

Knut Schwippert

Was wird aus den Büchern? Sozialer Hintergrund von Lernenden und Bildungsungleichheit aus Sicht der international vergleichenden Erziehungswissenschaft

Zusammenfassung

Internationale empirische Bildungsstudien sind vor allem dazu angelegt, auf Eigenschaften von Bildungssystemen aufmerksam zu werden, die sich erst im Vergleich mit anderen Bildungssystemen zeigen, da beispielsweise innerhalb eines Systems keine Varianz bezüglich eines Merkmals besteht oder aufgrund einer fehlenden Normvorstellung das Ausmaß eines Effekts innerhalb eines Systems nicht bewertet werden kann. Eine besondere Aufmerksamkeit hat bei diesen Betrachtungen die Frage der Abhängigkeit des Schulerfolgs vom sozialen Hintergrund erfahren. Zur Erfassung des sozialen Hintergrunds hat sich insbesondere in westeuropäisch geprägten Staaten eine Orientierung an der Kapitaltheorie von Bourdieu etabliert, in der eine Unterscheidung des Status anhand von ökonomischen, sozialen und kulturellen Merkmalen (Kapital) vorgenommen wird. In internationalen Bildungsstudien hat sich die Abfrage des heimischen Buchbesitzes als gut zu erhebender und robuster Indikator für kulturelles und ökonomisches Kapital bewährt. Vor dem Hintergrund der rasanten Entwicklung digitaler Medien und dessen Einzug in private Haushalte und Schulen stellt sich die Frage, ob der Buchbesitz überhaupt noch als geeigneter Indikator für den ökonomisch-kulturellen Hintergrund von Familien genutzt werden kann. Der Beitrag geht dieser Frage anhand einer Analyse der deutschen Teildatensätze der Internationalen Grundschul-Lese-Untersuchung (IGLU) von 2001 bis 2016 nach. Anhand von probabilistischen Testmodellen wird die Messeigenschaft der Büchervariablen im Verlauf von 15 Jahren nachgezeichnet. Hierbei stellt sich die Information über den heimischen Buchbesitz als robuster und stabiler Indikator zur Beschreibung des bildungsrelevanten familiären Hintergrunds im zeitlichen Verlauf dar.

Prof. Dr. Knut Schwippert, Fakultät für Erziehungswissenschaft, EW 1: Allgemeine und International Vergleichende Erziehungswissenschaft, Universität Hamburg, Von-Melle-Park 8, 20146 Hamburg, Deutschland
E-Mail: knut.schwippert@uni-hamburg.de

Schlagworte*Heimischer Buchbesitz; kultureller Hintergrund; digitale Medien***What's about the books? Social background of students and educational opportunities from the perspective of international large-scale surveys****Abstract**

International large-scale assessments are designed to identify characteristics of educational systems which only can be identified by comparing different systems. The comparison is necessary once within a country characteristic does not show any variance or if normative standards are not available to that a characteristic could be related. Special attention in this context is given to the relation of social background of students and their success in school. As theoretical framework – especially in European countries – the work from Pierre Bourdieu is often used in which social, cultural and economic capitals are distinguished. In international large-scale surveys the question to students and/or their parents about the books at home is tradition since some decades. This indicator is an easy to survey and robust indicator for cultural and economic capital. Taken into account the accelerating availability of digital media in the households the question rises, if the question on the books at home is still appropriate. This questions lies in the focus of the article. Answers are sought by analyzing background information and form the German data of the Progress in International Reading Literacy Study (PIRLS) in the four cycles from 2001 to 2016. By using models of probabilistic item response theory the characteristics of the indicators available – which address the cultural and economic background of the students – are investigated over a 15 year period. The results show that the information about the books at home is still a robust and stable indicator for the description of the educational background of the students.

Keywords*Books at home; Cultural background; Digital media***1. Entwicklung und Ziele von groß angelegten Schulvergleichsuntersuchungen**

Ende der 1950er Jahre wurde durch eine Gruppe von interessierten Personen die Frage aufgegriffen, ob Bildungssysteme anhand quantitativer empirischer Untersuchungen verglichen werden können. Im Rahmen einer 1959 bis 1961 durchgeführten Machbarkeitsstudie in 12 Staaten (Foshay, 1962) wurden sowohl Leistungsdaten als auch Hintergrundmerkmale erhoben. Aus dieser Initiative entwickelte sich die International Association for the Evaluation of Educational

Achievement (IEA), die bis heute namhafte internationale Bildungsstudien initiiert und organisiert (Husén & Postlethwaite, 1996). Da die IEA als Wegbereiter solcher Studien angesehen werden kann, liegt der Fokus dieses Beitrags auf der Initiierung und Entwicklung von IEA-Studien. Die damals wie heute im Fokus stehende Frage ist, anhand welcher Merkmale Staaten bzw. deren Bildungssysteme angemessen verglichen werden können. Eine immer wiederkehrende und in vielen der teilnehmenden Staaten formulierte Frage richtet sich darauf, ob bzw. in welchem Umfang der Lernerfolg von Schülerinnen und Schülern von Merkmalen des sozialen Hintergrunds abhängt. Zum einen geht es um die Frage, wie stark sich der Zusammenhang zwischen den Staaten gleicht bzw. unterscheidet und ob ggf. weitere Merkmale identifiziert werden können, die die Unterschiedlichkeit zwischen den Staaten erklären helfen. Dazu werden die Tests und Befragungsinstrumente in die jeweiligen Landessprachen übersetzt. Auch wenn dies mit hohem Aufwand (Übersetzungskontrolle und Rückübersetzungen zur Qualitätssicherung) betrieben wird (Ebbs & Wry, 2017), bleibt offen, ob die augenscheinlich gleich erhobenen Informationen auch in allen Staaten Gleiches erfassen – mit anderen Worten, ob sie gleich zu interpretieren sind und damit auch als valide Information zum Vergleich herangezogen werden können. Dies sind einerseits materielle Ausstattungsmerkmale und andererseits auch Merkmale von Prozessen bzw. Personen. Der Ansatz, anhand einiger empirisch gut zu erfassender Indikatoren bzw. Skalen im Rahmen eines Modells ein reduziertes empirisches Abbild der Wirklichkeit darzustellen, ist zum Standard quantitativer empirischer Bildungsforschung geworden (Rost, 2004). Im Sinne der Sparsamkeitsprämisse der Wissenschaft wird hierbei Wert daraufgelegt, eine reale Situation bzw. einen realen Prozess mit so wenigen Merkmalen wie möglich zu erfassen und zu analysieren. Bei diesem Ansatz geht es nicht – wie bei anderen empirischen Vorgehensweisen – um die umfangreiche und möglichst vollständige Beschreibung von Situationen oder Prozessen, sondern um die exemplarische und sparsame Beschreibung von Gesetzmäßigkeiten, die in Modellen als reduziertes Abbild der Wirklichkeit beschrieben und geprüft werden kann.

Neben der inhaltlichen und der am jeweiligen Erkenntnisinteresse ausgerichteten Auswahl der zu erfassenden Informationen spielt deren empirische Güte eine zentrale Rolle. So wird die Anwendung der (Haupt-)Testgütekriterien Objektivität, Reliabilität und Validität nicht mehr hinterfragt – was nicht heißt, dass diese Kriterien in ihrer Auslegung in den letzten Jahrzehnten gleichgeblieben sind (Kane, 2013). Die modellorientierte Entwicklung von groß angelegten internationalen Bildungsstudien unterliegen der Herausforderung, die für die Modellprüfungen notwendigen Informationen sowohl national als auch international so zu erfassen, dass Vergleiche sowohl innerhalb als auch zwischen Staaten möglich werden und darüber hinaus – bei Trenduntersuchungen – auch Analysen über die Zeit hinweg realisiert werden können.

1.1 Vergleich – von was?

Die Intention der an der Initiierung internationaler empirischer Bildungsstudien beteiligten Wissenschaftlerinnen und Wissenschaftler sind nicht die oft aus diesen Untersuchungen (i. d. R. von Dritten) abgeleiteten *Liga-Tabellen*. Ziel und Zweck dieser Untersuchungen ist es, den Horizont zu erweitern, von dem auf das eigene (nationale) Bildungssystem geschaut werden kann. Auch wenn kein normativer Diskurs oder gesellschaftlicher Konsens bestimmter Phänomene vorliegt, kann im Vergleich mit anderen Staaten beurteilt werde, ob zum Beispiel die Abhängigkeit des Lernerfolgs der Schülerinnen und Schüler vom sozialen Hintergrund als gering, mittel oder stark zu charakterisieren ist. Es gibt Merkmale, die national innerhalb von Bildungssystemen kaum variieren, wodurch eine empirische Prüfung der Wirkung bzw. Wirksamkeit innerhalb des Systems nicht möglich ist. Erst wenn diese Merkmale (unter sonst gleichen Bedingungen) in ihrer Variation über Staatengrenzen hinweg international vergleichend betrachtet werden, lassen sich Aussagen über mögliche Ursachen- und Wirkungszusammenhänge formulieren.

Je nach politischem, historischem und kulturellem Hintergrund hat sich in jedem Staat ein individuelles Bildungssystem entwickelt und etabliert. Ein entsprechender Vergleich von intentional und historisch bedingt variierenden Systemen ist eine Herausforderung. Jeder geplante Vergleich unterliegt der Frage „kann man Äpfel mit Birnen“ vergleichen? Wilfried Bos antwortete einmal auf diese Frage: „Ja – wenn mich der Anteil von Vitamin C in 100 Gramm Fruchtfleisch interessiert, kann ich Äpfel mit Birnen vergleichen“. Die Kunst des internationalen Vergleichs von Bildungssystemen liegt somit darin, relevante und auch tatsächlich vergleichbare Merkmale in den Blick zu nehmen und unter Berücksichtigung verschiedener (bekannter) Rahmenbedingungen in Relation zu setzen. In den international groß angelegten empirischen Bildungsstudien werden neben Leistungsdaten auch Rahmenbedingungen anhand von Fragebögen erfasst, die sich – je nach Fragestellung und Fokus der Untersuchung – an verschiedene Akteure im Bildungssystem richten und aus verschiedenen Wissenschaftsdisziplinen speisen. Im Fokus von IEA-Studien stehen oft Schülerinnen und Schüler, in der Regel ihre Lehrkräfte und die Schulleitungen, aber auch ihre Eltern oder Personen aus der Bildungsadministration und/oder -politik. Die durch die Befragung dieser Akteure geschaffene Datenbasis kann einerseits generelle Überblicke und eine Gegenüberstellung von Merkmalen der Bildungssysteme erlauben und andererseits auch Vergleiche von Bildungssystemen unter Dokumentation des Vergleichsrahmens ermöglichen. Die IEA ist hierbei traditionell an dem gemeinsamen Erkenntnisinteresse der nationalen Bildungsexpertinnen und Bildungsexperten der beteiligten Staaten orientiert. Somit gelingt es (gerechtfertigte) Vergleiche zu beschreiben und gleichzeitig die Grenzen der Vergleiche zu dokumentieren.

1.2 Tests

Ein Hauptaugenmerk im Rahmen von groß angelegten internationalen Schulvergleichsuntersuchungen liegt auf der Betrachtung der durch Schülerinnen und Schüler erreichten Kompetenzen. Kompetenzen werden hierbei als die Fähigkeit beschrieben, in variablen Situationen neuen Herausforderungen begegnen zu können (Klieme & Hartig, 2007; Weinert, 1999). Seit den ersten Studien der IEA liegt hierbei der Fokus auf den in Regelschulsystemen als zentral identifizierten Domänen: Mathematik, Naturwissenschaft (in Deutschland insbesondere: Physik, Chemie, Geographie, Biologie), Lesen, Schreiben und erste Fremdsprache (wenn diese Englisch ist). In späteren Untersuchungen sind – in Reflexion internationaler gesellschaftlicher Veränderungen – unter anderem auch Untersuchungen zur politischen Bildung, zur Klassenraumführung und (zunächst) Verfügbarkeit und (später) Nutzung von Computern hinzugekommen (Goy, van Ackeren & Schwippert, 2008).

Unter Berücksichtigung nationaler Curricula wird ein gemeinsamer internationaler Rahmen zur Erfassung der Kompetenzen von Schülerinnen und Schülern gesteckt. In diesem Rahmen werden Erwartungen formuliert, was Schülerinnen und Schüler sicher auf mittlerem oder auf hohem Niveau bearbeiten können sollten. Anhand von fachlich variierenden Testaufgaben kann so ein Leistungsspektrum empirisch abgebildet werden. Aktuelle Definitionen der in den Tests erfassten Kompetenzen lauten exemplarisch für die Progress in International Reading Literacy Study (PIRLS) und für die International Computer and Information Literacy Study 2013 (ICILS 2013) wie folgt:

- „The PIRLS definition of reading literacy is grounded in IEA’s 1991 study, in which reading literacy was defined as ‘the ability to understand and use those written language forms required by society and/or valued by the individual’” (Mullis, Martin & Sainsbury, 2015, S. 11).
- Die ICILS 2013-Studie untersucht, welche Kompetenzen Jugendliche im Umgang mit Computern und Informationstechnologien haben, die notwendig sind, sich aktiv in einer technologischen Welt bewegen zu können. Computer and information literacy (ICT) ist definiert als: „an individual’s ability to use computers to investigate, create and communicate in order to participate effectively at home, at school, in the workplace and in society“ (Fraillon, Schulz & Ainley, 2013, S. 17).

Beiden Definitionen ist gemein, dass sie den Lebensweltbezug der betrachteten Kompetenzen betonen.¹ Somit lassen sich die betrachteten Kompetenzen als notwendige – wenn auch nicht hinreichende – Voraussetzungen für Bildung verstehen. Nach Klafkis (2007) Auffassung ist Bildung die Befähigung zur vernünftigen Selbstbestimmung, Mitbestimmung und Solidaritätsfähigkeit, die zur Gewinnung von Individualität und Gemeinschaftlichkeit führt. Solch ein Verständnis von

1 Vgl. auch die entsprechende Definitionen zur Naturwissenschaft (Jones, Wheeler & Centurino, 2013, S. 26) und zur Mathematik (Grønmo, Lindquist, Arora & Mullis, 2013, S. 11).

Bildung setzt in einer modernen Gesellschaft die Beherrschung verschiedener Kulturtechniken voraus. Um in dieser Gesellschaft selbständig, selbstbestimmt aber auch verantwortungsvoll agieren zu können, rücken national wie auch international Kulturtechniken in den Fokus, „die als grundlegende Prämissen für die Teilhabe an gesellschaftlicher Kommunikation durch schulische Arbeit universalisiert werden müssen – als historische Gestalt dessen, was heute allgemeine Bildung heißen kann“ (Tenorth, 2005, S. 26): Erst lesen und schreiben zu können, eröffnet den Individuen die Möglichkeit, an einer modernen Informationsgesellschaft teil zu haben. Auch Kenntnisse in Mathematik und den Naturwissenschaften stellen wichtige Voraussetzungen dar, sich selbstbestimmt und verantwortungsvoll in der Gesellschaft zu bewegen, sich in und mit ihr zu entwickeln (vgl. Schwippert, 2015).

Bei der Erfassung von Kompetenzen in groß angelegten internationalen empirischen Bildungsstudien geht es nicht um eine detaillierte Individualdiagnostik, sondern um die Erfassung von Kompetenzständen in den betrachteten Bildungssystemen. Entsprechend wird das Studiendesign so optimiert, dass valide Aussagen über Gruppen von Schülerinnen und Schülern getroffen werden und nicht differenzierte Aussagen in Bezug auf einzelne Personen. Dies hat den positiven Nebeneffekt, dass – abstrakt betrachtet – die individuelle Tagesform der Schülerinnen und Schüler keine solche Rolle spielt wie beispielsweise im Rahmen von (zentralen) Abschlussprüfungen. Aufgrund der umfangreichen Befragung von Schülerinnen und Schülern kann davon ausgegangen werden, dass sich Tagesform abhängige Unterschiede zwischen „wahrer“ Kompetenz und gezeigter Kompetenz (diese wird auch als Performanz bezeichnet) im Mittel über alle Beobachtungen hinweg ausgleichen. Lediglich systematische Beeinflussungen (z. B. durch *Priming* erzeugte bewusste negativ-Manipulation der Ergebnisse oder in Gruppen induzierte Verweigerungen) würden das Gesamtbild verfälschen. Aber selbst wenn solche Abweichungen einen Einfluss auf die Gesamtergebnisse haben würden, wäre die dahinterliegende Haltung bzw. Einstellung auch als ein Merkmal des Bildungssystems zu interpretieren. Jedoch wäre in diesem Fall wichtig, dass solche Merkmale identifiziert und dokumentiert werden.

1.3 Hintergrundmerkmale

Wie bereits erwähnt ist das Ziel von groß angelegten Schulvergleichsuntersuchungen nicht das Ranking von Staaten. Vielmehr werden Merkmale der Bildungssysteme miteinander in Beziehung gesetzt, um so – aus der Interpretation von Ähnlichkeiten und Unterschieden zwischen Bildungssystemen – Schlussfolgerungen für die Reform oder Weiterentwicklung von Bildungs- bzw. Schulsystemen, Unterricht und Rahmenbedingungen des Lehrens und Lernens ziehen zu können. Je nach Wahrnehmung von be- bzw. anstehenden Herausforderungen für das Bildungssystem werden hierzu bewährte Modelle zur Erklärung von Unterschieden innerhalb und zwischen Bildungssystemen – je nach Ebene des Erkenntnisinteresses – herangezogen.

Schon seit den ersten Untersuchungen der IEA wurden Vergleiche und Relationen vorgestellt, ohne Rangreihen der beteiligten Staaten vorzunehmen (Foshay, 1962; Husén, 1967). Im Fokus stand vielmehr die Identifikation von Mustern im Sinne von Gemeinsamkeiten bzw. Unterschieden, aufgrund deren Reformen bzw. Innovationen in den jeweiligen Bildungssystemen initiiert werden konnten. Eine besondere Rolle hat hierbei seit der ersten internationalen empirischen Bildungsstudie die Frage gespielt, inwieweit die individuellen Voraussetzungen von Schülerinnen und Schülern im Schulsystem eine Rolle spielen. Hierbei steht die Frage im Raum, ob familiär bedingte (soziale) Unterschiede sich durch Schule und Unterricht verringern, gleichbleiben oder sogar vergrößern.

Im Zuge der Weiterentwicklung der IEA-Studien wurde das Repertoire an Hintergrundmerkmalen der Schülerinnen und Schüler – insbesondere orientiert an Befunden aus der Soziologie – weiter ausdifferenziert. Hierbei spielt auch in den aktuellen Studien eine differenzielle Betrachtung des sozialen, kulturellen und ökonomischen Kapitals (vgl. insb. Bourdieu, 1977; Bourdieu, 1983; Coleman, 1988) der Schülerinnen und Schüler international eine zentrale Rolle. Diese soll in diesem Beitrag im Fokus stehen. Eine nähere Betrachtung von Untersuchungen, die sich der Bedeutung und Wirkung von Faktoren auf Ebene der (nationalen) Bildungspolitik bzw. -administration, der Schulen bzw. des Unterrichts und den damit im Zusammenhang stehenden Tätigkeiten von Lehrkräften im internationalen Kontext widmen, muss anderen Beiträgen vorbehalten bleiben, da der Rahmen dieses Beitrags sonst überschritten würde.

Treibende Kraft bei der Analyse von Bildungssystemen ist die stetige Auseinandersetzung mit Fragen zur sozialen Ungleichheit, welche sich kontinuierlich durch das sich wandelnde Verständnis von Unterschiedlichkeit – also Heterogenität – in Schule und Unterricht aktualisiert (vgl. Schwippert, 2011). Das sich ändernde Verständnis ist nicht nur auf sich gesellschaftlich wandelnde Anforderungen (z. B. Gleichheit von Jungen und Mädchen; Inklusion von Schülerinnen und Schülern mit Beeinträchtigungen im Regelschulsystem) zurückzuführen, sondern auch auf veränderte Aufgaben, die als wandelnde Rahmenbedingungen im zeitlichen Verlauf unterschiedliche Herausforderungen induzieren (z. B. Berücksichtigung der Kinder von angeworbenen ausländischen Arbeitnehmern; fluchtbedingter Migration; Konzentration der Bevölkerung in Ballungsräumen).

Als übliches Repertoire zur Erfassung individueller Heterogenität bei Schülerinnen und Schülern werden neben dem Geschlecht und Migrationserfahrungen in der Familie insbesondere auch bildungsrelevante Hintergrundmerkmale erhoben. Hierzu zählen der Bildungshintergrund der Eltern, die Ausstattung mit bildungsrelevanten Gütern und die Ausstattung mit Gütern, die von dem ausgeübten Beruf der Eltern und der hierdurch induzierten ökonomischen Situation abhängen.

Bereits in der internationalen Reading Literacy Study der IEA, deren Haupterhebung ab 1990 stattfand, wurde die Frage nach dem heimischen Buchbesitz als eine der zentralen und erklärungskräftigsten Indikatoren für das in der Familie vorhandene kulturelle und (auch) ökonomische Kapital erfragt (Elley, 1994; Lehmann, Peek, Pieper & Stritzky, 1991). Der besondere Charme dieses

Indikators liegt darin begründet, dass er von Schülerinnen und Schülern ganz unterschiedlicher Klassenstufen erfragt werden kann, im Gegensatz zu Variablen wie Bildungsabschlüsse und Berufe der Eltern, die von diesen erfragt werden müssen. Im internationalen Vergleich kann damit sowohl in Grundschulen als auch im Sekundarbereich gezeigt werden, wie stark in den verschiedenen beteiligten Staaten die Testergebnisse vom kulturellen bzw. ökonomischen Status der Familien abhängen – und auch in Deutschland erweist sich diese Variable als bedeutsam bei der Untersuchung von Unterschieden zwischen Schülerinnen und Schülern (Stubbe & Goy, 2013). So zeigt sich in Deutschland in der seit 2001 alle fünf Jahre durchgeführten IGLU-Studie und der seit 2007 im vierjährigen Turnus durchgeführten Trends in International Mathematics and Science Study (TIMSS), dass die durch den Buchindikator abgebildete Bildungsungleichheit in Grundschulen, wenn überhaupt nur in wenigen anderen Staaten größer ausfällt als in Deutschland (Hußmann, Stubbe & Kasper, 2017; Stubbe, Schwippert & Wendt, 2016).

1.4 Herausforderung internationaler Vergleich und Dokumentation von Entwicklungen

He und van de Vijver (2013b) weisen darauf hin, dass eine zu stark datenbezogene technischen Interpretation von Skalen insbesondere im Kontext international vergleichender oder auch längsschnittlich angelegter empirischer Bildungsstudien zu kurz greifen kann. Sie mahnen eine umsichtige Konstruktion von Fragen und Skalen an, die in Studien eingesetzt werden, wenn Staaten mit unterschiedlichen Kulturen im Bildungssystem gemeinsam in den Blick genommen werden sollen (so wie es in internationalen Vergleichsstudien üblich ist). He und van de Vijver (2013a) verweisen darüber hinaus darauf, dass neben der Notwendigkeit einer genauen Analyse der Items und Skalen (z.B. durch DIF-Modelle (differential item functioning), konfirmatorischen Faktoranalysen oder anhand von Mehrebenenanalysen) auch auf eine standardisierte und international einheitliche Testdurchführung zu achten ist – wie dies etwa bei den IEA-Studien wie TIMSS und PIRLS üblich ist. Darüber hinaus empfehlen sie eine Dekomposition möglicher Quellen für beobachtete Unterschiede (Varianzen), die möglicherweise auf die Konstruktion, die Auswertungsmethode oder auf Einzelitems zurückgeführt werden können. Sie weisen darauf hin, dass die Identifikation unterschiedlicher Varianzquellen notwendig ist, um so Effekte, die auf Individuen, Gruppen oder Staaten zurückzuführen sind, unterscheiden zu können. Einige dieser Unterschiede könnten dann durch kulturvergleichende Analysen erklärt werden, andere wiederum könnten auf tatsächliche (objektive) Unterschiede zwischen den Vergleichsgruppen zurückführbar sein. He und van de Vijver (2013a) haben in ihren Untersuchungen gezeigt, dass bei empirischen Vergleichsuntersuchungen kulturbedingte Hauptfaktoren bei Likert-Skalen identifizierbar waren, welche (unter Berücksichtigung der Überprüfung von Messinvarianz) für kulturvergleichende Analysen genutzt werden konnten. Bei vertiefenden Analysen von internatio-

nal eingesetzten Skalen erscheint es daher notwendig, bei der Interpretation der Befunde auch auf mögliche weitere – bisher nicht explizit modellierte oder beschriebene latente Konstrukte – zu achten und diese auch bei der Interpretation zu berücksichtigen.

Neben der Herausforderung bei der Identifizierung von Indikatoren bzw. Skalen zum internationalen Vergleich von Staaten in groß angelegten Vergleichsuntersuchungen ergibt sich aufgrund des Designs der Studien als Trend-Untersuchungen eine weitere Herausforderung. Seit 2001 wird PIRLS (mit Deutscher Beteiligung seit 2001)² und seit 1995 TIMSS (mit deutscher Beteiligung in der Grundschule seit 2007)³ in wiederkehrenden Zyklen durchgeführt. Wenn Vergleiche einerseits zwischen den Staaten aber auch im zeitlichen Verlauf (insb. innerhalb der Staaten) betrachtet werden sollen, sind diese Indikatoren und Skalen gleichzuhalten. Um mit den Studien jedoch sich wandelnde Bedingungen in dynamischen Schulsystemen angemessen zu erfassen, muss hierbei der Gestaltung und Weiterentwicklung der Hintergrundfragebögen eine besondere Aufmerksamkeit zukommen. In diesem Zusammenhang stellt sich gerade heute die Frage, welche Bedeutung Büchern – in diesem Kontext als einem der zentralen Indikatoren für die Dokumentation von Bildungsungleichheiten – in einer sich medial rasant wandelnden Welt für das Lehren und Lernen von Schülerinnen und Schülern zukommt und welche Möglichkeiten bei der Modellierung bestehen, wenn Veränderungen der Indikatoren festgestellt werden.

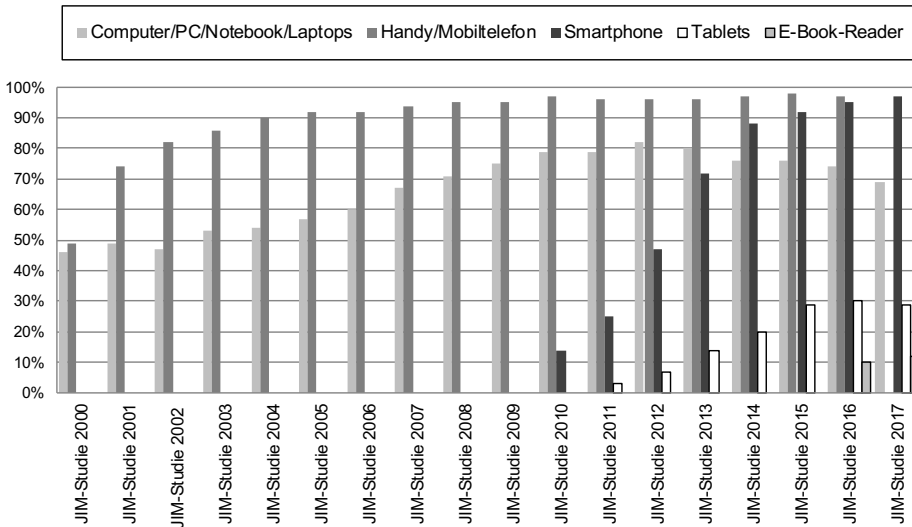
2. Ableitung der Fragestellung und empirische Grundlage

Wie zuvor skizziert nimmt seit Jahren die standardmäßig abgefragte Zahl des heimischen Buchbesitzes eine zentrale Rolle in internationalen Vergleichsuntersuchungen ein. Bei der Interpretation multivariater Zusammenhänge offenbaren sich vielschichtige Bezüge. Neben der numerischen Information über den Umfang des Buchbestandes umfasst die Information über den Buchbesitz auch Merkmale des kulturellen, des sozialen und auch des ökonomischen Kapitals einer Familie (Schulte, Hartig & Pietsch, 2014). Sie steht darüber hinaus auch als Indikator für die Anwendung von Kulturtechniken. Somit steht der Indikator für den heimischen Buchbesitz auch für eine bildungsorientierte Haltung.

2 PIRLS/IGLU (Bos et al., 2007; Bos et al., 2003; Bos, Tarelli, Bremerich-Vos & Schwippert, 2012; Hußmann, Wendt, Bos et al., 2017).

3 TIMSS (Bos et al., 2008; Bos, Wendt, Köller & Selter, 2012; Wendt et al., 2016).

Abbildung 1: Aufstellung zum Wandel der Verfügbarkeit digitaler Medien (eigene Darstellung auf Grundlage der JIM-Ergebnisse von 2000 bis 2017)



Anmerkung. Die Information zum E-Book-Reader wurde erst ab 2016 bei der Befragung berücksichtigt.

Vielfach ist belegt, dass der Indikator des heimischen Buchbesitzes einen überraschend hohen Anteil an Leistungsvarianz auf Individualebene erklärt (z. B. Schwippert, Wendt & Tarelli, 2012) – betrachtet man die Entwicklung der Verbreitung digitaler Medien, erscheint es aus heutiger Sicht nicht mehr reversibel, dass die bisher durch den Buchbesitz erfasste Information über die (zeitgemäße) Anwendung von Basiskompetenzen durch die Betrachtung der Nutzung von digitalen Medien zunächst ergänzt und langfristig möglicherweise auch abgelöst wird. Abbildung 1 aus der Zusammenfassung der von 2000 bis 2017 durchgeführten „Jugend, Information, (Multi-)Media“ (JIM) Studien (mpfs, n.d.) ist zu entnehmen, dass bis 2012 ein kontinuierlicher Zuwachs an Computern zu verzeichnen ist, der ab 2010 durch den Aufwuchs von Smartphones bzw. ab 2011 durch Tablet-Computer ergänzt wird. Diese Medien haben – neben Büchern – eine wichtige Rolle bei der Informationsbeschaffung in Lernkontexten übernommen. Explizit ist ab 2016 auch die Nutzung von E-Book-Readern aufgeführt, deren Verbreitung und Nutzung sicherlich mit einer Abnahme der Anzahl von Printmedien in Haushalten einhergehen wird. Damit ist zu vermuten, dass der Nutzung von digitalen Medien zukünftig eine entsprechende Erklärungskraft bei der Analyse von (Bildungs-) Disparitäten zukommen wird, wie heute noch dem Buchbesitz.

Um zu untersuchen, inwieweit sich die Bedeutung des heimischen Buchbesitzes im Kontext der Erfassung des kulturellen und ökonomischen Kapitals von Familien in Deutschland in den letzten 15 Jahren (zw. 2001 und 2016) verändert hat, wird dessen Eigenschaft im Kontext der Modellierung des latenten Konstrukts des kulturellen und ökonomischen Kapitals anhand der vier Erhebungszyklen von PIRLS

in Deutschland untersucht. Somit wird zunächst *ex post* der Frage nachgegangen, wie sich der Status des heimischen Buchbesitzes empirisch abbildet, um anschließend *ex ante* interpretierend der Frage „Was wird aus den Büchern?“ nachzugehen.

Wenn Skalen in verschiedenen Zyklen einer Untersuchung in den Hintergrundfragebögen wiederholt aufgenommen wurden, bieten probabilistische Testmodelle (item response theory – IRT) gute Möglichkeiten, diese einerseits über Staaten hinweg und andererseits auch im Laufe der Zyklen zu modellieren und diese damit sowohl international als auch im Längsschnitt als vergleichbare Kennwerte verfügbar zu machen. Wie bei Mullis, Martin und Hooper (2016) beschrieben, wurde auch im Kontext der TIMSS- und PIRLS-Untersuchungen IRT-Modelle zur Konstruktion von Skalen eingesetzt.

Geeignete IRT-Modelle bieten die Möglichkeit, auch bei Variation von einzelnen Items zwischen den Staaten oder über die Studienzyklen hinweg, diese bei der Berechnung der Gesamtskalen berücksichtigen zu können (*missing by design*). Anhand der Analysen kann dann festgestellt werden, ob die in den Skalen berücksichtigten Items in der erwarteten Relation zu den anderen erfassten Variablen stehen und damit in der Zusammenfassung das theoretisch abgeleitete latente Konstrukt empirisch angemessen erfasst wird. Diese Analysen der Skalen bietet die Möglichkeit, Abweichungen festzustellen, die in den Testeigenschaften der Items liegen. Werden solche Abweichungen (*bias*) identifiziert, können diese Items entweder aus den Modellen entfernt werden oder – wenn möglich – durch angepasste IRT-Modelle weiterhin berücksichtigt werden. Je nach Ausprägung und Eigenschaft der Abweichung kann dies auch auf eine notwendige Modifizierung der Befragungsinstrumente für zukünftige Erhebungen hinweisen. Um in jedem der realisierten Studienzyklen angemessene und reliable Antworten zu erhalten, bedeuten diese Veränderungen in den Messmodellen bzw. den Skalen jedoch auch, dass ein direkter Vergleich über die Studienzyklen hinweg nicht mehr zulässig wäre. Wird auf die Veränderung der Skaleneigenschaften bei der Interpretation der Befunde keine Rücksicht genommen, kann dies zu Fehlinterpretationen bei der Bewertung von Unterschieden führen. Darüber hinaus stellen Skalen, die nur auf einer geringen Anzahl von Items basieren, eine besondere Herausforderung dar. Gustafsson und Rosén (2014) weisen darauf hin, dass dies insbesondere für Skalen gilt, die im Längsschnitt einen Wandel erfahren haben und nur wenige Ankeritems zur Verbindung der verschiedenen Studienzyklen aufweisen. In ihrer Schlussfolgerung weisen Gustafsson und Rosén (2014) daher darauf hin, dass insbesondere der Auswahl der Ankeritems besondere Beachtung zukommen muss. Die Büchervariable ist bei den nachfolgend vorgestellten Analysen eine dieser Ankeraufgaben, sodass die Betrachtung ihrer psychometrischen Eigenschaften im Fokus liegen wird.

Für die Analysen werden die deutschen Datensätze der IEA PIRLS-Untersuchung (in Deutschland als Internationale Grundschul-Lese-Untersuchung – IGLU bekannt) von 2001 (n = 7633), 2006 (n = 7899), 2011 (n = 4000) und 2016

(n = 3959) genutzt.⁴ Die Schülerinnen und Schüler wurden im Rahmen einer systematischen Stichprobenziehung so bestimmt, dass die Stichprobe als repräsentativ für Deutschland gelten kann (für technische Details zur Stichprobenziehung siehe u.a. LaRoche, Joncas & Foy, 2017). Im Rahmen der IGLU-Studie wurden in Deutschland Viertklässlerinnen und Viertklässler am Ende des Schuljahrs (jeweils im Frühjahr des angegebenen Jahrs der Haupterhebung) befragt. In allen Zyklen wurden den Kindern insgesamt zwei Texte im Umfang von ca. 4 Seiten (abhängig von der Bebilderung der Texte) vorgelegt, zu denen im Anschluss 12 bis 17 Fragen gestellt wurden (Hußmann, Wendt, Kasper, Bos & Goy, 2017). Im Anschluss wurde den Kindern nach der Testung ein Fragebogen zur Erfassung verschiedener Hintergrundinformationen vorgelegt. Testungen und Befragungen wurden durch kindgerechte Pausen unterbrochen. In PIRLS/IGLU wurden zusätzlich auch die Eltern der beteiligten Viertklässlerinnen und Viertklässler befragt, sodass diese Informationen als zusätzliche Hintergrundinformationen über die Kinder zur Verfügung standen.

In allen Zyklen liegen Informationen vor über (1) den heimischen Buchbesitz (5-stufig), (2) den höchsten Bildungsabschluss der Eltern (3-stufig; aus Elternbefragung) und die Information (dichotom), ob (3) zu Hause immer Deutsch gesprochen wird, (4) es einen Computer gibt, (5) das Kind einen eigenen Schreibtisch hat, (6) es eine Tageszeitung und (7) es einen Rasenmäher im Haushalt gibt und ob (8) das Kind ein eigenes Zimmer hat. Damit liegen über alle vier Studienzyklen acht Ankeraufgaben vor. Weitere Indikatoren zur ökonomischen und kulturellen Situation sind z.T. einmalig in einzelnen Zyklen, aber auch wiederholt erfragt worden. Tabelle 1 gibt Auskunft über die Verfügbarkeit der Merkmale für die vier Studienzyklen.

Aus den Hintergrundfragebögen der Schülerinnen und Schüler und ihrer Eltern werden die Informationen für die in diesem Beitrag im Fokus stehenden Analysen (unter Berücksichtigung des *sampling-weights* HOUWGT (Student House Weight)) genutzt. Hierzu werden die durch die Befragungen ermittelten Einzelinformationen zum ökonomischen und kulturellen Hintergrund der Kinder bzw. deren Familien herangezogen, um hieraus einen kombinierten Kennwert zu bilden. Als bewährtes und robustes IRT-Modell wird hierzu das ein-parametrische logistische Modell (Rasch-Modell) in drei Variationen herangezogen. Als Kriterium zur Berücksichtigung der Hintergrundmerkmale (Items) in den Modellen wurde der gewichtete Fit (MNSQ) herangezogen. Bei allen nachfolgend beschriebenen Modellen weisen die berücksichtigten Items⁵ einen sehr guten Fit auf ($0.9 < \text{MNSQ} < 1.1$). Zur Berechnung ist das Programm ConQuest genutzt worden (Wu, Adams, Wilson & Haldane, 2007).

4 Die Daten sind am 1.4.2018 vom PIRLS-Server am Boston College herunter geladen worden [www.pirls.org].

5 Lediglich das Item zur Verfügbarkeit des Mobiltelefons (2006 und 2011) wurde aufgrund schlechter Fit-Werte von den Analysen ausgeschlossen.

Tabelle 1: Erfragte Merkmale der Haushalte in Deutschland in PIRLS 2001, 2006, 2011 und 2016 im Überblick

	2001	2006	2011	2016
Bücher im Haushalt	*	*	*	*
Höchster Bildungsabschluss Eltern	*	*	*	*
Sprachgebrauch	*	*	*	*
Computer	*	*	*	*
Eigener Schreibtisch	*	*	*	*
Tageszeitung	*	*	*	*
Rasenmäher	*	*	*	*
Persönliche Bücher	*	*	*	*
Geschirrspülmaschine	*			
Eigenes Zimmer	*	*	*	*
Computer-Lernprogramme	*	*	*	
Internetzugang	*		*	*
Wörterbuch	*			
Kunstgegenstände (Bilder)	*			
Zweites Auto	*	*	*	
Klavier		*	*	*
Haushaltshilfe				*

Anmerkung. 2006 und 2011 wurde auch nach einem eigenen Mobiltelefon gefragt. Dieses Item wurde aufgrund schlechter FIT-Werte nicht berücksichtigt.

- I) Gemeinsame Modellierung der Hintergrundvariablen mit einheitlichen Itemparametern in allen vier Erhebungszyklen. Berücksichtigung der Itemschwierigkeit bei dichotomen Items und Modellierung von polytomen Items durch Bestimmung der Übergangsschwellen zwischen den Kategorienfunktionen (vgl. Rost, 2004).
(ConQuest Modell: $\text{item} + \text{item}^*\text{step}$).
- II) Gemeinsame Modellierung aller erfragten Hintergrundvariablen mit Ermittlung der generellen Itemschwierigkeiten und der Unterschiede zwischen den Erhebungszyklen. Zusätzlich Bestimmung zyklenspezifischer Schwellenparameter.
(ConQuest Modell: $\text{item} + \text{ncycle} + \text{item}^*\text{step} + \text{item}^*\text{step}^*\text{ncycle}$).

III) Gemeinsame Modellierung ausschließlich der Hintergrundvariablen, die in allen Erhebungszyklen gefragt wurden. Die polytomen Items (Bücher und höchster Bildungsabschluss der Eltern) wurden zur Verbesserung der Fit-Werte dichotomisiert;⁶ anschließend wurden die Itemschwierigkeiten, die Zykleneffekte und die zyklenspezifischen Itemparameter modelliert. (ConQuest Modell: $\text{item} + \text{ncycle} + \text{item} * \text{ncycle}$).

3. Ergebnisse

Model I: Prüfung allgemeiner Itemeigenschaften unter Berücksichtigung der Daten aus vier Zyklen: Im ersten Analyselauf wurden die Itemparameter aller berücksichtigten Variablen bestimmt. Tabelle 2 sind die Itemschwierigkeiten zu entnehmen. Für die polytomen Items zum heimischen Buchbesitz (ASBGBOOK) und des höchsten Bildungsabschlusses der Eltern (H_ISCED_R) wurden zusätzlich die Schwellenparameter berechnet. In Abbildung 2 sind für alle dichotomen Items die Itemschwierigkeiten und für die polytomen Items die Schwellenparameter gemäß Thurstone (Wu et al., 2007) abgetragen. Dieser ist zu entnehmen, dass sich die vier Schwellen der Büchervariablen (1.1 bis 1.4) gleichmäßig in der Verteilung der Itemschwierigkeiten verteilen. Insgesamt zeigt sich eine gute Streuung der Items im Vergleich zur Verteilung der Personenparameter, auch wenn eine etwas dichtere Indikatorisierung im mittleren Bereich wünschenswert wäre.

Model II: Differenzielle Schwellenparameter der Büchervariablen nach Zyklen: In einem weiteren Modell wird der generelle Unterschied der Itemschwierigkeiten für die vier Zyklen überprüft. Die Analysen (vgl. Tabelle 3) zeigen, leichte Verschiebungen der Itemparameter für die Erhebungen von 2001, 2006, 2011 und 2016 – diese sind jedoch numerisch als sehr gering anzusehen (auch wenn sie aufgrund der sehr großen Stichprobenumfänge als signifikant gekennzeichnet sind). Diese Unterschiede könnten auch auf Unterschiede in den Schülerinnen- und Schülerstichproben in den verschiedenen Zyklen zurückzuführen sein – eine Gegenüberstellung der mittleren Personenparameter für die Zyklen (2001 = 1,068; 2006 = 1,053; 2011 = 1,054; 2016 = 1,066) zeigt jedoch nur marginale Unterschiede, die deutlich unter den festgestellten Unterschieden in den mittleren Itemschwierigkeiten der vier Zyklen (vgl. Tabelle 4) liegen.

6 Dichotomisierung: Buchbesitz 0 = bis 100 Bücher vs. 1 = mehr als 100 Bücher; höchster Bildungsabschluss 0 = bis Universitätsbesuch ohne BA Abschluss vs. 1 = Universitätsabschluss BA und höher.

Tabelle 2: Itemschwierigkeiten und Fit-Werte der Skalierung der Hintergrundmerkmale für die Deutschen PIRLS-Daten von 2001, 2006, 2011 und 2016

Item	Item-Schwierigkeit	S.E.	weighted MNSQ
Bücher im Haushalt	0,785	0,007	1,08
Sprachgebrach	-0,685	0,011	1,02
Computer	-1,172	0,011	0,96
Eigener Schreibtisch	-1,918	0,012	0,96
Persönliche Bücher	-2,282	0,012	0,95
Tageszeitung	0,586	0,010	0,96
Geschirrspülmaschine	0,072	0,012	0,94
Eigenes Zimmer	-0,731	0,011	0,96
Computer-Lernprogramme	0,097	0,010	1,04
Internetzugang	0,276	0,011	1,06
Wörterbuch	-1,522	0,012	1,00
Kunstgegenstände (Bilder)	-0,487	0,012	1,03
Rasenmäher	0,384	0,010	0,91
Zweites Auto	1,079	0,010	0,98
Klavier	2,487	0,011	0,95
Haushaltshilfe	1,871	0,012	1,07
Höchster Bildungsabschluss Eltern	1,160	0,043	1,02

Abbildung 2: Stem-and-leaf Darstellung von Personenparametern und Itemschwierigkeiten für die Skalierung der Hintergrundmerkmale für die Deutschen PIRLS-Daten von 2001, 2006, 2011 und 2016

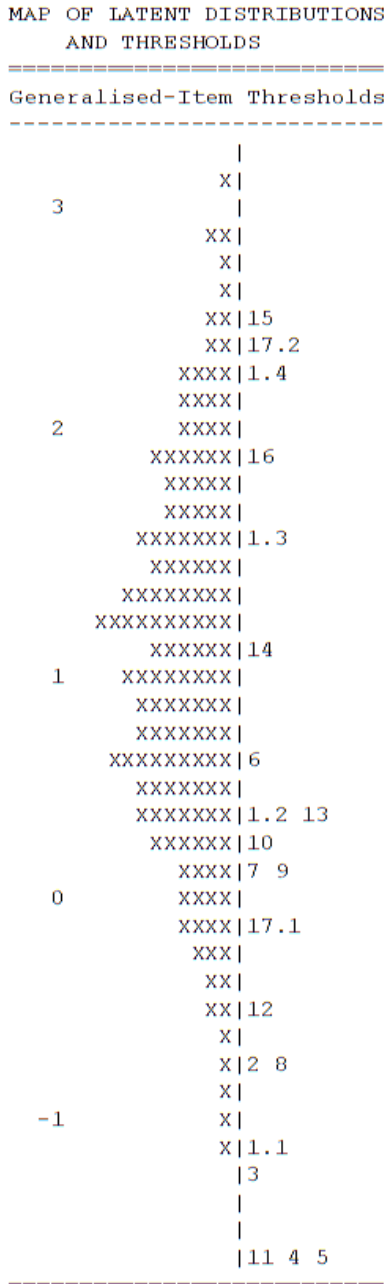
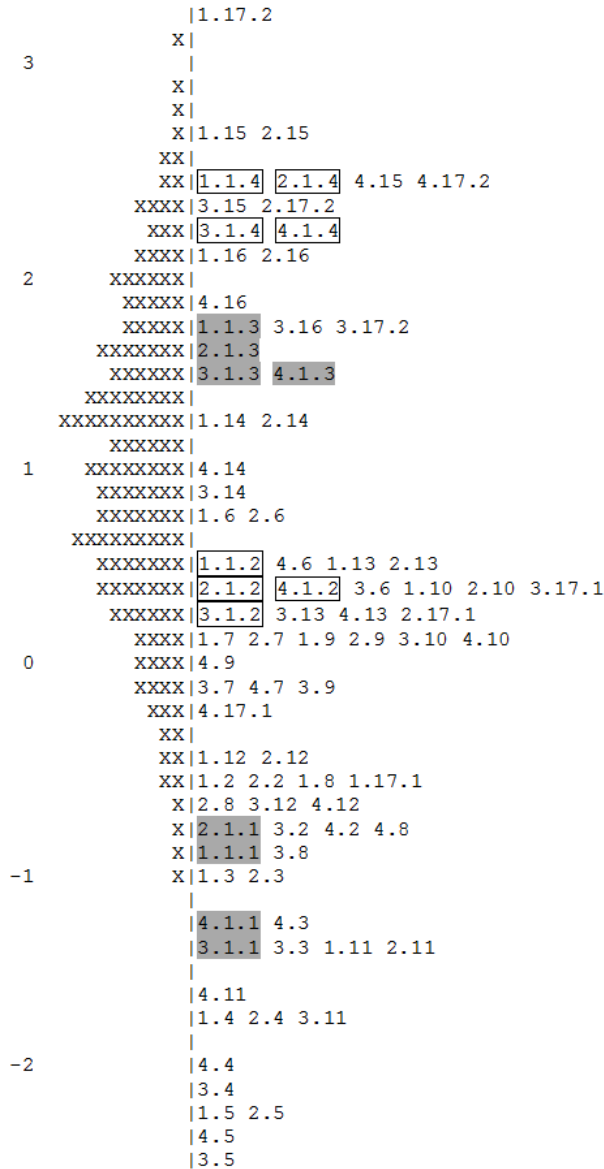


Tabelle 3: Genereller Zyklen-Effekt bei der Modellierung der Itemschwierigkeiten für die Skalierung der Hintergrundmerkmale für die Deutschen PIRLS-Daten von 2001, 2006, 2011 und 2016

Zyklus	Item-Schwierigkeit	S.E.	weighted MNSQ
1 - PIRLS 2001	0,125	0,006	1,02
2 - PIRLS 2006	0,119	0,006	0,97
3 - PIRLS 2011	-0,156	0,007	0,98
4 - PIRLS 2016	-0,088	0,011	0,95

In Abbildung 3 sind die zyklenspezifischen Itemschwierigkeiten der Büchervariable (Thurstonian thresholds) herausgehoben, die sich als nur wenig variierend erweisen. Lediglich die Parameter für die erste Schwelle (x.1.1) zeigen sich etwas stärker gespreizt als für die Schwellen 2, 3 und 4. Da ConQuest die zyklenspezifischen Itemschwierigkeiten (Item*Zyklus) aufgrund systematisch fehlender Werte für bestimmte Variablen (siehe Tabelle 1 – Items, die in verschiedenen Zyklen nicht erfragt wurden) nicht ausgibt, ist abschließend noch ein Modell vorgestellt, welches ausschließlich auf den Variablen basiert, die in allen vier Zyklen erhoben wurde.

Abbildung 3: Verortung der Schwellenparameter (stem-and-leaf plot) für den heimischen Buchbesitz in den Deutschen PIRLS-Daten von 2001, 2006, 2011 und 2016



Model III: Vollständiges DIF-Modell für in allen Zyklen erfragte Variablen: In allen vier PIRLS-Zyklen wurden in Deutschland acht Variablen einheitlich erfasst. Basierend auf diesen Variablen werden im Model III die generelle Itemschwierigkeit, die Unterschiede in den Itemschwierigkeiten in den vier Zyklen und die darüber hinausgehenden zyklenspezifischen Item-Variationen (DIF) berechnet (ConQuest Modell: $\text{item} + \text{ncycle} + \text{item} * \text{ncycle}$). Als Ergebnis der Berechnung bietet ConQuest eine Dekomposition der drei Itemparameter-Komponenten. Durch die Zusammenführung der drei Komponenten lassen sich die Itemschwierigkeitswerte je Item und Zyklus berechnen. Diese item- und zyklusspezifische Zusammenführung der von ConQuest ausgegebenen Parameter ist in Abbildung 4 visualisiert.

Abbildung 4: Zyklenspezifische Itemschwierigkeiten für die Merkmale, die in den Erhebungszyklen 2001, 2006, 2011 und 2016 erfragt wurden

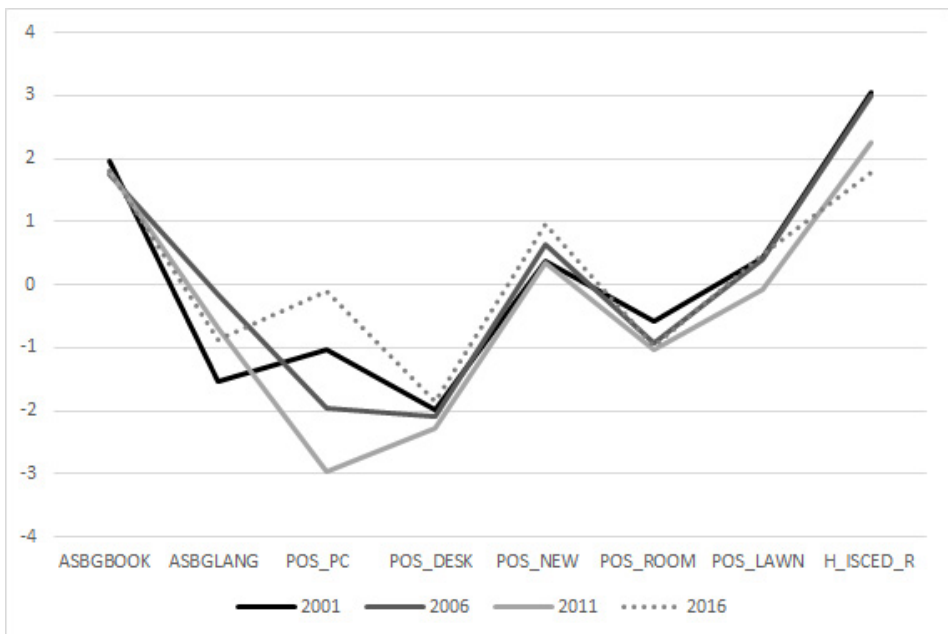


Abbildung 4 ist zu entnehmen, dass die zyklenspezifische Variation der Itemschwierigkeit bei der Büchervariable (ASBGBOOK) besonders gering ausfällt. Etwas größere Schwankungen zeigen sich bei der Information nach dem eigenen Schreibtisch (_DESK), der Verfügbarkeit einer Tageszeitung (_NEW), des eigenen Zimmers (_ROOM) und auch beim Rasenmäher (_LAWN). Etwas stärkere Unterschiede zeigen sich beim heimischen Sprachgebrauch (ASBGLANG) und dem höchsten Bildungsabschluss der Eltern (H_ISCED_R). Eine auffallend hohe Spreizung weist in dem Modell der Parameter für die Information über die Verfügbarkeit des Computers (_PC) in den verschiedenen Studienzyklen auf. Mit diesen Befunden kann gezeigt werden, dass die insbesondere in der Diskussion ste-

hende Büchervariable bei den vorliegenden Studienzyklen als stabiler Indikator und Ankeritem bei der Modellierung von kulturellem und ökonomischem Hintergrund genutzt werden kann. Allerdings verweist die starke Schwankung und damit die variierende Bedeutung des Computerbesitzes auf die Sensitivität des Konstrukts, wenn die inhaltliche Bedeutung der einzelnen Items über die Zeit variiert.⁷ Die erwarteten Wechselwirkungen zwischen Buchbesitz und Computer (inkl. Tablet, Smartphone und ggf. E-Book-Reader) verweisen auf die Notwendigkeit, die erwarteten Bedeutungsverschiebungen empirisch im Blick zu behalten.

Vergleich der Modelle: Zum Vergleich der Modelle, die auf der gleichen Personen- und Itemauswahl basieren, wurden Modellanpassungstests (goodness of fit) gerechnet. Tabelle 4 enthält die entsprechenden Werte für die drei Modelle.

Tabelle 4: Modellanpassung für die Bestimmung der zyklusunspezifischen und zyklusspezifischen Skalierungsläufe

	Modell I	Modell II	Modell III
AIC	326145,73	324821,99	166787,11
BIC	326323,14	325152,63	167053,23
CAIC	326345,14	325193,63	167086,23

Die durchweg niedrigeren Werte der drei Indices im Modell III (vgl. Tabelle 4) verweisen auf die beste Anpassung des sparsamsten Modells – während im direkten Vergleich der Modelle I und II, in denen alle verfügbaren Ressourcen als Items aufgenommen waren, die differenzierte Modellierung der Itemschwierigkeiten nach den Erhebungszyklen eine bessere Passung für das Modell II aufweist.

4. Diskussion

Bei der Bestimmung der Itemschwierigkeitsparameter für die als polytom kodierete Büchervariable zeigen sich zwar numerische (und aufgrund der sehr großen Stichproben auch signifikante) Unterschiede zwischen den Erhebungszyklen, die jedoch inhaltlich als wenig bedeutsam interpretiert werden können. Somit konnte gezeigt werden, dass die verschiedenen Abstufungen der Büchervariablen im Verlauf der letzten 15 Jahre praktisch keine Veränderung erfahren haben. Lediglich für die unterste Schwelle – am Übergang von 10 und weniger zu mehr als 10 Büchern – zeigt sich eine gewisse Spreizung über die Jahre. Hierbei ist jedoch zu beachten, dass die unterste Kategorie von unter 10 % der Schülerinnen und Schüler angegeben wird und entsprechend die für die verschiedenen Studienzyklen ermittelten Schwellen nicht nur eine Folge der Messeigenschaften des Items sind, sondern

⁷ Während sich 2001, 2006 und 2011 generell nach der Verfügbarkeit eines Computers erkundigt wurde, ist 2016 nach dem Besitz eines eigenen Computers gefragt worden.

auch von einer sich wandelnden Zusammensetzung der Schülerschaft abhängt (vgl. hierzu Hußmann, Stubbe et al., 2017).

Betrachtet man die Skalen insgesamt und damit auch die Komposition der berücksichtigten Indikatoren zum ökonomischen und kulturellen Status, erweisen sich diese als gut geeignet, um Unterschiede zwischen den Schülerinnen und Schülern in den jeweiligen Erhebungszyklen aber auch über diese Zyklen im zeitlichen Verlauf hinweg zu beschreiben. Der Modellvergleich zeigte eine besonders gute Passung des sparsamsten Modells an. Eine Entscheidung hierfür wäre aber eine empiristische Entscheidung, die außer Acht lässt, dass sich durch das Auslassen von Items auch die Interpretation des modellierten latenten Konstrukts ökonomischen und kulturellen Kapitals inhaltlich verändert und in diesem Fall auch einengt. Gerade wenn von einer sich über die Zeit verändernden Gesellschaft mit wandelnden Entscheidungsmöglichkeiten betreffend der individuellen Ausstattung mit Bildungsgütern ausgegangen wird, erscheint eine Fokussierung nur auf die gemeinsamen (Anker-)Items inhaltlich nicht angemessen. Vielmehr sollte die im Rahmen der Skalierung zulässige Variation von Items über die Zeit genutzt werden und gleichzeitig die mögliche zeitlich bedingte Interpretationsvalidität der Skalen in den Fokus rücken (Kane, 2013). Betreffend der Frage „Was wird aus den Büchern?“ kann zusammenfassend festgestellt werden, dass für den beobachteten Zeitraum von 2001 bis 2016 eine Betrachtung der Bildungsungleichheit auch mit der Information über den heimischen Buchbestand angemessen und gerechtfertigt erscheint. Somit kann die häufig geäußerte Frage, ob die Büchervariable zur Beschreibung des sozialen Hintergrunds sich schon gewandelt hätte, verneint werden. Allerdings zeigt die Berücksichtigung von Items, wie zum Beispiel die Frage nach dem heimischen Computer (dessen Bedeutsamkeit im zeitlichen Verlauf eine Veränderung erfahren hat), dass eine besondere Berücksichtigung bei der Modellierung angezeigt ist. Um diese Variable, die zu den jeweiligen Messzeitpunkten gute Messeigenschaften aufweist, nicht aus dem Modell entfernen zu müssen, wäre eine Alternative, diese Variable nicht übergreifend als Ankeritem zu verwenden, sondern jeweils als zyklusspezifisches Item (messzeitpunktspezifisch) zu berücksichtigen. So würde eine sich nachzeichenbare Veränderung des Items insbesondere als ökonomisches Merkmal im Modell Berücksichtigung finden. Gerade vor dem Hintergrund der im Beitrag skizzierten Veränderung in der Verfügbarkeit von digitalen Medien, die – wird die bisherige Entwicklung weiter in die Zukunft projiziert – zunehmend Printmedien substituieren könnten, ist abzusehen, dass auch mit der Information über den heimischen Buchbesitz zukünftig entsprechend umgegangen werden muss. Daher wird bei den Trendstudien PIRLS 2021 und TIMSS 2019 diese sich wandelnde Funktion der Büchervariablen zur Beschreibung des sozialen Hintergrunds zu überprüfen und in der Befunddarstellung zu reflektieren sein. Gleichzeitig ist zu eruieren, welche alternativen bildungsrelevanten Güter zukünftig in den Blick genommen werden müssen, um eine angemessene Darstellung von Unterschieden im sozialen Hintergrund realisieren zu können. Bis dato kann jedoch festgestellt werden, dass

die Information über den heimischen Buchbesitz als Ankeritem in der Betrachtung des sozialen Hintergrunds im Zeitraum von 2001 bis 2016 genutzt werden kann.

4.1 Ausblick

Sicherlich wird der heimische Buchbesitz weiterhin eine erklärungskräftige Variable von Bildungsdisparitäten sein –, aber die inhaltliche Interpretation des Zusammenhangs und damit der Validität wird sich nachhaltig durch die Verfügbarkeit und Nutzung digitaler Medien wandeln. Aus theoretischer Perspektive (vgl. Bourdieu, 1983) verrät die Büchervariable sowohl etwas über das kulturelle als auch das ökonomische Kapital von Familien. Empirisch konnten diese vielschichtigen Zusammenhänge nachgewiesen werden (vgl. u. a. Ehmke & Jude, 2010; Stubbe et al., 2016). Der heimische Buchbesitz stellt somit bisher eine wichtige und vielschichtige Ressource dar, die jedoch zur Steuerung bzw. in diesem Zusammenhang zur Kompensation von Bildungsbenachteiligungen – wenn überhaupt – zunächst nur in sehr geringem Umfang genutzt werden kann. So wäre bei dem Versuch kompensatorisch zu wirken selbstverständlich, dass die isolierte Steigerung der Bücher in einem Haushalt (durch Geschenke bzw. Subventionen) nicht mit einer tatsächlichen Steigerung des kulturellen bzw. ökonomischen Kapitals (im Sinne Bourdieus) und in der Folge somit auch nicht mit einer Steigerung der in groß angelegten Schulvergleichsuntersuchung im Fokus stehenden Kompetenzen einhergeht. Die wirksame Größe ist und bleibt der bildungsrelevante Umgang mit Büchern bzw. zukünftig auch mit digitalen Medien, die zumindest vergleichbare Inhalte bieten. Die Verfügbarkeit von digitalen Medien und die Kompetenz, diese nutzen zu können, stellen bereits heute einen zu berücksichtigenden und in absehbarer Zukunft (von 5 bis 10 Jahren) sicherlich zunehmend wichtigeren Faktor im Bildungskontext dar. Jedoch ist wie bisher beim heimischen Buchbestand auch bei digitalen Medien nicht allein der Besitz der Faktor, der die Bildungsunterschiede bei den Schülerinnen und Schülern erklären kann (Lorenz & Gerick, 2014). Erst bei der Betrachtung des Zusammenspiels von Verfügbarkeit und Nutzung gewinnen wir Erkenntnisse darüber, wie die selbständige Gestaltung und Entwicklung des eigenen Handelns in der (informierten Wissens-)Gesellschaft wirken.

4.2 Grenzen der vorliegenden Betrachtungen

Die in dem Beitrag vorgelegten Betrachtungen fokussieren auf die praktische Umsetzung von Messmodellen in einem Bildungssystem. In Erweiterung internationaler Analysen zur Messinvarianz von Hintergrundmerkmalen (Wendt, Kasper & Trendtel, 2017), bei dem die Einheitlichkeit der Modellierung im internationalen Vergleich im Fokus stand, sind – vor dem Hintergrund der hier vorgestellten Befunde – in einem nächsten Schritt internationale Vergleiche vorzunehmen, bei denen soziale, ökonomische und kulturelle Merkmale zwischen den betrach-

teten Staaten variieren und somit auch als weitere Varianzkomponenten zu berücksichtigen sind. Bei einem solchen globalen Vergleich wäre zu berücksichtigen, dass die – auch in diesem Beitrag – herangezogenen empirisch vielfach replizierten Erklärungsmodelle ihren Ursprung in einer insbesondere nord-westlich geprägten Tradition haben. Es wäre somit zu prüfen, ob und wie entsprechenden, Modellierungen für andere Regionen der Welt anzupassen wären.

Literatur

- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.) (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K. & Valtin, R. (Hrsg.) (2007). *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. 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., Tarelli, I., Bremerich-Vos, A. & Schwippert, K. (Hrsg.) (2012). *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Wendt, H., Köller, O. & Selter, C. (Hrsg.) (2012). *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bourdieu, P. (1977). Cultural reproduction and social reproduction. In J. Karabel & A. H. Halsey (Hrsg.), *Power and ideology in education* (S. 487–511). New York, NY: Oxford University Press.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (Soziale Welt, Sonderbd. 2, S. 183–198). Göttingen: Schwartz.
- Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, 95–120.
- Ebbs, D. & Wry, E. (2017). Translation and layout verification for PIRLS 2016. In M. O. Martin, I. V. S. Mullis & M. Hooper (Hrsg.), *Methods and procedures in PIRLS 2016*. Zugriff am 04.06.2018 unter https://pirls.bc.edu/publications/pirls/2016-methods/P16_MP_Chap7_Translation_Layout_Verification.pdf
- Ehmke, T. & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 231–253). Münster: Waxmann.
- Elley, W. B. (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Exeter: Pergamon.
- Foshay, A. W. (1962). The background and the procedures of the twelve-country study. In A. W. Foshay, R. L. Thorndike, F. Hotyat, D. A. Pidgeon & D. A. Walker (Hrsg.), *Educational achievements of thirteen-year-olds in twelve countries. Results of an international research project, 1959–61* (S. 7–19). Hamburg: Unesco Institute for Education.
- Fraillon, J., Schulz, W. & Ainley, J. (2013). *International Computer and Information Literacy Study. Assessment framework*. Zugriff am 04.06.2018 unter https://research.acer.edu.au/ict_literacy/9/

- Goy, M., van Ackeren, I. & Schwippert, K. (2008). Ein halbes Jahrhundert internationale Schulleistungsstudien. Eine systematische Übersicht. *Tertium Comparationis*, 14(1), 77–107.
- Grønmo, L. S., Lindquist, M., Arora, A. & Mullis, I. V. S. (2013). TIMSS 2015 mathematics framework. In I. V. S. Mullis & M. O. Martin (Hrsg.), *TIMSS 2015 assessment frameworks* (S. 11–27). Zugriff am 03.02.2017 unter http://timssandpirls.bc.edu/timss2015/downloads/T15_Frameworks_Full_Book.pdf
- Gustafsson, J.-E. & Rosén, M. (2014). Quality and credibility of international studies. In R. Strietholt, W. Bos, J.-E. Gustafsson & M. Rosén (Hrsg.), *Educational policy evaluation through international comparative assessments* (S. 19–32). Münster: Waxmann.
- He, J. & van de Vijver, F. J. R. (2013a). A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Differences*, 55, 794–800.
- He, J. & van de Vijver, F. J. R. (2013b). Methodological issues in cross-cultural studies in educational psychology. In G. A. D. Liem & A. B. I. Bernado (Hrsg.), *Advancing cross-cultural perspectives on educational psychology* (S. 39–55). Charlotte, NC: Information Age Publishing Inc.
- Husén, T. (Hrsg.) (1967). *International study of achievement in mathematics. A comparison of twelve countries/international project for the evaluation of educational achievement (IEA)*. Stockholm: Almqvist & Wiksell.
- Husén, T. & Postlethwaite, T. N. (1996). A brief history of the International Association for the Evaluation of Educational Achievement (IEA). *Assessment in Education*, 3(2), 129–141.
- Hußmann, A., Stubbe, T. C. & Kasper, D. (2017). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195–217). Münster: Waxmann.
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M., McElvany, N., Stubbe, T. C. & Valtin, R. (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Hußmann, A., Wendt, H., Kasper, D., Bos, W. & Goy, M. (2017). Ziele, Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU 2016). In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 29–77). Münster: Waxmann.
- Jones, L. R., Wheeler, G. & Centurino, V. A. S. (2013). TIMSS 2015 science framework. In I. V. S. Mullis & M. O. Martin (Hrsg.), *TIMSS 2015 assessment frameworks* (S. 29–58). Zugriff am 03.02.2017 unter http://timssandpirls.bc.edu/timss2015/downloads/T15_Frameworks_Full_Book.pdf
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 55, 1–73. Zugriff am 03.02.2017 unter onlinelibrary.wiley.com/doi/10.1111/jedm.11200/epdf
- Klafki, W. (2007). *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik*. Weinheim: Beltz.
- Klieme, E. & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. *Zeitschrift für Erziehungswissenschaft, Sonderheft 8*, 11–29.
- LaRoche, S., Joncas, M. & Foy, P. (2017). Sample design in PIRLS 2016. In M. O. Martin, I. V. S. Mullis & M. Hooper (Hrsg.), *Methods and procedures in PIRLS*

2016. Zugriff am 04.06.2018 unter https://pirls.bc.edu/publications/pirls/2016-methods/P16_Methods_and_Procedures.pdf
- Lehmann, R. H., Peek, R., Pieper, I. & Stritzky, R. v. (1991). Lesefähigkeit und Lesegewohnheiten von Achtkläßlern. *Deutschunterricht*, 44(6), 410–417.
- Lorenz, R. & Gerick, J. (2014). Neue Technologien und die Leseleistung von Grundschulkindern. In B. Eickelmann, R. Lorenz, M. Vennemann, J. Gerick & W. Bos (Hrsg.), *Grundschule in der digitalen Gesellschaft. Befunde aus den Schulleistungstudien IGLU und TIMSS 2011* (S. 59–71). Münster: Waxmann.
- mpfs – Medienpädagogischer Forschungsverbund Südwest (n.d.). *Basisuntersuchungen zur Mediennutzung von Kindern und Jugendlichen. JIM-Studie*. Zugriff am 04.06.2018 unter <https://www.mpfs.de/studien/?tab=tab-18-1>
- Mullis, I. V. S., Martin, M. O. & Hooper, M. (2016). Measuring changing educational contexts in a changing world: Evolution of the TIMSS and PIRLS questionnaires. In M. Rosén, K. Y. Hansen & U. Wolff (Hrsg.), *Cognitive abilities and educational outcomes* (S. 207–222). Cham: Springer International Publishing.
- Mullis, I. V. S., Martin, M. O. & Sainsbury, M. (2015). PIRLS 2016 reading framework. In I. V. S. Mullis & M. O. Martin (Hrsg.), *PIRLS 2016 assessment framework* (S. 11–29). Zugriff am 03.02.2017 unter http://timss.bc.edu/pirls2016/downloads/P16_FW_Chap1.pdf
- Rost, J. (2004). *Testtheorie. Testkonstruktion*. Bern: Huber.
- Schulte, K., Hartig, J. & Pietsch, M. (2014). Der Sozialindex für Hamburger Schulen. In D. Fickermann & N. Maritzen (Hrsg.), *Grundlagen für eine daten- und theoriegestützte Schulentwicklung. Konzeption und Anspruch des Hamburger Instituts für Bildungsmonitoring und Qualitätsentwicklung (IfBQ)* (HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Bd. 13; S. 67–80). Münster: Waxmann.
- Schwippert, K. (2011). Von der Beschreibung zur Interpretation. Konzepte der Betrachtung von Chancengleichheit. In C. Tarnai (Hrsg.), *Sozialwissenschaftliche Forschung in Diskurs und Empirie* (S. 35–48). Münster: Waxmann.
- Schwippert, K. (2015). Wie viel Bildung steckt in groß angelegten Schulvergleichsuntersuchungen? In H.-J. Fischer, H. Giest & K. Michalik (Hrsg.), *Bildung im und durch Sachunterricht* (Probleme und Perspektiven des Sachunterrichts, Bd. 25; S. 39–50). Bad Heilbrunn: Klinkhardt.
- Schwippert, K., Wendt, H. & Tarelli, I. (2012). Lesekompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 191–207). Münster: Waxmann.
- Stubbe, T. C. & Goy, M. (2013). Besitzen wollen und sich leisten können. Analysen zur dimensional getrennten Erfassbarkeit von ökonomischem und kulturellem Kapital in Familien. In K. Schwippert, M. Bonsen & N. Berkemeyer (Hrsg.), *Schul- und Bildungsforschung. Diskussionen, Befunde und Perspektiven. Festschrift für Wilfried Bos* (S. 203–222). Münster: Waxmann.
- Stubbe, T. C., Schwippert, K. & Wendt, H. (2016). Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland und im internationalen Vergleich* (S. 229–316). Münster: Waxmann.
- Tenorth, H.-E. (2005). Grundbildung – Institutionelle Restriktion oder legitimes Programm? In M. Götz & K. Müller (Hrsg.), *Grundschule zwischen den Ansprüchen der Individualisierung und Standardisierung* (S. 17–30). Wiesbaden: Springer.
- Weinert, F. E. (1999). *Concepts of competence. Definition and selection of competencies*. Zugriff am 04.06.2018 unter <https://pdfs.semanticscholar.org/8b88/efa9d45e0a4b605aea6e5e3b9ec640bebo89.pdf>

- Wendt, H., Bos, W., Selter, C., Köller, O., Schwippert, K. & Kasper, D. (Hrsg.) (2016). *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Wendt, H., Kasper, D. & Trendtel, M. (2017). Assuming measurement invariance of background indicators in international comparative educational achievement studies: A challenge for the interpretation of achievement differences. *Large-scale Assessments in Education*, 5(10).
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest version 2.0. Generalised item response modelling software*. Camberwell: Australian Council for Educational Research.