

Sprachkompetenz und Mathematikleistung – Empirische Untersuchung sprachlich bedingter Hürden in den Zentralen Prüfungen 10

Susanne Prediger · Nadine Wilhelm · Andreas Büchter · Erkan Gürsoy · Claudia Benholz

Zusammenfassung Anhand der Zentralen Prüfung 10 Mathematik in Nordrhein-Westfalen wird durch Varianz-, Kovarianz- und Regressionsanalysen gezeigt, dass sprachliche Kompetenz unter allen sozialen und sprachlichen Faktoren den stärksten Zusammenhang zur Mathematikleistung hat. DIF-Analysen sowie Analysen schriftlicher Bearbeitungen auf Basis von Interviewanalysen ermöglichen ein tieferes Verständnis des starken Zusammenhangs und damit einen Beitrag zur empirischen Untersuchung von sprachlich bedingten Hürden in den abschlussrelevanten Prüfungen. Sie verweisen insbesondere auch auf die kognitive Funktion von Sprache jenseits von Lesehürden.

Schlüsselwörter Mathematikleistung · Sprachkompetenz · Sprachlich bedingte Hürden · DIF-Analyse

Mathematics Subject Classification (2010) 97D60 · 97D70

Language proficiency and mathematics achievement – Empirical investigation of obstacles in a high stakes test, the central exam ZP10

Abstract The article shows by analysis of (co-)variance in high stakes test items (of the central exam ZP 10 in North-Rhine Westphalia) that language proficiency is the background factor with the strongest connection to mathematics achievement, among all social and linguistic background factors. A differential functional analysis, analysis of written products and clinical interviews offer a deeper understanding of this connection and therefore a contribution to the empirical investigation of obstacles for students with low language proficiency in high stakes tests which refer also to the cognitive function of language beyond reading obstacles.

Key words Mathematics achievement · Language Proficiency · Language obstacles · DIF-Analysis

1. Einleitung

Ausgangspunkt des Artikels ist der Befund international vergleichender Schulleistungsstudien, dass schulische Leistungen in Deutschland stärker als in anderen Ländern an den familiären Hintergrund der Schülerinnen und Schüler gebunden sind (Baumert u. Schümer 2001). Während jedoch die meisten empirischen Studien und offiziellen Statistiken in Deutschland auf herkunftsbezogene soziale und sprachbiographische Faktoren fokussieren, legen internationale Befunde nahe, auch die Sprachkompetenz als Faktor stärker einzubeziehen, und zwar nicht nur als Lesekompetenz, sondern in einem breiteren Sinne (Secada 1992; Abedi 2006).

Die vorliegende Studie untersucht daher anhand der nordrhein-westfälischen Zentralen Prüfungen am Ende der Klasse 10 (ZP10-Mathematik) im Jahr 2012, welche der sozialen und sprachlichen Faktoren den stärksten Zusammenhang zur Mathematikleistung haben. Dazu wurden in einem ersten Schritt die Mathematikleistungen von 1495 Schülerinnen und Schülern in Beziehung gesetzt zu Daten zum Migrationshintergrund, zum sozioökonomi-

schen Status, zur Mehrsprachigkeit und zum Zeitpunkt des Deutscherwerbs, aber auch zur Sprachkompetenz und Lesekompetenz.

Itemanalysen ermöglichen in einem zweiten Schritt darüber hinaus, den vorgefundenen starken Zusammenhang zwischen Mathematikleistung und Sprachkompetenz durch eine gründliche Spezifizierung der sprachlich bedingten Hürden in den Items genauer zu verstehen. Als sprachlich bedingt werden dabei zum einen Hürden bezeichnet, die bei der Aufgabenbearbeitung direkt in der Prüfungssituation aus sprachlichen Gründen entstehen, z. B. durch Leseschwierigkeiten. Zum anderen werden darunter aber auch konzeptuelle und prozessuale Hürden gefasst, an denen sprachlich schwache Lernende auch im besonderen Maße scheitern, die jedoch aus Hürden in den Lernprozessen vorher resultieren, im Aufbau von Konzepten und prozessbezogenen Kompetenzen. Zur Erläuterung, inwiefern diese auch sprachlich bedingt sein können, wird auf die kognitive Funktion von Sprache in den Lernprozessen der Jahre vor der Prüfung rekurriert.

2. Theoretische und empirische Hintergründe:

Relevanz sozialer und sprachlicher Faktoren für Mathematikleistung

2.1 Hintergrundfaktoren zur Erfassung von Benachteiligung

Zahlreiche empirische Studien zeigen, dass sozial oder sprachlich benachteiligte Jugendliche in Deutschland erheblich geringere Fachleistungen erreichen als ihre Klassenkameraden; dies gilt insbesondere für das Fach Mathematik (Baumert u. Schümer 2001). Dabei werden in den einzelnen Studien unterschiedliche soziale und sprachliche Hintergrundfaktoren herangezogen, mit denen die Benachteiligung erfasst und ein Zusammenhang zur Mathematikleistung hergestellt wird:

- Nationalität., d. h. Staatsangehörigkeit (z. B. Autorengruppe Bildungsberichterstattung 2012; Mikrozensus 2011),
- Migrationshintergrund als Auskunft über Geburtsländer der Betroffenen und ihrer Eltern (z. B. OECD 2007; Autorengruppe Bildungsberichterstattung 2012; Tarelli et al. 2012; u. v. a.),
- Mehrsprachigkeit bzw. Nichtpassung von Familien- und Unterrichtssprache (z. B. OECD 2007, S. 120; Heinze et al. 2007; Burns u. Shadoian-Gersing 2010; Ufer et al. 2013; Haag et al. 2013),
- sozioökonomischer Status (z. B. Bos et al. 2003; Ehmke et al. 2004; Werning et al. 2008), sowie
- Lesekompetenz (Rindermann 2006; Leutner et al. 2004, S. 167ff.; Knoche u. Lind 2004, S. 206; Bos et al. 2012, S. 237ff.), die auch als möglicher Mediator für die anderen Hintergrundmerkmale diskutiert wird (Walzebug 2014).

Während die meisten deutschen Untersuchungen auf Herkunftsfaktoren oder Lesekompetenz fokussieren, lassen Ergebnisse empirischer Studien aus anderen Ländern vermuten, dass eine breiter gefasste *Sprachkompetenz* ein bedeutender Einflussfaktor für die Mathe-

matikleistung sein könnte (Pimm 1987; Secada 1992, S. 638; Abedi 2006). Um dem genauer nachgehen zu können, muss zunächst die Beziehung zwischen Sprache und Mathematiklernen genauer gefasst werden.

2.2 Sprachkompetenz und sprachlich bedingte Hürden

Die Untersuchung der zentralen Rolle von Sprache im Mathematikunterricht hat in der Mathematikdidaktik eine lange Tradition (Pimm 1987; Ellerton u. Clarkson 1996; Maier u. Schweiger 1999). Sie gewinnt mit Blick auf herkunftsbedingte Leistungsdisparitäten jedoch eine neue Dimension, die im Folgenden knapp skizziert werden soll (vgl. Prediger u. Özdil 2011 für einen Überblick).

Der *Zusammenhang zwischen Herkunft und Sprachkompetenz* wird vor allem durch die Unterscheidung von Alltagssprache und Bildungssprache bzw. von BICS (basic interpersonal communication skills) und CALP (cognitive academic language proficiency, vgl. Cummins 2000) theoretisch erfasst: Viele Kinder und Jugendliche mit Migrationshintergrund oder bildungsfernen Elternhäusern erwerben zwar zuweilen solide alltagssprachliche Kompetenzen im Deutschen, aber oft in zu geringem Umfang bildungssprachliche Kompetenzen, auf die es im schulischen Kontext ankommt (Schleppegrell 2004; Gogolin 2009; Morek u. Heller 2012). Das bildungssprachliche Register ist gegenüber dem alltagssprachlichen gekennzeichnet durch spezifische lexikalische Anforderungen (wie Präfixverben, Komposita, normierte Fachbegriffe und hohe lexikalische Dichte) sowie zahlreiche grammatikalisch-syntaktische und diskursive Merkmale, die hohe Verdichtungen und Dekontextualisierungen ermöglichen (ebd.). Dementsprechende Kompetenzen werden im Alltag nicht automatisch erworben.

Auch wenn das Konstrukt Bildungssprache noch weiterer Ausdifferenzierungen bedarf, die derzeit Gegenstand intensiver linguistischer, psychologischer und erziehungswissenschaftlicher Forschungen sind (aktueller Überblick z. B. Redder u. Weinert 2013), zeigen bereits einige Untersuchungen, wie sich Einschränkungen in der bildungssprachlichen Kompetenz auf das Mathematiklernen und die Mathematikleistung auswirken können (Kaiser u. Schwarz 2003; Heinze et al. 2007; Gellert 2011; Rösch u. Paetsch 2011; Ufer et al. 2013; Prediger 2013; Prediger u. Wessel 2013). Diese Studien bilden für die vorliegende Untersuchung einen wichtigen Hintergrund, da sie zeigen, dass Konzeptualisierungen von Sprachkompetenz jenseits der Lesekompetenz notwendig sind. Leitend ist dabei der Gedanke der *Dualität von kommunikativer und kognitiver Funktion von Sprache* (Maier u. Schweiger 1999, S. 18; Morek u. Heller 2012).

Der Zusammenhang von Sprachkompetenz und Mathematikleistung wurde für Tests bisher zumeist im Sinne der *kommunikativen Funktion von Sprache* diskutiert: Einige Lernende haben größere Schwierigkeiten, Test-Items sinnkonstituierend zu lesen und können daher ihre mathematischen Kompetenzen nicht zeigen. Sprachlich bedingte Hürden bei der Test-Bearbeitung in diesem Sinne sind *Lesehürden*, die gerade in der amerikanischen Diskussion (Abedi 2006; Brown 2005; Wolf u. Leon 2009; Martiniello 2009; u.v.a.) vor allem als sprachliche Biases betrachtet werden, also als Hinweise auf mangelnde Validität der Tests, die dann nicht mehr mathematische, sondern Lesekompetenzen prüfen. In dieser

Perspektive ist es sinnvoll, sprachliche Kompetenz als *Lesekompetenz* zu konzeptualisieren und Lesehürden mit dem Ziel zu identifizieren, sie für die Herstellung eines fairen und nur auf mathematische Kompetenzen fokussierten Tests zu eliminieren (Abedi 2006).

Sprache besitzt jedoch neben der kommunikativen Funktion auch eine *kognitive Funktion*, d. h. sie ist gleichzeitig ein kognitives Werkzeug in Lern- und Denkprozessen (Maier u. Schweiger 1999, S. 18). Um die kognitive Funktion von Sprache theoretisch zu erklären, wurden linguistische Modelle entwickelt, die die Beziehungen zwischen Denken und Sprechen unterschiedlich konzeptualisieren (vgl. Morek u. Heller 2012 für einen Überblick zur dort epistemisch genannten Funktion). Dabei wird ihre kognitive Funktion insbesondere für das bildungssprachliche Register immer wieder betont, weil gerade die spezifischen sprachlichen Möglichkeiten der Verdichtung und Dekontextualisierung komplexere kognitive Prozesse erlauben (Halliday 1993; Schleppegrell 2004; Thürmann, Vollmer u. Pieper 2010; Morek u. Heller 2012).

In Bezug auf die kognitive Funktion von Sprache können sprachlich bedingte Hürden in Mathematiktests daher für sprachlich schwache Lernende auch auftauchen, wenn die Aufgabenstellung der Items zwar in den Anforderungen verstanden ist, die Anforderungen aber dennoch kognitiv nicht bewältigt werden können. Sprachlich bedingte Hürden sind in dieser Perspektive somit nicht nur Lesehürden, sondern auch andere Hürden, an denen *sprachlich schwache* Lernende im Prozess der Aufgabenbearbeitung gehäuft scheitern. Sie sind zum Teil eher auf längerfristige, sprachlich bedingte Einschränkungen in den Lern- und Denkprozessen vor dem Test zurückzuführen, wie gezeigt werden kann.

Einen Hinweis auf die längsschnittliche, längerfristige Relevanz der kognitiven Funktion von Sprache geben Heinze et al. (2007) in einer Studie zur Entwicklung der Mathematikleistung von Klasse 1 nach Klasse 2. Darin korrelieren der sozioökonomische Status und die Mehrsprachigkeit zwar im querschnittlichen Vergleich am stärksten mit der Mathematikleistung. Für die längsschnittliche Leistungsentwicklung bilden dagegen (neben den kognitiven Grundfertigkeiten) die sprachlichen Kompetenzen (hier erfasst als Hörverständnis und Verfügbarkeit von Wortschatz) den wichtigsten Prädiktor; dies gilt vor allem für Aufgaben zum konzeptuellen Verständnis (Ufer et al. 2013). Die Autoren verstehen dies als Indiz dafür, dass sich „sprachliche Defizite kumulativ negativ auf die Lernzuwächse in den Sachfächern [wie Mathematik] aus[wirken]“ können (Herwartz-Emden 2003, S. 692).

Da für die kognitive Funktion von Sprache nicht nur Lesekompetenz relevant ist (Duarte et al. 2011, S. 39), kann insgesamt festgehalten werden, dass Sprachkompetenz in dieser Perspektive breiter gefasst werden und sowohl lexikalisch-semantische (d.h. auf den Wortschatz und seine Bedeutungen bezogene) als auch grammatikalische Qualifikationen in Sprachrezeption und *-produktion* mit einbeziehen muss, die eng miteinander verknüpft sind.

Wie diese kombiniert erfasst werden können, wurde in der Linguistik und Spracherwerbsforschung seit 30 Jahren theoretisch begründet konzeptualisiert und mit dem für diese Zwecke vielfach bewährten Testinstrument der C-Tests operationalisiert (Grotjahn 1992). Die lexikalisch-semantischen und grammatikalischen Charakteristika der Bildungssprache lassen sich dabei berücksichtigen, wenn den C-Tests bildungssprachliche Texte

zugrunde gelegt werden (Daller 1999). Zwar kann eine solche Operationalisierung von bildungssprachlicher Kompetenz die Breite und Tiefe des Konstrukts nur teilweise erfassen (für die genaueren Hintergründe dieses Konstrukts muss auf die umfangreiche Diskussion in Linguistik und Spracherwerbsforschung verwiesen werden, vgl. Redder u. Weinert 2013). Dennoch ermöglicht sie in linguistisch akzeptierter Weise die Effekte der im Lernprozess wirksamen tatsächlichen Sprachkompetenz zu erfassen. Auf die genannten Vorarbeiten zur Operationalisierung bildungssprachlicher Kompetenz wird daher hier in der Instrumentenwahl zurückgegriffen. Die konkrete Identifizierung sprachlich bedingter Hürden jenseits der Lesehürden erfolgt dann datengeleitet aus den Itemanalysen und Analysen von Bearbeitungsprozessen.

2.3 Sprachlich bedingte Hürden in literacy-basierten Tests

In Anbetracht der existierenden empirischen Befunde müsste sich der Zusammenhang von Lese- bzw. Sprachkompetenz und Mathematikleistung gerade in solchen Tests zeigen, die nicht nur kontextfrei Rechenfertigkeiten prüfen, sondern im Sinne einer *mathematical literacy* nach PISA (OECD 2007; Neubrand 2001) den verständigen und flexiblen Umgang mit Mathematik, indem das Finden von Lösungsansätzen in außer- oder innermathematischen Kontexten sowie konzeptuelles Verständnis betont wird. Zu diesen sogenannten literacy-basierten Tests zählen Büchter und Pallack (2012, S. 63) auch die Zentralen Prüfungen 10 aus Nordrhein-Westfalen, auf die die hier dokumentierte Untersuchung fokussiert. Die Existenz spezifischer Hürden für sprachlich schwache Lernende in Prüfungen mit dieser kontext- und damit insbesondere textbezogenen Ausrichtung wurde für amerikanische literacy-basierte Tests nachgewiesen (Brown 2005) und wird nun für diese Prüfungen untersucht.

Gerade im Hinblick auf *mathematical literacy* haben Kaiser und Schwarz (2003) gefordert, selbst Lesehürden nicht allein als Gefährdung von Validität zu betrachten, weil das sinnentnehmende Lesen für den verständigen Umgang mit Mathematik in Texten zentral ist. Daher erscheint es (auch mit Blick auf die aktuellen curricularen Vorgaben) als lohnend, sowohl Lesehürden als auch andere Hürden für sprachlich schwache Lernende genauer zu identifizieren, um in Anschlussprojekten gezieltere Lerngelegenheiten zur Überwindung dieser Hürden auf dem Weg zum Erwerb von *mathematical literacy* zu ermöglichen.

2.4 Forschungsfragen

Ausgehend vom aktuellen Forschungsstand wird der Zusammenhang der Mathematikleistung zum Faktor Sprachkompetenz untersucht, und zwar im Vergleich zu anderen herkunftsbedingten Faktoren am Beispiel der nordrhein-westfälischen Zentralen Prüfungen ZP10-Mathematik auf dem Niveau des Mittleren Schulabschlusses, also am Ende der Vollzeitschulpflicht. Dabei sollen unterschiedliche Hürden identifiziert werden, die auf die rein kommunikative oder auch kognitive Funktion von Sprache zurückgeführt werden können. Verfolgt werden dabei folgende Forschungsfragen:

- F1. Welche sozialen und sprachlichen Hintergrundfaktoren haben den stärksten Zusammenhang zur Mathematikleistung in den Zentralen Prüfungen 10?
- F2. Bei welchen Items haben viele Lernende Schwierigkeiten, bei welchen besonders die sprachlich schwachen Lernenden?
- F3. Welche Hürden zeigen sich bei den für viele Lernende schweren Items und bei den für sprachlich Schwache besonders schweren Items?

Dabei ist die Forschungsfrage F2 eine Hilfsfrage, die die Bearbeitung der Forschungsfrage F3 ermöglicht.

3. Forschungsdesign und Methoden

In einem Mixed-Methods-Design wurden zunächst Leistungsdaten der ZP10-Mathematik 2012 mit Hintergrundvariablen in Beziehung gesetzt. Diese statistischen Analysen wurden für die theoretisch wie empirisch besonders ergiebige Hintergrundvariable *Sprachkompetenz* durch DIF-Analysen (zum Auffinden des Differential Item Functioning) ergänzt und schließlich mit Analysen von schriftlichen Bearbeitungen und Beobachtungen aus videographierten Bearbeitungsprozessen trianguliert. Durch einen pragmatischen Methodeneinsatz konnten so innerhalb einer Untersuchung statistische Effekte auch inhaltlich tiefergehend erklärt werden.

3.1 Erhebungsinstrumente der quantitativen Untersuchung

Als *abhängige Variable* für die Untersuchung wurde *Mathematikleistung* im Sinne von institutionell festgelegtem Erfolg („*achievement*“) in einer auf Literacy orientierten Prüfung betrachtet. Die Auswertungsgrundlage für die *Mathematikleistung* stellen die Bewertungen der Lehrkräfte in den ZP10-Mathematik 2012 (Mittlerer Schulabschluss) dar. Das Besondere des Datensatzes liegt also darin, dass er unter normalen Feldbedingungen einer abschlussrelevanten Prüfung erhoben wurde. Der hohen extrinsischen Motivation der Schülerinnen und Schüler bei der Bearbeitung, die viele Lehrkräfte berichten, steht dabei eine fehlende empirische Kontrolle der Auswertungsobjektivität gegenüber. Die Reliabilitätsbetrachtungen können sich auf Ergebnisse aus anderen Untersuchungen der gleichen Prüfungsart stützen (Büchter u. Pallack 2012).

Die von den Lehrerinnen und Lehrern eingetragenen 1 bis 5 Punkte pro Item wurden in einem Bewertungsbogen erfasst, daraus ergeben sich für die Abschlussprüfung die Klausurnoten. Für die hier vorliegende Untersuchung wurde eine Teilaufgabe aufgrund unterschiedlicher Anforderungen in zwei Items getrennt und für jedes der 27 Items inhaltlich entschieden, ab welcher erreichten Punktzahl es als im wesentlichen richtig gelöst gilt (für das Dichotomisierungskriterium und dessen Legitimation vgl. ebd.). Als *Itemscore* einer Person wurde die Anzahl der im wesentlichen richtig gelösten Items erfasst.

Auf dieser Basis wurden die Leistungsdaten in einem eindimensionalen dichotomen Rasch-Modell skaliert und die in diesem Modell geschätzten Personenfähigkeiten (WLE: Weighted Likelihood Estimates) in den weiteren Analysen als metrisch skalierte Messwerte für die Mathematikleistung verwendet (vgl. Abschnitt 3.3).

Die unabhängigen Variablen wurden entlang der in Abschnitt 2.1 skizzierten Literaturlage ausgewählt: Als erstes Set unabhängiger Variablen (*soziale Hintergrundfaktoren*) wurden im Vorfeld der Prüfungen die Familienhintergründe in einem Erhebungsbogen über Selbstauskünfte der Lernenden erfasst. Neben Alter und Geschlecht umfasste der Bogen folgende Hintergrundfaktoren, die aus Gründen der Vergleichbarkeit jeweils dreistufig ordinal verwendet werden:

- *Migrationshintergrund*, operationalisiert über die Geburtsländer der Befragten und ihrer Eltern (wie allgemein üblich z. B. in PISA, vgl. OECD 2007); dreistufig ordinale Verwendung (Ausprägungen: 1. Generation – 2. Generation – 3. Generation/ohne Migrationshintergrund).
- *Sozioökonomischer Status (SES)*, operationalisiert durch den fünfstufigen, visualisierten Book-at-Home-Index (bei Paulus 2009 mit Retest-Reliabilität von $r = .80$; auch genutzt in TIMSS von Schnabel u. Schwippert 2000, S. 269); zusammengefasste dreistufig ordinale Verwendung (Ausprägungen: niedrig – mittel – hoch).
- *Zeitpunkt des Deutscherwerbs* als Operationalisierung von Familiensprachen und Spracherwerbstypen (De Houwer 2009); dreistufig ordinale Verwendung (Ausprägungen: nur Deutsch in der Familie – Deutsch neben weiterer Sprache vor dem Kindergarten in der Familie erworben – Deutsch erst ab dem Kindergarten oder noch später erworben).
- Nicht erhoben werden konnten (unter den Feldbedingungen der abschlussrelevanten High-Stakes-Prüfung mit ihren spezifischen institutionellen Rahmungen) die kognitiven Grundfertigkeiten. Dieser methodischen Einschränkung steht die hohe Authentizität und schulpolitische Relevanz genau dieser Prüfungen unter Echtbedingungen im einwohnerstärksten Bundesland gegenüber.

Für das zweite Set unabhängiger Variablen wurden in der Untersuchung *sprachliche Kompetenzen* in zweierlei Weise konzeptualisiert. Beide Tests messen zwar zusammenhängende, aber nicht identische Konstrukte, wobei pragmatisch jeweils ein metrisches Messniveau angenommen wird:

- *Lesekompetenz*, operationalisiert mit Hilfe der 14 Items zum Leseverstehen in den zur ZP10-Mathematik parallelen Prüfungsarbeiten im Fach Deutsch (Mittlerer Schulabschluss). Dieser Test weist allerdings eine schlechte interne Konsistenz auf ($\alpha = 0.54$ bei der Stichprobengröße $n = 1066$). Ein Grund hierfür ist, dass bei der Entwicklung zentraler Prüfungen die Berücksichtigung eines breiten Anforderungsspektrums Vorrang vor einem aus testtheoretischer Sicht homogenen Konstrukt hat. Für heuristische Betrachtungen wurden die Ergebnisse, die die Schülerinnen und Schüler bei diesem Test erzielt haben, dennoch zunächst verwendet, wobei die Anzahl richtig bearbeiteter Items als Testwert verwendet wurde. Die vertieften Analysen wurden dann – aus statistischen und inhaltlichen Gründen – mit dem deutlich reliableren C-Test durchgeführt.

- *Sprachkompetenz in Deutsch*, operationalisiert durch einen C-Test, der sich zur zeitökonomischen und standardisierten Erfassung eines komplexen Konstrukts von Sprachkompetenz ohne Reduktion auf einzelne sprachliche Teilfertigkeiten mit hinreichender Reliabilität bewährt hat (Grotjahn 1992). Der eingesetzte C-Test (Baur u. Spettmann 2010) bestand aus fünf bildungssprachlich anspruchsvollen Texten, wobei die Texte mit mathematisch relevanten Kontexten empirisch die gleichen Schwierigkeits- und Zusammenhangswerte erzielten wie die ohne mathematische Kontexte. Der C-Test zeigte über die fünf Texte eine gute interne Konsistenz ($\alpha = 0.86$ bei der Stichprobengröße $n = 698$). Als Testwert für vertiefende Regressions- und Kovarianzanalysen wurde die Gesamtzahl richtig bearbeiteter Lücken (in fünf bearbeiteten Texten) verwendet. Auf Basis der C-Testwerte wurden zudem drei gleich große, sozialnormbezogene Gruppen (abkürzend „sprachlich schwaches / mittleres / starkes Drittel“) sowie für einfach interpretierbare DIF-Analysen zwei gleich große, sozialnormbezogene Gruppen (Mediansplit in abkürzend „sprachlich schwache / starke Hälfte“) gebildet.

3.2 Stichprobe

Die Basis für die empirische Untersuchung bildete eine Stichprobe von 1495 Schülerinnen und Schülern der zehnten Jahrgangsstufe in 67 Mathematik-Erweiterungskursen (Leistungsniveau mit dem Ziel mittlerer Schulabschluss) aus insgesamt 19 Gesamtschulen (vgl. Tab. 1).

Tab. 1 Überblick zur Stichprobe und zu den Teilstichproben

Merkmal	Gruppenbildung	Verteilung
Gesamt	19 Gesamtschulen, 67 Mathematik-Erweiterungskurse	N = 1495
Alter (n=1489)	17 Jahre und älter 16 Jahre 15 Jahre	311 (21 %) 984 (66 %) 194 (13 %)
Geschlecht (n=1487)	männlich weiblich	774 (52 %) 713 (48 %)
Migrations- hintergrund (n=1480)	1. Generation 2. Generation ab 3. Generation / ohne	152 (10 %) 623 (42 %) 705 (48 %)
Sozioökonomischer Status (n=1493)	niedriger SES („keine“ oder „sehr wenige“ Bücher) mittlerer SES („genug Bücher für ein Regalbrett“) hoher SES („drei Regale“ oder „ganze Regalwand voll“)	509 (34 %) 488 (33 %) 496 (33 %)
Zeitpunkt des Deutscherwerbs (n=1486)	mehrsprachig, Deutsch ab Kindergarten mehrsprachig, Deutsch vor Kindergarten einsprachig, nur Deutsch	289 (19 %) 538 (36 %) 659 (44 %)
Teilstichproben		
Test zur Lesekompetenz	nur Teilnehmende an Deutsch-Erweiterungskursen	n = 1066
C-Test zur Sprachkompetenz	unverzerrte Auswahl von Schulen	n = 698

Diese Stichprobe hat eine für den Ballungsraum Ruhrgebiet repräsentative Zusammensetzung der Schülerschaft von Gesamtschulen hinsichtlich sozialer und familiärer Hintergrundfaktoren und hinsichtlich der Leistungsergebnisse in Lernstandserhebungen und

Zentralen Prüfungen. Die Leistungsrepräsentativität in der nicht-gymnasialen Zielgruppe wurde von Empirikern des Ministeriums für Schule und Weiterbildung NRW anhand interner Daten überprüft und bestätigt; Gymnasien schreiben die Prüfung nicht mit.

Von den 1495 Lernenden der Gesamtstichprobe besuchten 1066 auch einen Deutsch-Erweiterungskurs; nur diese Teilstichprobe hat auch die ZP10-Deutsch auf dem Niveau des Mittleren Schulabschlusses absolviert, die verbleibenden 429 Lernenden in Deutsch-Grundkursen schrieben einen anderen Deutsch-Test auf dem Niveau des Hauptschulabschlusses nach Klasse 10. Da die Prüfungsaufgaben zum Leseverstehen für den Hauptschulabschluss unabhängig von den Aufgaben des Mittleren Schulabschlusses sind, können sie nicht auf einer gemeinsamen Skala betrachtet werden.

Daher wird die Lesekompetenz im Folgenden nur für die größere Teilstichprobe „Deutsch-Erweiterungskurs“ berücksichtigt; sie weist aufgrund ihrer Zusammensetzung etwas bessere Ergebnisse im Mathematiktest als die Gesamtstichprobe auf ($M = 11,5$; $SD = 4,5$ im Vergleich zu $M = 10,9$; $SD = 4,7$ in der Gesamtstichprobe, vgl. unten Tabelle 2). C-Test-Ergebnisse liegen von 698 Lernenden vor, da mit Blick auf den testbedingten Unterrichtsausfall kurz vor den Zentralen Prüfungen nur ein Teil der Schulen an dieser zusätzlichen Erhebung teilgenommen hat. Die Mathematikleistung dieser Teilstichprobe ($M = 11,1$; $SD = 4,7$, vgl. Tabelle 2) unterscheidet sich aber nicht signifikant von der Mathematikleistung der Gesamtstichprobe.

3.3 Modellierung der Mathematikleistung und statistische Auswertungsverfahren

Die Leistungsdaten zu den ZP10-Mathematik wurden im eindimensionalen dichotomen Rasch-Modell skaliert (Rost 2004, S. 115ff.). Diese Skalierung hat sich in Vorstudien als tragfähig für die Modellierung von Leistungsdaten aus zentralen Prüfungen erwiesen (Büchter und Pallack 2012). Durch die Skalierung werden die Leistungsdaten aller Lernenden und die Aufgabenschwierigkeiten auf einer gemeinsamen metrischen Skala erfasst. Die bei der Skalierung geschätzten *Personenfähigkeiten* (WLE) geben die in diesem Modell gemessenen Werte an (mit einer Normierung der Skala auf die mittlere Itemschwierigkeit 0). Für jede Aufgabe wird die *Itemschwierigkeit* (als Logit) angegeben, die im Rasch-Modell auf der gemeinsamen metrischen Skala mit der Personenfähigkeit gemessen wurde.

Die Kennwerte der Rasch-Skalierung belegen, dass das eindimensionale dichotome Rasch-Modell gut zu den dichotomisierten Prüfungsdaten passt. Die Item-Fits Weighted Mean Square (MNSQ: Mean square – gewichtete Abweichungsquadrate) liegen für alle Items zwischen 0.93 und 1.12, also ausnahmslos deutlich innerhalb des Intervalls [0.80; 1.20], das etwa im Rahmen der PISA-Studien als zulässiger Bereich für diese Kennwerte betrachtet wird (vgl. OECD 2009, S. 355). Die WLE-Reliabilität ist mit 0.79 zufriedenstellend. Die Itemschwierigkeiten liegen zwischen -2.50 für das leichteste und 2.70 für das schwerste Item (mit 0 als mittlerer Itemschwierigkeit).

Neben den aufbereiteten Leistungsdaten aus der Rasch-Skalierung werden im Folgenden auch die in der Prüfung erzielten Punktzahlen als Rohwerte berücksichtigt, da diese

wesentliche noten- und damit bildungsgangrelevante Kenngrößen darstellen, also ihre Bedeutung im Kontext der Zentralen Prüfungen direkt sichtbar ist.

Für die *Analysen zum Zusammenhang von Hintergrundfaktoren und Mathematikleistung (Forschungsfrage F1)* wurden Varianz-, Regressions- und Kovarianzanalysen durchgeführt. Dabei wurden die Rasch-skalierten Leistungsdaten (WLE) als abhängige Variable sowie die Faktoren zum Familienhintergrund und die Lese- bzw. Sprachkompetenz als unabhängige Variablen betrachtet:

- Zunächst wurden in getrennten Modellen isoliert die Effekte der einzelnen betrachteten unabhängigen Variablen bestimmt und die jeweiligen Anteile aufgeklärter Varianz (η^2 bzw. R^2) miteinander verglichen. Dabei wurden für die kategorial oder ordinal skalierten Hintergrundfaktoren einfaktorielle Varianzanalysen (ANOVA: Analysis of Variance) und für die Lese- bzw. Sprachkompetenz lineare Regressionen verwendet.
- Anschließend wurde mit den statistisch und inhaltlich tragfähigen unabhängigen Variablen (*SES*, *Zeitpunkt des Deutscherwerbs* und *Sprachkompetenz*, s. o.) eine Kovarianzanalyse (ANCOVA: Analysis of Covariance) durchgeführt.

Für die *zweite Forschungsfrage (F2) nach den absoluten und relativen Schwierigkeiten* wurde mit DIF-Analysen untersucht, welche Items der ZP10-Mathematik im Rahmen des eindimensionalen dichotomen Rasch-Modells „statistisch unerwartet schwierig“ für die Hälfte der sprachlich schwachen Lernenden (gemäß C-Test) waren. Für diese Gruppe werden im Rahmen der DIF-Analyse die nach der Rasch-Skalierung erwarteten Lösungshäufigkeiten für die einzelnen Items bestimmt, indem die mittleren Personenfähigkeitswerte der jeweiligen Gruppe und die Schwierigkeit der jeweiligen Items zueinander in Beziehung gesetzt werden. Die so ermittelten, theoretisch zu erwartenden Schwierigkeitswerte wurden dann mit den beobachteten Schwierigkeitswerten der Gruppe verglichen, es entsteht der sogenannte DIF-Wert (ähnlich bei Abedi 2006; Haag et al. 2013).

3.4 Methoden für die Analysen der individuellen Bearbeitungen

Um Aufschlüsse zur *Spezifizierung konkreter sprachlich bedingter Hürden* zu erlangen, die einige Items schwer bzw. besonders schwer für sprachlich schwache Lernende machen (*Forschungsfrage F3*), wurden die als relativ schwer identifizierten Items in einem bzgl. Leistung und Hintergrundfaktoren repräsentativen Subsample von 195 schriftlichen Test-Bearbeitungen genauer analysiert. Die schriftlichen Bearbeitungen wurden dazu hinsichtlich der Bewältigung zentraler Bearbeitungsschritte des jeweiligen Items codiert, dazu wurden nach Sichtung der Bearbeitungen jeweils ein aufgabenspezifisches Codiersystem entwickelt, in dem die zentralen Bearbeitungsschritte erfasst wurde. Zur Kontrastierung der sprachlich schwachen und starken Hälfte wurde die unterschiedliche Hürdenbewältigung in Drop-Out-Raten quantifiziert. Aus den Analysen der Drop-Out-Raten bei den Bearbeitungsschritten ließen sich die Hürden für unterschiedliche Sprachkompetenzgruppen genauer bestimmen.

Darüber hinaus wurden 47 Bearbeitungsprozesse von je ein bis zwei Lernenden im klinischen Interviewsetting videographiert und systematisch qualitativ bzgl. der auftauchenden Hürden (47 x 30-45 min. Videomaterial) analysiert. Die Ergebnisse dieser qualitativen Analysen werden andernorts ausführlicher dargestellt (Prediger et al. 2013; Gürsoy et al. 2013; Wilhelm 2015). Hier werden ausgewählte Ergebnisse nur argumentativ stützend und inhaltlich erklärend genutzt.

Durch einen Vergleich der verschiedenen Hürden in den schriftlichen Bearbeitungen und den videographierten Bearbeitungsprozessen wurden vier Kategorien von Hürden gebildet:

- (1) als *Lesehürden* werden Hürden kategorisiert, die rein in der Erfassung des Aufgabentextes liegen,
- (2) als *prozessuale Hürden* solche gefasst, die kognitiv anspruchsvolle Prozessschritte umfassen,
- (3) *konzeptuelle Hürden* werden in Bearbeitungsschritten identifiziert, die konzeptuelles Verständnis verlangen (z. B. Grundvorstellungen zu mathematischen Konzepten), und
- (4) als *rechnerische Hürden* schließlich solche, die nur in den Prozessen der innermathematischen Durchführung verortet sind.

Die prozessualen und rechnerischen Hürden tauchen in späteren Prozessschritten auf als das reine Lesen, konzeptuelle Hürden dagegen sind oft mit Lesehürden eng verbunden, weil ein Vorverständnis der mathematischen Struktur notwendig ist, um gezielt den Text zu erschließen. Rechnerische Hürden traten zwar auf, waren für sprachlich schwache Lernende aber nicht von größerer Bedeutung als für sprachlich Starke, sie werden daher hier nicht näher betrachtet.

4. Ergebnisse der Analysen

4.1 Zusammenhänge zwischen Hintergrundfaktoren und Mathematikleistung

Gruppenunterschiede in der Mathematikleistung werden in Tabelle 2 gezeigt, wobei neben den Gruppen, die sich aus verschiedenen Hintergrundfaktoren ergeben, auch die unterschiedlichen aus den oben genannten Gründen gebildeten Teilstichproben dargestellt werden. Als Mathematikleistung werden einerseits die mittleren mit dem Rasch-Modell erfassten Personenfähigkeitswerte (durchschnittliche WLE) der jeweiligen Gruppen angegeben, andererseits die notenrelevanten Gesamtpunktzahlen. Aus Gründen der Vergleichbarkeit wurden, sofern dies sinnvoll möglich war, jeweils drei Gruppen gebildet. Dabei lassen sich im großen mittleren Leistungsbereich WLE-Unterschiede von 0,2 in etwa als ein richtig gelöstes Item mehr bzw. weniger interpretieren. Bei den Punktzahlen entsprechen 11-12 Punkte einem Unterschied von einer Notenstufe; die Note „ausreichend“ wurde bei 38-49 Punkten vergeben. Die erste Spalte der Tabelle führt die untersuchten Faktoren auf. Für jeden Faktor wurden die in der zweiten Spalte aufgeführten Gruppen gebildet, deren Verteilung in der dritten Spalte angegeben wird.

Bzgl. der Mathematikleistung zeigen sich für jeden Faktor hoch signifikante Differenzen zwischen der starken und der schwachen Gruppe (jeweils in einem Scheffé-Test mit $p < 0.001$ als Post-hoc-Test nach einer einfaktoriellem ANOVA mit signifikantem F-Test).

Tab. 2 Gruppenunterschiede bzgl. verschiedener Hintergrundfaktoren und Zusammenhang zur Mathematikleistung (jeweils 3 Gruppen)

Hintergrundfaktoren	Kategorienbildung	Verteilung bei allen Lernenden (wie in Tabelle 1)	Mittlere M-Test-Gesamtpunkte (max. 85) m (SD)	Mittlere Leistungsdaten (WLE), m (SD)
Gesamt	19 Gesamtschulen	N=1495	43,5 (13,6)	-0,62 (1,07)
Geschlecht (n=1487)	männlich	774 (52,1 %)	45,3 (14,0)	-0,44 (1,09)
	weiblich	713 (47,9 %)	41,3 (12,8)	-0,81 (1,02)
Migrationshintergrund (n=1480)	1. Generation	152 (10,3 %)	41,3 (13,6)	-0,77 (1,13)
	2. Generation	623 (42,1 %)	40,9 (13,5)	-0,81 (1,06)
	ab 3. Generation	705 (47,6 %)	46,2 (13,0)	-0,41 (1,03)
Sozioökonomischer Status (n=1493)	niedriger SES	509 (34,1 %)	41,9 (14,0)	-0,74 (1,13)
	mittlerer SES	488 (32,7 %)	42,9 (12,9)	-0,67 (1,04)
	hoher SES	496 (33,2 %)	45,7 (13,4)	-0,43 (1,01)
Zeitpunkt	ab Kindergarten	289 (19,4 %)	39,5 (13,7)	-0,91 (1,09)
Deutscherwerb (n=1486)	vor Kindergarten	538 (36,2 %)	42,2 (13,5)	-0,71 (1,07)
	nur Deutsch	659 (44,3 %)	46,3 (13,0)	-0,40 (1,02)
Sprachliche Kompetenz (C-Test, n=698)	schwach	235 (33,7 %)	37,3 (13,4)	-1,04 (1,08)
	mittel	233 (33,4 %)	44,2 (12,6)	-0,53 (1,01)
	stark	230 (33,0 %)	50,3 (11,4)	-0,11 (0,94)
	<i>gesamt</i>	<i>698 (100 %)</i>	<i>43,9 (13,6)</i>	<i>-0,56 (1,08)</i>
Lesekompetenz (D-E-Kurse, n=1066)	schwach	365 (34,2 %)	40,3 (12,9)	-0,85 (1,02)
	mittel	405 (38,0 %)	46,6 (12,6)	-0,39 (0,95)
	stark	296 (27,8 %)	50,0 (12,5)	-0,14 (0,99)
	<i>gesamt</i>	<i>1066 (100 %)</i>	<i>45,4 (13,3)</i>	<i>-0,48 (1,02)</i>

Diese Gruppendifferenzen sind jedoch unterschiedlich hoch, wie die vierte und fünfte Spalte der Tabelle 2 zeigen: Die Unterschiede bzgl. der sozialen Faktoren betragen jeweils weniger als 7 Punkte, also etwa eine halbe Note Differenz in der abschlussrelevanten Klausurnote, bzgl. Geschlecht 4 Punkte, Migrationshintergrund 5,3 Punkte, SES 3,8 Punkte und Zeitpunkt des Deutscherwerbs 6,8 Punkte. Im Gegensatz dazu weisen die bzgl. der sprachlichen Faktoren gebildeten Gruppen höhere Differenzen auf: In der (etwas homogeneren) Teilstichprobe der Deutsch-E-Kurs-Schülerinnen und -Schüler variiert die Mathematikleistung der Lesekompetenzdrittel um 9,7 Punkte. In der Teilstichprobe der C-Test-Absolventen beträgt die Differenz zwischen schwachem und starkem Drittel 13 Punkte. Die durchschnittliche Punktzahl des sprachlich schwachen Drittels (im Sinne des Konstrukts *Sprachkompetenz* gemessen durch den eingesetzten C-Test) liegt genau auf der Notengrenze zwischen „mangelhaft“ und „ausreichend“, während die durchschnittliche Punktzahl des sprachlich starken Drittels gerade noch für ein „befriedigend“ reicht. Damit liegt bzgl. der Sprachkompetenz ein Gruppenunterschied von mehr als einer ganzen Notestufe vor.

Statistisch systematischer werden die unterschiedlichen Effekte der sozialen und sprachlichen Faktoren durch die Varianz- und Regressionsanalysen anhand der Aufklärung

der beobachteten Varianz der Mathematikleistung gezeigt (vgl. Tabelle 3): Während die sozialen Faktoren jeweils Varianzteile zwischen 1 % und 3 % aufklären, erweisen sich die sprachlichen Faktoren mit 10 % bzw. 14 % aufgeklärter Varianz als deutlich erklärungs-mächtiger.

Tab. 3 Varianzaufklärung (*Mathematikleistung*) bzgl. verschiedener Hintergrundvariablen

Hintergrundfaktor	Verfahren	Varianzaufklärung	df	F	Signifikanz
Sozioökonomischer Status (n = 1493)	ANOVA	$\eta^2 = 0.02$	2	11.44	p < 0.01
Migrationshintergrund (n = 1480)	ANOVA	$\eta^2 = 0.03$	2	25.64	p < 0.01
Zeitpunkt des Deutscherwerbs (n = 1486)	ANOVA	$\eta^2 = 0.03$	2	26.24	p < 0.01
Lesekompetenz* (n = 1066)	lineare Regression	$R^2 = 0.10$	1	111.08	p < 0.01
Sprachkompetenz (C-Test) (n = 698)	lineare Regression	$R^2 = 0.14$	1	124.44	p < 0.01

* Dieser Test weist mit $\alpha = 0.54$ eine schlechte interne Konsistenz auf (s. o.); er wurde in die Tabelle aufgenommen, um heuristisch vergleichende Betrachtungen durchführen zu können.

In der isolierten Betrachtung der unabhängigen Variablen zeigt sich im Bereich der sprachlichen Faktoren, dass die Sprachkompetenz einen größeren Varianzanteil aufklärt als die (ohnehin kaum reliabel gemessene) Lesekompetenz und somit potenziell erklärungs-mächtiger ist. Dies ist auch inhaltlich plausibel, da das Konstrukt *Sprachkompetenz* über die Lesekompetenz hinausgehende sprachliche Fähigkeiten berücksichtigt, die geeignet sind, Unterschiede in der Mathematikleistung zu erklären (vgl. Abschnitt 2.2). Daher werden die vertiefenden Analysen mit dem Faktor *Sprachkompetenz* durchgeführt.

Bei den sozialen Hintergrundfaktoren zeigt sich, dass der *Migrationshintergrund* und der *Zeitpunkt des Deutscherwerbs* etwa gleich große Varianzanteile aufklären. Beide Fak-toren hängen eng zusammen, wobei für den *Zeitpunkt des Deutscherwerbs* eine nahe-liegende inhaltliche Deutung des Zusammenhangs zur Mathematikleistung vorliegt. Daher wurde in der Kovarianzanalyse (vgl. Tabelle 4), die den Einfluss der sozialen Hintergrund-faktoren auf die *Mathematikleistung* unter Berücksichtigung der Kovariate *Sprachkompe-tenz* bestimmt, von diesen beiden zusammenhängenden Faktoren nur der *Zeitpunkt des Deutscherwerbs* berücksichtigt.

Die Kovarianzanalyse zeigt, dass der *Sozioökonomische Status* bei Kontrolle der *Sprachkompetenz* keinen eigenständigen signifikanten Beitrag zur Aufklärung von Varianz in der Mathematikleistung leistet ($F(2, 682) = 1.38$; $p = 0.25$). Dies kann mit Blick auf entsprechende soziolinguistische Befunde so gedeutet werden, dass der Einfluss des *Sozio-ökonomischen Status* vor allem über Unterschiede in der *Sprachkompetenz* wirkt. Der *Zeitpunkt des Deutscherwerbs* übt hingegen auch bei Kontrolle der Sprachkompetenz ei-nen signifikanten Einfluss auf die Mathematikleistung aus ($F(2, 682) = 9.29$, $p < 0.01$). Eine mögliche (wenn auch empirisch weiter zu fundierende) Deutung für diesen Befund führt auf die allgemeine Bedeutung der kognitiven Funktion von Sprache für das Mathe-matiklernen zurück: Wenn verschiedene Schülerinnen und Schüler zum Testzeitpunkt (10. Schuljahr) zwar die gleiche Sprachkompetenz haben, mit dem Erwerb der Unterrichts-

sprache aber zu unterschiedlichen Zeitpunkten ihrer Bildungsbiographie begonnen haben, dann könnte es sein, dass die geringere Sprachkompetenz in weiter zurückliegenden Schuljahren dazu geführt hat, dass bestimmte mathematische Kompetenzen nur partiell erworben werden konnten.

Tab. 4 Kovarianzanalyse (*Mathematikleistung* mit Kovariate *Sprachkompetenz*; $n = 692$)

Quelle der Varianz	Quadratsumme	df	F	Signifikanz	partielles η^2
Korrigiertes Modell	133.79	9	15.02	$p < 0.01$	0.17
Sprachkompetenz (C-Test)	49.02	1	50.12	$p < 0.01$	0.07
Sozioökonomischer Status (SES)	2.70	2	1.38	$p = 0.25$	0.00
Zeitpunkt des Deutscherwerbs	18.16	2	9.29	$p < 0.01$	0.03
SES * Zeitpunkt des Deutscherwerbs	0.78	4	0.20	$p = 0.94$	0.00
Fehler	666.98	682			
Korrigierte Gesamtvariation	800.77	691			

Insgesamt ergibt sich als Ergebnis dieser ersten statistischen Analysen, dass die Sprachkompetenz unter allen betrachteten sozialen und sprachlichen Faktoren den größten Beitrag zur statistischen Erklärung von Unterschieden in der Mathematikleistung hat. Da sich dieser Befund stimmig in den oben diskutierten internationalen Forschungsstand einfügt (vgl. Abschnitt 2.2), wurde die im C-Test operationalisierte Sprachkompetenz daher für alle weiteren Analysen ins Zentrum gerückt.

4.2 Itemanalysen zu relativen und absoluten Schwierigkeiten

Zur Bearbeitung der Forschungsfrage F2 sind in der folgenden Tabelle 5 die relativen und absoluten Schwierigkeiten der Items in je zwei Darstellungen abgedruckt: Die Lösungshäufigkeiten in der Gesamtstichprobe lassen erkennen, welche Items für alle Lernenden schwer sind; die Lösungshäufigkeiten in dem sprachlich schwachen und sprachlich starken Drittel machen Gruppenunterschiede sichtbar und können in Bezug zu den für diese Gruppen berichteten Gesamtestwerten in Tabelle 2 gesetzt werden. Daneben sind die Itemschwierigkeiten auf der Rasch-Skala angegeben sowie die DIF-Werte als Maß für die relative Schwierigkeit für die sprachlich Schwachen, die unten gesondert erläutert werden. Mit Blick auf eine möglichst gute Interpretierbarkeit wurden bei der DIF-Analyse nur zwei gleichgroße Gruppen gebildet, die sprachlich schwache und die sprachlich starke Hälfte, da die DIF-Werte dann in beiden Gruppen den gleichen Betrag haben und sich genau durch das Vorzeichen, also die Richtung der Schwierigkeitsverschiebung, unterscheiden.

Als grundsätzliche Tendenz zeigt sich erwartungsgemäß, dass das sprachlich schwache Drittel für die meisten Items erkennbar niedrigere Lösungshäufigkeiten erreicht als das sprachlich starke Drittel. Mithilfe der DIF-Analyse können nun diejenigen Items identifiziert werden, die darüber hinaus auch „statistisch unerwartet schwierig“ für die sprachlich schwache Hälfte der Stichprobe sind (im Vergleich zur Itemschwierigkeit, die aufgrund des Gesamtergebnisses der Gruppe zu erwarten gewesen wären, abkürzend „relativ schwieriger“). In der letzten Spalte von Tabelle 5 werden für jedes Item die DIF-Werte, d. h. die gruppenspezifischen Verschiebungen der Itemschwierigkeiten (beobachtete Itemschwierigkeit im Vergleich zur erwarteten Itemschwierigkeit) für die sprachlich

schwache Hälfte angegeben. Zum Beispiel lässt sich für Item 1a ablesen, dass das Item für die sprachlich schwache Hälfte um 0,177 Einheiten auf der Skala, auf der die Aufgabenschwierigkeiten und die Personenfähigkeiten metrisch erfasst werden, schwieriger wäre als für die Gesamtstichprobe (und für die sprachlich starke Hälfte um 0,177 Einheiten leichter).

Tab. 5 Überblick über absolute und relative Schwierigkeit der Items

Items	Kurzcharakterisierung des Inhalts	Lösungshäufigkeiten			Item-schwierigkeit (im Rasch-Modell)	DIF-Wert (sprachlich schwache Hälfte)
		in der Gesamtstichprobe	in sprachlich schwacher Hälfte	in sprachlich starker Hälfte		
1 Basiskompetenzen						
1a	Münzturm schätzen (Fermi)	31 %	23 %	44 %	0,311	0,177*
1b	Bakterienwachstum (Term aufstellen)	46 %	38 %	49 %	-0,424	-0,087
1c	Partyhut (Kegelberechnung)	24 %	23 %	34 %	0,720	-0,023
1d11	Tabellen (Berechnung der Seitenlänge)	74 %	72 %	75 %	-1,841	-0,284*
1d12	Tabellen (Berechnung des Umfangs)	84 %	85 %	83 %	-2,539	-0,444*
1d2	Tabellenkalkulation (Formel)	12 %	8 %	13 %	1,716	-0,063
1e	Internetnutzung (Tabellen lesen)	69 %	64 %	73 %	-1,558	-0,138
2 Kraftstoffverbrauch (Quadratische Funktion)						
2a1	x-Wert zu gegebenen y-Wert ablesen	68 %	66 %	70 %	-1,508	-0,246*
2a2	Prozentualer Vergleich zweier Fkt-Werte	12 %	9 %	17 %	1,646	0,071
2b1	Verbrauchsrechnung Dreisatz / Division	68 %	62 %	78 %	-1,526	0,083
2b2	Literberechnung (Multiplikation)	33 %	28 %	43 %	0,207	0,034
2c1	y-Wertbestimmung mit quadr. Gl.	50 %	46 %	55 %	-0,612	-0,146
2c2	x-Wertbestimmung mit quadr. Gl.	11 %	10 %	13 %	1,804	-0,146
3 Krake Paul (Wahrscheinlichkeit)						
3a	Simulation	49 %	37 %	60 %	-0,558	0,209*
3b	Baumdiagramm zweistufig	59 %	56 %	76 %	-1,063	0,178*
3c	Wahrscheinlichkeit zweistufig	68 %	64 %	82 %	-1,486	0,158
3d	Komplementäreignis berechnen	17 %	14 %	25 %	1,276	0,069
3e1	Wahrscheinlichkeitsrechnung mehrstufig	53 %	45 %	64 %	-0,736	0,107
3e2	Werte in Koordinatensystem übertragen	37 %	29 %	45 %	0,030	0,079
3e3	Wieso Graph nicht durchzeichnen	5 %	4 %	11 %	2,701	0,261
3f	Formel für W(n)	9 %	5 %	11 %	1,997	0,180
4 Treppenstufen (Algebra)						
4a1	2,60 : 14 (Kontext eindenken)	39 %	36 %	52 %	-0,068	0,029
4a2	Steigung in % berechnen	23 %	20 %	27 %	0,841	-0,083
4a3	Steigungswinkel berechnen (Tangens)	34 %	32 %	41 %	0,161	-0,118
4b1	Faustformel überprüfen	52 %	46 %	55 %	-0,701	-0,132
4b2	Randwerte für Faustformel bestimmen	51 %	41 %	59 %	-0,685	0,074
4b3	Beispielwerte durch Probieren finden	10 %	10 %	10 %	1,897	0,204

Ein Minuszeichen vor dem Betrag der Verschiebung bedeutet, dass der Wert für die Itemschwierigkeit in dieser Gruppe sinkt, die Aufgabe dieser Gruppe also relativ leicht fällt gegenüber den durch den Gesamttest bestimmbar erwartungen. Mithilfe des nicht abgedruckten Standardfehlers wurden die Verschiebungen der Itemschwierigkeiten auf

statistische Signifikanz untersucht¹. Diejenigen Items, deren DIF-Werte auf dem 5 %-Niveau signifikant werden, sind in der letzten Spalte der Tabelle 5 mit einem Stern (*) markiert.

Als für sprachlich Schwache relativ leicht stellten sich somit drei Items heraus (1d11, 1d12 Tabellenkalkulation und 2a1 Kraftstoffverbrauch am Diagramm ablesen). Nur drei Items (1a Münzturm, 3a und 3b Krake Paul) erwiesen sich als statistisch signifikant relativ schwerer. Ein weiteres Item (2a2 Kraftstoffverbrauch prozentual vergleichen) zeigte zwar keinen signifikanten DIF-Wert, ist aber aus sprachlichen Gründen für alle Lernenden sehr schwer. Es ist daher für die Untersuchung sprachlich bedingter Hürden ebenfalls interessant.

<p>Item 1a: Münzturm (relativ schwerer für sprachlich Schwache) Schätze, wie viele Kilometer hoch ein Turm aus 2,4 Milliarden 1-Cent-Münzen ungefähr wäre. Beschreibe, wie du vorgegangen bist.</p>																																																															
<p>Items 1d12 und 1d12: Tabellenkalkulation (relativ leichter für sprachlich Schwache) Malak untersucht mithilfe einer Tabellenkalkulation Rechtecke mit dem Flächeninhalt 144 cm^2. Dabei berechnet er für unterschiedliche Längen der Seite a die zugehörige Länge der Seite b und den zugehörigen Umfang.</p>		<table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> </tr> </thead> <tbody> <tr> <td>1</td> <td colspan="3">Gegebener Flächeninhalt des Rechtecks: 144 cm^2</td> </tr> <tr> <td>2</td> <td></td> <td></td> <td></td> </tr> <tr> <td>3</td> <td></td> <td></td> <td></td> </tr> <tr> <td>4</td> <td>Seite a [in cm]</td> <td>Seite b [in cm]</td> <td>Umfang [in cm]</td> </tr> <tr> <td>5</td> <td>1,5</td> <td>96,0</td> <td>195,0</td> </tr> <tr> <td>6</td> <td>3,0</td> <td>48,0</td> <td>102,0</td> </tr> <tr> <td>7</td> <td>4,5</td> <td></td> <td>73,0</td> </tr> <tr> <td>8</td> <td>6,0</td> <td>24,0</td> <td>60,0</td> </tr> <tr> <td>9</td> <td>7,5</td> <td>19,2</td> <td>53,4</td> </tr> <tr> <td>10</td> <td>9,0</td> <td>16,0</td> <td></td> </tr> <tr> <td>11</td> <td>10,5</td> <td>13,7</td> <td>48,4</td> </tr> <tr> <td>12</td> <td>12,0</td> <td>12,0</td> <td>48,0</td> </tr> <tr> <td>13</td> <td>13,5</td> <td>10,7</td> <td>48,3</td> </tr> <tr> <td>14</td> <td>15,0</td> <td>9,6</td> <td>49,2</td> </tr> </tbody> </table>			A	B	C	1	Gegebener Flächeninhalt des Rechtecks: 144 cm^2			2				3				4	Seite a [in cm]	Seite b [in cm]	Umfang [in cm]	5	1,5	96,0	195,0	6	3,0	48,0	102,0	7	4,5		73,0	8	6,0	24,0	60,0	9	7,5	19,2	53,4	10	9,0	16,0		11	10,5	13,7	48,4	12	12,0	12,0	48,0	13	13,5	10,7	48,3	14	15,0	9,6	49,2
	A	B	C																																																												
1	Gegebener Flächeninhalt des Rechtecks: 144 cm^2																																																														
2																																																															
3																																																															
4	Seite a [in cm]	Seite b [in cm]	Umfang [in cm]																																																												
5	1,5	96,0	195,0																																																												
6	3,0	48,0	102,0																																																												
7	4,5		73,0																																																												
8	6,0	24,0	60,0																																																												
9	7,5	19,2	53,4																																																												
10	9,0	16,0																																																													
11	10,5	13,7	48,4																																																												
12	12,0	12,0	48,0																																																												
13	13,5	10,7	48,3																																																												
14	15,0	9,6	49,2																																																												
<p>(1) Berechne für die Zellen B7 und C10 die fehlenden Werte. (2) Gib für B5 und C5 jeweils eine Formel an.</p>																																																															
<p>Items 2a: Kraftstoffverbrauch (2a1 relativ leichter für sprachlich Schwache, 2a2 schwer für alle) Der Kraftstoffverbrauch wird für Fahrzeuge durch den durchschnittlichen Verbrauch in Litern (ℓ) auf einer Strecke von 100 Kilometern angegeben. Der Kraftstoffverbrauch eines Autos hängt vor allem von der gefahrenen Geschwindigkeit ab. Das Diagramm zeigt den Kraftstoffverbrauch für ein Auto, das im höchsten Gang gefahren wird. Daher beginnt der Graph bei 70 km/h.</p>																																																															
<p>(1) Wie schnell fährt das Auto durchschnittlich, wenn es 11 ℓ auf 100 km verbraucht? (2) Um wie viel Prozent liegt der Verbrauch bei 180 km/h über dem Verbrauch bei 100 km/h? Notiere deine Rechnung.</p>																																																															
<p>Items 3a und 3b: Krake Paul (relativ schwerer für sprachlich Schwache) Bei der Fußball-WM 2010 wurde der Krake Paul international berühmt. Vor jedem Fußballspiel [...] suchte [Paul] sich einen der beiden Futtertöpfe aus. Seine Wahl wurde dann von den Medien als „Vorhersage“ des Gewinners des Fußballspiels gedeutet. [...] Mathematisch betrachtet handelt es sich bei den „Vorhersagen“ also um einen Zufallsversuch mit zwei gleich wahrscheinlichen Ergebnissen.</p>																																																															
<p>a) Erkläre, wie man diesen Zufallsversuch mithilfe eines Würfels simulieren kann. b) Zeichne ein Baumdiagramm, das die Wahrscheinlichkeiten für zwei Vorhersagen angibt.</p>																																																															

Abb. 1 Auffällige Items der Zentralen Prüfungen 10 Mathematik NRW 2012

¹ Die Signifikanz des DIF-Wertes hängt von der Differenz auf der metrischen Fähigkeit-Schwierigkeit-Skala ab, aber auch von der Itemschwierigkeit: Bei sehr leichten oder sehr schwierigen Items sind größere Verschiebungen erforderlich, um signifikante Ergebnisse zu erzielen, als bei Items mittlerer Schwierigkeit.

Diese sieben auffälligen Items sind in Abb. 1 abgedruckt. Die relativ leichten Items (die für sprachlich schwache Lernende einen signifikanten negativen Wert hatten) werden hier kurz angesprochen, die relativ schwierigen im nächsten Abschnitt ausführlicher diskutiert.

Item 1d11 und *1d12* zielen auf einfaches Wissen zur Tabellenkalkulation, über das fast alle Lernende (unabhängig von der Sprachkompetenz) zu verfügen scheinen, die dazu im Unterricht überhaupt eine Lerngelegenheit hatten. Sie fallen den sprachlich Schwachen daher nicht schwerer als den sprachlich Starken. Für *Item 2a1* (Kraftstoffverbrauch am Diagramm ablesen) zeigte sich in den videographierten Bearbeitungsprozessen, dass viele Lernende mit einer oberflächlichen Standardbearbeitung den gesuchten Wert zum gegebenen Funktionswert aus dem Diagramm ablesen können, ohne den funktionalen Zusammenhang durchdringen zu haben. Dies kann eine Erklärungshypothese für die relative Leichtigkeit des Items für die sprachlich schwache Hälfte der Stichprobe geben, die hier signifikant besser als statistisch erwartbar abschneidet.

4.3 Rekonstruierte Hürden in den auffälligen Items

Die Analysen von 195 schriftlichen Bearbeitungen und 47 videographierten Bearbeitungsprozessen der als schwer bzw. relativ schwer identifizierten Items 1a, 2a2, 3a und 3b ermöglichte die Rekonstruktion von wiederkehrenden Hürden, die in vier Typen eingeteilt werden konnten (vgl. Abschnitt 3.4): Lesehürden, prozessuale Hürden, konzeptuelle Hürden und rechnerische Hürden. Da letztere nicht in besonderer Weise sprachlich schwache Lernende betrafen, werden sie nicht als sprachlich bedingte Hürden betrachtet. Die drei erstgenannten sollen daher als Ergebnis der Analysen jeweils exemplarisch konkretisiert werden.

Lesehürden auf Satzebene am Beispiel des Items 2a2 (Kraftstoffverbrauch)

Item 2a2 (Kraftstoffverbrauch prozentual vergleichen, abgedruckt in Abb. 1) war für alle Lernende eine schwierige Aufgabe (vgl. Tabelle 5), die von insgesamt 12 % der Lernenden gelöst wurde, von nur 9 % der sprachlich schwachen Hälfte.

Die Analysen der videographierten Bearbeitungsprozesse verweisen immer wieder auf Lesehürden auf Satzebene: Lesehürden erzeugen hier nicht einzelne unbekannte Worte, sondern die komplexen Satzstrukturen durch geschachtelte Präpositionalphrasen („Um wie viel Prozent liegt der Verbrauch *bei* 180 km/h *über* dem Verbrauch *bei* 100 km/h?“, Hervorhebung eingefügt), die mehrere Hauptsätze ersetzen und die komplexen Relationen somit in einer kurzen Frage verdichten.

Exemplarisch deutlich wird diese Schwierigkeit an dem Transkriptauszug von Berna, einer Sechzehnjährigen mit Familiensprache Türkisch, die laut C-Test-Ergebnissen zum sprachlich schwachen Drittel gehört:

5 B [liest die Aufgabe leise für sich durch, 14 sec] Also hier ist das ja, glaube ich, dann die Aufgabe, dass wir dann herausfinden sollen, wie viel Prozent, ähm, 180 von 100 Kilometern sind.

...

19 B [2 sec Pause] Also der- wir haben ja diesen Verbrauch, äh 100 km/h.

20 I Mhm.

- 21 B So. Und wenn dann jemand mit dem Auto fährt, dann fährt der 100- äh 180 km/h. Und wir sollen dann rausfinden, wie viel Prozent das über dem Normalverbrauch ist.
 ... [setzt in Prozentformel 180 und 100 ein]
 33 B [3 sec Pause] Ähm, der Verbrauch liegt bei 55 Prozent [2 sec Pause] über dem Verbrauch bei 100 km/h [lacht] Ähm. [4 sec Pause] Antwortsatz hätte ich jetzt nicht.

Berna vereinfacht die Frage

Um wie viel Prozent liegt der Verbrauch bei 180 km/h über dem Verbrauch bei 100 km/h?
 zu

Wie viel Prozent ist 180 über 100?

(in Zeile 21) und rechnet (in Zeile 33)

Wie viel Prozent ist 100 von 180?

Die verkürzte Präpositionalphrase „Verbrauch bei 100 km/h“ erkennt sie nicht als „Funktionswert der Verbrauchsfunktion für die Geschwindigkeit 100“, sondern identifiziert die 100 mit dem Verbrauch selbst (Zeile 19, 21). Den Satzbaustein „Um wie viel ... liegt über“ wiederholt sie zwar sprachlich (Zeile 21), berechnet aber einen einfachen Anteil (zwischen Zeile 21 und 33). Damit hat sie beide relevanten Relationen in dem Satz nicht adäquat identifiziert und kann im Mathematisierungsprozess nicht mehr tragfähig voranschreiten. Dass ihr schließlich die Interpretation des Gerechneten schwer fällt, formuliert sie selbst durch „Antwortsatz hätte ich jetzt nicht“ (Zeile 33). Sprachlich stärkere Lernende dagegen richten eher den Fokus auf die Relationen, was eine tragfähige Mathematisierung erst möglich macht (vgl. Gürsoy et al. 2013 und Wilhelm 2015 für weitere Analysen).

Um zu prüfen, inwieweit sich Bernas Schwierigkeiten zur Überwindung der Lesehürden auch in der Breite zeigen, wurden die 195 schriftlichen Bearbeitungen codiert bzgl. der Bewältigung der notwendigen Bearbeitungsschritte im Lösungsprozess. Die Kontrastierung der Drop-Out-Raten in den Bearbeitungsschritten ist wie folgt zu lesen: Als nicht bewältigt wurde ein Bearbeitungsschritt codiert, wenn er nicht tragfähig bearbeitet oder ausgelassen wurde. Dabei bezieht sich die Drop-Out-Rate als bedingte relative Häufigkeit bei jedem Schritt jeweils auf die Gesamtheit derjenigen Bearbeitungen, die im Lösungsprozess so weit gekommen sind, dass grundsätzlich eine erfolgreiche Bearbeitung dieses Schritts möglich gewesen wäre. Daher kann von der so definierten Drop-Out-Rate auf relevante Hürden zurück geschlossen werden.

So wie Berna erkennen 79 % der sprachlich schwachen Hälfte und 63 % der sprachlich Starken an dem „Verbrauch bei 100“ nicht, dass sie für die Beantwortung der Frage die Verbrauchswerte zu den Geschwindigkeitsangaben im Diagramm bestimmen müssen. Beide Werte dann auch für die Rechnung zu nutzen, daran scheitern weitere 64 % bzw. 38 % (viele nutzen stattdessen nicht angemessene proportionale Ansätze).

Tab. 6 Überblick über Drop-Outs in Bearbeitungsschritten von Item 2a2 (Kraftstoffverbrauch) für Lernende der sprachlich schwachen und starken Hälfte

Bearbeitungsschritt	Drop-Out-Rate im jeweiligen Bearbeitungsschritt für...	
	... die sprachlich schwache Hälfte	... die sprachlich starke Hälfte
Entnahme der Verbrauchswerte aus Diagramm	79 %	63 %
Umgang mit Einheiten	31 %	25 %
Nutzung beider Verbrauchswerte für Rechnung	64 %	38 %
Übersetzung in eine Rechnung	79 %	70 %
Deutung des Ergebnisses	84 %	78 %

Wer den Satzbaustein „Um wie viel ... liegt über...“ nicht erkennt, kann die Daten nicht in die richtige Rechnung übersetzen, auch hier scheitern wiederum 79 % bzw. 70 %. Dabei sind nicht nur Lesehürden relevant wie bei Berna, sondern auch konzeptuelle Hürden, d.h. das Verständnis, was ein prozentualer Zuwachs ist. Außer dem rein rechnerischen Bearbeitungsschritt „Umgang mit Einheiten“, bei dem in beiden Sprachdritteln relativ wenige Drop-Outs zu verzeichnen sind (31 % und 25 %), können bei den vier anderen Bearbeitungsschritten jeweils sowohl Lesehürden (um was geht es überhaupt?) als auch konzeptuelle Hürden eine Rolle spielen (z. B. bei funktionalen Zusammenhängen werden zwei Größen zueinander in Beziehung gesetzt). Dies lässt sich in den schriftlichen Bearbeitungen nicht trennen. Insgesamt zeigen sich außer für den Schritt „Nutzung beider Verbrauchswerte für eine Rechnung“ nur wenige Unterschiede zwischen den Gruppen.

Auch wenn also in diesem Item 2a2 die potentiellen Lesehürden und konzeptuellen Hürden eng miteinander verknüpft sind, ist das Item doch für Lesehürden typisch mit seiner syntaktischen Komplexität und der hohen Bedeutung der relationstragenden Präpositionen. Auch in anderen Items entstehen zuweilen Lesehürden durch Verdichtungen, deren syntaktische Komplexität nicht überwunden werden kann. Die Verdichtungen tragen dazu bei, dass hohe Satz- und Textlängen (entgegen oft von Lehrkräften geäußerten Annahmen) in der getesteten Altersstufe keine schwierigkeitsgenerierenden Merkmale bilden. Die Bedeutung von Präpositionalkonstruktionen zum Beschreiben komplexer Beziehungen dagegen scheint typisch für mathematische Aufgabentexte, so betont Jorgensen, „It is difficult to think of teaching mathematics without the use of prepositions“ (Jorgensen 2011, S. 324). Insofern liegt hierin eine typische Schwierigkeit der deutschen Sprache von hoher linguistischer Relevanz (Grießhaber 1999), gerade für die Mathematik.

Prozessuale Hürden am Beispiel des Item 1a (Münzturm)

Item 1a ist mit einer Lösungshäufigkeit von 31 % für alle Lernenden tendenziell schwierig, aber für die sprachlich Schwachen besonders mit 23 % Lösungshäufigkeit (mit DIF-Wert genau an der Grenze zur statistischen Signifikanz auf dem 5 %-Niveau).

Die Aufgabenstellung bei Item 1a („Schätze, wie viele Kilometer hoch ein Turm aus 2,4 Milliarden 1-Cent-Münzen ungefähr wäre. Beschreibe, wie du vorgegangen bist.“) konnten fast alle Lernenden in den Interviews richtig reformulieren. Das ließ den Schluss zu, dass das Item keine hohen Leseanforderungen enthält, und führte zur zweiten Vermutung, dass die sprachlich schwachen Lernenden an der Darstellungsleistung der Lösungsweg-Erläuterung gehäuft scheitern könnten. Bei der Erfassung der Bearbeitungsschritte des

Subsamples von 195 Bearbeitungen wurde daher die Erläuterung der Schätzung extra erfasst.

Tabelle 7 zeigt die unterschiedlichen Drop-Out-Raten der sprachlich schwachen und starken Hälften in den einzelnen Bearbeitungsschritten. Während bei der „Erfassung der Fragestellung“ in beiden Gruppen wenige Lernende ausstiegen, zeigt sich im zweiten Lösungsschritt ein großer Unterschied: Von den sprachlich Starken starteten nur 27 % mit einer nicht tragfähigen Schätzung der Münzhöhe, von den sprachlich Schwachen dagegen 56 %. Sie schätzten statt 1-2 mm meist 8-12 mm, was darauf hindeutet, dass sie ohne Bildung eines adäquaten Situationsmodells zum Durchmesser als erste verfügbare Schätzgröße griffen (vgl. Reusser 1989 zur Trennbarkeit von episodischem Aufgabenverständnis und Situationsmodell). Die Interviewanalysen zeigten, dass alle Lernenden, die zum Zeichnen ihres Modells aufgefordert wurden, also zur Explizierung ihres Situationsmodells, dieses sofort revidierten (Wilhelm 2015). Dadurch konnte auch die Vermutung ausgeschlossen werden, dass die Lernenden explizit von einer alternativen Realisierung des Münzturms ausgehen.

Die zweite Vermutung, dass die schriftliche Erklärung, also die Sprachproduktion eine Hürde für sprachlich Schwache sein könnte, stimmt für 36 % der an diesem Schritt angekommenen sprachlich Schwachen (und nur 12 % der Starken). Sie hat daher eine gewisse Bedeutung, doch weniger als das Problem des Bildens eines Situationsmodells. Damit erweisen sich auch in diesem Item sowohl die kommunikative als auch die kognitive Funktion als relevant, mit stärkerem Gewicht auf der kognitiven.

Tab. 7 Überblick über Drop-Outs in Bearbeitungsschritten von Item 1a Münzturm für Lernende der sprachlich schwachen und starken Hälfte

Bearbeitungsschritt	Drop-Out-Rate im jeweiligen Bearbeitungsschritt für...	
	... die sprachlich schwache Hälfte	... die sprachlich starke Hälfte
Erfassen der Fragestellung	12 %	4 %
Schätzen der Münzhöhe	56 %	27 %
Erläutern der Schätzung	36 %	12 %
Wahl der Rechenoperation	34 %	20 %
Stellengerechte Multiplikation	33 %	13 %
Umrechnung Mill/Einheiten	76 %	62 %

Die Bildung des Situationsmodells zeigte sich auch bei anderen Items wie Item 3b (Krake Paul) als eine für sprachlich Schwache schwer überwindbare Hürde, die nicht im Text allein begründet liegt, sondern in den nachfolgenden, mit dem Modellierungsprozess verbundenen kognitiven Verarbeitungsschritten. Weitere prozessuale Hürden sind zum Beispiel bei Item 3b der kognitiv anspruchsvolle Prozessschritt des Festlegens eines kohärenten Ereignisraums als zentraler Schritt der Modellierung (Wilhelm 2015).

Konzeptuelle Hürden in Items 2a2 (Kraftstoffverbrauch) und Item 3b (Krake Paul)

Das Item 2a2 („Um wie viel Prozent liegt der Verbrauch bei 180 km/h über dem Verbrauch bei 100 km/h?“) zeigte nicht nur Lesehürden, sondern auch konzeptuelle Hürden, und zwar in den Bearbeitungsschritten „Nutzung der Verbrauchswerte für Rechnung“

(64 % gegenüber 38 % Drop-Out, vgl. Tabelle 6) und „Übersetzung in eine Rechnung“. Hier schneiden gemäß der Interviewanalysen diejenigen Lernenden schlecht ab, die kein konzeptuelles Verständnis (also keine Grundvorstellungen) zum prozentualen Vergleich aktivieren können, um die Beziehung angemessen zu strukturieren. Ein Beispiel bot bereits Berna, die im Interview das Item 2a2 (Kraftstoffverbrauch) bearbeitet, indem sie den Prozentwert 100 km/h und den Grundwert 180 km/h in die Formel für den Prozentsatz einsetzt und auf diesem Weg 55 % ermittelt (s. o.). Da sie ihren Weg nicht erklären kann, kann auch nicht von einem abweichenden Aufgabenverständnis ausgegangen werden. In der gleichen Aufgabe überwinden andere Lernende die konzeptuelle Hürde nicht, dem funktionalen Zusammenhang zwischen Geschwindigkeit und Verbrauch Bedeutung zu verleihen, weil sie nicht die Grundvorstellung aktivieren können, dass Funktionen immer zwei Größen verbinden (vgl. Wilhelm 2015).

Auch *Item 3a und 3b* (Krake Paul), die in der DIF-Analyse beide als relativ schwer für die sprachlich schwache Hälfte mit signifikantem DIF-Wert identifiziert wurden, beinhalten tiefliegende konzeptuelle Hürden (vgl. Wilhelm 2015). Dies zeigt etwa die Bearbeitung zu Item 3b der Zehntklässlerin Delia, die ebenfalls zum sprachlich schwachen Drittel gehört. Delia kann ihre Entscheidungen zum Zeichnen des Baumdiagramms (in Abb. 2) nicht konzeptuell begründen kann:

- 3 D Ja, also, ähm. Wegen 2 Vorhersagen. Also das ist halt eine [zeigt auf Ast 1 im Baumdiagramm] und das ist halt eine [zeigt auf Ast 2]. [2 sec Pause]
 Und ähm, weil's ja dann am Anfang 2 gibt's, würde ich oben 2 [zeigt auf die 2/6 über Ast 1] und unten halt-
 Ich weiß nicht warum ich die 6 genommen hab, aber- [1 sec Pause] wegen dem Würfel.

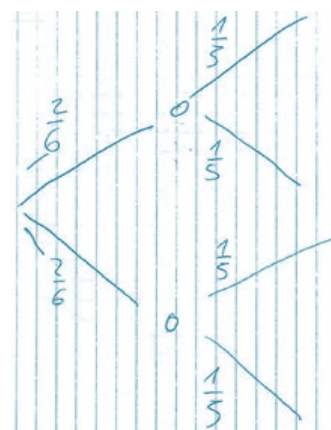


Abb. 2 Delias Baumdiagramm

Ebenso wie bei Delia zeigen viele der videographierten Bearbeitungsprozesse erhebliche Defizite der sprachlich schwachen Lernenden im konzeptuellen Verständnis von zentralen stochastischen Konzepten wie mehrstufige Zufallsversuche oder Simulation, während wiederum die Rephasierung des Textes gut möglich ist (Wilhelm 2015).

Insgesamt erweisen sich also prozessuale und konzeptuelle Hürden als zentral, gerade für die Erklärung der relativen Schwierigkeiten der als signifikant identifizierten Items. Dagegen spielen Lesehürden eine untergeordnete Rolle bei Items mit signifikantem DIF-Wert.

5. Diskussion der Ergebnisse

Da die Untersuchung nur auf eine spezifische Prüfung (ZP10-Mathematik auf dem Niveau Mittlerer Schulabschluss in NRW im Jahr 2012) bezogen ist, muss eine Übertragbarkeit der Ergebnisse auf andere Prüfungen – insbesondere solche, die das Literacy-Konzept weniger stark berücksichtigen – in Anschlussstudien noch geklärt werden. Dennoch geben die Analysen bereits jetzt interessante Befunde zu den leitenden Forschungsfragen.

5.1 Zusammenhänge zwischen Hintergrundfaktoren und Mathematikleistung

F1. Welche sozialen und sprachlichen Hintergrundfaktoren haben den stärksten Zusammenhang zur Mathematikleistung in den Zentralen Prüfungen 10?

Mit den Varianz- und Regressionsanalysen lassen sich ebenso wie mit den Kovarianzanalysen die bestehenden Befunde zu herkunftsbedingten Leistungsdisparitäten ausdifferenzieren, denn sie zeigen, dass die Sprachkompetenz einen erheblich stärkeren Zusammenhang zur Mathematikleistung in den ZP10-Mathematik aufweist, als Faktoren zum sozialen Hintergrund (SES, Migrationshintergrund und Zeitpunkt des Deutscherwerbs). Dabei erweist sich in dieser Untersuchung die lexikalische und grammatische Komponenten umfassende rezeptive und produktive Sprachkompetenz als wichtiger als die – in dieser Untersuchung allerdings kaum reliabel gemessene – rein rezeptive Lesekompetenz (vgl. Tabelle 2): Während das (mit C-Test erfasste) sprachlich schwache Drittel mit einer durchschnittlichen Gesamtpunktzahl von 37,3 auf der Notengrenze zwischen „mangelhaft“ und „ausreichend“ liegt, erreicht das sprachlich starke Drittel mit durchschnittlich 50,3 Punkten die Note „befriedigend“.

Die im amerikanischen Sprachraum oft festgestellte Relevanz der Sprachkompetenz für die Mathematikleistung (Abedi 2006; Secada 1992) konnte somit auch im deutschen Sprachraum vorgefunden werden, und zwar nicht nur, wie meist bisher, bzgl. der Sprachhintergründe oder der Lesekompetenz, sondern bzgl. einem breiter verstandenen und bildungssprachlich ausgerichteten Konstrukt von Sprachkompetenz. Dieser empirische Befund legt nahe, Sprachkompetenz in Large-Scale Assessments, aber auch in der Bildungsberichterstattung konsequenter einzubeziehen, um die Bedeutung sozialer Disparitäten, vermittelt über Sprachkompetenz, erfassen zu können.

Von *sozialer* Benachteiligung ist dennoch auch weiter zu sprechen, weil Sprachkompetenz gemäß breit akzeptierter soziolinguistischer Befunde stark von familiär bedingten Lerngelegenheiten abhängt und somit ein soziales Phänomen ist (Cook-Gumperz 1973, S. 1). Doch während die Relevanz des SES und des Migrationshintergrunds vor allem globale bildungs- und sozialpolitische Konsequenzen haben kann, verweist die Bedeutung der Sprachkompetenz konkret auf Handlungsfelder im Unterricht, ist also ein *didaktisch höchst relevanter Angriffspunkt* zur Reduktion sozialer Ungleichheit. Dass ein entsprechend ausgerichteter sprachsensibler Mathematikunterricht vermutlich gleichzeitig allen Lernenden zugutekommen kann, macht ihn zum besonders wichtigen Handlungsfeld für Unterrichtsentwicklung, auch wenn die Vermutung noch zu überprüfen ist.

Gleichwohl müssen auch die methodischen Grenzen der Untersuchung betrachtet werden: Die Stichprobe ist lediglich für den mittleren Schulabschluss repräsentativ, es fehlen die Lernenden von Gymnasial- und Hauptschulniveau. Die Reliabilitätsproblematik der Bewertung durch die Lehrkräfte, die relativ grobe Erfassung des SES durch die Bücheraufgabe und die fehlende Kontrolle der kognitiven Grundfertigkeiten bilden Einschränkungen, die bei der Interpretation der Ergebnisse berücksichtigt werden müssen. Eine Anschlussuntersuchung sollte daher außerhalb der abschlussrelevanten High-Stakes-Bedingungen durchgeführt werden, um diese Grenzen zu überwinden. Wünschenswert wäre

auch eine ausdifferenziertere Operationalisierung des Faktors Sprachkompetenz, auch wenn einige Linguisten dies für nicht möglich halten.

5.2 Identifikation von sprachlich bedingten Hürden

Um konkretere Konsequenzen für den Unterricht ziehen zu können, waren Analysen auf Itemebene notwendig, um die Hürden genauer zu spezifizieren. Denn die statistische Analyse der Zusammenhänge allein kann keinen Aufschluss darüber geben, wie genau die Sprachkompetenz auf die Mathematikleistung Einfluss nimmt oder ob der Zusammenhang nur durch gemeinsame andere Faktoren, wie z. B. der hier nicht kontrollierten kognitiven Grundfertigkeit entsteht.

In den Analysen der auffälligen Items, ihrer schriftlichen Bearbeitungen und videographierten Bearbeitungsprozesse konnte dagegen der Zusammenhang zwischen Sprachkompetenz und Bearbeitungserfolg genauer verstanden werden. Zwar ist auch die Identifikation und Typisierung von Hürden keineswegs durch die erfolgten Analysen abgeschlossen und unmittelbar verallgemeinerbar, sie liefert dennoch interessante erste Kategorien, deren weitere Beforschung auch für andere Tests lohnenswert erscheint: Rekonstruiert werden konnten drei Arten von Hürden für sprachlich schwache Lernende:

- *Lesehürden* in der Texterschließung, insbesondere durch komplexe Satzstrukturen und morphologische Hürden mit der zentralen Bedeutung der Präpositionen (Gürsoy 2013; Grießhaber 1999). Sie verweisen nicht nur auf *language biases* (Abedi 2006), die aus den Prüfungen eliminiert werden sollten (auch wenn solche als Projektergebnis durchaus spezifizierbar sind, vgl. Gürsoy et al. 2013), sondern daneben auch auf angemessene Leseanforderungen, denen einige Lernende noch nicht in ausreichendem Maße gewachsen sind; hierfür müssen entsprechend fokussierte Lerngelegenheiten entwickelt werden.
- *prozessuale Hürden* bei kognitiv anspruchsvolleren Prozessen, z. B. im Bilden des Situationsmodells oder definitorischem Sich-Festlegen (Wilhelm 2015; ähnlich auch bei Duarte et al. 2011) und
- *konzeptuelle Hürden* im konzeptuellen Verständnis der erschließenden mathematischen Konzepte (Wilhelm 2015; ähnlich auch bei Ufer et al. 2013).

Die hier exemplarisch skizzierten Ergebnisse der Tiefenanalysen zeigen, dass sprachlich schwache Lernende in Prüfungssituationen nicht nur Hürden im Leseprozess erfahren, die der kommunikativen Funktion von Sprache zuzuschreiben sind. Vielmehr scheitert eine vergleichsweise große Zahl dieser Jugendlichen auch an Hürden im weiteren Lösungsprozess. Zu ähnlichen Ergebnissen kommen auch DIF-Analysen zu Vera 3 (Haag et al. 2013).

Diese Hürden scheinen sich jedoch nicht auf kurzfristige Probleme der spezifischen Prüfungssituation zu beziehen, sondern eher auf langfristig kumulierte Defizite im Überwinden prozessualer und konzeptueller Hürden. Sie geben daher erste Hinweise zur Erklärung des statistisch beobachtbaren Phänomens der *kumulativ wachsenden Schwierigkeiten*

sprachlich schwacher Lernender (Herwartz-Emden 2003, S. 692), die eher auf die kognitive Funktion von Sprache als auf die kommunikative Funktion zurück zu führen sein scheinen. Dafür geben vertiefende Analysen in *Lernprozessen* interessante Hinweise (Prediger 2013; Zindel 2015).

Gerade die Befunde zu prozessualen und konzeptuellen Hürden erscheinen von großer Bedeutung für das Feld, die eine genauere Analyse der sprachlichen Einschränkungen in Lernprozessen zur dringlichen Forschungsnotwendigkeit machen. Weitere mathematik- und sprachdidaktische Forschung zu diesem schulpraktisch höchst bedeutsamen Thema wird in den nächsten Jahren notwendig sein, um die verschiedenen Lesehürden, prozessualen und konzeptuellen Hürden in ihren Wirkungen genauer einzugrenzen und zu verstehen. Daran müssen sich Entwicklungsforschungsstudien zur Entwicklung von Fördermaßnahmen anschließen (vgl. Prediger u. Özdil 2011 für Forschungs- und Entwicklungsbedarfe sowie Prediger u. Wessel 2013 für eine erste Realisierung).

Dank

Die Studie wurde mit Mitteln des Ministeriums für Schule und Weiterbildung von Nordrhein-Westfalen gefördert. Den Inhalt der Veröffentlichung verantworten ausschließlich die Verfassenden. Herzlich gedankt wird allen beteiligten Schulen, Lehrkräften und Lernenden. Wir danken außerdem den Gutachterinnen und Gutachtern sowie der Herausgeberin für ihr sehr sorgfältiges Durchdenken. Sie haben erheblich beigetragen, den Artikel weiter zu entwickeln in Bezug auf argumentative Stringenz und Nachvollziehbarkeit.

Literatur

- Abedi, J. (2006). Language issues in item-development. In S. M. Downing, & T. M. Haladyna (Hrsg.), *Handbook of test development* (S. 377-398). Mahwah, NJ: Erlbaum.
- Autorengruppe Bildungsberichterstattung (2012). *Bildung in Deutschland. Ein indikatorengestützter Bericht mit einer Analyse zu Perspektiven des Bildungswesens im demografischen Wandel*. Bielefeld: Bertelsmann.
- Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323-410). Opladen: Leske + Budrich.
- Baur, R. S., & Spettmann, M. (2010). Lesefertigkeiten testen und fördern. In C. Benholz, G. Kniffka, & E. Winters-Ohle (Hrsg.), *Fachliche und sprachliche Förderung von Schülern mit Migrationsgeschichte* (S. 95-114). Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G., & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Wendt, H., Ünlü, A., Valtin, R., Euen, B., Kasper, D., & Tarelli, I. (2012). Leistungsprofile von Viertklässlerinnen und Viertklässlern in Deutschland. In W. Bos, H. Wendt, O. Köller, & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 269-301). Münster: Waxmann.
- Brown, C. L. (2005). Equity of Literacy Based Math Performance Assessments for English Language learners. *Bilingual Research Journal*, 29(2), 337-363.
- Büchter, A., & Pallack, A. (2012). Methodische Überlegungen und empirische Analysen zur impliziten Standardsetzung durch zentrale Prüfungen. *Journal für Mathematik-Didaktik*, 33(1), 59-85.
- Burns, T., & Shadoian-Gersing, V. (2010). The importance of effective teacher education for diversity. In OECD (Hrsg.), *Educating Teachers for Diversity. Meeting the Challenge* (S. 19-40). Paris: OECD.
- Cook-Gumperz, J. (1973). *Social control and socialization: A study of class differences in the language of maternal control*. London: Routledge & Kegan Paul.
- Cummins, J. (2000). *Language, power and pedagogy*. Clevedon: Multi lingual Matters.
- Daller, H. (1999). *Migration und Mehrsprachigkeit*. Frankfurt et al.: Peter Lang.
- De Houwer, A. (2009). *An Introduction to Bilingual Development*. Bristol: Multilingual Matters.

- Duarte, J., Gogolin, I., & Kaiser, G. (2011). Sprachlich bedingte Schwierigkeiten von mehrsprachigen Schülerinnen und Schülern bei Textaufgaben. In S. Prediger, & E. Özdi (Hrsg.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit* (S. 35-54). Münster: Waxmann.
- Ehmke, T., Hohensee, F., Heidemeier, H., & Prenzel, M. (2004). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost, & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 225-253). Münster: Waxmann.
- Ellerton, N., & Clarkson, P. (1996). Language Factors in Mathematics Teaching and Learning. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Hrsg.), *International Handbook of Mathematics Education* (Vol. 4, S. 987-1033). Dordrecht: Kluwer.
- Gellert, U. (2011). Mediale Mündlichkeit und Dekontextualisierung. Zur Bedeutung und Spezifik von Bildungssprache im Mathematikunterricht der Grundschule. In S. Prediger, & E. Özdi (Hrsg.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit* (S. 97-116). Münster: Waxmann.
- Gogolin, I. (2009). Zweisprachigkeit und die Entwicklung bildungssprachlicher Fähigkeiten. In I. Gogolin, & U. Neumann (Hrsg.), *Streitfall Zweisprachigkeit – The Bilingualism Controversy* (S. 263-280). Wiesbaden: Verlag für Sozialwissenschaften.
- Grießhaber, W. (1999). Präpositionen als relationierende Prozeduren. In A. Redder, & J. Rehbein (Hrsg.), *Grammatik und mentale Prozesse* (S. 241-260). Tübingen: Stauffenburg.
- Grotjahn, R. (1992). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Band 1. Bochum: Broockmeyer.
- Gürsoy, E. (2013). *Textkohäsion und Textkohärenz in den Zentralen Prüfungen 10 Mathematik (NRW) - Eine multiperspektivische Untersuchung zu Kohäsionsmitteln im Deutschen und Türkischen und zur Repräsentation von Kohäsion bei türkisch-deutschsprachigen Schülerinnen und Schülern*. Dissertation. Essen: Universität Duisburg-Essen.
- Gürsoy, E., Benholz, C., Renk, N., Prediger, S., & Büchter, A. (2013). Erlös = Erlösung? – Sprachliche und konzeptuelle Hürden in Prüfungsaufgaben. *Deutsch als Zweitsprache*, 13(1), 14-24.
- Haag, N., Hept, B., Stanat, P., Kuhl, P., & Pant, H. A. (2013). Second language learners' performance in mathematics: Disentangling the effects of academic language features. *Learning and Instruction*, 28(0), 24-34.
- Halliday, M.A.K. (1993). Some grammatical problems in scientific English. In M.A.K. Halliday & J.R. Martin (Hrsg.), *Writing science: Literacy and discursive power* (S. 69-85). Pittsburgh: University Press.
- Heinze, A., Herwartz-Emden, L., & Reiss, K. (2007). Mathematikkennntnisse und sprachliche Kompetenz bei Kindern mit Migrationshintergrund zu Beginn der Grundschulzeit. *Zeitschrift für Pädagogik*, 53(4), 562-581.
- Herwartz-Emden, L. (2003). Einwandererkinder im deutschen Bildungswesen. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer, & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (S. 661-709). Reinbek: Rowohlt.
- Jorgensen, R. (2011). Language, culture and learning mathematics: A Bourdieuan analysis of indigenous learning. In C. Wyatt-Smith, J. Elkins, & S. Gunn (Hrsg.), *Multiple perspectives on difficulties in learning literacy and numeracy* (S. 315-329). Dordrecht: Springer.
- Kaiser, G., & Schwarz, I. (2003). Mathematische Literalität unter einer kulturell-sprachlichen Perspektive. *Zeitschrift für Erziehungswissenschaft*, 6(3), 357-377.
- Knoche, N., & Lind, D. (2004). Bedingungsanalysen mathematischer Leistungen: Leistungen in den anderen Domänen, Interesse, Selbstkonzept und Computernutzung. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA 2000* (S. 205-226). Wiesbaden: Verlag für Sozialwissenschaften.
- Leutner, D., Klieme, E., Meyer, K., & Wirth, J. (2004). Problemlösen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost, & U. Schiefele (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147-175). Münster: Waxmann.
- Maier, H., & Schweiger, F. (1999). *Mathematik und Sprache. Zum Verstehen und Verwenden von Fachsprache im Unterricht*. Wien: oebv & hpt.
- Martiniello, M. (2009). Linguistic complexity, schematic representations and differential item functioning for English language learners in math tests. *Educational assessment*, 14(3-4), 160-179.
- Mikrozensus (2011). *Bevölkerung nach Migrationsstatus regional*. Hrsg. von Statistischen Ämtern des Bundes und der Länder. https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/MigrationIntegration/BevoelkerungMigrationsstatus5125203117004.pdf?_blob=publicationFile. Gesehen 31. Mai 2013.
- Morek, M., & Heller, V. (2012). Bildungssprache – Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs. *Zeitschrift für angewandte Linguistik*, 57(1), 67-101.

- Neubrand, M. (2001). PISA – „Mathematische Grundbildung“ / „mathematical literacy“ als Kern einer internationalen und nationalen Leistungsstudie. In G. Kaiser, N. Knoche, D. Lind, & W. Zillmer (Hrsg.), *Leistungsvergleiche im Mathematikunterricht. Ein Überblick über aktuelle Studien* (S. 177-194). Hildesheim: Franzbecker.
- OECD (2007). *PISA 2006. Science competencies for tomorrow's world*. Vol. 2: Data. Paris: OECD.
- OECD (2009). *PISA 2006. Technical Report*. Paris: OECD.
- Paulus, C. (2009). Die „Bücheraufgabe“ zur Bestimmung des kulturellen Kapitals bei Grundschulern. http://psydok.sulb.uni-saarland.de/volltexte/2009/2368/pdf/BA_Artikel.pdf. Gesehen 5. Mai 2015.
- Pimm, D. (1987). *Speaking mathematically. Communication in mathematics classrooms*. London, New York: Routledge.
- Prediger, S. (2013). Darstellungen, Register und mentale Konstruktion von Bedeutungen und Beziehungen. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. J. Vollmer (Hrsg.), *Sprache im Fach – Sprachlichkeit und fachliches Lernen* (S. 167-183). Münster: Waxmann.
- Prediger, S., & Özdil, E. (Hrsg.). (2011). *Mathematiklernen unter Bedingungen der Mehrsprachigkeit – Stand und Perspektiven zu Forschung und Entwicklung*. Münster: Waxmann.
- Prediger, S., & Wessel, L. (2013). Fostering second language learners' constructions of meanings for fractions – Design and effects of a language- and mathematics-integrated intervention. *Mathematics Education Research Journal*, 25(3), 435-456.
- Prediger, S., Renk, N., Büchter, A., Gürsoy, E., & Benholz, C. (2013). Family background or language disadvantages? Factors for underachievement in high stakes tests. In A. Lindmeier, & A. Heinze (Hrsg.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, S. 49-56). Kiel: PME.
- Redder, A., & Weinert, S. (Hrsg.) (2013). *Sprachförderung und Sprachdiagnostik. Interdisziplinäre Perspektiven*. Münster: Waxmann.
- Reusser, K. (1989). *Vom Text zur Situation zur Gleichung. Kognitive Simulation von Sprachverständnis und Mathematisierung beim Lösen von Textaufgaben*. (Habilitationsschrift). Bern: Universität Bern.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? *Psychologische Rundschau*, 57(2), 69-86.
- Rösch, H., & Paetsch, J. (2011). Sach- und Textaufgaben im Mathematikunterricht als Herausforderung für mehrsprachige Kinder. In S. Prediger, & E. Özdil (Hrsg.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit* (S. 55-76). Münster: Waxmann.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Bern: Hans Huber.
- Schnabel, K. U. & Schwippert, K. (2000). Einflüsse sozialer und ethnischer Herkunft beim Übergang in die Sekundarstufe II und den Beruf. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Band 1 (S. 261-281). Opladen: Leske + Budrich.
- Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah: Lawrence Erlbaum.
- Secada, W. G. (1992). Race, ethnicity, social class, language and achievement in mathematics. In D. A. Grouws (Hrsg.), *Handbook of Research on Mathematics Teaching and Learning* (S. 623-660). New York: MacMillan.
- Tarelli, I., Schwippert, K., & Stubbe, T. C. (2012). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 247-268.). Münster: Waxmann.
- Thürmann, E., Vollmer, H., & Pieper, I. (2010). *Language(s) of schooling: Focusing on vulnerable learners*. (= The linguistic and educational integration of children and adolescents from migrant backgrounds. Studies and resources 2). Straßbourg: Council of Europe. http://www.coe.int/t/dg4/linguistic/Source/Source2010_ForumGeneva/2-VulnerLLearnersThurm_EN.pdf. Gesehen 13. Februar 2015.
- Ufer, S., Reiss, K., & Mehringer, V. (2013). Sprachstand, soziale Herkunft und Bilingualität: Effekte auf Facetten mathematischer Kompetenz. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach – Sprachlichkeit und fachliches Lernen* (S. 167-184). Münster: Waxmann.
- Walzbeug, A. (2014). Is there a language-based social disadvantage in solving mathematical items? *Learning, Culture and Social Interaction*, 3(2), 159-169.
- Werning, R., Löser, J. M., & Urban, M. (2008). Cultural and social diversity: An analysis of minority groups in German schools. *The Journal of Special Education*, 42(1), 47-54.
- Wilhelm, N. (2015, in Vorbereitung). Hürden in Löseprozessen von mathematischen Prüfungsaufgaben durch sprachlich schwache Lernende – Quantitative und qualitative Analysen (Arbeitstitel). Dissertation, TU Dortmund.

- Wolf, M. K. & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. *Educational Assessment*, 14(3-4), 139-159.
- Zindel, C. (2015, im Druck). „Wenn ich wüsste, was davon was ist...“ – konzeptuelle und sprachliche Hürden bei funktionalen Abhängigkeiten. Erscheint in H. Linneweber-Lammerskitten (Hrsg.), *Beiträge zum Mathematikunterricht*. Münster: WTM Verlag.