

NEPS SURVEY PAPERS

Lena Nusser und Markus Messingschlager

ERFASSUNG KOGNITIVER GRUNDFÄHIG-
KEITEN BEI SCHÜLERINNEN UND SCHÜLERN
AN FÖRDERSCHULEN IN STARTKOHORTE 4
(KLASSE 9)

NEPS Survey Paper No. 33
Bamberg, Januar 2018

Survey Papers of the German National Educational Panel Study (NEPS)

at the Leibniz Institute for Educational Trajectories (LifBi) at the University of Bamberg

The NEPS Survey Paper Series provides articles with a focus on methodological aspects and data handling issues related to the German National Educational Panel Study (NEPS).

The NEPS Survey Papers are edited by a review board consisting of the scientific management of LifBi and NEPS.

They are of particular relevance for the analysis of NEPS data as they describe data editing and data collection procedures as well as instruments or tests used in the NEPS survey. Papers that appear in this series fall into the category of 'grey literature' and may also appear elsewhere.

The NEPS Survey Papers are available at <https://www.neps-data.de> (see section "Publications").

Editor-in-Chief: Corinna Kleinert, LifBi/University of Bamberg/IAB Nuremberg

Contact: German National Educational Panel Study (NEPS) – Leibniz Institute for Educational Trajectories – Wilhelmsplatz 3 – 96047 Bamberg – Germany – contact@lifbi.de

Erfassung kognitiver Grundfähigkeiten bei Schülerinnen und Schülern an Förderschulen in Startkohorte 4 (Klasse 9)

*Lena Nusser & Markus Messingschlager
Leibniz-Institut für Bildungsverläufe*

E-Mail-Adresse der Erstautorin:

lena.nusser@lifbi.de

Bibliographische Angaben:

Nusser, L. & Messingschlager, M. (2018). Erfassung kognitiver Grundfähigkeiten bei Schülerinnen und Schülern an Förderschulen in Startkohorte 4 (Klasse 9) (NEPS Survey Paper No. 33). Bamberg: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel. doi:10.5157/NEPS:SP33:1.0

Inhaltsverzeichnis

1. Einleitung.....	3
2. Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf Lernen und besondere Herausforderung in der Testung.....	3
3. Ziele des Beitrags.....	5
4. Methode.....	5
4.1 Stichproben	5
4.2 Instrumente.....	6
4.2.1 Erfassung der kognitiven fluiden Leistungsfähigkeit.....	6
4.2.2 Weitere Kontrollmaße.....	7
4.3 Methodisches Vorgehen	8
5. Ergebnisse	11
5.1 Fehlende Werte.....	11
5.2 Lösungswahrscheinlichkeiten.....	15
5.3 Summenwerte	16
5.4 Reliabilität.....	17
5.5 Konstruktvalidität	17
5.6 Messäquivalenz	18
6. Diskussion.....	21
Literatur.....	23

1. Einleitung

Das Nationale Bildungspanel (NEPS; Blossfeld, von Maurice & Schneider, 2011) erfasst in sechs Alterskohorten (von früher Kindheit bis ins hohe Erwachsenenalter) verschiedene Facetten kognitiver Kompetenzen (Weinert et al., 2011). Diese umfassen neben domänenspezifischen Kompetenzen (z.B. in den Bereichen Lesen und Mathematik) auch domänenübergreifende kognitive Fähigkeiten (vgl. Artelt, Weinert & Carstensen, 2013). Die kognitiven Grundfähigkeiten werden im NEPS auf der Grundlage einer Expertise von Brunner, Lang und Lüdtke (2014) mit zwei Kurztests zur Messung von Wahrnehmungsgeschwindigkeit und schlussfolgerndem Denken erfasst (Lang, Kamin, Rohr, Stünkel & Willinger, 2014). Diese Instrumente werden in allen Alterskohorten des NEPS eingesetzt¹. In Startkohorte 4 wurde dieser Test nicht nur an Regelschulen bei Schülerinnen und Schülern der Jahrgangsstufe 9 administriert, sondern ebenfalls an Förderschulen mit Schwerpunkt Lernen. Im Rahmen von Machbarkeitsstudien sollte evaluiert werden, ob und wenn ja, unter welchen Bedingungen Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf im Bereich Lernen (SPF-L) aussagefähig und vergleichbar in die längsschnittlichen Erhebungen des NEPS einbezogen werden können (Heydrich, Weinert, Nusser, Artelt & Carstensen, 2013). Da es sich bei dieser Schülergruppe um eine besonders heterogene Stichprobe mit sehr unterschiedlichen Kompetenzprofilen sowie einem breit streuenden Fähigkeitsspektrum handelt, stellt die valide und vergleichbare Messung kognitiver Fähigkeiten eine besondere Herausforderung dar. Eine zielgruppenspezifische Adaptation von Testmaterial kann unter Umständen dessen Validität beeinträchtigen (siehe Cormier, Altman, Shyyan & Thurlow, 2010). Daher wurde in den vorliegenden Studien zur Erfassung der kognitiven Grundfähigkeiten in Förderschulen und Regelschulen dasselbe Instrument unverändert eingesetzt. Dies soll eine möglichst große Vergleichbarkeit zwischen beiden Schülergruppen gewährleisten. Dieser Beitrag fasst zentrale Ergebnisse zur Datenqualität und Testgüte dieses Instruments bei Schülerinnen und Schülern mit SPF-L zusammen. Zudem werden Analysen zur Messinvarianz des eingesetzten Tests zur Erfassung der kognitiven Grundfähigkeiten nach Lang und Kollegen (2014) zwischen Schülergruppen an Hauptschulen und bei Schülerinnen und Schülern an Förderschulen im 9. Jahrgang berichtet.

2. Herausforderung in der Testung von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf Lernen

Für Schülerinnen und Schüler ist ein sonderpädagogischer Förderbedarf Lernen (SPF-L) anzunehmen, wenn sie „in ihrer Lern- und Leistungsentwicklung so erheblichen Beeinträchtigungen unterliegen, dass sie auch mit zusätzlichen Lernhilfen der allgemeinen Schulen nicht ihren Möglichkeiten entsprechend gefördert werden können“ (KMK - Sekretariat der Ständigen Konferenz der Kultusminister der Länder, 1999, S. 4). Diese Kinder und Jugendlichen stellen in Deutschland mit gut 44 % die größte Gruppe jener mit einem sonderpädagogischen Förderbedarf dar (Autorengruppe Bildungsbericht, 2016). Sie werden überwiegend in einem nach Förderschwerpunkten ausdifferenzierten und historisch gewachsenen Förderschulsystem beschult und unterrichtet. Das Vorliegen eines SPF-L wird als überdauernde und weitreichende Einschränkung bei der Bewältigung schulischer Anforderungen beschrieben (Klauer & Lauth, 1997), die sich im Besonderen bei dem „Erwerb

¹ Das Instrument für jüngere Kinder (bis zur zweiten Klassen) in den Startkohorten 1 und 2 weist eine leicht veränderte Darbietung sowie leichtere Items auf (vgl. Lang, Kamin, Rohr, Stünkel & Willinger, 2014).

kognitiv-verbaler und abstrakter Inhalte“ (Grünke & Grosche, 2014, S. 76) zeigt. In den Lernhilfe-Schulen soll diesen Schülerinnen und Schülern eine adäquate Förderung zuteilwerden, nicht nur um ihre kognitiven und sprachlichen Defizite auszugleichen, sondern auch ihre emotionale und soziale Entwicklung zu unterstützen (KMK, 1999). Die Schülerschaft an Förderschulen mit dem Schwerpunkt Lernen stellt sich als eine sehr heterogene Gruppe mit einer stark ausgeprägten Leistungsspanne sowie intra- und interindividuell verschiedenen Kompetenzprofilen dar (Bos, Müller & Stubbe, 2010; Müller, Stubbe & Bos, 2013). Auch zwischen Schülerinnen und Schülern an Regel- und Förderschulen gibt es bereits im Grundschulalter deutliche Unterschiede, zum Beispiel, hinsichtlich ihres sozialen Hintergrundes, aber auch betreffend ihrer kognitiven Grundfähigkeiten, sowohl im verbalen als auch im räumlichen Fähigkeitsbereich (Kocaj, Kuhl, Kroth, Pant & Stanat, 2014). Ähnliche Ergebnisse finden auch Bos, Gartmeier und Gröhlich (2009) in der Sekundarstufe: Schülerinnen und Schüler mit SPF-L schneiden bei Aufgaben zu räumlichen kognitiven Grundfähigkeiten deutlich schlechter ab als Schülergruppen an Haupt- und Realschulen. Kognitive Grundfähigkeiten stehen in einem engen Zusammenhang mit dem Erwerb domänenspezifischer Kompetenzen von Individuen (Weinert et al., 2011). Daher handelt es sich dabei um ein zentrales Maß, welches die Entwicklung von Kompetenzen erklären und beeinflussen kann.

Das schlechtere Abschneiden der Schülerinnen und Schüler mit SPF-L bei (Kompetenz-)Tests kann auf verschiedene Beeinträchtigungen, wie schwächere sprachliche Fähigkeiten, geringere Lernleistungen, aber auch eine reduzierte Aufmerksamkeitsspanne zurückgeführt werden (Grünke & Grosche, 2014). Da diese Schülergruppe bei Lernprozessen weniger effektive Strategien anwendet, beispielsweise bei der Organisation einzelner Schritte (Klauer & Lauth, 1997), können sich Schwierigkeiten durch eine oberflächliche Informationsentnahme insbesondere bei der Bearbeitung von Testaufgaben zeigen (Hessels, 2009; Scruggs, Bennion & Lifson, 1985). Zudem weisen diese Schülerinnen und Schüler ein weniger ausgeprägtes Instruktionsverständnis auf (Hessels-Schlatter, 2002; Zielinski, 1996), sodass sie Schwierigkeiten haben, vorgegebenen Anleitungen und Bearbeitungshinweisen zu folgen und korrekt auf die vorliegenden Testaufgaben anzuwenden. Daher ist eine Anleitung, die mit Blick auf Tests für Regelschülerinnen und -schüler konzipiert wurde, möglicherweise nicht hinreichend verständlich für Schülerinnen und Schüler mit SPF-L. Es ist anzunehmen, dass eine mangelhaft verstandene Instruktion, die weniger explizite Anleitungshinweise enthält, die Testbearbeitung und -performanz negativ beeinträchtigen kann (Nusser & Weinert, 2017; Wong, Wong & LeMare, 1982).

Neben den beschriebenen dispositionellen Ausgangsbedingungen der Schülerinnen und Schüler mit SPF-L können auch verschiedene Testmerkmale einen Einfluss auf die Testleistung haben. Sind Testaufgaben hinsichtlich ihrer Schwierigkeit nicht angemessen für die entsprechende Zielpopulation, können Reliabilität und Validität beeinträchtigt werden (vgl. Nusser, 2018). Eine differenzierte Erfassung kognitiver Grundfähigkeiten auch im niedrigen Leistungsbereich stellt eine zentrale Herausforderung für die Testung von Schülerinnen und Schüler mit SPF-L dar. Häufig werden bei der Konstruktion von Intelligenztests und anderen diagnostischen Instrumenten keine zielgruppenspezifischen Voraussetzungen berücksichtigt und somit gewisse Zugangsfertigkeiten, wie zum Beispiel Lesefertigkeit, Aufmerksamkeit sowie eine korrekte Anwendung von Testanweisungen, vorausgesetzt (Renner & Mickley, 2015). Ohne vorherige Prüfung der Güte eines Testverfahrens für diese Schülergruppen ist eine vergleichbare Interpretation der Testergebnisse nicht möglich und somit der

diagnostische Wert fragwürdig. Um den Zugang zu und das Verständnis von Testaufgaben für beeinträchtigte Schülergruppen zu verbessern sowie konstruktirrelevante und behinderungsbezogene Einschränkungen zu verringern, werden in vielen internationalen Studien im Rahmen von Kompetenztests Akkommodationen implementiert (Koretz & Barton, 2004; Thurlow, Ysseldyke & Silverstein, 1995). Wenn aber die Testadministration und -darbietung angepasst wird, kann sich die Testung zwischen verschiedenen Schülergruppen derart stark unterscheiden, dass eine übergreifende Standardisierung nicht mehr gegeben ist. Es stellt sich dann die Frage, ob durch die Adaptation die Testinstrumente möglicherweise so stark verändert wurden, dass sie nicht mehr zu vergleichbaren Messungen führen (Cormier et al., 2010). Die Diskussion von Akkommodationen und deren Effekten auf Validität und Vergleichbarkeit ist aufgrund heterogener Forschungsbefunde bislang noch nicht abschließend geklärt, sodass es keine eindeutige Entscheidung bezüglich der Implementation spezifischer Anpassungen für neuentwickelte Testverfahren gibt, die eine valide und vergleichbare Erfassung von kognitiven Fähigkeiten ermöglichen.

3. Ziele des Beitrags

Angesichts der Herausforderung eine valide Messung kognitiver Grundfähigkeiten bei Schülerinnen und Schülern mit SPF-L zu realisieren, soll im Folgenden die Erfassung bei dieser Zielpopulation im Vergleich zu anderen Schülergruppen näher untersucht werden. Kognitive Grundfähigkeiten stellen ein wichtiges Kontrollmaß für viele Forschungsfragen der empirischen Bildungsforschung dar. Eine valide und vergleichbare Messung des Konstrukts stellt eine zentrale Voraussetzung dar, um sinnvolle Interpretationen von möglichen Gruppenunterschieden vornehmen zu können. Unter der übergreifenden Perspektive einer möglichen Einbeziehung dieser Schülergruppe in längsschnittliche Large-Scale-Untersuchungen, wird in diesem Beitrag das für die NEPS-Studie entwickelte Instrument zur Messung von Wahrnehmungsgeschwindigkeit und schlussfolgerndem Denken anhand von Stichproben aus Regelschulen und Förderschulen Lernen in der 9. Jahrgangsstufe genauer betrachtet. Insbesondere werden die Testgüte des eingesetzten Verfahrens (wie Reliabilität und Validität) sowie die Vergleichbarkeit der Ergebnisse zwischen Förder- und Regelschülerinnen und -schülern in diesem Altersbereich näher untersucht. Zudem sollen besondere Bearbeitungsprobleme auf Individualebene identifiziert werden, welche die Vorgabe des Tests bei Schülerinnen und Schülern mit SPF-L erschweren.

4. Methode

4.1 Stichproben

Für die Ziehung der Förderschulstichprobe im Rahmen des NEPS wurde ein Oversampling angestrebt. Insgesamt konnten im gesamten Bundesgebiet 103 Institutionen in der Startkohorte 4 für die NEPS-Erhebung gewonnen werden (für Ziehungsprozeduren siehe Aßmann et al., 2011). Im Jahrgang 9 an Förderschulen gaben 1.186 Schülerinnen und Schüler ihr Einverständnis zur Teilnahme an der Studie. An der Erhebung im Frühjahr 2011 nahmen 990 Schülerinnen und Schüler an Förderschulen teil. Als Vergleichsgruppe wurden die Jugendlichen an Hauptschulen aus der NEPS-Haupterhebung an Regelschulen im Jahrgang 9 ausgewählt. Diese Stichprobe umfasste insgesamt 3.805 Schülerinnen und Schüler. Für 3.419 Schülerinnen und Schüler an Hauptschulen lagen Testdaten vor. Tabelle 1 gibt einen Überblick über die Zusammensetzung der Stichproben, die in den folgenden Auswertungen berücksichtigt wurden.

Tabelle 1: NEPS-Stichprobe Klasse 9

	Förderschule	Hauptschule
N	990	3.419
Anteil weiblich [%]	43,9	49,5
Alter <i>M (SD)</i>	16,0 (0,63)	15,9 (0,70)
Altersrange	14,4 – 18,9	13,8 – 19,1

4.2 Instrumente

4.2.1 Erfassung der kognitiven Grundfähigkeit

Die Erfassung der kognitiven Grundfähigkeit erfolgte über zwei Kurzindikatoren, den Bild-Zahlen-Test (NEPS-BZT) zur Erfassung der Wahrnehmungsgeschwindigkeit und den Matrizentest (NEPS-MAT) zur Erfassung des schlussfolgernden Denkens, die eigens für die kohärente, längsschnittliche Messung über den Lebenslauf für das NEPS entwickelt wurden (Lang et al., 2014). Beide Subtests werden von der 5. Jahrgangsstufe bis zum hohen Erwachsenenalter unverändert eingesetzt. Die Testgüte des NEPS-BZT und NEPS-MAT wurde in einer Ergänzungsstudie von Lang et al. (2014) auch für die vorliegende Altersgruppe (nicht aber für Schülerinnen und Schüler an Förderschulen) mit zufriedenstellenden Ergebnissen geprüft.

Der Subtest NEPS-BZT (Lang et al., 2014) erfasst als Speed-Test die Wahrnehmungsgeschwindigkeit der Personen. Dabei gilt es bildhaften Symbolen zugeordnete Zahlen möglichst schnell zu erkennen und richtig zu reproduzieren. Der Test besteht aus drei Runden. In jeder Runde wird ein Lösungsschlüssel vorgegeben, der neun verschiedenen Bildern die Ziffern von 1 bis 9 zuordnet. Unterhalb des Lösungsschlüssels befinden sich drei Zeilen mit insgesamt 31 sich wiederholenden Bildern des Lösungsschlüssels, wobei die zugehörigen Ziffern fehlen. Aufgabe ist es, in jeder Runde innerhalb von 30 Sekunden nacheinander zu den 31 Bildern möglichst viele der fehlenden Ziffern zu ergänzen. Aufgrund der begrenzten Bearbeitungszeit erzielen Personen mit einer schnelleren Wahrnehmungsgeschwindigkeit mehr richtige Antworten (Lang et al., 2014). Die Anzahl der korrekten Antworten wird daher als Indikator für die Wahrnehmungsgeschwindigkeit verwendet.

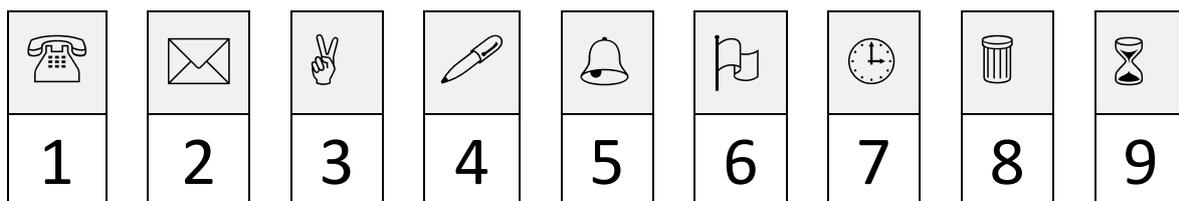


Abbildung 1: Beispiel Subtest NEPS BZT (Lang et al., 2014, S. 9)

Der Subtest NEPS-MAT (Lang et al., 2014) erfasst als Power-Test mit einem Matrizen-Test nach Raven (1977) das logisch-schlussfolgernde Denken (reasoning). Dieser Test umfasst insgesamt drei Aufgabenblöcke mit jeweils vier Matrizenaufgaben. Jede Matrizenaufgabe besteht aus mehreren waagrecht und senkrecht angeordneten geometrischen Elementen, wobei ein Feld der Matrix frei bleibt. Aufgabe ist es, die logischen Regeln, denen die Anordnung der geometrischen Elemente folgen, zu erschließen und aus jeweils sechs Antwortoptionen die richtige Ergänzung für das frei gebliebene Feld auszuwählen. Die Bearbeitungszeit von 3 Minuten pro Runde ist großzügig bemessen, da der NEPS-MAT die maximale Leistungsfähigkeit und nicht die Leistungsfähigkeit unter Zeitdruck erfassen soll (Lang et al., 2014). Als Ergebnis des Tests wird die Anzahl der korrekt gelösten Aufgaben genutzt.

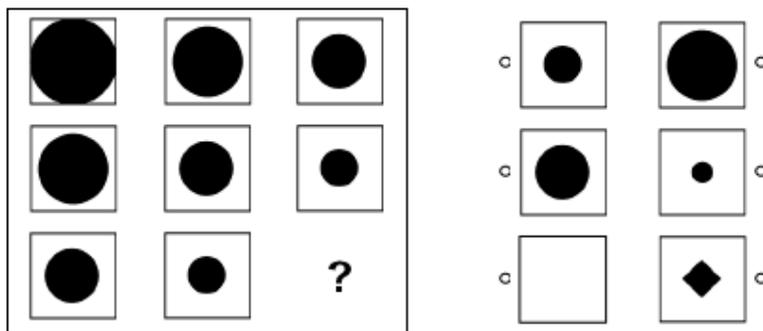


Abbildung 2: Beispiel Subtest NEPS-MAT (Lang et al., 2014, S. 10)

Um die Erläuterungen zu den Aufgaben verständlicher zu gestalten, wurde an Förderschulen zur visuellen Unterstützung ein Demonstrationsposter im Großformat A0 eingesetzt. Ausgebildete Testleiterinnen und Testleiter konnten daran die Aufgabenbearbeitung veranschaulichen und die Testbearbeitung im Rahmen der standardisierten Instruktion erläutern. Weitere Anpassungen der Testinhalte und -administration wurden nicht implementiert, da diese konstruktverändernd wirken können (z.B. mehr Testzeit). Mit dem Ziel der Vergleichbarkeit der erfassten kognitiven Grundfähigkeiten zwischen Schülerinnen und Schülern mit SPF-L und anderen Schülerinnen und Schülern wurde auf spezifische Anpassungen des Tests im Vergleich zur Regelschule verzichtet.

4.2.2 Weitere Kontrollmaße

Neben dem Instrument von Lang et al. (2014) zur Erfassung der kognitiven Grundfähigkeit wurden weitere Konstrukte erhoben. In der Förderschule wurden als zusätzliche Kontrollmaße zwei Subtests aus dem kognitiven Fähigkeitstests 4-12+R (KFT; Heller & Perleth, 2000) administriert. Hierbei wurden ein verbaler Subtest (V3: Wortanalogien) sowie ein nonverbaler Subtest (N2: Figurenanalogien) eingesetzt, die zur Kurzversion des Verfahrens gehören. Um eine adäquatere Passung der Itemschwierigkeiten zu den Fähigkeiten der Schülerinnen und Schüler an Förderschulen zu erreichen, wurden in einer sogenannten ‚out-of-level‘-Testung (Koretz & Barton, 2004) die Items der 8. Jahrgangsstufe administriert.

Zusätzlich wurde in beiden Stichproben die Lesegeschwindigkeit der Schülerinnen und Schüler erhoben. Hierzu wurde ein für das NEPS entwickelter Test eingesetzt (Zimmermann, Gehr, Artelt & Weinert, 2012). Innerhalb von zwei Minuten mussten die Schülerinnen und Schüler den Inhalt von 51 Aussagesätzen als wahr oder falsch beurteilen. Die Anzahl der korrekt gelösten Aufgaben wurde als Indikator für die Lesegeschwindigkeit der Personen herangezogen.

4.3 Methodisches Vorgehen

Zur näheren Betrachtung der Güte der Testung wurden unterschiedliche Zugänge gewählt. Hierbei wurden je nach Ziel einerseits die Bearbeitung der Testitems auf Individualebene und andererseits die Ergebnisse auf Populationsebene untersucht, um allgemeinere Aussagen über das Verfahren bei der Schülergruppe in 9. Klassen mit SPF-L generieren zu können. Beide Perspektiven sollen erlauben, sowohl das Testverfahren und seine Eignung für die Erfassung der kognitiven Grundfähigkeiten bei Schülerinnen und Schülern an Förderschulen Lernen zu beurteilen, als auch die Administration und damit einhergehenden Schwierigkeiten auf Individualebene näher zu beleuchten. Auf Grund der eingangs skizzierten Herausforderungen bei der Testung von Schülerinnen und Schülern mit SPF-L bezüglich des Verständnisses der Testinstruktionen und der korrekten Bearbeitung von Testaufgaben wurde das Auftreten fehlender Werte genauer betrachtet. Als ein Indikator für mögliche Verständnisschwierigkeiten wurde der Anteil nicht instruktionskonformer Antworten näher untersucht. Als nicht instruktionskonform wurden z.B. Antworten außerhalb des vorgegebenen Zahlenspektrums von 1 bis 9 beim NEPS-BZT sowie Mehrfachnennungen beim NEPS-MAT eingestuft. Für den Subtest NEPS-MAT wurde zusätzlich der Anteil nicht erreichter und nicht bearbeiteter Items kumuliert herangezogen, um Hinweise auf Bearbeitungsschwierigkeiten zu erhalten. Zur Identifikation individueller Bearbeitungsmuster im Sinne von Verweigerung oder auch Nichtverständnis erfolgte eine Betrachtung fehlender Werte auf Individualebene. Diesbezüglich stellte sich die Frage, wie viele Schülerinnen und Schüler die Aufgaben gar nicht oder den Test nicht gemäß der vorangestellten Testinstruktion korrekt bearbeitet haben.

Um die Passung zwischen den Itemschwierigkeiten und den Personenfähigkeiten einzuschätzen, wurden die Anteile richtiger Lösungen betrachtet (Anzahl korrekter Antworten an allen instruktionskonformen Antworten). Dies ermöglichte zusätzlich auffällige itemspezifische Schwierigkeiten zu identifizieren. Ergänzend wurden die deskriptiven Ergebnisse der erreichten Summenwerte beider Stichproben betrachtet. Aufgrund der beschriebenen Ausgangsvoraussetzungen der Schülerinnen und Schüler an Förderschulen Lernen kann davon ausgegangen werden, dass die gezeigten Leistungen in den Testverfahren geringer ausfallen als jene der Schülerschaft an Hauptschulen.

Für die Überprüfung der Reliabilität des NEPS-BZT wurde mit der Odd-Even-Methode die Split-Half-Reliabilität berechnet. Die interne Konsistenz des NEPS-MAT wurde mit Hilfe des Cronbach's Alpha kalkuliert.

Der Frage der Validität des Testverfahrens wurde sich anhand der zusätzlich erhobenen Kontrollmaße angenähert. Zur Überprüfung der konvergenten und divergenten Validität wurden die beiden Subtests NEPS-BZT und NEPS-MAT mit anderen Maßen zur Erfassung der kognitiven Grundfähigkeit sowie der Lesegeschwindigkeit in Beziehung gesetzt. Als Hinweis für die Validität des NEPS-MAT kann eine starke Korrelation mit dem nonverbalen Subtest des KFT gewertet werden, während eher geringe Zusammenhänge mit dem verbalen KFT-Subtest sowie der Lesegeschwindigkeit angenommen werden. Für den NEPS-BZT als Geschwindigkeitsmaß wird – trotz der verschiedenen Inhalte (nonverbal vs. verbal) – ein moderater Zusammenhang mit der Lesegeschwindigkeit erwartet.

Abschließend wurde die Messäquivalenz des NEPS-BZT und NEPS-MAT zwischen der Stichprobe der Schülerinnen und Schüler mit SPF-L und der Hauptschulstichprobe überprüft. Mit Hilfe konfirmatorischer Mehrgruppenvergleiche wurde untersucht, ob in den beiden

Stichproben die strukturellen Beziehungen zwischen Indikatoren und latenten Variablen vergleichbar waren und somit Messinvarianz besteht (Steenkamp & Baumgartner, 1998). Faktorielle Invarianz stellt eine zentrale Voraussetzung für vergleichende Analysen dar, da davon ausgegangen werden kann, dass keine systematischen Unterschiede des Messmodells über die Gruppen hinweg bestehen.

Für diese Modellierung orientierten wir uns an der Struktur des von Lang et al. (2014) postulierten Messmodells für die fluide kognitive Leistungsfähigkeit. Dieses geht von zwei latenten Faktoren (Wahrnehmungsgeschwindigkeit und schlussfolgerndes Denken) aus, die jeweils auf drei Indikatoren beruhen. Die manifesten Variablen sind die jeweiligen Summenscores je Runde der beiden Subtests. Da der Subtest zur Wahrnehmungsgeschwindigkeit (NEPS-BZT) auf Grund der Speedkomponente eine hohe Anzahl fehlender Werte für die jeweils letzten Items jeder Runde aufweist, fließen, wie bei Lang und Kollegen (2014), nicht Einzelitems, sondern Rundenscores (Anzahl korrekter Antworten pro Runde) in das Messmodell ein. Da die Items des NEPS-BZT alle vergleichbar hohe Lösungswahrscheinlichkeiten aufweisen (siehe Abschnitt 5.2.1) ist es nicht relevant, welche Items von den Schülerinnen und Schüler gelöst wurden – ausschlaggebend ist die Anzahl der korrekten Antworten. Im NEPS-MAT hingegen haben die Items unterschiedliche Schwierigkeiten, so dass eine entsprechende Spezifizierung im Modell es ermöglicht, die heterogenen Anforderungen an die Schülerinnen und Schüler und die Messinvarianz der einzelnen Items zu prüfen. Somit wird das Modell von Lang und Kollegen (2014) für die hier vorliegenden Analysen geringfügig angepasst (s. Abbildung 3). Das schlussfolgernde Denken (NEPS-MAT) wird anders als im ursprünglich Modell (Lang et al., 2014) mit den 12 Einzelitems als Indikatoren modelliert. Diese Modulierung erlaubt es, die unterschiedlichen Anforderungen und Inhalte der beiden Tests besser zu berücksichtigen. Zur Identifizierung der Modelle wurden die Varianzen der latenten Faktoren auf 1 fixiert.

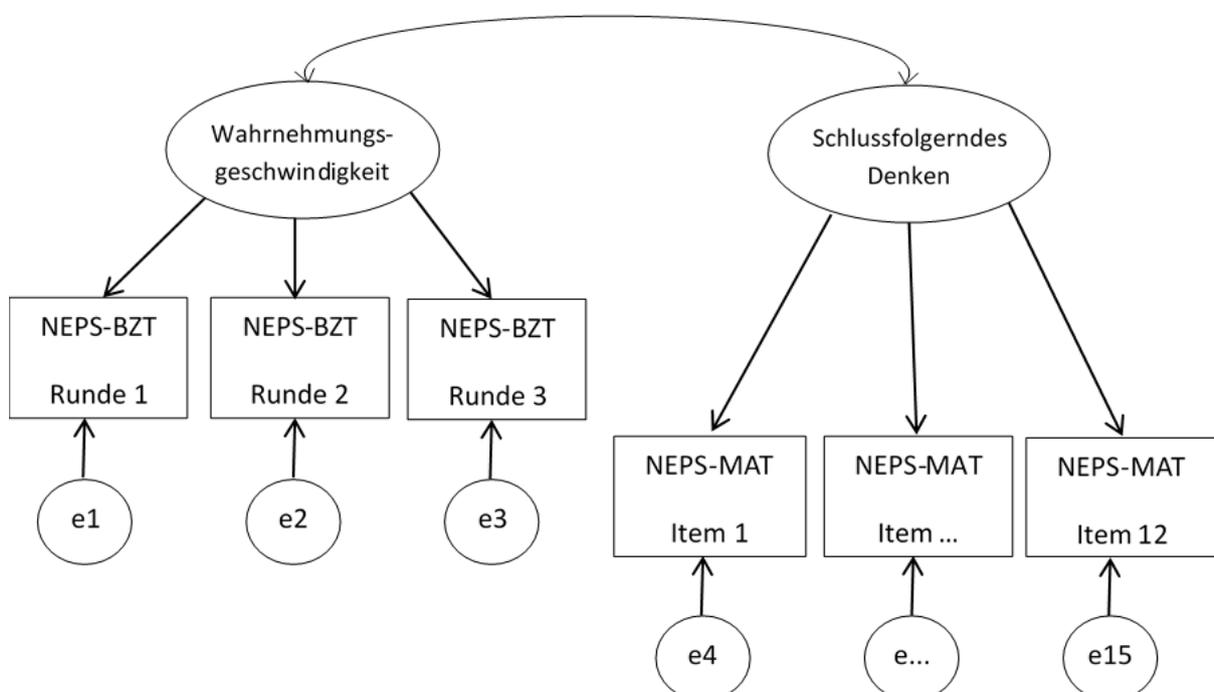


Abbildung 3: Messmodell der fluiden kognitiven Leistungsfähigkeit (adaptiert nach Lang et al., 2014)

Die in Abbildung 3 dargestellte Faktorstruktur wird als Baseline-Modell zunächst in den einzelnen Gruppen getrennt geschätzt, bevor in folgenden Analysen mit simultaner Schätzung

schrittweise die Invarianz der Parameter geprüft wird (Sass, 2011). Im Zuge der Prüfung der konfiguralen Messinvarianz werden identische Messbeziehungen zwischen Indikatoren und den Faktoren der Wahrnehmungsgeschwindigkeit und dem schlussfolgernden Denken bei gleicher Anzahl freier und fixierter Parameter angenommen. Sind ebenfalls die Faktorladungen in den beobachteten Gruppen invariant, spricht man von metrischer Messinvarianz. In diesem Fall kann von einer identischen Beziehung zwischen den Indikatoren und ihren latenten Faktoren in den Gruppen ausgegangen werden. Für die skalare Invarianz und damit starker faktorieller Invarianz, die vergleichende Analysen der latenten Mittelwerte erlaubt, werden die Intercepts (bei kontinuierlichen Daten) als auch die Thresholds (bei kategorialen Daten) der Indikatoren in den Gruppen gleichgesetzt. Sind diese Annahmen erfüllt, besteht in beiden Gruppen der gleiche Zusammenhang zwischen den Indikatoren (bzw. Testitems) und den gemessenen Konstrukten. Zudem existieren keine itemspezifischen Schwierigkeitsunterschiede zwischen den Gruppen (Sass, 2011). Es wurde ein Weighted Least Square Schätzer (WLSMV) verwendet, der im Vergleich zu Maximum Likelihood-Schätzern für Modelle mit kategorialen Daten akkuratere Faktorladungen schätzt (Li, 2014). Die genestete Struktur der Daten (Schülerinnen und Schüler in Schulen) wurde in den Analysen berücksichtigt, indem eine Korrektur des Standardfehlers erfolgte. Alle Modelle wurden mit der Software Mplus 7 (Muthén & Muthén, 2012) geschätzt.

Zur Beurteilung der Modellpassung wurden die gängigen Fitindizes herangezogen: RMSEA ($<0,05$), CFI ($>0,95$) und TLI ($>0,95$) (Hu & Bentler, 1999). Zur Überprüfung der Modellgüte der restriktiveren Modelle im Vergleich zu den Modellen mit weniger Restriktionen und somit mehr freien Parametern, wurde der Differenzwert der relevanten Fitindizes mit entsprechend etablierten Regeln genutzt: Δ RMSEA ($\leq -0,015$), Δ CFI ($\leq -0,010$) (Cheung & Rensvold, 2002) und Δ TLI ($\leq -0,010$; Chen, 2007).

5. Ergebnisse

5.1 Fehlende Werte

5.1.1 NEPS-BZT

Konstruktbedingt weisen die letzten Items jeder Runde dieses Geschwindigkeitstests einen hohen Anteil fehlender Werte auf, da die Schülerinnen und Schüler innerhalb der vorgegebenen Testzeit von 30 Sekunden diese nicht mehr bearbeiten können. Diese Tatsache hat vor allem Auswirkung auf das Testergebnis insgesamt. Erwähnenswert ist jedoch die Tatsache, dass es in beiden Stichproben Schülerinnen und Schüler gab, die insgesamt gar kein Item bzw. in einzelnen Runden kein Item bearbeitet haben.

In der Förderschulstichprobe fand sich eine Person (0,1 %), die gar kein Item bearbeitet hat, und weitere zwei Schülerinnen und Schüler (0,2 %), die insgesamt die Items zweier Runden unbeantwortet ließen. 21 der teilnehmenden Personen (2,1 %) gaben in einer Runde keine Antwort. In der Hauptschulstichprobe waren es zwei Personen (0,1 %), die den gesamten Test nicht bearbeiteten, fünf Personen (0,2 %) ließen zwei Runden unbeantwortet und 21 (0,6 %) weitere Personen haben eine Runde unbeantwortet gelassen.

Die Betrachtung übersprungener Items ist insofern interessant, da diese als Hinweise für bestimmte Bearbeitungsweisen dienen können. In beiden Stichproben fanden sich bei einem sehr geringen Anteil der Schülerinnen und Schüler gewisse Muster übersprungener Items, die vermuten lassen, dass die teilnehmenden Jugendlichen eine andere Strategie bei der Bearbeitung der Items genutzt haben als die instruierte aufeinanderfolgende, sequentielle Beantwortung der Items. So kann angenommen werden, dass sie nicht jedes der Items in der vorgegebenen Reihenfolge bearbeitet haben, sondern sich an dem dargebotenen Antwortschlüssel orientiert und die Items entsprechend dieser Anordnung beantwortet haben.

Der Anteil nicht instruktionskonformer Antworten beim NEPS-BZT war in beiden Stichproben sehr gering (vgl. Abbildungen 4 und 5). Der Großteil der Schülerinnen und Schüler gaben ausschließlich instruktionskonforme Antworten, während ein kleiner Teil der Gruppe vereinzelt ein bis wenige nicht instruktionskonforme Antworten abgaben. In der Förderschulstichprobe war lediglich für eine/n Schüler/in ausschließlich nicht instruktionskonforme Antworten bei diesem Subtest zu finden.

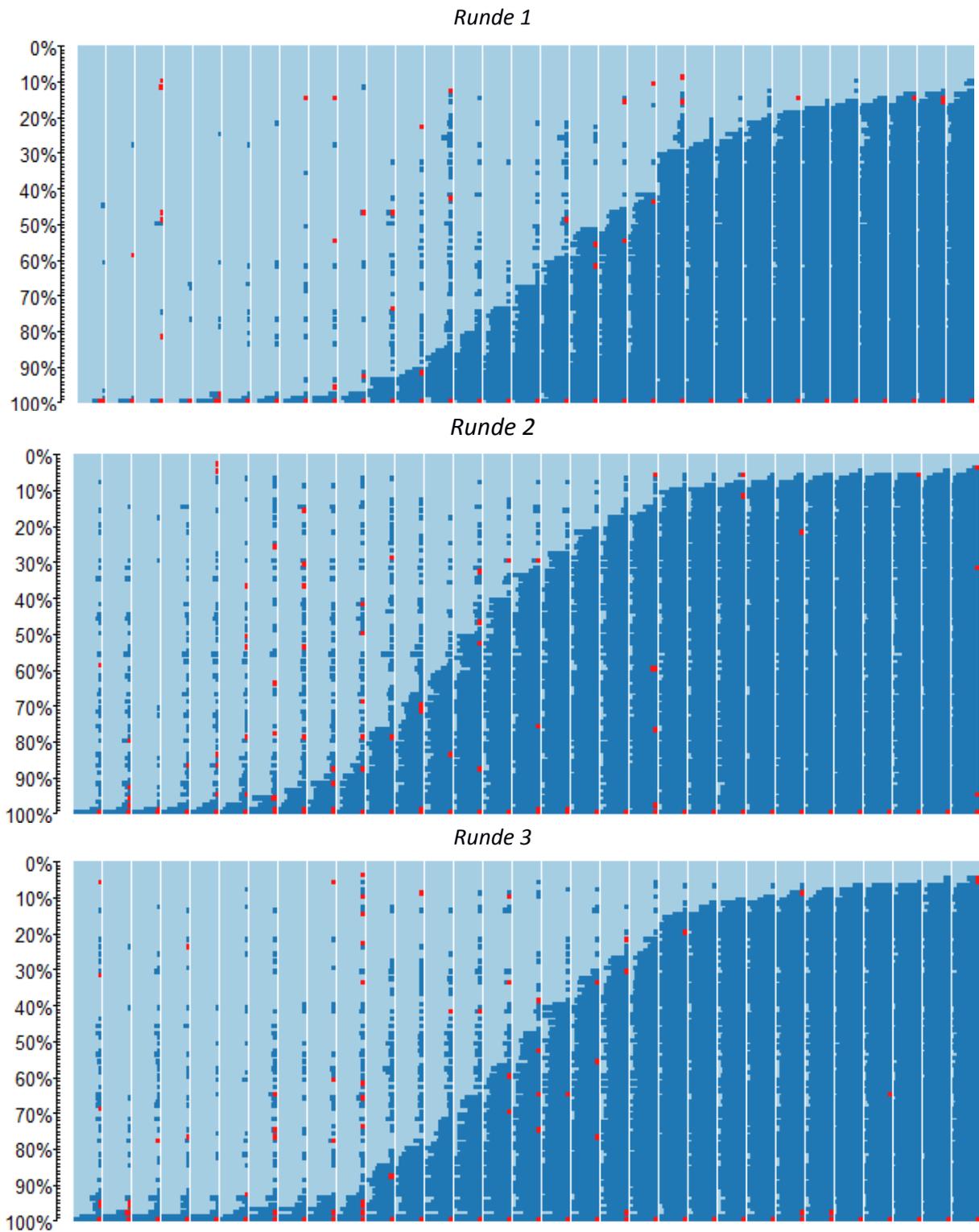


Abbildung 4. Antwortmuster in den drei Runden des NEPS-BZT für Schülerinnen und Schüler an Förderschulen.

Anmerkungen: x-Achse = Personen; y-Achse = Items 1-31; hellblau = bearbeitet; blau = nicht bearbeitet; rot = nicht instruktionskonforme Antwort; 10 Schülerinnen und Schüler pro Itemreihe.

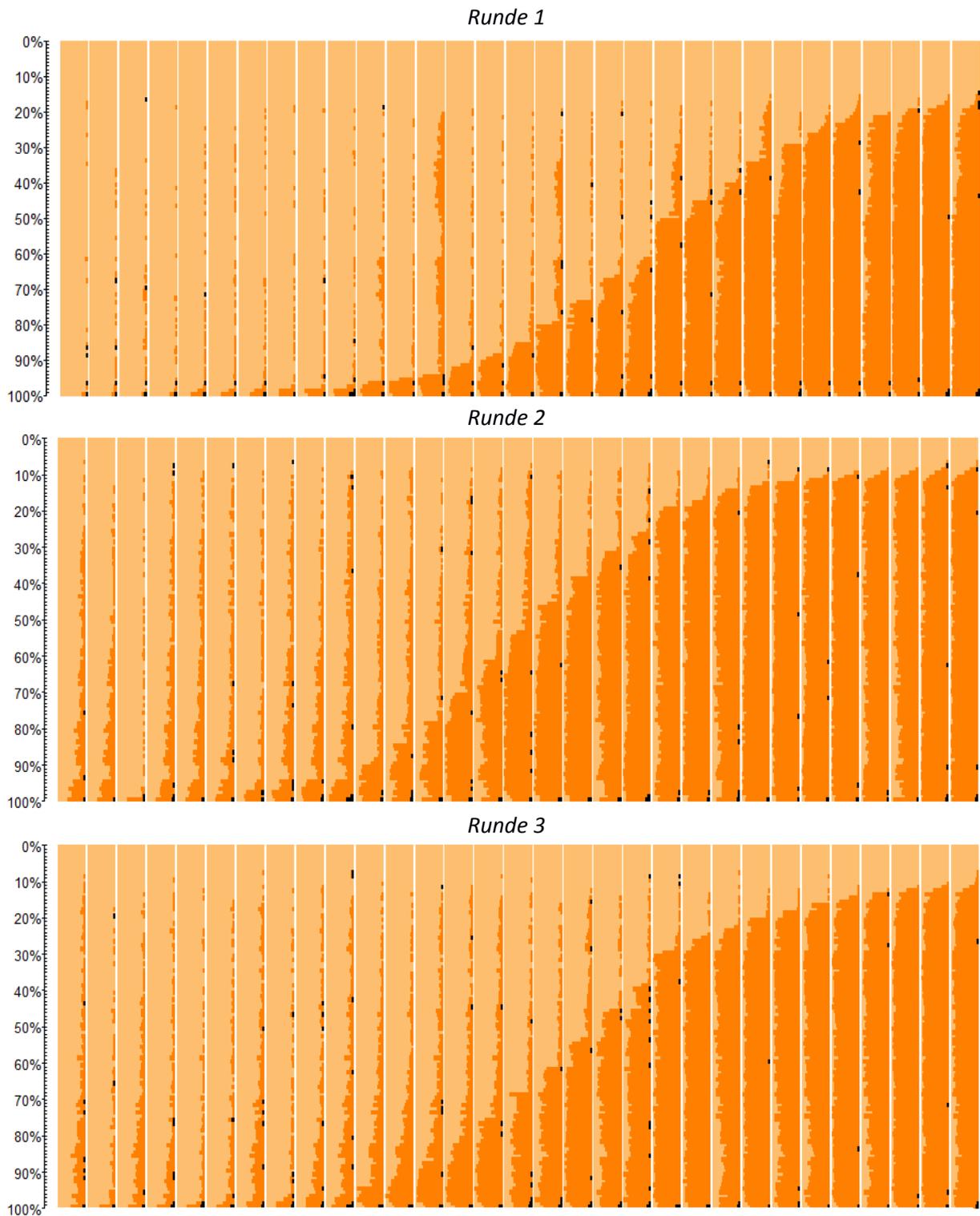


Abbildung 5. Antwortmuster in den drei Runden des NEPS-BZT für Schülerinnen und Schüler an Hauptschulen.

Anmerkungen: x-Achse = Personen; y-Achse = Items 1-31; hellorange = bearbeitet; orange = nicht bearbeitet; schwarz = nicht instruktionskonforme Antwort; 34 Schülerinnen und Schüler pro Itemreihe.

5.1.2 NEPS-MAT

Insgesamt zwei Schülerinnen und Schüler der Förderschulstichprobe (0,2 %) haben keines der Items des NEPS-MAT bearbeitet. Weitere drei Schülerinnen und Schüler (0,3 %) haben die Items einer Runde unbeantwortet gelassen. Für die Stichprobe an Hauptschulen war je eine Person zu finden (0,03 %), die alle drei bzw. zwei Runden nicht bearbeitet hat, während weitere 10 Schülerinnen und Schüler (0,3 %) in einer der drei Runden kein Item bearbeiteten.

Der Anteil fehlender Werte war in der Förderschulstichprobe im Vergleich zu den Schülerinnen und Schülern an Hauptschulen bei allen 12 Items erhöht. Vor allem die Anteile der nicht instruktionskonform beantworteten Items variierten zwischen den beiden Schulgruppen. Während in der Förderschulstichprobe von 2,12 bis 4,34 % der Antworten pro Item nicht gemäß der Instruktion abgegeben wurden, variierte der Anteil bei der Gruppe an Hauptschulen zwischen 0,94 bis 1,90 % (siehe Abbildung 6).

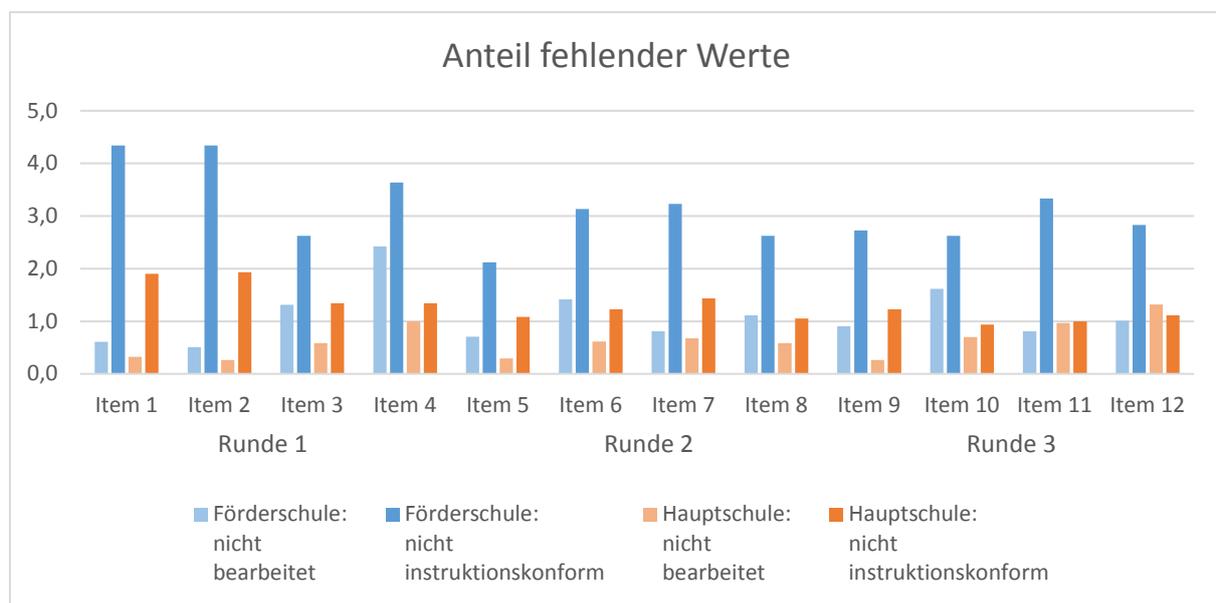


Abbildung 6: Anteil fehlender Werte pro Item des NEPS-MAT (blau = Förderschule; orange = Hauptschule).

Etwa 8,1 % der Schülerinnen und Schüler mit SPF-L ($n = 80$) gaben mindestens eine nicht instruktionskonforme Antwort, während in der Hauptschulstichprobe nur 3,2 % ($n = 111$) eine oder mehr nicht instruktionskonforme Antworten aufwiesen (siehe Abbildung 7). Eine Häufung dieses Antwortverhaltens war bei einzelnen Schülerinnen und Schülern festzustellen. In beiden Stichproben waren Personen zu finden, die alle Items nicht gemäß der Instruktion beantworteten.

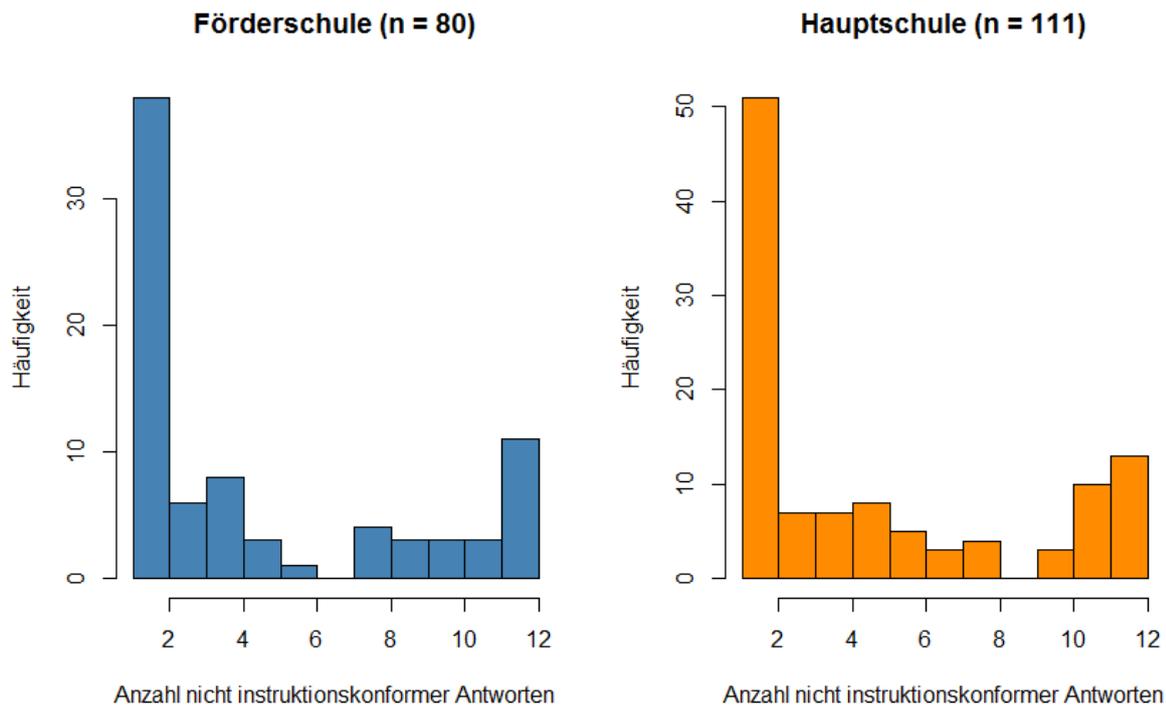


Abbildung 7. Anzahl nicht instruktionskonformer Antworten pro Schülerin bzw. Schüler mit mindestens einer nicht instruktionskonformer Antwort für NEPS-MAT in Förderschulen und Hauptschulen (blau = Förderschule; orange = Hauptschule).

5.2 Lösungswahrscheinlichkeiten

5.2.1 NEPS-BZT

Die einzelnen Items des NEPS-BZT sind konstruktbedingt sehr leicht zu lösen. Ähnlich wie in der Hauptschulstichprobe (Lösungswahrscheinlichkeit $p = 0,91$ bis $0,99$) lagen die Lösungswahrscheinlichkeiten in der Förderschulstichprobe für alle 93 Items zwischen $p = 0,87$ und $0,99$.

5.2.2 NEPS-MAT

Wie Abbildung 8 zu entnehmen ist, zeigten die Lösungswahrscheinlichkeiten für alle 12 Items dieses Subtests in beiden Stichproben ein vergleichbares Muster. Somit waren die relativen Schwierigkeiten der einzelnen Items innerhalb beider Stichproben vergleichbar. Dennoch wiesen alle Items für Schülerinnen und Schüler an den Förderschulen im Vergleich zu den Hauptschulen geringere Lösungswahrscheinlichkeiten auf. Von den Schülerinnen und Schülern mit SPF-L beantworteten durchschnittlich 20 % weniger Personen die Aufgaben richtig im Vergleich zur Hauptschulgruppe. Die Lösungswahrscheinlichkeiten der Schülerinnen und Schüler mit SPF-L variierten von $p = 0,20$ bis $0,76$, für die Hauptschulstichprobe von $p = 0,40$ bis $0,81$.

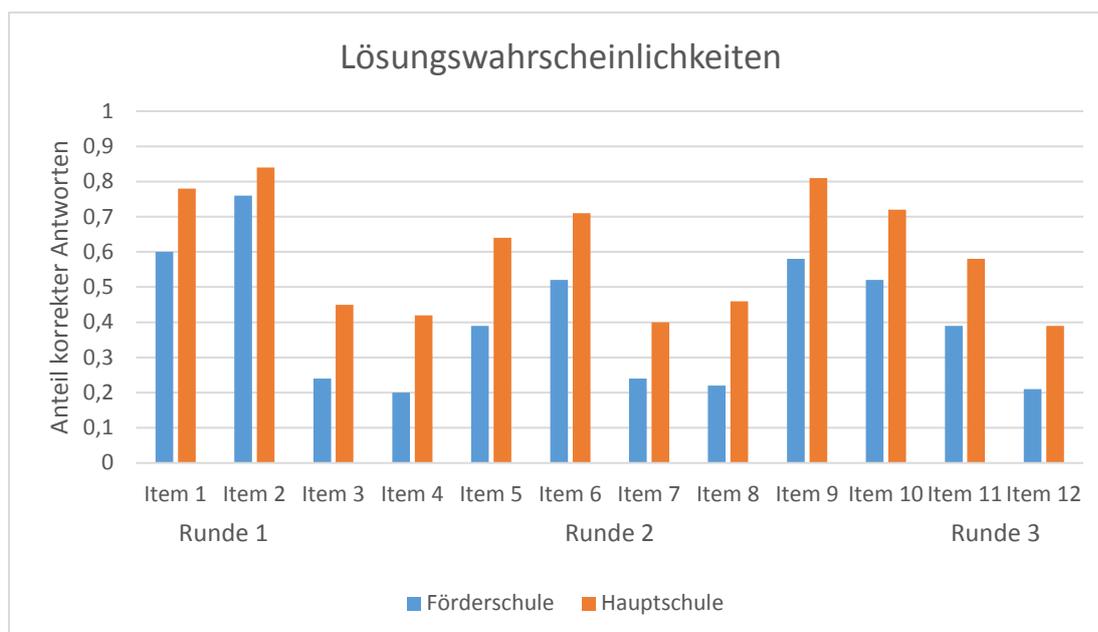


Abbildung 8: Lösungswahrscheinlichkeiten für NEPS-MAT im 9. Jahrgang.

5.3 Summenwerte

Tabelle 2 zeigt die Mittelwerte der Gesamtscores im NEPS-BZT und NEPS-MAT für beide Stichproben. Im Vergleich zur Hauptschule wurden in der Förderschule erwartungsgemäß im Mittel signifikant niedrigere Werte erreicht. Bei der Betrachtung der Verteilung der Gesamtwerte ist für die Förderschulstichprobe eine rechtsschiefe Verteilung zu erkennen (siehe Abbildung 9 und 10), es waren jedoch keine Bodeneffekte zu finden.

Tabelle 2: Kennwerte zur Verteilung der Gesamtwerte im NEPS-BZT und NEPS-MAT

	NEPS-BZT		NEPS-MAT	
	Förderschule	Hauptschule	Förderschule	Hauptschule
<i>M</i>	48,69	56,93	4,74	7,08
<i>SD</i>	14,80	15,78	2,56	2,60
<i>Min</i>	5	2	0	0
<i>Max</i>	93	93	12	12
<i>Teststatistik</i>	$t(1691,7) = 15,185; p < 0,001;$ $d = 0,54$		$t(1596,7) = 25,139; p < 0,001;$ $d = 0,91$	

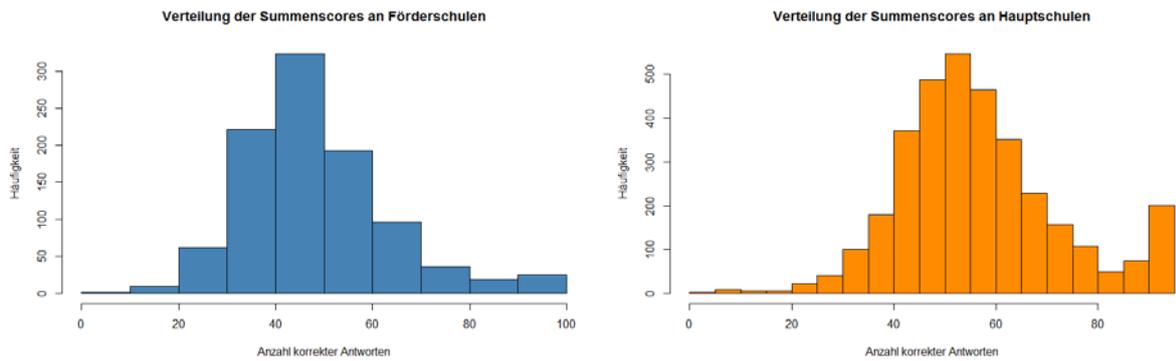


Abbildung 9. Verteilung NEPS-BZT (blau = Förderschule; orange = Hauptschule)

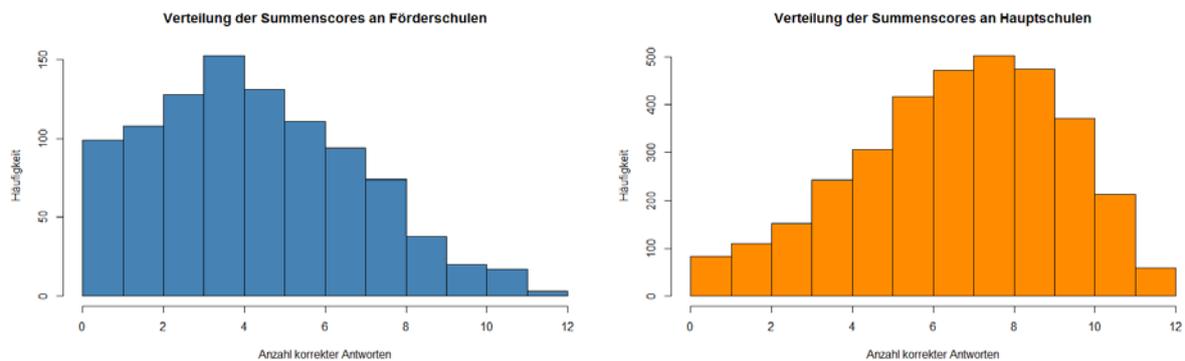


Abbildung 10. Verteilung NEPS-MAT (blau = Förderschule; orange = Hauptschule)

5.4 Reliabilität

Die Reliabilitäten des NEPS-BZT lagen in beiden Stichproben bei $r = 0,96$ und waren somit zufriedenstellend. Für den NEPS-MAT lagen die internen Konsistenzen für die Stichproben bei $\alpha = 0,79$ und somit in einem akzeptablen Bereich.

5.5 Konstruktvalidität

Da diese Machbarkeitsstudie an Förderschulen innerhalb des NEPS nicht als klassische Validierungsstudien angelegt war, ist eine Überprüfung der Konstruktvalidität mit den vorliegenden Daten nur eingeschränkt möglich. Hinweise auf die konvergente und divergente Validität des NEPS-BZT und NEPS-MAT liefern Korrelationen zwischen diesen Kurzindikatoren und den Subtests des Kognitiven Fähigkeitstests 4-12+R (Heller & Perleth, 2000) sowie der Lesegeschwindigkeit (Zimmermann et al., 2012). Der KFT wurde im 9. Jahrgang nur in der Förderschule administriert. Dementsprechend kann hier lediglich von Ergebnissen aus der Förderschule berichtet werden. Da nicht alle Summenwerte normalverteilt waren, wurde der Rangkorrelationskoeffizient nach Spearman (Spearman's Rho) berechnet.

5.5.1 NEPS-BZT

Insbesondere für den NEPS-BZT ist die Beurteilung der konvergenten und divergenten Validität nur sehr eingeschränkt möglich, da in den Erhebungen kein weiteres Instrument zur Erfassung der Wahrnehmungsgeschwindigkeit eingesetzt wurde. Während der NEPS-BZT als Speed-Test konstruiert ist, ist der KFT eher als Power-Test zu betrachten (Heller & Perleth, 2000). Folglich sind die Zusammenhänge zwischen dem NEPS-BZT und den Subtests des KFT

relativ gering, wobei diese zu dem nonverbalen Subtest des KFT (N2) stärker ausgeprägt ist ($r = 0,10$, $p < 0,05$). Die schwächste und nicht signifikante Korrelation besteht zwischen dem NEPS-BZT und dem KFT V3 ($r = 0,01$). Auch dies ist erwartungsgemäß, da der KFT V3 als Wortschatztest eine ausgeprägte kristalline Komponente aufweist, während der NEPS-BZT fluide kognitive Fähigkeiten erfasst. Der Zusammenhang des NEPS-BZT mit der Lesegeschwindigkeit fällt deutlich größer aus. Beiden Tests liegt eine Speed-Komponente zugrunde, die sich in den moderaten signifikanten Korrelationen in beiden Stichproben von $r = 0,26$ bzw. $r = 0,24$ ausdrückt. Diese beiden Zusammenhänge unterscheiden sich nicht signifikant voneinander ($z = 0,59$; $p = 0,28$; siehe Tabelle 3).

5.5.2 NEPS-MAT

Der Zusammenhang zwischen dem NEPS-MAT und dem nonverbalen Subtest des KFT ist erwartungsgemäß höher ($r = 0,42$) als zwischen dem NEPS-MAT und dem verbalen Subtest des KFT ($r = 0,33$). Die ausgeprägte Korrelation zwischen dem NEPS-MAT und dem KFT N2 kann als Hinweis auf die konvergente Validität des Tests gewertet werden. Ein Zusammenhang mit der Lesegeschwindigkeit ist in der Hauptschulstichprobe nicht vorhanden, jedoch gibt es eine kleine signifikante Korrelation in der Förderschulstichprobe ($r = 0,11$), die sich auch signifikant von dem Zusammenhang in der Hauptschulstichprobe unterscheidet ($z = 2,78$; $p < 0,05$; siehe Tabelle 3).

Tabelle 3: Korrelationen zwischen den Kurzindikatoren zur Erfassung der kognitiven Grundfähigkeiten (NEPS-BZT und NEPS-MAT) und weiteren Kontrollvariablen

Förderschule	NEPS-MAT	LG	KFT N2	KFT V3
NPS-BZT	0,07*	0,26*	0,10*	0,01
NEPS-MAT		0,11*	0,42*	0,33*
Hauptschule				
NPS-BZT	0,07	0,24*		
NEPS-MAT		0,01		

Anmerkungen. LG: Lesegeschwindigkeit, KFT N2: Figurenanalogie; KFT V3: Wortanalogie; * $p < .05$.

5.6 Messäquivalenz

Zunächst wurde für beide Stichproben das aufgestellte Messmodell (siehe Abbildung 1) in einer konfirmatorischen Faktorenanalyse spezifiziert. Zusätzlich wurde eine Korrelation der Kovarianzen Items 6 und 10 modelliert. Diese beiden Items weisen inhaltlich starke Ähnlichkeiten bezüglich der Matrizengestaltung als auch der zu identifizierenden logischen Abfolge auf, sodass ein systematischer Zusammenhang der Messfehler angenommen und in das Modell aufgenommen wird. Sowohl für die Förderschulgruppe als auch die Hauptschulgruppe zeigt das Modell sehr gute Fitindices (siehe Tabelle 4). Für die Indikatoren des NEPS-MAT fallen niedrige Ladungen von $< 0,6$ auf. Bereits in der Ergänzungsstudie von Lang et al. (2014) sind ähnlich niedrige Ladungen bei der Modellierung des NEPS-MAT mit drei Rundenwerten als Indikatoren berichtet. In beiden Stichproben ist kein Zusammenhang

zwischen den beiden latenten Faktoren Wahrnehmungsgeschwindigkeit (NEPS-BZT) und schlussfolgerndes Denken (NEPS-MAT) zu beobachten.

Der Mehrgruppenvergleich zur konfiguralen Messinvarianz kann mit guten Fitindices verifiziert werden (siehe Tabelle 5). Auch die metrische Messinvarianz mit identischen Faktorladungen in beiden Gruppen führt nicht zu einer signifikanten Modellverschlechterung ($\Delta\text{TLI} = 0,006$; $\Delta\text{CFI} < 0,002$; $\Delta\text{RMSEA} = 0,002$). Die Annahme gleicher Intercepts bzw. gleicher Thresholds für die skalare Invarianz kann nicht bestätigt werden ($\Delta\text{TLI} = 0,017$; $\Delta\text{CFI} = 0,018$; $\Delta\text{RMSEA} = 0,004$). Werden die Items 3 und 4 nicht beiden Gruppen frei geschätzt, kann partielle skalare Invarianz angenommen werden ($\Delta\text{TLI} = 0,007$; $\Delta\text{CFI} = 0,008$; $\Delta\text{RMSEA} = 0,002$). Diese beiden Items sind bereits mit sehr geringen Faktorladungen aufgefallen (siehe Tabelle 4).

Tabelle 4: Standardisierte Angaben der Faktorladungen, Faktorkorrelationen und Fitindices.

Förderschule (n = 990)			Hauptschule (n = 3.327)		
Faktorladungen			Faktorladungen		
	WG	SD		WG	SD
NEPS-BZT Runde 1	0,615*		NEPS-BZT Runde 1	0,669*	
NEPS-BZT Runde 2	0,890*		NEPS-BZT Runde 2	0,835*	
NEPS-BZT Runde 3	0,804*		NEPS-BZT Runde 3	0,827*	
NEPS-MAT Item 1		0,643*	NEPS-MAT Item 1		0,569*
NEPS-MAT Item 2		0,560*	NEPS-MAT Item 2		0,557*
NEPS-MAT Item 3		0,289*	NEPS-MAT Item 3		0,177*
NEPS-MAT Item 4		0,318*	NEPS-MAT Item 4		0,338*
NEPS-MAT Item 5		0,415*	NEPS-MAT Item 5		0,386*
NEPS-MAT Item 6		0,597*	NEPS-MAT Item 6		0,560*
NEPS-MAT Item 7		0,479*	NEPS-MAT Item 7		0,439*
NEPS-MAT Item 8		0,369*	NEPS-MAT Item 8		0,423*
NEPS-MAT Item 9		0,604*	NEPS-MAT Item 9		0,704*
NEPS-MAT Item 10		0,581*	NEPS-MAT Item 10		0,575*
NEPS-MAT Item 11		0,415*	NEPS-MAT Item 11		0,580*
NEPS-MAT Item 12		0,506*	NEPS-MAT Item 12		0,547*
Faktorkorrelationen			Faktorkorrelationen		
WG	1,00		WG	1,00	
SD	0,059	1,00	SD	0,094*	1,00
Fitindices			Fitindices		
χ^2	df	χ^2/df	χ^2	df	χ^2/df
123,839	88	1,41	241,404	88	2,74
TLI	CFI	RMSEA	TLI	CFI	RMSEA
0,960	0,967	0,020	0,959	0,951	0,023

Anmerkung. WG = Wahrnehmungsgeschwindigkeit, SD = schlussfolgerndes Denken; * p<0,05.

Tabelle 5: Prüfung der Messinvarianz.

Modell	χ^2	df	χ^2/df	TLI	CFI	RMSEA	Δ TLI	Δ CFI	Δ RMSEA
Konfigurale Invarianz	358,021	176	2,03	0,955	0,962	0,022			
Metrische Invarianz	362,586	191	1,90	0,961	0,964	0,020	0,006	0,002	0,002
Skalare Invarianz	463,838	204	2,27	0,944	0,946	0,024	0,017	0,018	0,004
Partielle skalare Invarianz	413,429	202	2,05	0,954	0,956	0,022	0,007	0,008	0,002

6. Diskussion

Dieses NEPS-Survey Paper betrachtete die Gütekriterien sowie die Messinvarianz eines Verfahrens zur Erfassung kognitiver Grundfähigkeiten, welches für das NEPS von Lang und Kollegen (2014) entwickelt wurde. Im Fokus standen eine Stichprobe von Schülerinnen und Schüler, die eine 9. Klasse an Förderschulen Lernen besuchten sowie eine Vergleichsgruppe an Hauptschulen. Zur Beurteilung der Testgüte wurden verschiedene Aspekte untersucht. Dazu gehören das Auftreten fehlender bzw. nicht instruktionskonformer Werte, die Lösungswahrscheinlichkeiten, die Reliabilität der Skalen, divergente sowie konvergente Validität als auch die Überprüfung der Messinvarianz.

Der NEPS-BZT ist ein Symbol-Zahlen-Test, der die Wahrnehmungsgeschwindigkeit der Zielpersonen erfasst. Die Bearbeitung des Subtests gemäß der Testinstruktion schien für Schülerinnen und Schüler an Förderschulen keine spezifische Herausforderung darzustellen. Der Anteil unplausibler bzw. nicht instruktionskonformer Angaben war nicht auffällig erhöht. Die Schwierigkeit der Items war auf Grund der Testanlage als niedrig einzuschätzen, was sich auch in den Lösungswahrscheinlichkeiten von über $p = 0,87$ ausdrückte. Auch die Summenscores der richtig gelösten Aufgaben ließen für die Gruppe der Förderschülerinnen und -schüler keine Bodeneffekte vermuten, sodass davon ausgegangen werden kann, dass dieses Format nicht mit besonderen gruppenspezifischen Herausforderungen einhergeht, welche die Erfassung der Wahrnehmungsgeschwindigkeit beeinträchtigen würden. Die Reliabilität war sehr hoch. Erwartungskonform konnte ein signifikanter, moderater Zusammenhang mit dem Kontrollmaß Lesegeschwindigkeit festgestellt werden sowie niedrige bzw. keine Korrelationen mit den Subtests des Kognitiven Fähigkeitstest 4-12+R.

Die Erfassung des schlussfolgernden Denkens unter dem Einsatz des NEPS-MAT als Matrizentest führte zu ähnlichen Ergebnissen. Auf Grund der Betrachtung der fehlenden und nicht instruktionskonformen Antworten, die zwar für die Schülerinnen und Schüler mit SPF-L im Vergleich zu Lernenden an Hauptschulen höher ausfielen, ist jedoch nicht davon auszugehen, dass es grundsätzliche Schwierigkeiten bei der Bearbeitung der Testaufgaben gab. Schülerinnen und Schüler mit SPF-L erreichten geringere Punktzahlen als die Gruppen, die eine Hauptschule besuchten. Dieser Befund schlug sich auch in den Lösungswahrscheinlichkeiten nieder. Der Test schien im unteren Fähigkeitsbereich, und damit für die besondere Gruppe der Schülerinnen und Schüler mit SPF-L, nicht ausreichend zu differenzieren. Die Reliabilität fiel für beide Stichproben lediglich in einen akzeptablen Bereich, war jedoch vergleichbar mit der Ergänzungsstudie von Lang et al. (2014). Die Autoren begründen die relativ geringe interne Konsistenz des NEPS-MAT in ihrer Studie mit der

Heterogenität des Tests hinsichtlich der Schwierigkeit der Testitems (siehe auch Abschnitt 5.2).

Insgesamt waren die Ergebnisse zur konvergenten und divergenten Validität erwartungskonform. Die Korrelationsmuster mit zusätzlichen Kontrollmaßen (Lesegeschwindigkeit sowie verbaler und nonverbaler Subtest des Kognitiven Fähigkeitstest 4-12+R) schienen die Validität des Testverfahrens zu stützen, wobei hier kein Vergleich mit den Korrelationsmustern der Hauptschulstichprobe möglich war.

Es wurden nur geringe Einschränkungen für die Vergleichbarkeit der Testergebnisse festgestellt. Orientiert am adaptierten Modell nach Lang et al. (2014) konnte in allen Substichproben faktorielle Validität angenommen werden, da die Modellpassungen sehr gute Fitindices aufwiesen. In Mehrgruppenvergleichen konnte die konfigurale und metrische Messinvarianz zwischen den Stichproben an Förderschulen und Hauptschulen bestätigt werden. Werden Schwellenwerte zwei der 12 Items (16.6 %) nicht über die beiden Gruppen hinweg restringiert, kann partielle skalare Messinvarianz angenommen werden. Trotz dieser freien Parameter können Mittelwertsvergleiche zwischen den Gruppen sinnvoll interpretiert werden, da die Mehrzahl der Items für den Subtest NEPS-MAT messinvariant sind (Byrne, Shavelson & Muthén, 1989; Steenkamp & Baumgartner, 1998).

In Anbetracht der vorgestellten Ergebnisse kann der Test als angemessen für einen Einsatz bei dieser Zielpopulation (Schülerinnen und Schüler in 9. Klassen an Förderschulen Lernen) beurteilt werden. Es kann davon ausgegangen werden, dass diese Schülerinnen und Schüler die Testinstruktion ausreichend verstehen, um die Testaufgaben im intendierten Sinne zu bearbeiten. Die Messung führt zu guten bis akzeptablen Reliabilitäten und scheint valide die kognitiven Grundfähigkeiten dieser Personen zu erfassen. Die Voraussetzungen für vergleichende Analysen mit Hauptschulstichproben sind gegeben. Obwohl der Test bzw. die beiden Subtests nicht mit Blick auf Schülerinnen und Schüler mit SPF-L entwickelt (Lang et al., 2014; Renner & Mickley, 2015) und in Rahmen der NEPS-Erhebung keine spezifischen Anpassungen (Koretz & Barton, 2004; Thurlow et al., 1995) implementiert wurden. Außer einer visuellen Unterstützung des Instruktionsverständnisses durch ein Demonstrationsposter hat der Einsatz des Testverfahrens zu aussagekräftigen Ergebnissen geführt.

Es handelt sich beim NEPS-BZT und NEPS-MAT um Kurzindikatoren für die Wahrnehmungsgeschwindigkeit und das schlussfolgernde Denken. Diese Verfahren sind jedoch weder für die Gesamtpopulation noch für die spezifische Gruppe von Schülerinnen und Schülern mit SPF-L für individualdiagnostische Zwecke geeignet (Lang et al., 2014). Eine Diagnostik von spezifischen kognitiven Einschränkungen ist daher nicht möglich. Dieses Instrument kann jedoch für Aussagen über Gruppen sinnvoll genutzt werden.

Literatur

- Artelt, C., Weinert, S. & Carstensen, C. H. (2013). Assessing competencies across the lifespan within the German National Educational Panel Study (NEPS)-Editorial. *Journal for educational research online*, 5 (2), 5–14.
- Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A. et al. (2011). Sampling designs of the National Educational Panel Study: challenges and solutions. *Zeitschrift für Erziehungswissenschaft*, 14 (S2), 51–65. doi:10.1007/s11618-011-0181-8
- Autorengruppe Bildungsbericht. (2016). *Bildung in Deutschland 2016: ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*.
- Blossfeld, H.-P., von Maurice, J. & Schneider, T. (2011). The National Educational Panel Study: need, main features, and research potential. *Zeitschrift für Erziehungswissenschaft*, 14 (2), 5–17. doi:10.1007/s11618-011-0178-3
- Bos, W., Gartmeier, M. & Gröhlich, C. (Hrsg.). (2009). *KESS 7. Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen zu Beginn der Jahrgangsstufe 7*. Münster: Waxmann.
- Bos, W., Müller, S. & Stubbe, T. C. (2010). Abgehängte Bildungsinstitutionen: Hauptschulen und Förderschulen. In G. Quenzel & K. Hurrelmann (Hrsg.), *Bildungsverlierer. Neue Ungleichheiten* (S. 375–397). Wiesbaden: Springer.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research* (Methodology in the social sciences). New York: Guilford Press.
- Brunner, M., Lang, F. R. & Lüdtke, O. (2014). *Erfassung der fluiden kognitiven Leistungsfähigkeit über die Lebensspanne im Rahmen der National Educational Panel Study: Expertise* (NEPS Working Paper No. 42.). Bamberg: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel. Zugriff am 16.1.2017. Verfügbar unter: https://www.neps-data.de/Portals/0/Working%20Papers/WP_XLII.pdf
- Byrne, B. M., Shavelson, R. J. & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological bulletin*, 105 (3), 456.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14 (3), 464–504. doi:10.1080/10705510701301834
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9 (2), 233–255. doi:10.1207/S15328007SEM0902_5
- Cormier, D. C., Altman, J., Shyyan, V. & Thurlow, M. L. (2010). A summary of the research on the effects of test accommodations: 2007-2008. Technical Report 56. *National Center on Educational Outcomes, University of Minnesota*.
- Grünke, M. & Grosche, M. (2014). Lernbehinderung. In G.W. Lauth, M. Grünke & J.C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen* (S. 76–89). Göttingen: Hogrefe.
- Heller, K. A. & Perleth, C. (2000). *KFT 4 - 12 + R: kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision; Manual*. Beltz Test. Verfügbar unter: https://books.google.de/books?id=Nv0Y_gAACAAJ
- Hessels, M. G. P. (2009). Estimation of the predictive validity of the HART by means of a dynamic test of geography. *Journal of Cognitive Education and Psychology*, 8 (1), 5–21. doi:10.1891/1945-8959.8.1.5

- Hessels-Schlatter, C. (2002). A dynamic test to assess learning capacity in people with severe impairments. *American Journal on Mental Retardation*, 340–351.
- Heydrich, J., Weinert, S., Nusser, L., Artelt, C. & Carstensen, C. H. (2013). Including students with special educational needs into large-scale assessments of competencies: Challenges and approaches within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online*, 5 (2), 217–240.
- 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. doi:10.1080/10705519909540118
- Klauer, K. J. & Lauth, G. W. (1997). Lernbehinderungen und Leistungsschwierigkeiten bei Schülern (Enzyklopädie der Psychologie, Serie Pädagogische Psychologie). In F.E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 701–738). Göttingen: Hogrefe.
- KMK - Sekretariat der Ständigen Konferenz der Kultusminister der Länder (Hrsg.). (1999). *Empfehlungen zum Förderschwerpunkt Lernen* (Beschluss der Kultusministerkonferenz vom 01.10.1999.). Berlin. Verfügbar unter: <http://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2000/sopale.pdf>
- Kocaj, A., Kuhl, P., Kroth, A. J., Pant, H. A. & Stanat, P. (2014). Wo lernen Kinder mit sonderpädagogischem Förderbedarf besser? Ein Vergleich schulischer Kompetenzen zwischen Regel- und Förderschulen in der Primarstufe. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66 (2), 165–191. doi:10.1007/s11577-014-0253-x
- Koretz, D. & Barton, K. (2004). Assessing students with disabilities: Issues and evidence. *Educational Assessment*, 9 (1–2), 29–60. doi:10.1080/10627197.2004.9652958
- Lang, F. R., Kamin, S., Rohr, M., Stünkel, C. & Willinger, B. (2014). *Erfassung der fluiden kognitiven Leistungsfähigkeit über die Lebensspanne im Rahmen des Nationalen Bildungspanels: Abschlussbericht zu einer NEPS-Ergänzungsstudie* (NEPS Working Paper No. 43.). Bamberg: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel.
- Li, C.-H. (2014). *The performance of MLR, USLMV, and WLSMV estimation in structural regression models with ordinal variables*. Michigan State University. Zugriff am 9.3.2017. Verfügbar unter: https://etd.lib.msu.edu/islandora/object/etd%3A3268/datastream/OBJ/download/The_performance_of_MLR_USLMV_and_WLSMV_estimation_in_structural_regression_models_with_ordinal_variables.pdf
- Müller, S., Stubbe, T. C. & Bos, W. (2013). Leistungsheterogenität angemessen berücksichtigen: Konzeption der Kompetenzmessung an Förderschulen mit dem Förderschwerpunkt Lernen im Rahmen von PARS-F. *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven*, 17, 264–296.
- Muthén, B. O. & Muthén, L. K. (2012). *Mplus - Version 7*. Los Angeles, CA.
- Nusser, L. (2018). *Valide und vergleichbare Erfassung bildungsrelevanter Konstrukte bei Schülerinnen und Schülern mit einem sonderpädagogischen Förderbedarf Lernen*. Bamberg: Otto-Friedrich-Universität.
- Nusser, L. & Weinert, S. (2017). Instructions in test-taking: An appropriate approach for students with special educational needs. *Journal of Cognitive Education and Psychology*, 16 (3), 227–240. doi:http://dx.doi.org/10.1891/1945-8959.16.3.227
- Raven, J. C. (1977). *Standard Progressive Matrices: Sets A, B, C, D & E*. San Antonio, TX: Harcourt.

- Renner, G. & Mickley, M. (2015). Berücksichtigen deutschsprachige Intelligenztests die besonderen Anforderungen von Kindern mit Behinderung? *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 64, 88–103.
- Sass, D. A. (2011). Testing Measurement Invariance and Comparing Latent Factor Means Within a Confirmatory Factor Analysis Framework. *Journal of Psychoeducational Assessment*, 29 (4), 347–363. doi:10.1177/0734282911406661
- Scruggs, T. E., Bennion, K. & Lifson, S. (1985). An analysis of children's strategy use on reading achievement tests. *The Elementary School Journal*, 85 (4), 479–484.
- Steenkamp, J.-B. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25 (1), 78–107.
- Thurlow, M. L., Ysseldyke, J. E. & Silverstein, B. (1995). Testing accommodations for students with disabilities. *Remedial and Special Education*, 16 (5), 260–270. doi:10.1177/074193259501600502
- Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T. & Carstensen, C. H. (2011). Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft*, 14 (S2), 67–86. doi:10.1007/s11618-011-0182-7
- Wong, B. Y. L., Wong, R. & LeMare, L. (1982). The effects of knowledge of criterion task on comprehension and recall in normally achieving and learning disabled children. *Journal of Educational Research*, 76, 119–126.
- Zielinski, W. (1996). Lernschwierigkeiten (Enzyklopädie der Psychologie, Serie Pädagogische Psychologie). In F.E. Weinert (Hrsg.), *Psychologie des Lernens und der Instruktion* (Band 2, S. 369–402). Göttingen: Hogrefe.
- Zimmermann, S., Gehrler, K., Artelt, C. & Weinert, S. (2012). *The assessment of reading speed in grade 5 and grade 9. Status: 2012* (Research Data). Bamberg: University of Bamberg, National Educational Panel Study (NEPS). Zugriff am 31.3.2017. Verfügbar unter: https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/1-0-0/com_rs_2012_en.pdf

Diese Arbeit nutzt Daten des Nationalen Bildungspanels (NEPS): Startkohorte Klasse 9, doi:10.5157/NEPS:SC4:1.0.0. Die Daten des NEPS wurden von 2008 bis 2013 als Teil des Rahmenprogramms zur Förderung der empirischen Bildungsforschung erhoben, welches vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde. Seit 2014 wird NEPS vom Leibniz-Institut für Bildungsverläufe e.V. (LifBi) an der Otto-Friedrich-Universität Bamberg in Kooperation mit einem deutschlandweiten Netzwerk weitergeführt.