

### **Psychometrische Qualität der deutschsprachigen ABAS-III Lehrerversion (5 – 21 Jahre). Ergebnisse der Erstpilotierung.**

**Zusammenfassung:** Das ABAS-III dient der Erfassung adaptiver Kompetenzen und wurde in der englischen Originalversion für eine Bevölkerung von 0 bis 89 Jahren normiert (Harrison & Oakland, 2015). In dieser Untersuchung wird die *deutschsprachige* Übersetzung des ABAS-III für 5 bis 21 jährige Schüler und Schülerinnen (Lehrerversion) *erstmalig* anhand einer deutschen Pilotierung von  $N = 217$  Kindern und Jugendlichen psychometrisch evaluiert. Die internen Konsistenzen der 10 Einzelskalen sind mit  $\alpha$  zwischen .89 und .96 durchweg gut bis hervorragend. Das einfaktorielle Gesamtkonstrukt der adaptiven Kompetenz lässt sich mit einer konfirmatorischen Faktorenanalyse ( $\chi^2(135) = 435.19, p < .001; RMSEA = .104, 90\% \text{ CI } [.093, .115]; CFI = .77; SRMR = .078$ ) akzeptabel plausibilisieren. Weiterhin wird das Drei-Faktoren-Domänenmodell mittels CFA überprüft und weist mit  $\chi^2(132) = 404.3, p < .001; RMSEA = .10, 90\% \text{ CI } [.093, .115]; CFI = .80$  und  $SRMR = .078$  eine ebenfalls akzeptable Anpassung auf. Aufgrund des geringen Stichprobenumfangs werden pro Subskala 2 Items nach *Restreliabilität* ausgewählt (a). Alternativ wird *Item-Parceling* genutzt, um die Anzahl freier Parameter zu reduzieren (b). Die Ergebnisse zeigen, dass das 1-Faktor-Modell schlechter an die Datenstruktur angepasst ist und das Domänenmodell vorzuziehen ist. Einschränkungen sowie praktische Implikationen der Studie werden abschließend diskutiert.

**Schlüsselwörter:** Pilotierung, ABAS-III, adaptive Kompetenzen, Kinder und Jugendliche, psychometrische Testgüte

**Abstract:** ABAS-III is an assessment system for adaptive skills and normed for a sample of the population of 0 to 89 years Old in the U.S. (Harrison & Oakland, 2015). The actual analysis is the *first* psychometric evaluation of the German version of ABAS-III for the subsample of 5 to 21 years old pupils (Teacher Form) in a pilot study of  $N = 217$  children and adolescents. Internal consistency of the 10 subscales is very good to excellent with  $\alpha$  from .89 to .96. The general factor of adaptive skills is reproduced with acceptable model fit in CFA with ( $\chi^2(135) = 435.19, p < .001; RMSEA = .104, 90\% \text{ CI } [.093, .115]; CFI = .77; SRMR = .078$ ). The model with three distinct domains of adaptive behaviour (conceptual, social, practical) is also reproduced with acceptable model fit in CFA with  $\chi^2(132) = 404.3, p < .001; RMSEA = .10, 90\% \text{ CI } [.093, .115]; CFI = .80$  and  $SRMR = .078$ . Due to small sample size, two items per subscale are chosen for CFA with respect to *post-reliability* (a). Furthermore *Item-Parceling* is used for reducing the number of free parameters (b). The results of CFA show the model with *three behaviour domains* is better regarding model fit to the simpler model with *general factor* of adaptive skills. Limitations as well as practical implications of the study are discussed in conclusion.

**Keywords:** pilot, ABAS-III, adaptive skills, children and adolescents, psychometric test quality

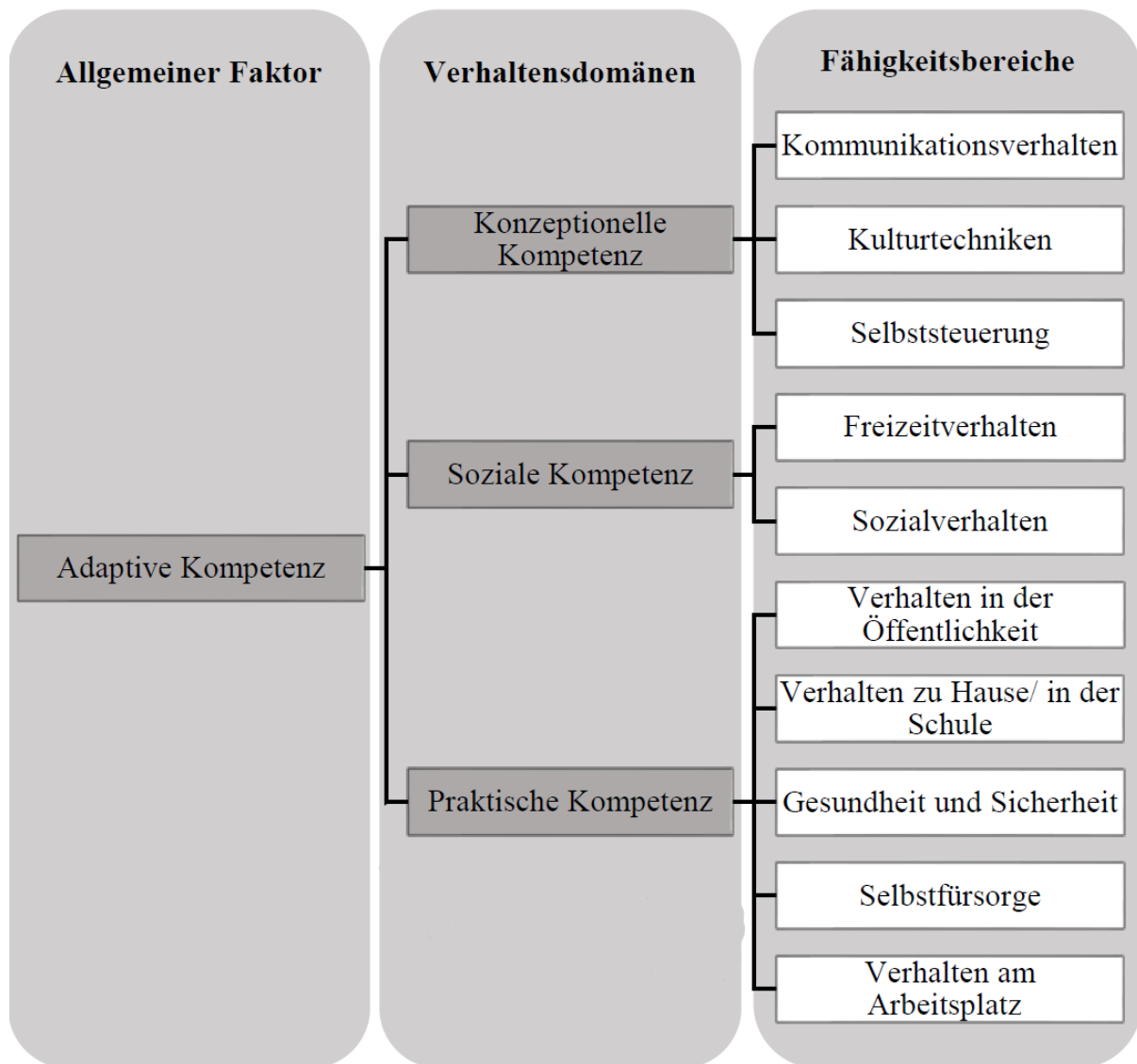
## Das ABAS-III als Messinstrument zur Erfassung adaptiver Kompetenzen:

Adaptive Kompetenzen beschreiben im Gegensatz zur Intelligenz als abstrakt-kognitive Problemlösefähigkeit konzeptionelle, soziale und praktische Fähigkeiten um sich im alltäglichen Leben zu behaupten und ein möglichst selbstbestimmtes Leben in der Gesellschaft führen zu können. Sie werden im Laufe des Lebens erworben und erweitert und ihr individuelles Ausmaß bestimmt maßgeblich den Grad der persönlichen Autonomie und der Bewältigung typischer gesellschaftlicher Anforderungen (Heber, 1959; Tassé, 2013; Hangauer et al., 2013). Der Diagnostik und Förderung dieser Kompetenzen ist daher insbesondere in sonderpädagogischen Hilfesystemen wie z.B. der (Förder-)Schule besondere Bedeutung beizumessen (Orthmann Bless & Zurbriggen, 2017). Als fundamentale Grundlage dieser Diagnostik und der daraus abgeleiteten Prozessevaluation ist ein psychometrisch hochwertiges Messinstrument unabdingbar. Das ABAS-III in der Lehrerversion ist ein Messinstrument zur personendifferenzierten Erfassung des Ausmaßes adaptiver Kompetenzen von Kindern und Jugendlichen im schulischen Kontext und ist für die englischsprachige Originalversion für einen Altersbereich von 5 bis einschließlich 21 Jahren normiert. Die psychometrische Qualität mittels Cronbachs Alpha liegt für die 10 einzelnen Subskalen des englischsprachigen ABAS-III zwischen  $\alpha = .86$  und  $\alpha = .99$  ( $\emptyset = .93$ ) und weist damit auf eine gute bis exzellente interne Konsistenz der jeweiligen Skalen hin. Die übergeordneten Verhaltensdomänen „Konzeptionell“, „Sozial“ und „Praktisch“ weisen einheitlich eine exzellente Messgenauigkeit mit  $\alpha = .97$  auf. Für den Gesamtfaktor „adaptive Kompetenz“ liegt ebenfalls mit  $\alpha = .99$  eine exzellente Messgenauigkeit vor (Harrison & Oakland, 2015).

Allgemein wird im ABAS-III von einem Gesamtkonstrukt adaptiver Kompetenzen ausgegangen, dass sich auf einer kleinteileren Ebene wiederum in drei Domänen adaptiver Kompetenzen aufteilen lässt. Bei den drei Domänen handelt es sich um „Konzeptionelle Kompetenzen“, „Soziale Kompetenzen“ und „Praktische Kompetenzen“. Diese Domänen werden jeweils durch ein Bündel an ordinalskalierten Likert-Items gemessen, die für konkrete Anforderungssituationen abfragen, ob eine Verhaltensweise durch den/die beurteilte/n Schüler/in gezeigt wird. Es ergeben sich folgende mögliche Ausprägungen auf der Skala: 0 (nicht in der Lage das Verhalten zu zeigen), 1 (nie oder fast nie, wenn erforderlich), 2 (manchmal, wenn erforderlich), 3 (immer oder fast immer, wenn erforderlich). Weiterhin liegt eine Zusatzkategorie vor, die ausweist, ob die jeweilige Beurteilung auf einer Vermutung basiert (Harrison & Oakland, 2015).

Die Likert-Items sind inhaltlichen Unterskalen adaptiver Kompetenz zugeordnet, zum Beispiel wird die konzeptionelle Verhaltensdomäne durch die Subskalen „Kommunikationsverhalten“, „Kulturtechniken“ und „Selbststeuerung“ operationalisiert.

Das folgende Schaubild (siehe Abbildung 1) skizziert den *Generalfaktor* adaptive Kompetenz, alle drei *Domänen* der adaptiven Kompetenz und deren zugrundeliegenden *Subskalen*:



**Abbildung 1:** Faktorenstruktur des ABAS-III für das einfaktorische Messmodell und das dreifaktorielle Domänenmodell.

### Fragestellung und Überblick der Pilotierung:

In folgendem Beitrag wird die erste Evaluation der psychometrischen Eigenschaften der deutschen Lehrerversion des ABAS-III anhand einer aktuell erhobenen Pilotstichprobe von  $N = 217$  Kindern und Jugendlichen durchgeführt. Konkret wird die faktorielle Validität der theoretisch postulierten Messmodelle (1-Faktor-Gesamt-Modell und Drei-Faktoren-Domänenmodell) sowie die Dimensionalität und die Reliabilität des Gesamtfaktors „adaptive Kompetenz“, der drei Verhaltensdomänen „konzeptionell“, „sozial“, „praktisch“ und der Einzelskalen mittels Cronbachs  $\alpha$  überprüft. Alternativ wird neben Cronbachs  $\alpha$  zusätzlich das psychometrisch weniger restriktive  $\omega$  nach McDonald für die Berechnung betrachtet, um eine plausiblere Schätzung der tatsächlichen jeweiligen Skalenreliabilität zu erhalten (McDonald, 1999).

## Methodik

Im Folgenden werden die eingesetzte Testbatterie, der deutschsprachige Übersetzungsprozess, das Auswahlverfahren der Stichprobenrekrutierung sowie die angewandten statistischen Analysen beschrieben.

### Testbatterie

Die Erhebung umfasste neben dem ABAS-III Pilotierungsinstrument und soziodemografischen Angaben zum/zur Schüler/in und zur beurteilenden Lehrkraft, ein Anschreiben an die Lehrkräfte sowie Erziehungsberechtigten der Schüler/innen sowie eine Datenschutzerklärung und Eidesstattliche Versicherung.

### Übersetzung

Die Übersetzung des ABAS-III (u.a. Lehrkräfteversion) folgte in Anlehnung an die Test Adaptation Guidelines der International Test Commission (ITC, 2005). Es wurde das übliche Verfahren der Rückübersetzung angewandt und mit weiteren Schritten kombiniert. In einem ersten Schritt erfolgte die Übersetzung der ABAS-III Originalversion u.a. für Lehrer/innen durch die deutsche Arbeitsgruppe (Bienstein, Döpfner, Sinzig<sup>†</sup>). Die Arbeitsgruppe setzte sich aus Expert/innen mit unterschiedlichen Berufsqualifikationen (u.a. aus Medizin, Psychologie, (Heil-/Sonder-)Pädagogik, diplomierte Übersetzer/in), mit Expertise u.a. im Bereich psychologisch-pädagogischer Diagnostik, Testadaptation und kulturvergleichender Studien (u.a. ausgewiesene Kenntnisse in Bezug auf die beteiligten Kulturen und zu messenden Konstrukte) zusammen. Jede Unstimmigkeit wurde diskutiert, bis eine Einigung erzielt wurde. In einem weiteren Schritt wurde die deutsche Übersetzung durch eine weitere diplomierte Übersetzerin rückübersetzt und mit der Originalversion verglichen. Die Rückübersetzung und vereinzelt Abweichungen sowie Anmerkungen wurden erneut in der Arbeitsgruppe besprochen und angepasst.<sup>1</sup>

In einem nächsten Schritt erfolgte der Einbezug von (angehenden) Praktiker/innen. Diese umfasste eine Befragung von Lehramtsanwärter/innen ( $N = 26$ ) sowie von bereits in der Praxis tätigen Lehrer/innen sowie Eltern von Kindern/Jugendlichen (im Alter von 5-21 Jahre;  $N = 13$ ), die eine Bewertung der deutschen Übersetzung u.a. mit Blick auf die Instruktions- und Itemverständlichkeit und Testökonomie vornahmen. Die Rückmeldungen wurden in der Arbeitsgruppe unter Berücksichtigung der Originalversion noch einmal besprochen und letzte Anpassungen in Bezug auf Bedeutungsgleichheit, Genauigkeit und Relevanz im Abgleich mit der Originalversion verglichen und sofern indiziert, in die deutsche Übersetzung eingearbeitet.

---

<sup>1</sup> Aufgrund äußerer Bedingungen, konnten weitere bzw. abschließende Anmerkungen zur Rückübersetzung durch den/die Herausgeber/in (Harrison & Oakland<sup>†</sup>) erst nach der ersten Pilotierung berücksichtigt werden.

### Stichprobenrekrutierung:

Die im Folgenden untersuchte Pilotstichprobe basiert auf einer Beurteilung von insgesamt  $N = 217$  Kindern und Jugendlichen durch eine betreuende Lehrkraft an allgemeinbildenden Schulen in Nordrhein-Westfalen im Zeitraum von 10/2015-07/2016. Die Auswahl der Schulen erfolgte willkürlich nach Gelegenheit und innerhalb der jeweiligen Schule erfolgte die Auswahl der zu beurteilenden SchülerInnen durch die Lehrkraft per einfacher Zufallsauswahl. Bei der Stichprobe handelt es sich daher um eine nicht-repräsentative Gelegenheitsstichprobe.

Die Interviews wurden durch Studierende der Bildungswissenschaften der Universität zu Köln durchgeführt. Die Interviewer/innen durchliefen im Anschluss an eine Einführungsvorlesung und ein Seminar zu Grundlagen der Diagnostik ein fallbezogenes Vertiefungsangebot zum ABAS-III in Kleingruppen (Universität zu Köln, 2015). Dieses umfasste eine mehrstündige theoretische und praktische Schulung zu adaptiven Kompetenzen, den Grundlagen der Testkonstruktion/Entwicklung des ABAS-III und zur Testadministration (u.a. Durchführung, Auswertung und Interpretation).

### Statistische Analysen:

Zur Überprüfung der psychometrischen Qualität des deutschsprachigen ABAS-III werden für die Pilotstichprobe die internen Konsistenzen mittels Cronbachs Alpha ( $\alpha$ ) und alternativ McDonalds Omega ( $\omega$ ) ermittelt. Hier wird die Reliabilität der 10 Subskalen des ABAS-III bestimmt. Weiterhin werden die Reliabilität der drei Domänen adaptiver Kompetenz (konzeptionell, sozial, praktisch) sowie des Gesamtkonstrukts adaptiver Kompetenz berechnet. Der Vorteil von McDonalds  $\omega$  gegenüber Cronbachs  $\alpha$  besteht darin, dass die restriktive und häufig nicht erfüllte Annahme der *essenziellen Tau-Äquivalenz* für die korrekte Schätzung der internen Konsistenz einer untersuchten Einzelskala keine Notwendigkeit mehr darstellt. Die Schätzung der internen Konsistenz nach Cronbach ist nur dann plausibel, wenn alle zugrundeliegenden Items der jeweiligen Einzelskala die *gleichen* Faktorladungen auf den implizierten gemeinsamen Faktor aufweisen. McDonalds  $\omega$  kommt auch bei variierenden Ladungsstärken der einzelnen Items zu einer korrekten Schätzung der Skalenreliabilität und ist daher methodologisch vorzuziehen (McDonald, 1999; Green & Yang, 2009; Dunn et al., 2014; McNeish, 2018).

Im Anschluss werden insgesamt vier *konfirmatorische Faktorenanalysen* (CFA) durchgeführt, um den Gesamtfaktor „Adaptive Kompetenzen“ sowie die dreifaktorielle Domänenstruktur des ABAS-III (konzeptionell, sozial, praktisch) messtheoretisch zu prüfen. Bei den CFAs muss aufgrund eines sehr ungünstigen Verhältnisses von  $N = 217$  beurteilten Personen zu rund 200 manifesten Items eine *Reduktion* auf 2 Items pro Subskala erfolgen, um die modellimplizierten Kovarianzmatrizen mittels Maximum-Likelihood (ML) zu schätzen.

Zur Auswahl der jeweils 2 Items pro Subskala werden auf Grundlage der einfachen Reliabilitätsanalysen das jeweils beste bzw. schlechteste Item nach *Restreliabilität* ausgewählt, d.h. das Item für das nach Ausschluss das *maximale* bzw. *minimale* Restalpha resultiert. Als alternative Modellspezifikation werden mithilfe von *Item-Parceling* inhaltlich passende Bündel von Items per personenspezifischer Mittelwertbildung aggregiert, um diese in den beiden Faktorenanalysen (Gesamtkonstrukt und Domänenmodell) als manifeste

Indikatoren zu verwenden. Inhaltliches Item-Parceling meint, dass alle ordinalen Likert-Items, die zur selben theoretischen Subskala (Kommunikationsverhalten, Kulturtechniken, Selbstfürsorge, etc.) gehören per Mittelwertbildung auf Personenebene zusammengefasst werden. Dies hat den statistischen Vorteil, dass (a) ein feineres Messniveau erreicht wird, was insbesondere bei grobstufigen Likert-Items sinnvoll ist, um ein näherungsweise metrisches Messniveau zu erreichen (Bandalos, 2002) und (b) die Anzahl zu schätzender Parameter substantiell zu reduzieren um bei kleinen Stichproben stabile Schätzungen der Parameter sowie der Modellgütemaße zu realisieren (Bandalos & Finney, 2001; Little et al., 2002).

Der Umfang der Gesamtstichprobe beträgt dabei  $N = 217$  Kinder und Jugendliche. Eine davon abweichende ausgewiesene Fallzahl ergibt sich aufgrund fehlender Werte auf den jeweils relevanten Items der zugrundeliegenden Domänen bzw. einzelnen Subskalen.

## Beschreibung der Stichprobe

Zur Beschreibung der erhobenen Pilot-Stichprobe von Kindern und Jugendlichen wird im Folgenden auf die univariate Geschlechterverteilung, Altersverteilung, besuchte Klassenstufen, den Migrationshintergrund sowie Vorliegen einer Behinderung bzw. Einschränkung eingegangen:

### Angaben zum/zur Schüler/in:

Es ergibt sich mit einem Anteil von rund 59% männlichen zu rund 41% weiblichen Kindern und Jugendlichen eine leicht von der Parität abweichende Verteilung in der Stichprobe (Tabelle 1).

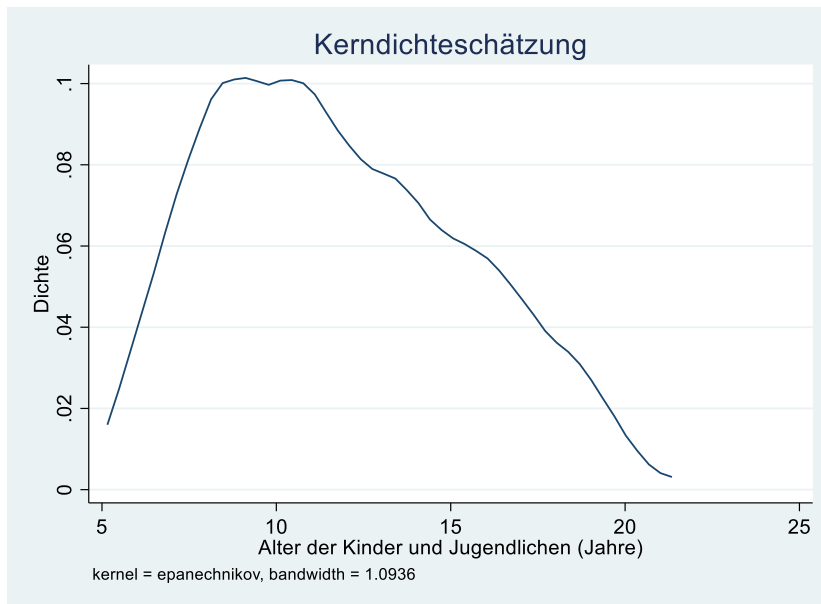
**Tabelle 1:** Verteilung des Geschlechts der Kinder und Jugendlichen.

Geschlecht	Absolute Häufigkeit	Prozent
Männlich	125	58.69
Weiblich	88	41.31
Total	213	100

Mit einem Mittelwert von  $M = 11.80$  Jahren ( $SD = 3.55$  Jahre) und einem Median von 11.38 Jahren resultiert eine leicht *rechtsschiefe* Verteilung des Alters, d.h. mehr als 50 % der Kinder und Jugendlichen sind jünger als der Mittelwert von  $M = 11.80$  Jahren. Das beobachtete Minimum liegt bei 6.25 Jahren und das Maximum bei 20.25 Jahren (Tabelle 2 & Abbildung 2).

**Tabelle 2:** Altersverteilung für die Kinder und Jugendlichen.

Merkmal	Mittelwert (SD)	Median	Min	Max	n (gültig)
Alter	11.80 (3.55)	11.38	6.25	20.25	212



**Abbildung 2:** Geschätzte Verlaufsform der Altersverteilung.

Die große Mehrheit der Kinder und Jugendlichen befindet sich mit rund 88 % anteilig höchstens in der Mittelstufe (bis einschließlich 10. Klasse). Eine Minderheit von rund 9 % der Kinder und Jugendlichen ist in der Oberstufe bzw. auf einer berufsbildenden Schule (Tabelle 3).

**Tabelle 3:** Verteilung der Kinder und Jugendlichen über die Schulklassen.

Schulklasse	Absolute Häufigkeit	Prozent
Kindergarten	2	0.94
1. Klasse	19	8.92
2. Klasse	25	11.74
3. Klasse	25	11.74
4. Klasse	21	9.86
5. Klasse	14	6.57
6. Klasse	30	14.08
7. Klasse	9	4.23
8. Klasse	20	9.39
9. Klasse	11	5.16
10. Klasse	11	5.16
11. Klasse	10	4.69
12. Klasse	6	2.82
13. Klasse	1	0.47
Berufsschule	3	1.41
Andere	6	2.82
Total	213	100

**Tabelle 4:** Verteilung der Kinder und Jugendlichen nach Migrationshintergrund.

Migrationshintergrund	Absolute Häufigkeit	Prozent
Ja	61	28.64
Nein	152	71.36
Total	213	100

Rund 29 % der beurteilten Kinder und Jugendlichen weisen einen Migrationshintergrund auf. Für rund 71 % der Kinder und Jugendlichen wird ein Migrationshintergrund verneint (Tabelle 4).

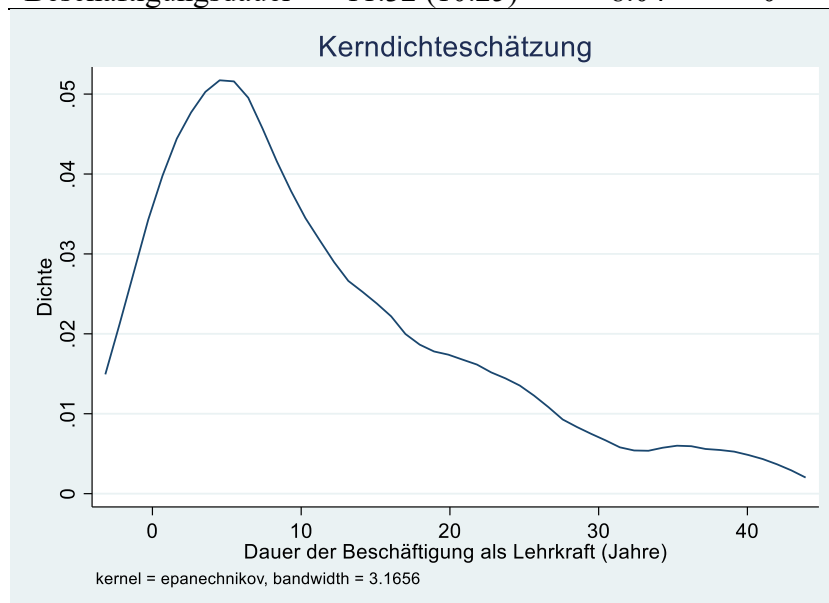
**Tabelle 5:** Auftretenshäufigkeit einer Behinderung/Einschränkung für die Kinder und Jugendlichen.

Behinderung/Einschränkung	Absolute Häufigkeit	Prozent
Nein	168	78.87
Ja	45	21.13
Total	213	100

Rund 79 % der beurteilten Kinder und Jugendlichen weisen keine Behinderung bzw. Einschränkung auf. Für rund 21 % ist eine Behinderung oder Einschränkung genannt (Tabelle 5). Exemplarisch sind hier ASS ( $n = 3$ ), ADHS ( $n = 2$ ), das Down-Syndrom ( $n = 2$ ) sowie Dyskalkulie ( $n = 2$ ) zu nennen.

**Tabelle 6:** Dauer der Berufstätigkeit als Lehrkraft.

Merkmal	Mittelwert (SD)	Median	Min	Max	n (gültig)
Beschäftigungsdauer	11.32 (10.25)	8.04	0	40.75	210

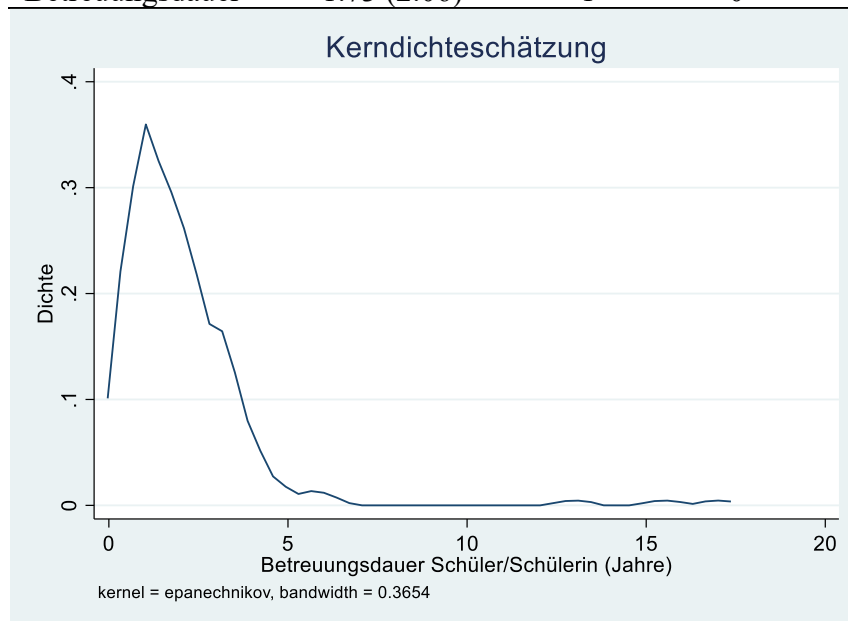


**Abbildung 3:** Geschätzte Verlaufsform der Dauer der Berufstätigkeit.

Es liegt mit einem Mittelwert von  $M = 11.32$  Jahren ( $SD = 10.25$  Jahre) und einem Median von rund 8 Jahren eine rechtsschiefe Verteilung der Beschäftigungsdauer vor, d.h. mehr als 50 % der Lehrkräfte sind kürzer beschäftigt als der Mittelwert von  $M = 11.32$  Jahren. Das beobachtete Minimum liegt bei 0 Jahren und das Maximum bei 40.75 Jahren (Tabelle 6 & Abbildung 3).

**Tabelle 7:** Betreuung des Schülers/der Schülerin als Lehrkraft in Jahren.

Merkmal	Mittelwert (SD)	Median	Min	Max	n (gültig)
Betreuungsdauer	1.73 (2.06)	1	0	17	202



**Abbildung 4:** Geschätzte Verlaufsform der Betreuungsdauer.

Es liegt mit einem Mittelwert von  $M = 1.73$  Jahren ( $SD = 2.06$  Jahre) und einem Median von einem Jahr eine rechtsschiefe Verteilung der Betreuungsdauer vor, d.h. mehr als 50 % der Lehrkräfte betreuen ihre Schüler und Schülerinnen kürzer als der Mittelwert von  $M = 1.73$  Jahren. Das beobachtete Minimum liegt bei 0 Jahren und das Maximum bei 17 Jahren (Tabelle 7 & Abbildung 4).

**Tabelle 8:** Beziehung zum Schüler/zur Schülerin.

Beziehung	Absolute Häufigkeit	Prozent
Klassenlehrer	100	46.95
Fachlehrer	50	23.47
Förderschullehrer	14	6.57
Sonstiges	49	23.00
Total	213	100

Mit  $n = 100$  (46.95%) sind die Beurteilenden überwiegend Klassenlehrkräfte. Danach folgen mit  $n = 50$  (23.47%) und  $n = 49$  (23%) Fach- bzw. sonstige Lehrkräfte und mit  $n = 14$  (6.57%) noch Förderschullehrkräfte (Tabelle 8).

## Ergebnisse der psychometrischen Analysen für das ABAS-III

Im Folgenden werden die internen Konsistenzen für die zehn Subskalen sowie das Ein- und Drei-Faktoren-Modell des ABAS-III psychometrisch anhand der Pilot-Stichprobe ( $N = 217$ ) evaluiert und die relevanten Kennwerte tabellarisch dargestellt. Hierfür wird Cronbachs Alpha ( $\alpha$ ) und McDonalds Omega ( $\omega$ ) bestimmt:

Interne Konsistenz mittels Cronbachs Alpha und McDonalds Omega:

**Tabelle 9:** Interne Konsistenzen für die 10 einzelnen Subskalen des ABAS-III.

<b>Subskala:</b>	<b>Gültiges n:</b>	<b><math>\alpha</math> (gerundet):</b>	<b><math>\omega</math> (gerundet):</b>
Kommunikation (KV)	206	.94	.94
Verhalten in der Öffentlichkeit (ViO)	210	.90	.89
Schulische Fertigkeiten (SF)	210	.96	.96
Verhalten im Schulalltag (VS)	207	.94	.95
Gesundheit und Sicherheit (GS)	210	.89	.90
Freizeitverhalten (FV)	209	.91	.91
Selbstfürsorge (SF)	211	.91	.92
Selbststeuerung (SS)	209	.95	.95
Sozialverhalten (SV)	209	.95	.95
Verhalten am Arbeitsplatz (VA)	7	.96	-

Für alle theoretischen Subskalen des deutschen ABAS-III liegt mit Alpha zwischen .90 und .96 sehr gute bis hervorragende interne Konsistenz vor. Einzig die Subskala „Gesundheit und Sicherheit“ fällt mit einem Alpha von .89 minimal ab (Tabelle 9). Insgesamt lässt sich anhand der internen Konsistenzen das theoretische Messmodell auf Ebene der jeweiligen Einzelskalen sehr plausibel begründen. Die alternative Betrachtung mittels Omega weist mit Werten  $\geq .90$  ebenfalls auf sehr gute Messgenauigkeiten der implizierten Einzelskalen hin.

Der Koeffizient lässt sich für die Subskala „Verhalten am Arbeitsplatz“ aufgrund der geringen Stichprobengröße ( $n = 7$ ) allerdings nicht berechnen (Tabelle 9).

**Tabelle 10:** Interne Konsistenz des Gesamtkonstrukts (ohne Verhalten am Arbeitsplatz).

<b>Gesamtkonstrukt (außer VA):</b>	<b>Gültiges n:</b>	<b><math>\alpha</math> (gerundet):</b>	<b><math>\omega</math> (gerundet):</b>
Adaptive Kompetenz	191	.99	.98

Auch das Gesamtkonstrukt der adaptiven Kompetenz wird für die deutsche ABAS-III Lehrerversion mit einem Alpha größer .95 mit einer hervorragenden internen Konsistenz abgebildet, weiterhin bestätigt sich die hervorragende interne Konsistenz ebenfalls plausibel durch ein Omega größer .95 (Tabelle 10). Daher wird messtheoretisch das Gesamtkonstrukt bestätigt.

**Tabelle 11:** Interne Konsistenzen der drei Domänen des ABAS-III (ohne Verhalten am Arbeitsplatz).

<b>Domäne:</b>	<b>Gültiges n:</b>	<b><math>\alpha</math> (gerundet):</b>	<b><math>\omega</math> (gerundet):</b>
Konzeptionelle Fähigkeiten	201	.98	.98
Soziale Kompetenzen	206	.96	.96
Praktische Kompetenzen	201	.97	.97

Alle drei übergeordneten Fähigkeitsdomänen des deutschen ABAS-III weisen mit Alpha größer .95 hervorragende interne Konsistenzen auf und werden durch Omega größer .95 zusätzlich plausibilisiert (Tabelle 11). Das theoretische Messmodell wird daher auch auf Ebene der drei Verhaltensdomänen jeweils hervorragend bestätigt.

#### Standardmessfehler (SEM) des Generalfaktors und der drei Domänen:

Im Anschluss an die Berechnung und Prüfung der internen Konsistenzen des Gesamtfaktors und der Verhaltensdomänen werden nun die daraus resultierenden Standardmessfehler (SEM) für diese latenten Konstrukte sowie die 95%-Konfidenzintervalle für die Punktschätzungen der wahren Testwerte bestimmt und dargestellt. Als Berechnungsgrundlage wird die Skalenreliabilität  $r_{xx}$  nach McDonald ( $\omega$ ) in folgender Formel zur Bestimmung des SEM verwendet (AERA, APA & NCME, 1985):

$$SEM = S * \sqrt{1 - r_{xx}}$$

wobei  $S$  = empirische Standardabweichung der Testwerteskala

Es ergeben sich die folgenden Standardmessfehler und 95%-Konfidenzintervalle für den wahren personenspezifischen Testwert der jeweiligen Domäne bzw. des Generalfaktors adaptiver Kompetenz. Zur besseren Einordnung der statistischen Unsicherheit der jeweiligen Schätzung wird für jedes Konstrukt die theoretische Spannweite (Range) der Skala dargestellt:

**Tabelle 12:** Standardmessfehler und 95%-Konfidenzintervalle der Schätzung für Generalfaktor und die drei Verhaltensdomänen.

<b>Konstrukt</b>	<b><math>S</math></b>	<b>SEM</b>	<b>95%-KI</b>	<b>Range</b>	<b><math>N</math></b>
Generalfaktor	79.03	10	$P \pm 19.6$	[0; 522]	213
Konzeptionelle Kompetenz	36.17	5.52	$P \pm 10.82$	[0; 195]	213
Soziale Kompetenz	18.01	3.50	$P \pm 6.83$	[0; 114]	213
Praktische Kompetenz	28.54	4.81	$P \pm 9.43$	[0; 213]	213

Für den Generalfaktor adaptiver Kompetenz ergibt sich ein plausibler Wertebereich für den tatsächlichen Testwert von  $\pm 19.6$  Punkten um den Schätzwert. Für konzeptionelle Kompetenz liegt diese Spanne bei  $\pm 10.82$  Punkten, für soziale Kompetenz bei  $\pm 6.83$  Punkten und für die praktische Domäne bei  $\pm 9.43$  Punkten (Tabelle 12). Zu beachten ist, dass die Standardmessfehler auf Grundlage der gesamten Pilot-Stichprobe bestimmt werden, d.h. keine *altersspezifische* Berechnung innerhalb von Subgruppen erfolgt. Dies ist einerseits dem geringen Stichprobenumfang geschuldet und andererseits der allgemeinen Komplexität der Darstellung. Eine altersspezifische Darstellung der Skalenreliabilität nach Cronbach und McDonald und der daraus abgeleiteten Standardmessfehler sowie 95%-Konfidenzintervalle erfolgt auf Grundlage der Normierungsstichprobe innerhalb des noch unveröffentlichten Manuals der deutschen Adaptation des ABAS-III.

#### Prüfung der Faktorenstruktur des Gesamtkonstrukts und der drei Domänen:

Im Folgenden werden die konfirmatorischen Faktoranalysen für die Pilotstichprobe für das Ein- und Dreifaktoren-Modell dargestellt, wobei Variante (a) sich auf die reduzierte Itemanzahl nach dem Kriterium der *Restreliabilität* bezieht und Variante (b) auf die personenspezifische Mittelwertaggregation der inhaltlichen Item-Pakete (*Item-Parceling*). Die zugrundeliegenden Items für Variante (a) werden im elektronischen Anhang zusätzlich tabellarisch anhand der für die Auswahl relevanten Restreliabilität aufgeschlüsselt.

#### Gesamtkonstrukt (1-Faktor-Modell)

##### Variante mit Reduktion auf zwei Items (a):

**Tabelle 12:** Modellgüte der 1-Faktor-CFA für Gesamtkonstrukt „adaptive Kompetenz“.

X <sup>2</sup> -Wert (p-Wert)	$\chi^2/df$ :	RMSEA:	CFI:	SRMR:	Gültige Fallzahl:
435.19 (df = 135) (p < .001)	3.22	.104	.77	.078	207

**Tabelle 13:** Faktorladungen der ausgewählten Indikatoren der jeweiligen Subskalen.

Faktor:	Item:	Ladung (standardisiert):
Adaptive Kompetenz	KV-13	.69
	KV-02	.20
	ViO-12	.57
	ViO-02	.48
	KT-11	.62
	KT-01	.47
	VS-20	.78
	VS-22	.51
	GS-13	.60
	GS-12	.19
	FV-15	.68
	FV-02	.43
	SF-07	.57
	SF-18	.49

SS-17	.73
SS-05	.46
SV-15	.72
SV-11	.61

Anmerkungen: KV= Kommunikationsverhalten, ViO = Verhalten in der Öffentlichkeit, KT = Kulturtechniken, VS = Verhalten in der Schule, GS = Gesundheit und Sicherheit, FV = Freizeitverhalten, SF = Selbstfürsorge, SS = Selbststeuerung, SV = Sozialverhalten.

In einer Maximum-Likelihood-basierten konfirmatorischen Faktorenanalyse für das einfaktorielle Gesamtkonstrukt resultiert mit einem *RMSEA* von .10, einem *CFI* von .77 und einem *SRMR* von .078 nach den üblichen Konventionen eine noch akzeptable Anpassung an die Stichprobenstruktur. Mit einem  $\chi^2$  (135) von 435 ( $p < .001$ ) wird die Nullhypothese einer perfekten Modellanpassung verworfen (Tabelle 9). Dieses Gütemaß ist allerdings mit Vorsicht zu interpretieren, da es sehr anfällig für die Größe der Stichprobe ist und eine perfekte Modellanpassung aus praktischer Sicht kein notwendiges Kriterium für ein gutes Messmodell ist (Hu & Bentler, 1999). Die standardisierten Faktorladungen liegen zwischen .19 und .78 ( $\bar{O} \sim .54$ ), wobei im Durchschnitt positive Ladungen mit größerer Magnitude überwiegen.

Dies deutet auf eine grundsätzliche Plausibilität der einfaktoriellen Messung allgemeiner adaptiver Kompetenz durch die ausgewählten Items hin, wenngleich die Modellgüte der CFA verbesserungswürdig ist (Tabelle 12 & Tabelle 13).

#### Variante mit Item-Parceling (b):

**Tabelle 14:** Modellgüte der 1-Faktor-CFA für Gesamtkonstrukt „adaptive Kompetenz“.

X <sup>2</sup> -Wert (p-Wert)	$\chi^2$ /df:	RMSEA:	CFI:	SRMR:	Gültige Fallzahl:
367(df = 27) (p < .001)	13.59	.244	.836	.06	213

**Tabelle 15:** Faktorladungen der gemittelten Indikatoren der jeweiligen Subskalen mittels Item-Parceling (b).

Faktor:	Subskala (gemittelt)	Ladung (standardisiert):
Adaptive Kompetenz	Kommunikationsverhalten (KV)	.89
	Verhalten in Öffentlichkeit (ViO)	.81
	Kulturtechniken (KT)	.77
	Verhalten in der Schule (VS)	.88
	Gesundheit und Sicherheit (GS)	.84
	Freizeitverhalten (FV)	.85
	Selbstfürsorge (SF)	.75
	Selbststeuerung (SS)	.91
	Sozialverhalten (SV)	.85

Die Modellgüte für das einfaktorielle Konstrukt „adaptiver Kompetenz“ weist in der alternativen Berechnung mittels Item-Parceling (b) mit  $\chi^2$  (27) = 367 ( $p < .001$ ) ebenfalls eine signifikant schlechtere Anpassung an die Stichprobe auf als das perfekt saturierte Modell.

Der Quotient aus  $\chi^2$ -Teststatistik und der Anzahl an Freiheitsgraden ( $df = 27$ ) überschreitet den kritischen Schwellenwert von maximal 5 für einen akzeptablen Modellfit mit 13.59 deutlich (Bollen & Long, 1993; Wheaton et al., 1977).

Der *RMSEA* ist mit .24 deutlich größer als bei der ersten Variante (a) der Berechnung und deutet absolut auf eine schlechte Modellanpassung hin, allerdings ist der *RMSEA* für Modelle mit geringer Anzahl an Freiheitsgraden ( $df = 27$ ) und geringem Stichprobenumfang systematisch nach oben verzerrt, weshalb er für dieses Modell mit Vorsicht zu interpretieren ist (Kenny et al., 2015). Der *CFI* hingegen fällt mit rund .84 relativ besser aus als in der ersten Variante (a) und weist auf einen noch akzeptablen Modellfit hin. Auch der *SRMR* ist mit .06 relativ kleiner als in der ersten Variante und induziert eine gute Modellanpassung an die Datenstruktur (Hu & Bentler, 1999).

Die standardisierten Faktorladungen der theoretischen Subskalen auf den gemeinsamen Faktor „adaptive Kompetenz“ liegen zwischen .75 und .91 ( $\bar{O} \sim .84$ ), was das theoretische Messmodell eines gemeinsamen Faktors hinter den Subskalen der jeweiligen Verhaltensbereiche deutlich plausibilisiert, da diese Ladungen alle positiv und stark ausfallen (Tabelle 14 & Tabelle 15).

### Domänen (3-Faktoren-Modell)

#### Variante mit Reduktion auf zwei Items (a):

**Tabelle 16:** Ergebnisse der 3-Faktoren-CFA für die drei Verhaltensdomänen.

$\chi^2$ -Wert (p-Wert)	$\chi^2/df$ :	RMSEA:	CFI:	SRMR:	Gültige Fallzahl:
404.3(df = 132) (p < .001)	3.06	.10	.795	.078	207

**Tabelle 17:** Faktorladungen der ausgewählten Indikatoren der jeweiligen Subskalen (a).

Faktor (Domäne):	Item:	Ladung (standardisiert):
Konzeptionell	KV-13	.64
	KV-02	.16
	SS-17	.69
	SS-05	.45
	KT-11	.64
	KT-01	.45
Sozial	FV-15	.73
	FV-02	.44
	SV-15	.77
	SV-11	.61
Praktisch	VS-20	.74
	VS-22	.50
	SF-07	.56
	SF-18	.46
	GS-13	.58
	GS-12	.16
	ViO-12	.59

Anmerkungen: KV= Kommunikationsverhalten, ViO = Verhalten in der Öffentlichkeit, KT = Kulturtechniken, VS = Verhalten in der Schule, GS = Gesundheit und Sicherheit, FV = Freizeitverhalten, SF = Selbstfürsorge, SS = Selbststeuerung, SV = Sozialverhalten.

Die konfirmatorische Faktorenanalyse für das dreifaktorielle Domänenmodell weist mit Berechnung nach Variante (a) einen *RMSEA* von .10, einen *CFI* von .795 und einen *SRMR* von .078 auf. Damit liegt eine noch akzeptable Modellanpassung an die Stichprobenstruktur vor.

Die Nullhypothese der perfekten Modellanpassung wird erneut mit einem  $\chi^2(132) = 404.3$  ( $p < .001$ ) verworfen, sollte allerdings aufgrund vorgenannter Kritik nicht überbewertet werden.

Der Quotient aus  $\chi^2$ -Teststatistik und der Anzahl an Freiheitsgraden liegt bei 3.06 und liegt damit unter dem konventionellen Schwellenwert von maximal 5 für eine akzeptable Modellanpassung (Bollen & Long, 1993; Wheaton et al., 1977). Daraus lässt sich ebenfalls die Plausibilität des postulierten Messmodells ableiten.

Die Faktorladungen für die ausgewählten zwei Items pro Subskala liegen insgesamt zwischen .16 und .77 ( $\emptyset \sim .54$ ), wobei die Ladungen für die Domäne „Konzeptionell“ zwischen .16 und .69 liegen ( $\emptyset \sim .51$ ), für „Sozial“ zwischen .44 und .77 ( $\emptyset \sim .64$ ) und für „Praktisch“ zwischen .16 und .74 ( $\emptyset \sim .51$ ). Insgesamt liegen innerhalb jeder Domäne hauptsächlich mittlere bis starke positive Ladungen auf den jeweiligen gemeinsamen Faktor vor. Dies plausibilisiert die Annahme eines theoretischen Messmodells über die drei Verhaltensdomänen (Tabelle 16 & Tabelle 17).

### Variante mit Item-Parceling (b):

**Tabelle 18:** Ergebnisse der 3-Faktoren-CFA für die drei Verhaltensdomänen.

$\chi^2$ -Wert (p-Wert)	$\chi^2/df$ :	RMSEA:	CFI:	SRMR:	Gültige Fallzahl:
329.59 (df = 24) (p < .001)	13.73	.245	.853	.057	213

**Tabelle 19:** Faktorladungen der gemittelten Indikatoren der jeweiligen Subskalen mittels Item-Parceling (b).

Faktor (Domäne):	Subskala (gemittelt):	Ladung (standardisiert):
Konzeptionell	Kommunikationsverhalten (KV)	.88
	Kulturtechniken (KT)	.78
	Selbststeuerung (SS)	.90
Sozial	Freizeitverhalten (FV)	.90
	Sozialverhalten (SV)	.90
Praktisch	Verhalten in Öffentlichkeit (ViO)	.83
	Verhalten in der Schule (VS)	.88
	Gesundheit und Sicherheit (GS)	.82
	Selbstfürsorge (SF)	.73

Die konfirmatorische Faktorenanalyse für die drei latenten Verhaltensdomänen (Konzeptionell, Sozial, Praktisch) weist mittels Indikatorbildung durch Item-Parceling (b) einen *RMSEA* von .245 auf, was eine deutlich schlechtere Modellanpassung suggeriert, als bei Variante (a). Allerdings gilt erneut aufgrund der geringen Anzahl an Freiheitsgraden und dem relativ geringen Stichprobenumfang, dass der *RMSEA* systematisch nach oben verzerrt ist und daher mit Vorsicht in Bezug auf die Modellgüte zu interpretieren ist (Kenny et al., 2015). Der *CFI* liegt bei  $\sim .85$  und der *SRMR* bei .057, was für eine ordentliche Modellanpassung an die Datenstruktur der Stichprobe zu bewerten ist und im Vergleich zum einfaktoriellen Modell (Gesamtfaktor Adaptive Kompetenz) für eine relativ bessere Anpassung spricht. Die Nullhypothese perfekter Anpassung muss mit einem  $\chi^2(24) = 329.59$  ( $p < .001$ ) deutlich verworfen werden, allerdings muss die generelle Kritik an der Aussagekraft dieses Gütemaßes erneut berücksichtigt werden. Der Quotient aus  $\chi^2$ -Teststatistik und Freiheitsgraden überschreitet den konventionellen Cut-Off „guter“ Modellanpassung von maximal 5 mit 13.73 deutlich (Tabelle 18). In der vorhergehenden Berechnungsvariante (a) wird dieser Grenzwert deutlich unterschritten und eine gute Anpassung des Modells induziert. Dies ist eine Konsequenz des *Item-Parceling*, da diese Methode die Anzahl zu schätzender Modellparameter im Vergleich zur vorherigen Variante mit jeweils zwei Items pro Subskala substantiell reduziert.

Die Ladungen der einbezogenen Indikatoren schwanken für den Faktor „Konzeptionell“ zwischen .78 und .90 ( $\bar{\emptyset} \sim .85$ ). Für den Faktor „Sozial“ weisen beide Indikatoren eine Ladung von .90 auf. Für die Domäne „Praktisch“ liegen die Ladungen zwischen .73 und .88 ( $\bar{\emptyset} \sim .815$ ). Für alle drei Verhaltensdomänen liegen daher einheitlich stark positive Ladungen der manifesten Indikatoren auf die jeweiligen latenten Faktoren vor. Diese Ladungsstruktur plausibilisiert das postulierte theoretische Messmodell (Tabelle 19).

## Fazit und Limitationen:

Es konnte gezeigt werden, dass das theoretische Messmodell des ABAS-III zur Erfassung adaptiver Kompetenzen auch in der deutschsprachigen Lehrerversion für 5 bis 21-Jährige in einer Pilot-Stichprobe von Kindern und Jugendlichen plausibel repliziert werden kann. Die interne Konsistenz nach Cronbach ist für das Gesamtkonstrukt, das Domänenmodell (beide jeweils ohne Verhalten am Arbeitsplatz) und die insgesamt 10 Einzelskalen als sehr gut bis hervorragend einzustufen. Dieser Befund stellt sich unter Verwendung des weniger restriktiven Omegas nach McDonald als robust dar, da beide Koeffizienten annähernd identisch ausfallen. Die Magnitude des Standardmessfehlers (SEM) ist für die Verhaltensdomänen und den Generalfaktor vorläufig als hinreichend klein für eine relativ präzise Individualdiagnostik zu bezeichnen, da die Spanne der Unsicherheit mit dem jeweiligen 95%-Konfidenzintervall im Verhältnis zur theoretischen Spannweite der jeweiligen Skala relativ klein ausfällt (Tabelle 12). Allerdings ist zu beachten, dass die Analysen global auf der *gesamten* Pilotierung stattfinden und keine Berechnung innerhalb der definierten Altersklassen des englischsprachigen Originals durchgeführt wird. Dies ist der geringen Stichprobengröße geschuldet. Daher ist der Befund nur als erster Hinweis auf eine brauchbare Individualdiagnostik mit dem ABAS-III (Lehrerversion) zu werten.

Das einfaktorielle Messmodell des Gesamtkonstrukts adaptiver Kompetenzen weist nach den üblichen Maßzahlen der Modellgüte ( $\chi^2$ , *RMSEA*, *CFI* und *SRMR*) allerdings nur eine akzeptable Modellanpassung auf (Tabelle 12 & Tabelle 14). Dieser Befund weicht von den Ergebnissen von Aricak und Oakland (2010) ab, die für die englischsprachige Lehrerversion des ABAS-II eine gute Modellanpassung aufzeigen können. Für das Drei-Faktoren-Domänenmodell (ohne Verhalten am Arbeitsplatz) wird in der hier vorliegenden Analyse allerdings eine relativ bessere Modellgüte erreicht (Tabelle 10). Dieser Befund steht konträr zur psychometrischen Analyse des ABAS-II, wo für das Domänenmodell adaptiver Kompetenzen keine brauchbare Modellanpassung aufgezeigt werden kann (Aricak & Oakland, 2010). Allerdings deckt sich der Befund der relativen Überlegenheit des Domänenmodells gegenüber dem Generalfaktormodell mit den psychometrischen Befunden des strukturgleichen ABAS-III in der englischsprachigen Lehrerversion (Harrison & Oakland, 2015).

Dies ist erfreulich, da das Domänenmodell messtheoretisch differenzierter ist, d.h. im Vergleich zum Generalfaktormodell eine Messung von drei übergeordneten Verhaltensbereichen systematisiert und erlaubt. Weiterhin liegt damit eine relative psychometrische Übereinstimmung mit dem normierten ABAS-III (Lehrerversion) im englischen Original vor (Harrison & Oakland, 2015).

Es muss weiterhin kritisch betrachtet werden, dass in dieser Analyse keine Untersuchung der drei hierarchischen Stufen von *Messinvarianz* (konfigural, metrisch, skalar) erfolgt, d.h. nicht geprüft wird, ob sich die postulierten Messmodelle systematisch für verschiedene Subgruppen der Stichprobe unterscheiden (z.B. zwischen weiblichen und männlichen Kindern und Jugendlichen). *Messinvarianz* ist eine Voraussetzung für sinnvolle Subgruppenvergleiche auf Unterschiede in der Ausprägung der latenten Faktoren (hier einmal Gesamtkonstrukt und die der drei Domänen). Aufgrund der geringen verbleibenden Stichprobengröße bei separater Modellschätzung in relevanten Subgruppen und der Tatsache, dass für die gesamte Stichprobe nur Modelle mit entweder zwei manifesten Items pro Subskala (a) oder inhaltlichem Item-Parceling durch Mittelwertaggregation (b) geschätzt werden können, erübrigen sich Berechnungen auf relevanten Teilstichproben und bleiben zukünftiger psychometrischer Forschung zum ABAS-III in der Lehrerversion vorbehalten.

Zusätzlich ist im absoluten Vergleich und der qualitativen Einordnung der allgemeinen Modellpassung zu berücksichtigen, dass die psychometrischen Vergleichswerte der englischsprachigen Originalversion mittels vorab berechneter *Rohwertsummen* auf den jeweiligen theoretischen Subskalen bestimmt sind, d.h. die Anzahl manifester Indikatoren in der CFA nicht den einzelnen Items des ABAS-III entsprechen, sondern den aggregierten individuellen Testwerten auf den Subskalen. Dadurch verringert sich einerseits die Anzahl final zu schätzender Parameter substanziell und andererseits wird aufgrund der hohen Skalenreliabilität der Einzelskalen und der Verwendung dieser aggregierten Scores im anschließenden Faktormodell unerklärte Varianz auf Ebene der einzelnen Items zugunsten geteilter Kovarianz der manifesten Skalenwerte verringert. Dadurch lassen sich die besseren Fit-Indizes in den referenzierten Publikationen zum englischen ABAS-III erklären.

Als weitere Limitation dieser Evaluation ist zu nennen, dass die untersuchte Pilot-Stichprobe in einem nicht-zufälligen Auswahlverfahren an Kölner Schulen rekrutiert wurde, und damit inferenzstatistische Generalisierungen auf eine größere Population von Schülerinnen und

Schülern mittels Signifikanzwerten und Konfidenzintervallen keine mathematische Grundlage besitzen. Eine grundsätzliche Plausibilität der Messmodelle und des eingesetzten Tests kann allerdings aufgrund der akzeptablen Fit-Indizes (insbesondere *SRMR*) und der insgesamt recht hohen und *einheitlich* positiven Faktorladungen sowie der sehr guten Alpha-Maße nach Cronbach und zusätzlich der Omega-Koeffizienten nach McDonald gefolgert werden.

## Literatur:

AERA, APA, & NCME (1985). *Standards for educational and psychological testing*. Washington, D. C.: American Psychological Association. p. 94.

Aricak, O., T. & Oakland, T. (2010). Multigroup Confirmatory Factor Analysis for the Teacher Form, Ages 5 to 21, of the Adaptive Behavior Assessment System–II. *Journal of Psychoeducational Assessment* 28(6), 578–584.

Bandalos, D. L. & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Hg.), *Advanced structural equation modeling: New developments and techniques*, 269-296. Mahwah, NJ: Lawrence Erlbaum.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9, 78 - 102.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park: Sage Publications.

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(3), 399–412. <https://doi.org/10.1111/bjop.12046>

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74, 121–135. <https://doi.org/10.1007/s11336-008-9098-4>

Hangauer, J., Worcester, J., & Armstrong, K. H. (2013). Models and Methods of Assessing Adaptive Behavior. In D. H. Saklofske, V. L. Schwann & C. R. Reynolds (Eds.), *The Oxford Handbook of Child Psychological Assessment* (Oxford Library of Psychology, S. 651–665). Oxford: Oxford University Press USA.

Harrison, P.L. & Oakland, T. (2015). *Adaptive Behavior Assessment System, Third Edition [Manual]*. Torrance, CA: Western Psychological Services.

Heber, R. (1959). A manual on terminology and classification in mental retardation. *American Journal of Mental Deficiency, Suppl* 64(2), 1–111.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

International Test Commission (2005). International Guidelines on Test Adaptation. URL: [www.intestcom.org](http://www.intestcom.org)

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486-507. <https://doi.org/10.1177/0049124114543236>

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*, 9, 151 - 173.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>

Universität zu Köln (2015). *Vorlesungsverzeichnis und Modulhandbuch für den Studiengang Bildungswissenschaften*. URL: [https://www.hf.uni-koeln.de/data/hflehrestudium/File/SSC\\_Paedagogik/Bildungswissenschaften/Modulhandbuecher/UzK\\_HF\\_MHB\\_G\\_MEd\\_BiWi\\_Version\\_3.pdf](https://www.hf.uni-koeln.de/data/hflehrestudium/File/SSC_Paedagogik/Bildungswissenschaften/Modulhandbuecher/UzK_HF_MHB_G_MEd_BiWi_Version_3.pdf)

Tassé, M. J. (2013). Adaptive Behavior. In M. L. Wehmeyer (Hrsg.), *The Oxford handbook of positive psychology and disability* (Oxford Library of Psychology, S. 105–115).

Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. (1977). Assessing Reliability and Stability in Panel Models. *Sociological Methodology*, 8 (1), 84-136.

Zurbriggen, C., & Orthmann Bless, D. (2017). Adaptive Behavior Assessment System II - eine erste Überprüfung der psychometrischen Eigenschaften der deutschen Erwachsenenversion - In: *Empirische Sonderpädagogik 9 (2017) 1*, S. 3-18.