



LifBi WORKING PAPERS

Michael Bayer und Sabine Zinn

ZUR ENTWICKLUNG DES ZUSAMMEN-
HANGS UND DER ERKLÄRBARKEIT VON
LEISTUNGEN UND KOMPETENZEN BEI
SCHÜLERINNEN UND SCHÜLERN DER
SEKUNDARSTUFE I

LifBi Working Paper No. 77
Bamberg, November 2018

Working Papers of the Leibniz Institute for Educational Trajectories (LifBi)

at the University of Bamberg

The LifBi Working Papers series publishes articles, expert reports, and findings related to data collected and studies conducted at the Leibniz Institute for Educational Trajectories—first and foremost, the National Educational Panel Study (NEPS) in Germany.

LifBi Working Papers are edited by the LifBi Board of Directors and the Heads of the LifBi Departments. The series started in 2011 under the name “NEPS Working Papers” and was renamed in 2017 to broaden the range of studies which may be published here.

Papers appear in this series as work in progress and may also appear elsewhere. They often present preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

Any opinions expressed in this series are those of the author(s) and not those of the LifBi management or the NEPS Consortium.

The LifBi Working Papers are available at www.lifbi.de (see section “Institute > Publications”). LifBi Working Papers based on NEPS data are also available at www.neps-data.de (see section “Data Center > Publications”).

Editor-in-Chief:

Corinna Kleinert, LifBi/University of Bamberg

Editorial Board:

Cordula Artelt, LifBi/University of Bamberg

Christian Aßmann, LifBi/University of Bamberg

Jutta von Maurice, LifBi

Sabine Weinert, LifBi/University of Bamberg

Contact:

Leibniz Institute for Educational Trajectories

Wilhelmsplatz 3

96047 Bamberg

Germany

contact@lifbi.de

Zur Entwicklung des Zusammenhangs und der Erklärbarkeit von Leistungen und Kompetenzen bei Schülerinnen und Schülern der Sekundarstufe I

Bayer, Michael
Leibniz-Institut für Bildungsverläufe

Zinn, Sabine
Leibniz-Institut für Bildungsverläufe

E-Mail-Adresse des Erstautors:

michael.bayer@lifbi.de

Bibliographische Angaben bei deutschsprachigen Papers:

Bayer, M. & Zinn, S. (2018). *Zur Entwicklung des Zusammenhangs und der Erklärbarkeit von Leistungen und Kompetenzen bei Schülerinnen und Schülern der Sekundarstufe I* (LifBi Working Paper No. 77). Bamberg, Deutschland: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel.

Zur Entwicklung des Zusammenhangs und der Erklärbarkeit von Leistungen und Kompetenzen bei Schülerinnen und Schülern der Sekundarstufe I

Abstract

Immer wieder diskutiert die Öffentlichkeit über eine vermeintliche Inflation guter Noten und einen sich dahinter verbergenden Rückgang des schulischen Anforderungsniveaus. Der Beitrag untersucht, ob sich eine solche Entwicklung in der Sekundarstufe I im Vergleich zweier 9-Klasse-Kohorten nachweisen lässt und wodurch sich die individuellen Noten in den Fächern Mathematik und Deutsch erklären lassen. In den Analysen mit Daten des Nationalen Bildungspanels (NEPS) zeigt sich im Untersuchungszeitraum von 2010/11 und 2014/15 keinerlei Verbesserungen der Durchschnittsnoten und auch keine signifikante Zunahme des Zusammenhangs zwischen Noten und Kompetenzen. Gleichzeitig wird aber deutlich, dass die Klassenebene, und hier insbesondere die Lehrkraft, die zentrale Erklärungsebene für das Zustandekommen von Noten darstellt.

There is an ongoing public discussion on an alleged inflation of excellent grades and the underlying process of decreasing levels of academic standards. In this paper, we test this assumption via an analysis with data on two ninth-grade cohorts. Moreover, we try to explain how marks in mathematics and in German can be predicted. With data from the National Educational Panel Study (NEPS), we find that between 2010/11 and 2014/15 there is no evidence for an improvement of average grades, and also not for an increasing correlation between grades and competencies. However, our results show that the strongest effects in explaining grades can be found on the level of school classes and particularly with regard to teachers.

Schlagworte

Schulnoten, Kompetenzen, Sekundarstufe I, Kohortenvergleich

1. Empirischer und konzeptioneller Ausgangspunkt

Der Erwerb von Bildungszertifikaten wie beispielsweise von Schulzeugnissen stellt ein wichtiges Ergebnis des Absolvierens von Bildungsetappen dar. Das Zeugnis soll die Leistungsfähigkeit der Person dokumentieren und fungiert als ein Signal an nachfolgende Bildungsetappen. Dadurch wird aber auch deutlich, dass das Zustandekommen des Zeugnisses bestimmten Kriterien, wie etwa Objektivität, genügen sollte. Das Zeugnis bzw. die Note als ein Indikator der Leistungsbeurteilung macht es notwendig zu klären, was unter Schulleistung zu verstehen ist. Dies ist auch deshalb von großer Relevanz, weil sowohl die Schülerinnen und Schüler wie auch die Schule einer Vielzahl an politischen und gesellschaftlichen Einflüssen ausgesetzt sind, die nicht nur die Leistungserzeugung, sondern auch die Definition dessen was unter Leistung verstanden wird, betreffen.

Die Erforschung dieses Prozesses steht bereits seit längerem im Zentrum pädagogischer Forschungszugänge (Fend 1980). Vor allem der Zuverlässigkeit der Beurteilung der Schülerleistungen durch die Lehrkraft galt sehr früh die Aufmerksamkeit (vgl. Ingenkamp 1971). In diesen Forschungen wurden sehr wichtige Befunde fixiert, etwa die besseren Noten und Leistungen von Mädchen im Vergleich zu Jungen, der in Gymnasien beobachtbare Trend zu durchschnittlich schlechteren Noten über die Schulzeit hinweg sowie die großen Unterschiede zwischen den Schulformen hinsichtlich der Leistungsfähigkeit von Schülerinnen und Schülern und nicht zuletzt die Kontextabhängigkeit der Lehrerbeurteilung, welche Noten über Klassen oder gar Schulen hinweg nur schwer vergleichen lassen.

Unabhängig von diesen Befunden und letztlich auch von den Einschränkungen der Objektivität von Noten als Leistungsindikatoren, bilden diese nach wie vor das zentrale Beurteilungs- und Messkriterium der individuellen Schülerleistung.¹ Gleichzeitig findet immer wieder eine Diskussion über eine vermeintliche Noteninflation an deutschen Schulen statt und insbesondere der Deutsche Philologenverband (DPhV) ist eine durchaus kritische Instanz hinsichtlich der sogenannten *Bestnotenentwicklung*, welche die Schulabschlüsse, etwa das Abitur aber auch das Studium, entsprechend entwerten würde. Diesen Warnrufen unterliegt eine Einschätzung, die davon ausgeht, dass der Prozess der Erzeugung von Leistungsbeurteilungen sich von der wahren Leistung entkoppelt. Man kann diese in der Öffentlichkeit stattfindenden Diskussionen nunmehr mit einem Blick in die entsprechenden Abschlussnotenstatistiken relativ schnell entkräften, da sich etwa hinsichtlich der Durchschnittsnoten beim Abitur bundesweit in den letzten 10 Jahren keinerlei Trend zu einer Verbesserung konstatieren lässt.²

Hinter diesen Diskussionen scheinen sich jedoch mehr als nur ein fehlendes Verständnis bzw. Wissen über die empirische Situation zu verbergen. So findet sich in einer aktuellen Studie der Konrad-Adenauer-Stiftung unter dem Titel „Ausbildungsreife und Studierfähigkeit“ folgende Einschätzung: „Denn trotz gestiegener guter Schulabschlüsse steigt die Anzahl der jungen Menschen, die gleich zu Beginn einer Berufsqualifikation in Unternehmen oder Hörsälen mit fehlenden Grundlagenkompetenzen hinsichtlich Sprache und Mathematik zu kämpfen

¹ Noten stellen einen durchaus ernstzunehmenden Indikator etwa für Ausbildungserfolg dar (vgl. Velten/Schnitzler 2011), was zumindest darauf hindeutet, dass sie innerhalb des Bildungs- und Ausbildungssystems als brauchbarer Einschätzungsindikator für Übergänge dienen können.

² Die von den Kultusministerien der Bundesländer erhobenen und von den Statistischen Landesämtern aufbereiteten Daten über die Abiturnoten zeigen keine Anhaltspunkte, dass es in den letzten Jahren zu einer Verbesserung der Durchschnittsnoten gekommen ist (<https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html>).

haben.“ (Henry-Huthmacher/Hoffmann 2016: 5). Hier wird ein Fehlen bestimmter Grundlagenkompetenzen bemängelt.

Es sind unserer Einschätzung nach zwei zentrale Fragen, die bei der Untersuchung des Zusammenhangs zwischen Leistungen und Kompetenzen von Schülerinnen und Schülern bearbeitet werden müssen. *Erstens* ist zu klären was unter (Schul-)Leistung zu verstehen ist bzw. inwieweit sich das vorherrschende Verständnis von Schulleistung selbst verändert hat. *Zweitens* gilt es zu untersuchen wie die verstärkte Nutzung des Kompetenzbegriffs auch und gerade als Messindikator für Leistungsbereitschaft bzw. Leistungsfähigkeit die innerschulischen Prozesse in einer Weise verändert hat, welche bei gleichbleibender oder gesteigener Schulleistung (in Form von Noten) Aussagen über fehlende Ausbildungsreife bzw. Studierfähigkeit erklärbar machen. Während die erste Frage im Rahmen der hier verwendeten empirischen Daten vor allem in Richtung der Analyse des Prozesses der Erklärung von Schulnoten bearbeitet werden kann, bedarf eine umfassende Beantwortung der zweiten Frage darüber hinaus Daten, welche es ermöglichen innerschulische Prozesse unter Fokussierung der beteiligten Akteure zu rekonstruieren.

Die Beantwortung dieser Fragen ist allein schon deshalb wichtig, weil Bildung in unserer Gesellschaft einen wichtigen Stellmechanismus für die Verteilung von Lebenschancen darstellt und Veränderungen in diesem Bereich Auswirkungen auf andere gesellschaftliche Felder, etwa Arbeitsmärkte, nach sich ziehen. In modernen, auf Erwerbsarbeit basierenden Gesellschaften werden Lebenschancen vornehmlich über Arbeitsmärkte verteilt (Becker/Hadjar 2017) wobei die Zugänge zu Arbeitsmärkten idealtypisch in starker Abhängigkeit von erbrachten *Bildungsleistungen* stehen. Die Gesamtkette von Bildung, Arbeitsmarkt, Berufsposition und Einkommen soll Ausdruck der erbrachten unterschiedlichen Leistungen der Individuen unter Bedingungen von Chancengleichheit sein.³ Jenseits der Frage, inwieweit sich meritokratische Funktionsweisen identifizieren lassen, hat die Orientierung an dieser Leitidee dazu geführt, dass das Bildungssystem als Institution zur Erzeugung und Absicherung von relativen Lebenschancen zunehmend an Bedeutsamkeit gewonnen hat (Solga 2005).

2. Bildung, Leistung, Kompetenz

Die drei Konzepte *Bildung*, *Leistung* und *Kompetenz* stehen im Zentrum bildungswissenschaftlicher und bildungspolitischer Auseinandersetzungen (für einen Überblick vgl. Bayer 2015). In den entsprechenden Arbeiten werden insbesondere Verschiebungen hinsichtlich bildungspolitischer Schwerpunktsetzungen thematisiert, die mit Stichworten wie internationale Vergleichbarkeit oder bildungspolitischer Wettbewerb einhergehen. Das heißt, es wird ein Prozess konstatiert, den Münch (2012) auf einer makrosozialen Ebene als Wandel vom Sozial- zum Wettbewerbsstaat bezeichnet.

Alle drei Konzepte sind zentral für die Beantwortung der Frage nach den Beurteilungskriterien bzw. dem Zustandekommen des Lehrerurteils über die Schülerleistung. Letztendlich ausschlaggebend für die Beurteilung der erbrachten Schülerleistung durch die Lehrkraft sind idealerweise die durch die Kultusministerkonferenz (KMK) festgelegten *Bildungsziele*. Diese werden mittlerweile vor allem unter dem Begriff der abschlussbezogenen *Bildungsstandards*

³ Dass diese meritokratische Vorstellung eher ein Idealbild darstellt und die soziale Wirklichkeit hiervon durchaus abweicht wird in entsprechenden Forschungen zum Thema Bildungsgerechtigkeit breit diskutiert (vgl. Hadjar 2008).

diskutiert (vgl. Klieme et al. 2003; kritisch v.a. Gruschka 2006). Hierbei umfassen die abschlussbezogenen Bildungsstandards „fachliche und fachübergreifende Basisqualifikationen, die für die weitere schulische und berufliche Ausbildung von Bedeutung sind und die anschlussfähiges Lernen ermöglichen“ (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland 2005: 7).⁴ Beides, sowohl die Bildungsziele wie auch die hieraus abgeleiteten abschlussbezogenen Bildungsstandards, finden ihre inhaltliche Verankerung im Verständnis des Begriffs der Bildung.

Der *Bildungsbegriff* ist äußerst vielschichtig. Unter dem Titel „Bildung. Mehr als Fachlichkeit“ formuliert der *Aktionsrat Bildung* einen aus den drei Aspekten *Wissen, Kompetenz* und *Persönlichkeitsbildung* bestehenden mehrdimensionalen Bildungsbegriff (vbw 2015: 23). Damit ordnet sich der Rat bewusst in eine kontinentale historische Tradition ein, die immer schon die beiden Aspekte „Bildung als Erwerb von Fertigkeiten bzw. Kompetenzen“ und „Bildung als Persönlichkeitsentwicklung“ als konstitutiv für das gesamte Bildungsgeschehen ansieht.

Bildung soll demnach sowohl den Erwerb von Kompetenzen als auch die Herausbildung der Persönlichkeit umfassen. Die OECD-Arbeitsgruppe „Definition and Selection of Key Competencies“ nutzt für die Beschreibung von Persönlichkeitsentwicklung vor allem den Begriff der Reflexivität als *wesentliche* Schlüsselkompetenz (DeSeCo 2005). Im Deutschen Qualifikationsrahmen für lebenslanges Lernen (DQR) wird Kompetenz definiert als „Fähigkeit und Bereitschaft des Einzelnen, Kenntnisse und Fertigkeiten sowie persönliche, soziale und methodische Fähigkeiten zu nutzen und sich durchdacht sowie individuell und sozial verantwortlich zu verhalten. Kompetenz wird in diesem Sinne als umfassende Handlungskompetenz verstanden.“ (DQR 2011: 9). Das hier verwendete Verständnis von Kompetenz als Handlungskompetenz ist jedoch grundlegend zu unterscheiden von dem Kompetenzverständnis, welches sich etwa in den großen Schulleistungstudien findet und welches ein kognitives Dispositionskonstrukt darstellt (Fleischer et al. 2013).

Frey und Bonsen (2013) konnten zeigen, dass Schülerinnen und Schüler mit schlechten Fachnoten mit einer höheren Wahrscheinlichkeit auch in Bezug auf ihr Arbeits- und Sozialverhalten eine weniger positive Beurteilung durch die Lehrkräfte erhalten. Dieser Befund weist darauf hin, dass Lehrkräfte bei der Gesamtbeurteilung einer Schülerin oder eines Schülers nicht zwingend zwischen fachbezogenen Kompetenzen und überfachlichen auf die Persönlichkeitswerdung abzielenden sozialen Kompetenzen einer Schülerin oder eines Schülers unterscheiden. Dieser Befund deutet ferner darauf hin, dass die Leistungsbeurteilung einer Person in der Schulpraxis von einem Zusammenspiel aus fachlichen und sozialen Kompetenzen sowie subjektiven Lehrerurteilen abzuhängen scheint.

Heller definierte den Begriff *Schulleistung* aus einer dezidiert pädagogischen Position: „Mit dem Begriff der Schulleistung sei hier das gesamte Leistungsverhalten, soweit es im Kontext schulischer Bildungsbemühungen virulent wird, angesprochen. Dabei verdienen der dynamische Aspekt (Lernprozeß) und der statische Aspekt (Lernprodukt) gleichermaßen Beachtung — entgegen manchen Begriffsvorstellungen, die nur das Ergebnis der Schülerleistung und nicht auch deren Bedingungsgefüge im Auge haben.“ (Heller 1974: 14).

⁴ Detailliertere Informationen zur Ausgestaltung der abschlussbezogenen Bildungsstandards finden sich unter anderem in einem Argumentationspapier des Sekretariats der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005) und auf der Webseite der KMK (<https://www.kmk.org/themen/qualitaetssicherung-in-schulen/bildungsstandards.html#c2604>).

Das Verständnis von Leistungen in der Schule steht hierbei in einem direkten Abhängigkeitsverhältnis von den beiden – auch bei Heller notierten – Primärfunktionen der Schule, auf die auch die bereits oben benannten Formulierungen des Aktionsrats Bildung abzielen: Persönlichkeitsbildung und Sicherung des volkswirtschaftlichen Leistungspotentials (vgl. hierzu auch Fend 2008).

Der seit den Befunden des PISA-Konsortiums (2001) geführte Bildungsdiskurs in Deutschland fokussiert insbesondere auf die systematische Entwicklung von *Kompetenzen*. Somit liegt die Vermutung nahe, dass eine Schwerpunktverlagerung stattgefunden hat von dem, was als Bildungsdimension des *Wissens* formuliert wurde (vgl. hierzu auch Klafki 1985), hin zu einem Verständnis von *Bildung als Kompetenzerwerb*.

Wir vermuten, dass die Vielschichtigkeit des Kompetenzbegriffs bei den verschiedenen Akteuren des Bildungssystems (wie Lehrern, Professoren und Ausbildern) zu sehr unterschiedlichen Kompetenzverständnissen und somit Erwartungen an die Schüler und deren Befähigungen beim Schulabschluss führen. Wir vermuten ferner, dass sich seit dem ersten *PISA-Schock* 2003 im deutschen Schulsystem ein Bildungsverständnis verankert, welches entgegen konzeptioneller Forderungen von einer reinen Fokussierung auf inhaltliche oder gar normative Dimensionen abweicht und zusätzlich das gute Abschneiden bei staatlich angeordneten Kompetenztests wie PISA in den Mittelpunkt stellt. Zudem ist es, dies unsere dritte Vermutung, kaum möglich Persönlichkeitsentwicklung mit dem Begriff soziale Kompetenz wiederzugeben ohne zentrale Komponenten des deutlich umfassenderen Konzeptes des Persönlichkeitsbegriffs im Kontext eines funktionalistischen Kompetenzverständnisses zu vernachlässigen (vgl. hierzu auch Grunert 2012).⁵

Ziel dieses Artikels ist es unter Verwendung der Daten des Nationalen Bildungspanels (NEPS) diese Vermutungen zu untersuchen. Hierfür beleuchten wir zeitliche Veränderungen von Zusammenhängen bereichsspezifischer Kompetenzen und schulischer Leistungsbeurteilung in der Sekundarstufe I und die Wirkung von Kontextfaktoren, die mutmaßlich neben der Kompetenz einen signifikanten Einfluss auf die Notenvergabe haben. Zuvor werden im nachstehenden Abschnitt allerdings die bereits aufgeworfenen Problematiken mit den jeweiligen Forschungsständen abgeglichen.

3. Bildung als Gleichzeitigkeit von Kompetenzentwicklung und Leistungsbeurteilung

Jenseits der aktuell sehr stark auf den Kompetenzbegriff setzenden Schulleistungsstudien wie PISA, TIMSS⁶ oder NEPS, stehen Noten und Notenvergabe nach wie vor im Fokus bildungswissenschaftlicher Forschungen. Auch in diesen Forschungen findet sich meistens eine Trennung des Schulleistungsbegriffs vom Notenbegriff. Trautwein u.a. machen etwa darauf aufmerksam, dass gerade „in der Gegenüberstellung der prädikativen Kraft von Schulnoten und Schulleistung [...] besondere Chancen für die empirische Bildungsforschung (liegen)“ (Trautwein et al. 2008: 101). Kompetenzen – unabhängig davon, ob es sich hierbei um curriculumnahe kognitive Kompetenzen oder aber um Leistungen im Sinne des Literacy-Konzepts (Artelt et al. 2003) handelt – stehen allerdings in einem allenfalls moderaten

⁵ Diese Vermutung liegt jedoch außerhalb des Fokus dieses Beitrags.

⁶ TIMSS („Trends in International Mathematics and Science Studie“) ist eine internationale Bildungsvergleichsstudie, die alle vier Jahre am Ende der Grundschulzeit das mathematische und naturwissenschaftliche Grundverständnis von Schülerinnen und Schülern testet.

Zusammenhang mit Schulnoten. Dieses Phänomen konnte bereits Ingenkamp (1967) festhalten und konstatierte einerseits eine generell schlechte prognostische Qualität der Lehrerurteile (Ingenkamp 1971). Andererseits bemängelte er den Bildungsauftrag von Schulen, der bis in die siebziger Jahre hinein nicht die Entwicklung und Förderung fachspezifischer Kompetenzen beinhaltete. Der Zusammenhang zwischen den bei standardisierten Leistungstests ermittelten bereichsspezifischen Kompetenzen und den bereichsäquivalenten Schulnoten zeigte sich bereits in den PISA Studien als in der Tendenz eher schwach bis moderat. Es lässt sich hier jedoch eine deutliche Zunahme zwischen den Befunden der PISA-Studie aus dem Jahr 2000 und der PISA-Studie von 2012 feststellen und dies vor allem im Bereich nicht-gymnasialer Schulformen (Hauptschule, Realschule, integrierte Schulformen). Insbesondere in Hauptschulen hat sich der (korrelative) Zusammenhang zwischen der Leseleistung und der Deutschnote von $-.14$ auf $-.32$ deutlich verstärkt (Autorengruppe Bildungsberichterstattung 2016: 94). Es sind insbesondere die sozial schwächeren Herkunftsschichten, die mit 44 Punkten (im Vergleich zu 21 Punkten bei den oberen Statusgruppen) einen überdurchschnittlichen Zugewinn bezüglich der Zusammenhangsstärke zeigen.

4. Strukturelle Einflussfaktoren auf die schulische Leistungsbewertung

Insbesondere die Analyse der Leistungsbeurteilung und Notengebung als solche steht im Fokus sowohl (schul-)pädagogischer wie auch bildungssoziologischer Fragestellungen. Für Letztere stellen der durch die PISA-Studien virulent gewordene Zusammenhang von sozialer Herkunft und Bildungschancen sowie die hier festgestellte Persistenz der Ungleichheitsreproduktion zentrale Forschungsanliegen dar. Dietrich und Fricke weisen nicht zu Unrecht darauf hin, „dass sich die Reproduktion sozialer Ungleichheit wesentlich in der Praxis schulischer Leistungsbewertung vollzieht“ (Dietrich/Fricke 2013: 283). Die Autoren arbeiten in ihrer Studie heraus, dass die Orientierung am Leistungsprinzip in der Notengebung durch die Lehrkraft in hohem Maße durch „das kollegiale Klima an der Schule, die Schülerschaft und [...] auch durch die Schulform bedingt ist“ (Dietrich/Fricke 2013: 287).

4.1 Lehrkräfte als Leistungsermittler und -beurteiler

Viele in der Öffentlichkeit ausgetragenen Diskussionen bezüglich Schulnoten- bzw. Abschlussnotenstatistiken simplifizieren den komplexen Prozess des Zustandekommens von Noten, indem sie relevante Einflussfaktoren wie eben die Lehrkraft ausklammern. Wie Helbig und Morar (2017) jedoch betonen, spielen bei der Notengebung zwei Akteure eine zentrale Rolle: der Schüler und die Lehrkraft⁷. Mit Daten zu Noten in der vierten Klasse einer auf die Bundesländer Hessen und Bayern beschränkten Studie können sie zeigen, dass die Einschätzungen der Lehrkräfte über die Begabung, das Arbeitsverhalten sowie die Wahrnehmung der elterlichen Unterstützung einer Schülerin bzw. eines Schülers die bedeutsamsten Prädiktoren für eine Erklärung des Einflusses der sozialen Herkunft auf die Varianz in den Noten darstellen. Damit wird zumindest deutlich, dass die Einschätzungen der Lehrkraft erstens wichtig für das Leistungsbeurteilungsgeschehen sind, sich zweitens nicht in einem auf kognitive Kompetenzen beschränkten Beurteilungsraum bewegen und drittens die Lehrerbeurteilung sowohl bezüglich der kognitiven als auch bezüglich der psychosozialen Dimension in einer sozial distinkten Weise erfolgt. Dies macht auch klar, dass die Wirkungen

⁷ Wir werden weiter unten jedoch zeigen, dass dies nicht die einzigen Akteure sind, die für das Zustandekommen von Leistungs- und Notenverteilungen relevant sind.

von Bildungsreformen immer auch durch die Lehrkräfte getrichert werden.⁸ Diese Befunde geben gleichzeitig auch eine Antwort auf die fehlenden Zusammenhänge, die zwischen (durch die Schüler) selbsteingeschätzten psychosozialen und motivationalen Merkmalen und den Schulnoten bestehen, wie sie etwa in der Studie von Maaz, Baeriswyl und Trautwein (2011: 38ff.) berichtet werden. Nicht die von den Schülern selbst berichteten Motivationen und Arbeitsweisen erzeugen Diskrepanzen, sondern die durch die Lehrkraft wahrgenommenen Motivationen und Arbeitsweisen der Schülerinnen und Schüler.

4.2 Bildungssystemische Rahmenbedingungen und Veränderungen

Lehrkräfte handeln auf Klassenebene, welche sich jedoch, je nach Schultyp bzw. auch nach regionaler Verortung der Schule, durch eine Vielzahl an Merkmalen auszeichnet, deren Wirkung auf die Benotungspraxis und demnach auch die Verteilung der Noten vielschichtig ist (vgl. Brühwiler/Helmke/Schrader 2017).

Insofern ist es nicht verwunderlich, dass die vorhandenen Studien zu Prozessen der Leistungsbeurteilung in Schulen sich häufig entweder auf den Bereich der Grundschulen beziehen (etwa Maaz/Baeriswyl/Trautwein 2011) oder aber auf besondere Gruppen wie etwa Gymnasiasten fokussieren (Becker/Birkelbach 2010). In beiden Fällen spielt das im Sekundarbereich relevante, stratifizierte Schulsystem auf der Ebene der Schulformen bzw. -typen eher eine vernachlässigbare Rolle. Eine ganzheitliche Untersuchung der Kompetenz- und Leistungsentwicklung von Schülerinnen und Schülern in Deutschland erfordert jedoch die Berücksichtigung von bundeslandspezifischen Strukturmerkmalen sowie deren Veränderung über die Zeit, da Bildung in Deutschland in der Steuerungshoheit der Länder liegt. Betrachtet man etwa die durchaus bedeutsamen Unterschiede zwischen den Bundesländern hinsichtlich der Verteilung der Schülerinnen und Schüler entlang der erworbenen Abschlussarten (Autorengruppe Bildungsberichterstattung 2016: 98), so wird deutlich, dass sich hier auf Bundeslandebene die Schülerpopulationen in den Schularten unterschiedlich zusammensetzen. Entsprechend ist die Beachtung dieser systemischen Komponente in der Analyse des Notengebungsprozesses unabdingbar.

5. Erklärungsmodell

Die bisherigen Ausführungen stellen heraus, dass das Benotungsgeschehen nicht nur in seinen Erzeugungsfaktoren komplex ist, sondern dass diese Erzeugungsfaktoren auch auf unterschiedlichen Ebenen zu verorten sind. Zudem muss berücksichtigt werden, dass das schulische Bildungsgeschehen über die Zeit deutlichen Veränderungsprozessen ausgesetzt ist. Beides, die systembedingte und die zeitliche Struktur soll in den folgenden Analysen in den Blick genommen werden, um herauszuarbeiten, welche Faktoren im Hinblick auf die Note als Leistungsindikator im systembedingten und zeitlichen Vergleich erklärungskräftig sind.

In der nachstehenden Abbildung finden sich Einflussebenen, die sich entweder in den bisherigen Analysen und Diskussionen als wirkungsvoll gezeigt haben oder aber aus konzeptionell-theoretischen Gründen als wirkmächtig vermutet werden.

⁸ Natürlich kennt man spätestens seit der Studie von Hattie (2009) die Bedeutsamkeit der Lehrkräfte.

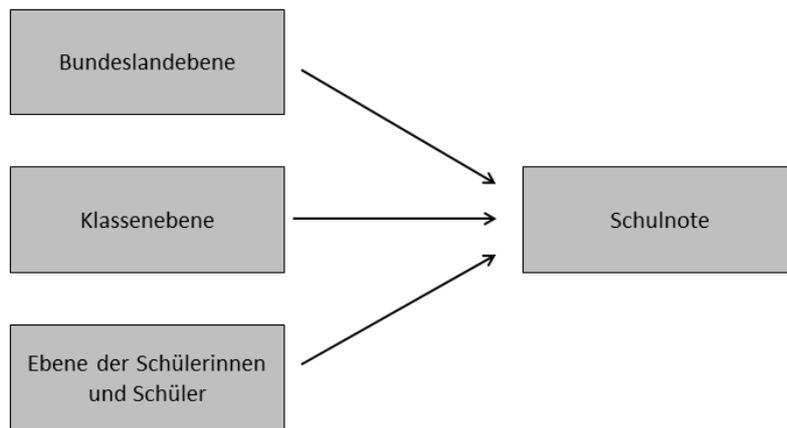


Abbildung 1: Multivariates Erklärungsmodell

Im Zentrum stehen hierbei zwei Fragen: erstens die Frage nach dem Zusammenhang von Kompetenz (Schülerebene) und Note und zweitens die Frage nach der Veränderung dieses Zusammenhangs über die Zeit (im konkreten Fall, der Vergleich von Schülerinnen und Schülern in der neunten Jahrgangsstufe in den Schuljahren 2010/11 und 2014/15). Der diesbezüglich zu untersuchende Prozess wird - wie bereits ausführlich herausgestellt - von Faktoren auf drei Hierarchieebenen beeinflusst (vgl. Abbildung 1), nämlich von bundeslandspezifischen Aspekten, der Schul- bzw. der Klassenebene und der Individualebene.

6. Daten und Konstrukte

Die folgenden Analysen wurden mit Daten des Nationalen Bildungspanels (NEPS) berechnet⁹ (Blossfeld/Roßbach/von Maurice 2011), wobei hier zwei Neunte-Klasse-Erhebungen aus den Schuljahren 2010/11 (Startkohorte 4, 2. Welle) und 2014/15 (Startkohorte 3, 5. Welle) genutzt wurden. Beide Erhebungen fanden im Frühjahr des entsprechenden Schuljahres statt. Die zugehörigen Schul- und Schülerstichproben wurden repräsentativ für das Schuljahr 2010/11 aus der Menge aller in Deutschland staatlich anerkannten Schulen in den Jahrgangsstufen 5 bzw. 9 gezogen.¹⁰ Hierbei erfolgte eine explizite Stratifizierung nach Schulform (Gymnasium, Realschule, Hauptschule, integrierte Gesamtschule, Schule mit mehreren Bildungswegen).¹¹

⁹ Diese Arbeit nutzt Daten des Nationalen Bildungspanels (NEPS): Startkohorte Klasse 5, doi:10.5157/NEPS:SC3:7.0.0 und Startkohorte Klasse 9, doi:10.5157/NEPS:SC4:9.1.0. Die Daten des NEPS wurden von 2008 bis 2013 als Teil des Rahmenprogramms zur Förderung der empirischen Bildungsforschung erhoben, welches vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde. Seit 2014 wird NEPS vom Leibniz-Institut für Bildungsverläufe e.V. (LIfBi) an der Otto-Friedrich-Universität Bamberg in Kooperation mit einem deutschlandweiten Netzwerk weitergeführt.

¹⁰ Die Schülerinnen und Schüler, die im Schuljahr 2014/15 die 9. Jahrgangsstufe besuchen, befanden sich zum Großteil im Schuljahr 2010/11 in der 5. Jahrgangsstufe. Sitzenbleiber und Überspringer sind von den folgenden Analysen und bedürfen einer gesonderten Untersuchung. Zudem enthalten die NEPS-Daten keine entsprechenden Informationen für das Schuljahr 2014/15, da der Fokus der zugehörigen Studie auf der Gruppe der Schülerinnen und Schüler liegt, die im Schuljahr 2010/11 die 5. Jahrgangsstufe besuchten.

¹¹ Die Stichproben enthalten auch Schülerinnen und Schüler in Förderschulen. Für die vorliegende Untersuchung werden diese allerdings von den weiteren Betrachtungen ausgeschlossen. Der Grund hierfür liegt in der strukturell und inhaltlich völlig differenten Beschulung von Schülerinnen und Schülern in Förderschulen, die einer gesonderten Analyse bedürfen. Auch die Überziehungsstichprobe von Kindern mit Migrationshintergrund wird wegen ihrer sehr geringen Ausschöpfungsquote von weiteren Betrachtungen ausgeschlossen. Gleiches gilt auch für Schülerinnen und Schüler in Freien Waldorfschulen. Hier war die Ausschöpfung bei den Schulen mit 5.

Zudem wurde implizit nach Bundesland, Urbanisierungsgrad und Trägerschaft stratifiziert. Die Ziehung von Schulen mit einer 5. bzw. 9. Jahrgangsstufe fand größenproportional zur Anzahl an Schülerinnen und Schülern, die in diesen Jahrgangsstufen beschult wurden, statt. Im Schuljahr 2012/13 wurde die Stichprobe der Fünftklässler in der 7. Jahrgangsstufe aufgefrischt. Insgesamt konnten 203 Schulen mit einer 5. Jahrgangsstufe ergänzt um 86 Schulen in der 7. Jahrgangsstufe sowie 545 Schulen mit einer 9. Jahrgangsstufe rekrutiert werden. An jeder der rekrutierten Schulen wurden zufällig zwei 5. bzw. 7. bzw. 9. Klassen herausgegriffen. Falls nur eine Klasse vorhanden war, wurde diese ausgewählt.¹² Die NEPS-Daten enthalten für das Schuljahr 2010/11 Noten- und Kompetenzinformationen für 12.604 Neuntklässler in 968 Klassen an 518 Schulen. Für das Schuljahr 2014/15 enthalten sie entsprechende Informationen für 4.007 Neuntklässler in 539 Klassen an 223 Schulen. Da die Teilnahme an den NEPS-Studien nicht verpflichtend ist, gibt es Ausfall sowohl auf Schul- als auch auf Schülerebene. Eine Möglichkeit um diesen Ausfall geeignet auszugleichen ist die Verwendung von Surveygewichten. Diese setzen sich zusammen aus den sogenannten Designgewichten auf Schul- und Schülerebene und Adjustierungsfaktoren, die den Ausfall von Schulen und Schülerinnen bzw. Schülern kompensieren.¹³ Hierbei gibt das Designgewicht einer Schule bzw. einer Schülerin oder eines Schülers an, wie viele gleichartige Elemente das Stichprobenobjekt in der Zielpopulation repräsentiert. Adjustierungsfaktoren werden derart bestimmt, dass die gewichteten Daten bzgl. relevanter Schul- und Schülereigenschaften wie dem Geschlecht, der zu Hause gesprochenen Sprache, der Schulform und bestimmten Regionalmerkmalen ebenso verteilt sind wie die Zielpopulation der Schülerinnen und Schüler. Informationen zum Stichprobendesign und zur Erstellung von Design- und Ausfallgewichten finden sich in u.a. in Steinhauer et al. (2015). Stichprobengrößen und Ausfallprozesse werden in Zinn et al. (2018) beschrieben.

Unter Bezug auf das theoretische Modell fanden in den Analysen folgende Konstrukte bzw. Variablen Eingang.

Als *abhängige Variablen* werden die Halbjahresnoten in den beiden Fächern Mathematik und Deutsch herangezogen. Die entsprechenden Informationen wurden von den Schülerinnen und Schülern nach der Kompetenztestung erfragt. Auf der *Ebene der Schülerinnen und Schüler* liegt unser Hauptaugenmerk auf dem Zusammenhang zwischen den Noten in Mathematik und Deutsch und den jeweiligen bereichsspezifischen Kompetenzen in *Mathematik* und *Lesen*. Die Mathematikkompetenz wurde mittels 22 Multiple Choice Fragen in einem eigens für NEPS entwickelten Test erfasst (vgl. Neumann et al. 2013). Zur Erfassung der Lesekompetenz wurden die Schülerinnen und Schüler gebeten 5 kurze Texte aus verschiedenen Bereichen (literarische und kommentierende Texte sowie Informations-, Hinweis- und Werbetext) zu lesen und zu reflektieren (vgl. Gehrter et al. 2012). Administriert wurden die Tests in

Jahrgangsstufen im Schuljahr 2010/11 so gering, dass eine sinnvolle statistische Auswertung nicht möglich ist. Daher sind auch Schülerinnen und Schüler in Freien Waldorfschulen von der Analyse ausgenommen.

¹² Somit können mit NEPS-Daten Effekte auf Klassenebene bestimmt werden und nicht nur, wie etwa mit PISA-Daten, auf Schulebene.

¹³ Die hierfür notwendigen Informationen wurden den amtlichen Schuldaten entnommen, die zum Zwecke der Stichprobenziehung und -validierung dem NEPS zur Verfügung gestellt wurden. Randanpassungen wurden auf Basis von Informationen der Fachserie 11 durchgeführt. Die Fachserie 11 dokumentiert die wichtigsten statistischen Eckzahlen aller allgemeinen bildenden staatlich anerkannten Schulen in Deutschland und wird jedes Jahr vom statistischen Bundesamt zur Verfügung gestellt, vgl. <https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/AllgemeinbildendeSchulen.html>.

Testgruppen an den Schulen, die die Schülerinnen und Schüler besuchten. Beide Tests erreichen gute IRT-basierte Reliabilitäten (Adams 2004). Der Mathematiktest weist eine Reliabilität von 0.79 auf und der Lesekompetenztest eine Reliabilität von 0.75. Weitere Informationen bzgl. der psychometrischen Eigenschaften der Tests (inklusive Angaben zur Dimensionalität und Messinvarianz) finden sich in Durcharadt und Gerdes (2013) sowie in Haberkorn et al. (2012). Alle Testergebnisse wurden mittels Methoden der Item Response Theorie skaliert (siehe Pohl/Carstensen 2013) und Kompetenzscores als *Weighted Maximum Likelihood* (WLE) Schätzer abgeleitet (vgl. Fischer et al. 2016).

Ein immer wieder diskutierter Erklärungsfaktor für einen allgemeinen Anstieg des durchschnittlichen Notenniveaus in Schulen ist der sogenannte „Flynn-Effekt“, ein allgemeiner Anstieg der durchschnittlichen Intelligenz (vgl. zum Flynn-Effekt z.B. die Metaanalyse von Pietschnig/Voracek 2015). Um die generelle Intelligenz einer Schülerin oder eines Schülers zu quantifizieren und in den Analysen zu kontrollieren, nutzen wir ein Maß zu den schlussfolgernden Grundfähigkeiten (DGCF). Diese wurden im NEPS mittels eines sogenannten Matrixtests mit 12 Fragen erhoben (vgl. Brunner/Lang/Lüdtke 2014). Die in den Matrizen des Tests dargestellten bildlichen Formen sind sehr gut geeignet, um fluide Intelligenz zu messen. Somit gestattet der im NEPS verwendete Matrixtest eine valide Approximation der generellen Intelligenz (Wilhelm 2005). Im Schnitt lösten die Schülerinnen und Schüler $M=8,56$ ($SD = 2,46$) Fragen richtig. Die kategoriale *omega* totale Reliabilität (Green und Yang 2009) dieses Maßes der schlussfolgernden Grundfähigkeiten beträgt $\omega_{total} = 0,74$.

Daneben finden die Merkmale *Geschlecht*, *Alter* der Schülerinnen und Schüler sowie der *Migrationshintergrund* Beachtung sowie der *Bildungshintergrund*, ausgedrückt durch den Bildungsabschluss der Mutter¹⁴ (operationalisiert anhand der CASMIN-Klassifizierung). Der Migrationshintergrund wird hierbei durch den Generationenstatus einer Schülerin bzw. eines Schülers abgebildet (vgl. Olczyk/Will/Kristen 2014). Konkret gehören Schülerinnen und Schüler, die im Ausland geboren wurden, der 1. Migrantengeneration an. Gleichermaßen gehören Schülerinnen und Schüler, die in Deutschland geboren wurden aber mit mindestens einem im Ausland geborenen Elternteil, der 2. Migrantengeneration an. Für die vorliegenden Analysen definieren wir eine Schülerin oder einen Schüler als Person mit Migrationshintergrund, falls sie Generationenstatus 1 oder 2 besitzt. Bezüglich des Bildungshintergrundes unterscheiden wir vier CASMIN Kategorien¹⁵ für den mütterlichen Bildungsabschluss: (i) kein Abschluss [1a], Hauptschulabschluss ohne berufliche Ausbildung [1b], Mittlere Reife ohne berufliche Ausbildung [2b], (ii) Hauptschulabschluss und berufliche Ausbildung [1c], Mittlere Reife und berufliche Ausbildung [2a], (iii) Fachhochschulreife/Abitur ohne berufliche Ausbildung [2c_gen], Fachhochschulreife/Abitur und berufliche Ausbildung [2c_voc] und (iv) Fachhochschulabschluss [3a], Hochschulabschluss [3b].

Auf der *Klassenebene* sind es die Indikatoren *Schulform* (Gymnasium vs. Nicht-Gymnasium), die *Anzahl* der Schüler und Schülerinnen in der Klasse sowie die *Anteile* an Schülerinnen und Schülern mit *Migrationshintergrund* bzw. mit *niedrigem sozialen Herkunftsstatus* sowie

¹⁴ Die Nutzung ausschließlich des mütterlichen Bildungsabschlusses hat einen sowohl datenbedingten wie auch inhaltlichen Grund. Einerseits lässt sich nicht in jedem Fall zweifelsfrei der Bildungsabschluss des Vaters in den Daten identifizieren und andererseits zeigt sich in Studien immer wieder, dass insbesondere die Bildung der Mutter einen zentralen Einfluss auf die Entwicklung von Kindern hat (Stevenson & Baker 1987; Youniss & Smollar 1985).

¹⁵ Genauere Informationen zur CASMIN Klassifizierung finden sich, unter anderem, in Brauns und Steinmann (1997).

allgemein der *Anteil an weiblichen Schülerinnen* in der Klasse, welche in die multivariaten Analysen Eingang finden.

Auf der *Bundeslandebene* nutzen wir die durch Helbig und Nikolai (2015) entwickelte Klassifikation von *Schulsystemtypen*, wobei hier zwischen „modernen Strukturen“, „traditionellen Mischformen“ und „modernisierten Mischformen“ unterschieden wird. Schulsystemtypen unterscheiden sich etwa hinsichtlich von schulrechtlichen Regelungen, den in einem System vorgesehenen Schulformen sowie der Zugänge zu unterschiedlichen Schulformen. Damit einher gehen unterschiedliche Lern- und Entwicklungsmilieus (Baumert et al. 2006), welche als auswirkungsreich hinsichtlich der Leistungsentwicklung von Schülerinnen und Schülern angenommen werden können. Tabelle 1 zeigt die Zuordnung der einzelnen Bundesländer zu den verschiedenen Ausprägungen der Schulsystemtypen.

Tabelle 1: Zuordnung der Bundesländer zu Schulsystemtypen und Freiwilligkeit von Lehrerfortbildung.

Bundesland	Schulsystemtyp Stand 2009-10¹⁶	Lehrerfortbildung Stand 2010-13¹⁷
Berlin & Brandenburg	Mischtyp-Modernisiert	verpflichtend
Baden-Württemberg	Mischtyp-Traditionell	verpflichtend
Bayern	Mischtyp-Traditionell	verpflichtend
Bremen	Modernisierte Strukturen	verpflichtend
Hessen	Modernisierte Strukturen	verpflichtend
Hamburg	Modernisierte Strukturen	verpflichtend
Mecklenburg-Vorpommern	Modernisierte Strukturen	verpflichtend
Niedersachsen	Mischtyp-Modernisiert	verpflichtend
Nordrhein-Westfalen	Modernisierte Strukturen	verpflichtend
Rheinland-Pfalz	Modernisierte Strukturen	freiwillig
Schleswig-Holstein	Modernisierte Strukturen	freiwillig
Saarland	Modernisierte Strukturen	verpflichtend
Sachsen	Mischtyp-Traditionell	verpflichtend
Sachsen-Anhalt	Mischtyp-Modernisiert	freiwillig
Thüringen	Mischtyp-Modernisiert	verpflichtend

Als einen weiteren Indikator auf der Ebene von Bundesländern nutzen wir die bundeslandspezifische Information über die Freiwilligkeit von *Lehrerfortbildungen* (vgl. Tabelle 1). Wir gehen davon aus, dass die Freiwilligkeit einen Proxy für die bundeslandspezifische Fähigkeit darstellt, auf die Ebene der Lehrkräfte einen bildungssteuernden Effekt auszuüben. Es gibt mittlerweile eine ausdifferenzierte Forschung zu Lehrerfortbildungsfragen etwa im Kontext neuer Steuerungsmodelle im Bildungssystem (siehe etwa Fussangel et al. 2010). Die meisten Studien, die das Thema fokussieren, konzentrieren sich hierbei auf Fragen der individuellen Teilnahmemotivation (Richter et al. 2018) oder des Einflusses auf die Entwicklungen und Leistungen der Schülerinnen und Schüler (Yonn et al. 2007).

Eine Zusammenstellung aller verwendeten Variablen inklusive Mittelwerten, Standardabweichungen und Anteilen findet sich in Tabelle 2.

¹⁶ Das Schuljahr 2009/10 ist das letzte Schuljahr, für das Helbig und Nikolai (ebd.) ihre Schulsystemtyp-Klassifizierung bereitstellen. Für nachfolgende Jahre liegt keine Klassifizierung mehr vor.

¹⁷ Die Angaben zur Lehrerfortbildung wurden den entsprechenden Seiten der Kultusministerien entnommen.

7. Deskriptive Analysen

In einer ersten deskriptiven Analyse betrachten wir die Veränderung der Notenverteilung in Mathematik und Deutsch in den gewichteten NEPS-Daten im Schulformvergleich (d.h. Gymnasium bzw. Gymnasialzweig an integrierten Gesamtschulen vs. Realschule bzw. Realschulzweig an integrierter Gesamtschule oder an Schule mit mehreren Bildungswegen vs. Hauptschule bzw. Hauptschulzweig an integrierter Gesamtschule oder an Schule mit mehreren Bildungswegen vs. unklarer Ausrichtung¹⁸) über die Zeit.

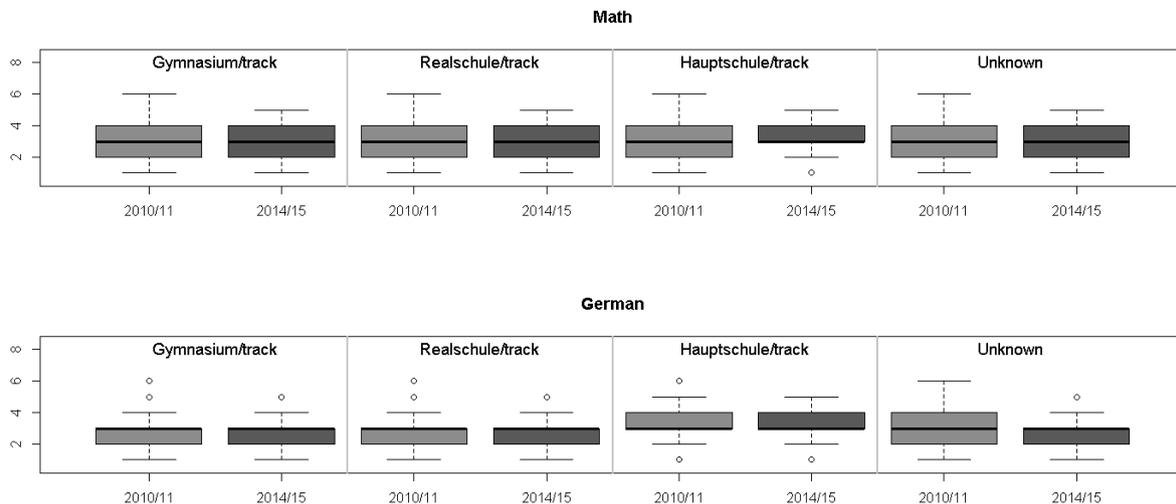


Abbildung 2: Verteilung der Mathematik- und Deutschnoten in den Schuljahren 2010/11 und 2014/15 nach Schulformen (dargestellt mittels Boxplots; Abkürzungen GY: Gymnasium, RS: Realschule, HS: Hauptschule)

Abbildung 2 zeigt diese Verteilung in Form von sogenannten *Boxplots*. Wir sehen, dass die Note 3 die Durchschnittsnote in Mathematik und Deutsch ist (gekennzeichnet durch die schwarze horizontale Linie) unabhängig von der Schulform bzw. dem besuchten Schulformzweig und dem Schuljahr. Zudem ist die größere Varianz bei den Mathematiknoten im Vergleich zu den Deutschnoten offensichtlich (ersichtlich sowohl an der Länge des *Interquartile Range* dargestellt durch die graue Box als auch an der Länge der sogenannten *Whiskers* abgebildet durch die schwarzen vertikalen Linien). Abgesehen von einer Fallzahl bedingten Varianzverringerung bei den Mathematiknoten im Hauptschulbereich und bei den Deutschnoten unklaren Schulformzuordnungen vom Schuljahr 2010/11 zum Schuljahr 2014/15¹⁹, finden sich keine Veränderungen der Notenverteilungen über die Zeit.

¹⁸ Unklare Schulformzuordnungen in den NEPS-Daten entstehen dadurch, dass Schulforminformationen von den Schülerinnen und Schülern bzw. ihren Eltern erfragt werden. Hierbei kommt es zu Ungenauigkeiten und/oder Widersprüchen, die eine eindeutige Zuordnung einer Schülerin oder eines Schülers zu einer Schulform nicht immer zulassen.

¹⁹ Insgesamt ist die Fallzahl der Schülerinnen und Schüler im Schuljahr 2014/10 deutlich geringer als im Schuljahr 2010/11. Zudem sind die beiden Notenbereiche 1 und 6 generell dünn besiedelt. All das in Kombination mit den niedrigen Fallzahlen bei Hauptschülern im Schuljahr 2014/15 führt zu den gezeigten (gewichteten) Notenverteilungen.

Eine Untersuchung der korrelativen Beziehung²⁰ zwischen der bereichsspezifischen Kompetenz und der jeweiligen Fachnote im Vergleich der beiden 9.-Klasse-Kohorten 2010/11 und 2014/15 zeigt einen geringen Zusammenhang zwischen beiden Maßen über alle betrachteten Schulformen hinweg (vgl. Abbildung 3). Der korrelative Zusammenhang schwankt zwischen 0.41 bei der Mathematiknote und -kompetenz im Schuljahr 2014/15 und 0.31 bei der Deutschnote und Lesekompetenz im Schuljahr 2010/11. Der signifikant höhere Zusammenhang im Fach Mathematik im Vergleich zum Fach Deutsch ist unverkennbar.

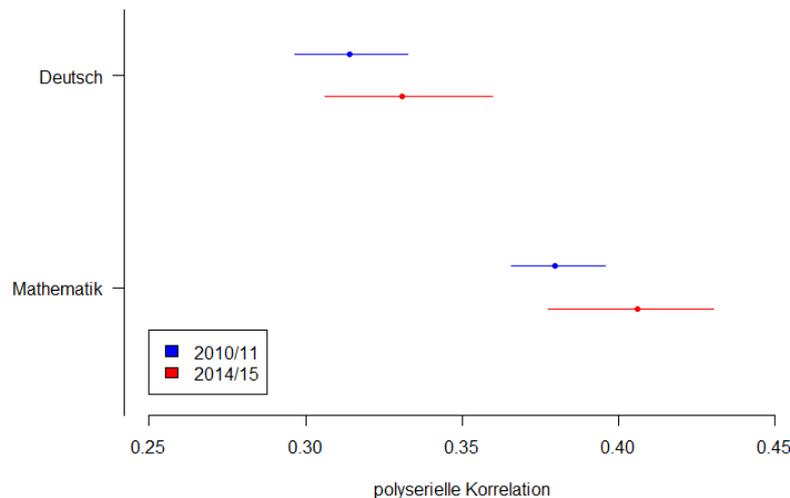


Abbildung 3: Kompetenz-Noten-Zusammenhang gemessen anhand polyserieller Korrelation für die Fächer Mathematik und Deutsch für die Schuljahre 2010/11 und 2014/15 (horizontale Linien markieren die 95% Konfidenzintervalle)

Auch eine Aufschlüsselung nach Schulform zeigt keine signifikanten Unterschiede zwischen beiden Kohorten (vgl. Abbildung 4). Zusammenfassend zeigen unsere Auswertungen, dass sich mit den NEPS-Daten der mit den PISA-Daten ermittelte Befund einer Zunahme des korrelativen Zusammenhangs zwischen Noten in den beiden Fächern Mathematik und Deutsch und den Mathematik- und Lesekompetenzen in den verschiedenen Schulformen nicht bestätigen lässt.

²⁰ Der korrelative Zusammenhang wurde durchgängig durch die Berechnung der polyseriellen Korrelation zwischen der ordinalen Variable „Note“ und der stetigen Größe „Kompetenzscore“ (Cox 1974) ermittelt. Die Noten wurden hierfür entsprechend umkodiert.

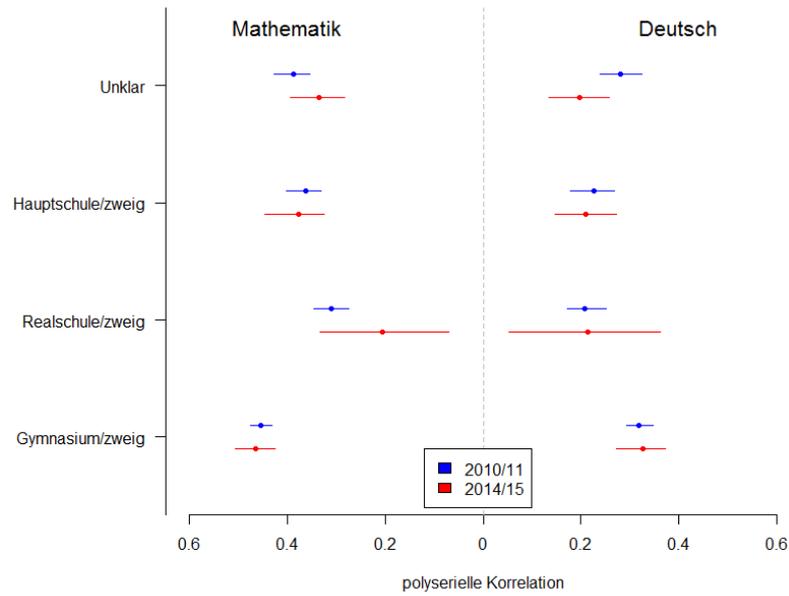


Abbildung 4: Kompetenz-Noten-Zusammenhang gemessen anhand polyserieller Korrelation für die Fächer Mathematik und Deutsch für die Schuljahre 2010/11 und 2014/15 nach Schulformen (horizontale Linien markieren die 95% Konfidenzintervalle)

Tabelle 2: Mittelwerte, Anteile und Standardabweichungen der im multivariaten Modell betrachteten Variablen (gewichtet und ungewichtet)

	Schuljahr 2010/11			Schuljahr 2014/15			Standardabweichung (gewichtet)
	Mittelwert bzw. Anteil (ungewichtet)	Standardabweichung (ungewichtet)	Mittelwert bzw. Anteil (gewichtet)	Mittelwert bzw. Anteil (ungewichtet)	Standardabweichung (ungewichtet)	Mittelwert bzw. Anteil (gewichtet)	
Deutschnote	2,874	0,832	2,914	0,857	2,761	0,838	2,863
Mathematiknote	2,980	1,024	3,019	1,047	2,893	1,033	3,023
Lesekompetenz	0,000	1,000	-0,004	1,000	0,000	1,000	-0,157
Mathematikkompetenz	0,000	1,000	0,003	0,997	0,000	1,000	-0,189
Alter	15,581	0,620	15,607	0,629	15,357	0,495	15,438
CASMIN Mutter							
1a, 1b, 2b	0,058	--	0,061	--	0,028	--	0,047
1c, 2a	0,296	--	0,298	--	0,234	--	0,232
2c	0,111	--	0,110	--	0,112	--	0,094
3a, 3b	0,091	--	0,093	--	0,115	--	0,084
Keine Information	0,444	--	0,438	--	0,510	--	0,544
Geschlecht							
weiblich	0,499	--	0,482	--	0,502	--	0,476
männlich	0,501	--	0,518	--	0,498	--	0,524
Migrationshintergrund							
ja	0,255	--	0,255	--	0,204	--	0,225
nein	0,745	--	0,745	--	0,796	--	0,775
Lehrerbildung							
verpflichtend	0,901	--	0,890	--	0,901	--	0,906
freiwillig	0,099	--	0,110	--	0,099	--	0,094
Schulsystem-Typ							
Mod. Strukturen	0,520	--	0,511	--	0,497	--	0,466
Mischform- traditionell	0,331	--	0,347	--	0,347	--	0,380
Mischform-modernisiert	0,149	--	0,143	--	0,156	--	0,154
Schulform							
Gymn./gymn. Zweig	0,360	--	0,362	--	0,482	--	0,377
Realschule/-zweig	0,228	--	0,266	--	0,236	--	0,259
Hauptschule/-zweig	0,220	--	0,195	--	0,060	--	0,162
Ausrichtung unklar	0,192	--	0,177	--	0,222	--	0,202
Anzahl SuS in Klasse	24,280	4,845	24,432	4,917	25,274	3,938	24,974
Anteil SuS in Klasse							
niedrigem soz. Status	32,342	27,754	32,403	27,745	20,949	19,172	20,888
Mädchen	46,531	13,655	46,783	14,233	49,093	12,937	47,903
Migrationshintergrund	23,865	23,651	24,225	24,344	18,360	18,438	18,923

Anmerkungen: SuS Schülerinnen und Schüler. †Für 24 von 12604 SuS im Schuljahr 2010/11 und 3 von 4007 SuS im Schuljahr 2014/15 liegen keine Testwerte zum logischen Schlussfolgern vor. Diese sind von der Analyse ausgenommen.

8. Multivariate Analysen

Im Nachfolgenden soll der korrelative Zusammenhang zwischen den Deutsch und Mathematiknoten in der 9. Jahrgangsstufe in den Schuljahren 2010/11 und 2014/15 im Hinblick auf mögliche Erklärungsursachen genauer untersucht werden. Hierfür leiten wir aus den zuvor besprochenen theoretischen Befunden ein geeignetes multivariates Regressionsmodell her. Anhand dieses Modells versuchen wir, den Einfluss der zuvor als relevant identifizierten Faktoren auf die Notenbildung empirisch abzuleiten. Hierzu gehen wir schrittweise vor. Das heißt, in einem ersten Schritt modellieren wir den Einfluss der Kompetenz auf die jeweilige Note (Modell 1). Dann nehmen wir die Fähigkeit zum schlussfolgernden Denken als Erklärungsvariable in das Modell auf (Modell 2). Wir erweitern in einem nächsten Schritt das Modell um die Individualmerkmale einer Schülerin bzw. eines Schülers (Alter, Geschlecht und Migrationshintergrund) sowie um den mütterlichen Bildungsabschluss (Modell 3). Anschließend (Modell 4) ergänzen wir Kontextinformationen auf Bundeslandebene (d.h. die Verpflichtung zur Lehrerfortbildung den Schulsystemtyp). Den potentiellen Einfluss des Klassenkontextes (wie zum Beispiel durch die Lehrkraft) modellieren wir in zwei weiteren Schritten. Einerseits durch die Hinzunahme einer latenten Größe (*random effect*), die Niveauunterschiede in der Notenverteilung auf Klassenebene abbilden soll (Modell 5). Andererseits fügen wir Kontextinformationen auf Klassenebene (konkret die Anzahl der Schülerinnen und Schüler in der Klasse, den Anteil an Schülerinnen und Schülern mit Migrationshintergrund und mit niedrigem sozio-ökonomischen Status sowie den Anteil an Mädchen in der Klasse) hinzu (Modell 6). Diese Art des hierarchischen Modellaufbaus dient der Identifizierung von Faktoren, die maßgeblich zur Notenbildung beitragen, wobei wir davon ausgehen, dass zuerst individuelle Leistungsmerkmale und der Bildungshintergrund, dann die Strukturmerkmale des Schulsystems (abgebildet durch die Bundeslandebene), und schließlich der Klassenkontext wesentlich sind. In unserer Analyse betrachten wir zunächst vier Fragestellungen, die wir nachfolgend zusammenführen: den Zusammenhang der Mathematiknote und der Mathematikkompetenz im Schuljahr 2010/11, den Zusammenhang der Mathematiknote und der Mathematikkompetenz im Schuljahr 2014/15, den Zusammenhang der Deutschnote und der Lesekompetenz im Schuljahr 2010/11 und den Zusammenhang der Deutschnote und der Lesekompetenz im Schuljahr 2014/15. Eine Gegenüberstellung der Ergebnisse der verschiedenen Modelle zeigt dann die zeitliche Veränderung in den Einflüssen der betrachteten Variablen auf die Notenbildung. Eine für das dargestellte Problem geeignete statistische Modellklasse ist ein *ordered probit* Regressionsmodell mit drei Hierarchieebenen (Schülerinnen und Schüler, Klasse bzw. Schule und Bundesland, siehe auch Abbildung 1). Wir nutzen (kalibrierte) Surveygewichte, um für ungleiche Auswahlwahrscheinlichkeiten und fehlende Werte auf Schul-, Klassen- und Schülerebene zu kompensieren.²¹

²¹ Diese Gewichte wurden eigens für die vorliegende Anwendung nach den Vorgaben in Steinhauer et al. (2015) berechnet.

Tabelle 3: Zusammenhang der Zeugnisnote (Halbjahreszeugnis) in Mathematik und Mathematikkompetenz im Schuljahr 2010/11, zusätzlich korrigiert für SES, Verpflichtung von Lehrerfortbildung, Schulform und Typ des Schulsystems im jeweiligen Bundesland (beta-Koeffizienten)

Variablen	Referenzkategorie	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Ebene							
SuS							
Kompetenz Mathematik							
Schlussfolgern (DGCF)		-0,416**	-0,369**	-0,356**	-0,356**	-0,470**	-0,534**
Alter		-	-0,100**	-0,089**	-0,089**	-0,134**	-0,137**
CASMIN				0,141**	0,142**	0,128**	0,171**
1a, 2a	1a, 1b, 2b	-	-	0,006	0,008	-0,043	-0,039
2c		-	-	0,072	0,072	-0,025	0,031
3a, 3b		-	-	0,026	0,025	-0,084	-0,110
Keine Information		-	-	0,023	0,023	-0,043	0,022
Geschlecht	weiblich						
männlich		-	-	0,037	0,036	0,039	0,030
Migrationshintergrund	nein						
ja		-	-	0,021	0,018	0,007	-0,022
Ebene							
Bundesland							
Lehrerfortbildung	verpflichtend						
freiwillig		-	-	-	-0,041	-0,088*	0,045
Schulsystem-Typ	Mod. Strukturen						
Mischform- traditionell		-	-	-	-0,017	0,028	0,010
Mischform- modernisiert		-	-	-	-0,024	-0,002	-0,042
Ebene							
Klasse							
Schulform	Andere						
Gymnasium		-	-	-	-	-	0,446**
Anzahl SuS in Klasse		-	-	-	-	-	0,017**
Anteil SuS in Klasse							
niedrigem Soz. Status		-	-	-	-	-	-0,001
Mädchen		-	-	-	-	-	-0,006**
Migrationshintergrund		-	-	-	-	-	0,001
Std.Dev Random Effekt auf Klassenebene		-	-	-	-	0,416	0,412
BIC		34862,77	34786,99	34756,90	34656,39	33864,59	12867,26
Stichprobengröße[†]		N=12580		N=4787			

Anmerkungen: SuS: Schülerinnen und Schüler. †Von den 15.629 SuS, die an SC4 Welle 1 teilgenommen haben, haben 12.580 SuS Angaben zur Kompetenzmessung, zur besuchten Schulform und Messungen des logischen Schlussfolgerns und 4787 (entspricht 38,1% der Fälle) zusätzlich Informationen auf Klassenebene bzgl. Anteil von SuS mit niedrigem sozio-ökonomischem Status, mit Migrationshintergrund, Anteil Mädchen in der Klasse und Anzahl SuS in der Klasse. Signifikanzcodes: ** mit $p < 0.05$ und * mit $0.05 < p < 0.1$.

Tabelle 4: Zusammenhang der Zeugnisnote (Halbjahreszeugnis) in Mathematik und Mathematikkompetenz im Schuljahr 2014/15, zusätzlich korrigiert für SES, Verpflichtung von Lehrerfortbildung, Schulform und Typ des Schulsystems im jeweiligen Bundesland (beta-Koeffizienten)

Variablen	Referenzkategorie	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Ebene							
SuS							
Kompetenz Mathematik		-0,359**	-0,318**	-0,336**	-0,338**	-0,504**	-0,664**
Schlussfolgern (DGCF)		-	-0,064**	-0,056	-0,051	-0,086*	-0,146*
Alter				0,065	0,068	0,171	0,349*
CASMIN	1a, 1b, 2b						
1c, 2a		-	-	-0,015	0,007	0,218	0,364
2c		-	-	-0,045	-0,044	-0,003	-0,011
3a, 3b		-	-	-0,052	-0,048	0,073	-0,257
Keine Information		-	-	0,009	0,034	0,046	-0,108
Geschlecht	weiblich						
männlich		-	-	0,321**	0,323**	0,286**	0,506**
Migrationshintergrund	nein						
ja		-	-	0,070	0,049	0,021	0,225*
Ebene							
Bundesland							
Lehrerfortbildung	verpflichtend						
freiwillig		-	-	-	0,014	0,093	-0,131
Schulsystem-Typ	Mod. Strukturen						
Mischform- traditionell		-	-	-	-0,047	-0,198**	-0,094
Mischform- modernisiert		-	-	-	0,166	-0,029	-0,289**
Ebene							
Klasse	Andere						
Gymnasium		-	-	-	-	-	0,884**
Anzahl SuS in Klasse		-	-	-	-	-	0,027**
Anteil SuS in Klasse		-	-	-	-	-	0,002
niedrigem Soz. Status		-	-	-	-	-	0,005
Mädchen		-	-	-	-	-	0,002
Migrationshintergrund		-	-	-	-	-	0,002
Std.Dev Random Effekt auf Klassenebene		-	-	-	-	0,736	0,861
BIC		11043,17	11039,12	11005,51	10999,10	10077,3	3734,75
Stichprobengröße[†]				N=4004			N=1512

Anmerkungen: SuS: Schülerinnen und Schüler. †Von den 5.586 SuS, die an SC3 Welle 6 teilgenommen haben, haben 4.004 SuS Angaben zur Kompetenzmessung, zur besuchten Schulform und Messungen des logischen Schlussfolgerns und 1512 (entspricht 38,1% der Fälle) zusätzlich Informationen auf Klassenebene bzgl. Anteil von SuS mit niedrigem sozio-ökonomischem Status, mit Migrationshintergrund, Anteil Mädchen in der Klasse und Anzahl SuS in der Klasse. Signifikanzcodes: ** mit $p < 0,05$ und * mit $0,05 < p < 0,1$.

Tabelle 5: Zusammenhang der Zeugnisnote (Halbjahreszeugnis) in Deutsch und Lesekompetenz im Schuljahr 2010/11, zusätzlich korrigiert für SES, Verpflichtung von Lehrerfortbildung, Schulform und Typ des Schulsystems im jeweiligen Bundesland (beta-Koeffizienten)

Variablen	Referenzkategorie	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Ebene							
SuS							
Kompetenz Deutsch		-0,331**	-0,338**	-0,267**	-0,266**	-0,304**	-0,345**
Schlussfolgern (DGCF)		-	0,017	0,012	-0,014	-0,005	-0,009
Alter				0,144**	0,145**	0,117**	0,073**
CASMIN	1a, 1b, 2b						
1c, 2a		-	-	-0,036	0,025	-0,005	0,067
2c		-	-	-0,098*	0,099*	-0,107	0,064
3a, 3b		-	-	-0,230**	-0,233**	-0,237**	-0,197
Keine Information		-	-	-0,005	0,008	-0,001	0,074
Geschlecht	weiblich						
männlich		-	-	0,432*	0,434**	0,468**	0,429**
Migrationshintergrund	nein						
Ja		-	-	0,117**	0,118**	0,144**	0,136**
Ebene							
Bundesland							
Lehrerfortbildung	verpflichtend						
Freiwillig		-	-	-	-0,013	0,048	-0,001
Schulsystem-Typ	Mod. Strukturen						
Mischform- traditionell		-	-	-	-0,142**	0,023	0,091
Mischform- Modernisiert		-	-	-	-0,029	-0,144**	-0,163
Ebene							
Klasse	Andere						
Gymnasium (GY)		-	-	-	-	-	-0,053
Anzahl SuS in Klasse		-	-	-	-	-	0,004
Anteil SuS in Klasse							
niedrigem Soz. Status		-	-	-	-	-	0,002
Mädchen		-	-	-	-	-	-0,003*
Migrationshintergrund		-	-	-	-	-	-0,003
Std.Dev Random Effekt auf		-	-	-	-	0,452	0,493
Klassenebene							
BIC		30670,14	30677,17	30071,29	30046,05	28927,98	11042,52
Stichprobengröße[†]				N=12580			N=4787

Anmerkungen: SuS: Schülerinnen und Schüler. †Von den 15.629 SuS, die an SC4 Welle 1 teilgenommen haben, haben 12.580 SuS Angaben zur Kompetenzmessung, zur besuchten Schulform und Messungen des logischen Schlussfolgerns und 4787 (entspricht 38,1% der Fälle) zusätzlich Informationen auf Klassenebene bzgl. Anteil von SuS mit niedrigem sozio-ökonomischem Status, mit Migrationshintergrund, Anteil Mädchen in der Klasse und Anzahl SuS in der Klasse. Signifikanzcodes: ** mit $p < 0.05$ und * mit $0.05 < p < 0.1$.

Tabelle 6: Zusammenhang der Zeugnisnote (Halbjahreszeugnis) in Deutsch und Lesekompetenz im Schuljahr 2014/15, zusätzlich korrigiert für SES, Verpflichtung von Lehrerfortbildung, Schulform und Typ des Schulsystems im jeweiligen Bundesland (beta-Koeffizienten)

Variablen	Referenzkategorie	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Ebene							
SuS							
Kompetenz Deutsch		-0,367**	-0,365**	-0,340**	-0,341**	-0,497**	-0,401**
Schlussfolgern (DGCF)		-	0,002	-0,023	-0,025	-0,013	-0,165
Alter				0,064	0,061	0,083	0,245*
CASMIN	1a, 1b, 2b						
1c, 2a		-	-	-0,207	-0,214	-0,553**	0,025
2c		-	-	-0,349**	-0,365**	-0,682**	-0,078
3a, 3b		-	-	-0,342*	-0,359*	-0,829**	-0,372
Keine Information		-	-	-0,212*	-0,222*	0,542**	-0,209
Geschlecht	weiblich						
männlich		-	-	0,686**	0,692**	0,728**	0,798**
Migrationshintergrund	nein						
Ja		-	-	-0,116	-0,110	-0,157	0,171
Ebene							
Bundesland							
Lehrerfortbildung	verpflichtend						
Freiwillig		-	-	-	0,081	-0,185	0,496**
Schulsystem-Typ	Mod. Strukturen						
Mischform- traditionell		-	-	-	-0,027	-0,070	-0,208
Mischform- Modernisiert		-	-	-	0,022	0,112	0,308*
Ebene							
Klasse	Andere						
Gymnasium (GY)		-	-	-	-	-	-0,171
Anzahl SuS in Klasse		-	-	-	-	-	-0,043**
Anteil SuS in Klasse		-	-	-	-	-	-0,009**
niedrigem Soz. Status		-	-	-	-	-	0,009*
Mädchen		-	-	-	-	-	0,010**
Migrationshintergrund		-	-	-	-	-	-
Std.Dev Random Effekt auf		-	-	-	-	0,701	0,775
Klassenebene							
BIC		9829,88	9838,16	9494,27	9515,69	8555,20	3314,86
Stichprobengröße[†]				N=4004			N=1512

Anmerkungen: SuS: Schülerinnen und Schüler. †Von den 5.586 SuS, die an SC3 Welle 6 teilgenommen haben, haben 4.004 SuS Angaben zur Kompetenzmessung, zur besuchten Schulform und Messungen des logischen Schlussfolgerns und 1512 (entspricht 38,1% der Fälle) zusätzlich Informationen auf Klassenebene bzgl. Anteil von SuS mit niedrigem sozio-ökonomischem Status, mit Migrationshintergrund, Anteil Mädchen in der Klasse und Anzahl SuS in der Klasse. Signifikanzcodes: ** mit $p < 0.05$ und * mit $0.05 < p < 0.1$.

9. Diskussion der Befunde

In den deskriptiven Analysen der Veränderung der Durchschnittsnoten konnte gezeigt werden, dass entgegen der Diskussionen, wie sie vor allem in der bildungspolitischen Öffentlichkeit stattfinden, keine signifikanten Notenverbesserungen sichtbar sind. Auch in der schulformspezifischen Aufgliederung zeigt sich, dass die Durchschnittsnote überall bei 3 liegt. Falls es tatsächlich zu einer Verbesserung etwa der Abiturnoten über die Zeit gekommen sein sollte, kann entsprechend angenommen werden, dass diese Veränderungen ausschließlich im Sekundarbereich II stattgefunden haben und gegebenenfalls weiter stattfinden.

Eine zentrale Vermutung unserer Analysen bezog sich vor allem auf die Zusammenhänge zwischen der Notenvergabepraxis und den Kompetenzen, wobei sich in den NEPS-Daten zeigt, dass die korrelativen Zusammenhänge insgesamt nicht sehr hoch sind und die auf den ersten Blick zu vermutende Zunahme des Zusammenhangs über die Zeit sich statistisch keineswegs absichern lässt. Die Korrelationswerte für das Schuljahr 2010/11 passen in ihrer Höhe (in diesem Fall für das Fach Deutsch und die Lesefähigkeiten) zu den im Bildungsbericht 2016 berichteten Werten auf Basis von PISA-Daten aus dem Jahr 2012 (vgl. Autorengruppe Bildungsberichterstattung 2016 S. 94). Allerdings finden sich entsprechend des Bildungsberichtes in den PISA-Daten auch klare Indizien für einen Veränderungsprozess hin zu einem stärker ausgeprägten Zusammenhang zwischen Kompetenzen und Noten. Auf Basis der hier verwendeten NEPS-Daten kann dieser Befund jedoch nicht bestätigt werden. Wenn es den unterstellten Veränderungsprozess dennoch gibt, so verläuft er wohl deutlich langsamer als die Darstellungen im Bildungsbericht vermuten lassen. Da im Bildungsbericht keine Konfidenzintervalle der PISA Befunde berichtet werden, kann die Unsicherheit in der Schätzung auch nicht beurteilt werden. Eine Verifizierung eines solchen Prozesses könnte somit erst durch die Hinzuziehung einer weiteren jüngeren Kohorte statistisch abgesichert werden. Entsprechend müssen die immer wieder diskutierten Annahmen über eine verstärkte Kompetenzbasiertheit des schulischen Bildungsbereichs und insbesondere der schulischen Leistungsbeurteilung an dieser Stelle zurückgewiesen werden.

Die multivariaten Erklärungsmodelle belegen den signifikanten Einfluss der bereichsspezifischen Kompetenzen auf die jeweiligen Fachnoten (Tabellen 2 bis 5). Die jeweiligen Noten in den Fächern Deutsch und Mathematik lassen sich also zu einem großen Teil durch die bereichsspezifischen Kompetenzen erklären, wobei dies im Fach Mathematik eindeutig stärker ausgeprägt ist als im Fach Deutsch. Dieser Umstand ist nicht zuletzt dem deutlich breiteren Profil des Faches Deutsch im Vergleich zur gemessenen Lesekompetenz geschuldet. Die Fähigkeiten zum schlussfolgernden Denken (als ein Proxy für fluide Intelligenz) zeigen im Fach Mathematik im jeweiligen Endmodell (Modelle 6 in Tab. 2 und 3) signifikante Effekte auf die Note. Ein derartiger Einfluss von Intelligenz auf Schulleistungen wurde bereits in einer Vielzahl von Studien gezeigt (vgl. etwa Deary et al. 2007; Rost 2009).

Der Bildungshintergrund der Schülerinnen und Schüler besitzt im Fach Mathematik keinerlei Erklärungskraft, während sich in den Analysen zu den Noten in Deutsch zeigt, dass dieser durchaus vorhanden ist, seinen signifikanten Einfluss jedoch bei Einbezug der *Klassenebene* und damit auch klassenkompositorischer Elemente verliert. Die entlang sozialer Herkunftsmerkmale durch eine doch deutliche Homogenität geprägten Schulformen in Deutschland lassen darüber hinaus direkt wirksame Bildungsherkunftseffekte auch nicht erwarten. Im jeweiligen Endmodell (Modell 6) verschiebt sich jedoch der Effekt der Bildungsherkunft der Schülerinnen und Schüler für die Note in Deutsch insbesondere im Schuljahr 2014/15 zum Teil auf die Ebene der klassenkompositorischen Merkmale wie die

soziale und ethnische Zusammensetzung der jeweiligen Schulklasse. In der Klassenkomposition spiegeln sich also die Herkunftseffekte wieder, was wegen der sozialen Segregation des gegliederten Schulsystems in Deutschland so auch erwartbar ist (vgl. Szczesny/Watermann 2011).

Betrachtet man die Veränderungen über die Zeit, so zeigt sich, dass neben der deutlichen Zunahme des Einflusses der jeweiligen Kompetenz auf die Fachnote, es hinsichtlich des Erklärungsfaktors Gymnasialklasse ebenfalls zu einer Verstärkung eines allerdings bereits 2010/11 gegenläufigen Befundes kommt. So zeigt die Zugehörigkeit zu einer Gymnasialklasse im Fach Mathematik unter Kontrolle der Mathematikkompetenzen einen negativen Effekt auf die Note, der sich im Kohortenvergleich zudem deutlich verstärkt, während sich dieser Effekt im Fach Deutsch 2010/11 und 2014/15 in umgekehrter Wirkrichtung, aber zu beiden Zeitpunkten auf einem nicht-signifikanten Niveau zeigt.

Eines der zentralen Ergebnisse der multivariaten Mehrebenenanalyse zeigt sich zudem im nicht erklärten Klasseneffekt, der insgesamt sehr hoch ist und über die Zeit hinweg sogar noch zunimmt und dies unabhängig vom Fach. Die Kontrolle der jeweiligen Klassenkomposition (Schulformzugehörigkeit der Klasse sowie die soziale und ethnische Zusammensetzung wie auch der Anteil an Mädchen in der Klasse) ändern an der Stärke dieses Effekts wenig. Hier scheint es sich um einen Lehrkräfteeffekt zu handeln, welcher unter Kontrolle sämtlicher Modellvariablen auf den unterschiedlichen Ebenen einen sehr hohen Anteil an der Gesamterklärung aufweist. Insgesamt beobachten wir von 2010/11 auf 2014/15 zudem ein starkes Anwachsen des Klasseneffektes sowohl auf die Mathematik- als auch die Deutschnote (von einer Standardabweichung (SD) von 0,41 in Mathematik im Schuljahr 2010/11 auf eine SD von 0,86 im Schuljahr 2014/15 und von einer SD von 0,49 in Deutsch im Schuljahr 2010/11 auf eine SD von 0,75 im Schuljahr 2014/15), wobei der Zuwachs im Fach Mathematik etwas prononcierter ausfällt.²² Die Lehrkraft ist demnach in ihrer jeweiligen Benotungspraxis zentral für die Notenvarianz in den beiden Fächern und dies unabhängig von den schüler-, klassenkompositions- und auch strukturbezogenen Merkmalen.

In Bezug auf die Mathematiknoten zeigt sich, dass die schulformspezifische Klassenzugehörigkeit (dadurch operationalisiert, ob eine Schülerin oder ein Schüler ein Gymnasium bzw. einen gymnasialen Zweig besucht) unter Kontrolle der Kompetenzen einen deutlichen Effekt auf die Note aufweist. Die im Übergang zur Sekundarstufe I vorgenommene leistungsbezogene Selektion der Schülerinnen und Schüler bei gleichzeitiger Schulformunabhängigkeit des Benotungsschemas lässt einen solchen Effekt zumindest wahrscheinlich sein. Umso erstaunlicher ist jedoch, dass dieser Effekt im Fach Deutsch in keiner der beiden Kohorten gefunden werden kann. Dies legt die Vermutung nahe, dass der Einfluss der Lehrkraft im Fach Deutsch unabhängig von der konkreten Schulform stärker wirkt als im Fach Mathematik. Auf der schulstrukturellen Ebene wird deutlich, dass die traditionellen Mischformen (siehe Tabelle 1) vor allem im Gymnasialbereich positive Effekte auf die Noten zeigen (in Mathematik im Schuljahr 2014/15), während sich bei modernisierten Mischformen insbesondere im nicht-gymnasialen Bereich entsprechende Effekte auf Noten nachweisbar sind (im Schuljahr 2014/15 in beiden Fächern, wenn das Gymnasium kontrolliert wird). Blickt man entlang der Bundesländer, die sich diesem Typus zuordnen lassen auf die

²² Wir beobachten zwischen 2010/11 und 2014/15 keinen signifikanten Anstieg des Anteils der Varianz zwischen den Klassen an der Gesamtvarianz der Mathematik- und Deutschnoten. Somit sind die beobachteten Anstiege des Klasseneffektes nicht durch Änderungen in der Klassenkomposition zu begründen.

zwischen 2010 und 2015 stattgefundenen Reformen, so zeigt sich, dass hier entweder neue Schulformen eingeführt wurden, wie etwa die Gemeinschaftsschule, oder aber Schulformen zusammengelegt wurden, wie etwa die in Niedersachsen im Schuljahr 2011/12 eingeführte Oberschule als Fusion von Haupt- und Realschule. Bildungspolitische Maßnahmen haben also durchaus Effekte bis hinab zur Ebene der Benotung von Schülerinnen und Schülern.

Ebenfalls bemerkenswert ist die deutliche Ausprägung eines Geschlechtereffektes sowohl auf die Deutschnote (beide Erhebungsjahre) als auch die Mathematiknote (Schuljahr 2014/15). Dass dieser Effekt zumindest im Fach Deutsch bekannt ist wurde jüngst wieder von Kessels und Heyder (2017) gezeigt. In Mathematik gibt es sehr unterschiedliche Ergebnisse über das Vorliegen eines Geschlechtereffektes in der Forschung (vgl. für einen Überblick Voyer/Voyer 2014). Hier ist jedoch durchaus kritisch anzumerken, dass in den Analysen zur Bestimmung eines Geschlechtereffektes nicht immer individuelle Kompetenzunterschiede ausreichend kontrolliert werden. Was sich im detektierten Geschlechtereffekt nun tatsächlich ausdrückt, darüber besteht keine abschließende Einigkeit. Dass es sich hierbei um einen Beurteilungseffekt auch unter Einbezug etwaiger geschlechtsspezifischer Stereotype handelt, darauf deuten die Befunde von Lorenz et al. (2016) hin, die einen solchen Geschlechtereffekt unter Kontrolle der entsprechenden Kompetenzen im Hinblick auf die Leistungserwartung von Lehrkräften identifizieren konnten.²³ In dieser Studie wurde darüber hinaus die von den Schülerinnen und Schülern selbst berichtete *Anstrengungsbereitschaft* in die Analysen einbezogen, die ebenfalls einen bedeutsamen Effekt auf die Leistungserwartungen durch die Lehrkräfte zeigten. Inwiefern sich hier selbstberichtete und durch die Lehrkraft wahrgenommene Anstrengungsbereitschaft decken, dürfte sicherlich für spätere Analysen durchaus interessant sein.

10. Fazit und Limitationen

Ausgangspunkt der durchgeführten Analysen war die Frage ob und inwieweit sich im Vergleich zweier Schulkohorten Veränderungen im Zusammenhang zwischen individuellen Leistungsfähigkeiten und der sich in Noten ausdrückenden Leistungsbeurteilung abbilden, über die sowohl innerhalb der fachwissenschaftlichen wie auch bildungspolitischen Diskussionen mit teils durchaus kontroversen Positionen verhandelt wird. Diese Frage wurde eingebettet in eine Debatte, die unter dem Stichwort der Bildungsgerechtigkeit, die Prozesse innerhalb des Bildungssystems sowohl auf konzeptioneller wie auch empirischer Ebene begleitet und in der es beispielweise um Fragen der aktuellen aber auch zukünftigen Ausbildungs- bzw. Studierfähigkeiten geht. Auf konzeptioneller Ebene wurden hierbei zwei Schwerpunkte in den Debatten benannt. Einerseits die Diskussionen über den *richtigen* Bildungsbegriff und andererseits die diagnostischen Beiträge zu empirischen Trends insbesondere im Bereich der Noten- und Kompetenzentwicklung. Im ersten Diskussionsstrang finden sich auch Diagnosen zu einer immer stärkeren (nur) auf Kompetenzentwicklung abzielenden Umstrukturierung des Bildungssystems. Die in diesem Kontext stark gemachten Thesen einer Veränderung des deutschen Schulsystems in Richtung einer verstärkten Orientierung an schulübergreifenden Kompetenzen lässt sich jedoch mit den hier verwendeten Daten so nicht bestätigen. Sicherlich ist nicht auszuschließen, dass ein solcher Prozess erst dann sichtbar werden würde, wenn man einen deutlich weiter gesteckten zeitlichen Rahmen ansetzt.

²³ Die diesem Beitrag zugrundeliegenden Daten beziehen sich jedoch ausschließlich auf Schulen in NRW.

Unsere Analysen der NEPS-Daten zeigen hingegen eine deutliche Ausprägung und Zunahme der Bedeutsamkeit des Klasseneffektes auf die Notengebung und das unter Kontrolle sowohl der individuellen (kognitiven) Kompetenzen wie auch der klassenkompositorischen Merkmale der Schülerinnen und Schüler. Dieses Resultat wird von uns dahingehend interpretiert, dass sich die im konzeptionellen Teil diskutierte Frage nach dem Bildungsverständnis zumindest soweit empirisch aufklären lässt, als dass man konstatieren muss, dass es jenseits der über Kompetenzen gemessenen Leistungsfähigkeit Beurteilungsmerkmale gibt, welche in die Notengebung Eingang finden. Damit zeigt sich im Sinne der von Heller (1974) vorgenommenen Unterscheidung von Lernprodukt und Lernprozess, das Letzterem durchaus eine bedeutsame Funktion in der Leistungsbeurteilung zukommt. Ob und inwieweit in diese Beurteilungen bildungsrelevante Merkmale im Sinne von Bildung als Persönlichkeitsentwicklung Eingang finden und sie damit kongruent zu den Bildungszielen sind, ist eine nach wie vor offene Frage. Das Ausmaß des Klasseneffektes, der in den Analysen trotz Kontrolle zentraler Variablen einen gewichtigen Beitrag an der Gesamterklärung der Varianz zwischen den Fachnoten aufweist, macht deutlich, dass es die Klassenebene ist, auf der sich im Sinne der eingangs formulierten Thesen die Lebenschancen von Schülerinnen und Schülern entscheiden.

Hier zeigen sich aber auch unverkennbar die Grenzen der Daten des Nationalen Bildungspanels, die mit ihrer starken Fokussierung auf das Individuum in seinem individuellen Bildungsverlauf und seiner individuellen Kompetenzentwicklung an dieser Stelle wenig Informationen bereitstellen, mit denen dieser Effekt hinsichtlich seiner Komposition analysiert werden könnte. Dass etwa die von den Schülerinnen und Schülern selbstberichteten Lernmotivationen oder auch Anstrengungsbereitschaften nicht die relevante Information darstellen, sondern vielmehr die diesbezüglich durch die Lehrkraft wahrgenommenen Ausprägungen dieser Merkmale erklärende Kraft haben, wurde unter Hinweis auf die Befunde von Maaz, Baeriswyl und Trautwein (2011) bereits ausgeführt. Das impliziert, dass der ausgemachte Klasseneffekt nur durch eine systematische Erfassung dieser Einschätzungen auf Ebene der Lehrkraft weiter aufgeschlüsselt werden könnte. Es ist also nach wie vor so, dass das Zustandekommen der Noten zwar durchaus mit den Leistungsfähigkeiten der Schülerinnen und Schülern zusammenhängen. In deutlich gewichtigerer Weise gehen jedoch darüber hinaus andere Beurteilungskriterien in die Benotung ein. Nähere Aufklärung, was genau auf Klassenebene passiert, könnte nur durch ein entsprechendes Monitoring der Lehrkräfte erreicht werden, insbesondere weil diese Ebene für die Umsetzung von Bildungszielen und das Erreichen von Bildungsstandards zentral ist.

In diesen Kontext gehören auch die Befunde zu dem deutlichen Geschlechtereffekt, der durchaus bekannt ist. Jedoch sollte dieser Effekt in einer zukünftigen Studie hinsichtlich seiner möglichen individuellen Ursachen, aber auch einer sich darin möglicherweise widerspiegelnden geschlechtsspezifischen Lehrer-Schüler-Interaktion analysiert werden. Hierzu gehört etwa die genauere Untersuchung der sozialen Kompetenz wie auch der (fachspezifischen) Selbstkonzepte von Schülerinnen und Schülern.

Insgesamt zeigen unsere Auswertungen, dass die in Teilen der Öffentlichkeit verlaufenden Diskussionen an wichtigen Stellen an der empirischen Situation vorbeigehen. Es steht jedoch außer Frage, dass es im deutschen Schulsystem wichtige bisher nicht ausreichend beleuchtete Entwicklungen gibt, die in zukünftigen Studien in den Blick genommen werden müssen.

Literatur

- Adams, Raymond J. (2005): Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31: 162-172.
- Arbeitskreis DQR (2011): *Deutscher Qualifikationsrahmen für lebenslanges Lernen*. Berlin.
- Artelt, Cordula, Brunner, Martin, Schneider, Wolfgang, Prenzel Manfred & Neubrand, Michael (2003): Literacy oder Lehrplanvalidität? — Ländervergleiche auf der Basis lehrplanoptimierter PISA-Tests, in: Jürgen Baumert et al. (Hrsg.), *PISA 2000 — Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Wiesbaden: VS Verlag. S. 77–108.
- Autorengruppe Bildungsberichterstattung (2016): *Bildung in Deutschland 2016*. Bielefeld: Bertelsmann Verlag.
- Baumert, Jürgen, Stanat, Petra & Watermann, Rainer (2006): Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus, in: Jürgen Baumert, Petra Stanat & Rainer Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen*. Wiesbaden: VS Verlag. S. 95–188.
- Bayer, Michael (2015): Bildung, Leistung und Kompetenz, in: Andreas Lange, Christine Steiner, Sabine Schutter & Herwig Reiter (Hrsg.), *Handbuch Kindheits- und Jugendsoziologie*. Wiesbaden: Springer VS. S. 1–13.
- Becker Dominik & Birkelbach Klaus (2010): Intelligenz und Schulleistung als Kontextmerkmale: Big-Fish-Little-Pond-oder Reflected-Glory-Effekt? Eine Mehrebenen-Analyse von Lehrerurteilen, in: Tilo Beckers, Klaus Birkelbach, Jörg Hagenah & Ulrich Rosar (Hrsg.), *Komparative empirische Sozialforschung*. Wiesbaden: Springer VS. S. 113–141.
- Becker, Rolf & Hadjar, Andreas (2017): Meritokratie – Zur gesellschaftlichen Legitimation ungleicher Bildungs-, Erwerbs- und Einkommenschancen in modernen Gesellschaften, in: Rolf Becker (Hrsg.), *Lehrbuch der Bildungssoziologie*. Wiesbaden: Springer VS. S. 33–62.
- Blossfeld, Hans-Peter, Roßbach, Hans-Günther & von Maurice, Jutta (2011): *Education as a Lifelong Process – The German National Educational Panel Study (NEPS)*. Zeitschrift für Erziehungswissenschaft: Sonderheft 14.
- Brauns, Hildegard & Steinmann, Susanne (1997): Educational reform in France, West-Germany, the United Kingdom and Hungary: Updating the CASMIN educational classification (ZUMA Nachrichten No. 23). Mannheim: MZES.
- Brühwiler, Christian, Helmke, Andreas & Schrader, Friedrich-Wilhelm (2017): Determinanten der Schulleistung, in: Martin K.W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion*. Wiesbaden: Springer VS. S. 291–314.
- Brunner, Martin, Lang, Frieder R. & Lüdtke, Oliver (2014): *Erfassung der fluiden kognitiven Leistungsfähigkeit über die Lebensspanne im Rahmen der National Educational Panel Study: Expertise* (NEPS Working Paper 42). Bamberg: NEPS.
- Cox, N. R. (1974). Estimation of the correlation between a continuous and a discrete variable. *Biometrics* 30(1): 171–178.
- Deary, Ian J., Strand, Steve, Smith, Pauline & Fernandes, Cres (2007): Intelligence and educational achievement. *Intelligence* 35(1): 13–21.

- DeSeCo (2005): *Definition und Auswahl von Schlüsselkompetenzen. Zusammenfassung*. OECD. <http://www.oecd.org/pisa/35693281.pdf>. Zugegriffen am 13.02.2018.
- Dietrich, Fabian & Fricke, Uwe (2013): Reproduktion sozialer Ungleichheit im Prozess schulischer Leistungsbewertung, in: Fabian Dietrich, Martin Heinrich & Nina Thieme (Hrsg.), *Bildungsgerechtigkeit jenseits von Chancengleichheit*. Wiesbaden: Springer VS. S. 259-292.
- Duchhardt, Christoph, & Gerdes, Annkathrin (2013): *NEPS Technical Report for Mathematics – Scaling results of Starting Cohort 4 in ninth grade* (NEPS Working Paper No. 22). Bamberg: NEPS.
- Fend, Helmut (1980): *Theorie der Schule*. Göttingen: Urban und Schwarzenberg.
- Fend, Helmut (2008): Dimensionen von Qualität im Bildungswesen. *Zeitschrift für Pädagogik* 54: 190-209.
- Fischer, Luise, Rohm, Theresa, Gnamb, Timo, & Carstensen, Claus H. (2016): *Linking the data of the competence tests* (NEPS Survey Paper No. 1). Bamberg: LfBi.
- Fleischer, Jens, Koeppen, Karoline, Kenk, Martina, Klieme, Eckhard & Leutner, Detelf (2013): Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. *Zeitschrift für Erziehungswissenschaften* 16(Supplement 1): 5–22.
- Frey, Kristina A. (2013): *Soziale Kompetenz. Eine Fragebogenerfassung in der Grundschule*. Münster: Waxmann.
- Frey, Kristina A. & Bonsen, Martin (2013): Zur Diskussion von Kopfnoten als „Chance“ für Schüler mit schlechten Fachnoten, in: Knut Schwippert, Martin Bonsen & Nils Berkemeyer (Hrsg.), *Schul- und Bildungsforschung. Diskussionen, Befunde und Perspektiven*. Münster: Waxmann, S. 91–109.
- Fussangel, Kathrin, Rürup, Matthias & Gräsel, Cornelia (2010): Lehrerfortbildung als Unterstützungssystem. In: Herbert Altrichter & Katharina Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*. Wiesbaden: Springer VS. S. 327–354.
- Gehrer, Karin, Zimmermann, Stefan, Artelt, Cordula, & Weinert, Sabine (2012): *The assessment of reading competence (Including sample items for grade 5 and 9) (Research Data)*. Bamberg: NEPS.
- Green, Samuel B. & Yang, Yanyung (2009): Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika* 74(1): 155–167.
- Grunert, Cathleen (2012): *Bildung und Kompetenz*. Wiesbaden: Springer VS.
- Gruschka, Andreas (2006): Bildungsstandards oder das Versprechen, Bildungstheorie in empirischer Bildungsforschung aufzuheben. *Pädagogische Korrespondenz* 35: 5–22.
- Haberkorn, Kerstin, Pohl, Steffi, Hardt, Katinka, & Wiegand, Elena (2012): *NEPS Technical Report for Reading – Scaling Results of Starting Cohort 4 in Ninth Grade (NEPS Working Paper No. 16)*. Bamberg: NEPS.
- Hadjar, Andreas (2008): *Meritokratie als Legitimationsprinzip*. Wiesbaden: Springer VS.
- Hattie, John (2009): *Visible learning. A synthesis of over 800 metaanalyses relating to achievement*. London: Routledge.

- Helbig, Marcel & Morar, Tatiana (2017): *Warum Lehrkräfte sozial ungleich bewerten*. Discussion Paper P 2017-005. Berlin: WZB.
- Helbig, Marcel & Nikolai, Rita (2015): *Die Unvergleichbaren. Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Bad Heilbrunn: Klinkhardt.
- Heller, Kurt (1974): Einleitung und Übersichtsreferat, in: Kurt Heller (Hrsg.), *Leistungsbeurteilung in der Schule*. Heidelberg: Quelle & Meyer. S. 13–36.
- Henry-Huthmacher, Christine & Hoffmann, Elisabeth (2016): *Ausbildungsreife & Studierfähigkeit*. Berlin: Konrad-Adenauer-Stiftung.
- Ingenkamp, Karlheinz (1967): *Schulleistungen - damals und heute: Meinungen und Untersuchungen zur Veränderung des Leistungsniveaus unserer Schuljugend*. Weinheim: Beltz.
- Ingenkamp, Karlheinz (1971): *Die Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz.
- Kessels, Ursula & Heyder, Anke (2017): Die Wertschätzung schulischer Anstrengung als Mediator von Geschlechtsunterschieden in Noten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 49 (2): 86–97.
- Klafki, Wolfgang (1985): *Neuere Studien zur Bildungstheorie und Didaktik*. Weinheim: Beltz.
- Klieme, Eckhard, Avenarius, Hermann, Blum, Werner, Döbrich, Peter, Gruber, Hans ... & Vollmer, Helmut J. (2003): *Zur Entwicklung nationaler Bildungsstandards*. Bildungsforschung Band 1. Berlin: BMBF.
- Kölller, Olaf & Baumert, Jürgen (2002): Das Abitur – immer noch ein gültiger Indikator für die Studierfähigkeit? *Aus Politik und Zeitgeschichte* B26: 12–19.
- Lorenz, Georg, Gentrup, Sarah, Kristen, Cornelia, Stanat, Petra, & Kogan, Irena (2016): Stereotype bei Lehrkräften?: Eine Untersuchung systematisch verzerrter Lehrererwartungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 68(1): 89–111.
- Maaz, Kai, Baeriswyl, Franz & Trautwein, Ulrich (2011): *Herkunft zensiert? Leistungsdiagnostik und soziale Ungleichheiten in der Schule*. Vodafone Stiftung Deutschland.
- Münch, Richard (2012): Mit dem PISA-Express in die globale Wissensgesellschaft, in: Hans-Georg Soeffner (Hrsg.), *Transnationale Vergesellschaftungen*. Wiesbaden: Springer VS. S. 695–706.
- Neumann, Irene, Duchhardt, Christoph, Grüßing, Meike, Heinze, Aiso, Knopp, Eva, & Ehmke, Timo (2013): Modeling and assessing mathematical competence over the lifespan. *Journal for Educational Research Online* 5(2): 80–109.
- Olczyk, Melanie, Will, Gisela & Kristen, Cornelia (2014): *Immigrants in the NEPS: Identifying generation status and group of origin (NEPS Working Paper No. 41a)*. Bamberg: LIfBi.
- Pietschnig, Jakob & Voracek, Martin (2015): One Century of Global IQ Gains: A Formal Meta-Analysis of the Flynn Effect (1909–2013). *Perspectives on Psychological Science* 10(3): 282–306.
- Pohl, Steffi & Carstensen, Claus. H. (2012): *NEPS Technical Report – Scaling the Data of the Competence Tests (NEPS Working Paper No. 14)*. Bamberg: NEPS.

- Richter, Eric, Richter, Dirk & Marx, Alexandra (2018): Was hindert Lehrkräfte an Fortbildungen teilzunehmen? Eine empirische Untersuchung der Teilnahmebarrieren von Lehrkräften der Sekundarstufe I in Deutschland. *Zeitschrift für Erziehungswissenschaft*.
- Rost, Detlef H. (2009): *Intelligenz: Fakten und Mythen*. Weinheim: Beltz.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005): *Bildungsstandards der Kultusministerkonferenz Erläuterungen zur Konzeption und Entwicklung*. Darmstadt: Wolters Kluwer. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Konzeption-Entwicklung.pdf
- Solga, Heike (2005): Meritokratie – die moderne Legitimation ungleicher Bildungschancen, in: Peter A. Berger & Heike Kahlert (Hrsg.), *Institutionalisierte Ungleichheiten? Stabilität und Wandel von Bildungschancen*. Weinheim/München: Juventa. S. 19–38.
- Steinhauer, Hans W., Aßmann, Christian, Zinn, Sabine, Goßmann, Solange, & Rässler, Susanne (2015): Sampling and weighting cohort samples in institutional contexts. *AStA Wirtschafts-und Sozialstatistisches Archiv* 9(2): 131–157.
- Stevenson, David L. & Baker, David P. (1987): The Family-School Relation and the Child's School Performance. *Child Development* 58(5): 1348–1357.
- Szczesny, Markus & Watermann, Rainer (2011): Differenzielle Einflüsse von Familie und Schulform auf Leseleistung und soziale Kompetenzen. *Journal for educational research online* 3 (1): 168–193.
- Trautwein Ulrich, Lüdtke Oliver, Becker Michael, Neumann Marko & Nagy Gabriel (2008): Die Sekundarstufe I im Spiegel der empirischen Bildungsforschung: Schulleistungsentwicklung, Kompetenzniveaus und die Aussagekraft von Schulnoten, in: Elisabeth Schlemmer & Herbert Gerstberger (Hrsg.), *Ausbildungsfähigkeit im Spannungsfeld zwischen Wissenschaft, Politik und Praxis*. Wiesbaden: Springer VS. S. 91-107.
- vbw – Vereinigung der Bayrischen Wirtschaft (2015): *Bildung mehr als Fachlichkeit*. Münster: Waxmann.
- Velten, Stefanie & Schnitzler, Annalisa (2011): Prognose von Ausbildungserfolg- Welche Rolle spielen Schulnoten und Einstellungstests? *BiBB BWP* 6/2011.
- Voyer, Daniel & Voyer, Susan D. (2014): Gender Differences in Scholastic Achievement: A Meta-Analysis. *Psychological Bulletin*. Advance online publication. <http://dx.doi.org/10.1037/a0036620>.
- Wilhelm, Oliver (2005): Measuring reasoning ability. In Oliver Wilhelm & Randall W. Engle (Hrsg.), *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage Publications. S. 373–392.
- Youniss, James & Smollar, Jacqueline (1985): *Adolescent relations with mothers, fathers and friends*. Chicago: University of Chicago Press.
- Zinn, Sabine, Würbach, Ariane, Steinhauer, Hans W. & Hammon, Angelina (2018): *Attrition and Selectivity of the NEPS Starting Cohorts: An Overview of the Past 8 Years (NEPS Survey Paper No. 34)*. Bamberg: LfBi.