferenzstrukturmessung geht davon aus, dass sich die Beurteilungsobjekte durch ein oder mehrere Attribute beschreiben lassen. Beurteilungsobjekte werden als Bündel von Attributen aufgefasst.

Ein Stellenangebot lässt sich beispielsweise an den Attributen Name des Unternehmens, Jahresgehalt, Weiterbildungsangebot und Abwechslungsgrad der Tätigkeit differenzieren. Die jeweiligen Attribute werden mit zwei oder mehreren Stufen operationalisiert. Bei nicht-stetigen Attributen, wie dem Namen des Unternehmens, sind die Abstände zwischen den einzelnen Stufen nicht definiert. Genau genommen gilt dies für stetige Attribute, welche durch einige wenige Stufen repräsentiert werden, ebenso. Das stetige Attribut Jahresgehalt kann beispielsweise mit fünf diskreten Beobachtungspunkten zwischen 20.000 Euro und 55.000 Euro operationalisiert werden.

Jeder einzelnen Attributausprägung wird in der Präferenzmessung ein numerischer Wert zugeordnet, der für die Auskunftsperson den für die spezielle Attributstufe subjektiv wahrgenommenen Teilnutzen repräsentiert. Die Abstände eines stetigen Attributs können bei Kenntnis des Nutzenverlaufs des Attributs über verschiedene Methoden einer Interpolation geschätzt werden.

In der multiattributen Präferenzstrukturmessung lässt sich individuell bestimmen, welche Wichtigkeit ein Attribut zur Beurteilung eines Gesamtobjekts aufweist und welchen Nutzen einzelne Ausprägungen eines Attributs stiften. Mit Hilfe von Bewertungs- und Verknüpfungsfunktionen werden die einzelnen Teilnutzenwerte zu dem Gesamtnutzen eines Beurteilungsobjekts zusammengefügt. Die Gesamtheit aller Teilnutzenwerte einer Person wird unter der Spezifikation eines zugrunde gelegten Präferenzstrukturmodells als Präferenzstruktur der Person bezeichnet. Die Abbildung der Präferenz eines Menschen durch ein numerisches Relativ stellt in diesem Sinne die multiattribute Präferenzstrukturmessung, oder kurz Präferenzmessung, dar. Green und Srinivasan (1990) bezeichnen die Messung des Ausmaßes, in dem die diversen Charakteristika eines Objekts zu dessen Gesamtattraktivität beitragen, als Präferenzstrukturmessung (preference structure measurement).

Zur Bestimmung der individuellen Präferenzstruktur einer Person werden Auskünfte dieser Person benötigt, die entweder aus Beobachtungen in Form von Marktdaten oder in Form von Befragungen erhoben werden können. Bei der Berücksichtigung von getätigten Kauf- oder Wahlentscheidungen wird von offenbarten Präferenzen (revealed preferences) gesprochen (z. B. Ben-Akiva et al., 1994). Es wird also, nachdem sich eine Person für ein bestimmtes Beurteilungsobjekt entschieden hat, auf die Präferenz der Auskunftsperson zurückgeschlossen. Die Präferenzmessung mittels Marktdaten unterliegt einigen methodischen Schwierigkeiten. Beispielsweise liegen nur in seltenen Fällen ausreichende

Mengen an Entscheidungsdaten vor, um Aussagen auf der individuellen Probandenebene zu treffen. Zudem sind die Präferenzen einer Auskunftsperson bei realen Entscheidungen konfundiert mit weiteren Faktoren wie der Verfügbarkeit und der Bewerbung eines Produkts. Es können mit realen Entscheidungsdaten keine neuen, noch nicht verfügbaren Produktkonzepte getestet werden. Schließlich bietet die Auswahlentscheidung für ein Objekt keine Informationen über die Präferenzdifferenzen der Auskunftsperson zwischen den übrigen, nicht ausgewählten Alternativen.

Die über eine Befragung erhobenen Präferenzdaten werden als ausgedrückte Präferenzen (stated preferences) bezeichnet und vermeiden die Nachteile der offenbarten Präferenzdaten. Jedoch müssen die in einer Befragung geäußerten Präferenzen einer Auskunftsperson nicht mit ihrem realen Auswahlverhalten bei der Entscheidung zwischen Objekten übereinstimmen. In der vorliegenden Meta-Analyse werden Untersuchungen berücksichtigt, die Verfahren zur Erhebung und Auswertung von Daten aus Befragungen, also ausgedrückte Präferenzen, verwenden.

Die häufig zu beobachtende Diskrepanz zwischen der ausgedrückten Präferenz einer Auskunftsperson und dessen realer Kauf- oder Wahlentscheidung kann analog diskutiert werden zu der Frage, aus welchen Gründen sich in vielen Untersuchungen lediglich ein geringer Zusammenhang zwischen der Einstellung einer Person und dessen Verhalten gefunden hat. Ajzen und Fishbein (1980) führen die Unterschiede in der Einstellung und dem gezeigten Verhalten auf die geringe Korrespondenz zwischen der eher globalen Messung der Einstellung und der sehr spezifischen Beobachtung realen Verhaltens in einer bestimmten Situation zurück. Das Verhalten stellt immer eine bestimmte Art von Handlung dar, richtet sich immer auf ein spezifisches Ziel und ist in einen Gesamtkontext eingebettet. Zudem können sich über die Zeit Veränderungen ergeben, wenn der Zeitpunkt der Einstellungsmessung weit von dem Zeitpunkt des gezeigten Verhaltens entfernt liegt.

Gemäß der Korrespondenzhypothese von Ajzen und Fishbein lässt sich ein enger Zusammenhang zwischen Einstellung und Verhalten genau dann finden, wenn sich die Einstellungs- und Verhaltensmessungen im Hinblick auf die vier Spezifikationen der Art der Handlung, des Ziels, des Kontexts und der Zeit entsprechen. In der Präferenzmessung führt die Idee der Korrespondenzhypothese zu der Forderung, dass die Befragungssettings den realen Entscheidungssituationen möglichst ähnlich sein sollen.

Eckes und Six (1994) kommen allerdings in einer Meta-Analyse auf der Basis von 501 Studien zur Relation von Einstellung und Verhalten zu dem Schluss, dass „das lange tradierte Urteil, Einstellungen stünden bestenfalls in einer schwachen Beziehung zu Verhalten, den tatsächlich bestehenden Zusammenhängen nicht gerecht wird“ (S. 253). Auch nach Durchsicht mehrerer Meta-Analysen der Einstellungs-Verhaltens-Forschung stellen sie fest, dass „Einstellungs- und Verhaltensmaße keineswegs so niedrig miteinander korrelieren, wie es in der Folge früherer narrativer Reviews immer wieder behauptet wurde“ (Six & Eckes, 1996, S. 7).

Übertragen auf die Präferenzmessung werfen die Erkenntnisse der Einstellungs-Verhaltens-Forschung die Frage auf, in wieweit die ausgedrückte Präferenz einer Auskunftsperson mit dem realen Entscheidungsverhalten korreliert. Diese Frage wird in dieser Meta-Analyse untersucht.

2.2 Zentrale Begrifflichkeiten

Bevor in diesem Kapitel detaillierter auf den Prozess der Präferenzmessung eingegangen wird, sollen zunächst einige zentrale Begrifflichkeiten eingeführt werden.

Der bereits verwendete Begriff des *Objekts* bezeichnet den Entscheidungs- oder Beurteilungsgegenstand, auf den sich die Präferenz bezieht. Es kann sich dabei um ein Produkt, eine zu bewertende Person oder um immaterielle Konzepte wie Ideen, Organisationen, Gesetze, Dienstleistungen etc. handeln. Der synonym genutzte Begriff des Produkts stellt eine Einengung des Begriffs Objekt dar, ist aber gerade in der Kaufverhaltensforschung häufig anzutreffen. Bisweilen wird der Begriff Objekt ersetzt durch Option oder Alternative. Mit diesen Varianten wird betont, dass mindestens zwei Objekte zur Beurteilung vorliegen, was bei der Präferenzbestimmung stets der Fall sein muss, da es sich um Einstellungen bezüglich der relativen Bevorzugung von Objekten handelt. Die Objekte werden anhand von Attributen definiert.

Unter einem *Attribut* eines Objekts versteht man eine entscheidungsrelevante Eigenschaft oder Beurteilungsdimension des Objekts. Der Begriff Merkmal wird synonym verwendet. In der Präferenzmessung werden Objekte in der Regel durch ein Bündel an Attributen beschrieben. Attribute können für die Beurteilung des Gesamtobjekts unterschiedliche Wichtigkeiten aufweisen. Es wird aus Gründen der Effizienz der Datenerhebung darauf

geachtet, dass in der Regel nur entscheidungsrelevante Eigenschaften des Objekts in die Messung eingehen. Attribute können unterschiedliche Skalenniveaus aufweisen. Es ist aufgrund der üblichen Anwendung des Teilnutzenmodells häufig der Fall, metrische Attribute über einzelne Attributstufen zu diskretisieren.

Unter *Attributstufen* werden die möglichen Ausprägungen eines Objekts auf einem spezifischen Attribut verstanden. Die Attributstufen stellen üblicher Weise eine diskrete Unterteilung der Eigenschaftsdimension dar. Beispielsweise kann das metrische Attribut Jahresgehalt durch die Attributstufen 20.000, 30.000 und 40.000 Euro repräsentiert werden. Die Attributstufen müssen dabei nicht gleichabständig sein. Ebenso sind qualitativ umschriebene Attributstufen denkbar, z. B. kann das Attribut Weiterbildungsmöglichkeiten mit den Attributstufen gering, mittel, hoch definiert werden. Nicht selten sind dichotome Attribute anzutreffen. Beispielsweise kann das Attribut Aufstiegsmöglichkeiten lediglich die Differenzierung vorhanden, nicht vorhanden vorsehen.

Im Rahmen der wahlbasierten Conjoint-Analyse werden den Probanden Wahlaufgaben mit einer begrenzten Anzahl an Alternativen vorgelegt. Ein Wahlset wird als *Szenario*, die Summe aller Wahlsets als Gesamtszenario bezeichnet. Die Häufigkeit der Wahl einer Alternative gegenüber allen Wahlentscheidungen des Szenarios stellt den Wahlanteil einer Alternative dar. Die Wahlanteile aller Optionen eines Szenarios summieren sich zu eins. Der Schwierigkeitsgrad einer Wahlaufgabe bestimmt sich maßgeblich durch die wahrgenommene Nutzenähnlichkeit der in dem Szenario enthaltenen Wahlalternativen (Brocke, 2006). Zur Generierung effizienter Untersuchungsdesigns kommt der Nutzenbalancierung der Wahlsets eine zentrale Bedeutung zu (z. B. Großmann et al., 2006). In nutzenbalancierten Szenarios werden die Wahlalternativen derart zusammengestellt, dass alle Optionen gemäß der a priori ermittelten individuellen Präferenzstruktur des Probanden eine vergleichbare Wahlwahrscheinlichkeit aufweisen.

Das theoretische Konstrukt des *Nutzens* wird in der Präferenzmessung intensiv genutzt. Unter dem Nutzen eines Objekts wird der subjektive Wert verstanden, welchen eine Auskunftsperson einer Attributstufe, einem Attribut oder einem Bündel an Attributen, dem Objekt, zuschreibt. Es wird zwischen dem Gesamtnutzen eines ganzheitlichen Objekts und den Teilnutzenwerten unterschieden. Letztere beschreiben den Beitrag, welchen die spezifischen Attributausprägungen zum wahrgenommenen Gesamtnutzen des Objekts beitragen. Es wird davon ausgegangen, dass eine Attributstufe jeweils mehr oder weniger

nutzenstiftend ist und somit positiv oder negativ zur Präferenz der Auskunftsperson beiträgt. Beispielsweise ist es naheliegend, dass ein Stellenangebot mit einem hohen Jahresgehalt ceteris paribus einem Stellenangebot mit einem niedrigen Jahresgehalt vorgezogen wird. Bei weitem nicht bei allen Attributen lassen sich die subjektiven Teilnutzenwerte den objektiven Attributstufen so linear zuordnen wie bei den quantitativen Attributen Kosten, Einkommen, Preis etc. Zur Definition des Zusammenhangs zwischen den objektiven Attributstufen und deren subjektiven Repräsentationen, den Teilnutzenwerten, werden Nutzenfunktionen verwendet.

Mit einer *Nutzenfunktion* wird festgelegt, wie einer Eigenschaftsausprägung ein Teilnutzenwert zugeordnet wird. Eine Nutzenfunktion beschreibt den Zusammenhang zwischen den objektiven Attributstufen und dem subjektiven Nutzen, welchen eine Auskunftsperson der Attributstufe zuschreibt. Die Nutzenfunktion wird auch als Bewertungsfunktion oder Präferenzfunktion bezeichnet. Nutzenfunktionen können, je nach Attribut und Proband, in ihrem Verlauf variieren. Grundsätzlich können Nutzenfunktionen jede beliebige Form annehmen. Es ist denkbar, einen bestimmten Verlauf der Nutzenfunktion zugrunde zu legen (z. B. linear, logarithmisch, exponentiell etc.).

Das übliche Modell in der multiattributen Präferenzstrukturmessung ist das *Teilnutzenmodell*, das jeder Attributstufe einen subjektiven Teilnutzenwert zuordnet. Bei der Anwendung des Teilnutzenmodells werden keine Restriktionen an den Verlauf der Nutzenfunktion aufgenommen. Die Funktion ist nur für die diskreten Attributstufen definiert. Ausgehend von der Nutzenzuordnung einzelner Attributausprägungen kann der Verlauf der Nutzenfunktion geschätzt werden. Zur Ermittlung des Nutzenverlaufs zwischen den definierten Skalenstufen des Teilnutzenmodells können Methoden der Interpolation herangezogen werden.

Zur Darstellung der Gesamtpräferenzen für Objekte ist eine *Verknüpfungsfunktion* erforderlich, die angibt, wie die Teilnutzen einzelner Eigenschaften zur jeweiligen Gesamtobjektpräferenz zusammengefasst werden. Verknüpfungsfunktionen werden auch als Präferenzintegrationsfunktionen bezeichnet. Eine Verknüpfungsfunktion stellt die Entscheidungsregel algebraisch dar, nach der eine Auskunftsperson bei der Beurteilung eines Gesamtobjekts vorgeht. Unter anderem ist es denkbar, dass eine Auskunftsperson ihre Präferenzen nach den auf den ersten Blick wichtigsten Attributen festlegt, Ausschlusskriterien definiert oder die Teilnutzenwerte eines jeden Objekts summarisch zusammenfasst.

Unter *Präferenzen* werden überdauernde und relationale Einstellungen zu Beurteilungsobjekten verstanden. Präferenzen sind das Ergebnis eines kognitiven Prozesses der Beurteilung von Objekten und drücken immer eine relative Bewertung eines Objekts zu einem oder mehreren anderen Objekten aus. Damit unterscheiden sie sich von dem Konstrukt der Einstellungen.

Einstellungen sind affektive oder kognitive Bewertungen gegenüber einem Objekt und können auch losgelöst von anderen Objekten existieren. Einstellungen und Präferenzen können sich verändern, es wird beiden jedoch eine gewisse Stabilität über einen Zeitraum unterstellt. Während Einstellungen als kognitive Repräsentationen eines Objekts bezeichnet werden, sind Präferenzen relationale, kognitive Repräsentationen eines Objekts.

Abzugrenzen ist der Begriff der Präferenz zudem von dem der *Verhaltensabsicht*. Eine Verhaltensabsicht setzt eine positive Einstellung gegenüber einem Objekt und eine entsprechende Präferenz für ein Objekt relativ zu anderen voraus. Darüber hinaus müssen jedoch auch weitere, absichtsbeeinflussende Faktoren wie soziale Normen und von der Person antizipierte, situative Rahmenbedingungen berücksichtigt werden.

Ob eine Verhaltensabsicht schließlich tatsächlich zu einem den Präferenzen entsprechenden *Entscheidungsverhalten* führt, hängt entscheidend von den situativen Rahmenbedingungen, dem Kontext der Entscheidungssituation ab. Das Entscheidungsverhalten beruht auf einer relationalen, kognitiven Repräsentation der Objekte und der kognitiven Repräsentation der vorhandenen Situation.

Es ist offensichtlich, dass von der Präferenz einer Person nicht stringent auf dessen Entscheidungsverhalten geschlossen werden kann. Es wird allerdings angenommen, dass die Präferenz ein guter Prädiktor für das Kaufverhalten ist, was die Popularität von Präferenzmessverfahren im Marketing erklärt. In anderen Feldern kommt der Präferenzermittlung hingegen bereits ein Wert an sich zu, beispielsweise in der Berufsberatung oder bei der Generierung von Anforderungsprofilen. Wie im dritten Kapitel beschrieben wird, messen Studien zur prognostischen Validität eines Präferenzmessverfahrens die Korrespondenz zwischen den ermittelten Präferenzstrukturen von Antwortpersonen und ihrem realen Entscheidungsverhalten.

Die bislang eingeführten Begriffe werden in der weiteren Arbeit durch die folgenden Notationen repräsentiert.

h	Auskunftsperson h , mit $h = 1, \dots, H$
A_i	Alternative / Objekt i , mit $i = 1, \dots, I$
x_j	Attribut j , mit $j = 1, \dots, J$
x_{jk}	Attributstufe k des Attributs j , mit $k = 1, \dots, K_j$
U_i	Subjektiver Gesamtnutzen, den die Auskunftsperson dem Objekt A_i zumisst
$u(x_{jk})$	Subjektiver Teilnutzen, den die Auskunftsperson der Attributstufe x_{jk} zumisst

2.3 Modellierung multiattributer Präferenzen

2.3.1 Das Präferenzstrukturmodell

Die Präferenzen einer Person werden bei der Präferenzmessung in Form eines Präferenzstrukturmodells abgebildet. Ein Präferenzstrukturmodell besteht aus den Nutzenfunktionen der einzelnen Eigenschaften eines Objekts und deren Verknüpfung. Es fasst die Nutzenfunktion und die Verknüpfungs- bzw. Präferenzintegrationsfunktion in einem Modell zusammen und bestimmt den Gesamtnutzen des Objekts A_i .

Mit Präferenzstrukturmodellen wird der Annahme Rechnung getragen, dass sich der Präferenzbildungsprozess in zwei Teilschritte untergliedern lässt. In einem Teilprozess werden den einzelnen Attributstufen mittels der Nutzenfunktion Teilnutzenwerte zugeordnet. In dem anderen Teilprozess werden die Teilnutzenwerte über die Verknüpfungsfunktion zu dem Gesamtnutzen des Objekts A_i zusammengefasst.

Die Nutzenfunktion gibt an, welchen Nutzen eine Antwortperson einer Attributstufe x_{jk} zumisst. Sie ordnet also jeder Eigenschaftsausprägung auf der individuellen Ebene einen reellen Teilnutzenwert zu. Die Verknüpfungsfunktion eines Präferenzstrukturmodells gibt an, wie die Teilnutzen einzelner Attributstufen zur Gesamtpräferenz für ein Objekt A_i zusammengefasst werden. Für die Verknüpfungsfunktion gibt es unterschiedliche Formen, die jeweils eine bestimmte Annahme über die zugrunde liegenden Entscheidungsregeln abbilden. Entscheidungsregeln lassen sich in kompensatorische und nichtkompensatorische Modelle unterteilen.

2.3.2 Nichtkompensatorische Entscheidungsregeln

Die konjunktive, disjunktive und lexikographische Entscheidungsregel gelten als wesentliche nichtkompensatorische Entscheidungsregeln. Bei der *konjunktiven Entscheidungsregel* wird davon ausgegangen, dass ein Proband für jedes Attribut eine Mindestanforderung definiert, die erfüllt sein muss, damit das Gesamtobjekt A_i akzeptiert wird. Sobald ein Attribut eine nichtakzeptable Ausprägung aufweist, wird das gesamte Objekt abgelehnt.

Auch bei der *disjunktiven Entscheidungsregel* definiert die Antwortperson für alle Attribute zu erfüllende Mindestausprägungen. Es wird jedoch davon ausgegangen, dass ein Produkt akzeptiert wird, sobald ein Attribut die festgelegte Mindestanforderung erreicht, disjunkt von den Ausprägungen und Wichtigkeiten der übrigen Eigenschaften.

Die *lexikographische Entscheidungsregel* geht von der Annahme aus, dass ein Entscheider die Objekte zunächst nur anhand des individuell als wichtigstes erachteten Attributs vergleicht. Sollten zwei oder mehrere Gesamtobjekte auf dem wichtigsten Attribut die meist präferierte Ausprägung aufweisen, so wird das nächstwichtigste Attribut in den Entscheidungsprozess mit einbezogen usw.

Es erscheint plausibel, dass Menschen gerade in komplexen Entscheidungssituationen auf nichtkompensatorische Entscheidungsregeln zurückgreifen. Das konjunktive und das disjunktive Modell weisen die Problematik auf, dass sich keine eindeutigen Präferenzrangfolgen ableiten lassen, wenn nicht aufwändiger Weise die Mindestanforderungen sukzessive angehoben werden (Böcker, 1986).

In der Präferenzstrukturmessung kommen nichtkompensatorischen Modellen vor allem im Rahmen von zweistufigen Präferenzbildungsprozessen Bedeutung zu. So können Probanden zunächst gebeten werden, nichtakzeptable Attributstufen oder nichtrelevante Attribute zu identifizieren, die dann von der weiteren Präferenzmessung ausgeschlossen werden. Die empirische Forschung hat allerdings gezeigt, dass Personen nicht durchweg konsistent Objekte zurückweisen, welche Attributausprägungen enthalten, die zuvor als inakzeptabel eingestuft worden waren.

2.3.3 Kompensatorische Entscheidungsregeln

Die kompensatorischen Entscheidungsregeln gehen davon aus, dass eine ungewünschte Ausprägung eines Attributs durch die mehr oder weniger gewünschten Ausprägungen der übrigen Attribute kompensiert werden kann, ohne dass das Gesamtobjekt abgelehnt wird. Bei den kompensatorischen Entscheidungsregeln erfolgt eine Verknüpfung der Teilnutzenwerte in der Regel additiv. Theoretisch ist auch ein multiplikatives Verknüpfungsmodell denkbar, dieses kommt aber in der praktischen Anwendung nicht vor.

Die Verknüpfungsfunktion für das additive Präferenzintegrationsmodell lässt sich wie folgt darstellen:

$$U_i = \sum_{j=1}^J u(x_{ij}) \quad (1)$$

Grundsätzlich ist eine jede Kombination aus den diversen Nutzen- und Verknüpfungsfunktionen zu einem Präferenzstrukturmodell denkbar. Aufgrund der Einfach- und Robustheit werden in der Regel additive Modelle zusammen mit dem Teilnutzenwertmodell verwendet. Es wird angenommen, dass sich der Nutzen eines Objekts als Summe der Nutzenwerte der Attributausprägungen ergibt. Wird die Nutzenfunktion des Teilnutzenmodells in die additive Verknüpfungsfunktion gemäß Formel (1) integriert, ergibt sich das in der Conjoint-Analyse am häufigsten verwendete additive Teilnutzenmodell:

$$U_i = \sum_{j=1}^J \sum_{k=1}^{K_j} b_{jk} x_{ijk} \quad (2)$$

- mit:
- U_i : Subjektiver Gesamtnutzen, den die Auskunftsperson dem Objekt A_i zuschreibt
 - b_{jk} : Teilnutzenwert für die Ausprägungsstufe k des Attributs j (zu schätzende Modellparameter)
 - x_{ijk} : Dummy-Variable mit dem Wert 1, wenn die Eigenschaft k des Attributs j bei dem Objekt i auftritt; sonst 0

2.4 Verfahren der Präferenzstrukturmessung

Objekte werden in der Präferenzstrukturmessung als Bündel von Attributen mit jeweils spezifischen Attributausprägungen betrachtet. Die Präferenz eines Probanden zu einem Objekt ergibt sich aus der Verknüpfung der Nutzenwerte, welche dieser den jeweiligen Attributausprägungen zumisst. Es ist die Aufgabe der Präferenzmessmethoden, die Teilnutzenwerte eines Probanden zu bestimmen. Dies kann über eine direkte Abfrage der Teilnutzenwerte oder eine ganzheitliche Bewertung von Gesamtobjekten erfolgen.

Aus den zahlreichen Möglichkeiten, die Präferenzstruktur von Individuen zu messen, werden diejenigen Methoden herausgegriffen, welche die größte Bedeutung und Verbreitung in der praktischen Anwendung erfahren haben. Wittink, Vriens und Burhenne (1994) berichten beispielsweise für Europa, dass 95 % der von ihnen registrierten 1.000 Anwendungsstudien eine der folgenden Verfahrensgruppen verwendet haben: traditionelle Conjoint-Analyse, wahlbasierte Conjoint-Analyse, adaptive Conjoint-Analyse (ACA), hybride Conjoint-Analyse oder direkte, kompositionelle Verfahren. Darauf folgende Erhebungen kommen zu ähnlichen Ergebnissen (Melles & Holling, 1998; Hartmann & Sattler, 2002).

Diese fünf Verfahrensgruppen werden in der Meta-Analyse berücksichtigt und werden daher in diesem Unterkapitel näher beschrieben.

2.4.1 Kompositionelle Nutzenmessung

Bei der kompositionellen Nutzenmessung werden die Antwortpersonen gebeten, jede Eigenschaftsausprägung separat zu bewerten. Die erhobenen Teilnutzenwerte werden dann gemäß einer üblicher Weise additiven Verknüpfungsregel zu den Gesamtnutzenwerten der Objekte zusammengesetzt (komponiert). Die kompositionelle Nutzenmessung wird synonym auch als direkte Nutzenmessung oder Nutzenmessung mittels self-explicated-Modellen bezeichnet, da die Probanden in der kompositionellen Nutzenmessung ihre Präferenzstruktur auf der Ebene der Teilnutzenwerte direkt benennen.

Unter den zahlreich denkbaren Formen der kompositionellen Nutzenmessung stellt das 2-Stufen-Modell das gängigste dar, da es durch die Integration in die Software ACA zur adaptiven Conjoint-Analyse eine weite Verbreitung gefunden hat. In der ersten Phase wird die Antwortperson gebeten, zu jeder Ausprägung eines Attributs die Erwünschtheit dieser

Attributstufe zu bestimmen. In einer zweiten Phase sollen dann die Attributwichtigkeiten benannt werden, also die Bedeutungen, welche den Attributen relativ zueinander zur Bestimmung des Gesamtnutzens aus der Sicht der Antwortperson zukommen.

Formal ausgedrückt ergibt sich der Gesamtnutzen eines Objekts A_i nach einem zweistufigen Modell der kompositionellen Nutzenmessung unter Annahme einer additiven Verknüpfung nach dem Teilnutzenmodell wie folgt:

$$U_i^{\text{komp}} = \sum_{j=1}^J \sum_{k=1}^{K_j} w_j^{\text{komp}} d_{jk}^{\text{komp}} x_{ijk} \quad (3)$$

mit: U_i^{komp} : Subjektiver Gesamtnutzen, den die Auskunftsperson dem Objekt A_i gemäß des zweistufigen Modells der direkten Nutzenmessung zumisst
 w_j^{komp} : Direkt erfragtes Bedeutungsgewicht des Attributs j
 d_{jk}^{komp} : Direkt erfragte Erwünschtheit der Ausprägungsstufe k des Attributs j
 x_{ijk} : Dummy-Variable mit dem Wert 1, wenn die Eigenschaft k des Attributs j bei dem Objekt i auftritt; sonst 0.

Der Gesamtnutzen ist nach Formel (3) die Summe aus den Teilnutzenwerten, errechnet aus dem Produkt des erfragten Bedeutungsgewichts eines Attributs mit dem Erwünschtheitswert der Ausprägungsstufe des Attributs. Bei den einstufigen Modellen der Nutzenmessung wird auf die Gewichtung der Attribute verzichtet und lediglich der Erwünschtheitswert der Ausprägungsstufen abgefragt. Vereinzelt werden auch nur die Attributgewichte erhoben.

Um bei der Berechnung der Teilnutzenwerte gemäß der Formel (3) zu sinnvollen Ergebnissen zu gelangen, müssen die Erwünschtheitswerte d_{jk}^{komp} der einzelnen Attributstufen mindestens intervallskaliert, die Attributwichtigkeiten w_j^{komp} ratioskaliert sein. Es ist fraglich, ob die erforderlichen Skalenniveaus bei den direkten Befragungen unterstellt werden können. Einige kompositionelle Verfahren arbeiten aufgrund dieser Problematik mit dem Konstantsummenverfahren, welches einen absoluten Nullpunkt beinhaltet.

Ein prototypisches Verfahren der Gruppe der direkten Nutzenmessung wurde von Srinivasan (1988) vorgestellt. Die Datenerhebung gliedert sich bei diesem Verfahren in drei Schritte:

Zunächst wird der Proband mit allen Attributen und deren Ausprägungsstufen vertraut gemacht und gebeten, solche Attributstufen zu benennen, welche in jeder Hinsicht inakzeptabel sind. Als inakzeptabel gelten solche Attributausprägungen, die zu einer

Ablehnung eines Objekts führen würden, unabhängig davon, wie die übrigen Attribute ausgeprägt sind.

In einem zweiten Schritt wird der Proband gebeten, dasjenige Attribut mit der größten Bedeutung für die Gesamtbeurteilung des Objekts auszuwählen und mit der Kennziffer 100 zu versehen. Alle übrigen Attribute sollen dann in ihrer Wichtigkeit relativ zu dem wichtigsten Attribut eine Kennziffer unter 100 erhalten, wobei ein Wert von 0 bedeutet, dass dieses Attribut keine Bedeutung für die Bewertung aufweist.

Der letzte Schritt der Datenerhebung besteht schließlich darin, dass der Proband gebeten wird, für alle grundsätzlich akzeptablen Ausprägungsstufen innerhalb eines Attributs Bewertungen vorzunehmen in der Art, dass die meist präferierte Ausprägungsstufe jeweils den Wert 100 erhält und die übrigen Ausprägungsstufen relativ dazu einen Wert zwischen 0 und 100.

Die kompositionelle Nutzenmessung wurde vielfach kritisiert (z. B. Green & Srinivasan, 1990; Vriens, 1995). Aus theoretischer Sicht wurde beispielsweise bemängelt, dass die Frage nach der Bedeutsamkeit der Attribute mehrdeutig interpretiert werden kann, da die Begrifflichkeit der Bedeutsamkeit konfundiert ist mit benachbarten Konstrukten wie der Relevanz, der Salienz oder der Wertigkeit eines Attributs (Böcker, 1986). Die Einschätzung der Attributwichtigkeiten unterliegt dem Attributstufeneffekt (number-of-levels-effect), demnach Attribute mit einer größeren Anzahl an Attributstufen bei gleicher Spannweite des Attributs systematisch als wichtiger eingeschätzt werden (z. B. Wittink, Krishnamurti & Nutter, 1982; Wittink, Krishnamurti & Reibstein, 1989; McCullough, 1999; Wittink & Seetharaman, 1999).

Weiterhin wird kritisiert, dass die Antwortpersonen mit der unrealistischen Aufgabe der Bewertungen von einzelnen Attributstufen konfrontieren werden, anstatt ganzheitliche Objektpräferenzen abzugeben. Gerade im Hinblick auf sozial sensitive Attribute wie den Preis oder die Marke ist es denkbar, dass soziale Normen und soziale Erwünschtheitseffekte zu einer verfälschten Nutzenerklärung führen. Weiterhin zeigte sich, dass bei einer direkten Befragung die eher unwichtigen Attribute in ihrer Wichtigkeit systematisch überschätzt und die wichtigen Attribute in ihrer Bedeutung unterschätzt werden (Slovic, Fleissner & Bauman, 1972; Scott & Wright, 1976).

Auch wenn die Verfahren der kompositionellen Nutzenmessung vielfältig kritisiert worden sind, kommt ihnen in der Präferenzstrukturmessung ein zentraler Stellenwert zu. Wenn

berücksichtigt wird, dass self-explicated-Modelle sowohl eigenständig als auch als ein Prozessschritt in hybriden Modellen eingesetzt werden, zeigen Studien, dass in 52 % (Wittink, Vriens & Burhenne, 1994) bzw. 48 % (Hartmann & Sattler, 2002) aller in Wissenschaft und Praxis durchgeführten Präferenzstrukturmessungen kompositionelle Nutzenmessungen zum Einsatz kommen.

Der zentrale Vorteil der Verfahren der direkten Nutzenmessung wird in der Einfachheit ihrer Modellannahmen auf der Basis der aus der Einstellungsforschung bekannten Erwartungswert-Modelle (expectancy value model) und der Einfachheit der Datenerhebung gesehen (Wilkie & Pessemier, 1973; Eisenführ & Weber, 1994). Zudem hat sich in mehreren empirischen Studien eine überraschende Robustheit einzelner self-explicated-Modelle im Hinblick auf die Prognosegüte gezeigt (z. B. Srinivasan & Park, 1997).

Die Frage nach der höheren Validität von dekompositionellen Verfahren der Nutzenmessung gegenüber den kompositionellen Methoden beschäftigt die Forschung seit dem Aufkommen der dekompositionellen Messverfahren im Verlauf der 1970er Jahre und konnte durch empirische Vergleichsstudien noch nicht eindeutig beantwortet werden. Die in dieser Arbeit vorgestellte Meta-Analyse greift diese Frage auf.

2.4.2 Dekompositionelle Nutzenmessung

Die Grundidee der dekompositionellen Nutzenmessung sieht vor, von Präferenzurteilen einer Antwortperson zu ganzheitlichen Objekten auf die die Präferenz determinierenden Elemente der Attributwichtigkeiten und Teilnutzenwerte zu schließen. Hierzu werden die abgegebenen Gesamtpräferenzen in die Teilnutzenstruktur zerlegt (dekomponiert). Es wird angenommen, dass die Präferenzurteile einer Person durch den wahrgenommenen Nutzen der zu bewertenden Objekte determiniert sind. Der Gesamtnutzen eines Objekts ergibt sich aus der linearen Kombination der aus den Daten zu schätzenden Teilnutzenwerte. Die Präferenzstrukturmodellierung im Sinne der dekompositionellen Nutzenmessung stellt eine lineare Kombination von bekannten Regressionsfunktionen dar.

Die Ermittlung der Teilnutzenwerte und Attributwichtigkeiten muss bei der dekompositionellen Präferenzmessung derart erfolgen, dass die gemäß der zugrunde gelegten Verknüpfungsfunktion zusammengefassten Einzelwerte die durch die Antwortperson

abgegebene Präferenz für das Gesamtobjekt möglichst genau reproduzieren (kalibrieren). Das Ausmaß der Fähigkeit einer Präferenzmessmethode zur Abbildung der geäußerten Objektpräferenzen durch das Kalibrierungsmodell wird als Anpassungsgüte (model-fit) des Modells bezeichnet. Die Ermittlung der Teilnutzenwerte und der Attributwichtigkeiten aus den ganzheitlichen Objektbewertungen des Probanden erfolgt bei der dekompositionellen Nutzenmessung mittels statistischer Schätzverfahren im Rahmen des verallgemeinerten linearen Modells, wie beispielsweise einer Kleinste-Quadrate-Regression, varianzanalytischer Verfahren oder multinomialer logistischer Regression.

Die dekompositionellen Verfahren der Nutzenmessung wurden als Reaktion auf die Nachteile der kompositionellen Nutzenmessung entwickelt. Sie überwinden die genannten Schwierigkeiten der direkten Nutzenmessung. Insbesondere wird positiv bewertet, dass die Beurteilung von ganzheitlichen Objekten eher einer realen Entscheidungssituationen entspricht als die Bewertung separater Attribute und deren Ausprägungen.

Mit Verfahren der dekompositionellen Nutzenmessung lassen sich neben den Haupteffekten auch Wechselwirkungen zwischen den Attributen berücksichtigen. Interaktionseffekte treten auf, wenn die Nutzenbewertungen der einzelnen Attribute nicht unabhängig voneinander erfolgen. Eine entscheidungsrelevante Wechselwirkung ist beispielsweise denkbar zwischen den Attributen Farbe und Marke eines Automobils, wenn eine Person grundsätzlich ein graues Fahrzeug präferiert, jedoch einen roten Ferrari einem grauen vorzieht. Präferenzmodelle, die derartige Bewertungsprozesse abbilden können, verwenden Polynome. Bei Wechselwirkungen liefern die Attribute nicht nur isoliert, sondern in Kombination zueinander einen eigenständigen Nutzenbeitrag.

Die Berücksichtigung von Wechselwirkungen erfordert die Schätzung von zusätzlichen Parametern. Sollen alle möglichen Wechselwirkungen berücksichtigt werden, so erfordert dies ein vollfaktorielles Untersuchungsdesign. In der Regel werden nur ausgewählte Interaktionen zwischen Attributen berücksichtigt, für die eine Wechselwirkung aufgrund von Vorannahmen oder Voruntersuchungen erwartet werden kann.

In der praktischen Anwendung von Präferenzmessungen werden Interaktionen zumeist nicht berücksichtigt (Wittink, Vriens & Burhenne, 1994). Dies ist darauf zurückzuführen, dass Interaktionsmodelle komplexer und schwieriger zu interpretieren sind und mehr Parameter zu schätzen sind. Zudem geht die genauere Abbildung von Entscheidungsregeln offenbar nicht

zwangsläufig einher mit einer höheren Prognosegüte (z. B. Akaah & Korgaonkar, 1983; Green, 1984; Hagerty, 1986; Elrod, Louviere & Davey, 1992).

Besteht das Ziel der Präferenzmessung in der Prognose von Präferenzurteilen, so gelten additive Modelle als ausreichend robust (z. B. Green & Srinivasan, 1978; Böcker, 1986). Soll jedoch der Entscheidungsprozess der Präferenzbeurteilung möglichst adäquat modelliert werden, ist die Berücksichtigung von Attributwechselwirkungen unablässlich.

Nicht alle Verfahren der Präferenzmessung sind in der Lage, Interaktionen zu erheben. In der Meta-Analyse werden aus Vergleichbarkeitsgründen nur Daten von Conjoint-Analysen verwendet, welche keine Interaktionsterme berücksichtigt haben.

Als Negativaspekte der indirekten Nutzenmessung werden in erster Linie die höheren Anforderungen an den Durchführer der Präferenzmessung einschließlich eines größeren Geld- und Zeitaufwands sowie die Gefahr der kognitiven Überlastung der Probanden angeführt (z. B. Vriens, 1995). Es wird davon ausgegangen, dass Antwortpersonen nur eine begrenzte Anzahl an Objektbeschreibungen mit einer begrenzten Anzahl an Attributen und Attributstufen gleichzeitig bewerten können.

Bei einer Überlastung der Informationsverarbeitungskapazitäten können Vereinfachungsstrategien (simplifying-effects; Wright, 1975) zum Einsatz kommen, die jedoch unter Umständen nicht mehr durch die im Präferenzstrukturmodell angenommenen Entscheidungsregeln abgebildet werden. Das prominenteste Verfahren der dekompositionellen Nutzenmessung ist die Conjoint-Analyse.

Die traditionelle Conjoint-Analyse

Der Begriff der Conjoint-Analyse (CA) wurde in den 1970er Jahren durch einige viel beachtete Publikationen insbesondere im Forschungsbereich des Marketings geprägt (z. B. Green & Rao, 1971; Green, Carmone & Wind, 1972; Johnson, 1974; Green & Srinivasan, 1978). Die Ursprünge der Conjoint-Analyse sind im Conjoint-Measurement zu sehen, das in den 1960er Jahren aus dem Umfeld der mathematischen Psychologie Verbreitung fand. Die ersten theoretischen Beiträge zum Conjoint-Measurement kamen von Luce und Tukey (1964), Krantz (1964) und Tversky (1967). Aus psychologischer Sicht war es von Interesse, klinische Urteilsbildungsprozesse zu modellieren. Durch die Einführung der Ideen des

Conjoint-Measurements in die Marketingliteratur durch Green und Rao (1971) fand die Conjoint-Analyse zunehmende Aufmerksamkeit. Heute gilt die Conjoint-Analyse als das beliebteste und verbreitetste Instrument, um das Kaufverhalten von Konsumenten zu untersuchen bzw. vorherzusagen. (Green, Krieger & Wind, 2001).

Das Conjoint-Measurement ermöglicht die Schätzung von intervallskalierten Teilnutzenwerten auf der Basis von ganzheitlichen Präferenzurteilen einer Person, auch wenn die Präferenzurteile selbst nur ein ordinales Skalenniveau aufweisen. Das Conjoint-Measurement untersucht die Frage, unter welchen Bedingungen eine ordinale Präferenzrangfolge eindeutig durch eine additive Präferenzrelation repräsentiert werden kann. Dazu wurden Axiome formuliert, unter welchen Bedingungen von den ordinalen Präferenzdaten auf eine metrische Teilnutzenstruktur geschlossen werden kann. Das Conjoint-Measurement überprüft die Modellannahmen und geht bei einer Erfüllung der notwendigen und hinreichenden Bedingungen von der Angemessenheit der Repräsentation der Präferenzurteile durch die ermittelten Nutzenwerte aus. Das Conjoint-Measurement überprüft die Eignung von Präferenzmodellen zur Abbildung von Präferenzurteilen einer Person.

Im Gegensatz zum Conjoint-Measurement wird in der Conjoint-Analyse von einem statistischen Modell mit einem Fehleranteil ausgegangen. Der Schwerpunkt der Conjoint-Analyse wird auf die Überprüfung der Anpassungsgüte des zugrundegelegten Präferenzstrukturmodells gelegt. Es wird überprüft, wie exakt sich die erhobenen Präferenzdaten durch das Kalibrierungsmodell reproduzieren lassen. Die Conjoint-Analyse ermittelt möglichst exakt die Schätzparameter eines vorab angenommenen Präferenzstrukturmodells.

Unter Conjoint-Analyse werden, in Anlehnung an die Definitionen von Green und Srinivasan (1978) und Schweikl (1985), alle dekompositionellen Methoden der indirekten Nutzenmessung verstanden, die dazu dienen, aus den ganzheitlichen Präferenzurteilen einer Antwortperson über eine Menge von Beurteilungsobjekten die Parameter eines vorher festgelegten Präferenzmodells zu schätzen. Es handelt sich um Spezialfälle des verallgemeinerten linearen Modells.

Die *rangreihenbasierte Conjoint-Analyse* stellt historisch gesehen die älteste Verfahrensform dar. Bei der rangreihenbasierten Conjoint-Analyse werden die befragten Personen gebeten, eine Teilmenge aus allen Objekten gemäß ihrer Präferenzen in eine Rangreihe zu bringen. Eine Weiterentwicklung stellt die Abfrage der Präferenz der Probanden mittels Einstufungsskalen dar.

Bei der auf *Einstufungsskalen basierten Conjoint-Analyse* werden die Antwortpersonen gebeten, ihre Präferenz anhand einer mehrstufigen kategorialen oder einer kontinuierlichen Ratingskala anzugeben. Verbreitet sind zum Beispiel sogenannte „likelihood of purchase“-Skalen, wonach der Proband auf einer 100er-Skala seine Wahrscheinlichkeit angibt, dieses Objekt zu erwerben. Im Gegensatz zur rangreihenbasierten Conjoint-Analyse bietet die einstufigsskalenbasierte Conjoint-Analyse den Vorteil, dass Informationen über die Stärke der Präferenz der antwortenden Person vorliegen.

Zur Analyse der Präferenzdaten, die auf Einstufungsskalen basieren, werden mit der Methode der kleinsten Quadrate bzw. der metrischen Varianzanalyse statistische Auswertungsverfahren aus dem allgemeinen linearen Modell herangezogen. Auch Daten aus dem nicht-metrischen Antwortformat der Rangreihen werden häufig unzulässiger Weise über eine Varianzanalyse ausgewertet. Für die Bestimmung der Teilnutzenwerte auf der Basis von rangreihenbasierten Präferenzdaten stehen nicht-metrische Schätzverfahren zur Verfügung wie beispielsweise die lineare Optimierung (LINMAP) nach Srinivasan und Shocker (1973), die monotone Varianzanalyse (MONANOVA) nach Kruskal (1965) oder Verfahren auf Basis der optimalen Skalierung (z. B. ADDALS / MORALS; Takane, Young & de Leeuw, 1976).

Bei der Ermittlung der Teilnutzenwerte ist es möglich, Einschränkungen (constraints) im Hinblick auf den Verlauf der Nutzenfunktionen festzulegen. Nicht immer antworten alle Probanden gemäß von a-priori-Erwartungen und geben beispielsweise an, dass sie einen höheren Preis gegenüber einem niedrigen präferieren. Einschränkungen stellen sicher, dass die erhobenen Daten gemäß der a-priori-Erwartungen korrigiert werden. Empirische Vergleichsstudien fanden in der Mehrheit eine verbesserte Validität für conjoint-analytische Untersuchungen mit Einschränkungen, allerdings waren die Unterschiede in der Regel gering und selten signifikant (z. B. Srinivasan et al., 1983; Van der Lans et al., 1992, Moore, Metha & Pavia, 1994).

Die wahlbasierte Conjoint-Analyse

Bei der wahlbasierten Conjoint-Analyse (Discrete Choice Analysis) wird das Präferenzurteil in Form von diskreten Auswahlentscheidungen erhoben (Ben-Akiva & Lerman, 1985). Die Antwortperson wird darum gebeten, aus einem Set von Objekten die präferierte Option auszuwählen. Von diesen Wahlaufgaben werden der Antwortperson mehrere in unterschiedlicher Zusammenstellung der Alternativen pro Wahlset vorgelegt. Die Sets bestehen in der Regel aus drei oder vier alternativen Objektbeschreibungen, wobei eine Basisalternative oder eine „Kein Kauf“-Alternative wiederkehrend in den Sets enthalten sein kann. Das Antwortformat der diskreten Wahlen gilt gegenüber der auf Rangordnungen und Bewertungsskalen basierten Conjoint-Analyse als praxisnäher.

Das diskrete Wahlverhalten kann über unterschiedliche Modelle abgebildet werden, von denen sich das Multinomiale Logit-Modell (MNL-Modell) weitgehend durchgesetzt hat. Die Modellierung von diskretem Wahlverhalten geht davon aus, dass Entscheider die Option mit dem größten Zufallsnutzen auswählen. Es wird angenommen, dass sich der Nutzen U_i , den eine Person einer Alternative i zuschreibt, additiv zusammensetzt aus der deterministischen Komponente V_i der ausgedrückten Präferenz und einer Zufallskomponente ε_i , in der unbeobachtbare Einflüsse bzw. der generell probabilistische Charakter von Wahlen berücksichtigt werden. Unterschiedliche Annahmen über die Verteilung der stochastischen Zufallskomponenten ε_i führen zu unterschiedlichen Wahlmodellen. Die Annahme einer multivariaten Normalverteilung führt zum Multinomialen Probit-Modell (MNP-Modell), die Annahme einer Gumbel-Verteilung zum Multinomialen Logit-Modell (MNL-Modell).

Das MNL-Modell lässt sich auf das Wahlaxiom von Luce (1959) zurückführen, nachdem sich die Wahrscheinlichkeit der Wahl einer Alternative A_i aus dem Verhältnis des wahrgenommenen Nutzens der Alternative zu der Summe der Nutzenwerte aller Alternativen eines Wahlsets C ergibt:

$$P(A_i | C) = \frac{U_i}{\sum_{i' \in C} U_{i'}} \quad (4)$$

- mit:
- $P(A_i | C)$: Wahrscheinlichkeit der Wahl der Alternative i aus dem Wahlset C
 - U_i : Subjektiver Gesamtnutzen der Alternative i aus dem Wahlset C
 - $U_{i'}$: Subjektiver Gesamtnutzen der Alternative i' aus dem Wahlset C

Das Wahrscheinlichkeitsverhältnis von zwei Alternativen A_1 und A_2 ist nach diesem Axiom unabhängig von allen übrigen Alternativen im dem Wahlset und ergibt sich gemäß Formel (4) nach Kürzen der identischen Nennerterme alleine aus dem Quotienten der Nutzenwerte der beiden Alternativen, also aus dem Quotienten von U_1 und U_2 . Diese Eigenschaft des Wahlxioms heißt Independence from Irrelevant Alternatives, IIA, und bedeutet, dass Wahlsituationen nicht modelliert werden können, in denen beispielsweise eine neu eingeführte Alternative einer im Set enthaltenen ähnlichen Alternative mehr Wahlwahrscheinlichkeit entzieht als einer unähnlichen Alternative.

Das MNL-Modell unterliegt analog zum Luce-Axiom der IIA-Eigenschaft. Es stellt eine Erweiterung des Luce-Modells um einen stochastischen Fehlerterm mit Gumbel-Verteilung dar:

$$P(A_i | C) = \frac{e^{\mu V_i}}{\sum_{i' \in C} e^{\mu V_{i'}}} \quad (5)$$

- mit:
- $P(A_i | C)$: Wahrscheinlichkeit der Wahl der Alternative i aus dem Wahlset C
 - V_i : Deterministischer Gesamtnutzen der Alternative i aus dem Wahlset C
 - $V_{i'}$: Deterministischer Gesamtnutzen der Alternative i' aus dem Wahlset C
 - μ : Gumbel-Skalenparameter

Obwohl das Multinomiale Logit-Modell den Einschränkungen der IIA-Eigenschaft unterliegt, hat es sich aufgrund der einfacheren Berechnung der Parameter gegenüber dem Multinomialen Probit-Modell in der Praxis durchgesetzt. Die Schätzung der Nutzenwerte erfolgt im MNL-Modell in der Regel über das Maximum-Likelihood-Prinzip.

Diskrete Wahlaufgaben liefern keine Informationen über die Intensität der Präferenz oder die Präferenz unter den nicht gewählten Objekten. Es müssen sehr viele Wahlaufgaben vorgelegt werden, um zu stabilen Schätzungen der Teilnutzenwerte auf individueller Ebene zu gelangen. Huber und Zwerina (1996), Graßhoff et al. (2004) oder Großmann et al. (2005) haben effiziente Untersuchungsdesigns entwickelt, welche es bei einer noch überschaubaren Anzahl an Wahlsets pro Proband ermöglichen, individuelle Präferenzwerte zu ermitteln.

Mit dem Aufkommen leistungsstarker Computer und der entsprechenden Software haben sich im Wesentlichen zwei Verfahren zur Bestimmung von individuellen Nutzenwerten auf der

Basis von wahlbasierten Conjoint-Daten etabliert: die Hierarchische Bayes-Methode (HB) und die Individual Choice Estimation (ICE).

Die Hierarchische Bayes-Methode (HB) gelangt zu den individuellen Teilnutzenwerten durch eine Kombination der Informationen aus den ermittelten Teilnutzenwerten aller Antwortpersonen mit den spezifischen Wahlen des Einzelnen. Die Schätzung der individuellen Teilnutzenwerte erfolgt dabei mit Hilfe einer iterativen Gibbs-Sampling-Prozedur.

Einzelne, wissenschaftliche Anwendungen berichten valide Vorhersageergebnisse auf der individuellen Ebene mit Hilfe der Bayes-Methode (z. B. Moore, Gray-Lee & Louviere, 1998 oder Teichert, 2000). Toubia, Hauser und Simester (2003) nutzen die Bayes-Methode, um während des Antwortprozesses für jeden Probanden auf der Basis der bisherigen Wahlentscheidungen eine individualisierte Nutzenbalancierung der Wahlsets vorzunehmen. Dieses Vorgehen bezeichnen die Autoren als adaptive, wahlbasierte Conjoint-Analyse.

Das ICE-(Individual Choice Estimation)-Verfahren (Johnson, 1997) geht von einer Segmentierung der Antwortpersonen gemäß gleichartiger Präferenzstrukturen aus. Im Gegensatz zu klassischen Segmentierungsverfahren (s. Kapitel 2.5.1) wird aber nicht allen Segmentzugehörigen die gleiche Präferenzstruktur zugeordnet, sondern die für das Segment gültigen Nutzenwerte mit der individuellen Wahrscheinlichkeit, dass ein Proband diesem Segment zugehört, gewichtet. Einzelne Untersuchungen berichten von validen Ergebnissen (z. B. Orme & Hefl, 1999 oder Sawtooth, 1997).

In der durchgeführten Meta-Analyse wird überprüft, ob die Wahl des zur Ermittlung der Präferenzwerte auf individueller Ebene verwendeten Verfahrens einen Einfluss auf die Gütekriterien der Präferenzmessung ausübt.

2.4.3 Hybride Nutzenmessung

Hybride Conjoint-Modelle verbinden die kompositionelle Nutzenmessung mit der dekompositionellen Nutzenmessung, um die Bewertungsaufgaben für die Antwortpersonen auch bei einer höheren Anzahl an Attributen handhabbar zu gestalten. Die ersten Modelle der hybriden Nutzenmessung sehen vor, individuelle Nutzenfunktionen abzuleiten, in denen einige Aspekte der resultierenden Parameter auf der Basis der individuellen Daten und einige

auf der Basis der aggregierten Daten bestimmt werden. Heute ist es aufgrund effizienter Computerunterstützung möglich, vollständig individualisierte Teilnutzenwerte zu ermitteln.

In der hybriden Nutzenmessung werden von jeder Antwortperson drei verschiedene Arten von Daten erhoben: die Erwünschtheitsbewertungen der einzelnen Attributstufen, die Wichtigkeitsbewertungen der einzelnen Attribute sowie Präferenzurteile zu einer begrenzten Anzahl an Vollprofilen. Die optimale Gewichtung der direkt abgefragten Nutzenwerte im Verhältnis zu den conjoint-analytischen Nutzenwerten wird regressionsanalytisch bestimmt.

Eines der ursprünglichen Modelle der hybriden Nutzenmessung ohne Berücksichtigung von Interaktionstermen stellt sich wie folgt dar (Green, Goldberg & Montemayor, 1981):

$$U_i^{\text{hybrid}} = a + b U_i^{\text{komp}} + \sum_{j=1}^J \sum_{k=1}^{K_j} b_{jk} x_{ijk} \quad (6)$$

- mit:
- U_i^{hybrid} : Subjektiver Gesamtnutzen, den die Auskunftsperson dem Objekt A_i gemäß des Grundmodells der hybriden Nutzenmessung zumisst
 - U_i^{komp} : Subjektiver Gesamtnutzen, den die Auskunftsperson dem Objekt A_i gemäß der Formel (3) der kompositionellen Nutzenmessung zumisst
 - a : Abschnittsparameter
 - b : Regressionskoeffizient, der den Einfluss der direkten Nutzenparameter bestimmt
 - b_{jk} : Teilnutzenwert für die Ausprägungsstufe k des Attributs j (zu schätzende Modellparameter der dekompositionellen Nutzenmessung)
 - x_{ijk} : Dummy-Variable mit dem Wert 1, wenn die Eigenschaft k des Attributs j bei dem Objekt i auftritt; sonst 0.

In dem ersten Grundmodell der hybriden Nutzenmessung gemäß Formel (6) werden die dekompositionellen Parameter auf der aggregierten Gruppenebene und die kompositionellen Parameter auf der individuellen Ebene bestimmt. Neuere hybride Verfahren wie die *individuelle Hybrid-Conjoint-Analyse* (Baier & Säuberlich, 1997) können aufgrund effizienterer Designs und entsprechender Computerunterstützung alle Parameter auf der individuellen Ebene bestimmen. Weitere Verfahrensalternativen der hybriden Nutzenmessung sind die Customized Conjoint-Analysis (CCA) nach Srinivasan und Park (1997) und die Customized Computerized Conjoint-Analysis (CCC) nach Hensel-Börner und Sattler (2000).

Die *Customized Conjoint-Analysis (CCA)* nach Srinivasan und Park (1997) ermöglicht in einem ersten Schritt die Eliminierung von unakzeptablen Attributstufen, welche in der Folge vollständig von der Präferenzmessung ausgeschlossen werden. Für die verbleibenden Attributstufen werden die Erwünschtheitsbewertungen abgefragt, im Anschluss dann die Attributwichtigkeiten. Anders als im klassischen Ansatz der hybriden Nutzenmessung (Green, Goldberg, Montemayor, 1981) werden im dekompositionellen Teil der CCA nur die bis zu sechs wichtigsten Attribute, die sogenannten Kernattribute, berücksichtigt. Die conjoint-analytische Bewertungsaufgabe wird also an die einzelne Antwortperson angepasst, indem nur die individuell wichtigsten Attribute berücksichtigt werden.

Die Bestimmung der optimalen Gewichtung zwischen den dekompositionell und den kompositionell ermittelten Teilnutzenwerten für die Kernattribute erfolgt regressionsanalytisch. Hierzu werden in der Customized Conjoint-Analyse sogenannte Kalibrierungsprofile verwendet, die aus der Ermittlung der Präferenzdaten im ersten Schritt herausgehalten worden sind. Die gesamte Präferenzstruktur eines Probanden ergibt sich in der CCA für die Kernattribute aus der hybriden Nutzenmessung und für die übrigen Attribute ausschließlich aus der kompositionellen Nutzenmessung.

Die *Customized Computerized Conjoint-Analysis (CCC)* nach Hensel-Börner und Sattler (2000) erweitert die CCA in zwei Aspekten. Zum Einen erfolgt die Datenerhebung vollständig computergestützt, so dass der kompositionelle und der dekompositionelle Interviewteil direkt hintereinander durchgeführt werden können. Zum Anderen wird ein anderes Gewichtungsschema bei der Kombination der kompositionell und dekompositionell bestimmten Teilnutzenwerte verwendet.

Bei der CCA gehen die Nicht-Kerneigenschaften mit ihrem vollen Bedeutungsgewicht gemäß der direkt erklärten Nutzenwerte in die Präferenzstruktur ein, wohingegen die Kerneigenschaften mit der optimalen Gewichtung der self-explicated und der conjoint-analytischen Teilnutzenwerte eingehen. Letzteres kann dazu führen, dass einzelne Wichtigkeiten für die Kernattribute kleiner sind als einzelne Bedeutungsgewichte der Nicht-Kernattribute, insbesondere wenn in der Conjoint-Aufgabe einige Kernattribute als unwichtig im Vergleich zu den übrigen berücksichtigten Kernattributen eingestuft werden und der optimale Gewichtungsfaktor für die dekompositionell ermittelten Teilnutzenwerte hoch ist. In der CCC wird dieser Effekt vermieden, indem die direkt bestimmten Bedeutungsgewichte der

Nicht-Kernattribute ebenfalls mit dem optimalen Gewichtungsfaktor für die kompositionelle Nutzenmessung multipliziert werden.

Insgesamt haben sich auf der Basis der grundlegenden Idee der hybriden Nutzenmessung (Green, Goldberg & Montemayor, 1981; Green, 1984) zahlreiche Ansätze herausgebildet, von denen die adaptive Conjoint-Analyse (ACA; Johnson, 1987) aufgrund der Verfügbarkeit von Standardsoftware die größte Verbreitung erfahren hat. Im Gegensatz zu den vorab aufgeführten hybriden Conjoint-Analysen arbeitet ACA im dekompositionellen Teil der Datenerhebung nicht mit Vollprofilbewertungen im Sinne der traditionellen Conjoint-Analyse sondern mit metrisch skalierten Paarvergleichsaufgaben. In der durchgeführten Meta-Analyse werden die hybriden Verfahren getrennt von der adaptiven Conjoint-Analyse als eine separate Gruppe an Präferenzmessverfahren analysiert.

Die adaptive Conjoint-Analyse (ACA)

Die adaptive Conjoint-Analyse greift auf direkt abgefragte Nutzenwerte und Präferenzurteile zu Paarvergleichen zurück. Im Gegensatz zu klassischen Paarauswahlaufgaben werden bei ACA die Präferenzurteile mit dem metrischen Antwortformat einer Einstufungsskala abgefragt, wobei der Abstand von der Skalenmitte das Ausmaß der Präferenz für das bevorzugte Objekt repräsentiert.

Die Datenerhebung bei ACA erfolgt vollständig computergestützt und in den folgenden vier Schritten:

Zunächst wird die Antwortperson um die Beurteilung der Ausprägungsstufen der Attribute und anschließend um die Bewertung der Attributwichtigkeiten gebeten. Es handelt sich um ein einfaches 2-Stufen-Modell der kompositionellen Nutzenmessung, bei dem die abgegebenen rangordnungs- oder einstufigsskalenbasierten Präferenzwerte umskaliert werden auf eine Spannweite von $-0,5$ bis $+0,5$. Ist die Präferenzrangfolge der Attributausprägungen a priori bekannt, kann auf die Abfrage verzichtet werden.

Im Vorfeld kann die Antwortperson aufgefordert werden, nicht akzeptable Attributstufen zu eliminieren, welche dann nicht weiter berücksichtigt werden und damit das ACA-Interview verkürzen. ACA modelliert im Sinne einer kompensatorischen Entscheidungsregel aber auch die nicht akzeptierten Attributstufen und ordnet diesen sehr niedrige Teilnutzenwerte zu.

Im dekompositionellen Interviewteil werden eine begrenzte Anzahl an metrischen Paarvergleichsaufgaben vorgelegt, die adaptiv unter Berücksichtigung der bereits abgegebenen Präferenzurteile der Antwortperson zusammengestellt werden. In einem letzten Schritt der Datenerhebung muss die Antwortperson ihre Kaufwahrscheinlichkeit zu ausgewählten Vollprofil-Beschreibungen, den sogenannten Kalibrierungskonzepten, angeben. Die Vollprofil-Bewertungen dienen dazu, eine optimale Integration der aus dem kompositionellen und dem dekompositionellen Teil der Befragung gewonnenen Teilnutzenwerten zu erreichen.

Um die Anzahl der Paarvergleiche auch bei Untersuchungsplänen mit vielen Attributen gering zu halten, erfolgt die Präferenzmessung adaptiv. Zunächst wird sichergestellt, dass die Attribute möglichst gleich häufig in die Objekte eingehen und die Attributstufen möglichst gleich häufig verwendet werden. Aus den dann zur Verfügung stehenden Attributstufen werden diejenigen ausgewählt, die einen möglichst gleichen Gesamtnutzen der beiden Alternativen zur Folge haben. Die erwarteten Nutzenwerte der Objekte ergeben sich dabei adaptiv aus den zuvor aus dem kompositionellen Interviewteil ermittelten Präferenzwerten, die nach jeder zusätzlichen Paarvergleichsaufgabe neu geschätzt werden. Großmann (2003) hat optimale Designs für metrische Paarvergleichsaufgaben in der Conjoint-Analyse vorgelegt, mit denen die Anzahl der erforderlichen Paarvergleichsaufgaben weiter reduziert werden kann.

2.5 Die Weiterverarbeitung der individuellen Präferenzstrukturen

Die Conjoint-Analyse wird in der Regel auf individueller Ebene ausgewertet, um der Heterogenität von Präferenzen angemessen gerecht zu werden. Sind die individuellen Präferenzstrukturen bekannt, so lassen sich Wahlen simulieren und damit auf der aggregierten Ebene Wahl- bzw. Marktanteile vorhersagen. Für die Wahlanteilssimulation existieren unterschiedliche Methoden, zu denen das Maximum-Utility-Modell, das BTL-Modell, das Logit-Modell und das Probit-Modell gehören. Diese werden im Unterkapitel 2.5.2 näher beschrieben.

Eine weitere aus Anwendersicht wichtige Weiterverarbeitung von individuellen Präferenzstrukturen stellt die Aggregation auf Segmentebene dar, beispielsweise für die Durchführung von Marktsegmentierungen und zur Identifikation von in sich homogenen

Konsumentengruppen. Die in der Meta-Analyse unterschiedenen Verfahren der Faktorenanalyse, der Clusteranalyse, der a-priori-Segmentierung und der latent-class-Segmentierung werden im folgenden Unterkapitel genauer vorgestellt.

2.5.1 Die Segmentierungsverfahren

Der erste Ansatz zur Identifikation von homogenen Konsumentengruppen nutzt das a priori über die Antwortpersonen vorhandene Wissen (z. B. Wind et al., 1989; Sentis & Li, 2002). Beispielsweise können die Segmente auf der Basis von soziodemographischen Daten der Antwortpersonen oder direkt abgefragter oder beobachteter Kaufverhaltensinformationen der Probanden gebildet werden. Für jedes der festgelegten a-priori-Segmente wird dann eine Präferenzstruktur bestimmt. Die *a-priori-Segmentierung* verspricht gute Ergebnisse, wenn die Hintergrundinformationen einen engen Zusammenhang mit den individuellen Präferenzwerten aufweisen.

Bei den traditionellen zweistufigen Segmentierungsverfahren werden zunächst die Präferenzstrukturen auf individueller Ebene ermittelt, um dann in einem zweiten Schritt eine Einteilung der Antwortpersonen auf der Basis der Ähnlichkeit der Teilnutzenwerte vorzunehmen. Die Gruppierung erfolgt post-hoc mittels einer *Clusteranalyse*, in welche die Präferenzwerte als Inputdaten eingehen (z. B. Green & Krieger, 1991). Diese Vorgehensweise unterliegt der Gefahr, dass die Zuordnung der Probanden zu den Segmenten abhängig ist von der Zuverlässigkeit der Schätzung der individuell ermittelten Präferenzparameter. Ein alternatives zweistufiges Segmentierungsverfahren sieht daher vor, in einem ersten Schritt eine Clusteranalyse auf der Basis der Rohdaten der Präferenzmessung vorzunehmen, um dann in einem zweiten Schritt für jedes Segment über alle Antwortpersonen eine separate Präferenzstruktur zu bestimmen (z. B. Green & Srinivasan, 1978).

Sowohl die a-priori-Segmentierung als auch die zweistufige Segmentierung mittels Clusteranalyse unterliegen der Kritik, dass in einem ersten Schritt versucht wird, die Homogenität der Probanden innerhalb der Segmentierungslösung zu optimieren und in einem zweiten, davon unabhängigen Schritt die Passung der Präferenzstruktur innerhalb des Segments zu optimieren. Die von Hagerty (1985) vorgeschlagene Methode hingegen nimmt

die Segmentierung und die Schätzung der Conjoint-Parameter simultan vor, indem ein optimales Gewichtungsschema bestimmt wird, das den erwarteten Vorhersagefehler in einem Validierungsset minimiert. Das Gewichtungsschema stellt eine optimale Einteilung der Probanden dar und wird durch eine *Faktorenanalyse* bestimmt. Dabei werden keine diskreten, nicht-überlappenden Segmente gebildet, sondern grundsätzlich gehört jeder Proband zu allen Segmenten, wobei das Ausmaß der Zugehörigkeit zu den unterschiedlichen Segmenten durch den ersten Eigenvektor der Korrelationsmatrix der beobachteten Rohdaten der Präferenzmessung über alle Probanden bestimmt wird.

Die von DeSarbo et al. (1992) vorgeschlagene *latent-class-Segmentierung* sieht ebenfalls die simultane Bestimmung der Segmente und der Parameter der Präferenzstruktur für die Segmente vor. In einem iterativen Prozess aus Regressions- und Clusteranalysen werden die Segmente und deren Teilnutzenwerte geschätzt. Die Gruppierung eines Probanden zu den verdeckten Segmenten erfolgt über die Maximierung der a posteriori geschätzten Wahrscheinlichkeiten, dass dieser Proband zu einem der Segmente gehört. DeSarbo, Ramaswamy und Cohen (1995) haben eine latent-class-Segmentierung für wahlbasierte Conjoint-Daten vorgelegt.

2.5.2 Die Wahlanteilssimulatoren

Zur Simulation von Wahlen müssen die individuellen Präferenzstrukturen der Antwortpersonen in Wahlvorhersagen konvertiert werden. Die individuell vorhergesagten Wahlwahrscheinlichkeiten können dann auf der aggregierten Ebene als Wahl- bzw. Marktanteile interpretiert werden. Zur Transformation der individuellen Präferenzdaten in Auswahldaten stehen unterschiedliche Wahlanteilssimulatoren zur Verfügung, die bei Currim (1982) oder Finkbeiner (1988) ausführlich besprochen werden. Im Folgenden werden die in der Meta-Analyse berücksichtigten Wahlanteilssimulatoren beschrieben.

Das *Maximum-Utility-Modell*, auch als First-Choice-Regel bezeichnet, geht davon aus, dass ein rational entscheidender Konsument diejenige Alternative auswählt, welche unter den zur Verfügung stehenden Objekten den größten Nutzen stiftet. Die Wahlanteilsschätzung für ein Objekt ist der proportionale Anteil der Stichprobe, für welchen das Objekt den größten

Nutzen verspricht. Wenn P_{hi} eine Funktion bezeichnet, die jedem Objekt A_i in einem Gesamtszenario eine Wahlwahrscheinlichkeit zuordnet, dann gilt das Maximum-Utility-Modell wie folgt:

$$P_{hi} = \begin{cases} 1 & \text{falls } U_{hi} \geq U_{hi'} \\ 0 & \text{sonst} \end{cases} \quad (7)$$

mit: U_{hi} : Subjektiver Gesamtnutzen der Alternative i für die Antwortperson h
 $U_{hi'}$: Subjektiver Gesamtnutzen einer anderen Alternative i' aus allen zur Wahl stehenden Alternativen I für die Antwortperson h

Das Maximum-Utility-Modell wird bisweilen als zu deterministisch kritisiert, da bei dieser Vorgehensweise nur die Auswahlwahrscheinlichkeiten 1 und 0 vergeben werden und damit die Stärke der Präferenz gegenüber den übrigen Wahlalternativen nicht modelliert wird. Sind die Teilnutzenwerte einmal bestimmt, dann werden die individuellen Wahlvorhersagen deterministisch abgeleitet. Das Auswahlverhalten in einer realen Umgebung unterliegt jedoch zahlreichen Unsicherheitsfaktoren, so dass keineswegs sicher ist, dass in jedem Fall das Produkt mit dem höchsten Gesamtnutzen auch tatsächlich erworben wird. Die anderen Wahlsimulatoren fassen das Wahlverhalten probabilistisch auf.

Nach dem *BTL-Modell* (Bradley & Terry, 1952; Luce, 1959) ergibt sich die Auswahlwahrscheinlichkeit P_{hi} eines Objekts aus dem Quotienten des subjektiven Gesamtnutzens dieses Objekts und der Summe der subjektiven Nutzenwerte aller zur Wahl stehenden Alternativen I . Damit wird Objekten mit einem höheren Nutzen auch eine höhere Auswahlwahrscheinlichkeit zugeordnet, ohne jedoch die Wahl des Objekts mit dem größten Gesamtnutzen als sicher zu unterstellen:

$$P_{hi} = \frac{U_{hi}}{\sum_{i'=1}^I U_{hi'}} \quad (8)$$

Das *Logit-Modell* bestimmt die Auswahlwahrscheinlichkeit P_{hi} eines Objekts ebenfalls als das Verhältnis aus dem subjektiven Gesamtnutzenwert des Objekts und der Summe der Nutzenwerte aller zur Auswahl stehenden Objekte, setzt die Nutzenwerte jedoch im Gegensatz zur BTL-Regel in den Exponenten der Eulerschen Zahl e :

$$P_{hi} = \frac{e^{U_{hi}}}{\sum_{i'=1}^I e^{U_{hi'}}} \quad (9)$$

Die Nutzenwerte, welche als Inputdaten in die Wahlanteilssimulatoren eingehen, sind im besten Fall intervallskaliert. Die unter der Annahme eines Intervallskalenniveaus zulässigen linearen Transformationen verändern jedoch die Wahlanteilsvorhersagen im Logit- und BTL-Modell (Green & Krieger, 1988). Die Addition einer beliebigen Konstante zu den Nutzenwerten verändern die Wahlanteilsvorhersagen des BTL-Modells, die Multiplikation mit einer beliebigen Konstante verändert die Wahlanteilsvorhersagen des Logit-Modells. Die Maximum-Utility-Regel ist gegen lineare Transformationen der Nutzenwerte invariant. Weiterhin unterliegen das BTL-Modell und das Logit-Modell der IIA-Eigenschaft (s. Kapitel 2.4.2), was dazu führt, dass beide Modelle Situationen mit ähnlichen Alternativen im Auswahlset nicht angemessen handhaben.

Davey und Elrod (1991) haben ein *Probit-Modell* vorgeschlagen, das nicht die IIA-Eigenschaft aufweist und Wahlanteilsschätzungen liefert, die gegen lineare Transformationen invariant sind. Das Logit- und das Probit-Modell basieren als probabilistische Wahlanteilssimulatoren auf den Annahmen der Zufallsnutzentheorie, nach der sich der Nutzen einer Alternative additiv zusammensetzt aus einer durch die Teilnutzenwerte bestimmten, deterministischen Komponente und einer Zufallskomponente ε_i , in der unbeobachtbare Einflüsse bzw. der generell probabilistische Charakter von Wahlen berücksichtigt werden. Unter der Annahme einer unabhängigen und identischen Verteilung der Fehler (Gumbelverteilung) resultiert das Logit-Modell. Das Probit-Modell nimmt an, dass die Fehlerterme multivariat normalverteilt sind.

Bei wahlbasierten Conjoint-Daten ist der Einsatz eines Wahlsimulators nicht notwendig, da die Auswahlwahrscheinlichkeiten direkt aus dem zugrundeliegenden Logit- oder Probit-Modell abgeleitet werden können und die Schätzung der Nutzenwerte auf der aggregierten Ebene der Gesamtstichprobe erfolgt.

In der vorliegenden Meta-Analyse werden die oben umschriebenen Wahlanteilssimulatoren unterschieden und überprüft, in wieweit sie reale Wahlanteile treffend modellieren bzw. vorhersagen können.

3. Die Reliabilität und Validität der Präferenzstrukturmessung

Die Abgrenzung zwischen Reliabilitäts- und Validitätsmessung wird in der Präferenzstrukturmessung nicht einheitlich vorgenommen, so dass es nicht verwundert, wie vielfältig und widersprüchlich die Begriffe in der Literatur verwendet werden. In dieser Arbeit soll auf der Basis der Empfehlungen von Bateson, Reibstein und Boulding (1987) die Generalisierbarkeitstheorie (G-Theorie; Cronbach et al., 1972) als Rahmenkonzept zur Systematisierung der zahlreichen Studien zur Bestimmung der Gütekriterien von Präferenzmessverfahren verwendet werden.

Die Reliabilität eines Verfahrens zur Präferenzstrukturmessung bezeichnet den Grad der Genauigkeit, mit dem das Verfahren die Präferenzstruktur messen kann. Die Reliabilität ist umso höher, je geringer die Varianz der Messfehler ausfällt. Die in der Präferenzstrukturmessung üblichen Formen der Reliabilitätsbestimmung werden im Unterkapitel 3.2 besprochen.

Die Validität eines Verfahrens zur Präferenzstrukturmessung bezeichnet das Ausmaß der Fähigkeit des Verfahrens, genau das zu messen, was es zu messen vorgibt. Es soll möglichst exakt der wahre Wert des hypothetischen Konstrukts der Präferenz abgebildet werden. Ein systematischer Fehler der Messung soll möglichst klein sein. Die in der Präferenzstrukturmessung üblichen Vorgehensweisen zur Bestimmung der Validität eines Messverfahrens werden im Unterkapitel 3.3 besprochen.

3.1 Die Generalisierbarkeitstheorie als Rahmenkonzept zur Reliabilitätsbestimmung

Zur Bestimmung der Reliabilität von Präferenzmessverfahren wurde in den meisten bisherigen Untersuchungen die klassische Testtheorie (z. B. Lord & Novick, 1974) als Rahmenkonzept zugrunde gelegt, wonach sich ein beobachteter Werte additiv aus dem wahren Wert und einem Messfehler zusammensetzt. Es wird angenommen, dass für eine zufällig ausgewählte Person der Erwartungswert des Messfehlers gleich 0 ist, und der Messfehler nicht mit dem wahren Wert, einer anderen Variablen oder dem Messfehler einer anderen Variablen korreliert ist. Als Maßzahl für den relativen Einfluss von wahren Wert bzw. Messfehler dient der Reliabilitätskoeffizient, der als Quotient der Varianz der wahren Werte und der Varianz der beobachteten Werte festgelegt ist. Da die Varianz der wahren Werte

maximal mit der Varianz der beobachteten Werte identisch ist, kann der Reliabilitätskoeffizient nur Werte zwischen 0 und 1 annehmen. Diese klassische Definition der Reliabilität bezieht sich, ebenso wie die vier Annahmen über die Eigenschaften des Messfehlers, in der klassischen Testtheorie ausschließlich auf quantitative, mindestens intervallskalierte Personenvariablen.

Intervallskalierte Personenvariablen liegen bei der Präferenzstrukturmessung entweder beim Einsatz von Einstufungsskalen als abhängige Variable auf der Ebene der Inputdaten oder als geschätzte Teilnutzenwerte auf der Ebene der Outputdaten vor. Für alle anderen Formen der in der Präferenzstrukturmessung anzutreffenden Messgütebestimmungen kann die klassische Testtheorie nicht als Rahmenkonzeption dienen, da die Messdaten von Wahl- und Rangordnungsaufgaben kein Intervallskalenniveau erreichen. Dennoch sind in der Literatur zahlreiche Studien anzutreffen, die beispielsweise die Korrelation zwischen zwei Inputrangreihen als Reliabilitätsmaß heranziehen (z. B. Weisenfeld, 1989; Vriens, 1995).

Eine Synthese der diversen Erkenntnisse zur Reliabilität von Präferenzmessverfahren muss also immer auch berücksichtigen, unter welchen theoretischen Annahmen und mit welchen praktischen Vorgehensweisen die Reliabilitätskoeffizienten in den jeweiligen Untersuchungen ermittelt worden sind. Eine zentrale Unterscheidung ist dabei, ob die Reliabilität auf der Ebene des Antwortvektors oder der Ebene der Teilnutzenwerte bestimmt wurde. Auf der Ebene des Antwortvektors wird die Zuverlässigkeit des Messverfahrens überprüft, in wieweit die Probanden zu den Stimuli der Präferenzmessung bei einer zweiten Messung konsistente Antworten geben. Im Kontrast dazu bestimmt die Reliabilitätsmessung auf der Ebene der geschätzten Parameter die Zuverlässigkeit des gesamten Schätzverfahrens. Aufgrund des kompensatorischen Charakters der Conjoint-Analyse ist es denkbar, dass die Präferenzen für ganzheitliche Objekte sehr stabil, die einzelnen Teilnutzenwerte jedoch weniger stabil gemessen werden.

Bateson, Reibstein und Boulding (1987) haben gezeigt, dass die Ergebnisse von Reliabilitätsuntersuchungen von Präferenzmessverfahren uneinheitlich sind, und die Reliabilitätskoeffizienten eine erhebliche Spannweite von .36 bis .92 aufweisen. Es scheinen bei der Präferenzstrukturmessung zahlreiche Fehlerquellen vorzuliegen, welche die Zuverlässigkeit der Messung systematisch gefährden können.

Im Gegensatz zur klassischen Testtheorie wird der Messfehler in der Generalisierbarkeitstheorie weiter aufgespaltet. Die Generalisierbarkeitstheorie bietet eine Rahmenkonzeption für die Untersuchung und Konstruktion von zuverlässigen Tests. Sie wurde ursprünglich von Cronbach et al. (1972) als Antwort auf die Restriktionen des traditionellen Wahren-Wert-Modells der klassischen Testtheorie entwickelt und kann als deren Weiterentwicklung gelten.

Es wird in der Generalisierbarkeitstheorie untersucht, zu welchem Grad ein Test für bestimmte systematische Fehlerquellen, wie beispielsweise den Einfluss des Untersuchers, der Testsituation, der Itemeigenschaften etc., anfällig ist. Diese Fehlerquellen werden in der G-Theorie als Facetten bezeichnet. Mit der Durchführung von Untersuchungen, in denen potentielle Fehlerquellen systematisch variiert und simultan erhoben werden, wird der Einfluss der unterschiedlichen Facetten bestimmt. Die verwendeten Facetten können zusammen mit dem sogenannten Objekt der Messung als Varianzanalyse mit Random Effects modelliert werden. Die Facetten sowie das Objekt der Messung gehen als Faktoren bzw. Haupteffekte in die Varianzanalyse mit ein; darüber hinaus werden Interaktionen von zwei oder mehr Facetten berücksichtigt, die als geteilter Einfluss oder Schnittmenge von zwei oder mehr Facetten interpretiert werden.

Der universelle Wert einer Person μ_p als der messfehlerfreie, wahre Wert einer Person in der G-Theorie ist definiert als der Mittelwert des Messobjekts über die Facetten, die im Universum der Generalisierbarkeit repräsentiert sind. Die Aufgabe einer Messung ist die akkurate Schätzung des universellen Werts μ_p auf der Basis einer Stichprobe von Beobachtungen.

Das Ausmaß, in wieweit eine Messung generalisierbar ist, hängt davon ab, wie akkurat die Messung den universellen Wert schätzen kann. Die Genauigkeit einer Messung wird über die Varianz der Messwerte aus den als äquivalent angenommenen Messungen bestimmt. Die Stärke des Einflusses von unterschiedlichen Stimuli, Messverfahren, Beurteilern, Situationen etc. auf die beobachteten Messwerte bestimmt die Abhängigkeit der Messung von den Umgebungseinflüssen. Mit anderen Worten, wenn unterschiedliche Stimuli, Beurteiler etc. zu ähnlichen Beobachtungen führen, kann eine einzelne Messung als angemessen gelten, um auf das Universum an Beobachtungen zu schließen und die Messergebnisse zu generalisieren.

Ein varianzanalytisches Modell wird herangezogen, um zu beschreiben, wie der beobachtete Wert X_{pio} einer Person p in einem dem wahren Wert entsprechenden Teil und die Fehleranteile zerlegt wird.

Ein die beiden Facetten der Art des Items i und des Zeitpunkts o berücksichtigendes *Person x Item x Zeitpunkt* - Design stellt sich beispielsweise wie folgt dar.

$$X_{pio} = \mu + \mu_p^* + \mu_i^* + \mu_o^* + \mu_{pi}^* + \mu_{po}^* + \mu_{io}^* + \mu_{pio,e}^* \quad (10)$$

- mit:
- μ universeller Mittelwert, konstant für alle Personen
 - μ_p^* Testeffekt von Person p ; beschreibt die interindividuellen Differenzen in den universellen Werten als Differenz ($\mu_p^* - \mu$)
 - μ_i^* Testeffekt von Item i ; beschreibt den Fehlereinfluss dieser Facette
 - μ_o^* Testeffekt von Zeitpunkt o
 - μ_{pi}^* Testeffekt der Interaktion von Person p und Item i
 - μ_{po}^* Testeffekt der Interaktion von Person p und Zeitpunkt o
 - μ_{io}^* Testeffekt der Interaktion von Item i und Zeitpunkt o
 - $\mu_{pio,e}^*$ Residuums-Effekt, bestehend aus der Dreifach-Interaktion und dem zufälligen Fehler

Der universelle Wert μ_p einer Person entspricht dem Erwartungswert E der Testwerte über die Facetten und ergibt sich aus der Summe von μ_p^* und μ :

$$\mu_p = E_i E_o X_{pio} = \mu + \mu_p^* \quad (11)$$

Die Varianz der universellen Werte aller Personen in der Population wird bezeichnet als σ_p^2 und entspricht der Varianz der wahren Werte in der klassischen Testtheorie. Der Generalisierbarkeitskoeffizient G ist wie folgt definiert:

$$G := \frac{\sigma_p^2}{E\sigma_x^2} \quad (12)$$

- mit:
- σ_p^2 = Varianz der universellen Werte
 - $E\sigma_x^2$ = Erwartungswert der Varianz der Testwerte über die Facetten

Die traditionelle Reliabilitätstheorie geht davon aus, dass sich jeder beobachtete Messwert auf einen wahren Wert bezieht und aus einem einheitlichen, parallelen Universum von Beobachtungen stammt. Diese Annahme ist haltbar unter der Voraussetzung, dass die der Reliabilitätsbestimmung zugrunde liegenden Messungen exakt parallel sind, also die gleichen Mittelwerte, Varianzen und Interkorrelationen aufweisen.

In der Praxis ist diese Äquivalenzannahme der klassischen Testtheorie jedoch durch die unterschiedlichen Fehlerquellen gefährdet. Im Sinne der G-Theorie sind leichte Variationen der beiden der Reliabilitätsbestimmung zugrunde liegenden Messungen hingegen erwünscht, um den Einfluss der Fehlerquelle quantifizieren zu können.

3.2 Übliche Vorgehensweisen zur Reliabilitätsbestimmung in der Präferenzmessung

Die Abgrenzung zwischen Reliabilitäts- und Validitätsbestimmung wird in der Präferenzmessung nicht einheitlich vorgenommen. Zudem zeigt sich, dass vielfach Reliabilitätsmessungen im Rahmenkonzept der klassischen Testtheorie vorgenommen werden, die mit den Grundlagen der klassischen Testtheorie nicht in Einklang zu bringen sind. Zur Systematisierung der Befunde zur Reliabilität von Präferenzmessverfahren kann die G-Theorie als Rahmenkonzept dienen. Die G-Theorie verlangt keine Äquivalenzannahme und keine intervallskalierten Personenvariablen und spaltet die beobachtete Fehlervarianz in die unterschiedlichen Quellen auf.

Für die Präferenzstrukturmessung unterscheiden Bateson, Reibstein und Boulding (1987) bzw. in deren Erweiterung Vriens (1995) die folgenden acht systematischen Fehlerquellen:

- Generalisierbarkeit über die Zeit: Die gleichen Aufgaben werden den gleichen Probanden zu einem zweiten Zeitpunkt erneut vorgelegt und die Übereinstimmung zwischen den beiden Datenreihen auf Input- oder Outputebene festgestellt. Diese Form der Reliabilitätsmessung wird auch als *Test-Retest-Reliabilität* bezeichnet. Entscheidend ist bei dieser Form der Reliabilitätsbestimmung der zeitliche Abstand zwischen den beiden Messzeitpunkten. Er sollte groß genug sein, um Erinnerungseffekte ausschließen zu können, aber klein genug, um strukturelle Veränderungen in den Präferenzen der Probanden nicht mit einzubeziehen. Die in

der Forschung verwendeten Zeitintervalle reichen von wenigen Minuten, wenn beispielsweise die gleichen Aufgaben am Anfang und am Ende der Präferenzmessung wiederholt vorgelegt werden (z. B. Hansen & Wittink, 1995; Orme & King, 1998), bis hin zu zehn Monaten (Singh et al., 1998).

- Generalisierbarkeit über das Attributset: Das in der Präferenzmessung verwendete Set an Attributen kann minimal variiert und dann die Übereinstimmung der Beurteilungen für die in beiden Messungen vorhandenen Attribute bestimmt werden. Die Variation kann dabei derart erfolgen, dass ein Attribut durch ein neues Attribut im Set ersetzt wird (structural analysis; McCullough & Best, 1979; Baalbaki & Malhotra, 1995) oder ein neues Attribut hinzugefügt wird (embedding analysis).
- Generalisierbarkeit über die Ausprägungsstufen: Hier werden die Ausprägungsstufen der verwendeten Attribute bei zwei Messungen variiert. Beispielsweise kann bei einer Messung ein Attribut gleichabständige Ausprägungsstufen aufweisen und bei der anderen Messung ungleiche Abstände. Alternativ kann die Anzahl der Ausprägungsstufen eines Attributs variiert werden (z. B. Green, Wind & Jain 1972; Perrey, 1996).
- Generalisierbarkeit über das Stimulusset: Bei der Verwendung von fraktionierten Designs können den Probanden bei der zweiten Messung andere Stimuli aus dem gleichen Gesamtdesign vorgelegt werden (z. B. Parker & Srinivasan, 1976).
- Generalisierbarkeit über die Methode der Stimulusdarbietung: In den beiden Messungen können die Stimuli auf zwei unterschiedliche Weisen präsentiert werden, um den Einfluss dieser Facette zu bestimmen. Dahan und Srinivasan (2000) verglichen beispielsweise die Präferenzwerte für Prototypen, wenn diese im Internet oder real präsentiert wurden. Sattler (1994) untersuchte, ob reale oder hypothetische Prototypen eine unterschiedliche Wirkung auf die Antwortpersonen haben. Louviere et al. (1986) überprüften für Naturparks, ob Präferenzbestimmungen mit verbalen Umschreibungen der Parks zu den gleichen Ergebnissen gelangen wie Beurteilungen von Fotos.
- Generalisierbarkeit über die abhängigen Variablen: Es können Variationen der abhängigen Variablen vorgenommen und die ermittelten Teilnutzenwerte miteinander verglichen werden.

- Generalisierbarkeit über den Datenerhebungsprozess: Die Präferenzdaten können über eine Computerinteraktion, ein Telefoninterview, einen Fragebogen oder ein persönliches Interview erhoben werden. Mit dem Aufkommen der Möglichkeit, Präferenzdaten über das Internet zu erheben, wurde mehrfach überprüft, wie zuverlässig diese Art des Datenerhebungsprozesses ist (z. B. Meyer, 1998; Orme & King, 1998; Melles, Laumann & Holling, 2000; Dahan et al., 2001).
- Generalisierbarkeit über die Datenerhebungsmethode: Mit Datenerhebungsmethode wird hier das eingesetzte Verfahren zur Präferenzmessung, also beispielsweise eine traditionelle Conjoint-Analyse gegenüber einer hybriden Conjoint-Analyse, verstanden. Mit dieser Vorgehensweise wird der Einfluss des eingesetzten Präferenzmessverfahrens auf die Zuverlässigkeit der Parameterschätzungen überprüft. Ein Vergleich von Teilnutzenwerten, die mit zwei unterschiedlichen Präferenzmessverfahren ermittelt wurden, soll in dieser Arbeit als eine Untersuchung der konvergenten Validität der Verfahren gelten (s. Kapitel 3.3).

In der Präferenzstrukturmessung kommen im Wesentlichen drei Methoden für die Bestimmung der Zuverlässigkeit des Messinstruments zur Anwendung: die bereits angesprochene Test-Retest-Reliabilität, die Paralleltest-Reliabilität und die Test-Halbierungs-Reliabilität. Von den in der Meta-Analyse erfassten Reliabilitätsmaßen waren 72 % Test-Retest-Reliabilitätsmaße, 21 % Paralleltest-Reliabilitätsmaße und 7 % Test-Halbierungs-Reliabilitätsmaße.

Bei der *Paralleltest-Reliabilität* wird an der gleichen Stichprobe eine zweite Präferenzmessung mit einem exakt parallelen Verfahren vorgenommen. In der Präferenzstrukturmessung ist es aufgrund der zahlreichen Variationsquellen schwierig, exakt parallele Tests zu entwickeln, die den Äquivalenzanforderungen der klassischen Testtheorie entsprechen. Die in der Literatur anzutreffenden Untersuchungen zur Zuverlässigkeit von Messungen, wenn minimale Änderungen beispielsweise im Stimuluset vorgenommen werden, müssten streng genommen mit der G-Theorie analysiert werden.

Die *Test-Halbierungs-Reliabilität* verlangt im Gegensatz zur Paralleltest- und zur Test-Retest-Reliabilität keinen untersuchungstechnischen Mehraufwand. Es werden für jede Antwortperson zwei Präferenzstrukturen ermittelt, jeweils beruhend auf der Hälfte der

abgegebenen Präferenzdaten. Da die Reliabilität eines Tests mit zunehmender Testlänge zunimmt, unterschätzt die Test-Halbierungs-Methode die Reliabilität des Gesamttests systematisch. Mit der Spearman-Brown-Formel (vgl. z. B. Lienert & Ratz, 1998) kann der durch die Test-Halbierungs-Methode gewonnene Reliabilitätswert nachträglich für den Gesamttest aufgewertet werden. In der Präferenzstrukturmessung haben beispielsweise Scott und Wright (1976), Heeler, Okechuku und Reid (1979), Weisenfeld (1989) sowie Melles, Laumann und Holling (2000) die Test-Halbierungs-Methode zur Bestimmung der Reliabilität verwendet.

Die Abbildung 3.1 veranschaulicht die Abgrenzung von Modellanpassung und Reliabilität bei der Präferenzstrukturmessung (Reiners, 1996).

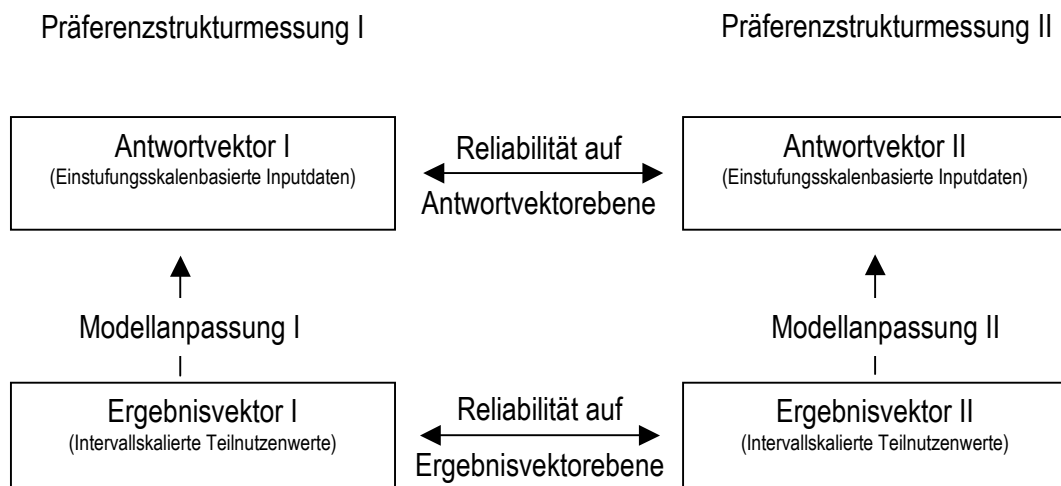


Abb.3.1: *Modellanpassung und Reliabilität bei der Präferenzstrukturmessung*
(Pfeil: Vorhersage; Doppelpfeil: Zusammenhang)

In Tabelle 3.1 werden die vier Arten der Reliabilitätsbestimmung, die in der Meta-Analyse berücksichtigt worden sind, in einer Übersicht zusammengefasst. Es ist erkennbar, dass die Bestimmung der Übereinstimmung von Teilnutzenwerten auf der Ergebnisvektorebene bei den in der Meta-Analyse berücksichtigten Studien am häufigsten als Reliabilitätsmaß angewendet worden ist.

Vorgehen zur Reliabilitätsbestimmung	Effektgröße	Anzahl der berücksichtigten Publikationen	Anzahl der berücksichtigten Effektgrößen	Prototypische Publikation ^a
Wiederholte Bewertung von identischen Vollprofilen mittels Einstufungsskala	Produkt-Moment-Korrelation (Pearson's r) zwischen den Einstufungen der ersten und der zweiten Datenerhebung, gemittelt über alle Probanden	9	39	Green, Krieger & Agarwal (1993)
Wiederholte Rangordnung von identischen Vollprofilstapeln	Rang-Korrelation (Spearman's ρ oder Kendall's τ) zwischen den Rangordnungen der ersten und der zweiten Datenerhebung, gemittelt über alle Probanden	9	27	Weisenfeld (1989)
Wiederholte Bestimmung von Teilnutzenwerte über eine weitere oder parallele Präferenzmessung	Produkt-Moment-Korrelation zwischen den bei der ersten und der zweiten Messung ermittelten Teilnutzenwerten, gemittelt über alle Probanden	22	91	Leigh, MacKay & Summers (1981)
Wiederholte Bestimmung von Attributwichtigkeiten über eine weitere oder parallele Präferenzmessung	Produkt-Moment-Korrelation zwischen den bei der ersten und der zweiten Messung ermittelten Attributgewichten, gemittelt über alle Probanden	3	10	Singh et al. (1998)

^a Eine detaillierte Aufstellung der jeweils berücksichtigten Studien findet sich im Anhang.

Tab.3.1: Übersicht über die Häufigkeiten der in der Meta-Analyse verwendeten Effektgrößen der Reliabilitätsbestimmung der Präferenzmessung

3.3 Übliche Vorgehensweisen zur Validitätsbestimmung in der Präferenzmessung

Die Validität eines Messverfahrens gibt an, in wie weit ein wahrer Wert ohne einen systematischen Fehler erhoben wird. Das relevante theoretische Konstrukt soll erschöpfend erfasst werden. Da es nicht möglich ist, eine absolute Lösung des Gültigkeitsproblems eines Messinstruments für ein theoretisches Konstrukt wie die Präferenz zu finden, muss sich die Validitätsbestimmung der Präferenzstrukturmessung auf die Überprüfung der Plausibilität beschränken, in wie weit das jeweilige Präferenzmessverfahren auch tatsächlich Präferenz misst.

3.3.1 Validitätsarten

Im Allgemeinen werden drei Arten der Validitätsbestimmung unterschieden: die Inhaltsvalidität, die Konstruktvalidität und die Kriteriumsvalidität (z. B. Lienert & Raatz, 1998; Müller-Hagedorn et al., 1993).

Die *Inhaltsvalidität* ist gegeben, wenn die Datenerhebung augenscheinlich die zu messende Präferenz erschöpfend erfasst. In der Präferenzstrukturmessung ist in diesem Sinne beispielsweise sicherzustellen, dass alle nutzenstiftenden Eigenschaften eines Objekts in der Präferenzmessung berücksichtigt werden. Weiterhin sind in der Präferenzstrukturmessung einige logische Plausibilitätsüberprüfungen üblich, wie beispielsweise der Vorzeichentest, mit dem überprüft wird, ob die ermittelten Teilnutzenwerte begründeten a-priori-Erwartungen entsprechen (z. B. Acito & Jain, 1980; Baalbaki & Malhotra, 1995; Ryan, 1999). Analog kann die Übereinstimmung der geschätzten Teilnutzenwerte mit den spontanen Verbalisationen der Probanden während der Bearbeitung der Aufgabe überprüft werden (z. B. Parker & Srinivasan, 1976).

Die *Konstruktvalidität* überprüft, ob die aus einem Zielkonstrukt abgeleiteten Hypothesen empirische Bestätigung finden. Es wird nicht ein einzelnes Außenkriterium festgelegt, sondern vielmehr ein Netz von Hypothesen über das zu validierende Konstrukt und dessen Relation zu anderen Variablen fortführend entwickelt und überprüft. In der Präferenzstrukturmessung wird dieser Form der Validitätsmessung teilweise damit Rechnung getragen, dass das Verhältnis der ermittelten Präferenz zu benachbarten Konstrukten wie der Kaufverhaltensabsicht oder tatsächlichem Wahlverhalten diskutiert wird. In der Regel wird die

Validität von Verfahren der Präferenzstrukturmessung jedoch im Sinne der Kriteriumsvalidität überprüft. Die Meta-Analyse fokussiert daher auf diese Art der Validitätsbestimmung.

Ein Messverfahren gilt im Sinne der *Kriteriumsvalidität* als valide, wenn das Ergebnis der Messung eines latenten Konstrukts mit der Ausprägung eines korrespondierenden, manifesten Kriteriums übereinstimmt. Bei der Kriteriumsvalidität lässt sich unterscheiden, zu welchem Zeitpunkt die zu validierenden Ergebnisse und das Validierungskriterium erhoben werden. Erfolgt die Messung des Validierungskriteriums zu einem späteren Zeitpunkt als die Messung des zu validierenden Konstrukts, so bemisst sich die Validität des Verfahrens an der Fähigkeit, die Ausprägung des Validierungskriteriums korrekt vorherzusagen. Diese Form wird als prognostische Validität (*predictive validity*) bezeichnet, im Gegensatz zur Übereinstimmungsvalidität (*concurrent validity*), bei der die Validierungskriterien mitlaufend zur Messung des zu validierenden Konstrukts erhoben werden.

3.3.2 Kategorisierung der in der Präferenzmessung verwendeten Validitätsarten

In der Praxis der Präferenzmessung wird der Begriff der prognostischen Validität nicht nur für die Vorhersage von zukünftigem Verhalten genutzt, sondern allgemeiner für die Güte von Verhaltensvorhersagen, unabhängig vom Zeitpunkt des gezeigten Verhaltens, eingesetzt. Beispielsweise sprechen Parker und Srinivasan (1976) von prognostischer Validität ihres Verfahrens, obwohl sie den Zusammenhang ihrer Präferenzdaten mit tatsächlichem Wahlverhalten untersuchten, welches zeitlich vor der Erhebung der Präferenzdaten lag. Page und Rosenbaum (1987) und Robinson (1980) reproduzierten gegenwärtige Marktbedingungen und bezeichneten dies ebenfalls als prognostische Validität. In der Folge soll daher weniger unterschieden werden, zu welchem Zeitpunkt das Validierungskriterium erhoben wurde, sondern ob es sich bei dem Kriterium um ein außerhalb oder innerhalb der Präferenzmessung erhobenes Kriterium handelt.

Als Außenkriterien werden in der Präferenzstrukturmessung beispielsweise tatsächliches Kaufverhalten oder Marktanteile zur Validierung der Präferenzdaten herangezogen. Als Innenkriterien dienen die Ergebnisse sogenannter Hold-out-Aufgaben oder alternativer Verfahren der Präferenzmessung. Hold-out-Aufgaben sind Wahl-, Rating- oder Rangordnungsaufgaben, wie sie den Antwortpersonen auch in der eigentlichen

Präferenzmessung vorgelegt werden. Die in den Hold-out-Aufgaben ermittelten Präferenzurteile werden nicht zur Bestimmung der individuellen Teilnutzenwerte der Probanden herangezogen. Analog zu den Hold-out-Aufgaben können die Ergebnisse aus einem anderen Verfahren der Präferenzmessung als Validierungskriterium herangezogen werden. Letztere Form der Validitätsbestimmung wird auch als konvergente Validität bezeichnet (Campbell & Fiske, 1959).

Wie in der Abbildung 3.2 veranschaulicht werden auf der Basis der aus einer ersten Präferenzstrukturmessung ermittelten Teilnutzenwerte eines Probanden dessen Antworten zu Aufgaben einer zweiten Präferenzstrukturmessung vorhergesagt. Die zweite Präferenzstrukturmessung kann wie beschrieben mittels Hold-out-Aufgaben oder einer alternativen Methode der Präferenzmessung erfolgen. Werden aus den aus der zweiten Messung gewonnenen Nutzenwerten die Inputdaten des ersten Messzeitpunkts zusätzlich vorhergesagt, liegt eine doppelte Kreuzvalidierung (cross-validity) vor.

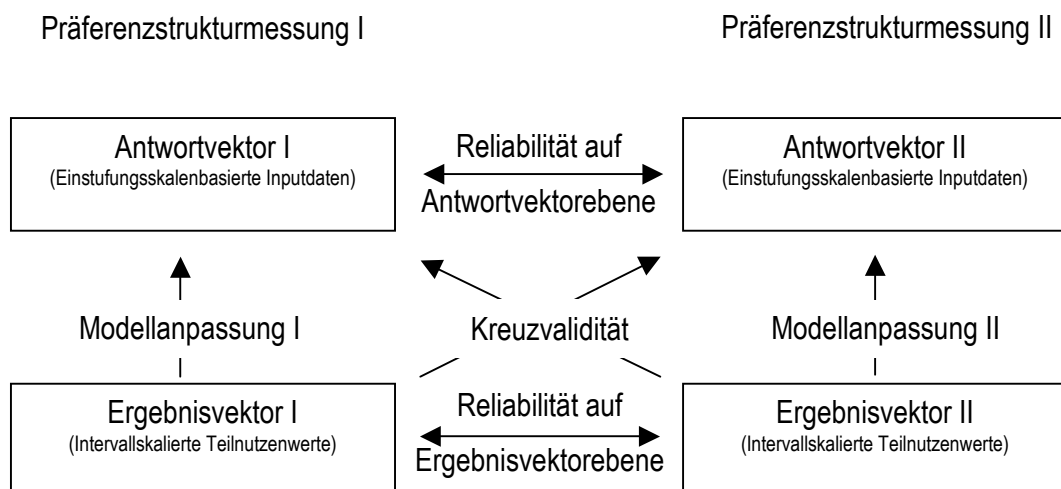


Abb.3.2: Modellanpassung, Reliabilität und Validität bei der Präferenzstrukturmessung (Pfeil: Vorhersage; Doppelpfeil: Zusammenhang)

Die Übereinstimmung der Präferenzmessung mit dem Validierungskriterium kann auf der Inputebene des Antwortvektors oder auf der Outputebene des Ergebnisvektors, also den geschätzten Präferenzstrukturmodells, erfolgen. Von den in der Meta-Analyse berücksichtigten Validitätskoeffizienten beruhen 82 % auf der Bestimmung des

Zusammenhangs zwischen vorhergesagten und beobachteten Inputdaten, 13 % auf der Korrelation von Teilnutzenwerten und 5 % auf der Korrelation von Attributwichtigkeiten.

Neben der Unterscheidung, ob es sich bei dem Validitätskriterium um ein Außen- oder ein Innenkriterium handelt, ist bei der Präferenzstrukturmessung zu berücksichtigen, ob die Validierung der Ergebnisse auf der Ebene der einzelnen Antwortpersonen oder auf aggregierter Gruppenebene erfolgt. Auf der aggregierten Ebene werden zur Validierung Markt- bzw. Wahlanteile vorhergesagt und mit den beobachteten Werten eines entsprechenden Validierungskriteriums verglichen. Auf der individuellen Ebene werden Präferenzdaten bzw. individuelle Wahlentscheidungen vorhergesagt und mit den beobachteten Werten verglichen. Die Abbildung 3.3 veranschaulicht die in dieser Arbeit verfolgte Kategorisierung der in der Präferenzstrukturmessung üblichen Formen der Validitätsbestimmung.

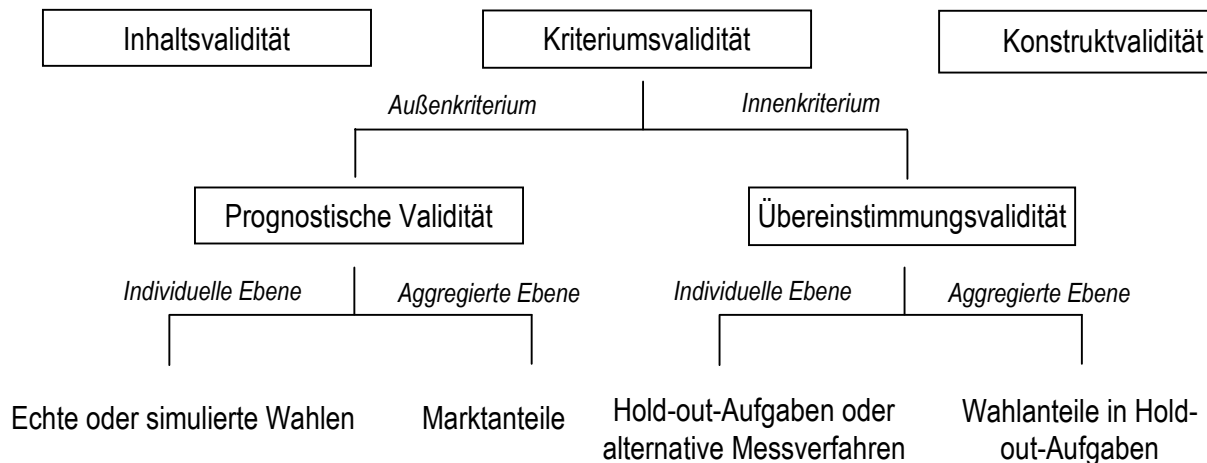


Abb.3.3: Kategorisierung der in der Präferenzstrukturmessung verwendeten Arten der Validitätsbestimmung

3.3.3 Die Maßzahlen der in der Präferenzmessung verwendeten Validitätsarten

Validitätskoeffizienten nehmen Werte zwischen 0 und 1 an, wobei die Kriteriumsvalidität maximal den Wert des geometrischen Mittels aus der Reliabilität des Messverfahrens und der Reliabilität des Kriteriums erreichen kann. Als Maßzahlen der Übereinstimmung zwischen der Vorhersage und der Beobachtung eines Kriteriums werden in der Präferenzstrukturmessung die Kennziffern des Produkt-Moment-Korrelationskoeffizienten, des Rangkorrelationskoeffizienten Spearman's rho und des Zusammenhangsmaßes Kendall's tau verwendet.

Die Koeffizienten werden jeweils für jeden Probanden separat bestimmt und dann über alle Probanden gemittelt. Neben diesen Zusammenhangsmaßen ist die Angabe der sogenannten Hitrate üblich. Die Hitrate gibt an, für welchen Anteil der Vorhersagen, über alle Probanden und Wahlsets zusammengefasst, der erste Präferenzrang in einem Set von mehreren Alternativen korrekt vorhergesagt werden konnte.

Auf der aggregierten Ebene werden beobachtete und vorhergesagte Markt- bzw. Wahlanteile miteinander verglichen. Als übliche Maßzahlen der Güte der Vorhersage dienen die mittlere absolute Abweichung (mean absolute error, MAE) oder die Wurzel der quadrierten mittleren Abweichung (root mean squared error, RMSE) zwischen vorhergesagten und beobachteten Markt- bzw. Wahlanteilen.

$$MAE = \frac{\sum_i |N_i - \hat{N}_i|}{\sum_i N_i} \quad RMSE = \sqrt{\frac{\sum_i (N_i - \hat{N}_i)^2}{\sum_i N_i}} \quad (13)$$

mit: N_i : Die Anzahl der beobachteten Wahlen von Objekt i

\hat{N}_i : Die Anzahl der vorhergesagten Wahlen von Objekt i

Der MAE hat den Vorteil, dass seine Maßzahl direkt die durchschnittliche Abweichung der Anteilsvorhersagen in absoluten Prozentpunkten angibt und demnach einfach zu interpretieren ist. Der MAE behandelt alle Abweichungen gleich, während der RMSE große Abweichungen durch die Quadrierung stärker gewichtet.

Neben dem MAE und dem RMSE werden vereinzelt auch noch andere Maßzahlen für die Güte der Wahlanteilsvorhersage eingesetzt, wie beispielsweise die relative Reduktion des mittleren quadrierten Fehlers (proportional reduction in mean squared error of prediction, PRMSEP; Elrod et al., 1992), Chi-Quadrat-Werte (Moore, Gray-Lee & Louviere, 1998) oder eine Produkt-Moment-Korrelation zwischen den Vektoren aus realen und vorhergesagten Marktanteilen (Bucklin und Srinivasan, 1991). Da diese Kennzahlen jedoch nur vereinzelt anzutreffen sind, fokussiert sich die Meta-Analyse auf die beiden üblicheren Maßzahlen des MAE und des RMSE.

Die in der Abbildung 3.3 dargestellte Kategorisierung der Arten der Validitätsbestimmung berücksichtigt nicht die in der Präferenzstrukturmessung häufiger anzutreffenden Begriffe der internen und externen Validität. Mit diesen Begriffen soll darauf verwiesen werden, ob das Validierungskriterium innerhalb oder außerhalb der Präferenzmessung erhoben wird, also gemäß der Kategorisierung aus Abbildung 3.3 ein Innen- oder Außenkriterium darstellt. Im eingeführten Begriffsverständnis (z. B. Campbell, 1976; Cook & Campbell, 1979) bezeichnen die interne und die externe Validität die Güte von empirischen Untersuchungsdesigns und nicht die Qualität von Messinstrumenten. Eine empirische Untersuchung gilt als intern valide, wenn ihre Ergebnisse eindeutig interpretierbar sind und nur eine plausible Erklärung für die Messergebnisse vorliegt. Die externe Validität bezeichnet den Grad der möglichen Generalisierbarkeit der Messergebnisse über das enge Untersuchungssetting und die betrachtete Stichprobe hinaus. In dieser Arbeit werden die Begriffe der internen und externen Validität nicht verwendet.

Die Bestimmung der Validität der Präferenzstrukturmessung mittels Hold-out-Aufgabe wird von Louviere (1988) als eine Form der Test-Retest-Reliabilitätsbestimmung und nicht der Validitätsmessung kritisiert. Die Kritik betont damit, dass es sich bei der Hold-out-Aufgabe um ein Innenkriterium handelt, das eine hohe Ähnlichkeit zu den Aufgaben der eigentlichen Präferenzmessung aufweist.

Wittink (2000) stellt demgegenüber heraus, dass die Hold-out-Validierungen zwar keine Rückschlüsse über die absolute Höhe der Vorhersagegüte der Präferenzmessverfahren für echte Wahlen zulässt, die relative Vorhersagestärke der Verfahren, wie sie sich in der Hold-out-Validierung zeigt, aber in einer realen Umgebung Bestand hat. Demnach müssen

Messverfahren, die bei einer Hold-out-Validierung besser als andere abschneiden, auch im Sinne der Vorhersage realer Entscheidungen überlegen sein. Die Untersuchung von Dahan et al. (2001) bestätigt diese Hypothese empirisch. Sie fanden eine hohe Übereinstimmung zwischen den Validierungsergebnissen mittels Hold-out-Aufgaben und echten Wahlen: Die Rangordnung der verglichenen Verfahren gemäß ihrer Kriteriumsvalidität stellte sich mit dem Innen- und dem Außenkriterium identisch dar.

Um dem Vorwurf der Reliabilitätsmessung im Zusammenhang mit Hold-out-Validierungen zu begegnen, empfiehlt Wittink (2000), die Ähnlichkeit der Beurteilungsaufgaben im Kalibrierungs- und im Validierungsset zu minimieren. Beispielsweise schlägt er vor, die Reihenfolge der Attribute in den Hold-out-Aufgaben zu verändern. Andere Autoren nutzten als Hold-out-Aufgabe Beschreibungen von realen Produkten (z. B. Nickerson, McClelland & Petersen, 1990; Lenk et al, 1996) oder reale Produkte bzw. Prototypen (z. B. Sattler, 1994; Tscheulin & Blaimont, 1993; Hensel-Börner & Sattler, 2000; Albrecht, 2000).

Veränderungen können auch in der abhängigen Variablen vorgenommen werden. Beispielsweise kann die Hold-out-Aufgabe in der Auswahl eines präferierten Objekts aus einem Set von Alternativen bestehen, während die Kalibrierungsdaten mittels Rangordnungen oder Einstufungsskalen erhoben werden. Wittink (2000) empfiehlt grundsätzlich, die Hold-outs als Wahlaufgabe zu gestalten, da die Auswahl eines Objekts reales Kaufverhalten realistischer abbildet als die Beurteilung der Stimuli mittels Rangordnung oder Einstufungsskala.

Durch die Ausgestaltung der Hold-outs als Wahlaufgabe kann zudem sichergestellt werden, dass kein Verfahren der Präferenzmessung in empirischen Vergleichsstudien bevorzugt wird. So vermuteten beispielsweise Huber et al. (1993), dass traditionelle Vollprofil-Verfahren gegenüber einer mit Teilprofilen arbeitenden adaptiven Conjoint-Analyse (ACA) in Validierungsstudien Vorteile hätten, da die Kalibrierungsaufgaben der traditionellen Conjoint-Analyse den entsprechenden Validierungsaufgaben ähnlicher seien als die ACA-Paarvergleiche. Einfache Urteilsheuristiken würden von der Kalibrierungs- auf die Validierungsaufgabe übertragen und eine Verzerrung der Validierungsergebnisse bewirken. Huber et al. führten aus diesem Grund verschiedene Variationen bei den Hold-out-Aufgaben ein. Unter diesen kontrollierten Bedingungen fanden Sie eine Überlegenheit von ACA bzw.

alternativen hybriden Modellen gegenüber der traditionellen Conjoint Analyse und einer direkten Nutzenmessung. Diese Tendenz findet sich in der durchgeführten Meta-Analyse bestätigt.

Neben der Frage, ob es sich bei der Hold-out-Validierung um eine Form der Reliabilitäts- oder der Validitätsbestimmung handelt, sind einige methodische Schwierigkeiten im Zusammenhang mit der Verwendung von Hold-out-Aufgaben zu diskutieren. Wie bereits dargestellt, hängt die Höhe der Kriteriumsvalidität nicht nur von der Reliabilität des Messverfahrens, sondern auch von der Reliabilität der Kriteriumsmessung ab. Holling, Melles und Reiners (1999) stellen heraus, dass die Reliabilität des Hold-out-Kriteriums vom Schwierigkeitsgrad der Hold-out-Aufgabe abhängt, der sich wiederum aus der Anzahl der im Hold-out-Set enthaltenen Alternativen, der Profilstärke der Stimuli und der Ähnlichkeit der Gesamtnutzenwerte der zu bewertenden Objekte ergibt.

Zahlreiche empirische Studien zeigen, dass ein bedeutender Teil der Probanden auf wiederkehrende Präferenzaufgaben nicht identisch antwortet. Beispielsweise stellten Huber et al. (1993) fest, dass nur 31 % von knapp 400 Antwortpersonen bei der unmittelbaren Wiederholung von acht Paarvergleichsaufgaben eine identische Auswahl trafen. Bryan et al. (2000) fanden bei der unmittelbaren Wiederholung von vier Paarvergleichen nur 57 % von 363 Antwortpersonen, die gleichbleibend antworteten. Van der Lens et al. (1992) zeigten, dass die Hitraten einer traditionellen CA und einer adaptiven CA von 57 % auf bis zu 82 % anstiegen, wenn nur die reliablen Antwortpersonen berücksichtigt wurden.

Um einen Anhaltspunkt für die durch die Einbußen in der Reliabilität gesetzte Obergrenze für Hitraten zu erhalten, empfehlen Orme, Alpert und Christensen (1997) die Wiederholung von einzelnen oder allen Hold-out-Aufgaben. Die Korrelation zwischen den Beurteilungen der identischen Hold-outs kann als Test-Retest-Reliabilität interpretiert werden. Orme et al. (1997) schlagen vor, die maximal zu erwartende Hitrate durch

$$Hitrate_{\max.} = \frac{1 + \sqrt{(2p - 1)}}{2} \quad (14)$$

zu berechnen, wenn mit p der Anteil der Probanden bekannt ist, die bei zwei wiederkehrenden Hold-out-Wahlaufgaben die identischen Entscheidungen treffen. Wenn

beispielsweise 85 % der Probanden gleichbleibend antworten, ist mit einer Hitrate von maximal 91,8 % zu rechnen.

Im Vergleich zur Bestimmung der Validität mittels eines Innenkriteriums wird der Validitätsbestimmung anhand eines externen Kriteriums in der Regel die größere Aussagekraft beigemessen (z. B. Louviere, 1988; Green & Srinivasan, 1990). Allerdings muss betont werden, dass die Messung von Präferenzen einerseits und die Beobachtung von tatsächlichem Wahlverhalten bzw. realen Marktanteilen andererseits zwei unterschiedliche Ebenen des Entscheidungsprozesses betrachten.

Wie in Kapitel 2.2 dargestellt, ist das theoretische Konstrukt der Präferenz abzugrenzen von der Verhaltensabsicht und dem direkt beobachtbaren Entscheidungsverhalten. Die Präferenz kann lediglich ein guter Prädiktor für tatsächliches Verhalten sein. Bezogen auf einen Marketingkontext hängt der Kauf eines Produktes nicht nur von der individuellen Präferenz, sondern auch von der subjektiv wahrgenommenen Präsenz (awareness) und der objektiven Verfügbarkeit (availability) des Produkts ab. Zudem können soziale Einflüsse, Mund-zu-Mund-Effekte und besondere Anstrengungen am Point-of-Sale das reale Kaufverhalten beeinflussen. Es kann daher nicht verwundern, dass bei der Validitätsbestimmung mittels Innenkriterium höhere Validitätskoeffizienten resultieren als bei Validierung anhand eines Außenkriteriums.

Zusammenfassend dargestellt werden in dieser Arbeit die folgenden Arten und Maßzahlen der Validitätsbestimmung von Präferenzmessverfahren berücksichtigt:

	<i>Individuelle Ebene</i>	<i>Aggregierte Ebene</i>
<i>Innenkriterium</i>	Hitraten und Übereinstimmungsmaße (Spearman's rho, Kendall's tau, Pearson's r) zwischen prognostizierten und beobachteten Präferenzdaten in Hold-out-Aufgaben bzw. konvergenten Messverfahren	MAE und RMSE zwischen prognostizierten und beobachteten Wahlanteilen in Hold-out-Aufgaben
<i>Außenkriterium</i>	Hitraten und Übereinstimmungsmaße (Spearman's rho, Kendall's tau, Pearson's r) zwischen prognostizierten und beobachteten realen, individuellen Wahlentscheidungen	MAE und RMSE zwischen prognostizierten und beobachteten realen Marktanteilen

Tab. 3.2: Tabellarische Übersicht über die in der Meta-Analyse verwendeten Arten und Maßzahlen der Kriteriumsvalidität der Präferenzmessung

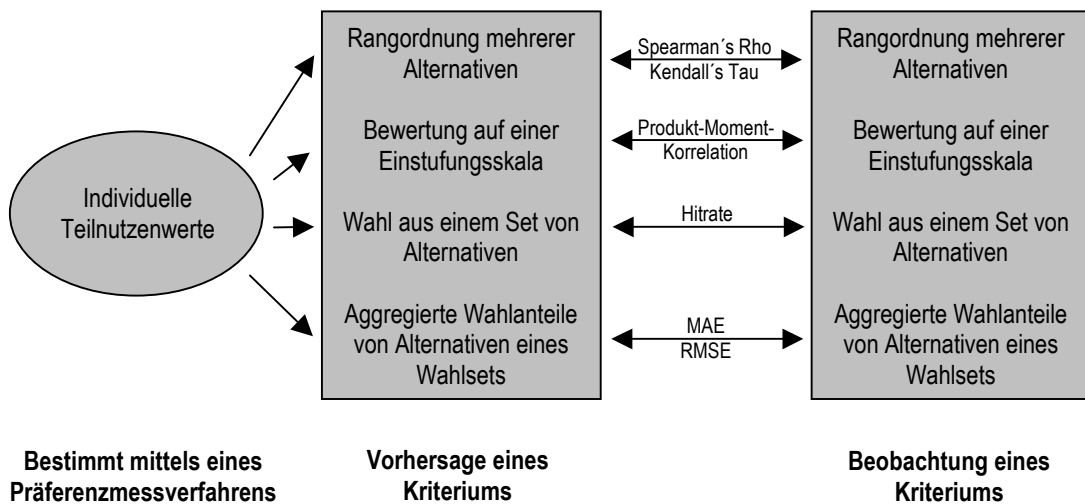


Abb.3.4: Grafische Übersicht über die in der Meta-Analyse verwendeten Arten und Maßzahlen der Kriteriumsvalidität der Präferenzmessung

Vorhergesagtes Kriterium	Effektgröße	Anzahl der berücksichtigten Publikationen	Anzahl der berücksichtigten Effektgrößen	Prototypische Publikation ^a
Individuelle Bewertung von Vollprofilen mittels einer Einstufungsskala	Produkt-Moment-Korrelation (Pearson's r) zwischen vorhergesagten und beobachteten Einstufungen, gemittelt über alle Probanden	34	178	Akaah (1991)
Individuelle Rangordnung von Vollprofilen	Rang-Korrelation (Spearman's rho oder Kendall's tau) zwischen vorhergesagten und beobachteten Rangordnungen, gemittelt über alle Probanden	25	97	Allenby, Arora & Ginter (1995)
Individuelle Teilnutzenwerte, die mit einer anderen Präferenzmessung ermittelt wurden	Produkt-Moment-Korrelation zwischen vorhergesagten und beobachteten Teilnutzenwerten, gemittelt über alle Probanden	12	112	Agarwal & Green (1991)
Individuelle Attributwichtigkeiten, die mit einer anderen Präferenzmessung ermittelt wurden	Produkt-Moment-Korrelation zwischen vorhergesagten und beobachteten Attributwichtigkeiten, gemittelt über alle Probanden	9	47	Akaah (1991)
Erster Rang bei der Rangordnung von Alternativen bzw. erste Wahl bei der Auswahl von Alternativen	Anteil der korrekt vorhergesagten ersten Wahlen an allen Wahlen, betrachtet über alle Probanden und Wahlaufgaben	74	469	Agarwal & Green (1991)
Aggregierte Wahlanteile von Alternativen aus einem oder mehreren Wahlsets	Mittlerer, absoluter Fehler (MAE) zwischen den vorhergesagten und den beobachteten Wahlanteilen, betrachtet über alle Probanden und Wahlaufgaben	30	229	Allenby, Arora & Ginter (1995) Moore, Gray-Lee & Louviere (1998)
Marktanteile / Verkaufsanteile realer Produkte	Mittlerer, absoluter Fehler (MAE) zwischen den auf aggregierter Ebene vorhergesagten und den beobachteten Marktanteilen	5	21	Bucklin & Srinivasan (1991) Conrad (1997) Orme & Heft (1999)
Aggregierte Wahlanteile von Alternativen aus einem oder mehreren Wahlsets	Wurzel des mittleren quadrierten Fehlers (RMSE) zwischen den vorhergesagten und den beobachteten Wahlanteilen, betrachtet über alle Probanden und Wahlaufgaben	4	51	Krieger, Green & Umesh (1998) Moore, Gray-Lee & Louviere (1998)

^a Eine detaillierte Aufstellung der jeweils berücksichtigten Studien findet sich im Anhang.

Tab. 3.3: Tabellarische Übersicht über die Häufigkeiten der in der Meta-Analyse verwendeten Effektgrößen der Kriteriumsvalidität der Präferenzmessung

4. Die Meta-Analyse als systematisches Verfahren zur Integration von Forschungsergebnissen

4.1 Die bisherige, zusammenfassende Forschung

Zur Verbesserung der Übersichtlichkeit der Forschung zur Messgüte der Präferenzmessverfahren haben Green und Srinivasan (1990) sowie Sattler und Hensel-Börner (2001) Zusammenfassungen publiziert. Die Autoren bedienten sich der Methode des literarischen bzw. narrativen Reviews, um den aktuellen Forschungsstand zusammenzutragen und verdichtend aufzubereiten.

Green und Srinivasan (1990) erwähnten insgesamt fünfzehn Studien mit Ergebnissen zur Validität von conjoint-analytischen Verfahren. Sie kommen zu dem Schluss, dass die empirischen Erkenntnisse auf die Validität der Conjoint Analyse als Vorhersagetechnik von präferenzgeleiteten Entscheidungen hinweisen. Sie fordern zeitgleich mehr Studien ein, die sich mit der Validität der Präferenzmessverfahren beschäftigen, insbesondere mit der Vorhersage von realen Kaufentscheidungen und Marktanteilen.

Sattler und Hensel-Börner (2001) haben vierzehn empirische Studien ausgewertet, welche die prognostische Validität von direkten Verfahren der Nutzenmessung mit conjoint-basierten Verfahren verglichen. Dazu zählten sie die Studien aus, in denen jeweils eine der beiden Verfahrensgruppen signifikant besser abschnitt. Im Ergebnis sahen sie keine Überlegenheit einer der Verfahrensgruppen. Ganz im Gegenteil schlussfolgerten sie, dass das verwendete Verfahren keinen eindeutigen, unidirektionalen Einfluss auf die prognostische Validität aufweise.

4.2 Die Zielsetzung und die Vorteile einer Meta-Analyse

Gegenüber narrativen Reviews wurde wiederholt die Kritik vorgebracht, dass sie nicht objektiv genug seien, da die Auswahl und die Gewichtung der zu berücksichtigten Studien subjektiv erfolge und dem Forscher zu viele Freiheiten lasse. Beaman (1991) betonte in einer Vergleichsstudie, dass Meta-Analysen gegenüber traditionellen Reviews objektiver seien, da die Integration der Forschungsergebnisse nicht auf einer rein sprachlichen, eher qualitativen

Ebene vorgenommen werde, sondern eine systematische Zusammenführung statistischer Indikatoren erfolge.

Das Ziel von Meta-Analysen ist die systematische Zusammenfassung von inhaltlich verwandten Studien, um die Stärke eines statistischen Effekts über die Einzelstudien hinweg gemeinsam abzuschätzen. Es wird dabei angestrebt, die Einzelergebnisse in ein einheitliches Effektmaß zu überführen, welches dann aufgrund der gepoolten Stichprobe eine breitere Aussagekraft aufweist.

Der Begriff der Meta-Analyse wurde von Glass (1976) eingeführt, der damit die Analyse von Analysen bezeichnet. Es haben sich unterschiedliche, methodische Ansätze zur Vorgehensweise bei einer Meta-Analyse herausgebildet. Die Ansätze von Schmidt und Hunter (1977, bzw. Hunter & Schmidt, 1990), Rosenthal und Rubin (1978) und Hedges und Olkin (1985) werden in diesem Kapitel aufgegriffen.

In Abgrenzung zu der Methode des narrativen Reviews bietet eine Meta-Analyse Vorteile:

Die Meta-Analyse gibt eine systematische Prozedur für die Zusammenfassung von Forschungsergebnissen vor. Sie ist damit nachvollziehbar, überprüfbar und replizierbar für Dritte.

Die wesentlichen Ergebnisse der Einzelstudien, insbesondere die statistischen Effektgrößen, werden in Meta-Analysen differenzierter und umfassender aufgenommen und ausgewertet als bei traditionellen Methoden der Zusammenfassung von Forschungsergebnissen. Während bei narrativen Reviews in der Regel eine qualitative Zusammenfassung der Ergebnisse oder ein reines Auszählen der statistisch signifikanten Resultate erfolgt, werden bei Meta-Analysen die Größe und die Richtung der Effektgrößen dokumentiert. Dies ist gerade daher von Bedeutung, da die Signifikanzen von Ergebnissen in Einzelstudien alleine durch Zufall (bei einer Vielzahl von abhängigen Variablen) oder in Abhängigkeit der Stichprobengröße und des damit verbundenen Stichprobenfehlers zustande gekommen sein können.

Die Integration der Forschungsergebnisse erfolgt bei der Meta-Analyse auf einer statistischen Basis. Die Einzelstudien werden dabei als Datenpunkte aufgefasst, die einer neuen, zusammenfassenden Analyse unterzogen werden. Daraus resultieren zwei Vorteile. Zum Einen weisen die Ergebnisse auf der Basis einer gepoolten Stichprobe eine größere, statistische Power auf, zum Anderen lassen sich Einflussgrößen auf die Höhe des

Effektmaes mit groerer Wahrscheinlichkeit identifizieren. Beispielsweise knnen Studiencharakteristika oder bestimmte Eigenschaften der unabhngigen Variable mit kodiert werden, um sie als Moderatorvariablen zu kontrollieren und einen mglicherweise statistisch signifikanten Einfluss auf die Hhe der abhngigen Variablen nachzuweisen.

Schlielich ermglichen die systematische Vorgehensweisen und die entsprechend mgliche Computeruntersttzung die Handhabung sehr groer Datenmengen. Whrend bei narrativen Reviews die Verarbeitungskapazitt des Forschers natrliche Grenzen setzt, sind Meta-Analysen mit mehreren hundert Studien immer noch gut zu handhaben.

4.3 Das Vorgehen bei einer Meta-Analyse

Der Prozess der Durchfhrung einer Meta-Analyse wird in sechs Schritten dargestellt.

Spezifikation der Forschungsfrage und Definition der Charakteristika der relevanten Forschungsliteratur. Die zentralen Kategorien der unabhngigen Variablen und die damit verbundenen Messgroen der abhngigen Variablen mssen eindeutig festgelegt werden. Die Problem- und Variablenspezifikation sollte eine eindeutige Richtschnur fr die Auswahl und den Einbezug von Studien vorgeben, um dem Vorwurf der Integration nicht zusammenfhrbarer Studien zu begegnen.

Literatursuche. Hier kommt es wesentlich darauf an, neben den ber elektronische Datenbanken rasch zugnglichen Publikationen auch die „graue“ Literatur zu identifizieren und aufzunehmen. Der Publikationsbias bezeichnet das Problem, dass Untersuchungen, welche zu signifikanten Ergebnissen gekommen sind, wahrscheinlicher publiziert werden. Konzentrieren sich Meta-Analysen ausschlielich auf die einschlgigen Verffentlichungen in dem Forschungsgebiet, so kann eine systematische Verzerrung der Ergebnisse auftreten. Das Ziel des Forschers muss es daher sein, mglichst auch die unpublizierte, sogenannte „graue“ Literatur zu identifizieren, wie z. B. interne Working und Research Papers, Conference Papers, Diplom- und Doktorarbeiten, etc.

Definition von cut-off-Kriterien, nach denen sich entscheidet, ob eine Studie in die Meta-Analyse aufgenommen werden soll. Hier gilt es, mglichst plausible Kriterien zu definieren und diese offenkundig zu machen, um subjektive Einflsse des Forschers zu

minimieren. Strenge Kriterien im Hinblick auf die qualitativen Ansprüche an die Einzelstudien schützen vor dem Vorwurf, dass Meta-Analysen qualitativ nicht besser sein können als die verwendeten Einzelstudien. Gemäß des Vorwurfs „garbage in – garbage out“ führt der Einbezug von methodisch zweifelhaften Studien zu einem falschen Gesamtergebnis der Meta-Analyse.

Entwicklung eines Kodierschemas und Kodierung der relevanten Informationen für jede Studie. Das Kodierschema gibt vor, welche Informationen aus den Studien extrahiert und dokumentiert werden sollen. Neben deskriptiven Informationen über die Studie wie Stichprobengrößen, Probandencharakteristika, Publikationsjahr etc. müssen in erster Linie Informationen über die abhängigen und die unabhängigen Variablen kodiert werden. Die abhängigen Variablen sind die statistischen Effektgrößen, die das Ausmaß eines Effekts der unabhängigen Variablen auf eine abhängige Variable bzw. die Stärke des Zusammenhangs zwischen zwei Variablen angeben.

Es werden zwei Familien von Effektgrößen unterschieden, die d-Indizes und die r-Indizes. Das d-Maß gibt den Mittelwertsunterschied zwischen zwei Gruppen, relativiert an der gepoolten Standardabweichung beider Gruppen (Hedges' g) oder der Streuung der Kontrollgruppe (Glass' Δ) an. Das d-Maß drückt somit den Abstand zwischen den untersuchten Gruppen in Einheiten der Standardabweichung aus.

Die Familie der r-Indizes enthält alle Formen von Zusammenhangsmaßen wie z. B. Pearson's r , Punkt-biserial- und Rangkorrelationen, Phi-Koeffizient etc. Diese Gruppe von Effektgrößen ist für eine Meta-Analyse von Validitätskennziffern relevant und wird daher in der Folge dieses Kapitels weiterverfolgt.

Die unabhängigen Variablen repräsentieren alle Informationen, die auf einen systematischen Einfluss auf die Effektgrößen hin analysiert werden sollen. Das sind in der Regel die unterschiedlichen Treatments wie beispielsweise die eingesetzten Lehr-, Behandlungs- oder Therapieformen oder, wie in der vorliegenden Arbeit, die unterschiedlichen Verfahren der Präferenzmessung. Darüber hinaus können aber auch weitere Faktoren gemäß theoretischer Vorannahmen auf ihren Einfluss überprüft werden, wie z. B. Probandencharakteristika, die Art des Untersuchungsdesigns etc.

Integration der Forschungsergebnisse durch eine statistische Analyse. In der Regel wird dazu auf der Basis der gepoolten Gesamtstichprobe eine Effektgröße durch die Mittelung der mit

den jeweiligen Stichprobengrößen gewichteten Einzeleffektgrößen berechnet. Zu der neuen Effektgröße können dann die Standardabweichung, die Varianz und ein Konfidenzintervall bestimmt und Signifikanztests durchgeführt werden. Für die statistische Integration der Forschungsergebnisse haben sich unterschiedliche Vorgehensweisen etabliert, von denen im folgenden Unterkapitel die Ansätze von Hedges und Olkin (1985) und von Hunter und Schmidt (1990) näher beschrieben werden.

Homogenitätsanalyse und Moderatorenidentifikation. Nach der Zusammenfassung der Einzelstudien muss überprüft werden, ob die eingegangenen Untersuchungen einer homogenen Grundgesamtheit entstammen. Die Annahme, dass alle berücksichtigten Studien aus einem homogenen Universum von möglichen Studium stammen, denen ein gemeinsamer Populationsparameter zugrunde liegt, kann über einen Q-Test überprüft werden, der im folgenden Unterkapitel näher beschrieben wird. Ein signifikantes Ergebnis des Q-Tests deutet auf eine heterogene Studienzusammensetzung hin. Praktisch bedeutet ein signifikantes Ergebnis des Q-Tests, dass die Suche nach Moderatorvariablen angezeigt ist. Dazu wird die Grundgesamtheit der Effektgrößen gemäß theoretischer Vorannahmen in Subgruppen untergliedert. Für jede Subgruppe können dann separate Meta-Analysen durchgeführt und erneute Homogenitätsüberprüfungen vorgenommen werden.

4.4 Die statistischen Vorgehensweisen bei einer Meta-Analyse

4.4.1 Korrelationskoeffizienten als Effektgrößen in der Meta-Analyse

Für die Integration der Forschungsergebnisse zur Messgüte der Präferenzmessverfahren wird auf die Familie der r -Indizes zurückgegriffen. Diese Gruppe von Effektgrößen ist für eine Meta-Analyse von Validitäts- und Reliabilitätskoeffizienten relevant und wird daher in der Folge dieses Kapitels weiterverfolgt. Der Stichprobenkorrelationskoeffizient r basiert auf n Datenpaaren (x_o, y_o) , $o = 1, \dots, n$. Im Falle der Kriteriumsvalidität von Präferenzmessverfahren ergibt sich ein Datenpaar beispielsweise durch die vorhergesagte und die beobachtete Ausprägung eines anhand einer Einstufungsskala abgefragten Präferenzurteils.

Der Stichprobenkorrelationskoeffizient r wird berechnet durch

$$r = \frac{\sum_{o=1}^n (x_o - \bar{x})(y_o - \bar{y})}{\sqrt{\sum_{o=1}^n (x_o - \bar{x})^2} \sqrt{\sum_{o=1}^n (y_o - \bar{y})^2}} \quad (15)$$

Für die Meta-Analyse lässt sich ein in einer Einzelstudie ermittelter Korrelationskoeffizient r als Realisierung einer Zufallsvariablen R auffassen. Das korrespondierende Paar von Zufallsvariablen zu (x_o, y_o) ist (X, Y) . In diesem Sinne ist ein in einer Studie ermittelter Korrelationskoeffizient als Indikator für die Stärke der Validität des Verfahrens ein stichprobenbasierter Schätzer für einen universalen Populationsparameter. Der korrespondierende Populationskorrelationskoeffizient ρ ist gegeben durch

$$\rho = \frac{E((X - E[X])(Y - E[Y]))}{\sigma_X \sigma_Y} \quad (16)$$

Schulze (2004) weist darauf hin, dass der Stichprobenkorrelationskoeffizient r ein negativ verzerrter Schätzer des Populationsparameters ist, insbesondere bei kleinen Stichprobengrößen und einem positiven Populationsparameter um $\rho = .58$. Zum Ausgleich dieses negativen Bias von r findet sich bei Olkin und Pratt (1985) eine Korrekturformel. Der unverzerrte Schätzer r^* für den Populationsparameter ρ , der UMVU (unique minimum variance unbiased estimator), ergibt sich in Annäherung wie folgt:

$$r^* = r \left(1 + \frac{1 - r^2}{2(n - 1 - 3)} \right) \quad (17)$$

Es wird kontrovers diskutiert, ob eine Fisher-z-Transformation des Stichprobenkorrelationskoeffizienten r in der Meta-Analyse ratsam ist (z. B. Hunter & Schmidt, 1990). Die Fisher-z-Transformation zu r ergibt sich durch

$$z = \frac{1}{2} \ln \frac{1+r}{1-r} \quad (18)$$

beziehungsweise in der inversen Richtung durch

$$r = \frac{e^{2z} - 1}{e^{2z} + 1} \quad (19)$$

Ein Vorteil der Fisher-z-Transformation wird darin gesehen, dass die Verteilung von z als annähernd normalverteilt gelten kann, wohingegen sich die Verteilung von r erst mit großen Stichproben ($n > 500$) einer Normalverteilung annähert. Die für die Berechnung von Konfidenzintervallen und die Durchführung von Signifikanztests bedeutende Varianz der jeweiligen Zufallsvariablen lässt sich für Z annähern durch

$$\sigma_Z^2 = \frac{1}{n-3} \quad . \quad (20)$$

Die Varianz von R wird angenähert durch

$$\sigma_R^2 = \frac{(1-\rho^2)^2}{n-1} \quad . \quad (21)$$

Im Gegensatz zur Varianz von R ist die Varianz von Z nicht von dem wahren Populationsparameter, sondern nur von der Stichprobengröße n abhängig und damit einfacher zu bestimmen.

Allerdings ist z ebenso wie r kein unverzerrter Schätzer für den zugrunde liegenden Populationsparameter. Für positive Ausprägungen von ρ und kleine Stichprobengrößen weist Fisher- z einen positiven Bias auf, der nach einer Empfehlung von Schulze (2004) wie folgt korrigiert werden sollte:

$$z^* = z - \frac{\tanh(z)}{2n - \frac{9}{2}} \quad . \quad (22)$$

4.4.2 Der Ansatz nach Hedges und Olkin

In dem von Hedges und Olkin (1985) vorgeschlagenen Vorgehen zur statistischen Zusammenführung der Effektgrößen werden die in den jeweiligen Studien angegebenen Korrelationskoeffizienten zunächst gemäß Formel (18) in Fisher- z -Werte transformiert. Die Varianzen der transformierten Effektgrößen z_i sind gegeben durch

$$\hat{\sigma}_{z_i}^2 = \frac{1}{n_i - 3} \quad . \quad (23)$$

Die mittlere Fisher-z-transformierte Effektgröße wird berechnet durch

$$\bar{z} = \frac{\sum_{i=1}^k (n_i - 3)z_i}{\sum_{i=1}^k (n_i - 3)} \quad (24)$$

Als Gewichte werden die reziproken Varianzen der Einzeleffektgrößen verwendet. Diese optimalen Gewichte minimieren die Varianz der gepoolten Effektgröße, da den genauesten Effektgrößen das größte Gewicht eingeräumt wird. Für die Bestimmung der Konfidenzintervalle wird der Standardfehler der mittleren Effektgröße benötigt, der sich bestimmt durch

$$\hat{\sigma}_{\bar{z}} = \frac{1}{\sqrt{N - 3k}} \quad , \text{ mit } N = \sum_{i=1}^k n_i \text{ und } k = \text{Anzahl der Effektgrößen.} \quad (25)$$

Die Grenzen eines 95%-Konfidenzintervalls ergeben sich dann wie folgt:

$$\begin{aligned} z_{\text{untereGrenze}} &= \bar{z} - 1,96\hat{\sigma}_{\bar{z}} \\ z_{\text{obereGrenze}} &= \bar{z} + 1,96\hat{\sigma}_{\bar{z}} \end{aligned} \quad (26)$$

Die mit einem positiven Bias belasteten Fisher-z-Werte können gemäß Formel (22) korrigiert werden. Die mittlere, Fisher-z-transformierte Effektgröße und die oberen und unteren Grenzen der Konfidenzintervalle werden schließlich gemäß Formel (19) in die r-Metrik zurück transformiert.

Die Annahme, dass alle integrierten Studien aus einem homogenen Universum von möglichen Studium stammen, denen ein gemeinsamer Populationsparameter zugrunde liegt, kann über einen Q-Test überprüft werden. Der über die Formel

$$Q = \sum_{i=1}^k (n_i - 3)(z_i - \bar{z})^2 \quad (27)$$

ermittelte, empirische Q-Wert folgt einer χ^2 -Verteilung mit k-1 Freiheitsgraden. Durch den Vergleich mit dem entsprechenden kritischen Wert der χ^2 -Verteilung wird getestet, ob die Annahme der Existenz eines gemeinsamen Populationsparameters gerechtfertigt ist. Ein signifikantes Ergebnis deutet auf eine heterogene Studienzusammensetzung hin. Es stellt sich dann die Frage, in wieweit die mittlere Effektgröße noch Aussagekraft besitzt. Sie stellt

zwar nach wie vor den Mittelwert der zusammengeführten Effektgrößen dar, eine Generalisierung des Ergebnisses auf das Universum aller möglichen Studien ist jedoch nicht mehr angezeigt.

Im Kontext der Präferenzmessverfahren ist es angesichts der zahlreichen Variationsmöglichkeiten der Präferenzhebung zu erwarten, dass ohne Berücksichtigung von Moderatoren nur ein geringer Anteil der beobachteten Varianz aufgeklärt werden kann. Anhand von theoretischen Vorannahmen lassen sich im Falle eines signifikanten Q-Tests Untergruppen bilden, für die dann separate Meta-Analysen berechnet und Homogenitätstests durchgeführt werden. Im Rahmen der Präferenzmessung bietet es sich an, die Art des verwendeten Präferenzmessverfahrens als Moderatorvariable zu überprüfen. Angesichts der Tatsache, dass unterschiedliche Arten der Reliabilitäts- bzw. Validitätsbestimmung in die Gesamt-Meta-Analyse eingehen (vgl. Kapitel 3), sollte zudem überprüft werden, ob die Art der Messgütebestimmung als Moderatorvariable fungiert. Die Prinzipien der Homogenitätsanalyse mittels Q-Test finden sich auch in dem Ansatz von Hunter und Schmidt (1990) wieder. Der Ansatz von Hunter und Schmidt wird in der vorliegenden Arbeit verwendet und daher im folgenden Unterkapitel näher beschrieben.

4.4.3 Der Ansatz nach Hunter und Schmidt

Der Ansatz von Rosenthal und Rubin (1978, 1982) ist in den Grundzügen vergleichbar mit dem von Hedges und Olkin vorgeschlagenen Vorgehen und wird hier nicht gesondert verfolgt. Demgegenüber weist der Ansatz von Hunter und Schmidt (1990) einige alternative Empfehlungen auf. Hunter und Schmidt haben sich insbesondere mit der Fragestellung der Generalisierbarkeit von Validitätskoeffizienten befasst. Die Generalisierbarkeit von Reliabilitäts- und Validitätskoeffizienten von Präferenzmessverfahren ist die Fragestellung der vorliegenden Arbeit. In der Folge wird daher der Ansatz von Hunter und Schmidt näher diskutiert und als Ansatz der Wahl für die Durchführung der Meta-Analyse verwendet.

Der Ansatz von Hunter und Schmidt ist nicht unumstritten (vgl. z. B. Schulze, 2004), findet aber in der Praxis eine weite Verbreitung. Zahlreiche Meta-Analysen, die beispielsweise im Journal of Applied Psychology der American Psychological Association publiziert wurden, verwenden den Ansatz von Hunter und Schmidt (z. B. Williams, McDaniel & Nguyen, 2006;

Dudley et al., 2006). Die von Bijmolt und Pieters (2001) zusammengefassten Meta-Analysen im Marketing verwenden mehrheitlich den Ansatz von Hunter und Schmidt.

Hunter und Schmidt gehen davon aus, dass sich die in den Korrelationen beobachtete Varianz zerlegen lässt in die wahre Varianz des Populationsparameters, die Varianz, die auf systematische Messfehler zurückzuführen ist und eine unsystematisch wirkende Varianz des Stichprobenfehlers.

Im Idealfall sollten die systematisch wirkenden Artefakte soweit quantifizierbar sein, dass sie einen möglichst hohen Anteil der beobachteten Varianz aufklären können. Es wird davon ausgegangen, dass sich die Artefakte durch unterschiedliche Korrekturen ausschalten lassen, so dass eine mögliche Residualvarianz vernachlässigbar wird und die Annahme der Generalisierbarkeit der Befunde aufrecht erhalten werden kann. Der verbleibende wahre Varianzanteil kann dann auf den Einfluss von Moderatorvariablen weiter überprüft werden. Bevor auf die Besonderheiten der Homogenitätsanalyse bei Hunter und Schmidt näher eingegangen wird, wird zunächst das Vorgehen zur Bestimmung der mittleren Effektgröße und der dazu gehörigen Varianz und Konfidenzintervalle dargestellt.

Die mittlere Effektgröße wird im Ansatz von Hunter und Schmidt ohne eine Fisher-z-Transformation bestimmt:

$$\bar{r} = \frac{\sum_{i=1}^k n_i r_i}{\sum_{i=1}^k n_i} \quad (28)$$

Die Varianz der mittleren Effektgröße ist gegeben durch

$$\hat{\sigma}_{\bar{r}}^2 = \frac{1}{k} \left(\frac{\sum_{i=1}^k n_i (r_i - \bar{r})^2}{\sum_{i=1}^k n_i} \right) \quad (29)$$

Für die Bestimmung der Konfidenzintervalle wird analog zu dem Ansatz von Hedges und Olkin der Standardfehler der mittleren Effektgröße verwendet, der sich ergibt als Quadratwurzel aus der Varianz der mittleren Effektgröße $\hat{\sigma}_{\bar{r}}^2$ gemäß Formel 29. Die Grenzen eines 95%-Konfidenzintervalls ergeben sich dann wie folgt:

$$\begin{aligned} r_{\text{untereGrenze}} &= \bar{r} - 1,96\hat{\sigma}_{\bar{r}} \\ r_{\text{obereGrenze}} &= \bar{r} + 1,96\hat{\sigma}_{\bar{r}} \end{aligned} \quad (30)$$

Analog zu dem Ansatz von Hedges und Olkin wird im Ansatz von Hunter und Schmidt ein Homogenitätstest durchgeführt, der praktisch dem Q-Test im Ansatz von Hedges und Olkin entspricht (Schulze, 2004). Ein empirischer Q-Wert für den Homogenitätstest bei Hunter und Schmidt wird ermittelt durch

$$Q = \frac{\sum_{i=1}^k (n_i - 1)(r_i - \bar{r})^2}{(1 - \bar{r}^2)^2} \quad (31)$$

Der empirische Q-Wert folgt einer χ^2 -Verteilung mit k-1 Freiheitsgraden. Durch den Vergleich mit dem entsprechenden kritischen Wert der χ^2 -Verteilung wird getestet, ob die Annahme der Existenz eines gemeinsamen Populationsparameters gerechtfertigt ist. Ein signifikantes Ergebnis deutet auf eine heterogene Studienzusammensetzung hin. Im Falle eines signifikanten Ergebnisses des Q-Tests in der vorliegenden Meta-Analyse soll die Grundgesamtheit der Validitäts- bzw. Reliabilitätskoeffizienten unterteilt werden anhand von zwei Faktoren. Zum Einen sollen separate Meta-Analysen für jede Gruppe an Präferenzmessverfahren vorgenommen werden. Zum Zweiten wird für jede Gruppe der Art der Validitäts- bzw. Reliabilitätsbestimmung jeweils eine separate Meta-Analyse durchgeführt. Für jede Gruppe kann dann separat die Homogenitätsannahme mittels Q-Test überprüft werden.

Eine Besonderheit im Ansatz von Hunter und Schmidt ist die Überprüfung der Homogenitätsannahme durch die 75%-Regel. Diese basiert auf der Modellannahme, dass die Variabilität der in den Einzelstudien beobachteten Effektgrößen nicht nur auf die zufällige Auswahl der Probanden in den einzelnen Studien, also einem Stichprobenfehler auf Probandenebene, zurückzuführen ist, sondern auch von der zufälligen Auswahl der in der Meta-Analyse berücksichtigten Studien, dem Stichprobenfehler auf Studienebene, abhängt.

Die beobachtete Varianz von r lässt sich nach Hunter und Schmidt wie folgt unterteilen:

$$S_r^2 = \hat{\sigma}_\rho^2 + \hat{\sigma}_e^2 \quad , \quad (32)$$

wobei $\hat{\sigma}_\rho^2$ einen Schätzer für die Varianz des Populationsparameters und $\hat{\sigma}_e^2$ einen Schätzer für die Stichprobenvarianz darstellt. Unter der Annahme der Homogenität müsste $\hat{\sigma}_\rho^2 = 0$ gelten und damit die beobachtete Varianz der Effektgrößen alleine auf den Stichprobenfehler auf Probandenebene zurückzuführen sein.

Hunter und Schmidt gehen davon aus, dass über den Stichprobenfehler hinaus gehende Fehlerquellen Artefakte sind, die grundsätzlich identifizierbar und damit einer Korrektur zugänglich sind. Es wird angestrebt, die in den Primärstudien aufgeführten Informationen über Artefakteinflüsse zu nutzen, um die Varianz des Populationsparameters möglichst getrennt von der Varianz zu analysieren, die sich durch systematisch wirkende Fehlerquellen ergibt. Von Hunter und Schmidt vorgelegte Artefaktkorrekturen beziehen sich beispielsweise auf den Einfluss einer eingeschränkten Reliabilität der Messung der abhängigen bzw. unabhängigen Variablen, der künstlichen Dichotomisierung kontinuierlicher Variablen, der imperfekten Konstruktvalidität der abhängigen bzw. unabhängigen Variablen oder von Range-Restriktionen bei den berücksichtigten Variablen. Eine detaillierte Aufstellung von insgesamt zehn Artefaktkorrekturen findet sich in Cooper und Hedges (1994).

Für die vorliegende Arbeit treffen die meisten Artefaktkorrekturen nicht zu. Für die Reliabilitätskorrektur von Hunter und Schmidt fehlen in den meisten Publikationen zu Präferenzmessverfahren die notwendigen Informationen. Zudem repräsentieren korrigierte Korrelationskoeffizienten ein Reliabilitäts- bzw. Validitätsmaß in einer sogenannten idealen Welt, in der die Populationsparameter ohne systematische Artefakteinflüsse bestimmt werden können. Für die Forschungspraxis stellen unkorrigierte mittlere Effektgrößen jedoch einen realistischeren Orientierungspunkt dar als ein um üblicher Weise auftretende Fehlerquellen bereinigtes Effektmaß. In der vorliegenden Meta-Analyse wird auf Artefaktkorrekturen nach Hunter und Schmidt verzichtet. Das durch die Meta-Analyse ermittelte mittlere Validitäts- bzw. Reliabilitätsmaß kann demnach auch im Sinne der absoluten Höhe als realistisches Benchmark für weitere Untersuchungen im Bereich der Präferenzmessmethoden dienen.

Sofern Artefakteinflüsse in einer Meta-Analyse unberücksichtigt bleiben, wird der Anteil der wahren Varianz an der beobachteten Varianz überschätzt, die Annahme einer homogenen

Grundgesamtheit aller Studien also vorschnell zurückgewiesen. Da es praktisch nicht möglich ist, alle artefaktbedingten Einflüsse zu kontrollieren, haben Hunter und Schmidt die 75%-Regel eingeführt. Es wird akzeptiert, dass in einer Meta-Analyse 25 % der beobachteten Varianz nicht alleine durch den Stichprobenfehler erklärt werden können. Mit anderen Worten soll der Quotient aus der Varianz des Stichprobenfehlers und der beobachteten Varianz nicht kleiner als .75 sein, um noch von einem homogenen Studienpool sprechen zu können. Die Argumentation sieht dabei vor, dass es dann keine echte Varianz des Populationsparameters gibt, wenn in einer Meta-Analyse die Varianz des Stichprobenfehlers mindestens 75 % der beobachteten Varianz ausmacht. Formal ausgedrückt stellt sich die 75%-Regel wie folgt dar:

$$x = \frac{\hat{\sigma}_e^2}{S_r^2} \quad (33)$$

mit: Homogenität, falls $x \geq .75$

Heterogenität, falls $x < .75$

Zur Berechnung der 75 % - Regel wird die Stichprobenvarianz geschätzt durch

$$\hat{\sigma}_e^2 = \frac{(1 - \bar{r}^2)^2}{(N/k) - 1} \quad (34)$$

Die beobachtete Varianz ist gegeben durch

$$S_r^2 = \hat{\sigma}_r^2 k = \frac{1}{N} \sum_{i=1}^k n_i (r_i - \bar{r})^2 \quad (35)$$

In der Kombination der Formeln (32), (34) und (35) lässt sich $\hat{\sigma}_\rho^2$ als Schätzer für die Heterogenitätsvarianz bestimmen. Die Quadratwurzel von $\hat{\sigma}_\rho^2$ stellt einen Schätzer für die Streuung des Populationsparameters dar und wird in der vorliegenden Arbeit ergänzend angegeben.

In der Zusammenfassung lässt sich festhalten, dass für die vorliegende Fragestellung der Integration der Forschungsergebnisse zur Reliabilität bzw. Validität von Präferenzmessverfahren der Ansatz von Hunter und Schmidt geeignet erscheint und für die Meta-Analyse verwendet wird. Es werden in der Folge Auswertungen auf der Ebene der gepoolten Gesamtstichprobe für die Reliabilität und Validität vorgenommen. Sofern bei der Durchführung der Homogenitätsanalyse mittels Q-Test gemäß Formel (31) und der

Anwendung der 75%-Regel gemäß Formel (33) von einer heterogenen Grundgesamtheit ausgegangen werden muss, werden Subgruppen gebildet anhand der Faktoren der Gruppe der Präferenzmessverfahren und der Art des verwendeten Reliabilitäts- bzw. Validitätsmaßes. Für die sich ergebenden Untergruppen werden separate Meta-Analysen durchgeführt.

Für jede der durchgeführten Meta-Analysen werden jeweils die gewichteten und ungewichteten mittleren Effektgrößen angegeben, die nach dem Ansatz von Hunter und Schmidt gemäß Formel (28) bestimmt wird. Weiterhin wird die Varianz der mittleren Effektgröße gemäß Formel (29) angegeben, die gemäß Formel (30) für die Berechnung der Grenzen eines 95%-Konfidenzintervalls um die jeweilige mittlere Effektgröße verwendet werden.

Schließlich werden in der Folge für jede der mittleren Effektgrößen die beobachtete Varianz gemäß Formel (35) und der Schätzer für die Stichprobenfehlervarianz gemäß Formel (34) angegeben. Mit diesen beiden Kennziffern lässt sich gemäß Formel (33) der Anteil der Stichprobenfehlervarianz an der beobachteten Varianz feststellen, der in den folgenden Ergebnistabellen mit $V\%$ dargestellt werden wird.

4.4.4 Integration der unterschiedlichen Effektgrößen der Messgütebestimmung der Präferenzmessverfahren

Wie im Kapitel 3 zusammengestellt, gibt es unterschiedliche Methoden zur Bestimmung der Messgüte von Präferenzmessverfahren. Die Tabellen 3.1 und 3.3 zeigen die unterschiedlichen Effektgrößen, die als Indikatoren für die Messgüte der Präferenzmessverfahren in der Meta-Analyse berücksichtigt werden sollen. Da es sich um unterschiedliche Maße handelt, muss im Verlauf der Analyse überprüft werden, ob die Art der Gütemessung einen Einfluss auf die Höhe der Validität bzw. Reliabilität aufweist. Sofern ein signifikanter Q-Test auf eine heterogene Grundgesamtheit schließen lässt, werden über eine gepoolte Auswertung hinaus separate Analysen für jede Unterart der Reliabilitäts- bzw. Validitätsbestimmung vorgenommen.

Für die integrierende Meta-Analyse aller Effektgrößen der Reliabilität bzw. der Validität ist eine Synthese der unterschiedlichen Arten von Effektgrößen notwendig. Als Basiseffektgröße

soll der Produkt-Moment-Korrelationskoeffizient r dienen. Die folgenden Formeln geben an, wie die unterschiedlichen Effektgrößen in ein zu Pearson's r äquivalentes Effektmaß transformiert werden können (Gilpin, 1993).

$$r = \frac{6}{\pi} * \arcsin\left(\frac{r_s}{2}\right) , \text{ mit } r_s \text{ als Rangkorrelationskoeffizient rho nach Spearman} \quad (36)$$

$$r = \frac{2}{\pi} * \arcsin \tau \quad , \text{ mit } \tau \text{ als Rang-Zusammenhangsmaß tau nach Kendall} \quad (37)$$

Die Hitraten können mit Cohen's h (Cohen, 1988; Cooper & Hedges, 1994) transformiert werden. Der Index Cohen's h stellt die Differenz zwischen einem beobachteten Anteil und einem theoretisch erwarteten Anteil dar, nachdem jeder der Anteile mittels Arcsinus-Transformation zu Radianten umgewandelt wurde:

$$h = 2 \arcsin(\sqrt{\text{Hitrate}}) - 2 \arcsin(\sqrt{\text{Zufallsrate}}) . \quad (38)$$

Die Hitrate stellt die empirische Trefferquote für die Vorhersage einer präferierten Alternative gemittelt über alle Probanden und Wahlsets dar. Als Zufallsrate wird die durch Zufallswahl zu erwartende Trefferquote eingesetzt, im Falle von beispielsweise 4 Alternativen pro Wahlset beträgt die Zufallsrate .25.

Die mit Formel (38) berechneten Cohen's h -Indizes können wie folgt in zu r äquivalente Effektmaße überführt werden:

$$r = \frac{e^{2h} - 1}{e^{2h} + 1} . \quad (39)$$

Neben der Transformation der Hitraten in ein zum Produkt-Moment-Koeffizienten äquivalentes Effektmaß kann eine zusammenfassende Vergleichbarkeit innerhalb der Hitraten mit unterschiedlichen zugrunde liegenden Zufallsraten über den Index II erreicht werden (Rosenthal, 1994). Der Index II gibt den Anteil der korrekten Vorhersagen aus einem Wahlset von zwei Alternativen an. Wenn ursprünglich mehr als zwei Alternativen im Wahlset waren, drückt II die Trefferquote derart aus, als wären nur zwei Alternativen im Wahlset gewesen.

$$\Pi = \frac{\textit{Hitrate}(k-1)}{\textit{Hitrate}(k-2)+1} . \quad (40)$$

Der Index Π ist geeignet, alle Hitraten so zu standardisieren, dass sie als Trefferquoten auf der theoretischen Basis von zwei Wahlalternativen ausgedrückt werden.

5. Aufbau und Durchführung der Meta-Analyse

Das Vorgehen bei der Meta-Analyse wird anhand der in Kapitel 4.3 dargestellten sechs Prozessschritte zur Durchführung einer Meta-Analyse beschrieben.

5.1 Spezifikation der Forschungsfrage und Definition der Charakteristika der relevanten Forschungsliteratur

Es ist das Ziel der Meta-Analyse, die bisherigen Forschungsergebnisse zur Bestimmung der Reliabilität und Validität von Präferenzmessverfahren zu integrieren. Dabei soll insbesondere überprüft werden, ob die Art des verwendeten Verfahrens zur Präferenzmessung einen Einfluss auf die Reliabilität und Validität der Ergebnisse hat. Weiterhin soll der Einfluss weiterer Faktoren des Untersuchungsdesigns auf die Höhe der Reliabilität bzw. Validität überprüft werden. Zu den diesbezüglich interessierenden Faktoren zählen die Art der Probanden (Studenten oder Nicht-Studenten), die Anzahl der im Design berücksichtigten Attribute und die Art der Validitäts- bzw. Reliabilitätsbestimmung.

5.2 Literatursuche

Als Ausgangspunkt für die Literatursuche diente die conjoint analysis literature list (CALL), eine Internetdatenbank, in der Melles (2000) über 1.300 Studien, die sich im erweiterten Sinne mit dem Thema der Präferenzmessung befasst haben, aufgeführt hat.

Zur Ergänzung wurden eine Internetrecherche und eine elektronische Datenbankrecherche vorgenommen. Als Suchbegriffe wurden jeweils verwendet: „conjoint“, „conjoint analysis“, „conjoint measurement“, „choice“, „discrete choice“, „choice-based“, „preference“, „preference measurement“, „Präferenzmessung“, „validity“ und „Validität“.

Die konsultierten Datenbanken waren: Econlit, Psyndex, Psycinfo, Medline, ERIC, SSCI, und SCI. Diese Datenbanken decken zusammen weltweit über 4.000 medizinische Fachzeitschriften, 3.300 betriebswirtschaftliche Fachzeitschriften, 1.400 sozialwissenschaftliche Fachzeitschriften, 500 psychologische Fachzeitschriften sowie über

1,8 Mio. technische Reports, Dissertationen, Bücher, Buchkapitel und Zeitschriftenartikel im psychologischen Bereich ab. Die Literatursuche erfolgte zwischen August 2001 und Mai 2005. Tabelle 4.1 zeigt die Verteilung der in den Meta-Analysen zur Reliabilität und Validität insgesamt berücksichtigten Effektgrößen im Hinblick auf die unterschiedlichen Arten der Publikation.

<i>Publikationsart</i>	<i>M</i>	<i>k</i>	%
Journalartikel	76	469	45.6
Arbeitspapier / Konferenzpapier	27	195	18.9
Buch oder Buchartikel	11	104	10.1
Dissertation / Habilitation / Diplomarbeit	7	103	10.0
Sonstige (z. B. Internet, Sawtooth Papers)	14	157	15.4
Total	135	1,028	100.0

Anmerkung: *M* = Anzahl der Publikationen, *k* = Anzahl der Effektgrößen, % = gültiger Anteil an Gesamtstichprobe

Tab. 5.1: Übersicht über die Herkunft der in der Meta-Analyse verwendeten Effektgrößen

5.3 Kodierung der Studienmerkmale

Auf der Basis der Literatursuche wurden Untersuchungen in die Meta-Analyse aufgenommen, wenn sie

- mindestens eines der Verfahren aus den folgenden fünf Gruppen der Präferenzmessung verwendet haben: traditionelle Conjoint-Analyse, self-explicated Ansatz, wahlbasierte Conjoint-Analyse, ACA oder andere hybride Methoden
und
- mindestens eines der Messgütekriterien zur Reliabilitäts- oder Validitätsbestimmung aus der Aufstellung der Effektgrößen gemäß der Tabellen 3.1 bzw. 3.3 berichteten.

Untersuchungen auf der Basis von Simulationsdaten wurden von der Meta-Analyse ausgeschlossen. Die Tabellen 3.1 bzw. 3.3 zeigen die Anzahl der in der Meta-Analyse zur Reliabilität bzw. zur Validität verwendeten Publikationen und Effektmaße. Eine detaillierte Aufstellung der in den einzelnen Meta-Analysen verwendeten Studien findet sich im Anhang.

Für die Meta-Analysen wurden aus den Studien Informationen zu insgesamt 51 Variablen kodiert. Die Dateneingabe erfolgte in eine SPSS-Datenmaske mit vorgegebenen Kategorien für die qualitativen Variablen. Jede Untersuchung wurde von zwei geschulten Kodierern unabhängig voneinander kodiert und die Ergebnisse miteinander verglichen.

Insgesamt wurden 36.540 Verschlüsselungen vorgenommen zu Variablen mit Studiencharakteristika und Beschreibungen der unabhängigen Variablen in der jeweiligen Untersuchung. Dabei wurden 578 Abweichungen registriert, was einer Rate der Nicht-Übereinstimmung von 1,6 % entspricht.

Die Verfahrensgruppe wurde von den beiden Kodierern in 7 von 1.044 Fällen abweichend voneinander verschlüsselt (Rate der Nicht-Übereinstimmung 0,7 %). Die relevanten Effektgrößen wurden in 17 von 1.044 Fällen abweichend voneinander eingegeben (Rate der Nicht-Übereinstimmung 1,7 %). Im Falle von Abweichungen wurden die Studien gemeinsam re-analysiert und in Abstimmung entschieden, welche Kodierung in der Meta-Analyse verwendet werden sollte.

5.4 Analyse der Datenbasis

Jahr der Publikation

Die berücksichtigten Studien sind zwischen 1972 und 2005 erschienen. Im Einzelnen verteilen sich die Primärstudien wie in Tabelle 5.2 dargestellt.

<i>Publikationsjahr</i>	<i>M</i>	<i>k</i>	<i>%</i>
1972 – 1974	3	4	.4
1975 – 1979	10	45	4.4
1980 – 1984	13	131	12.7
1985 – 1989	19	99	9.6
1990 – 1994	21	207	20.1
1995 – 1999	34	261	25.5
2000 – 5 / 2005	35	281	27.3
Total	135	1,028	100.0

Tab. 5.2: Übersicht über das Erscheinungsjahr der in der Meta-Analyse verwendeten Effektgrößen

Objektbereich

Von den berücksichtigten Effektgrößen stammen knapp ein Drittel (31,7 %) aus Studien, die als Objektbereich der Untersuchung technische Geräte, insbesondere Computer, Computer-Zubehör und Haushaltsgeräte, verwendeten. Überdurchschnittlich häufig finden gerade bei wissenschaftlichen Fragestellungen Apartments als Präferenzobjekte Verwendung (18,8 %). Nahrungsmittel und Jobs (jeweils 7,5 %) und Autos (4,7 %) folgen als weitere Kategorien.

Insgesamt zeigte sich ein breites Spektrum an Objektbereichen, die in den berücksichtigten Präferenzstudien Gegenstand der Untersuchung waren: 29,7 % der kodierten Effektgrößen ließen sich zu keiner nennenswerten übergeordneten Objektkategorie zusammenfassen. Tabelle 5.3 zeigt die Verteilung der Effektgrößen auf Studien der unterschiedlichen Untersuchungsobjekte.

<i>Objektbereich</i>	<i>k</i>	<i>%</i>
Technische Geräte und Zubehör	318	31.7
Apartments	188	18.8
Jobs	75	7.5
Nahrungsmittel	76	7.6
Autos	47	4.7
Sonstige	298	29.7
Total	1,002	100.0

Tab. 5.3: Übersicht über die untersuchten Objektbereiche der in der Meta-Analyse verwendeten Effektgrößen

Merkmalsgewinnung

Bei 67 % der in der Meta-Analyse verwendeten Effektgrößen konnte aus den Publikationen identifiziert werden, wie die Definition der Attribute und Ausprägungsstufen im Vorfeld der Untersuchung erfolgt ist. Ein Drittel der Publikationen enthielt keine Angabe zur Merkmalsgewinnung.

In gut einem Drittel der Fälle, die Angaben zur Merkmalsgewinnung enthielten, wurde eine einzelne Literaturrecherche zur Merkmalsgewinnung vorgenommen (35,0 %). Eine Vorstudie wurde bei 11,7 % der berücksichtigten Effektgrößen, eine Experten- bzw. Zielgruppenbefragung im Vorfeld der eigentlichen Untersuchung in 3,9 % der Fälle berichtet.

Knapp die Hälfte (47,5 %) der Effektgrößen, zu denen Angaben über die Art der Merkmalsgewinnung vorlagen, basieren auf Untersuchungen, in denen die Merkmale durch eine Kombination von Literaturrecherche und Vorstudie / Vorabbefragung definiert wurden. Tabelle 5.4 stellt die Daten zur Merkmalsgewinnung noch einmal übersichtlich dar. Der hohe Anteil an Studien, die Angaben über ein systematisches Vorgehen bei der Merkmalsgewinnung enthielten, kann als ein Indikator für die Qualität der berücksichtigten Untersuchungen gelten.

<i>Merkmalsgewinnung</i>	<i>k</i>	<i>%</i>
Literaturrecherche	242	35.0
Vorstudie	81	11.7
Experten- bzw. Zielgruppenbefragung	27	3.9
Kombination aus Literaturrecherche und Vorstudie bzw. Befragung	328	47.5
Sonstige	13	1.9
Total	691	100.0

Tab. 5.4: Übersicht über die Merkmalsgewinnung bei den Untersuchungen, die in die Meta-Analyse eingegangen sind

Probanden

Von den in der Meta-Analyse berücksichtigten Effektgrößen stammen 68,9 % aus Untersuchungen, die keine Studenten, sondern die direkte Zielgruppe des Objektbereichs als Probanden verwendeten. In 22,5 % der berücksichtigten Effektgrößen bildeten Studenten die Untersuchungsteilnehmer. In 8,6 % der Fälle konnte aus der Publikation nicht erkannt werden, um welche Art der Probanden es sich handelte.

In allen Studien, in denen Studenten als Probanden dienten, waren Objektbereiche Gegenstand der Untersuchung, die für Studenten eine persönliche Relevanz haben (z. B. Apartments, Jobs). Die Studenten bildeten die Zielgruppe für den jeweiligen Objektbereich.

Wie in der durchgeführten Meta-Analyse gezeigt werden wird, weisen Untersuchungen mit Studenten keine höhere Validität auf als Untersuchungen mit Nicht-Studenten. Eine mögliche Überschätzung der Gütemaße aufgrund des Einbezugs von Untersuchungen mit Studenten als Probanden scheint demnach in der vorliegenden Meta-Analyse nicht vorzuliegen.

Anzahl der Attribute

Die Anzahl der in den berücksichtigten Präferenzmessungen verwendeten Attribute liegt zwischen 2 und 25 Attributen mit einem Mittelwert von 6,5 Attributen. 71,8 % der Effektgrößen stammen aus Untersuchungen mit 4 bis 8 Attributen. Lediglich 3,8 % der berücksichtigten Effektgrößen stammen aus Untersuchungen mit mehr als 10 Attributen. Abbildung 5.1 zeigt die Häufigkeitsverteilung der berücksichtigten Effektgrößen in Abhängigkeit der Anzahl der in dem jeweiligen Untersuchungsdesign berücksichtigten Attribute.

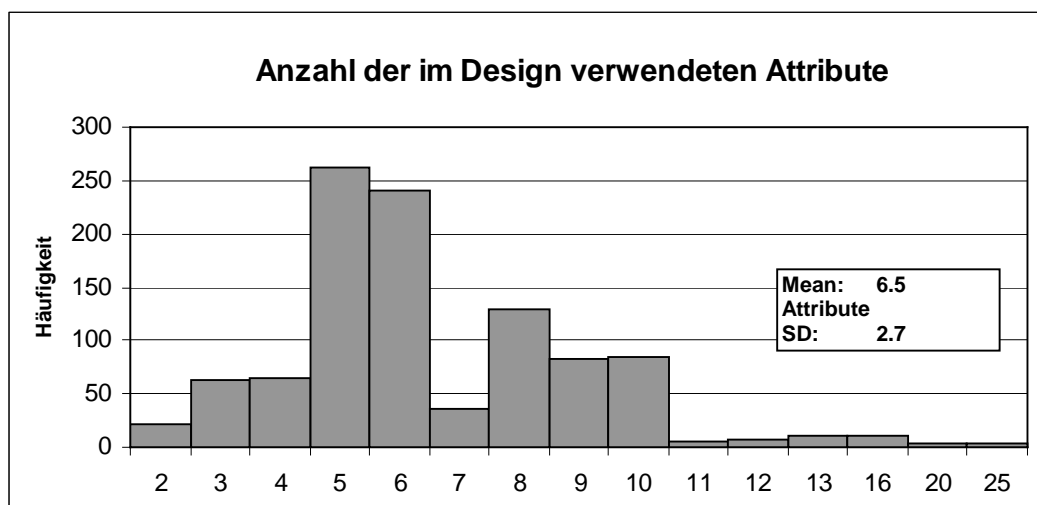


Abb. 5.1: Übersicht über die Größe der Untersuchungsdesigns der in der Meta-Analyse berücksichtigten Effektgrößen

Die in der Meta-Analyse berücksichtigten Untersuchungen, in denen die adaptive Conjoint-Analyse (ACA) zum Einsatz kam, weisen im Mittel eine größere Anzahl an Attributen (7,9) auf als self-explicated Ansätze (6,7), wahlbasierte Conjoint-Analysen (6,5) und traditionelle Conjoint-Analysen (5,9).

Hold-out-Aufgabe

Die Hold-out-Aufgabe wurde in der Regel nach der Kalibrierungsaufgabe gestellt (59,7 % der berücksichtigten Effektgrößen). In 11,1 % der Fälle wurde erst die Hold-out-Aufgabe und dann die Kalibrierungsaufgabe gestellt. Nicht üblich ist die Integration der Hold-out-Aufgabe

in die Kalibrierungsaufgabe (3,4 %). 19,3 % der Effektgrößen basieren auf Untersuchungen, welche die Hold-out-Aufgabe an einem anderen Tag vorgelegt haben.

In den meisten Fällen bestand die Hold-out-Aufgabe für die Probanden in der Wahl der meist präferierten Alternative aus einem Wahlset (46,1 %). Zudem wurden in der Hold-out-Aufgabe einstufigsskalenbasierte (26,7 %) bzw. rangordnungsbasierte (17,4 %) Präferenzdaten abgefragt.

Von den Effektgrößen, die auf wahlbasierte Hold-out-Aufgaben zurückgehen, wird in 19,4 % der Fälle berichtet, dass die Wahlsets nutzenbalanciert zusammengestellt wurden. In nutzenbalancierten Wahlsets dominiert keine Option die anderen Alternativen in allen Attributen. Für die Wahloptionen wird ein ähnlicher Gesamtnutzen unterstellt. Nutzenbalancierte Wahlsets gelten als anspruchsvolles Vorhersagekriterium.

Die Anzahl der Optionen in den Hold-out-Wahl-Aufgaben liegt zwischen 2 und 32, wobei nur 9 Effektgrößen (2,2 %) auf Wahlsets mit mehr als 10 Alternativen zurückgehen. Knapp die Hälfte der auf der Basis von Hold-out-Wahlaufgaben ermittelten Effektgrößen resultiert aus Hold-out-Wahlsets mit drei Optionen (49,2 %), der Modalwert liegt bei 3 Alternativen pro Wahlset, der Mittelwert bei 4,1 Alternativen. Tabelle 5.5 veranschaulicht die Häufigkeitsverteilung der Anzahl der Alternativen in den Hold-out-Wahlsets.

<i>Hold – out – Optionen</i>	<i>k</i>	<i>%</i>
2 Optionen	71	16.2
3 Optionen	216	49.2
4 Optionen	46	10.5
5 Optionen	48	10.9
6 bis 8 Optionen	9	2.1
9 bis 10 Optionen	40	9.2
mehr als 10 Optionen	9	2.2
Total	439	100.0

Tab. 5.5: Übersicht über die Anzahl der Wahlalternativen in den Hold-out-Szenarien für die in der Meta-Analyse berücksichtigten Effektgrößen

5.5 Vorgehen bei der Datenauswertung und der Homogenitäts- und Moderatorenanalyse

Die Datenanalyse erfolgt in drei Schritten. In einem ersten Schritt werden die unterschiedlichen Effektgrößen zu einem korrelationsäquivalenten Effektmaß gemäß der Formeln (36) bis (39) transformiert. Die korrelationsäquivalenten Effektmaße werden aus den in Kapitel 4.4.3 dargelegten Gründen mit dem Ansatz von Hunter und Schmidt (1990) zu einem gepoolten Effektmaß zusammengefasst. Wenn mehr als eine Effektgröße in der jeweiligen Primärstudie vorliegt, wird festgelegt, dass der Produkt-Moment-Korrelationskoeffizient verwendet wird. Liegt kein Produkt-Moment-Korrelationskoeffizient als Effektmaß vor, wird das zu den Rangkorrelationen äquivalente Effektmaß und erst dann das zur Hitrate äquivalente Effektmaß verwendet.

In einem zweiten Schritt wird überprüft, ob die zusammengeführten Effektmaße aus einem homogenen Universum von potenziellen Repräsentationen des universellen Populationsparameters stammen. Hierzu werden ein Q-Test und der 75%-Homogenitätstest nach Hunter und Schmidt durchgeführt. Sollten beide Tests auf eine heterogene Studienzusammenstellung schließen lassen, wird die Gesamtpopulation unterteilt in fünf Subgruppen in Abhängigkeit des Faktors des verwendeten Verfahrens der Präferenzmessung. Für jede der Verfahrensgruppen der Präferenzmessung werden separate mittlere Effektgrößen ermittelt. Die beobachteten Unterschiede in den Mittelwerten werden varianzanalytisch jeweils paarweise auf statistische Signifikanz überprüft. Mit Hilfe der Streuung der mittleren Effektgrößen werden Konfidenzintervalle für die durchschnittlichen Reliabilitäts- bzw. Validitätskoeffizienten bestimmt.

Abschließend werden weitere, potenzielle Einflussgrößen auf die Messgüte im Rahmen der Identifikation von Moderatorvariablen untersucht. Zunächst wird überprüft, ob die Art der Reliabilitäts- bzw. Validitätsbestimmung einen moderierenden Einfluss auf die Höhe der Effektgrößen aufweist. Abschließend werden über eine multiple Regression diverse Faktoren auf einen signifikanten Einfluss auf die Validitätshöhe untersucht. Relevante Einflussfaktoren werden im Ergebnisteil berichtet.

6. Ergebnisse der Meta-Analyse

Es werden zunächst die Ergebnisse der Meta-Analyse berichtet, die sich auf die Kriterien zur Bestimmung der Validität beziehen. Im zweiten Teil dieses Kapitels werden die Ergebnisse der Meta-Analyse zur Reliabilitätsbestimmung dargestellt.

6.1 Ergebnisse der Meta-Analyse für Effektgrößen der Validität

6.1.1 Ergebnisse der Meta-Analyse für Validitätsmaße auf der individuellen Ebene

Als Basis für die gepoolte Auswertung der unterschiedlichen Effektgrößen wird der Produkt-Moment-Korrelationskoeffizient verwendet. Tabelle 6.1 gibt die zusammengeführten, mittleren Effektgrößen für die berücksichtigten Verfahrensgruppen, jeweils ungewichtet und an der Stichprobengröße gewichtet, an.

<i>Verfahren</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95% CI</i>
CA	322	31,316	.67	.68	.0101	.033	.0030	.173	9.10	.66-.70
ACA	172	38,860	.75	.75	.0145	.036	.0009	.187	2.36	.72-.78
SE	103	10,643	.64	.62	.0200	.041	.0037	.193	9.03	.58-.66
CBCA	148	48,529	.80	.84	.0052	.004	.0003	.061	6.63	.83-.85
HYB	32	8,226	.67	.63	.0262	.022	.0014	.143	6.46	.58-.68
Total	777	137,574	.71	.75	.0062	.030	.0011	.170	3.62	.74-.76

Anmerkung: CA = Traditionelle Conjoint-Analyse, ACA = Adaptive Conjoint-Analyse, SE = Self-Explicated-Modelle, CBCA = Wahlbasierte Conjoint-Analyse, HYB = Hybride Conjoint-Analysen, *k* = Anzahl der Effektgrößen, *N* = Stichprobenumfang, \bar{r}_u = ungewichtete mittlere Korrelation, \bar{r}_g = gewichtete mittlere Korrelation, $\hat{\sigma}_{\bar{r}}$ = Streuung der mittleren Korrelation, S_r^2 = beobachtete Varianz, $\hat{\sigma}_e^2$ = Stichprobenfehlervarianz, $\hat{\sigma}_\rho$ = geschätzte Populationsstreuung, *V%* = Anteil der Stichprobenfehlervarianz an der beobachteten Varianz, *CI* = Konfidenzintervall.

Für $p < .05$ unterscheiden sich jeweils paarweise signifikant (mit Bonferroni-Adjustierung): CBCA/CA, CBCA/SE, CBCA/HYB, ACA/CA, ACA/SE.

Für $p < .05$ unterscheiden sich alle gewichteten Mittelwerte jeweils paarweise signifikant (mit Bonferroni-Adjustierung).

Tab. 6.1: Ergebnisse der Meta-Analyse für Effektgrößen der Validität

Sowohl in der gewichteten als auch der ungewichteten Meta-Analyse weisen wahlbasierte Conjoint-Analysen im Mittel die höchste Validität auf, gefolgt von ACA und anderen hybriden Verfahren bzw. der traditionellen Conjoint-Analyse. In beiden Auswertungsbedingungen weisen self-explicated-Ansätze die geringste mittlere Validität auf.

Für den Homogenitätstest wird ein Q-Wert gemäß Formel (31) ermittelt. Auf der Ebene der Gesamtstichprobe ergibt sich ein empirisches Q, das größer ist als der kritische Wert der Chi-Quadrat-Verteilung ($Q = 21.478$, $df = 776$, $p = .001$). Aufgrund des signifikanten Q-Tests muss die Annahme einer homogenen Grundgesamtheit zurückgewiesen werden.

Der Q-Test wird durch die 75%-Regel nach Hunter und Schmidt ergänzt. Auf der Ebene der Gesamtstichprobe beträgt die geschätzte Stichprobenvarianz gemäß der Formel (34) $\hat{\sigma}_e^2 = .0011$. Die beobachtete Varianz beträgt gemäß der Formel (39) auf der Ebene der Gesamtstichprobe $S_r^2 = .030$. Der Anteil der Stichprobenvarianz an der beobachteten Varianz beträgt somit 3,6 %. Demnach muss auch nach der 75 %-Regel die Annahme einer homogenen Studienstichprobe zurückgewiesen werden.

Die heterogene Grundgesamtheit wird unterteilt anhand des Faktors des verwendeten Verfahrens der Präferenzmessung. Es werden fünf Subgruppen gebildet, für die jeweils separate Mittelwerte und Konfidenzintervalle bestimmt werden. Die Ergebnisse sind in Tabelle 6.1 dargestellt.

Für die einzelnen Verfahrensgruppen werden jeweils separate Q-Werte ermittelt. Unter Berücksichtigung der entsprechenden Freiheitsgrade $k-1$ muss die Annahme einer homogenen Grundgesamtheit auch für die gebildeten Subgruppen der jeweiligen Präferenzmessverfahren zurückgewiesen werden. Dieser Befund der Q-Tests wird durch die 75%-Regel nach Hunter und Schmidt unterstützt. In Tabelle 6.1 sind die Anteile der jeweiligen Stichprobenvarianz an der jeweiligen beobachteten Varianz in der Spalte $V\%$ dargestellt. In keiner Subgruppe werden die geforderten 75 % Stichprobenvarianzanteile erreicht. Aufgrund der festgestellten Heterogenität ist die weitere Suche nach Moderatorvariablen angezeigt. In einem ersten Schritt werden hierzu die unterschiedlichen Formen der Validitätsbestimmung als Einflussfaktor auf die Höhe der Validität untersucht.

Die Gesamtstichprobe wird anhand der Faktoren des verwendeten Verfahrens der Präferenzmessung und des verwendeten Validitätskriteriums in kleinere Untergruppen geteilt.

Es werden separate Auswertungen vorgenommen für die Hitraten und die Korrelationskoeffizienten. Die Tabellen 6.2 bis 6.6 zeigen die Ergebnisse. Die Untergruppen werden separat auf Homogenität überprüft.

Verfahren	k	N	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	V%	95% CI
CA	150	18,313	.81	.83	.0084	.011	.0008	.100	7.13	.82-.85
ACA	118	23,655	.83	.79	.0073	.006	.0007	.074	11.52	.77-.80
SE	52	6,560	.81	.84	.0198	.020	.0007	.140	3.54	.80-.87
CBCA	132	48,165	.85	.85	.0036	.002	.0002	.039	12.28	.84-.86
HYB	17	3,178	.80	.84	.0233	.009	.0005	.094	5.17	.79-.88
Total	469	99,871	.83	.83	.0037	.007	.0005	.078	6.95	.82-.84

Anmerkung: Für $p < .05$ sind die Mittelwertsunterschiede im ungewichteten Auswertungsmodus nicht signifikant.

Für $p < .05$ unterscheidet sich der gewichtete Mittelwert CBCA von allen übrigen Mittelwerten jeweils paarweise signifikant (mit Bonferroni-Adjustierung). Der gewichtete Mittelwerte ACA unterscheidet sich von allen übrigen Mittelwerten jeweils paarweise signifikant. Alle übrigen Unterschiede sind nicht signifikant.

Für $p < .05$ sind die Mittelwertsunterschiede zwischen den gewichteten Hitraten jeweils paarweise signifikant (mit Bonferroni-Adjustierung).

Die Ergebnisse der Gruppe SE basieren auf einer überdurchschnittlichen Anzahl an Wahlalternativen im Hold-out-Set. In Tabelle 6.7 findet sich eine Auswertung analog zu der Auswertung 6.2, in der dieser Einfluss zugunsten einer besseren Vergleichbarkeit korrigiert wird.

Tab. 6.2: Ergebnisse der Meta-Analyse auf der alleinigen Basis der Hitraten-Äquivalente

Verfahren	k	N	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	V%	95% CI
CA	184	14,220	.63	.66	.0106	.021	.0041	.129	19.95	.64-.68
ACA	56	2,724	.67	.63	.0208	.024	.0075	.130	30.80	.59-.68
SE	66	5,520	.61	.62	.0209	.029	.0045	.156	15.78	.58-.66
CBCA	18	559	.72	.70	.0396	.028	.0085	.141	29.97	.63-.78
HYB	10	844	.67	.65	.0551	.030	.0040	.162	13.06	.54-.76
Total	334	23,867	.64	.65	.0085	.024	.0047	.138	19.80	.63-.67

Anmerkung: Für $p < .05$ sind die Mittelwertsunterschiede im ungewichteten Auswertungsmodus nicht signifikant.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte jeweils paarweise signifikant mit Ausnahme der Differenz CA/HYB.

Tab. 6.3: Ergebnisse der Meta-Analyse auf der alleinigen Basis von Pearson's r

Verfahren	k	N	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	V%	95% CI
CA	33	3,223	.71	.61	.0323	.034	.0040	.174	11.61	.55-.68
ACA	9	13,757	.51	.73	.0927	.077	.0001	.278	.19	.55-.91
SE	3	158	.60	.53	.1289	.050	.0099	.200	19.90	.28-.79
CBCA	0	0	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.
HYB	0	0	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.
Total	45	17,138	.66	.71	.0398	.071	.0007	.266	.93	.63-.78

Anmerkung: In diese Auswertung sind die Spearman-Rangkorrelationskoeffizienten ohne Transformation in ein zu Pearson's r äquivalentes Effektmaß eingegangen.

Für $p < .05$ ist im ungewichteten Auswertungsmodus ausschließlich der Mittelwertsunterschied zwischen CA und ACA signifikant.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte jeweils paarweise signifikant.

Tab. 6.4: Ergebnisse der Meta-Analyse auf der alleinigen Basis von Spearman's rho

Verfahren	k	N	$\bar{\tau}_u$	$\bar{\tau}_g$	$\hat{\sigma}_{\bar{\tau}}$	S_τ^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\tau$	V%	95% CI
CA	30	5,125	.52	.58	.0239	.017	.0026	.121	14.94	.54-.63
ACA	6	91	.46	.46	.0404	.010	.0441	k.A.	k.A.	.38-.54
SE	17	3,808	.47	.55	.0381	.025	.0022	.150	8.96	.47-.62
CBCA	9	473	.45	.45	.0310	.009	.0122	k.A.	k.A.	.39-.51
HYB	12	5,184	.74	.75	.0047	.000	.0005	k.A.	k.A.	.74-.76
Total	74	14,681	.53	.63	.0171	.022	.0019	.140	8.67	.59-.66

Anmerkung: $\hat{\sigma}_{\bar{\tau}}$ = Streuung des mittleren Effektmaßes, $\hat{\sigma}_\tau$ = geschätzte Streuung des wahren Populationsparameters tau. In diese Auswertung sind die Rangkorrelationsmaße Kendall's tau ohne Transformation in ein zu Pearson's r äquivalentes Effektmaß eingegangen.

Für $p < .05$ sind im ungewichteten Auswertungsmodus ausschließlich die Mittelwertsunterschiede von HYB zu den übrigen Verfahren jeweils paarweise signifikant.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte jeweils paarweise signifikant mit Ausnahme von ACA/CBCA.

Tab. 6.5: Ergebnisse der Meta-Analyse auf der alleinigen Basis von Kendall's tau

<i>Verfahren</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95% CI</i>
CA	238	20,977	.62	.61	.0110	.029	.0045	.155	15.71	.59-.63
ACA	71	16,572	.62	.70	.0321	.073	.0011	.268	1.50	.64-.77
SE	77	7,895	.58	.55	.0216	.036	.0047	.176	13.21	.51-.60
CBCA	27	947	.59	.55	.0440	.052	.0143	.195	27.40	.46-.64
HYB	22	6,028	.60	.55	.0165	.006	.0018	.065	29.26	.52-.59
Total	435	52,419	.61	.62	.0102	.045	.0031	.205	6.91	.60-.64

Anmerkung: Für $p < .05$ sind die Mittelwertsunterschiede im ungewichteten Auswertungsmodus nicht signifikant.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte ACA und CA jeweils paarweise von allen übrigen Mittelwerten. Die Unterschiede zwischen CBCA, SE und HYB sind jeweils paarweise nicht signifikant (mit Bonferroni-Adjustierung).

Tab. 6.6: Ergebnisse der Meta-Analyse auf der Basis der Korrelationskoeffizienten Pearson's r , Spearman's rho und Kendall's tau (ohne Hitraten-Äquivalente)

Für die Gruppen der verwendeten Präferenzgruppen bleiben die empirischen Q-Werte größer als die kritischen Chi-Quadrat-Werte, auch wenn separate Auswertungen für die Hitraten und die Korrelationskoeffizienten vorgenommen werden. Die Homogenitätstests lassen auf heterogene Subgruppen schließen. Die 75%-Regel nach Hunter und Schmidt unterstützt die Zurückweisung der Annahme homogener Subgruppen. Der Anteil der Stichprobenvarianz an der beobachteten Varianz beträgt im höchsten Fall 30,8 %.

Die auf der Basis der Hitraten ermittelten Korrelationsäquivalente (Tab. 6.2) sind in einer einfaktoriellen Varianzanalyse mit Bonferroni-Adjustierung ($p < .001$) jeweils paarweise signifikant höher als die Korrelationsäquivalente auf der Basis der Rangkorrelationen (Tab. 6.3 und 6.4) oder als die Produkt-Moment-Korrelationen (Tab. 6.5). Die Wahl des Validierungskriteriums, also ob Wahl-, Einstufungs- oder Rangordnungsaufgaben als Hold-out-Aufgaben präsentiert werden, hat einen signifikanten Einfluss auf die Höhe des Effektmaßes.

In allen Auswertungen weisen self-explicated-Ansätze eine unterdurchschnittliche Validitätshöhe auf. Eine Ausnahme davon bildet die Auswertung auf der alleinigen Basis der zu den Hitraten äquivalenten Korrelationsmaße (Tab. 6.2). Eine mögliche Erklärung für die Umkehr der Rangfolge der Verfahren ist in der unterschiedlichen Anzahl der Alternativen im Hold-out-Set zu sehen. Wie eine Detailanalyse zeigt, weist die Verfahrensgruppe der direkten Nutzenmessung im Durchschnitt 7,8 Alternativen pro Hold-out-Wahlset auf. Die

Korrelationsäquivalente auf der Basis von Hitraten der adaptiven Conjoint-Analyse und der wahlbasierten Conjoint-Analyse basieren auf durchschnittlich 3,4 bzw. 3,7 Alternativen im Hold-out-Set. Den Validitätsmaßen der hybriden bzw. traditionellen Conjoint-Analyse liegen in dieser Auswertung Hold-out-Aufgaben mit durchschnittlich 5,4 bzw. 5,6 Alternativen pro Wahlset zugrunde. Die Anzahl der Hold-out-Alternativen geht über die damit festgelegte Zufallshitrate in die Berechnung des korrelationsäquivalenten Effektmaßes ein, was darin begründet liegt, dass eine korrekte Vorhersage der ersten Wahl aus einem großen Set an Optionen einen höheren Schwierigkeitsgrad bedeutet.

Wie Brocke (2006) zeigen konnte, ist die Schwierigkeit einer Conjoint-Wahlaufgabe jedoch maßgeblich von der wahrgenommenen Nutzenähnlichkeit der Optionen und weniger durch die Anzahl der Optionen bestimmt. Aus dieser Überlegung heraus kann eine alternative Berechnung der Validitätsmaße auf der Basis von Hitraten darin bestehen, lediglich Korrelationsäquivalente zu verwenden, die aus Studien mit maximal zehn Wahlalternativen im Hold-out-Set resultierten. Wie Tabelle 6.7 zeigt, gehen unter der Prämisse, dass nur Validitätsmaße auf der Basis von maximal zehn Wahlalternativen im Hold-out-Set verwendet werden, 95 % der in Tabelle 6.2. dargestellten Auswertungstichprobe mit ein, die extremen Daten werden jedoch zugunsten einer besseren Vergleichbarkeit ausgeschlossen.

<i>Verfahren</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	V%	95% CI
CA	136	16,634	.81	.83	.0088	.011	.0008	.098	8.00	.81-.84
ACA	113	22,924	.83	.78	.0070	.001	.0007	.069	13.30	.77-.80
SE	37	4,851	.77	.81	.0251	.023	.0009	.150	3.89	.76-.86
CBCA	128	48,052	.85	.85	.0035	.002	.0002	.037	12.86	.84-.86
HYB	15	2,824	.80	.84	.0262	.010	.0005	.099	4.50	.79-.89
Total	429	95,285	.82	.83	.0038	.006	.0004	.076	7.27	.82-.83

Anmerkung: Die vorliegende Auswertung entspricht der Auswertung in Tab. 6.2 unter Ausschluss der korrelationsäquivalenten Effektmaße, die auf Hold-out-Aufgaben mit mehr als zehn Wahlalternativen basieren.

Für $p < .05$ sind im ungewichteten Auswertungsmodus die Mittelwertsunterschiede CA/CBCA, ACA/SE, SE/CBCA signifikant. Alle übrigen ungewichteten Mittelwerte unterscheiden sich jeweils paarweise nicht signifikant.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte jeweils paarweise signifikant (mit Bonferroni-Adjustierung).

Tab. 6.7: Ergebnisse der Meta-Analyse auf der alleinigen Basis der Hitraten-Äquivalente unter Ausschluss von Effektmaßen auf der Basis von Hold-out-Aufgaben mit mehr als zehn Optionen

Aus Tabelle 6.7 ist zu erkennen, dass die Verfahren der direkten Nutzenmessung analog zu den übrigen Auswertungen eine unterdurchschnittliche Validität aufweisen. Die wahlbasierte Conjoint-Analyse erreicht in allen Auswertungen die signifikant höchsten Validitätswerte. Eine Ausnahme bildet die Auswertung auf der Basis der Rangkorrelationen, in denen die wahlbasierten Conjoint-Analysen keine überdurchschnittliche Validität zeigen. Die rangbasierten Validitätsmaße der wahlbasierten Conjoint-Analyse beruhen auf einer vergleichsweise kleinen Stichprobe, was zu entsprechend breiten Konfidenzintervallen führt. Die traditionelle Conjoint-Analyse und die adaptive Conjoint-Analyse weisen in den Einzelauswertungen durchweg solide bis leicht überdurchschnittliche Validitätsmaße auf.

6.1.2 Einflussfaktoren auf die Höhe der Validität

Zur weiteren Untersuchung des Einflusses unterschiedlicher Faktoren auf die Höhe der ermittelten durchschnittlichen Validität wurden einfaktorielle Varianzanalysen mit den folgenden Faktoren gerechnet: Art der Publikation, Studentische vs. Nicht-studentische Stichprobe, Objektbereich der Präferenzmessung und Verfahren zur Ermittlung der individualisierten Präferenzdaten bei der wahlbasierten Conjoint-Analyse (ICE- vs. Hierarchical Bayes-Verfahren). Die Ergebnisse sind in diesem Kapitel ab Tabelle 6.9 aufgeführt.

Zuvor wird in Tabelle 6.8 der Zusammenhang ausgewählter Faktoren mit der Höhe der Validität dargestellt. Als Effektmaße der Validität wurde zum Einen das Korrelationsäquivalent auf der gepoolten Basis der integrierten Korrelationsmaße und Hitratenäquivalente verwendet. Zum Anderen wurde die Korrelation zwischen den ausgewählten Faktoren mit den Effektmaßen, die auf Pearson's r zurückgehen, und mit den Effektmaßen, die auf die Hitraten zurückgehen, separat berechnet und in der Tabelle aufgeführt.

Einen positiven Zusammenhang weist die Wahl einer wahlbasierten Conjoint-Analyse (CBCA) mit der Validitätshöhe auf. Die Wahl eines Verfahrens der direkten Nutzenmessung (SE-Verfahren) weist einen schwachen negativen Zusammenhang mit der Validitätshöhe auf. Ein jüngeres Publikationsjahr geht mit einer höheren durchschnittlichen Validität einher. Die Anzahl der Attribute in einem Design der Präferenzmessung sowie die Größe der Untersuchungsstichprobe und die Anzahl der Stimuli in der Hold-out-Aufgabe weisen einen untergeordneten Zusammenhang mit den Effektgrößen der Validität auf.

<i>Faktor</i>	$r_{ges.}$		$r_{nurKorr.}$		$r_{nurHitraten}$	
	<i>gew.</i>	<i>ungew.</i>	<i>gew.</i>	<i>ungew.</i>	<i>gew.</i>	<i>ungew.</i>
Verfahrensgruppe der Präferenzmessung (CBCA versus andere)	.400	.239	.054	.104	.220	.111
Verfahrensgruppe der Präferenzmessung (SE versus andere)	-.223	-.059	-.099	-.081	.017	-.140
Art der abhängigen Variablen (Pearson versus Hitrate bzw. Rangkorrelation)	-.266	-.289	k.A.	k.A.	k.A.	k.A.
Größe der relevanten Stichprobe	.076	.102	.168	.052	-.092	.037
Anzahl der im Untersuchungsdesign verwendeten Attribute (2 bis 25)	.047	.089	-.046	.010	-.295	-.101
Jahr der Publikation	.337	.276	.284	.319	.083	.203
Anzahl der Stimuli in der Hold-out-Aufgabe (2 bis 30)	-.206	-.207	.013	-.045	.139	-.032

Anmerkung: $r_{ges.}$ = Effektgrößen auf der gepoolten Basis der Hitraten-Äquivalente und der Korrelationskoeffizienten, $r_{nurKorr.}$ = Effektgrößen auf der alleinigen Basis der Korrelationskoeffizienten Pearson's r , $r_{nurHitraten}$ = Effektgrößen auf der alleinigen Basis der Hitraten-Äquivalente

Tab. 6.8: Korrelation ausgewählter Faktoren mit den Effektgrößen der Validität

Art der Publikation

<i>Faktor</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95% CI</i>
Journalartikel	378	42,221	.67	.68	.0088	.030	.0026	.164	8.83	.66-.70
Arbeitspapier / Konferenzpapier	91	26,206	.67	.75	.0236	.051	.0007	.224	1.33	.70-.79
Buch / Buchartikel	70	22,524	.66	.75	.0204	.029	.0006	.169	2.05	.71-.79
Dissertation / Diplomarbeit	88	2,900	.73	.75	.0159	.022	.0059	.127	26.86	.72-.78
Sonstige (z. B. Sawtooth Papers)	140	43,336	.84	.82	.0058	.005	.0003	.066	7.28	.81-.83
Total	767	137,187	.71	.75	.0062	.030	.0011	.170	3.57	.74-.76

Anmerkung: Für $p < .05$ sind die Unterschiede der gewichteten Mittelwerte jeweils paarweise signifikant mit Ausnahme des Unterschiedes zwischen Buch/Buchartikel und Dissertation/Diplomarbeit.

Für $p < .05$ sind die Unterschiede der ungewichteten Mittelwerte zwischen Sonstige und den übrigen Publikationsarten jeweils paarweise signifikant.

Tab. 6.9: Durchschnittliche Validität in Abhängigkeit der Publikationsart

Die in wissenschaftlichen Journals publizierten Studien weisen im Durchschnitt eine geringere Validität auf als Studien aus anderen Quellen. Die höchste durchschnittliche Validität weisen Studien auf, die im Internet publiziert worden sind. Der Publication Bias besagt, dass Studien mit signifikanten Ergebnissen mit größerer Wahrscheinlichkeit publiziert werden. Dieser Effekt kann mit den vorliegenden Daten nicht bestätigt werden.

Studentische vs. nicht-studentische Probanden

<i>Faktor</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95% CI</i>
Zielgruppe des Präferenzprodukts	301	94,499	.72	.76	.0104	.032	.0006	.178	1.76	.74-.78
Studenten, die als Zielgruppe gelten können	446	30,842	.69	.70	.0081	.029	.0038	.159	13.17	.68-.72
Total	747	125,341	.70	.75	.0066	.032	.0011	.176	3.56	.74-.76

Anmerkung: Für $p < .05$ sind die Unterschiede der gewichteten Mittelwerte jeweils paarweise signifikant mit Ausnahme des Unterschiedes zwischen Buch/Buchartikel und Dissertation/Diplomarbeit.

Für $p < .05$ sind die Unterschiede der ungewichteten Mittelwerte nicht signifikant.

Tab. 6.10: Durchschnittliche Validität in Abhängigkeit der Probandenart

Der bedeutende Anteil an in der Meta-Analyse berücksichtigten Probanden besteht aus der direkten Zielgruppe des Präferenzprodukts. Studien mit nicht-studentischen Probanden weisen im Mittel eine höhere Validität auf als Studien, in denen Studenten als Probanden zum Einsatz kamen. Die Studien mit Studenten als Probanden behandelten in der Regel Objektbereiche, die für Studenten als Zielgruppe Relevanz besitzen.

Objektbereich der Präferenzmessung

<i>Faktor</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95% CI</i>
Apartments	172	24,891	.72	.79	.0086	.013	.0010	.109	7.48	.78-.81
Autos	42	3,760	.60	.58	.0321	.043	.0050	.195	11.50	.52-.64
Jobs	66	2,696	.58	.58	.0321	.068	.0112	.239	16.45	.51-.64
Nahrungsmittel	32	7,132	.68	.66	.0289	.027	.0015	.159	5.46	.60-.71
Technische Geräte und Zubehör	236	45,477	.73	.78	.0102	.025	.0008	.155	3.29	.76-.80
Sonstiges	207	40,738	.72	.74	.0145	.044	.0010	.206	2.39	.71-.77
Total	755	124,694	.70	.75	.0066	.033	.0012	.178	3.52	.74-.76

Anmerkung: Für $p < .05$ sind die Unterschiede der gewichteten Mittelwerte jeweils paarweise signifikant mit Ausnahme des Unterschiedes zwischen Autos und Jobs.

Tab. 6.11: Durchschnittliche Validität in Abhängigkeit des Objektbereichs der Präferenzmessung

Die in der Meta-Analyse berücksichtigten Studien mit Autos, Jobs oder Nahrungsmitteln als Objektbereich der Präferenzmessung weisen im Mittel eine geringere Validität auf als die berücksichtigten Studien mit Apartments oder technischen Geräten.

ICE- vs. HB-Verfahren

Wie in Kapitel 2.4 beschrieben kommen zur Ermittlung individueller Teilnutzenwerte bei der wahlbasierten Conjoint-Analyse insbesondere zwei Verfahren zum Einsatz: das Individual Choice-Estimation- und das Hierarchical Bayes-Verfahren. Beide Verfahren erreichen hohe Validitätskoeffizienten. Im direkten Vergleich erreicht das HB-Verfahren im Mittel der berücksichtigten Untersuchungen eine signifikant höhere Validität als ICE. Tabelle 6.12 zeigt die Ergebnisse der Meta-Analyse.

<i>Faktor</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95% CI</i>
ICE-Verfahren (Individual Choice Estimation)	20	9,226	.86	.84	.0097	.002	.0002	.041	10.07	.82-.86
HB-Verfahren (Hierarchical Bayes)	43	11,425	.84	.86	.0065	.002	.0003	.039	14.15	.85-.87
Total	63	20,651	.85	.85	.0059	.002	.0002	.044	10.78	.84-.86

Anmerkung: Für $p < .05$ ist der Unterschied der gewichteten Mittelwerte signifikant.

Für $p < .05$ ist der Unterschied der ungewichteten Mittelwerte nicht signifikant.

Tab. 6.12: Durchschnittliche Validität in Abhängigkeit der Art der Individualisierung der Teilnutzenwerte bei der wahlbasierten Conjoint-Analyse

Segmentierungsverfahren

Wie in Kapitel 2.5 beschrieben werden unterschiedliche Verfahren zur Segmentierung der Probanden anhand ihrer Präferenzurteile verwendet. Neben den traditionellen Verfahren der Faktoren- und Clusteranalyse und der a-priori-Segmentierung kommt in den letzten Jahren das Verfahren der Latent-Class-Segmentierung verstärkt zum Einsatz. Im direkten Vergleich erreichen die Verfahren unterschiedlich hohe Validitätsmaße. Tabelle 6.13 gibt die mittleren Validitätskoeffizienten der in der Meta-Analyse verwendeten Untersuchungen an in Abhängigkeit des verwendeten Segmentierungsverfahrens.

<i>Verfahren</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95%CI</i>
Latent-Class-Segmentierung	65	26,703	.82	.84	.0096	.006	.0002	.076	3.61	.82-.86
a-priori-Segmentierung	6	1,090	.82	.80	.0209	.003	.0007	.044	26.63	.76-.84
Faktorenanalyse	20	2,282	.70	.69	.0213	.009	.0025	.081	27.03	.65-.73
Clusteranalyse	23	2,121	.69	.71	.0326	.024	.0027	.147	11.24	.64-.77
Keine Segmentierung	654	99,599	.67	.73	.0073	.035	.0014	.182	4.16	.72-.74
Total	768	131,795	.69	.75	.0063	.030	.0011	.170	3.71	.74-.76

Anmerkung: Für $p < .05$ sind die Mittelwertsunterschiede im ungewichteten Auswertungsmodus zwischen Latent-Class-Segmentierung und Clusteranalyse bzw. keine Segmentierung signifikant. Alle übrigen Mittelwertsunterschiede sind im ungewichteten Auswertungsmodus nicht signifikant.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte jeweils paarweise signifikant.

Tab. 6.13: Durchschnittliche Validität in Abhängigkeit der Verwendung eines Segmentierungsverfahrens

Die Segmentierungsverfahren erreichen im Mittel eine höhere Validität als wenn keine Segmentierungen vorgenommen worden sind. Die Faktorenanalyse weist im Durchschnitt ebenso wie die Clusteranalyse eine signifikant geringere Validität auf als Verfahren, die keine Segmentierung vorgenommen haben. Die höchste Validität weisen in dieser Auswertung Verfahren auf, die eine Latent-Class-Segmentierung vorgenommen haben. Auch a-priori-Segmentierungen weisen im Vergleich eine überdurchschnittliche Validitätshöhe auf.

6.1.3 Ergebnisse der Meta-Analyse für Validitätsmaße auf der aggregierten Ebene

Auf der aggregierten Ebene werden die Validitätsmaße zur Vorhersage von Wahlanteilen meta-analytisch zusammengefasst. Als Validitätsmaße werden die mittlere absolute Abweichung und die Wurzel der mittleren quadrierten Abweichung zwischen vorhergesagten und beobachteten Wahlanteilen gemäß Formel (13) herangezogen. Die Tabellen 6.14 und 6.15 geben die zusammengeführten, mittleren Effektgrößen für die berücksichtigten Verfahrensgruppen, jeweils ungewichtet und an der Stichprobengröße gewichtet, an.

<i>Verfahren</i>	<i>k</i>	<i>N</i>	MAE_u	MAE_g	$\hat{\sigma}_{MAE}$	S_{MAE}^2	95% CI
CA	66	8,611	5.33	5.15	.4185	11.56	4.33-5.97
ACA	48	17,846	9.23	7.29	.3833	7.05	6.54-8.05
SE	14	1,454	12.80	11.27	2.081	60.65	7.19-15.35
CBCA	110	26,146	5.17	4.75	.1918	4.05	4.37-5.12
HYB	10	1,588	6.67	5.46	.6231	3.88	4.24-6.68
Total	248	55,645	6.49	5.82	.1981	9.73	5.43-6.21

Anmerkung: MAE_u = mittlere absolute Abweichung (ungewichtetes Mittel), MAE_g = mittlere absolute Abweichung (gewichtetes Mittel).

Für $p < .05$ unterscheiden sich jeweils paarweise signifikant (mit Bonferroni-Adjustierung): CBCA/CA, CBCA/SE, CBCA/HYB, ACA/CA, ACA/SE.

Für $p < .05$ unterscheiden sich alle gewichteten Mittelwerte jeweils paarweise signifikant (mit Bonferroni-Adjustierung).

Tab. 6.14: Ergebnisse der Meta-Analyse für die Effektgröße der mittleren absoluten Abweichung (MAE) zwischen vorhergesagten und beobachteten Wahlanteilen

Sowohl in der gewichteten als auch der ungewichteten Meta-Analyse weisen wahlbasierte Conjoint-Analysen die geringste mittlere Abweichung bei der Wahlanteilsvorhersage auf, gefolgt von der traditionellen Conjoint Analyse und anderen hybriden Verfahren. Die Gruppe der direkten Nutzenmessung weist im Mittel eine signifikant geringere Validität auf.

<i>Verfahren</i>	<i>k</i>	<i>N</i>	$RMSE_u$	$RMSE_g$	$\hat{\sigma}_{RMSE}$	S^2_{RMSE}	95% <i>CI</i>
CA	30	5,138	.13	.13	.0163	.008	.10-.16
ACA	4	360	.12	.12	.0065	.000	.10-.13
SE	3	319	.12	.12	.0100	.000	.10-.14
CBCA	10	1,840	.05	.05	.0014	.000	.05-.05
HYB	4	514	.08	.08	.0112	.001	.06-.10
Total	51	8,171	.11	.11	.0110	.006	.09-.13

Anmerkung: $RMSE_u$ = Quadratwurzel der mittleren quadrierten Abweichung (ungewichtetes Mittel),

$RMSE_g$ = Quadratwurzel der mittleren quadrierten Abweichung (gewichtetes Mittel).

Für $p < .05$ ist der Mittelwertsunterschied im ungewichteten Auswertungsmodus zwischen CA/CBCA signifikant.

Die übrigen ungewichteten Mittelwerte unterscheiden sich jeweils paarweise nicht signifikant.

Für $p < .05$ unterscheiden sich alle gewichteten Mittelwerte jeweils paarweise signifikant (mit Bonferroni-Adjustierung) mit Ausnahme von ACA/SE.

Tab. 6.15: Ergebnisse der Meta-Analyse für die Effektgröße der Wurzel der mittleren quadrierten Abweichung (RMSE) zwischen vorhergesagten und beobachteten Wahlanteilen

Sowohl in der gewichteten als auch der ungewichteten Meta-Analyse weisen wahlbasierte Conjoint-Analysen den geringsten RMSE bei der Wahlanteilsvorhersage auf, gefolgt von hybriden Verfahren. Die Gruppen der traditionellen Conjoint-Analyse, der adaptiven Conjoint-Analyse und der direkten Nutzenmessung weisen eine unterdurchschnittliche Validität in dem Kriterium der Wurzel der mittleren quadrierten Abweichung auf.

Wie in Kapitel 2.5 beschrieben worden ist, werden Simulatoren verwendet, um auf der Basis individueller Nutzenwerte zu Wahl-/Marktanteilsschätzungen zu gelangen. Zahlreiche Studien, die MAE- oder RMSE-Effektmaße angegeben hatten, enthielten keine Angabe über die Art des verwendeten Wahlanteilssimulator. Für die übrigen wurde eine separate Auswertung vorgenommen, welche Simulatoren im Mittel zu geringeren Vorhersagefehlern führten. Die Ergebnisse sind in der folgenden Tabelle dargestellt.

<i>Verfahren</i>	<i>1stChoice</i>			<i>BTL</i>			<i>Logit</i>		
	<i>N</i>	<i>MAE_g</i>	<i>95%CI</i>	<i>N</i>	<i>MAE_g</i>	<i>95%CI</i>	<i>N</i>	<i>MAE_g</i>	<i>95%CI</i>
CA	5,980	4.06	3.35-4.77	384	13.60	6.60-20.60	2,247	6.60	5.73-7.48
ACA	5,851	7.84	6.78-8.90	191	21.73	13.02-30.43	2,227	6.89	4.63-9.15
SE	396	12.93	11.11-14.75	392	15.14	6.07-24.21	585	7.41	0.0-16.39
CBCA	5,724	5.91	5.01-6.81	0	k.A.	k.A.	11,012	3,83	3.46-4.19
HYB	346	8.35	7.65-9.05	272	7.10	6.56-7.65	970	3.96	3.56-4.37
Total	18,297	6.12	5.55-6.69	1,239	13.91	9.35-18.48	17,041	4.72	4.12-5.33

Anmerkung: Für $p < .05$ sind die Unterschiede zwischen den mittleren absoluten Abweichungen der drei Wahlanteilssimulatoren jeweils paarweise signifikant (mit Bonferroni-Adjustierung).

Tab. 6.16: Ergebnisse der Meta-Analyse für die Effektgröße der mittleren absoluten Abweichung (MAE) in Abhängigkeit des Verfahrens zur Wahlanteilssimulation

Die First-Choice-Regel wurde bei den in der Meta-Analyse berücksichtigten Studien am häufigsten verwendet. Sie erreicht über alle Verfahren hinweg betrachtet eine mittlere absolute Abweichung bei Wahlanteilsvorhersagen von 6,5 Prozentpunkten. Mit 4,3 absoluten Prozentpunkten weisen Logit-Verfahren den geringsten MAE auf. In dieser Auswertung werden mit der BTL-Regel die größten absoluten Abweichungen erzielt. Die berücksichtigte Probandenanzahl ist deutlich geringer als bei den beiden anderen Wahlanteilssimulatoren.

Die wahlbasierten Conjoint-Analysen erreichen sowohl in den Validitätsmaßen auf individueller Ebene als auch auf der Ebene der Vorhersage von Wahlanteilen die besten Ergebnisse. Auf der anderen Seite weist die Gruppe der direkten Präferenzmessverfahren durchweg unterdurchschnittliche Validitätsmaße auf. ACA zeigt bei den Wahlanteilsvorhersagen unterdurchschnittliche Validitätsmaße, erreicht bei den individuellen Vorhersagen jedoch mehrheitlich überdurchschnittliche Ergebnisse. Die traditionelle Conjoint-Analyse und die hybriden Verfahren weisen im Kriterium des MAE eine gute Vorhersagegüte von Wahlanteilen auf und erreichen in den übrigen Validitätsmaßen durchschnittliche Ausprägungen.

6.2 Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität

Die Rangkorrelationskoeffizienten werden mit Hilfe der Formeln (36) und (37) in Äquivalente zum Produkt-Moment-Korrelationskoeffizienten transformiert und in eine gemeinsame Meta-Analyse integriert. Die gepoolten Ergebnisse für Effektgrößen der Reliabilität sind in Tabelle 6.17 dargestellt.

<i>Verfahren</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95% CI</i>
CA	128	5,181	.62	.66	.0172	.038	.0081	.172	21.44	.63-.69
ACA ^a	23	1,501	.85	.78	.0365	.031	.0025	.168	8.09	.70-.85
SE	11	519	.83	.80	.0456	.023	.0028	.142	12.05	.71-.89
CBCA	1	27	.83	.83	.0000	.000	.0037	k.A.	k.A.	.83-.83
HYB	3	525	.78	.84	.0900	.024	.0005	.154	2.09	.66-.99
Total	166	7,753	.67	.70	.0153	.039	.0056	.182	14.42	.67-.73

Anmerkung: Für $p < .05$ sind die Mittelwertsunterschiede zwischen CA/ACA und CA/SE im ungewichteten Auswertungsmodus signifikant. Die übrigen ungewichteten Mittelwerte unterscheiden sich nicht signifikant.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte jeweils paarweise signifikant mit Ausnahme der Differenzen zwischen CBCA/SE, CBCA/HYB, CBCA/ACA und ACA/SE.

^a In der Gruppe von ACA ist eine Studie enthalten, die mit 7.813 Probanden 87 % der ACA-Stichprobe ausmacht. Diese Studie kann auch im Hinblick auf den mit .97 extremen Reliabilitätskoeffizienten als Ausreißer gelten. Diese Studie wurde in dieser und den folgenden Berechnungen berücksichtigt mit einer Stichprobengröße von 307, was der größten Stichprobe der übrigen Studien entspricht.

Tab. 6.17: Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität

Die Reliabilität der traditionellen Conjoint-Analyse ist im Durchschnitt der in der Meta-Analyse berücksichtigten Studien signifikant niedriger als die Reliabilität der übrigen Verfahrensgruppen. Über alle Verfahrensgruppen der Präferenzmessung hinweg betrachtet liegt die Reliabilität mit $r = .67$ im ungewichteten bzw. $r = .70$ im gewichteten Auswertungsmodus nur gering über den in Tabelle 6.6 dargestellten durchschnittlichen Validitätskoeffizienten in Höhe von $r = .61$ bzw. $r = .62$.

In der Folge wird unterschieden, ob sich die Reliabilitätsmaße aus der Messung auf der Inputebene oder der Outputebene ergeben haben. Gemäß der in Tabelle 3.1 dargestellten Kategorisierung der Reliabilitätsmaße werden die Zusammenhangsmaße in einer Meta-Analyse ausgewertet, welche die Übereinstimmung zwischen Input-Rangordnungen bzw. Input-Ratings zwischen zwei Messzeitpunkten ausdrücken. In einer zweiten Meta-Analyse werden die Reliabilitätsmaße zusammengefasst, welche den Zusammenhang zwischen ermittelten Outputdaten, also Teilnutzenwerten oder Attributwichtigkeiten, repräsentieren. Die Tabellen 6.18 und 6.19 zeigen die Ergebnisse der separaten Meta-Analysen zur Reliabilität auf Input- bzw. Outputvektorebene.

<i>Verfahren</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95% CI</i>
CA	57	2,789	.65	.63	.0184	.019	.0077	.107	40.17	.59-.66
ACA	4	599	.63	.58	.0477	.009	.0030	.078	32.50	.49-.67
SE	4	186	.74	.74	.0827	.027	.0045	.151	16.50	.58-.90
CBCA	0	0	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.
HYB	1	101	.52	.52	.0000	.000	.0053	k.A.	k.A.	.52-.52
Total	66	3,675	.65	.62	.0168	.019	.0069	.109	36.83	.59-.65

Anmerkung: Für $p < .05$ sind die Unterschiede der ungewichteten Mittelwerte nicht signifikant.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte jeweils paarweise signifikant (Bonferroni-adjustiert).

Tab. 6.18: Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität auf der Input-Ebene

Die Reliabilität der direkten Nutzenmessung ist auf der Input-Ebene die höchste. Die durchschnittliche Reliabilität der traditionellen Conjoint-Analyse und der adaptiven Conjoint-Analyse ist auf der Input-Ebene demgegenüber signifikant geringer.

Über alle Verfahren hinweg betrachtet ist die durchschnittliche Reliabilität auf der Input-Ebene geringer als auf der Output-Ebene, wie ein Vergleich der in den Tabellen 6.18 und 6.19 dargestellten Ergebnisse zeigt.

<i>Verfahren</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95%CI</i>
CA	71	2,392	.60	.70	.0283	.057	.0081	.221	14.20	.64-.75
ACA	19	902	.89	.90	.0121	.003	.0007	.045	25.46	.88-.93
SE	7	333	.89	.84	.0494	.017	.0019	.123	11.31	.74-.93
CBCA	1	27	.83	.83	.0000	.000	.0037	k.A.	k.A.	.83-.83
HYB	2	424	.91	.91	.0087	.000	.0001	.005	84.30	.90-.93
Total	100	4,078	.68	.78	.0212	.045	.0039	.202	8.71	.74-.82

Anmerkung: Für $p < .05$ sind die Mittelwertsunterschiede im ungewichteten Auswertungsmodus jeweils paarweise nicht signifikant mit Ausnahme der Unterschiede zwischen den Mittelwerten CA/ACA und CA/SE.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte jeweils paarweise signifikant (Bonferroni-adjustiert) mit Ausnahme der Unterschiede zwischen den Mittelwerten CBCA/ACA, CBCA/SE, CBCA/HYB und ACA/HYB.

Tab. 6.19: Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität auf der Output-Ebene

Die Gruppe der traditionellen Conjoint-Analysen weist signifikant niedrigere Reliabilitätskoeffizienten auf der Output-Ebene auf als Verfahren der übrigen Gruppen der Präferenzmessung. Die höchste Reliabilität weist die Gruppe der hybriden bzw. der adaptiven Conjoint-Analysen auf. Auch die Verfahren der direkten Nutzenmessung weisen im Durchschnitt der in der Meta-Analyse berücksichtigten Studien ausgeprägte Reliabilitätskoeffizienten auf.

Die Übereinstimmung zwischen ermittelten Teilnutzenwerten bzw. Attributwichtigkeiten bei zwei Messungen ist nach den vorliegenden Daten ausgeprägter als die Übereinstimmung zwischen zwei Erhebungen auf der Inputebene, also die Abgabe von Präferenzurteilen anhand von Rangordnungsaufgaben oder Einstufungsskalen. Der Einfluss des Faktors, ob das Übereinstimmungsmaß jeweils auf der Input- oder der Outputebene bestimmt wird, auf die absolute Höhe des Reliabilitätskoeffizienten wird näher untersucht. Die Ergebnisse sind in Tabelle 6.20 dargestellt.

<i>Bestimmungsart</i>	<i>k</i>	<i>N</i>	\bar{r}_u	\bar{r}_g	$\hat{\sigma}_{\bar{r}}$	S_r^2	$\hat{\sigma}_e^2$	$\hat{\sigma}_\rho$	<i>V%</i>	<i>95%CI</i>
Korrelation von Rangordnungsdaten	27	1,838	.67	.60	.0258	.018	.0062	.108	34.43	.55-.65
Korrelation von Ratingdaten	39	1,837	.64	.65	.0217	.018	.0074	.105	40.22	.60-.69
Korrelation von Teilnutzenwerten	90	3,773	.67	.77	.0227	.046	.0040	.206	8.68	.73-.82
Korrelation von Attributwichtigkeiten	10	305	.81	.87	.0437	.019	.0021	.130	10.89	.78-.95
Total	166	7,753	.67	.70	.0153	.039	.0056	.182	14.42	.67-.73

Anmerkung: Für $p < .05$ unterscheiden sich die ungewichteten Mittelwerte jeweils paarweise nicht signifikant.

Für $p < .05$ unterscheiden sich die gewichteten Mittelwerte jeweils paarweise signifikant.

Tab. 6.20: Ergebnisse der Meta-Analyse für Effektgrößen der Reliabilität in Abhängigkeit der Art der Reliabilitätsmessung

Die Art der Reliabilitätsmessung ist ein signifikanter Einflussfaktor auf die Höhe des Reliabilitätsmaßes. Die höchste Reliabilität wird von Studien angegeben, welche die Zuverlässigkeit des Verfahrens durch die Korrelation zwischen bei zwei Messungen ermittelten Attributwichtigkeiten ausdrücken. Signifikant niedriger fallen die Reliabilitätsmaße aus, wenn sie auf der Basis der Übereinstimmung von Input-Daten der Probanden berechnet werden.

Offensichtlich sind Verfahren der Präferenzmessung in gutem Maße in der Lage, die Teilnutzenstruktur von Probanden zuverlässig zu bestimmen. Weniger reliabel ist hingegen die Erhebung der Präferenzdaten auf der Inputebene. Die geringere Zuverlässigkeit der Datenerhebung selbst zeigt sich auch in der wiederholt berichteten eingeschränkten Konsistenz von Probanden bei der Abgabe von Präferenzurteilen zu identischen, wiederholt vorgelegten Stimuli (z. B. Huber et al. 1993; Orme & King, 1998).

In folgenden Untersuchungen sollten Angaben über die Reliabilität der Messung häufiger aufgenommen werden. Bei Angaben zur Reliabilität sollte erläutert werden, auf welche Art der Reliabilitätsbestimmung der Koeffizient ermittelt worden ist. Für Korrelationen von Teilnutzenwerten können nach den Ergebnissen der vorliegenden Meta-Analyse höhere Reliabilitätskoeffizienten erwartet werden als für Korrelationen auf der Ebene der Input-Daten.

7. Zusammenfassende Diskussion und Ausblick

Im ersten Teil dieses abschließenden Kapitels werden die Ergebnisse der Meta-Analyse zusammengefasst. Im zweiten Teil dieses Kapitels werden die Ergebnisse der Meta-Analyse diskutiert und Implikationen für die praktische Anwendung von Verfahren der Präferenzmessung formuliert.

7.1 Zusammenfassung der Ergebnisse der Meta-Analyse

Die Zielsetzung der vorliegenden Meta-Analyse bestand in der Synthese der Forschungsergebnisse der vergangenen 35 Jahre zur Reliabilität und Validität von Präferenzmessverfahren. Es wurden die folgenden Gruppen an Präferenzmessverfahren in der Meta-Analyse berücksichtigt: traditionelle Conjoint-Analyse mit Rangordnungs- oder einstufigsskalenbasierten Aufgaben, die wahlbasierte Conjoint-Analyse, Verfahren der kompositionellen Präferenzmessung, die adaptive Conjoint-Analyse und sonstige hybride Verfahren.

Insgesamt wurden 135 Untersuchungen in der Meta-Analyse berücksichtigt. Aus den verwendeten Untersuchungen konnten 1.247 Effektmaße codiert werden. In der Summe basieren die Ergebnisse der Meta-Analyse auf einer gepoolten Stichprobe von 135.000 Probanden.

7.1.1 Zusammenfassung der Ergebnisse der Meta-Analyse zur Reliabilität

Die folgenden in der Praxis üblichen Indikatoren für die Bestimmung der Reliabilität von Präferenzmessverfahren wurden in der Meta-Analyse berücksichtigt:

- 1.) Übereinstimmung von Präferenzurteilen zu identischen Rangordnungsaufgaben, ausgedrückt durch ein Zusammenhangsmaß Spearman's Rho oder Kendall's Tau
- 2.) Übereinstimmung von Präferenzurteilen zu identischen einstufigsskalenbasierten Aufgaben, ausgedrückt durch das Zusammenhangsmaß des Produkt-Moment-Korrelationskoeffizienten

- 3.) Übereinstimmung von durch zwei (Teil-)Messungen ermittelten Teilnutzenwerten, ausgedrückt durch das Zusammenhangsmaß des Produkt-Moment-Korrelationskoeffizienten
- 4.) Übereinstimmung von durch zwei (Teil-)Messungen ermittelten Attributwichtigkeiten, ausgedrückt durch das Zusammenhangsmaß des Produkt-Moment-Korrelationskoeffizienten oder Spearman's Rho oder Kendall's Tau

Die ersten beiden Formen der Reliabilitätsbestimmung werden als Reliabilitätsbestimmung auf der Ebene des Input-Vektors zusammengefasst. Die anderen beiden Formen können zusammenfassend als Reliabilitätsbestimmung auf der Ebene des Output-Vektors bezeichnet werden.

Die Ergebnisse der Meta-Analyse zeigen über alle Verfahrensgruppen der Präferenzmessung hinweg eine Reliabilität in Höhe von $r = .70$. Die Verfahrensgruppe der adaptiven Conjoint-Analyse erreicht mit $r = .78$ eine überdurchschnittliche Reliabilität. Auch die sonstigen hybriden Verfahren weisen mit $r = .84$ eine höhere Reliabilität auf als die übrigen Verfahren. Die Verfahren der direkten Nutzenmessung erreichen mit $r = .80$ eine signifikant höhere Reliabilität als die Gruppe der traditionellen Conjoint-Analyse mit $r = .66$.

In der Einzelbetrachtung zeigt sich, dass Untersuchungen, welche die Reliabilität auf der Ebene des Input-Vektors erhoben haben, eine signifikant geringere Reliabilität berichten als Untersuchungen, welche die Reliabilität auf der Ebene des Output-Vektors erhoben haben. Die Übereinstimmung zwischen Probandenantworten zu identischen Rangordnungs- bzw. einstufigsskalenbasierten Aufgaben betrug im Mittel $r = .62$. Die Übereinstimmung zwischen ermittelten Teilnutzenwerten oder Attributwichtigkeiten betrug demgegenüber im Mittel $r = .78$.

7.1.2 Zusammenfassung der Ergebnisse der Meta-Analyse zur Validität

Unter den zahlreichen Varianten zur Bestimmung der Validität von Präferenzmessverfahren wurden Kennziffern der kriterienorientierten Validität ausgewählt. Als Kriterien zur Validierung der Präferenzmessergebnisse wurden Vorhersagen individueller Präferenzurteile oder Wahlentscheidungen in der Meta-Analyse ebenso berücksichtigt wie Validitätsmaße zur Darstellung der Abweichungen bei der Vorhersage von Wahl-/Marktanteilen.

Auf der individuellen Ebene wurden die folgenden Validitätskriterien in der Meta-Analyse berücksichtigt:

- 1.) Übereinstimmung von vorhergesagten und beobachteten Wahlentscheidungen in einem oder mehreren Hold-out-Wahlsets, ausgedrückt durch die durchschnittliche Trefferquote über alle Probanden und Wahlen hinweg (Hitrate)
- 2.) Übereinstimmung von vorhergesagten und beobachteten Präferenzurteilen bei der Rangordnung von Alternativen in einem oder mehreren Hold-out-Stapeln, ausgedrückt durch den Rangkorrelationskoeffizienten Spearman's Rho oder durch Kendall's Tau
- 3.) Übereinstimmung von vorhergesagten und beobachteten Präferenzbeurteilungen von Alternativen in einem oder mehreren Hold-out-Stapeln anhand von Einstufungsskalen, ausgedrückt durch den Produkt-Moment-Korrelationskoeffizienten.

Unter Verwendung von Transformationsformeln konnten alle Validitätsmaße in ein zum Produkt-Moment-Korrelationskoeffizienten äquivalentes Effektmaß überführt und damit einer gemeinsamen statistischen Analyse zugänglich gemacht werden.

Im Durchschnitt erreichen die Präferenzmessverfahren eine Güte bei der Vorhersage der benannten Kriterien in Höhe von $r = .75$. Die signifikant höchsten Validitätsmaße erreicht die Gruppe der wahlbasierten Conjoint-Analyse mit $r = .84$ gefolgt von der adaptiven Conjoint-Analyse mit $r = .75$. Mittlere Ausprägungen des integrierten Effektmaßes der Validität weist die traditionelle Conjoint-Analyse mit $r = .68$ auf. Für die Gruppe der sonstigen hybriden Verfahren ergab sich ebenso wie für die Verfahren der direkten, kompositionellen Nutzenmessung eine signifikant geringere Validität in Höhe von $r = .63$ bzw. $r = .62$.

Die durchgeführten Homogenitätstests lassen auf eine heterogene Zusammensetzung der Studienstichprobe schließen. Im Kontext der Präferenzmessmethoden lässt sich die beobachtete Varianz in den berücksichtigten Koeffizienten ohne Berücksichtigung von Moderatoren nur eingeschränkt erklären. In der Einzelbetrachtung wird deutlich, dass neben der Gruppe des verwendeten Verfahrens der Präferenzmessung auch die Wahl des Validierungskriteriums einen signifikanten Einfluss auf die Höhe der Validität aufweist.

In Abhängigkeit, ob das Validitätskriterium als Hitrate oder als Korrelationskoeffizient angegeben worden war, ergeben sich unterschiedliche Validitätshöhen. Besonders ausgeprägt sind die korrelationsäquivalenten Maße, wenn sie auf der Basis von Hitraten ermittelt wurden, die wiederum aus Hold-out-Aufgaben mit vielen Wahlalternativen stammten. Eher geringere Validitätsmaße resultierten bei der Messung der Übereinstimmung zwischen vorhergesagten und beobachteten Rangordnungen oder Präferenzbeurteilungen anhand von Einstufungsskalen mittels Korrelationskoeffizienten.

Weitere Einflussfaktoren auf die Höhe der Validität wurden wie folgt identifiziert:

- 1.) Untersuchungen mit Studenten als Probanden weisen mit $r = .70$ eine signifikant geringere Validität auf als Studien mit der unmittelbaren Zielgruppe des Präferenzobjekts ($r = .77$). Eine Überschätzung der mit dieser Meta-Analyse ermittelten Reliabilitäts- bzw. Validitätsmaße aufgrund der Integration von Studien mit studentischen Probanden scheint nicht vorzuliegen.
- 2.) Aus wissenschaftlichen Journals entnommene Untersuchungen weisen eine signifikant geringere Validität auf als Untersuchungen, die aus anderen Quellen stammten. Gemäß des Publication Bias wäre ein entgegengesetzter Effekt zu erwarten gewesen.
- 3.) Der Objektbereich der Apartments wurde in den in der Meta-Analyse berücksichtigten Untersuchungen am häufigsten verwendet. Studien mit Apartments und mit technischen Geräten als Objektbereich weisen eine signifikant höhere durchschnittliche Validität auf als Studien mit Jobs, Autos oder Nahrungsmitteln.
- 4.) Innerhalb der Gruppe der wahlbasierten Conjoint-Analyse weisen Studien, welche das HB-Verfahren (Hierarchical Bayes) zur Individualisierung der Teilnutzenwerte verwendet haben, eine signifikant höhere Validität auf als Studien mit ICE-Verfahren (Individual Choice-Estimation). Beide Verfahren erreichen mit $r = .85$ bzw. $r = .83$ überdurchschnittliche Validitätsmaße.
- 5.) Bei den Segmentierungsmethoden erreicht das Verfahren der Latent-Class-Segmentierung mit $r = .84$ im Durchschnitt die ausgeprägteste Validität. Die traditionellen Methoden der Faktorenanalyse bzw. der Clusteranalyse weisen im Mittel der berücksichtigten Untersuchungen mit $r = .69$ bzw. $r = .71$ geringere Validitätshöhen auf als Studien, die keine Segmentierung vorgenommen haben.

Neben der Güte der Vorhersage von individuellen Präferenzurteilen wurde auch die Güte der Vorhersage von aggregierten Wahlen im Sinne von Marktanteilsschätzungen als Validitätskriterium untersucht. Als Effektmaße der Validität wurden in der Meta-Analyse die mittlere absolute Abweichung (MAE) und die Wurzel der mittleren quadrierten Abweichung (RMSE) zwischen vorhergesagten und beobachteten Markt-/Wahlanteilen herangezogen.

Im Mittel aller berücksichtigten Untersuchungen können Methoden der Präferenzmessung einzelne Markt-/Wahlanteile mit einer Abweichung von 5,8 absoluten Prozentpunkten korrekt vorhersagen. Eine überdurchschnittliche Prognosegüte weisen die in der Meta-Analyse berücksichtigten Verfahren der wahlbasierten Conjoint-Analyse mit $MAE = 4,7$, die traditionelle Conjoint-Analyse mit $MAE = 5,1$ und sonstige hybride Verfahren mit $MAE = 5,5$ auf. Eine geringere Prognosegüte weisen im Mittel die Verfahrensgruppen der adaptiven Conjoint-Analyse mit $MAE = 7,3$ und der direkten Nutzenmessung mit $MAE = 11,3$ auf.

Im Kriterium der Wurzel der mittleren quadrierten Abweichung weisen wahlbasierte Conjoint-Analysen die beste Vorhersagegüte auf mit $RMSE = 0,05$. Mittlere Ausprägungen weisen die hybriden Verfahren mit $RMSE = 0,08$ und die adaptive Conjoint-Analyse bzw. die Verfahren der direkten Nutzenmessung mit jeweils $RMSE = .12$ auf. Die traditionelle Conjoint-Analyse weist im Kriterium des mittleren quadrierten Abweichungsfehlers die geringste Validität auf mit $RMSE = .13$.

Zur Ermittlung von Markt-/Wahlanteilsschätzungen müssen Anteilssimulatoren auf der Basis der individuellen Teilnutzenwerte Annahmen über die realen Auswahlentscheidungen treffen. Eine Detailuntersuchung der von den in der Meta-Analyse berücksichtigten Studien verwendeten Wahlanteilssimulatoren zeigte, dass die First-Choice-Regel am häufigsten eingesetzt worden war. Die höchste Prognosegüte weist mit einer mittleren absoluten Abweichung von $MAE = 4,3$ das Logit-Verfahren auf. Die First-Choice-Regel resultierte im Mittel in einen Vorhersagefehler von $MAE = 6,5$. Die BTL-Regel wurde seltener verwendet und weist im Mittel einen $MAE = 13,8$ auf.

7.2 Diskussion der Ergebnisse der Meta-Analyse

Die Literatursuche zur Meta-Analyse erfolgte über einen Zeitraum von mehr als drei Jahren. Es wurden 135 Untersuchungen in die Meta-Analyse aufgenommen, von denen 55 % aus anderen Publikationsquellen als wissenschaftlichen Zeitschriften stammen. Trotz einer sorgfältigen Literatursuche und -analyse werden nicht alle im deutsch- und englischsprachigen Raum vorhandenen Untersuchungen zur Messgüte von Präferenzmessverfahren erfolgreich identifiziert worden sein. Angesichts einer Gesamtstichprobe von ca. 135.000 Probanden erscheint es aber unwahrscheinlich, dass einzelne noch nicht berücksichtigte Untersuchungen die Ergebnisse der Meta-Analyse grundlegend verändern würden.

Über die unterschiedlichen Messgütekriterien hinweg betrachtet erreicht die Gruppe der wahlbasierten Conjoint-Analyse die besten Ergebnisse. Sowohl bei der individuellen Vorhersage von Präferenzurteilen bzw. Wahlentscheidungen als auch insbesondere auf der Ebene der Vorhersage von aggregierten Wahlanteilen weisen die wahlbasierten Conjoint-Analysen im Mittel der berücksichtigten Studien die jeweils höchsten Validitätsmaße auf.

Die adaptive Conjoint-Analyse (ACA) erreicht in den Validitätskriterien auf der individuellen Ebene im Mittel eine höhere Validität als die traditionelle Conjoint-Analyse oder sonstige hybride Verfahren. Bei der Vorhersage von Wahlanteilen weist ACA eine überdurchschnittliche absolute Abweichung und die traditionelle Conjoint-Analyse die höchste mittlere quadrierte Abweichung auf. Beide Verfahren erreichen bei den individuellen Validitätskriterien bessere Ergebnisse als bei den Wahlanteilsvorhersagen.

Die Verfahren der kompositionellen Nutzenmessung weisen in nahezu allen Validitätskriterien unterdurchschnittliche und signifikant geringere Ergebnisse auf als die übrigen in der Meta-Analyse berücksichtigten Verfahren. Lediglich bei der Synthese der Hitraten als Validitätsmaß weisen die self-explicated-Verfahren eine überdurchschnittlich hohe Validität auf. Dieses Ergebnis ist allerdings maßgeblich durch die im Vergleich zu den anderen Verfahrensgruppen höhere Anzahl an Alternativen in den Hold-out-Wahlsets zu erklären. Werden lediglich Studien berücksichtigt, die nicht mehr als zehn Alternativen im Hold-out-Wahlset verwendet haben, weisen Verfahren der direkten Nutzenmessung auch im Kriterium der zu den Hitraten äquivalenten Korrelationskoeffizienten eine unterdurchschnittliche Validität auf.

Im Verlauf der etwa 35-jährigen Forschungsgeschichte der Präferenzmessung wurden diverse Vorschläge zur Verbesserung der Validität der Präferenzmessverfahren unterbreitet. Es wurden dabei ausgehend von den direkten Methoden der Nutzenmessung und dem klassischen Conjoint-Measurement über hybride und adaptive Messmethoden bis zur individualisierten wahlbasierten Conjoint-Analyse auch grundlegend neue Verfahren eingeführt. In der Meta-Analyse haben die zeitlich neueren Verfahren signifikant besser abgeschnitten als die traditionelle Conjoint-Analyse oder die Verfahren der direkten Nutzenmessung. Gemäß der Ergebnisse der Meta-Analyse hat die Forschung der vergangenen Jahre zu einer Verbesserung der Vorhersagegüte der Präferenzmessverfahren beigetragen.

Die Reliabilitätsergebnisse sind zum überwiegenden Teil überraschend gering. Das ungewichtete Gesamtmittel liegt beispielsweise bei $r = .67$, also niedriger als die durchschnittliche ermittelte Validität in Höhe von $r = .75$. Eine Erklärung für das höhere Validitätsmaß kann in der integrierten Berücksichtigung der Hitraten gesehen werden. Die Hitraten weisen im Durchschnitt eine Validität von $r = .83$ auf gegenüber dem auf der Basis der Korrelationen ermittelten Validitätsmaß von $r = .62$.

Die geringe Ausprägung der Reliabilität stellt sich besonders gravierend dar, wenn nur die Untersuchungen zusammengefasst werden, welche die Übereinstimmung von Präferenzurteilen auf der Input-Ebene angegeben haben. Auf der Ebene der Input-Daten ergibt sich eine mittlere Übereinstimmung in Höhe von $r = .62$. Die geringe Reliabilität steht in Einklang mit Untersuchungen, die identische Wahlaufgaben zu Beginn und zum Abschluss der Untersuchung vorlegten, und dabei eine erstaunlich häufige Veränderung der Wahlentscheidungen festgestellt haben (z. B. Huber et al., 1993). Die geringe Konsistenz der Probanden auf der Input-Ebene limitiert die Vorhersagegüte eines jeden Verfahrens der Präferenzmessung.

Um so erstaunlicher stellt sich vor diesem Hintergrund die Reliabilität in Höhe von $r = .79$ auf der Ebene des Output-Vektors dar. Offensichtlich sind die meisten Verfahren der Präferenzmessung gut in der Lage, die Teilnutzenstruktur von Probanden zuverlässig zu messen, auch wenn Probanden zu einzelnen Items nicht konsistent antworten.

Die Verallgemeinerung der Validitätsergebnisse der Meta-Analyse ist dadurch gefährdet, dass der überwiegende Teil der Effektmaße auf der Ermittlung der Vorhersagegüte mittels Hold-out-Aufgaben beruht. Lediglich 13 % der in der Meta-Analyse berücksichtigten Probanden wurden mit realen Produkten im Rahmen der Validitätsüberprüfung konfrontiert, die übrigen Probanden erhielten jeweils Produktbeschreibungen vorgelegt.

Einige Autoren bezeichnen die Messgütebestimmung mittels Hold-out-Aufgabe als eine Überprüfung der Reliabilität des Messverfahrens (z. B. Louviere, 1988). In diesem Sinne wäre eine ausgeprägte Hold-out-Validität eine notwendige, aber nicht hinreichende Bedingung für eine gute Vorhersage realer Auswahlentscheidungen oder Marktanteilen.

Dahan et al. (2001) haben innerhalb einer Untersuchung sowohl Hold-out-Aufgaben als auch reale Wahlentscheidungen als Validitätskriterien verwendet. Die relative Vorhersagegüte der untersuchten Verfahren der Präferenzmessung war bei beiden Validitätsüberprüfungen identisch. Die Ergebnisse von Dahan et al. lassen vermuten, dass die Rangordnung der Präferenzmessverfahren, wie sie sich in der Meta-Analyse dargestellt hat, auch bei realen Wahlentscheidungen Bestand haben würde. Eine Meta-Analyse auf der Basis von Validitätsstudien mit einem externen Kriterium könnte die These der identischen relativen Vorhersagegüte der Verfahrensgruppen bestätigen. Zum aktuellen Zeitpunkt liegen jedoch noch zu wenige Untersuchungen mit einem externen Validierungskriterium vor, um zu sinnvollen Generalisierungen zu gelangen.

Wie in Kapitel 2.1 beschrieben, unterscheiden sich die Konstrukte der Präferenz und der Entscheidung. Analog zu Fragen der Einstellungs-Verhaltens-Forschung ist ein enger Zusammenhang zwischen der Präferenz und einer tatsächlichen Wahlentscheidung nur unter bestimmten Ähnlichkeitsbedingungen zu erwarten. Zudem hängt eine reale Kaufentscheidung von zahlreichen Zusatzfaktoren wie beispielsweise der Verfügbarkeit und Präsenz eines Produkts, besonderer Anstrengungen am Point-of-Sale oder weiterer sozialer Einflüsse ab.

Auch wenn zahlreiche Faktoren eine reale Wahlentscheidung mit beeinflussen, weisen die Ergebnisse der Meta-Analyse darauf hin, dass mit der Kenntnis der Präferenzstruktur von Probanden eine gute Vorhersage individueller Wahlentscheidungen getroffen werden kann. Beispielsweise beträgt die durchschnittliche Hitrate bei Vorhersagen von Entscheidungen zwischen zwei Alternativen 78 %, eine signifikante Verbesserung gegenüber der zufällig zu erwartenden Hitrate von 50 %. Die Meta-Analyse unterstreicht über die zum Teil

anekdotenhaften Einzelberichte hinaus, dass Präferenzmessverfahren, und hier insbesondere die modernen Methoden der adaptiven bzw. wahlbasierten Conjoint-Analyse, ein valides Instrument zur Vorhersage von Wahlentscheidungen darstellen.

Die Deutlichkeit der Unterschiede in den ermittelten Reliabilitäts- bzw. Validitätskennziffern weist darauf hin, dass die in der durchgeführten Meta-Analyse festgestellte relative Rangordnung der Verfahren als stabil gelten kann. Die absoluten Höhen der jeweiligen Reliabilitäts- bzw. Validitätskennziffern können in folgenden Messgüte-Untersuchungen als Benchmarks herangezogen werden. Eine weitere Verbesserung der bestehenden Verfahren bzw. die Entwicklung neuer Methoden der Präferenzmessung sollten in diesem Sinne zu einer Verbesserung der mit dieser Meta-Analyse vorliegenden Maßzahlen der Reliabilität und Validität führen.

Abschließend können die Ergebnisse der Meta-Analyse überspitzt dargestellt wie folgt zusammengefasst werden:

Um eine gute Messgüte zu erreichen, sollte in nachfolgenden Präferenzmessungen eine wahlbasierte Conjoint-Analyse mit einem HB-Verfahren zur Individualisierung der Teilnutzenwerte verwendet werden. Im Bedarfsfall sollten zudem eine Segmentierung mit dem Latent-Class-Verfahren und eine Marktanteilssimulation mit der Logit-Methode vorgenommen werden.

Anhang A: Studien, die in der Meta-Analyse zur Validität auf der individuellen Ebene berücksichtigt worden sind (Ergebnisse s. Kapitel 6.1.1)

Korrelation zwischen vorhergesagten und beobachteten Präferenzdaten (Ratingskalen)

- Agarwal, M. K. (1989). How Many Pairs Should We Use in Adaptive Conjoint Analysis: An Empirical Analysis. In American Marketing Association (Hrsg.), *AMA Winter Educators' Conference Preceedings* (7-11). Chicago: American Marketing Association.
- Agarwal, M. K. (1991). An Empirical Comparison of Traditional Full-Profile Conjoint and Adaptive Conjoint Analysis. In Academy of Marketing Science (Hrsg.), *Proceedings of the 15th Conference*. Richmond: Academy of Marketing Science.
- Akaah, I. P. (1991). Predictive Performance of Self-Explicated, Traditional Conjoint, and Hybrid Conjoint Models under Alternative Data Collection Modes. *Journal of the Academy of Marketing Science*, 19, 309-314.
- Akaah, I. P. & Korgaonkar, P. K. (1983). An Empirical Comparison of the Predictive Validity of Self-Explicated, Huber-Hybrid, Traditional Conjoint, and Hybrid Conjoint Models. *Journal of Marketing Research*, 20, 187-197.
- Allenby, G. M., Arora, N. & Ginter, J. L. (1995). Incorporating Prior Knowledge into the Analysis of Conjoint Studies. *Journal of Marketing Research*, 32, 152-162.
- Baalbaki, I. B. & Malhotra, N. K. (1995). Standardization versus Customization in International Marketing: An Investigation Using Bridging Conjoint Analysis. *Journal of the Academy of Marketing Science*, 26, 182-194.
- Basu, A. K. & Hastak, M. (1990). Multiattribute Judgements Under Uncertainty: A Conjoint Measurement Approach. *Advances in Consumer Research*, 17, 554-562.
- Cattin, P., Gelfand, A. E. & Danes, J. (1983). A Simple Procedure for Estimation in a Conjoint Model. *Journal of Marketing Research*, 20, 29-35.
- Cattin, P., Hermet, G. & Pioche, A. (1982). Alternative Hybrid Models for Conjoint Analysis: Some Empirical Results. In R. K. Srivastava & E. D. Shocker (Hrsg.), *Analytical Approaches to Product and Market Planning: The Second Conference* (142-151). Cambridge: Marketing Science Institute.
- Cestre, G. & Darmon, R. Y. (1998). Assessing Consumer Preferences in the Context of New Product Diffusion. *International Journal of Research in Marketing*, 15, 123-135.

- Dahan, E., Hauser, J. R., Simester, D. & Toubia, O. (2001). *Application and Test of Web-Based Adaptive Polyhedral Conjoint Analysis*. Working paper Sloan School of Management, MIT, Cambridge.
- Geck, T. (2001). *Vergleich der Güte von Wahl- und Paarvergleichsmethoden in der Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Green, P. E., Goldberg, S. M. & Montemayor, M. (1981). A Hybrid Utility Estimation Model for Conjoint Analysis. *Journal of Marketing*, 45, 33-41.
- Green, P. E. & Helsen, K. (1989). Cross-Validation Assessment of Alternatives to Individual-Level Conjoint Analysis: A Case Study. *Journal of Marketing Research*, 26, 346-350.
- Green, P. E., Helsen, K. & Shandler, B. (1988). Conjoint Internal Validity Under Alternative Profile Presentations. *Journal of Consumer Research*, 15, 392-397.
- Green, P. E., Kim, J. & Shandler, B. (1990). *Predictive Accuracy of Classification and Regression Trees (CART) versus Conjoint Analysis*. Auf Anforderung direkt von den Autoren zu erhalten.
- Green, P. E., Krieger, A. M. & Agarwal, M. K. (1993). A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences. *Marketing Letters*, 4, 369-380.
- Green, P. E., Krieger, A. M. & Bansal, P. (1988). Completely Unacceptable Levels in Conjoint Analysis: A Cautionary Note. *Journal of Marketing Research*, 25, 293-300.
- Green, P. E., Krieger, A. M. & Schaffer, C. M. (1993). A Hybrid Conjoint Model with Individual-Level Interaction Estimation. *Advances in Consumer Research*, 20, 149-154.
- Green, P. E., Krieger, A. M. & Schaffer, C. M. (1993). An Empirical Test of Optimal Respondent Weighting in Conjoint Analysis. *Journal of the Academy of Marketing Science*, 21, 345-351.
- Hausruckinger, G. & Helm, R. (1996). Die Bedeutung des Country-of-Origin Effekts vor dem Hintergrund der Internationalisierung von Unternehmen. *Marketing ZFP*, 4, 267-278.
- Holling, H. & Schwabe, R. (2002). *Effiziente Versuchsplanung in der Conjoint-Analyse – Zwischenbericht zum DFG-Projekt*. Working Paper am Lehrstuhl für Evaluations- und Forschungsmethoden, Organisationspsychologie an der Universität Münster.
- Krieger, A. M., Green, P. E. & Umesh, U. N. (1998). Effect of Level of Disaggregation of Conjoint Cross Validations: Some Comparative Findings. *Decision Sciences*, 29, 1047-1058.

- Lenk, P. J., DeSarbo, W. S., Green, P. & Young, M. R. (1996). Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science*, 15, 173-191.
- Leptien, C. (1995). Incentives for Employed Inventors: An Empirical Analysis with Special Emphasis in the German Law for Employee's Inventions. *R&D Management*, 25, 213-225.
- Molin, E., Oppewal, H. & Timmermans, H. (2002). Conjoint Modeling of Residential Group Preferences: A Comparison of the Internal Validity of Hierarchical Information Integration Approaches. *Journal of Geographical Systems*, 4, 343-358.
- Moore, W. L. (1980). Levels of Aggregation in Conjoint Analysis: An Empirical Comparison. *Journal of Marketing Research*, 17, 516-523.
- Moore, W. L., Metha, R. B. & Pavia, T. M. (1994). A Simplified Method of Constraint Parameter Estimation in Conjoint Analysis. *Marketing Letters*, 5, 173-181.
- Mulye, R. (1998). An Empirical Comparison of Three Variants of the AHP and Two Variants of Conjoint Analysis. *Journal of Behavioral Decision Making*, 11, 263-280.
- Mummalaneni, V., Dubas, K. M. & Chao, C. (1996). Chinese Purchasing Managers' Preferences and Trade-offs in Supplier Selection and Performance Evaluation. *Industrial Marketing Management*, 25, 115-124.
- Neslin, S. A. (1981). Linking Product Features to Perceptions: Self-Stated Versus Statistically Revealed Importance Weights. *Journal of Marketing Research*, 18, 80-86.
- Oppewal, H. & Timmermans, H. (1999). Modeling Consumer Perception of Public Space in Shopping Centers. *Environment and Behavior*, 31, 45-65.
- Pavia, T. M., Moore, W., Mehta, R. B. & Ottum, B. D. (1996). *Can We Simplify the Conjoint Task: An Investigation of the Effect of Differing Numbers of Calibration Attributes on Validations*. Working Paper David Eccles School of Business, University of Utah, Salt Lake City.
- Schweickl, H. (1985). *Computergestützte Präferenzanalyse mit individuell wichtigen Produktmerkmalen*. Berlin: Duncker & Humblot.

Korrelation zwischen vorhergesagten und beobachteten Rangordnungen

- Acito, F. & Olshavsky, R. W. (1980). Limits to Accuracy in Conjoint Analysis. In Association for Consumer Research (Hrsg.), *Advances in Consumer Research* 8 (313-316). Provo: Association for Consumer Research.
- Agarwal, M. K. (1991). An Empirical Comparison of Traditional Full-Profile Conjoint and Adaptive Conjoint Analysis. In Academy of Marketing Science (Hrsg.), *Proceedings of the 15th Conference*. Richmond: Academy of Marketing Science.
- Albrecht, J. (2000). *Präferenzstrukturmessung - Ein empirischer Vergleich der Conjoint-Analyse mit einer kompositionellen Methode*. Europäische Hochschulschriften, Frankfurt: Peter Lang.
- Allenby, G. M., Arora, N. & Ginter, J. L. (1995). Incorporating Prior Knowledge into the Analysis of Conjoint Studies. *Journal of Marketing Research*, 32, 152-162.
- Baier, D. & Säuberlich, F. (1997). Kundennutzenschätzung mittels individueller Hybrid-Conjointanalyse. *Zeitschrift für betriebswirtschaftliche Forschung ZfbF*, 49, 951-972.
- Basu, A. K. & Hastak, M. (1990). Multiattribute Judgements Under Uncertainty: A Conjoint Measurement Approach. *Advances in Consumer Research*, 17, 554-562.
- Dahan, E., Hauser, J. R., Simester, D. & Toubia, O. (2001). *Application and Test of Web-Based Adaptive Polyhedral Conjoint Analysis*. Working paper Sloan School of Management, MIT, Cambridge.
- Eck, M. (2001). *Vergleich kompositioneller und dekompositioneller Verfahren in der multiattributiven Nutzenanalyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Geck, T. (2001). *Vergleich der Güte von Wahl- und Paarvergleichsmethoden in der Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Green, P. E. (1984). Hybrid Models for Conjoint Analysis: An Expository Review. *Journal of Marketing Research*, 21, 155-169.
- Green, P. E., Carmone, F. J. & Wind, Y. (1972). Subjective Evaluation Models and Conjoint Measurement. *Behavioral Science*, 17, 288-299.
- Green, P. E., Helsen, K. & Shandler, B. (1988). Conjoint Internal Validity Under Alternative Profile Presentations. *Journal of Consumer Research*, 15, 392-397.
- Hagerty, M. R. (1985). Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis. *Journal of Marketing Research*, 22, 168-184.

- Jain, A. K., Acito, F., Malhotra, N. K. & Mahajan, V. (1979). A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models. *Journal of Marketing Research*, 16, 313-322.
- Melles, T., Laumann, R. & Holling, H. (2000). *Validity and Reliability of Online Conjoint Analysis*. Working Paper presented at the Sawtooth Software Conference, Hilton Head, SC.
- Mulye, R. (1998). An Empirical Comparison of Three Variants of the AHP and Two Variants of Conjoint Analysis. *Journal of Behavioral Decision Making*, 11, 263-280.
- Nickerson, C. A. E., McClelland, G. H. & Petersen, D. M. (1990). Solutions to some Problems in the Implementation of Conjoint Analysis. *Behavior Research Methods, Instruments, & Computers*, 22, 360-374.
- Olshavsky, R. W. & Acito, F. (1980). An Information Processing Probe into Conjoint Analysis. *Decision Sciences*, 11, 451-470.
- Safizadeh, H. M. (1989). The Internal Validity of the Trade-off Method of Conjoint Analysis. *Decision Sciences*, 20, 451-461.
- Sheluga, D. A., Jaccard, J. & Jacoby, J. (1979). Preference, Search, and Choice: An Integrative Approach. *Journal of Consumer Research*, 6, 166-176.
- Srinivasan, V. & Jain, A. K. & Malhotra, N. K. (1983). Improving Predictive Power of Conjoint Analysis by Constrained Parameter Estimation. *Journal of Marketing Research*, 20, 433-438.
- Tscheulin, D. K. (1996). Determinanten der Validität der Conjoint-Analyse. In A. von Ahnen & T. Czernikowsky (Hrsg.), *Marketing und Marktforschung* (585-598). Hamburg: LIT-Verlag.
- Tscheulin, D. K. & Blaimont, C. (1993). Die Abhängigkeit der Prognosegüte von Conjoint-Studien von demographischen Probanden-Charakteristika. *Zeitschrift für Betriebswirtschaft ZfB*, 63 (8), 839-847.
- Vriens, M. (1995). *Conjoint Analysis in Marketing: Developments in Stimulus Representation and Segmentation Methods*. Capelle, NL: Labyrinth Publications.
- Willms, R. (2001). *Generierung von Anforderungsprofilen mittels Conjoint-Analyse und Discrete Choice-Analyse*. Diplomarbeit an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld.

Korrelation zwischen den vorhergesagten und ermittelten Teilnutzenwerten

- Agarwal, M. K. & Green, P. E. (1991). Adaptive Conjoint Analysis Versus Self-Explicated Models: Some Empirical Results. *International Journal of Research in Marketing*, 8, 141-146.
- Albrecht, J. (2000). *Präferenzstrukturmessung - Ein empirischer Vergleich der Conjoint-Analyse mit einer kompositionellen Methode*. Europäische Hochschulschriften, Frankfurt: Peter Lang.
- Baier, D. & Säuberlich, F. (1997). Kundennutzenschätzung mittels individueller Hybrid-Conjointanalyse. *Zeitschrift für betriebswirtschaftliche Forschung ZfbF*, 49, 951-972.
- Dorsch, M. J. & Teas, K. R. (1992). A Test of the Convergent Validity of Self-explicated and Decompositional Conjoint Measurement. *Journal of the Academy of Marketing Science*, 20, 37-48.
- Geck, T. (2001). *Vergleich der Güte von Wahl- und Paarvergleichsmethoden in der Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Holling, H. & Schwabe, R. (2003). *Effiziente Versuchsplanung in der Conjoint-Analyse – 2. Zwischenbericht zum DFG-Projekt*. Working Paper am Lehrstuhl für Evaluations- und Forschungsmethoden, Organisationspsychologie an der Universität Münster.
- Kamakura, W. & Ozer, M. (2001). A Multi-trait Multi-Method Validity Test of Partworth Estimates. In A. Gustafsson et al. (Hrsg.), *Conjoint Measurement, Methods and Applications* (251-277). Berlin: Springer.
- Leigh, T. W., MacKay, D. B. & Summers, J. O. (1981). On Alternative Experimental Methods for Conjoint Analysis. In K. B. Monroe (Hrsg.), *Advances in Consumer Research* (317-322). Ann Arbor: Association for Consumer Research.
- Melles, T. (1996). *Eine empirische Studie zur Optimierung der Paarvergleichsaufgabe im Rahmen der adaptiven Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Mulye, R. (1998): An Empirical Comparison of Three Variants of the AHP and Two Variants of Conjoint Analysis: *Journal of Behavioral Decision Making*, 11, 263-280.
- Sheluga, D. A., Jaccard, J. & Jacoby, J. (1979). Preference, Search, and Choice: An Integrative Approach. *Journal of Consumer Research*, 6, 166-176.

Willms, R. (2001). *Generierung von Anforderungsprofilen mittels Conjoint-Analyse und Discrete Choice-Analyse*. Diplomarbeit an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld.

Korrelation zwischen den vorhergesagten und ermittelten Attributwichtigkeiten

Akaah, I. P. (1991). Predictive Performance of Self-Explicated, Traditional Conjoint, and Hybrid Conjoint Models under Alternative Data Collection Modes. *Journal of the Academy of Marketing Science*, 19, 309-314.

Akaah, I. P. & Korgaonkar, P. K. (1983). An Empirical Comparison of the Predictive Validity of Self-Explicated, Huber-Hybrid, Traditional Conjoint, and Hybrid Conjoint Models. *Journal of Marketing Research*, 20, 187-197.

Albrecht, J. (2000). *Präferenzstrukturmessung - Ein empirischer Vergleich der Conjoint-Analyse mit einer kompositionellen Methode*. Europäische Hochschulschriften, Frankfurt: Peter Lang.

Dorsch, M. J. & Teas, K. R. (1992). A Test of the Convergent Validity of Self-explicated and Decompositional Conjoint Measurement. *Journal of the Academy of Marketing Science*, 20, 37-48.

Heeler, R. M., Okechuku, C. & Reid, S. (1979). Attribute Importance: Contrasting Measurements. *Journal of Marketing Research*, 16, 60-63.

Jaccard, J., Brinberg, D. & Ackerman, L. J. (1996). Assessing Attribute Importance: A Comparison of Six Methods. *Journal of Consumer Research*, 12, 463-468.

Scott, J. E. & Wright, P. (1976). Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations. *Journal of Marketing Research*, 13, 211-224.

Sheluga, D. A., Jaccard, J. & Jacoby, J. (1979). Preference, Search, and Choice: An Integrative Approach. *Journal of Consumer Research*, 6, 166-176.

Willms, R. (2001). *Generierung von Anforderungsprofilen mittels Conjoint-Analyse und Discrete Choice-Analyse*. Diplomarbeit an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld.

Anteil der korrekt vorhergesagten Wahlen / Hitrate

- Agarwal, M. K. (1991). An Empirical Comparison of Traditional Full-Profile Conjoint and Adaptive Conjoint Analysis. In Academy of Marketing Science (Hrsg.), *Proceedings of the 15th Conference*. Richmond: AMS.
- Agarwal, M. K. & Green, P. E. (1991). Adaptive Conjoint Analysis Versus Self-Explicated Models: Some Empirical Results. *International Journal of Research in Marketing*, 8, 141-146.
- Agarwal, M. K. & Rao, V. R. (1996). An Empirical Comparison of Consumer-Based Measures of Brand Equity. *Marketing Letters*, 7, 237-247.
- Akaah, I. P. & Korgaonkar, P. K. (1983). An Empirical Comparison of the Predictive Validity of Self-Explicated, Huber-Hybrid, Traditional Conjoint, and Hybrid Conjoint Models. *Journal of Marketing Research*, 20, 187-197.
- Albrecht, J. (2000). *Präferenzstrukturmessung - Ein empirischer Vergleich der Conjoint-Analyse mit einer kompositionellen Methode*. Europäische Hochschulschriften, Frankfurt: Peter Lang.
- Allenby, G. M., Arora, N. & Ginter, J. L. (1995). Incorporating Prior Knowledge into the Analysis of Conjoint Studies. *Journal of Marketing Research*, 32, 152-162.
- Anderson, J. C. & Donthu, N. (1988). A Proximate Assessment of the External Validity of Conjoint Analysis. In Frazier et al. (Hrsg.), *AMA Educators' Proceedings 54* (287-291). Chicago: AMA.
- Baier, D. & Säuberlich, F. (1997). Kundennutzenschätzung mittels individueller Hybrid-Conjointanalyse. *Zeitschrift für betriebswirtschaftliche Forschung ZfbF*, 49, 951-972.
- Böcker, F. & Schweikl, H. (1988). Better Preference Prediction with Individualized Sets of Relevant Attributes. *International Journal of Research in Marketing*, 5, 15-24.
- Cattin, P., Hermet, G. & Pioche, A. (1982). Alternative Hybrid Models for Conjoint Analysis: Some Empirical Results. In R. K. Srivastava and E. D. Shocker (Hrsg.), *Analytical Approaches to Product and Market Planning: The Second Conference* (142-151). Cambridge: Marketing Science Institute.
- Chrzan, K. (1998). *Full versus Partial Profile Choice Experiments: Aggregate and Disaggregate Comparisons*. Working Paper presented at the INFORMS Marketing Science Conference, Paris.

- Dahan, E., Hauser, J. R., Simester, D. & Toubia, O. (2001). *Application and Test of Web-Based Adaptive Polyhedral Conjoint Analysis*. Working paper Sloan School of Management, MIT, Cambridge.
- Ding, M., Grewal, R. & Liechty, J. (2004). *Incentive-Aligned Conjoint Analysis*. Working Paper Smeal College of Business Administration, Pennsylvania State University, Pennsylvania.
- Eck, M. (2001). *Vergleich kompositioneller und dekompositioneller Verfahren in der multiattributiven Nutzenanalyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Elrod, T., Louviere, J. J. & Davey, K. S. (1992). An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models. *Journal of Marketing Research*, 29, 368-377.
- Geck, T. (2001). *Vergleich der Güte von Wahl- und Paarvergleichsmethoden in der Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Green, P. E., Goldberg, S. M. & Montemayor, M. (1981). A Hybrid Utility Estimation Model for Conjoint Analysis. *Journal of Marketing*, 45, 33-41.
- Green, P. E. & Helsen, K. (1989). Cross-Validation Assessment of Alternatives to Individual-Level Conjoint Analysis: A Case Study. *Journal of Marketing Research*, 26, 346-350.
- Green, P. E., Helsen, K. & Shandler, B. (1988). Conjoint Internal Validity Under Alternative Profile Presentations. *Journal of Consumer Research*, 15, 392-397.
- Green, P. E., Kim, J. & Shandler, B. (1990). *Predictive Accuracy of Classification and Regression Trees (CART) versus Conjoint Analysis*. Auf Anforderung direkt von den Autoren zu erhalten.
- Green, P. E. & Krieger, A. M. (1995). Attribute Importance Weights Modification in Assessing a Brand's Competitive Potential. *Marketing Science*, 4, 253-270.
- Green, P. E., Krieger, A. M. & Agarwal, M. K. (1993). A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences. *Marketing Letters*, 4, 369-380.
- Green, P. E., Krieger, A. M. & Schaffer, C. M. (1993). A Hybrid Conjoint Model with Individual-Level Interaction Estimation. *Advances in Consumer Research*, 20, 149-154.
- Haijjer, R., Kamakura, W. & Wedel, M. (2000). Response Latencies in the Analysis of Conjoint Choice Experiments. *Journal of Marketing Research*, 37, 376-382.
- Hagerty, M. R. (1985). Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis. *Journal of Marketing Research*, 22, 168-184.

- Hansen, D. E. & Wittink, D. R. (1995). Combining Self-Explicated Priors with Conjoint Data Using Bayesian Regression. *Marketing Letters*, 6, 63-71.
- Hauser, J. R. & Shugan, S. M. (1978). Extended Conjoint Analysis with Intensity Measures and Computer Assisted Interviews: Applications to Telecommunication and Travel. *Advances in Consumer Research*, 5, 440-447.
- Hensel-Börner, S. & Sattler, H. (2000). Ein empirischer Validitätsvergleich zwischen der Customized Computerized Conjoint Analysis (CCC), der Adaptive Conjoint Analysis (ACA) und Self-Explicated Verfahren. *Zeitschrift für Betriebswirtschaft ZfB*, 70, 705-727.
- Holling, H. (2004). Präferenzmessung im Personalmarketing durch optimal geplante Conjoint-Analysen. In K.-C. Hamborg & H. Holling (Hrsg.), *Innovative Ansätze der Personal- und Organisationsentwicklung* (53-72). Göttingen: Hogrefe.
- Holling, H. & Schwabe, R. (2002). *Effiziente Versuchsplanung in der Conjoint-Analyse – Zwischenbericht zum DFG-Projekt*. Working Paper am Lehrstuhl für Evaluations- und Forschungsmethoden, Organisationspsychologie an der Universität Münster.
- Holling, H. & Schwabe, R. (2003). *Effiziente Versuchsplanung in der Conjoint-Analyse – 2. Zwischenbericht zum DFG-Projekt*. Working Paper am Lehrstuhl für Evaluations- und Forschungsmethoden, Organisationspsychologie an der Universität Münster.
- Hopkins, D. S., Larréché, J. & Massy, W. F. (1977). Constrained Optimization of a University Administrator's Preference Function. *Management Science*, 24, 4, 365-377.
- Huber, J. & Hansen, D. (1986). Testing the Impact of Dimensional Complexity and Affective Differences of Paired Concepts. In Association for Consumer Research (Hrsg.), *Advances in Consumer Research* (159-163). Provo: ACR.
- Huber, J., Wittink, D. R., Fiedler, J. A. & Miller, R. (1993). The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice. *Journal of Marketing Research*, 30, 105-114.
- Huber, J., Wittink, D. R., Johnson, R. M. & Miller, R. (1992). Learning Effects in Preference Tasks: Choice-Based versus Standard Conjoint. *Sawtooth Software Conference Proceedings*.
- Jain, A. K., Acito, F., Malhotra, N. K. & Mahajan, V. (1979). A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models. *Journal of Marketing Research*, 16, 313-322.

- Johnson, R. M. (1989). Assessing the Validity of Conjoint Analysis. In Sawtooth Software Inc. (Hrsg.), *Proceedings of the Sawtooth Software Conference 3* (273-280). Ketchum: Sawtooth Software.
- Johnson, R. M. (1997). Individual Utilities from Choice Data: A New Method. In Sawtooth Software Inc. (Hrsg.), *Proceedings of the Sawtooth Software Conference* (191-208). Ketchum: Sawtooth Software.
- Johnson, R. M. (1999). The Joys and Sorrows of Implementing HB Methods for Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Johnson, R. M. (2000). Monotonicity Constraints in Choice-Based Conjoint with Hierarchical Bayes. *Sawtooth Software Research Paper Series*.
- Johnson, R. M., Huber, J. & Orme, B. (2005). A Second Test of Adaptive Choice-Based Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Johnson, R. M., Orme, B., Huber, J. & Pinnell, J. (2005). Testing Adaptive Choice-Based Conjoint Designs. *Sawtooth Software Research Paper Series*.
- King, W. C., Hill, A. & Orme, B. (2005). The Importance Question in ACA: Can It Be Omitted?. *Sawtooth Software Research Paper Series*.
- Klemz, B. R. & Gruca, T. S. (2003). Dueling or the Battle Royale? The Impact of Task Complexity on the Evaluation of Entry Threat. *Psychology & Marketing*, 20 (11), 999-1016.
- Leigh, T. W., MacKay, D. B. & Summers, J. O. (1984). Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison. *Journal of Marketing Research*, 21, 456-462.
- Lenk, P. J., DeSarbo, W. S., Green, P. & Young, M. R. (1996). Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science*, 15, 173-191.
- McLauchlan, W. G. (1991). Scaling Prior Utilities in Sawtooth Software's Adaptive Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Meyer, L. (1998). *Predictive Accuracy for Conjoint Models for Data Collected by Means of World Wide Web Survey*. Verfügbar unter: <http://www.lucameyer.com/kul/title.htm> [26.08.2002].
- Molin, E., Oppewal, H. & Timmermans, H. (1999). Group-based versus Individual-based Conjoint Preference Models of Residential Preferences: A Comparative Test. *Environment and Planning A*, 31, 1935-1947.

- Moore, W. L., Gray-Lee, J. & Louviere, J. J. (1998). A Cross-Validity Comparison of Conjoint Analysis and Choice Models at Different Levels of Aggregation. *Marketing Letters*, 9, 195-207.
- Nickerson, C. A. E., McClelland, G. H. & Petersen, D. M. (1990). Solutions to some Problems in the Implementation of Conjoint Analysis. *Behavior Research Methods, Instruments, & Computers*, 22, 360-374.
- Orme, B. K., Alpert, M. I. & Christensen, E. (1997). Assessing the Validity of Conjoint Analysis – Continued. *Sawtooth Software Research Paper Series*.
- Orme, B. K. & King, C. W. (1998). Conducting Full-Profile Conjoint Analysis over the Internet. *Sawtooth Software Research Paper Series*.
- Parker, B. R. & Srinivasan, V. (1976). A Consumer Preference Approach to the Planning of Rural Health-care Facilities. *Operations Research*, 24, 991-1025.
- Patterson, M. & Frazier, C. (2000). *Cutoff Constrained Discrete Choice Models*. Working Paper presented at the Sawtooth Software Conference, Hilton Head, SC.
- Pullman, M. E., Dodson, K. J. & Moore, W. M. (1999). A Comparison of Conjoint Methods When There Are Many Attributes. *Marketing Letters*, 10, 125-138.
- Rosko, M. D., DeVita, M., McKenna, W. F. & Walker, L. R. (1985). Strategic Marketing Applications of Conjoint Analysis: An HMO Perspective. *Journal of Health Care Marketing*, 5 (4), 27-38.
- Sattler, E. (1994). Die Validität von Produkttests. *Marketing ZFP*, 1, 31-41.
- Sawtooth Software (1999). ACA/Hierarchical Bayes - Technical Paper. *Sawtooth Software Technical Paper Series*.
- Sawtooth Software (2000). CBC/Hierarchical Bayes Analysis - Technical Paper (version 3.0). *Sawtooth Software Technical Paper Series*.
- Sawtooth Software (2003). ACA/Hierarchical Bayes v2.0 - Technical Paper. *Sawtooth Software Technical Paper Series*.
- Sentis, K. & Li, L. (2002). One Size Fits All or Custom Tailored: Which HB Fits Better?. *Sawtooth Software Research Paper Series*.
- Sheluga, D. A., Jaccard, J. & Jacoby, J. (1979). Preference, Search, and Choice: An Integrative Approach. *Journal of Consumer Research*, 6, 166-176.
- Srinivasan, V. (1988). A Conjunctive-Compensatory Approach to the Self-Explication of Multiattributed Preferences. *Decision Sciences*, 19, 295-305.

- Srinivasan, V., Jain, A. K. & Malhotra, N. K. (1983). Improving Predictive Power of Conjoint Analysis by Constrained Parameter Estimation. *Journal of Marketing Research*, 20, 433-438.
- Srinivasan, V. & Park, C. S. (1997). Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement. *Journal of Marketing Research*, 34, 286-291.
- Teichert, T. (2000). Auswirkungen von Verfahrensalternativen bei der Erhebung von Präferenzurteilen. *Marketing ZFP*, 2, 145-159.
- Teichert, T. (2001). *Nutzenschätzung in Conjoint-Analysen: theoretische Fundierung und empirische Aussagekraft*. Wiesbaden: Gabler.
- Van der Lans, I. A., Wittink, D. R., Huber, J. & Vriens, M. (1992). Within- and across Attribute Constraints in ACA and Full Profile Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Voeth, M. (2000). *Nutzenmessung in der Kaufverhaltensforschung: Die hierarchisch individualisierte Limit Conjoint-Analyse (HILCA)*. Wiesbaden: Deutscher Universitätsverlag.
- Vriens, M. (1995). *Conjoint Analysis in Marketing: Developments in Stimulus Representation and Segmentation Methods*. Capelle, NL: Labyrinth Publications.
- Vriens, M., Opperwal, H. & Wedel, M. (1998). Ratings-based versus Choice-based Latent Class Conjoint Models - An Empirical Comparison. *Journal of Market Research Society*, 40 (3), 237-248.
- Willms, R. (2001). *Generierung von Anforderungsprofilen mittels Conjoint-Analyse und Discrete Choice-Analyse*. Diplomarbeit an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld.
- Zwerina, K. (1997). *Discrete Choice Experiments in Marketing: Use of Priors in Efficient Choice Designs and Their Application to Individual Preference Measurement*. Heidelberg: Physica.

Anhang B: Studien, die in der Meta-Analyse zur mittleren absoluten Abweichung von Wahlanteilsprognosen (MAE) berücksichtigt worden sind (Ergebnisse s. Kapitel 6.1.3)

- Agarwal, M. K. (1989). How Many Pairs Should We Use in Adaptive Conjoint Analysis: An Empirical Analysis. In American Marketing Association (Hrsg.), *AMA Winter Educators' Conference Proceedings* (7-11). Chicago: AMA.
- Allenby, G. M., Arora, N. & Ginter, J. L. (1995). Incorporating Prior Knowledge into the Analysis of Conjoint Studies. *Journal of Marketing Research*, 32, 152-162.
- Bucklin, R. E. & Srinivasan, V. (1991). Determining Interbrand Substitutability Through Survey Measurement of Consumer Preference Structures. *Journal of Marketing Research*, 28, 58-71.
- Chrzan, K. (1998). *Full versus Partial Profile Choice Experiments: Aggregate and Disaggregate Comparisons*. Working Paper presented at the INFORMS Marketing Science Conference, Paris.
- Chrzan, K. & Boeger, L. (1999). *Improving Choice Predictions Using a Cutoff-Constrained Aggregate Choice Model*. Working Paper presented at the INFORMS Marketing Science Conference, Syracuse.
- Conrad, T. (1997). *Preisbildung mittels der Conjoint-Analyse und eines Simulationsmodells am Beispiels eines Premiumanbieters der Automobilindustrie*. Dissertation an der wirtschaftswissenschaftlichen Fakultät der Universität Tübingen.
- Dahan, E. & Srinivasan, V. (2000). The Predictive Power of Internet-Based Product Concept Testing Using Visual Depiction and Animation. *Journal of Product Innovation Management*, 17, 99-109.
- Dellaert, B., Borgers, A., Louviere, J. J. & Timmermans, H. (2001). Using Conjoint Choice Experiments to Model Consumer Choices of Product Component Packages. In A. Gustafsson et al. (Hrsg.), *Conjoint Measurement, Methods and Applications* (387-414). Berlin: Springer.
- Elrod, T., Louviere, J. J. & Davey, K. S. (1992). An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models. *Journal of Marketing Research*, 29, 368-377.
- Finkbeiner, C. T. (1988). Comparison of Conjoint Choice Simulators. *Proceedings of the Sawtooth Software Conference*, 75-103.

- Hartmann, A. & Sattler, H. (2002). Commercial Use of Conjoint Analysis in Germany, Austria, and Switzerland. *Research Paper on Marketing and Retailing*, Universität Hamburg.
- Hauser, J. R. & Shugan, S. M. (1978). Extended Conjoint Analysis with Intensity Measures and Computer Assisted Interviews: Applications to Telecommunication and Travel. *Advances in Consumer Research*, 5, 440-447.
- Hensel-Börner, S. & Sattler, H. (2000). Ein empirischer Validitätsvergleich zwischen der Customized Computerized Conjoint Analysis (CCC), der Adaptive Conjoint Analysis (ACA) und Self-Explicated Verfahren. *Zeitschrift für Betriebswirtschaft ZfB*, 70, 705-727.
- Huber, J., Wittink, D. R., Fiedler, J. A. & Miller, R. (1993). The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice. *Journal of Marketing Research*, 30, 105-114.
- Johnson, R. M. (1997). Individual Utilities from Choice Data: A New Method. In Sawtooth Software Inc. (Hrsg.), *Proceedings of the Sawtooth Software Conference* (191-208). Ketchum: Sawtooth Software.
- Johnson, R. M. (1999). The Joys and Sorrows of Implementing HB Methods for Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Johnson, R. M. (2000). Monotonicity Constraints in Choice-Based Conjoint with Hierarchical Bayes. *Sawtooth Software Research Paper Series*.
- Johnson, R. M., Huber, J. & Orme, B. (2005). A Second Test of Adaptive Choice-Based Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Johnson, R. M., Orme, B., Huber, J. & Pinnell, J. (2005). Testing Adaptive Choice-Based Conjoint Designs. *Sawtooth Software Research Paper Series*.
- King, W. C., Hill, A. & Orme, B. (2005). The Importance Question in ACA: Can It Be Omitted?. *Sawtooth Software Research Paper Series*.
- Krieger, A. M., Green, P. E. & Umesh, U. N. (1998). Effect of Level of Disaggregation of Conjoint Cross Validations: Some Comparative Findings. *Decision Sciences*, 29, 1047-1058.
- Moore, W. L., Gray-Lee, J. & Louviere, J. J. (1998). A Cross-Validity Comparison of Conjoint Analysis and Choice Models at Different Levels of Aggregation. *Marketing Letters*, 9, 195-207.

- Oliphant, K., Eagle, T. C., Louviere, J. J. & Anderson, D. (1992). Cross-Task Comparison of Ratings-Based and Choice-Based Conjoint. *Proceedings of the Sawtooth Software Conference*, 383-404.
- Orme, B. K. & Baker, G. C. (2000). Comparing Hierarchical Bayes Draws and Randomized First Choice for Conjoint Simulations. *Sawtooth Software Research Paper Series*.
- Orme, B. K. & Heft, M. A. (1999). Predicting Actual Sales with CBC: How Capturing Heterogeneity Improves Results. *Sawtooth Software Research Paper Series*.
- Page, A. L. & Rosenbaum, H. F. (1987). Redesigning Product Lines with Conjoint Analysis: How Sunbeam Does It. *Journal of Product Innovation Management*, 4, 120-137.
- Patterson, M. & Frazier, C. (2000). *Cutoff Constrained Discrete Choice Models*. Working Paper presented at the Sawtooth Software Conference, Hilton Head, SC.
- Pullman, M. E., Dodson, K. J. & Moore, W. M. (1999). A Comparison of Conjoint Methods When There Are Many Attributes. *Marketing Letters*, 10, 125-138.
- Sattler, E., Gedenk, K. & Hensel-Börner, S. (2002). *Bandbreiten-Effekte bei multiattributiven Entscheidungen: Ein empirischer Vergleich von Verfahren zur Bestimmung von Eigenschaftsgewichten*. Working Paper am Institut für Handel und Marketing der Universität Hamburg.
- Sawtooth Software (1999). ACA/Hierarchical Bayes - Technical Paper. *Sawtooth Software Technical Paper Series*.
- Sawtooth Software (2000). CBC/Hierarchical Bayes Analysis - Technical Paper (version 3.0). *Sawtooth Software Technical Paper Series*.
- Sawtooth Software (2003). ACA/Hierarchical Bayes v2.0 - Technical Paper. *Sawtooth Software Technical Paper Series*.
- Van der Lans, I. A., Wittink, D. R., Huber, J. & Vriens, M. (1992). Within- and across Attribute Constraints in ACA and Full Profile Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Zwerina, K. (1997). *Discrete Choice Experiments in Marketing: Use of Priors in Efficient Choice Designs and Their Application to Individual Preference Measurement*. Heidelberg: Physica.

Anhang C: Studien, die in der Meta-Analyse zum Root Mean Squared Error (RMSE) von Wahlanteilsprognosen berücksichtigt worden sind (Ergebnisse s. Kapitel 6.1.3)

Krieger, A. M., Green, P. E. & Umesh, U. N. (1998). Effect of Level of Disaggregation of Conjoint Cross Validations: Some Comparative Findings. *Decision Sciences*, 29, 1047-1058.

Lenk, P. J., DeSarbo, W. S., Green, P. & Young, M. R. (1996). Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science*, 15, 173-191.

Moore, W. L., Gray-Lee, J. & Louviere, J. J. (1998). A Cross-Validity Comparison of Conjoint Analysis and Choice Models at Different Levels of Aggregation. *Marketing Letters*, 9, 195-207.

Pullman, M. E., Dodson, K. J. & Moore, W. M. (1999). A Comparison of Conjoint Methods When There Are Many Attributes. *Marketing Letters*, 10, 125-138.

Anhang D: Studien, die in der Meta-Analyse zur Reliabilität berücksichtigt worden sind (Ergebnisse s. Kapitel 6.2)

Korrelation zwischen den mit Einstufungsskalen erhobenen Inputdaten zweier Messungen

Cestre, G. & Darmon, R. Y. (1998). Assessing Consumer Preferences in the Context of New Product Diffusion. *International Journal of Research in Marketing*, 15, 123-135.

Green, P. E., Krieger, A. M. & Agarwal, M. K. (1993). A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences. *Marketing Letters*, 4, 369-380.

Hoeffler, S. (2003). Measuring Preferences for Really New Products. *Journal of Marketing Research*, 40, 406-420.

Johnson, R. M. (1989). Assessing the Validity of Conjoint Analysis. In Sawtooth Software (Hrsg.), *Proceedings of the Sawtooth Software Conference 3* (273-280). Ketchum: Sawtooth Software.

Leigh, T. W., MacKay, D. B. & Summers, J. O. (1981). On Alternative Experimental Methods for Conjoint Analysis. In K. B. Monroe (Hrsg.), *Advances in Consumer Research* (317-322). Ann Arbor: Association for Consumer Research.

Leigh, T. W., MacKay, D. B. & Summers, J. O. (1984). Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison. *Journal of Marketing Research*, 21, 456-462.

McLauchlan, W. G. (1991). Scaling Prior Utilities in Sawtooth Software's Adaptive Conjoint Analysis. *Sawtooth Software Research Paper Series*.

Pullman, M. E., Dodson, K. J. & Moore, W. M. (1999). A Comparison of Conjoint Methods When There Are Many Attributes. *Marketing Letters*, 10, 125-138.

Teas, K. R. (1987). Magnitude Scaling of the Dependent Variable in Decompositional Multiattribute Preference Models. *Journal of the Academy of Marketing Science*, 15, 64-73.

Korrelation zwischen den Rangordnungen zweier Messungen

- Acito, F. (1977). An Investigation of Some Data Collection Issues in Conjoint Measurement. In American Marketing Association (Hrsg.), *Educators' Conference Proceedings* (82-85). Chicago: American Marketing Association.
- Heeler, R. M., Okechuku, C. & Reid, S. (1979). Attribute Importance: Contrasting Measurements. *Journal of Marketing Research*, 16, 60-63.
- Kamakura, W. & Ozer, M. (2001). A Multi-trait Multi-Method Validity Test of Partworth Estimates. In A. Gustafsson et al. (Hrsg.), *Conjoint Measurement, Methods and Applications* (251-277). Berlin: Springer
- Leigh, T. W., MacKay, D. B. & Summers, J. O. (1984). Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison. *Journal of Marketing Research*, 21, 456-462.
- McCullough, J. & Best, R. (1979). Conjoint Measurement: Temporal Stability and Structural Reliability. *Journal of Marketing Research*, 16, 26-31.
- Melles, T. (1996). *Eine empirische Studie zur Optimierung der Paarvergleichsaufgabe im Rahmen der adaptiven Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Segal, M. N. (1982). Reliability of Conjoint Analysis: Contrasting Data Collection Procedures. *Journal of Marketing Research*, 19, 139-143.
- Vriens, M. (1995). *Conjoint Analysis in Marketing: Developments in Stimulus Representation and Segmentation Methods*. Capelle, NL: Labyrinth Publications.
- Weisenfeld, U. (1989). *Die Einflüsse von Verfahrensvariationen und der Art des Kaufentscheidungsprozesses auf die Reliabilität der Ergebnisse bei der Conjoint Analyse*. Berlin: Duncker und Humblot.

Korrelation von aus zwei Messungen ermittelten Teilnutzenwerten

- Baalbaki, I. B. & Malhotra, N. K. (1995). Standardization versus Customization in International Marketing: An Investigation Using Bridging Conjoint Analysis. *Journal of the Academy of Marketing Science*, 26, 182-194.
- Baier, D. & Säuberlich, F. (1997). Kundennutzenschätzung mittels individueller Hybrid-Conjointanalyse. *Zeitschrift für betriebswirtschaftliche Forschung ZfbF*, 49, 951-972.
- Bucklin, R. E. & Srinivasan, V. (1991). Determining Interbrand Substitutability Through Survey Measurement of Consumer Preference Structures. *Journal of Marketing Research*, 28, 58-71.
- Eck, M. (2001). *Vergleich kompositioneller und dekompositioneller Verfahren in der multiattributiven Nutzenanalyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Ernst, O. & Sattler, H. (2000). Multimediale versus traditionelle Conjoint-Analysen – Ein empirischer Vergleich alternativer Produktpräsentationsformen. *Marketing ZFP*, 2, 161-172.
- Geck, T. (2001). *Vergleich der Güte von Wahl- und Paarvergleichsmethoden in der Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Green, P. E., Krieger, A. M. & Agarwal, M. K. (1993). A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences. *Marketing Letters*, 4, 369-380.
- Green, P. E., Wind, Y. & Jain, A. K. (1972). Preference Measurement of Item Collections. *Journal of Marketing Research*, 9, 371-377.
- Hoeffler, S. (2003). Measuring Preferences for Really New Products. *Journal of Marketing Research*, 40, 406-420.
- Holling, H. (2004). Präferenzmessung im Personalmarketing durch optimal geplante Conjoint-Analysen. In K.-C. Hamborg & H. Holling (Hrsg.), *Innovative Ansätze der Personal- und Organisationsentwicklung* (53–72). Göttingen: Hogrefe.
- Holling, H. & Schwabe, R. (2002). *Effiziente Versuchsplanung in der Conjoint-Analyse – Zwischenbericht zum DFG-Projekt*. Working Paper am Lehrstuhl für Evaluations- und Forschungsmethoden, Organisationspsychologie an der Universität Münster.
- Leigh, T. W., MacKay, D. B. & Summers, J. O. (1981). On Alternative Experimental Methods for Conjoint Analysis. In K. B. Monroe (Hrsg.), *Advances in Consumer Research* (317-322). Ann Arbor: Association for Consumer Research.

- Leigh, T. W., MacKay, D. B. & Summers, J. O. (1984). Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison. *Journal of Marketing Research*, 21, 456-462.
- McCullough, J. & Best, R. (1979). Conjoint Measurement: Temporal Stability and Structural Reliability. *Journal of Marketing Research*, 16, 26-31.
- Melles, T. (1996). *Eine empirische Studie zur Optimierung der Paarvergleichsaufgabe im Rahmen der adaptiven Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Melles, T., Laumann, R. & Holling, H. (2000). *Validity and Reliability of Online Conjoint Analysis*. Working Paper presented at the Sawtooth Software Conference, Hilton Head, SC.
- Parker, B. R. & Srinivasan, V. (1976). A Consumer Preference Approach to the Planning of Rural Health-care Facilities. *Operations Research*, 24, 991-1025.
- Perry, J. (1996). Erhebungsdesign-Effekte bei der Conjoint-Analyse. *Marketing ZFP*, 2, 105-116.
- Rao, V. R. & Sattler, H. (2001). Measurement of Price Effects with Conjoint Analysis: Separating Informational and Allocative Effects of Price. In A. Gustafsson et al. (Hrsg.), *Conjoint Measurement, Methods and Applications* (47-66). Berlin: Springer.
- Segal, M. N. (1982). Reliability of Conjoint Analysis: Contrasting Data Collection Procedures. *Journal of Marketing Research*, 19, 139-143.
- Singh, J., Cuttler, L., Shin, M., Silvers, J. B. & Neuhauser, D. (1998). Medical Decision-Making and the Patient. *Medical Care*, 36, 8, 31-45.
- Weisenfeld, U. (1989). *Die Einflüsse von Verfahrensvariationen und der Art des Kaufentscheidungsprozesses auf die Reliabilität der Ergebnisse bei der Conjoint Analyse*. Berlin: Duncker und Humblot.

Korrelation von aus zwei Messungen ermittelten Attributwichtigkeiten

Akaah, I. P. (1991). Predictive Performance of Self-Explicated, Traditional Conjoint, and Hybrid Conjoint Models under Alternative Data Collection Modes. *Journal of the Academy of Marketing Science*, 19, 309-314.

Scott, J. E. & Wright, P. (1976). Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations. *Journal of Marketing Research*, 13, 211-224.

Singh, J., Cuttler, L., Shin, M., Silvers, J. B. & Neuhauser, D. (1998). Medical Decision-Making and the Patient. *Medical Care*, 36, 8, 31-45.

Literaturverzeichnis

- Acito, F. (1977). An Investigation of Some Data Collection Issues in Conjoint Measurement. In American Marketing Association (Hrsg.), *Educators' Conference Proceedings* (82-85). Chicago: AMA.
- Acito, F. & Jain, A. K. (1980). Evaluation of Conjoint Analysis Results: A Comparison of Methods. *Journal of Marketing Research*, 17, 106-112.
- Acito, F. & Olshavsky, R. W. (1980). Limits to Accuracy in Conjoint Analysis. In Association for Consumer Research (Hrsg.), *Advances in Consumer Research* 8 (313-316). Provo: ACR.
- Agarwal, M. K. (1989). How Many Pairs Should We Use in Adaptive Conjoint Analysis: An Empirical Analysis. In American Marketing Association (Hrsg.), *AMA Winter Educators' Conference Proceedings* (7-11). Chicago: AMA.
- Agarwal, M. K. (1991). An Empirical Comparison of Traditional Full-Profile Conjoint and Adaptive Conjoint Analysis. In Academy of Marketing Science (Hrsg.), *Proceedings of the 15th Conference*. Richmond: AMS.
- Agarwal, M. K. & Green, P. E. (1991). Adaptive Conjoint Analysis Versus Self-Explicated Models: Some Empirical Results. *International Journal of Research in Marketing*, 8, 141-146.
- Agarwal, M. K. & Rao, V. R. (1996). An Empirical Comparison of Consumer-Based Measures of Brand Equity. *Marketing Letters*, 7, 237-247.
- Ajzen, I. & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior*. New York: Prentice-Hall / Englewood Cliffs.
- Akaah, I. P. (1991). Predictive Performance of Self-Explicated, Traditional Conjoint, and Hybrid Conjoint Models under Alternative Data Collection Modes. *Journal of the Academy of Marketing Science*, 19, 309-314.
- Akaah, I. P. & Korgaonkar, P. K. (1983). An Empirical Comparison of the Predictive Validity of Self-Explicated, Huber-Hybrid, Traditional Conjoint, and Hybrid Conjoint Models. *Journal of Marketing Research*, 20, 187-197.
- Albrecht, J. (2000). *Präferenzstrukturmessung - Ein empirischer Vergleich der Conjoint-Analyse mit einer kompositionellen Methode*. Europäische Hochschulschriften, Frankfurt: Peter Lang.

- Allenby, G. M., Arora, N. & Ginter, J. L. (1995). Incorporating Prior Knowledge into the Analysis of Conjoint Studies. *Journal of Marketing Research*, 32, 152-162.
- Anderson, J. C. & Donthu, N. (1988). A Proximate Assessment of the External Validity of Conjoint Analysis. In Frazier et al. (Hrsg.), *AMA Educators' Proceedings 54* (287-291). Chicago: AMA.
- Baalbaki, I. B. & Malhotra, N. K. (1995). Standardization versus Customization in International Marketing: An Investigation Using Bridging Conjoint Analysis. *Journal of the Academy of Marketing Science*, 26, 182-194.
- Baier, D. & Säuberlich, F. (1997). Kundennutzenschätzung mittels individueller Hybrid-Conjointanalyse. *Zeitschrift für betriebswirtschaftliche Forschung ZfbF*, 49, 951-972.
- Basu, A. K. & Hastak, M. (1990). Multiattribute Judgements Under Uncertainty: A Conjoint Measurement Approach. *Advances in Consumer Research*, 17, 554-562.
- Bateson, J. E. G., Reibstein, D. & Boulding, W. (1987). Conjoint Analysis Reliability and Validity: A Framework for Future Research. In M. J. Houston (Hrsg.), *Review of Marketing* (451-481). Chicago: AMA.
- Beaman, A. L. (1991). An Empirical Comparison of Meta-Analytic and Traditional Reviews. *Personality and Social Psychology Bulletin*, 17, 252-257.
- Ben-Akiva, M., Bradley, M., Morikava, T., Novak, T., Oppewal, H. & Rao, V. (1994). Combining Revealed and Stated Preferences Data. *Marketing Letters*, 5, 4, 335-350.
- Ben-Akiva, M. & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT-Press.
- Bijmolt, T. H. A. & Pieters, R. G. M. (2001). Meta-Analysis in Marketing when Studies Contain Multiple Measurements. *Marketing Letters*, 12, 157-169.
- Böcker, F. (1986). Präferenzforschung als Mittel marktorientierter Unternehmensführung. *Zeitschrift für betriebswirtschaftliche Forschung zfbf*, 38, 543-574.
- Böcker, F. & Schweikl, H. (1988). Better Preference Prediction with Individualized Sets of Relevant Attributes. *International Journal of Research in Marketing*, 5, 15-24.
- Bradley, R. A. & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: The Method of Paired Comparisons. *Biometrika*, 39, 324-345.
- Brocke, M. (2006). *Präferenzmessung durch die Discrete Choice-Analyse : Effekte der Aufgabenkomplexität*. Wiesbaden: Deutscher Universitäts-Verlag.

- Bryan, S., Gold, L., Sheldon, R. & Buxton, M. (2000). Preference Measurement Using Conjoint Methods: An Empirical Investigation of Reliability. *Health Economics*, 9, 385-395.
- Bucklin, R. E. & Srinivasan, V. (1991). Determining Interbrand Substitutability Through Survey Measurement of Consumer Preference Structures. *Journal of Marketing Research*, 28, 58-71.
- Campbell, D. T. (1976). Psychometric Theory. In M. E. Dunnette (Hrsg.), *Handbook of Industrial and Organizational Psychology* (185-222). Chicago: Rand McNally.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, 56 (2), 81-105.
- Cattin, P., Gelfand, A. E. & Danes, J. (1983). A Simple Procedure for Estimation in a Conjoint Model. *Journal of Marketing Research*, 20, 29-35.
- Cattin, P., Hermet, G. & Pioche, A. (1982). Alternative Hybrid Models for Conjoint Analysis: Some Empirical Results. In R. K. Srivastava and E. D. Shocker (Hrsg.), *Analytical Approaches to Product and Market Planning: The Second Conference* (142-151). Cambridge: Marketing Science Institute.
- Cestre, G. & Darmon, R. Y. (1998). Assessing Consumer Preferences in the Context of New Product Diffusion. *International Journal of Research in Marketing*, 15, 123-135.
- Chrzan, K. (1998). *Full versus Partial Profile Choice Experiments: Aggregate and Disaggregate Comparisons*. Working Paper presented at the INFORMS Marketing Science Conference, Paris.
- Chrzan, K. & Boeger, L. (1999). *Improving Choice Predictions Using a Cutoff-Constrained Aggregate Choice Model*. Working Paper presented at the INFORMS Marketing Science Conference, Syracuse.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.
- Conrad, T. (1997). *Preisbildung mittels der Conjoint-Analyse und eines Simulationsmodells am Beispiels eines Premiumanbieters der Automobilindustrie*. Dissertation an der wirtschaftswissenschaftlichen Fakultät der Universität Tübingen.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- Cooper, H. & Hedges, L. V. (1994). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.

- Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley.
- Currim, I. S. (1982). Predictive Testing of Consumer Choice Models not Subject to Independence of Irrelevant Alternatives. *Journal of Marketing Research*, 19, 208-222.
- Dahan, E., Hauser, J. R., Simester, D. & Toubia, O. (2001). *Application and Test of Web-Based Adaptive Polyhedral Conjoint Analysis*. Working paper Sloan School of Management, MIT, Cambridge.
- Dahan, E. & Srinivasan, V. (2000). The Predictive Power of Internet-Based Product Concept Testing Using Visual Depiction and Animation. *Journal of Product Innovation Management*, 17, 99-109.
- Davey, K. S. & Elrod, T. (1991). *Predicting Shares from Preferences for Multiattribute Alternatives*. Working Paper Faculty of Business, University of Alberta, Edmonton.
- Dellaert, B., Borgers, A., Louviere, J. J. & Timmermans, H. (2001). Using Conjoint Choice Experiments to Model Consumer Choices of Product Component Packages. In A. Gustafsson et al. (Hrsg.), *Conjoint Measurement, Methods and Applications* (387-414). Berlin: Springer.
- DeSarbo, W. S., Ramaswamy, V. & Cohen, S. H. (1995). Market Segmentation With Choice-Based Conjoint Analysis. *Marketing Letters*, 6, 137-147.
- DeSarbo, W. S., Wedel, M., Vriens, M. & Ramaswamy, V. (1992). Latent Class Metric Conjoint Analysis. *Marketing Letters*, 3, 273-288.
- Ding, M., Grewal, R. & Liechty, J. (2004). *Incentive-Aligned Conjoint Analysis*. Working Paper Smeal College of Business Administration, Pennsylvania State University, Pennsylvania.
- Dorsch, M. J. & Teas, K. R. (1992). A Test of the Convergent Validity of Self-explicated and Decompositional Conjoint Measurement. *Journal of the Academy of Marketing Science*, 20, 37-48.
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E. & Cortina, J. M. (2006). A Meta-Analytic Investigation of Conscientiousness in the Prediction of Job Performance: Examining the Intercorrelations and the Incremental Validity of Narrow Traits. *Journal of Applied Psychology*, 91, 40-57.

- Eck, M. (2001). *Vergleich kompositioneller und dekompositioneller Verfahren in der multiattributiven Nutzenanalyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Eckes, T. & Six, B. (1994). Fakten und Fiktionen in der Einstellungs-Verhaltens-Forschung: Eine Meta-Analyse. *Zeitschrift für Sozialpsychologie*, 1994, 253-271.
- Elrod, T., Louviere, J. J. & Davey, K. S. (1992). An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models. *Journal of Marketing Research*, 29, 368-377.
- Eisenführ, F. & Weber, M. (1994). *Rationales Entscheiden*. (2. Aufl.). Berlin: Springer.
- Ernst, O. & Sattler, H. (2000). Multimediale versus traditionelle Conjoint-Analysen – Ein empirischer Vergleich alternativer Produktpräsentationsformen. *Marketing ZFP*, 2, 161-172.
- Finkbeiner, C. T. (1988). Comparison of Conjoint Choice Simulators. *Proceedings of the Sawtooth Software Conference*, 75-103.
- Fishbein, M. & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Boston: Addison-Wesley.
- Franke, N. (1999). High-Potentials - Conjointanalytische Identifikation und empirisches Realbild zukünftiger kaufmännischer Führungseliten. *Zeitschrift für Betriebswirtschaft*, 69, 889-911.
- Geck, T. (2001). *Vergleich der Güte von Wahl- und Paarvergleichsmethoden in der Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Gilpin, A. R. (1993). Table for Conversion of Kendall's Tau to Spearman's Rho within the Context of Measures of Magnitude of Effect for Meta-Analysis. *Educational and Psychological Measurement*, 53, 87-92.
- Glass, G. V. (1976). Primary, Secondary and Meta-Analysis of Research. *Educational Researcher*, 5, 3-8.
- Graßhoff, U., Großmann, H., Holling, H. & Schwabe, R. (2004). Optimal Designs for Main Effects in Linear Paired Comparison Models. *Journal of Statistical Planning and Inference*, 126, 361-376.
- Green, P. E. (1984). Hybrid Models for Conjoint Analysis: An Expository Review. *Journal of Marketing Research*, 21, 155-169.
- Green, P. E., Carmone, F. J. & Wind, Y. (1972). Subjective Evaluation Models and Conjoint Measurement. *Behavioral Science*, 17, 288-299.

- Green, P. E., Goldberg, S. M. & Montemayor, M. (1981). A Hybrid Utility Estimation Model for Conjoint Analysis. *Journal of Marketing*, 45, 33-41.
- Green, P. E. & Helsen, K. (1989). Cross-Validation Assessment of Alternatives to Individual-Level Conjoint Analysis: A Case Study. *Journal of Marketing Research*, 26, 346-350.
- Green, P. E., Helsen, K. & Shandler, B. (1988). Conjoint Internal Validity Under Alternative Profile Presentations. *Journal of Consumer Research*, 15, 392-397.
- Green, P. E., Kim, J. & Shandler, B. (1990). *Predictive Accuracy of Classification and Regression Trees (CART) versus Conjoint Analysis*. Auf Anforderung direkt von den Autoren zu erhalten.
- Green, P. E. & Krieger, A. M. (1988). Choice Rules and Sensitivity Analysis in Conjoint Simulators. *Journal of the Academy of Marketing Science*, 16, 114-127.
- Green, P. E. & Krieger, A. M. (1991). Segmenting Markets with Conjoint Analysis. *Journal of Marketing*, 55, 20-31.
- Green, P. E. & Krieger, A. M. (1995). Attribute Importance Weights Modification in Assessing a Brand's Competitive Potential. *Marketing Science*, 4, 253-270.
- Green, P. E., Krieger, A. M. & Agarwal, M. K. (1993). A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences. *Marketing Letters*, 4, 369-380.
- Green, P. E., Krieger, A. M. & Bansal, P. (1988). Completely Unacceptable Levels in Conjoint Analysis: A Cautionary Note. *Journal of Marketing Research*, 25, 293-300.
- Green, P. E., Krieger, A. M. & Schaffer, C. M. (1993). A Hybrid Conjoint Model with Individual-Level Interaction Estimation. *Advances in Consumer Research*, 20, 149-154.
- Green, P. E., Krieger, A. M. & Schaffer, C. M. (1993). An Empirical Test of Optimal Respondent Weighting in Conjoint Analysis. *Journal of the Academy of Marketing Science*, 21, 345-351.
- Green, P. E., Krieger, A. M. & Wind, Y. (2001). Thirty Years of Conjoint Analysis: Reflections and Prospects. *INFORMS Electronic Issue*.
- Green, P. E. & Rao, V. R. (1971). Conjoint Measurement for Quantifying Judgemental Data. *Journal of Marketing Research*, 8, 355-363.
- Green, P. E. & Srinivasan, V. (1978). Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research*, 5, 103-123.
- Green, P. E. & Srinivasan, V. (1990). Conjoint Analysis in Marketing: New Developments With Implications for Research and Practice. *Journal of Marketing*, 54, 3-19.

- Green, P. E., Wind, Y. & Jain, A. K. (1972). Preference Measurement of Item Collections. *Journal of Marketing Research*, 9, 371-377.
- Großmann, H. (2003). *Designs für Paarvergleiche in der metrischen Conjoint-Analyse*. Wiesbaden: Deutscher Universitäts-Verlag.
- Großmann, H., Holling, H., Brocke, M., Graßhoff, U. & Schwabe, R. (2005). On the Empirical Relevance of Optimal Designs for the Measurement of Preferences. In M. P. F. Berger & W. K. Wong (Hrsg.), *Applied Optimal Designs* (45-65). Chichester, UK: Wiley.
- Großmann, H., Holling, H., Graßhoff, U. & Schwabe, R. (2006). Optimal Designs for Asymmetric Linear Paired Comparisons with a Profile Strength Constraint. *Metrika*, 64, 109-119.
- Haijjer, R., Kamakura, W. & Wedel, M. (2000). Response Latencies in the Analysis of Conjoint Choice Experiments. *Journal of Marketing Research*, 37, 376-382.
- Hagerty, M. R. (1985). Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis. *Journal of Marketing Research*, 22, 168-184.
- Hagerty, M. R. (1986). The Cost of Simplifying Preference Models. *Marketing Science*, 5, 298-319.
- Hansen, D. E. & Wittink, D. R. (1995). Combining Self-Explicated Priors with Conjoint Data Using Bayesian Regression. *Marketing Letters*, 6, 63-71.
- Hartmann, A. & Sattler, H. (2002). Commercial Use of Conjoint Analysis in Germany, Austria, and Switzerland. *Research Paper on Marketing and Retailing*, Universität Hamburg.
- Hauser, J. R. & Shugan, S. M. (1978). Extended Conjoint Analysis with Intensity Measures and Computer Assisted Interviews: Applications to Telecommunication and Travel. *Advances in Consumer Research*, 5, 440-447.
- Hausrucking, G. & Helm, R. (1996). Die Bedeutung des Country-of-Origin Effekts vor dem Hintergrund der Internationalisierung von Unternehmen. *Marketing ZFP*, 4, 267-278.
- Hedges, L. V. & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Heeler, R. M., Okechuku, C. & Reid, S. (1979). Attribute Importance: Contrasting Measurements. *Journal of Marketing Research*, 16, 60-63.
- Hensel-Börner, S. & Sattler, H. (2000). Ein empirischer Validitätsvergleich zwischen der Customized Computerized Conjoint Analysis (CCC), der Adaptive Conjoint Analysis (ACA) und Self-Explicated Verfahren. *Zeitschrift für Betriebswirtschaft ZfB*, 70, 705-727.

- Hoeffler, S. (2003). Measuring Preferences for Really New Products. *Journal of Marketing Research*, 40, 406-420.
- Holling, H. (2004). Präferenzmessung im Personalmarketing durch optimal geplante Conjoint-Analysen. In K.-C. Hamborg & H. Holling (Hrsg.), *Innovative Ansätze der Personal- und Organisationsentwicklung* (53-72). Göttingen: Hogrefe.
- Holling, H., Melles, T. & Reiners, W. (1999). Evaluation verschiedener Formen von Paarvergleichen in der Conjoint-Analyse. In H. Holling & G. Gediga (Hrsg.), *Evaluationsforschung* (249-267). Göttingen: Hogrefe.
- Holling, H. & Schwabe, R. (2002). *Effiziente Versuchsplanung in der Conjoint-Analyse – Zwischenbericht zum DFG-Projekt*. Working Paper am Lehrstuhl für Evaluations- und Forschungsmethoden, Organisationspsychologie an der Universität Münster.
- Holling, H. & Schwabe, R. (2003). *Effiziente Versuchsplanung in der Conjoint-Analyse – 2. Zwischenbericht zum DFG-Projekt*. Working Paper am Lehrstuhl für Evaluations- und Forschungsmethoden, Organisationspsychologie an der Universität Münster.
- Hopkins, D. S., Larréché, J. & Massy, W. F. (1977). Constrained Optimization of a University Administrator's Preference Function. *Management Science*, 24, 4, 365-377.
- Huber, J. & Hansen, D. (1986). Testing the Impact of Dimensional Complexity and Affective Differences of Paired Concepts. In Association for Consumer Research (Hrsg.), *Advances in Consumer Research* (159-163). Provo: ACR.
- Huber, J., Wittink, D. R., Fiedler, J. A. & Miller, R. (1993). The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice. *Journal of Marketing Research*, 30, 105-114.
- Huber, J., Wittink, D. R., Johnson, R. M. & Miller, R. (1992). Learning Effects in Preference Tasks: Choice-Based versus Standard Conjoint. *Sawtooth Software Conference Proceedings*.
- Huber, J. & Zwerina, K. (1996). The Importance of Utility Balance in Efficient Choice Designs. *Journal of Marketing Research*, 33, 307-317.
- Hunter, J. E. & Schmidt, F. L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage.
- Jaccard, J., Brinberg, D. & Ackerman, L. J. (1996). Assessing Attribute Importance: A Comparison of Six Methods. *Journal of Consumer Research*, 12, 463-468.

- Jain, A. K., Acito, F., Malhotra, N. K. & Mahajan, V. (1979). A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models. *Journal of Marketing Research*, 16, 313-322.
- Johnson, R. M. (1974). Trade-off Analysis of Consumer Values. *Journal of Marketing Research*, 11, 121-127.
- Johnson, R. M. (1987). Adaptive Conjoint Analysis. In Sawtooth Software Inc. (Hrsg.), *Proceedings of the Sawtooth Software Conference 1* (253-265). Ketchum: Sawtooth Software.
- Johnson, R. M. (1989). Assessing the Validity of Conjoint Analysis. In Sawtooth Software Inc. (Hrsg.), *Proceedings of the Sawtooth Software Conference 3* (273-280). Ketchum: Sawtooth Software.
- Johnson, R. M. (1997). Individual Utilities from Choice Data: A New Method. In Sawtooth Software Inc. (Hrsg.), *Proceedings of the Sawtooth Software Conference* (191-208). Ketchum: Sawtooth Software.
- Johnson, R. M. (1999). The Joys and Sorrows of Implementing HB Methods for Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Johnson, R. M. (2000). Monotonicity Constraints in Choice-Based Conjoint with Hierarchical Bayes. *Sawtooth Software Research Paper Series*.
- Johnson, R. M., Huber, J. & Orme, B. (2005). A Second Test of Adaptive Choice-Based Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Johnson, R. M., Orme, B., Huber, J. & Pinnell, J. (2005). Testing Adaptive Choice-Based Conjoint Designs. *Sawtooth Software Research Paper Series*.
- Kamakura, W. & Ozer, M. (2001). A Multi-trait Multi-Method Validity Test of Partworth Estimates. In A. Gustafsson et al. (Hrsg.), *Conjoint Measurement, Methods and Applications* (251-277). Berlin: Springer.
- King, W. C., Hill, A. & Orme, B. (2005). The Importance Question in ACA: Can It Be Omitted?. *Sawtooth Software Research Paper Series*.
- Klemz, B. R. & Gruca, T. S. (2003). Dueling or the Battle Royale? The Impact of Task Complexity on the Evaluation of Entry Threat. *Psychology & Marketing*, 20 (11), 999-1016.
- Krantz, D. H. (1964). Conjoint Measurement: The Luce-Tukey axiomatization and some extensions. *Journal of Mathematical Psychology*, 1, 248-277.

- Krieger, A. M., Green, P. E. & Umesh, U. N. (1998). Effect of Level of Disaggregation of Conjoint Cross Validations: Some Comparative Findings. *Decision Sciences*, 29, 1047-1058.
- Kruskal, J. B. (1965). Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data. *Journal of the Royal Statistical Society, Series B*, 27, 251-263.
- Leigh, T. W., MacKay, D. B. & Summers, J. O. (1981). On Alternative Experimental Methods for Conjoint Analysis. In K. B. Monroe (Hrsg.), *Advances in Consumer Research* (317-322). Ann Arbor: Association for Consumer Research.
- Leigh, T. W., MacKay, D. B. & Summers, J. O. (1984). Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison. *Journal of Marketing Research*, 21, 456-462.
- Lenk, P. J., DeSarbo, W. S., Green, P. & Young, M. R. (1996). Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science*, 15, 173-191.
- Leptien, C. (1995). Incentives for Employed Inventors: an Empirical Analysis with Special Emphasis in the German Law for Employee's Inventions. *R&D Management*, 25, 213-225.
- Lienert, G. A. & Raatz, A. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Psychologie Verlags Union.
- Lord, F. M. & Novick, M. R. (1974). *Statistical Theories of Mental Test Scores*. Reading: Addison-Wesley Publishing Comp.
- Louviere, J. J. (1988). Conjoint Analysis Modelling of Stated Preferences. *Journal of Transport Economics and Policy*, 10, 93-119.
- Louviere, J. J., Schroeder, H., Louviere, C. H. & Woodworth, G. G. (1986). Do the Parameters of Choice Models Depend on Differences in Stimulus Presentation: Visual versus Verbal Presentation?. In Association of Consumer Research (Hrsg.), *Advances in Consumer Research* 14 (79-82). Provo: ACR.
- Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: John Wiley and Sons.
- Luce, R. D. & Tukey, J. W. (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. *Journal of Mathematical Psychology*, 1, 1-27.

- McCullough, D. (1999). The Number of Levels Effect: Psychological or Algorithmic?. *Sawtooth Software Conference Proceedings*.
- McCullough, J. & Best, R. (1979). Conjoint Measurement: Temporal Stability and Structural Reliability. *Journal of Marketing Research*, 16, 26-31.
- McLauchlan, W. G. (1991). Scaling Prior Utilities in Sawtooth Software's Adaptive Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Melles, T. (1996). *Eine empirische Studie zur Optimierung der Paarvergleichsaufgabe im Rahmen der adaptiven Conjoint Analyse*. Diplomarbeit am Fachbereich Psychologie der Universität Münster.
- Melles, T. (2000). *CALL – Conjoint Analysis Literature List*. Verfügbar unter: http://www.geocities.com/melles99/melles_e.html [31.05.2005].
- Melles, T. & Holling, H. (1998). *Einsatz der Conjoint-Analyse in Deutschland – Eine Befragung von Anwendern*. Working Paper am Lehrstuhl für Evaluations- und Forschungsmethoden, Organisationspsychologie an der Universität Münster.
- Melles, T., Laumann, R. & Holling, H. (2000). *Validity and Reliability of Online Conjoint Analysis*. Working Paper presented at the Sawtooth Software Conference, Hilton Head, SC.
- Meyer, L. (1998). *Predictive Accuracy for Conjoint Models for Data Collected by Means of World Wide Web Survey*. Verfügbar unter: <http://www.lucameyer.com/kul/title.htm> [26.08.2002].
- Molin, E., Oppewal, H. & Timmermans, H. (1999). Group-based versus Individual-based Conjoint Preference Models of Residential Preferences: A Comparative Test. *Environment and Planning A*, 31, 1935-1947.
- Molin, E., Oppewal, H. & Timmermans, H. (2002). Conjoint Modeling of Residential Group Preferences: A Comparison of the Internal Validity of Hierarchical Information Integration Approaches. *Journal of Geographical Systems*, 4, 343-358.
- Moore, W. L. (1980). Levels of Aggregation in Conjoint Analysis: An Empirical Comparison. *Journal of Marketing Research*, 17, 516-523.
- Moore, W. L., Gray-Lee, J. & Louviere, J. J. (1998). A Cross-Validity Comparison of Conjoint Analysis and Choice Models at Different Levels of Aggregation. *Marketing Letters*, 9, 195-207.
- Moore, W. L., Metha, R. B. & Pavia, T. M. (1994). A Simplified Method of Constraint Parameter Estimation in Conjoint Analysis. *Marketing Letters*, 5, 173-181.

- Müller-Hagedorn, L., Sewing, E. & Toporowski, W. (1993). Zur Validität von Conjoint-Analysen. *Zeitschrift für betriebswirtschaftliche Forschung zfbf*, 45, 123-148.
- Mulye, R. (1998). An Empirical Comparison of Three Variants of the AHP and Two Variants of Conjoint Analysis. *Journal of Behavioral Decision Making*, 11, 263-280.
- Mummalaneni, V., Dubas, K. M. & Chao, C. (1996). Chinese Purchasing Managers' Preferences and Trade-offs in Supplier Selection and Performance Evaluation. *Industrial Marketing Management*, 25, 115-124.
- Neslin, S. A. (1981). Linking Product Features to Perceptions: Self-Stated Versus Statistically Revealed Importance Weights. *Journal of Marketing Research*, 18, 80-86.
- Nickerson, C. A. E., McClelland, G. H. & Petersen, D. M. (1990). Solutions to some Problems in the Implementation of Conjoint Analysis. *Behavior Research Methods, Instruments, & Computers*, 22, 360-374.
- Niebergall, A. & Schulz, U. (1996). Evaluation von Expertenurteilen bei der Personalauswahl mittels Conjoint-Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, 40, 38-41.
- Oliphant, K., Eagle, T. C., Louviere, J. J. & Anderson, D. (1992). Cross-Task Comparison of Ratings-Based and Choice-Based Conjoint. *Proceedings of the Sawtooth Software Conference*, 383-404.
- Olkin, I. & Pratt, J. W. (1985). Unbiased Estimation of Certain Correlation Coefficients. *Annals of Mathematical Statistics*, 29, 201-211.
- Olshavsky, R. W. & Acito, F. (1980). An Information Processing Probe into Conjoint Analysis. *Decision Sciences*, 11, 451-470.
- Oppewal, H. & Timmermans, H. (1999). Modeling Consumer Perception of Public Space in Shopping Centers. *Environment and Behavior*, 31, 45-65.
- Orme, B. K., Alpert, M. I. & Christensen, E. (1997). Assessing the Validity of Conjoint Analysis – Continued. *Sawtooth Software Research Paper Series*.
- Orme, B. K. & Baker, G. C. (2000). Comparing Hierarchical Bayes Draws and Randomized First Choice for Conjoint Simulations. *Sawtooth Software Research Paper Series*.
- Orme, B. K. & Heft, M. A. (1999). Predicting Actual Sales with CBC: How Capturing Heterogeneity Improves Results. *Sawtooth Software Research Paper Series*.
- Orme, B. K. & King, C. W. (1998). Conducting Full-Profile Conjoint Analysis over the Internet. *Sawtooth Software Research Paper Series*.

- Page, A. L. & Rosenbaum, H. F. (1987). Redesigning Product Lines with Conjoint Analysis: How Sunbeam Does It. *Journal of Product Innovation Management*, 4, 120-137.
- Parker, B. R. & Srinivasan, V. (1976). A Consumer Preference Approach to the Planning of Rural Health-care Facilities. *Operations Research*, 24, 991-1025.
- Patterson, M. & Frazier, C. (2000). *Cutoff Constrained Discrete Choice Models*. Working Paper presented at the Sawtooth Software Conference, Hilton Head, SC.
- Pavia, T. M., Moore, W., Mehta, R. B. & Ottum, B. D. (1996). *Can We Simplify the Conjoint Task: An Investigation of the Effect of Differing Numbers of Calibration Attributes on Validations*. Working Paper David Eccles School of Business, University of Utah, Salt Lake City.
- Perry, J. (1996). Erhebungsdesign-Effekte bei der Conjoint-Analyse. *Marketing ZFP*, 2, 105-116.
- Pullman, M. E., Dodson, K. J. & Moore, W. M. (1999). A Comparison of Conjoint Methods When There Are Many Attributes. *Marketing Letters*, 10, 125-138.
- Rao, V. R. & Sattler, H. (2001). Measurement of Price Effects with Conjoint Analysis: Separating Informational and Allocative Effects of Price. In A. Gustafsson et al. (Hrsg.), *Conjoint Measurement, Methods and Applications* (47-66). Berlin: Springer.
- Reiners, W. (1996). *Multiattributive Präferenzstrukturmodellierung durch die Conjoint Analyse*. Hamburg: LIT-Verlag.
- Robinson, P. J. (1980). Applications of Conjoint Analysis to Pricing Problems. In D. B. Montgomery & D. R. Wittink (Hrsg.), *Market Measurement and Analysis: Proceedings of the 1979 ORSA/TIMS Conference on Marketing* (183-205). Cambridge: Marketing Science Institute.
- Rosenthal, R. (1994). Parametric Measures of Effect Size. In H. Cooper & L. V. Hedges (Hrsg.), *The Handbook of Research Synthesis* (231-244). New York: Russell Sage Foundation.
- Rosenthal, R. & Rubin, D. B. (1978). Interpersonal Expectancy Effects: The First 345 Studies. *The Behavioral and Brain Sciences*, 3, 377-415.
- Rosenthal, R. & Rubin, D. B. (1982). Comparing Effect Sizes of Independent Studies. *Psychological Bulletin*, 92, 500-504.
- Rosko, M. D., DeVita, M., McKenna, W. F. & Walker, L. R. (1985). Strategic Marketing Applications of Conjoint Analysis: An HMO Perspective. *Journal of Health Care Marketing*, 5 (4), 27-38.

- Ryan, M. (1999). Using Conjoint Analysis to Take Account of Patient Preferences and Go Beyond Health Outcomes: An Application to in Vitro Fertilisation. *Social Science & Medicine*, 48, 535-546.
- Safizadeh, H. M. (1989). The Internal Validity of the Trade-off Method of Conjoint Analysis. *Decision Sciences*, 20, 451-461.
- Sattler, E. (1994). Die Validität von Produkttests. *Marketing ZFP*, 1, 31-41.
- Sattler, E., Gedenk, K. & Hensel-Börner, S. (2002). *Bandbreiten-Effekte bei multiattributiven Entscheidungen: Ein empirischer Vergleich von Verfahren zur Bestimmung von Eigenschaftsgewichten*. Working Paper am Institut für Handel und Marketing der Universität Hamburg.
- Sattler, E. & Hensel-Börner, S. (2001). A Comparison of Conjoint Measurement with Self-Explicated Approaches. In A. Gustafsson et al. (Hrsg.), *Conjoint Measurement, Methods and Applications* (121-134). Berlin: Springer.
- Sawtooth Software (1997). Individual Choice Estimation (ICE) Modul for Choice-Based Conjoint. *Sawtooth Software Technical Paper Series*.
- Sawtooth Software (1999). ACA/Hierarchical Bayes - Technical Paper. *Sawtooth Software Technical Paper Series*.
- Sawtooth Software (2000). CBC/Hierarchical Bayes Analysis - Technical Paper (version 3.0). *Sawtooth Software Technical Paper Series*.
- Sawtooth Software (2003). ACA/Hierarchical Bayes v2.0 - Technical Paper. *Sawtooth Software Technical Paper Series*.
- Scott, J. E. & Wright, P. (1976). Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations. *Journal of Marketing Research*, 13, 211-224.
- Schmidt, F. L. & Hunter, J. E. (1977). Development of a General Solution to the Problem of Validity Generalization. *Journal of Applied Psychology*, 62, 529-540.
- Schulze, R. (2004). *Meta-analysis: A Comparison of Approaches*. Cambridge, MA: Hogrefe & Huber Publishers.
- Schulze, R., Holling, H. & Böhning, D. (2003). *Meta-analysis: New Developments and Applications in Medical and Social Sciences*. Seattle, WA: Hogrefe & Huber.
- Schweickl, H. (1985). *Computergestützte Präferenzanalyse mit individuell wichtigen Produktmerkmalen*. Berlin: Duncker & Humblot.

- Scott, J. E. & Wright, P. (1976). Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations. *Journal of Marketing Research*, 13, 211-224.
- Segal, M. N. (1982). Reliability of Conjoint Analysis: Contrasting Data Collection Procedures. *Journal of Marketing Research*, 19, 139-143.
- Sentis, K. & Li, L. (2002). One Size Fits All or Custom Tailored: Which HB Fits Better?. *Sawtooth Software Research Paper Series*.
- Sheluga, D. A., Jaccard, J. & Jacoby, J. (1979). Preference, Search, and Choice: An Integrative Approach. *Journal of Consumer Research*, 6, 166-176.
- Singh, J., Cuttler, L., Shin, M., Silvers, J. B. & Neuhauser, D. (1998). Medical Decision-Making and the Patient. *Medical Care*, 36, 8, 31-45.
- Six, B. & Eckes, T. (1996). Metaanalysen in der Einstellungs-Verhaltens-Forschung. *Zeitschrift für Sozialpsychologie*, 1996, 7-17.
- Slovic, P., Fleissner, D. & Bauman, W. S. (1972). Analyzing the Use of Information in Investment Decision Making: A Methodological Proposal. *Journal of Business*, 45, 283-301.
- Srinivasan, V. (1988). A Conjunctive-Compensatory Approach to the Self-Explication of Multiattributed Preferences. *Decision Sciences*, 19, 295-305.
- Srinivasan, V., Jain, A. K. & Malhotra, N. K. (1983). Improving Predictive Power of Conjoint Analysis by Constrained Parameter Estimation. *Journal of Marketing Research*, 20, 433-438.
- Srinivasan, V. & Park, C. S. (1997). Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement. *Journal of Marketing Research*, 34, 286-291.
- Srinivasan, V. & Shocker, A. D. (1973). Estimating the Weights for Multiple Attributes in a Composite Criterion Using Pairwise Judgements. *Psychometrika*, 38, 473-493.
- Takane, Y., Young, F. W. & de Leeuw, J. (1976). Non-metric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 42, 7-67.
- Teas, K. R. (1987). Magnitude Scaling of the Dependent Variable in Decompositional Multiattribute Preference Models. *Journal of the Academy of Marketing Science*, 15, 64-73.

- Teichert, T. (2000). Auswirkungen von Verfahrensalternativen bei der Erhebung von Präferenzurteilen. *Marketing ZFP*, 2, 145-159.
- Teichert, T. (2001). *Nutzenschätzung in Conjoint-Analysen: theoretische Fundierung und empirische Aussagekraft*. Wiesbaden: Gabler.
- Toubia, O., Hauser, J. R. & Simester, D. I. (2003). *Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis*. Working Paper Sloan School of Management, MIT, Cambridge.
- Tscheulin, D. K. (1996). Determinanten der Validität der Conjoint-Analyse. In A. von Ahnen & T. Czieskowsky (Hrsg.), *Marketing und Marktforschung* (585-598). Hamburg: LIT-Verlag.
- Tscheulin, D. K. & Blaimont, C. (1993). Die Abhängigkeit der Prognosegüte von Conjoint-Studien von demographischen Probanden-Charakteristika. *Zeitschrift für Betriebswirtschaft ZfB*, 63 (8), 839-847.
- Tversky, A. (1967). A General Theory of Polynomial Conjoint Measurement. *Journal of Mathematical Psychology*, 4, 1-20.
- Van der Lans, I. A., Wittink, D. R., Huber, J. & Vriens, M. (1992). Within- and across Attribute Constraints in ACA and Full Profile Conjoint Analysis. *Sawtooth Software Research Paper Series*.
- Voeth, M. (2000). *Nutzenmessung in der Kaufverhaltensforschung: Die hierarchisch individualisierte Limit Conjoint-Analyse (HILCA)*. Wiesbaden: Deutscher Universitätsverlag.
- Vriens, M. (1995). *Conjoint Analysis in Marketing: Developments in Stimulus Representation and Segmentation Methods*. Capelle, NL: Labyrinth Publications.
- Vriens, M., Oppewal, H. & Wedel, M. (1998). Ratings-based versus Choice-based Latent Class Conjoint Models - An Empirical Comparison. *Journal of Market Research Society*, 40 (3), 237-248.
- Weisenfeld, U. (1989). *Die Einflüsse von Verfahrensvariationen und der Art des Kaufentscheidungsprozesses auf die Reliabilität der Ergebnisse bei der Conjoint Analyse*. Berlin: Duncker und Humblot.
- Wilkie, W. L. & Pessemier, E. A. (1973). Issues in Marketing's Use of Multiattribute Attitude Models. *Journal of Marketing Research*, 10, 428-441.

- Williams, M. L., McDaniel, M. A. & Nguyen, N. T. (2006). A Meta-Analysis of the Antecedents and Consequences of Pay Level Satisfaction. *Journal of Applied Psychology*, 91, 392-413.
- Willms, R. (2001). *Generierung von Anforderungsprofilen mittels Conjoint-Analyse und Discrete Choice-Analyse*. Diplomarbeit an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld.
- Wiltinger, K. (1997). Personalmarketing auf der Basis von Conjoint Analysen. *Zeitschrift für Betriebswirtschaft*, 67, Ergänzungsheft (3), 55-78.
- Wind, J., Green, P. E., Shifflet, D. & Scarbrough, M. (1989). Courtyard by Marriott: Designing a hotel facility with consumer-based marketing models. *Interfaces*, 19, 25-47.
- Wittink, D. R. (2000). *Predictive Validity of Conjoint Analysis*. Working Paper presented at the Sawtooth Software Conference, Hilton Head, SC.
- Wittink, D. R. & Cattin, P. (1989). Commercial Use of Conjoint Analysis: An Update. *Journal of Marketing*, 53, 91-96.
- Wittink, D. R., Krishnamurthi, L. & Nutter, J. B. (1982). Comparing Derived Importance Weights Across Attributes. *Journal of Consumer Research*, 8, 471-474.
- Wittink, D. R., Krishnamurthi, L. & Reibstein, D. J. (1989). The Effect of Differences in the Number of Attribute Levels on Conjoint Results. *Marketing Letters*, 2, 113-123.
- Wittink, D. R. & Seetharaman, P. B. (1999). A Comparison of Alternative Solutions to the Number-of-Levels Effect. *Sawtooth Software Conference Proceedings*, 269-281.
- Wittink, D. R., Vriens, M. & Burhenne, W. (1994). Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections. *International Journal of Research in Marketing*, 11, 41-52.
- Wright, P. (1975). Consumer Choice Strategies: Simplifying vs. Optimizing. *Journal of Marketing Research*, 12, 60-67.
- Zwerina, K. (1997). *Discrete Choice Experiments in Marketing: Use of Priors in Efficient Choice Designs and Their Application to Individual Preference Measurement*. Heidelberg: Physica.

Lebenslauf

Marcus Heidbrink, geb. am 03. September 1973 in Münster

Ausbildung

- | | |
|-------------------|---|
| Seit 10/2000 | Aufbaustudium: Psychologie (Hauptfach) und Politikwissenschaften (Nebenfach) an der Westfälischen Wilhelms-Universität Münster, Angestrebter Abschluss: Dr. phil. |
| 10/1994 – 04/1999 | Studium: Psychologie an der Westfälischen Wilhelms-Universität Münster, Abschluss: Diplom-Psychologe |
| 07/1984 – 06/1993 | Schule: Wilhelm Hittorf Gymnasium Münster, Abschluss: Abitur |
| 08/1980 – 06/1984 | Grundschule: Gottfried von Cappenberg Grundschule Münster |

Beruflicher Werdegang

- | | |
|-------------------|---|
| Seit 01/2001 | Selbstständiger Unternehmensberater für Organisations- und Führungskräfteentwicklung, Köln |
| 04/1999 – 12/2000 | Unternehmensberater bei der Kienbaum Management Consultants GmbH / Geschäftsbereich Personalentwicklung, Düsseldorf |

Köln, 03. August 2006

Marcus Heidbrink