

Empirische Sonderpädagogik, 2017, Nr. 1, S. 3-18
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Adaptive Behavior Assessment System II – Eine erste Überprüfung der psychometrischen Eigenschaften der deutschen Erwachsenenversion

Carmen Zurbriggen & Dagmar Orthmann Bless

Universität Freiburg (Schweiz)

Zusammenfassung

Adaptive Kompetenzen sind eines der drei Definitionskriterien für geistige Behinderung. An entsprechenden Assessmentinstrumenten mangelt es jedoch bislang im deutschen Sprachraum. In diesem Beitrag erfolgt deshalb eine erste Überprüfung der psychometrischen Eigenschaften einer deutschen Übersetzung der Erwachsenenversion des Adaptive Behavior Assessment System II (ABAS II). Basierend auf einer Stichprobe von 299 Teilnehmenden ($M_{Alter} = 40.95$ Jahre, $SD_{Alter} = 14.90$ Jahre) werden die faktorielle Validität des ABAS II, die interne Konsistenz bzw. Reliabilität der Skalen sowie die Funktionsweise der Items überprüft. Die Ergebnisse konfirmatorischer Faktorenanalysen mit kategorial geordneten Indikatoren liefern Hinweise dafür, dass mit dem ABAS II adaptive Kompetenzen generell sowie die zehn Subdimensionen – nicht aber die drei Bereiche – erfasst werden. Die zehn Skalen weisen eine hohe interne Konsistenz auf. Analysen im Rahmen der Item-Response-Theorie verweisen auf eine gute Diskriminationsfähigkeit der Skalen, insbesondere im unteren Bereich des jeweiligen Merkmals, sowie auf unterschiedliche Schwierigkeitsgrade der Items. Bei einigen Items deuten die Ergebnisse ferner darauf hin, dass gewisse Antwortkategorien zusammengefasst werden sollten. Abschließend werden das Potenzial des ABAS II und Ansatzpunkte zu dessen Weiterentwicklung diskutiert.

Schlüsselwörter: Adaptive Kompetenzen, Assessment, geistige Behinderung, Erwachsene, Item-Response-Theorie

Adaptive Behavior Assessment System II – A first examination of the psychometric properties of the German adult version

Abstract

Adaptive behavior is one of three diagnostic criteria of intellectual disability. However, the lack of instruments for the corresponding assessment in German speaking countries is apparent. In this paper, thus, the psychometric properties of a German translation of the adult version of the Adaptive Behavior Assessment System (ABAS II) are tested. Based on a sample of 299 participants ($M_{age} = 40.95$ years, $SD_{age} = 14.90$ years), the factorial validity of the ABAS II, the internal consistency or reliability of the scales and the functioning of items are examined. Results of confirmatory categorical factor analyses suggest that the ABAS II measures adaptive behavior in general as well as the ten subdimensions, but not the three domains. The ten scales show high internal consistency. Item response theory analyses indicate that the scales in particular discriminate in the lower regions of the latent trait, and that the items have varying difficulty levels.

Moreover, the results imply that certain response categories of several items should be merged. To conclude, potential of the ABAS II and areas of its further development are discussed.

Key words: adaptive behavior, intellectual disability, assessment, adults, item response theory

Adaptive Kompetenzen: Definition, Bedeutung und Assessment

Adaptive Kompetenzen (adaptive behavior) sind konzeptuelle, soziale und praktische Fähigkeiten, welche im Laufe des Lebens erworben werden (z.B. Heber, 1959; Tassé, 2013). Sie manifestieren sich in der aktuellen Performanz in konkreten Lebenssituationen, z.B. in Bezug auf Kommunikation, Orientierung in der Gemeinschaft, Wohnen, Selbstfürsorge und soziale Anpassung. Sie bestimmen in entscheidendem Umfang das Ausmaß an persönlicher Autonomie sowie die Bewältigung alters- und kulturtypischer gesellschaftlicher Anforderungen. Adaptive Kompetenzen sind damit für die Lebensbewältigung des Individuums zentral, insbesondere im Erwachsenenalter.

Das Assessment adaptiver Kompetenzen ist vor allem bei Behinderungen und Beeinträchtigungen der Persönlichkeitsentwicklung evident. Seit Ende der 1950er Jahre werden adaptive Kompetenzen als eines der Definitionskriterien für geistige Behinderung (intellectual disability) aufgeführt (z.B. American Association on Intellectual and Developmental Disabilities [AAIDD], 2010; American Psychiatric Association [APA], 2013a). In einschlägigen Definitionen wurden sie dem traditionsreicheren Konstrukt der Intelligenz zugesellt und gegenüber diesem oft implizit oder explizit als zweitrangig angesehen. Im deutschen Sprachraum blieben sie im Bereich des Assessments häufig sogar unberücksichtigt. Neuere Entwicklungen werten den Stellenwert der adaptiven Kompetenzen im Rahmen der Diagnostik von geistiger Behinderung deutlich auf. So fordern z.B. Tassé et al. (2016): „the constructs of intellectual functioning and adaptive behavior and their

assessment must be weighed equally and considered jointly in the diagnosis of ID“ (S. 382). Im DSM-5 werden neu die Schweregrade der Beeinträchtigung anhand der Ausprägung adaptiver Kompetenzen und nicht mehr wie bisher anhand des Intelligenzquotienten bestimmt (APA, 2013a; 2013b). Begründet wird die neue Gewichtung einerseits mit messtechnischen Problemen bei der Differenzialdiagnostik im Bereich starker Einschränkungen der Intelligenz. Andererseits ergeben sich Art und Ausmaß von entwicklungsförderlichem, individuellem Support eher aus der Diagnostik der adaptiven Kompetenzen als aus einer Intelligenzdiagnostik. Der letztgenannte Aspekt gewinnt im Zusammenhang mit Veränderungen im Hilfesystem von einer fallpauschalisierten hin zu einer individualisierten Gestaltung von Leistungen zunehmend an Bedeutung (Orthmann Bless & Zurbriggen, 2017).

Für das Assessment adaptiver Kompetenzen steht eine Reihe von Verfahren für alle Altersgruppen zur Verfügung (vgl. z.B. Tassé, 2013). Standardisierte, an der Gesamtpopulation normierte Verfahren bieten dabei den Vorteil, beiden genannten Anwendungsbereichen, nämlich einerseits der Diagnosestellung einer geistigen Behinderung und andererseits der individuellen Hilfeplanung und Unterstützung dienen zu können.

Das Adaptive Behavior Assessment System

Das Adaptive Behavior Assessment System II (ABAS II; Harrison & Oakland, 2008) ist ein etabliertes, an der Gesamtpopulation normiertes Verfahren mit guten psychometrischen Eigenschaften (Burns, 2005). Es

existiert in verschiedenen Formen, welche insgesamt ein Altersspektrum von 0 bis 89 Jahren sowie die Anwendung durch verschiedene Beurteilende abdecken.

Der ABAS II misst adaptive Kompetenzen in folgender Struktur (Tabelle 1). Auf einer ersten Ebene werden zehn verschiedene Dimensionen, bestehend aus insgesamt 239 Items, unterschieden. Auf einer zweiten Ebene werden diese Dimensionen zu drei Bereichen, nämlich konzeptuellen, sozialen und praktischen adaptiven Kompetenzen zusammengefasst. Auf einer dritten Ebene werden die drei Bereiche zu einem Gesamtwert adaptiver Kompetenzen, dem sog. GAL-Wert (general adaptive level) zusammengeführt. Auf Ebene der zehn Dimensionen werden Standardwerte ($M = 10$, $SD = 3$) ermittelt. Für die drei Bereiche werden sog. Mischwerte ($M = 100$, $SD = 15$) und auf Ebene des Gesamtwertes der genannte GAL-Wert ($M = 100$, $SD = 15$) bestimmt (Tabelle 1).

Das Instrument entspricht mit dieser Grundstruktur der Terminologie und dem Konzept der adaptiven Kompetenzen von AAIDD und APA. Die Mehrdimensionalität der adaptiven Kompetenzen sowie deren

heute gebräuchliche Gruppierung in die drei Bereiche (konzeptuelle, soziale und praktische adaptive Kompetenzen) gehen bereits auf Heber (1959) zurück und haben sich im Laufe der Zeit konzeptionell kaum verändert. Zur Prüfung der dreifaktoriellen Struktur adaptiver Kompetenzen liegen einige ältere Studien vor (zusammenfassend bei Tassé et al., 2012). Bisher wurden allerdings keine alternativen Strukturprüfungen vorgenommen, beispielsweise hinsichtlich der zehn Dimensionen. Die Erwachsenenversion des ABAS II wurde zudem noch nicht speziell im Hinblick auf die psychometrischen Eigenschaften auf Itemebene untersucht.

Deutsche Übersetzung des ABAS II

Für das Erwachsenenalter gibt es bisher keine offizielle deutschsprachige Version des ABAS II. Deutsche Übersetzungen verschiedener Formen des Originals (für Kinder, Jugendliche und junge Erwachsene) wurden in Forschungsprojekten eingesetzt (z.B. Sermier Dessemontet et al., 2012). Für die

Tabelle 1: Struktur der adaptiven Kompetenzen gemäß ABAS II

10 Dimensionen	3 Bereiche	1 Gesamtwert
Kommunikation	Konzeptuelle Kompetenzen (KON)	Gesamtwert adaptiver Kompetenzen (general adaptive level; GAL)
Funktionale akademische Fähigkeiten		
Selbststeuerung		
Soziale Anpassung	Soziale Kompetenzen (SOZ)	
Freizeitverhalten		
Orientierung in der Gemeinschaft	Praktische Kompetenzen (PR)	
Wohnen		
Gesundheit und Sicherheit		
Selbstfürsorge		
Arbeit		
Standardwerte $M = 10$, $SD = 3$	Mischwerte $M = 100$, $SD = 15$	GAL-Wert $M = 100$, $SD = 15$

nachfolgende Analyse wurde der ABAS II für Erwachsene im Alter von 16 bis 89 Jahren in einer eigenen, deutschsprachigen Übersetzung (Orthmann Bless, 2013) verwendet, welche sich bereits in verschiedenen Forschungsprojekten praktisch bewährt hat (Orthmann Bless & Hellfritz, 2016; Orthmann Bless & Zurbruggen, 2017). Eine Prüfung der psychometrischen Eigenschaften für dieses Instrument steht noch aus.

Inzwischen wurde eine neue Version des ABAS (ABAS III; Harrison & Oakland, 2015) in englischer Sprache veröffentlicht. Diese ist in ihrer Struktur (Anzahl der Items, drei Bereiche, zehn Dimensionen) gegenüber dem ABAS II unverändert. 78-88% der Items sind identisch, die anderen Items wurden inhaltlich modernisiert. Zudem wurde das Instrument für die USA neu normiert. Die Veränderungen des ABAS III gegenüber dem ABAS II sind im Hinblick auf die hier angestrebte Prüfung psychometrischer Eigenschaften des Instrumentes von untergeordneter Bedeutung.

Zielsetzung

Zentrale Zielsetzung dieses Beitrages ist eine erste Überprüfung der deutschen Erwachsenenversion des Adaptive Behavior Assessment System II (ABAS II). Die Überprüfung soll folgende Kriterien betreffen: (1.) die faktorielle Validität und Dimensionalität des ABAS II, (2.) die interne Konsistenz bzw. Reliabilität der Skalen, (3.) die Funktionsfähigkeit der Skalen und Items auf unterschiedlichen Levels des jeweils gemessenen Konstrukts. Mit der Prüfung der genannten psychometrischen Eigenschaften sind Aussagen zur Qualität von Messungen im Rahmen des Assessments adaptiver Kompetenzen mit dem ABAS II (Erwachsenenversion) möglich.

Methode

Stichprobe

Für die Untersuchung wurden Institutionen aus der Deutschschweiz angefragt, welche Erwachsenen mit einer geistigen Behinderung eine Begleitung und Förderung im Alltag bieten. Die 13 kontaktierten Institutionen (mit stationärer Wohnform) sowie alle 11 Wohnschulen der Deutschschweiz willigten in die Untersuchung ein.

Bei der Auswahl der Teilnehmenden wurde innerhalb der Institutionen auf Varianz in Bezug auf Geschlecht, Alter und Ausprägung an Unterstützungsbedarf geachtet. Die ausgewählten Personen wurden mündlich und mit zusätzlicher Unterstützung von Piktogrammen oder schriftlich in vereinfachter Sprache über die Untersuchung informiert und um Teilnahme ersucht. Die Teilnehmenden sowie deren gesetzliche Vertretungen erhielten adäquate Informationen zum Datenschutz und unterzeichneten eine entsprechende Einverständniserklärung.

Die Stichprobe umfasst 299 Teilnehmende (davon 51.5% weiblich) aus 24 Institutionen in 14 Kantonen. Zum Zeitpunkt der Erhebung betrug das Alter im Durchschnitt 40.95 Jahre ($SD = 14.90$), wobei die jüngste Person 18 und die älteste 83 Jahre alt war (Näheres zur Stichprobe bei Orthmann Bless & Zurbruggen, 2017).

Messinstrument und Durchführung

Die adaptiven Kompetenzen der Teilnehmenden wurden mit der deutschen Übersetzung der Erwachsenenversion des ABAS II (Harrison & Oakland, 2008; Orthmann Bless, 2013) gemessen. Als Beurteilende fungierte jeweils eine professionelle Betreuungsperson aus der Institution. Das Durchschnittsalter der Betreuungspersonen (davon 64% weiblich) betrug 43 Jahre ($SD = 11.4$ Jahre).

Alle 239 Items sind auf einer konkreten Verhaltensebene operationalisiert. Es wird jeweils in standardisierter Weise beurteilt,

wie oft die Person eine Verhaltensweise korrekt und ohne Hilfe ausführt, und zwar dann, wenn diese Verhaltensweise angepasst und erforderlich ist. Das Antwortformat ist eine vierstufige Likert-Skala mit den Ausprägungen 0 („ist nicht fähig, das Verhalten zu zeigen“), 1 („zeigt Verhalten nie oder fast nie, wenn es erforderlich ist“), 2 („... manchmal, wenn es erforderlich ist“) und 3 („... immer oder fast immer, wenn es erforderlich ist“).

Testmodell

Das Antwortformat des ABAS II verlangt für die Untersuchung der psychometrischen Güte der Items nach einem polytomen test-theoretischen Modell. Ein geeignetes und weitläufig verwendetes Modell der Item-Response-Theorie (IRT) ist das Graded-Response-Modell (GRM; Samejima, 1969, 1997; für eine Anwendung vgl. z.B. Tassé et al., 2016).

Das GRM ist ein zweiparametrisches logistisches Modell, bei dem jedes Item (i) anhand eines Diskriminationsparameters (a_i) und j ($j = \text{Anzahl Kategorien} - 1$) Schwellen- oder Schwierigkeitsparametern (b_{ij}) spezifiziert wird. Der Diskriminationsparameter (a_i) beschreibt die nichtlineare Beziehung zwischen der (beobachteten) Item-Antwort und dem (unbeobachteten) latenten Merkmal θ und verdeutlicht damit die Fähigkeit eines Items, zwischen Personen mit unterschiedlichen Merkmalsausprägungen trennen zu können. Der Schwierigkeitsparameter (b_{ij}) wiederum gibt die Ausprägung auf dem latenten Merkmal θ an, welche benötigt wird, um mit einer Wahrscheinlichkeit von 50% über der Schwelle j zu antworten.

Im Rahmen des GRM entspricht die Wahrscheinlichkeit einer Person, mit einer Merkmalsausprägung θ in einer bestimmten Kategorie k eines bestimmten Items zu antworten, der Differenz zwischen der kumulativen Wahrscheinlichkeit einer Antwort in derselben Kategorie oder höher und der kumulativen Wahrscheinlichkeit einer Antwort in der nächsten Kategorie oder höher.

Ein Ziel des GRM ist nun, die Schwellen zwischen den Antwortkategorien auf dem Kontinuum des latenten Merkmals zu lokalisieren. Hierzu wird ein Item als eine Serie von $k - 1$ Dichotomien betrachtet (0 vs. 1, 2, 3; 0, 1 vs. 2, 3; 0, 1, 2 vs. 3). Für jede dieser Dichotomien wird ein zweiparametrisches Modell mit demselben Diskriminationsparameter innerhalb eines Items geschätzt. Für die tatsächliche Antwortwahrscheinlichkeit in einer bestimmten Kategorie wird schließlich die Differenz der kumulativen Wahrscheinlichkeit für angrenzende Kategorien berechnet (Näheres z.B. bei De Ayala, 2009; Embretson & Reise, 2000).

Analysen

Mit knapp 300 Personen ist die Stichprobe für die Zielpopulation zwar verhältnismäßig groß. Für die Überprüfung bestimmter psychometrischer Eigenschaften muss sie jedoch als (zu) klein eingestuft werden, so dass bei einigen Analyseschritten Anpassungen nötig waren.

(1.) Faktorielle Validität und Dimensionalität. Die Überprüfung der faktoriellen Struktur der deutschen Erwachsenenversion des ABAS II erfolgte anhand konfirmatorischer Faktorenanalysen mit kategorial geordneten Indikatoren (confirmatory categorical factor analysis; CCFA) in *Mplus* Version 7.4 (Muthén & Muthén, 1998-2015). Die Parameter wurden mit dem Weighted Least Square Means and Variances Estimator (WLSMV) geschätzt. Aufgrund der Stichprobengröße war eine Überprüfung der faktoriellen Struktur mit allen 239 Items des ABAS II allerdings nicht möglich. Deshalb wurden pro Skala jene vier Items mit der höchsten Faktorladung ausgewählt (gemäß Spezifizierung eines Messmodells für jede der zehn postulierten Dimensionen). Von der Bildung von Item-Päckchen (item parcels) wurde abgesehen, da die Dimensionalität - als Voraussetzung für die Art der Päckchenbildung (vgl. Little, Cunningham, Shahar & Widaman, 2002) - nicht überprüft werden

konnte respektive Gegenstand des Untersuchungsschrittes war. Im Einzelnen wurden folgende Modelle getestet: 1. CCFA mit einem generellen Faktor; 2. CCFA mit drei Faktoren erster Ordnung (first-order) gemäß den drei postulierten Bereichen konzeptuelle, soziale und praktische Kompetenzen; 3. hierarchische CCFA mit drei first-order Faktoren gemäß den drei Bereichen und einem übergeordneten Faktor (second-order); 4. CCFA mit zehn first-order Faktoren gemäß den zehn Dimensionen; 5. hierarchische CCFA mit zehn first-order Faktoren und einem second-order Faktor; 6. bi-faktorielle CCFA mit zehn Faktoren und einem generellen Faktor mit Faktorladungen auf alle

Items (Modelle 1-6, s. Abbildung 1). Sowohl bei den hierarchischen Modellen als auch beim bi-faktoriellen Modell wird davon ausgegangen, dass den Antworten zu allen Items ein globales Konstrukt zugrunde liegt, während die first-order Modelle auf der Annahme von distinktiven Facetten ohne gemeinsamem Kern beruhen. Beim bi-faktoriellen Modell wird zudem unmittelbar getestet, wie groß jener Anteil ist, welcher ausschließlich durch die spezifischen Faktoren erklärt werden kann (Morin, Arens & Marsh, 2016).

Zur Beurteilung der Modellgüte wurden neben dem χ^2 -Modelltest die drei folgenden approximativen Fit-Indizes herangezogen

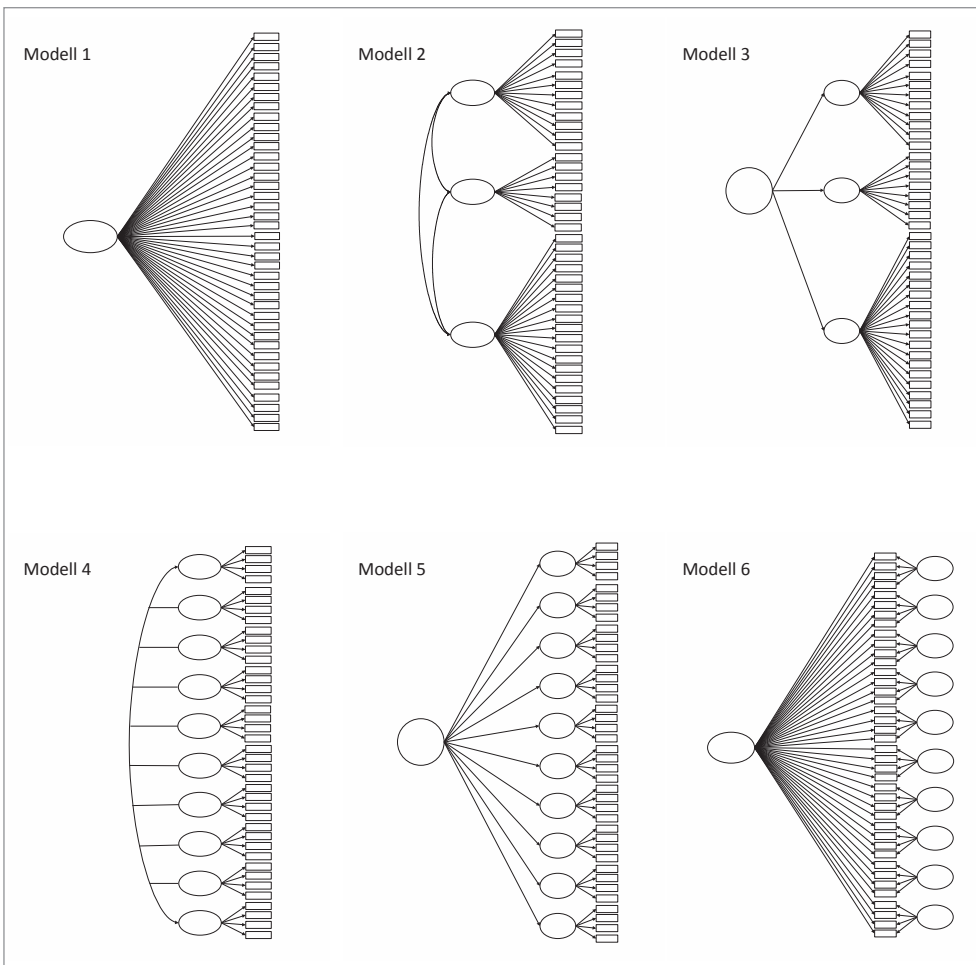


Abbildung 1: Modelle zur Überprüfung der Faktorenstruktur des ABAS II

gen, welche unterschiedliche und sich gegenseitig ergänzende Informationen zur Modellpassung abgeben (Beauducel & Wittmann, 2005): Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA) und Weighted Root Mean Square Residual (WRMR). Als Richtwerte für eine gute Übereinstimmung zwischen dem postulierten Modell und den Daten gelten CFI-Werte um .95 oder höher, RMSEA-Werte um .06 oder tiefer sowie WRMR-Werte von 1.0 oder tiefer (Hu & Bentler, 1999; vgl. auch Wang & Wang, 2012). Zusätzliche Informationen für die Beurteilung der Modellgüte und der Dimensionalität liefern die Höhe der Faktorladungen sowie Angaben über allfällige Fehlpassungen der Modelle (Heene, Hilbert, Draxler, Ziegler & Bühner, 2011; Ziegler & Hagemann, 2015). Durch die Reduktion von Items bieten die CCFA nur ansatzweise Aussagen zur Frage der Unidimensionalität aller Items einer Skala. Gleichwohl liefern sie erste Hinweise zur Faktorenstruktur und der Dimensionalität der deutschen Erwachsenenversion des ABAS-II. Durch die formale Ähnlichkeit von CCFA und GRM lassen sich des Weiteren die Parameterschätzungen ineinander übertragen (z.B. Wirth & Edwards, 2007).

(2.) Interne Konsistenz bzw. Reliabilität.

Als Maße der internen Konsistenz der Skalen wurden (Guttman-)Cronbachs Alpha (α) und McDonalds Omega (ω) herangezogen. Ein Vorteil des Koeffizienten ω ist, dass er auch bei Modellen τ -kongenerischer Variablen als Reliabilität interpretiert werden kann, während der Koeffizient α einer unteren Grenze entspricht (Dunn, Baguley & Brunson, 2014; McDonald, 2013). Cronbachs α wurde in SPSS (Version 24) und McDonalds ω im Rahmen von CCFA in *Mplus* gerechnet.

(3.) Itemanalysen. Die Funktionsfähigkeit der Items wurde für jede Skala separat anhand des Graded-Response-Modells (GRM) mit dem R-Paket *ltm* (Rizopoulos, 2006) überprüft. Die Schätzung der Parameter er-

folgte dabei mittels der marginalen Maximum-Likelihood-Schätzmethode. Da die Stichprobengröße hierfür als ausreichend eingestuft werden kann (vgl. De Ayala, 2009), wurden jeweils alle Items einer Skala einbezogen.

Ergebnisse

Faktorielle Validität und Dimensionalität

Tabelle 2 präsentiert die Fit-Indizes der CCFA-Modelle. Die Einfaktorstruktur scheint nicht gut zu den Daten zu passen (CFI = .910, RMSEA = .105, WRMR = 2.368). Die beiden Modelle mit drei Faktoren liefern zwar einen besseren, jedoch nicht ausreichend zufriedenstellenden Fit (CFI = .930, RMSEA = .093, WRMR = 2.105), während die drei Modelle mit zehn Faktoren alle ein sehr gutes Ausmaß an Passung mit den Daten aufweisen. Dabei vermag das first-order Modell mit zehn Faktoren die implizierte Struktur am besten zu reproduzieren (CFI = .979, RMSEA = .052, WRMR = 1.076). Das bi-faktorielle Modell weist nur geringfügig schlechtere Fitmaße auf (CFI = .978, RMSEA = .054, WRMR = 1.172).

Die Inspektion der Faktorladungen zeigt ein ähnliches Bild. Grundsätzlich haben die Items in allen Modellen substantielle Faktorladungen auf den entsprechenden Faktor erster Ordnung (Tabelle 2, letzte Spalte). In Modell 1 sind diese jedoch weniger ausgeprägt ($|\lambda| = .586$ bis $.921$, $M = .787$) als etwa in Modell 4 ($|\lambda| = .745$ bis $.971$, $M = .905$). Die bi-faktorielle CCFA erfordert eine differenziertere Beschreibung: Der generelle Faktor ist klar definiert mit signifikanten, relativ hohen Faktorladungen ($|\lambda| = .464$ bis $.896$, $M = .746$) und spricht damit für ein globales, übergeordnetes Konstrukt (adaptive Kompetenzen). Bei 36 der 40 Items zeigen sich signifikante, geringe bis ausgeprägte Faktorladungen auf den spezifischen Faktor ($|\lambda| = .013$ bis $.872$,

$M = .531$). Unter Berücksichtigung des generellen Faktors liefern demnach die meisten der einbezogenen Items - wenn auch in unterschiedlichem Ausmaß - spezifische Informationen.

Gesamthaft betrachtet ist vorläufig davon auszugehen, dass der ABAS II nicht drei Bereiche, sondern zehn mehr oder weniger distinktive Dimensionen abbildet.

Item- und Skalenstatistiken sowie Reliabilität

In Tabelle 3 sind als erstes die Deskriptivstatistiken der zehn Skalen aufgeführt. Die Skalenmittelwerte liegen allesamt über dem theoretischen Mittelwert. Die Trennschärfe-koeffizienten der Items liegen größtenteils über .30. Bei den drei Skalen Gesundheit und Sicherheit (GS), Arbeit (AR) sowie Selbstfürsorge (SS) weist jedoch je ein Item eine nicht zufriedenstellende Trennschärfe auf.

Die Skalen korrelieren alle signifikant miteinander. Die tiefste Korrelation ist zwischen der Skala Funktionelle akademische Fähigkeiten (FA) und der Skala Selbstfürsorge (SS) zu verzeichnen ($r = .52$), die höchste zwischen der Skala Orientierung in der Gemeinschaft (OG) und der Skala Gesundheit und Sicherheit (GS; $r = .87$). Die mittlere Interkorrelation beträgt .73.

Die zehn Skalen zeigen hohe interne Konsistenz. Wie Tabelle 3 zu entnehmen ist, liegt Cronbachs α zwischen .94 und .97 und erwartungsgemäß McDonalds ω etwas höher zwischen .96 und .99.

Itemanalysen mittels GRM

Die Darstellung der Ergebnisse aus den Itemanalysen zu allen 239 Items des ABAS II würde den Rahmen dieses Beitrages sprengen. Deshalb werden ausgewählte Befunde berichtet, welche einen möglichst umfassenden Einblick in die Funktionsfähigkeit der Skalen und Items gewähren sollen.

Abbildung 2 zeigt die Testinformativfunktion (TIF) von vier Skalen. Bei der Aus-

Tabelle 2: Fit-Indizes verschiedener geschätzter Modelle des ABAS II

CCFA-Modelle	χ^2	df	p	CFI	WRMR	RMSEA [V _{90%}]	CFit	$\lambda_{\min}/\lambda_{\max}$
1) 1 st -order, 1 Faktor	3105.92	740	< .001	.910	2.368	.105 [.101-.109]	< .001	.586/.921
2) 1 st -order, 3 Faktoren	2572.25	737	< .001	.930	2.105	.093 [.089-.097]	< .001	.633/.944
3) 2 nd -order, 3 Faktoren	2572.25	737	< .001	.930	2.105	.093 [.089-.097]	< .001	.633/.941
4) 1 st -order, 10 Faktoren	1243.72	695	< .001	.979	1.076	.052 [.048-.057]	.211	.745/.971
5) 2 nd -order, 10 Faktoren	1474.73	730	< .001	.972	1.338	.059 [.055-.064]	< .001	.729/.971
6) Bf-factor, 10 Faktoren	1280.52	700	< .001	.978	1.172	.054 [.049-.058]	.102	G.: 464/.896 S.: 107/.872

Anmerkungen. Die Schätzungen beziehen sich auf 40 Items (je 4 pro Dimension; Auswahl anhand Höhe der Faktorladungen). CCF: konfirmatorische Faktorenanalyse mit kategorial geordneten Indikatoren; CFI: Comparative Fit Index; WRMR: Weighted Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation; V_{90%}: 90-Prozent-Vertrauensintervall; CFit: p-Wert des Close-Fit-Tests (Wahrscheinlichkeit RMSEA $\leq .05$); l: standardisierte Faktorladungen.

Tabelle 3: Skalenstatistiken (*M*, *SD*), Interne Konsistenz bzw. Reliabilität (Cronbachs α , McDonalds ω), Trennschärfeffizienten (r_{it}) der Items sowie Interkorrelationen der Skalen

	<i>M</i>	<i>SD</i>	α	ω	(<i>VI</i> _{95%})	r_{it}	KO	OG	AF	WO	GS	FZ	SF	SS	SA
KO	55.28	14.35	.95	.97	(.96/.98)	.35-.75									
OG	42.67	18.85	.97	.98	(.98/.98)	.55-.82	.80								
AF	40.09	22.15	.97	.99	(.99/.99)	.58-.85	.77	.85							
WO	45.53	15.64	.96	.97	(.97/.98)	.46-.81	.61	.75	.62						
GS	40.47	12.58	.94	.96	(.96/.97)	.21-.78	.76	.87	.76	.83					
FZ	45.32	13.19	.94	.95	(.95/.96)	.36-.75	.78	.80	.73	.67	.79				
SF	63.67	11.58	.94	.98	(.97/.98)	.28-.72	.60	.67	.52	.80	.74	.59			
SS	51.23	15.68	.95	.97	(.96/.97)	.45-.75	.78	.84	.74	.77	.85	.79	.73		
SA	51.59	12.22	.94	.96	(.95/.96)	.35-.71	.73	.65	.55	.56	.67	.77	.57	.76	
AR	51.90	10.91	.94	.96	(.95/.97)	.27-.73	.69	.67	.60	.64	.72	.67	.61	.79	.74

Anmerkungen. *VI*_{95%}: 95-Prozent-Vertrauensintervall; r_{it} : Bereich der Trennschärfeffizienten (korrigierte Item-Skalen-Korrelationen). KO: Kommunikation (25 Items), OG: Orientierung in der Gemeinschaft (24 Items), AF: Funktionelle akademische Fähigkeiten (27 Items), WO: Wohnen (23 Items), GS: Gesundheit und Sicherheit (20 Items), FZ: Freizeit (23 Items), SF: Selbstfürsorge (25 Items), SS: Selbststeuerung (25 Items), AR: Soziale Anpassung (23 Items), SA: Arbeit (24 Items). Alle Korrelationen sind auf dem 0.01-Niveau (zweiseitig) signifikant.

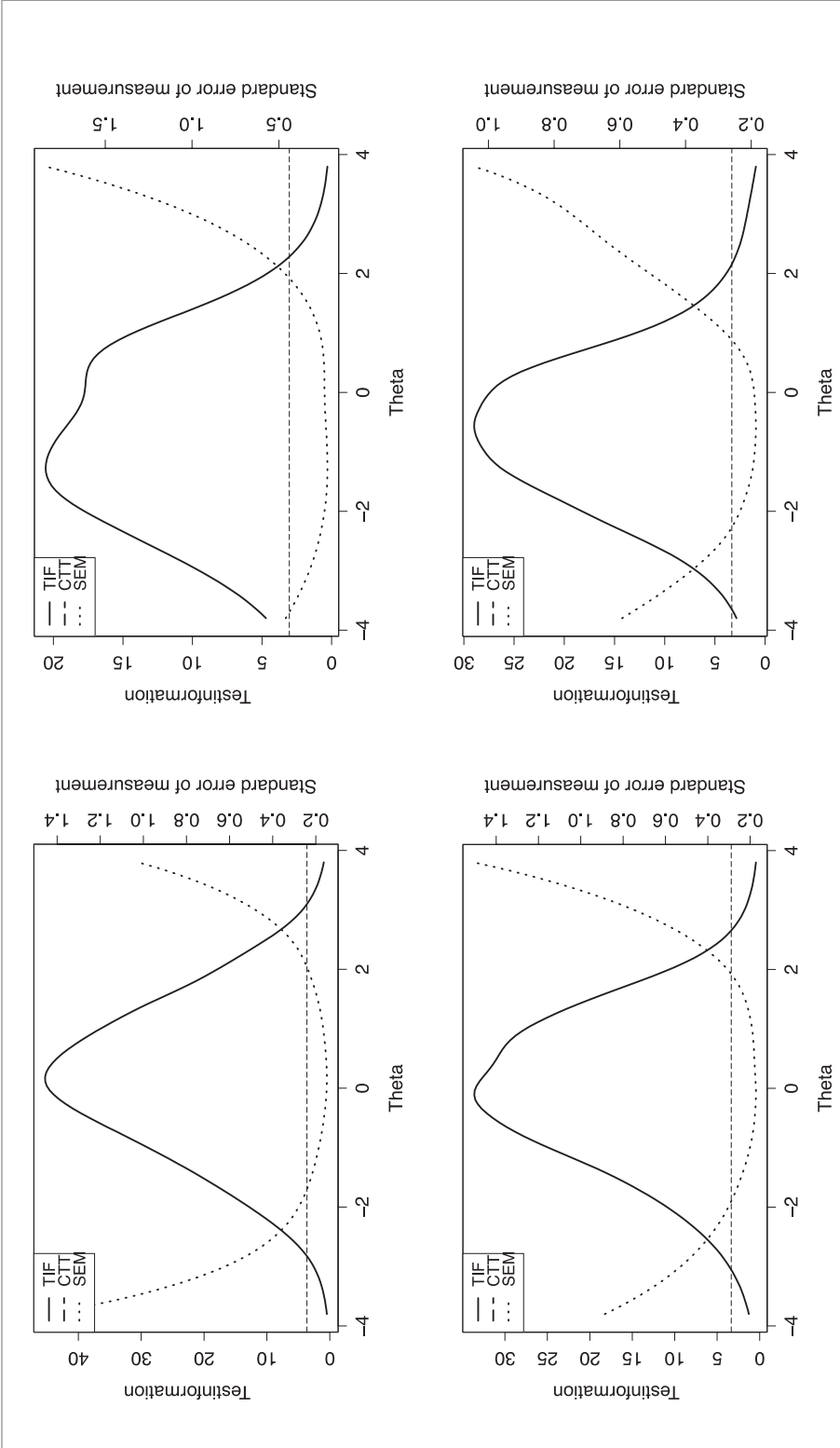


Abbildung 2: Testinformationsfunktion (TIF), durchschnittlicher Messfehler nach klassischer Testtheorie (CTT) und Standardmessfehler (SEM) vier ausgewählter Skalen

wahl wurde zum einen darauf geachtet, typische Muster zu verdeutlichen und zum anderen jeden der drei Bereiche (konzeptuell, sozial, praktisch) zu berücksichtigen.

Bei der Skala Funktionelle akademische Fähigkeiten (FA) vermögen die Items im mittleren Bereich (-2 bis 2) 86% der Information des latenten Merkmals (θ) zu liefern, wobei der grössere Anteil (46.8%) zwischen 0 und 2 liegt. Im untersten Bereich (< -2) liegen nur 5.5% der Information zu den funktionellen akademischen Fähigkeiten, im obersten Bereich (> 2) immerhin 8.5%. Zudem ist der Standardmessfehler im untersten und obersten Bereich stark erhöht. Ein ähnliches Bild ergibt sich bei der Skala Orientierung in der Gemeinschaft (OG). Die TIF ist jedoch im mittleren Bereich etwas breiter ausgebildet (42.1% vs. 44.3%) sowie etwas ausgeprägter im untersten Bereich (8.3% vs. 5.3%). Der Standardmessfehler ist insbesondere im obersten Teil stark erhöht. Die Items der Skala Wohnen (WO) bieten in der unteren Hälfte (< 0) mit knapp 66% einen großen Teil der Information von θ . Rund 50% liegen dabei zwischen -2 und 0. Zwischen 0 und 2 vermögen die Items noch 27.4% der adaptiven Kompetenzen in Bezug auf Wohnen zu erfassen. In diesem Bereich steigt zudem der Standardfehler bereits recht früh (etwa ab 1) an. Bei der Skala Soziale Anpassung ist die TIF ebenfalls im unteren Bereich stärker ausgeprägt. Auf die untere Hälfte entfallen sogar knapp 70% der Information von θ . Im untersten Bereich liegt der verhältnismäßig große Anteil von 26%. Im obersten Bereich vermögen die Items hingegen nur noch wenig Information (3.3%) zu fassen. Der Standardmessfehler ist zudem im untersten Bereich verhältnismäßig gering und nimmt erst im obersten Bereich stark zu.

Bei der Prüfung der Item-Informationenfunktionen und der Item-Antwortkategorie-Charakteristikkurven konnten insgesamt vor allem drei Erkenntnisse gewonnen werden, welche nun anhand eines Ausschnitts der Items der beiden Skalen Orientierung in der

Gemeinschaft (OG) und Soziale Anpassung (SA) exemplarisch aufgezeigt werden (Abbildung 3; Tabelle 4).

1.) Die Diskriminationsparameter a_i unterscheiden sich innerhalb einer Skala teilweise recht stark. Während beispielsweise das Item SA10 („hat ein gutes Verhältnis zu Familienmitgliedern“) mit $a_i = 0.81$ kaum zwischen Personen auf unterschiedlichen Levels des Merkmals zu differenzieren vermag, weist Item SA18 („macht anderen Komplimente für gute Taten oder Verhalten, z.B. Aufrichtigkeit oder Freundlichkeit“) mit $a_i = 2.41$ eine sehr hohe Differenzierungsfähigkeit auf.

2.) Die Schwierigkeitsparameter b_j nehmen jeweils die entsprechende Reihung auf dem latenten Kontinuum θ ein. So ergibt sich beispielsweise bei Item OG10 („beschreibt die allgemeine Adresse eines Reiseziels, z.B. in der Nähe des Bahnhofs“) die Reihenfolge: $b_1 = -0.76$, $b_2 = -0.13$, $b_3 = 0.91$. Für den ersten Schwierigkeitsparameter b_1 bedeutet dies, dass bei einem θ -Wert von -0.76 die Wahrscheinlichkeit 50% überschreitet, Kategorie 2 oder höher (statt Kategorie 1) zu wählen. Bei Item OG10 liegen die Schwellen demnach eher im mittleren Bereich von θ , d.h. das Item funktioniert hauptsächlich bei durchschnittlichen Kompetenzen bezüglich Orientierung in der Gemeinschaft. Bei Item OG13 („fragt eine Verkaufsperson um Hilfe, wenn sie im Geschäft etwas nicht finden kann“) sind die Schwellen hingegen eher im unteren Bereich, damit entspricht es einem relativ leichten Item. Item OG18 („ruft einen Handwerker, wenn z.B. die Heizung nicht mehr funktioniert“) wiederum erfordert ein höheres Kompetenzniveau und kann daher als relativ schwieriges Item bezeichnet werden.

3.) Die zweite und dritte Kategorie haben vielfach eine geringe Wahrscheinlichkeit gewählt zu werden, z.B. bei Item OG15 („besucht alleine Freunde in der Nachbarschaft“). Insbesondere die zweite Kategorie wird bei gleichem θ relativ selten gewählt. Teilweise fällt die zweite Kategorie sogar mit der ersten Kategorie zusammen,

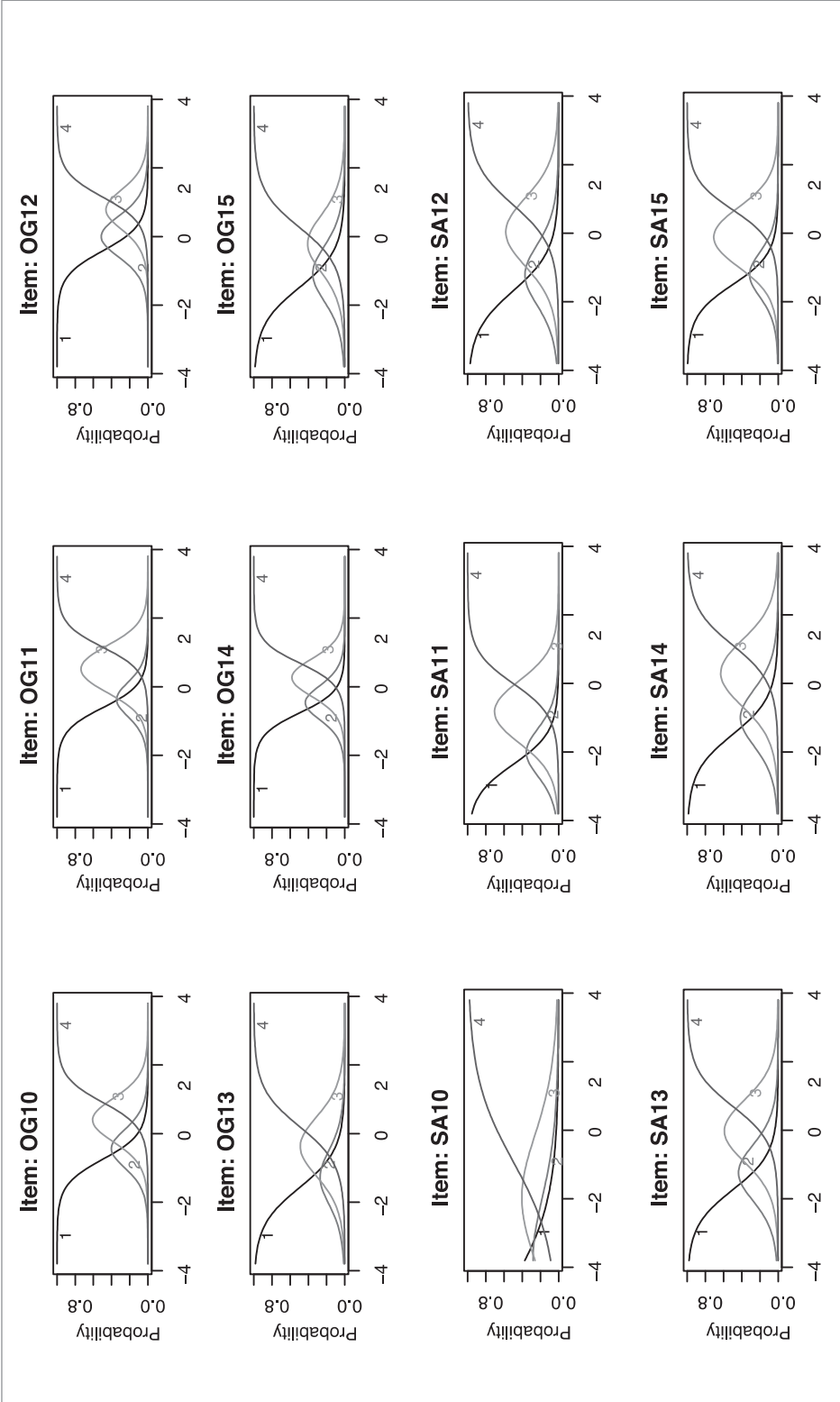


Abbildung 3: Item-Antwortkategorie-Charakteristikkurven (item response category characteristic curves) ausgewählter Items des ABAS II

Tabelle 4: Diskriminationsparameter (a_i) und Schwierigkeitsparameter (b_j) der Informationsfunktion zu ausgewählten Items

	a_i	b_1	b_2	b_3
Orientierung in der Gemeinschaft (OG)				
OG 10	2.72	-0.76	-0.13	0.91
OG 11	2.94	-0.62	-0.13	1.56
OG 12	2.71	-0.42	0.42	1.16
OG 13	1.82	-1.58	-0.97	0.21
OG 14	3.09	-0.77	-0.17	0.69
OG 15	1.82	-1.49	-0.68	0.29
Soziale Anpassung (SA)				
SA 10	0.81	-4.45	-3.02	-0.91
SA 11	2.11	-2.37	-1.66	0.00
SA 12	1.68	-1.69	-0.76	0.82
SA 16	2.09	-1.45	-0.41	0.92
SA 17	2.34	-1.75	-0.78	0.76
SA 18	2.41	-1.59	-0.75	0.65

z.B. bei Item OG11 („folgt der Wegbeschreibung Anderer zu nahe gelegenen Orten“) oder bei Item SA11 („zeigt Mitgefühl mit Anderen, wenn diese traurig oder wütend sind“).

Diskussion

Die vorliegende Überprüfung der deutschen Erwachsenenversion des ABAS II zeigt, dass mit dem Instrument die adaptiven Kompetenzen reliabel erfasst werden können. Zudem weist der ABAS II insbesondere im unteren Bereich der Kompetenzausprägung eine gute Diskriminationsfähigkeit auf und eignet sich somit für ein Assessment bei intellektuellen Beeinträchtigungen. Gleichzeitig deuten die Ergebnisse auch auf Schwachstellen des ABAS II hin, welche nun entlang der drei Kriterien der Zielsetzung methodenkritisch und mit Blick auf weiterführende Forschungsarbeiten diskutiert werden. Die entsprechenden prakti-

schen Implikationen werden anschließend zusammenfassend aufgezeigt.

Methodenkritische Diskussion

Die Ergebnisse stellen die postulierte Struktur des ABAS II – ein Konstrukt mit zehn Dimensionen und drei übergeordneten Bereichen – in Frage. Mit dem ABAS II scheinen adaptive Kompetenzen generell sowie die zehn Dimensionen im Besonderen erfasst zu werden. Die Ergebnisse der bi-faktoriellen CCFA unterstützen den Bedarf, diese Inhaltsspezifität im Modell zu berücksichtigen. Einschränkend gilt festzuhalten, dass die konfirmatorische Prüfung der faktoriellen Struktur basierend auf den je vier Items mit den höchsten Faktorladungen vorgenommen wurde. Geringere Faktorladungen führen tendenziell zu geringeren Fit-Werten (Heene et al., 2011), was allerdings auch für Modelle mit drei Faktoren zutrifft. Bei einer Überprüfung anhand einer größeren Stichprobe würden sich unter anderem bi-faktorielle explorative Strukturgleichungsmodelle anbieten, bei denen - im Gegensatz zu konfirmatorischen Modellen - Querladungen zugelassen sind (Morin et al., 2016).

Positiv zu werten ist die hohe interne Konsistenz der zehn Skalen. In Anbetracht der recht großen Anzahl Items pro Skala ist dieser Befund jedoch etwas zu relativieren. Zudem ergaben die Analysen mittels GRM, dass die Reliabilität häufig im obersten Bereich und bei einigen Skalen im untersten Bereich tiefer ist.

Die Mehrheit der Skalen vermag vor allem in den unteren und mittleren Bereichen des jeweiligen Konstrukts - wie etwa Kommunikation oder Orientierung in der Gemeinschaft - zu differenzieren bzw. zu diskriminieren. Eine Ausnahme bildet die Dimension der funktionellen akademischen Fähigkeiten, wo der größere Teil der Information eher in der oberen Hälfte des Merkmals liegt. Vor dem Hintergrund, dass die Parameterschätzungen im Rahmen von IRT stichprobenunabhängig sind, kann die Dis-

krinationsfähigkeit des Instruments insgesamt als angemessen eingestuft werden.

Als eine der Stärken der IRT gilt, individuelle Unterschiede auf einem zu Grunde liegenden Konstrukt - wie beispielsweise soziale Anpassung - mit der Wahrscheinlichkeit von Item-Antworten zu verknüpfen. Die Analysen auf Itemebene zeigten dabei eine gewisse Heterogenität innerhalb des ABAS II. So weisen die Items unterschiedliche Diskriminations- und Schwierigkeitsgrade auf.

Die Itemanalysen mittels GRM zeigten des Weiteren, dass eine Unterscheidung zwischen der ersten und zweiten Antwortkategorie („0 = ist nicht fähig, das Verhalten zu zeigen“ und „1 = zeigt das Verhalten nie oder fast nie, wenn erforderlich“) bei einigen Items fraglich ist. Hier bildet sich offensichtlich das inhaltlich-konzeptionelle Problem ab, nicht gezeigtes Verhalten mittels Beobachtung bewerten zu müssen. Dies deckt sich mit unseren Erfahrungen bei der Durchführung der Untersuchung. Die Respondenten hatten bei nicht beobachtbarem Verhalten Schwierigkeiten zu unterscheiden, ob die Person nicht fähig ist, das Verhalten zu zeigen, oder ob die Person das Verhalten (aus bestimmten Gründen) nie oder fast nie zeigt, wenn es erforderlich ist. Diese Problematik, die offensichtlich auch auf den Iteminhalt zurückzuführen ist, sollte weiter untersucht und im Falle entsprechender zusätzlicher Befunde durch eine Adaptation der Antwortkategorien bearbeitet werden.

Praktische Implikationen

Für die praktische Anwendung der Erwachsenenversion des ABAS II ergeben sich aus den Ergebnissen zur Strukturprüfung folgende Schlussfolgerungen. Sowohl für die Diagnostizierung einer geistigen Behinderung als auch für das individuelle Assessment im Rahmen von Hilfeplanung und Unterstützung ist eine Konzentration auf den Gesamtwert adaptiver Kompetenzen sowie auf die zehn Dimensionen zu empfehlen. Intra-

und interindividuelle Profilanalysen auf Ebene der zehn Dimensionen anstelle der Betrachtung der drei übergeordneten Bereiche liefern nicht nur spezifischere, sondern auch empirisch besser abgestützte Erkenntnisse.

Wie bereits in der methodenkritischen Diskussion angesprochen, weisen die Items unterschiedliche Diskriminations- und Schwierigkeitsgrade auf. Dieses Merkmal ist für die angestrebte Differenzialdiagnostik auf inter- und intraindividuelle Ebene als positiv zu bewerten.

Die gute Diskriminationsfähigkeit des ABAS II im unteren Bereich der Kompetenzausprägungen ist besonders für ein Assessment bei Erwachsenen mit intellektuellen Beeinträchtigungen von großem Wert. Mit dem ABAS II ist im Unterschied zu verfügbaren Verfahren der Intelligenzmessung auch eine Diagnostik im Bereich starker Normunterschreitungen möglich. Diese Eigenschaft prädestiniert das Instrument für die Bestimmung von Schweregraden einer geistigen Behinderung anhand der Ausprägung adaptiver Kompetenzen, welche aktuell empfohlen wird (APA, 2013b).

Fazit

Zusammenfassend hat die hier vorgelegte Überprüfung der deutschen Übersetzung des ABAS II neue Erkenntnisse bezüglich der faktoriellen Struktur, der Reliabilität der Skalen sowie zur Funktionsweise der Items erbracht und Ansatzpunkte für eine weitere Auseinandersetzung mit den psychometrischen Eigenschaften des Instrumentes identifiziert. Im Zusammenhang mit der großen Bedeutung eines qualitativ hochwertigen Assessments adaptiver Kompetenzen sollte an den oben genannten Punkten auch im deutschen Sprachraum weiter geforscht werden.

Literaturverzeichnis

- American Association on Intellectual and Developmental Disabilities. (2010). *Intellectual disability: Definition, classification, and systems of supports* (11th ed.). Washington, DC: Author.
- American Psychiatric Association. (2013a). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: Author.
- American Psychiatric Association. (2013b). *Highlights of changes from DSM-IV-TR to DSM-5*. Washington, DC: Author. Available from <http://www.psychiatry.org/psychiatrists/practice/dsm>
- Beauducel, A. & Wittmann, W. W. (2005). Simulation study of fit indexes in CFA based on data with slightly distorted sample structure. *Structural Equation Modeling*, 12(1), 41-75.
- Burns, M. K. (2005). Test review of the Adaptive Behavior Assessment System (2nd ed.). In R. A. Spies & B. S. Plake (Eds.), *The Sixteenth Mental Measurements Yearbook*. Lincoln, NE: Buros Center of Mental Measurement. Available from <http://www.unl.edu/buros/>
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- Dunn, T. J., Baguley, T. & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399-412.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Psychology Press.
- Harrison, P. L. & Oakland, T. (2008). *Adaptive Behavior Assessment System – Second Edition (ABAS II)*. Los Angeles: Western Psychological Services.
- Harrison, P. L. & Oakland, T. (2015). *Adaptive Behavior Assessment System – Third Edition (ABAS III)*. Los Angeles, CA: Western Psychological Services.
- Heber, R. (1959). A manual on terminology and classification in mental retardation. Monograph supplement. *American Journal of Mental Deficiency*, 64, 1–111.
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M. & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, 16(3), 219-336.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Little, T. D., Cunningham, W. A., Shahar, G. & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 151-173.
- McDonald, R. P. (2013). Modern test theory. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods. Volume 1: Foundations* (pp. 118-143). New York, NY: Oxford University Press.
- Morin, A. J. S., Arens, A. K. & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*, 23, 116-139.
- Muthén, L. K. & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Orthmann Bless, D. (2013). *Evaluationssystem für das Adaptive Verhalten, Fragebogen für Erwachsene 16 – 89 Jahre* (unveröffentlichte deutsche Übersetzung des Adaptive Behavior Assessment System – Second Edition, von P. L. Harrison & T. Oakland, 2008). Freiburg: Heilpädagogisches Institut der Universität Freiburg/Schweiz.
- Orthmann Bless, D. & Hellfritz, K.-L. (2016). *Eltern mit geistiger Behinderung und ihre Kinder unterstützen. Evaluation zur Begleiteten Elternschaft in Deutschland*.

- Freiburg: Heilpädagogisches Institut der Universität Freiburg/Schweiz.
- Orthmann Bless, D. & Zurbriggen, C. (2017). Zur Variabilität adaptiver Kompetenzen von Erwachsenen mit geistiger Behinderung. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 86(1), 41-55.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1-25.
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. Richmond, VA: Psychometric Society.
- Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). New York, NY: Springer.
- Sermier Dessemontet, R., Bless, G. & Morin, D. (2012). Effects of inclusion on the academic achievement and adaptive behaviour of children with intellectual disabilities. *Journal of Intellectual Disability Research*, 56(6), 579-587.
- Tassé, M. J. (2013). Adaptive behavior. In M. L. Wehmeyer (Ed.), *The Oxford handbook of positive psychology and disability* (pp. 105-115). Oxford: Oxford University Press.
- Tassé, M. J., Schalock, R. L., Balboni, G., Bersani, H., Borthwick-Duffy, S. A., Sprent, S., . . . Zhang, D. (2012). The construct of adaptive behavior: Its conceptualization, measurement, and use in the field of intellectual disability. *American Journal on Intellectual and Developmental Disabilities*, 117(4), 291-303.
- Tassé, M. J., Schalock, R. L., Thissen, D., Balboni, G., Bersani, H., Borthwick-Duffy, S. A., . . . Navas, P. (2016). Development and standardization of the Diagnostic Adaptive Behavior Scale: Application of the item response theory to the assessment of adaptive behavior. *American Journal on Intellectual and Developmental Disabilities*, 121(2), 79-94.
- Wang, J. & Wang, X. (2012). *Structural equation modeling. Applications using Mplus*. Chichester, UK: Wiley.
- Wirth, R. J. & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, 12(1), 58-79.
- Ziegler, M. & Hagemann, D. (2015). Testing the unidimensionality of items: Pitfalls and loopholes. *European Journal of Psychological Assessment*, 31(4), 231-237.

Carmen Zurbriggen

Universität Freiburg (Schweiz)
 Departement für Sonderpädagogik
 Petrus-Kanisius-Gasse 21
 1700 Fribourg
 Schweiz
 carmen.zurbriggen@unifr.ch

Erstmalig eingereicht: 28.03.2017

Überarbeitung eingereicht: 08.05.2017

Angenommen: 15.05.2017