Vorteile kognitiver Homogenisierung?

Replikation einer Studie von Hartmut Esser und Julian Seuring zu den Auswirkungen von strikter Leistungs differenzierung auf Schulleistungen und Bildungsungleichheit

Georg Lorenz¹

Sarah Lenz²

Januar 2022

¹ Georg Lorenz

Institut zur Qualitätsentwicklung im Bildungswesen (IQB) an der Humboldt-Universität zu Berlin

Unter den Linden 6

10099 Berlin, Germany

+(49) 0172 4517965

Georg.Lorenz@iqb.hu-berlin.de

² Sarah Lenz

Institut zur Qualitätsentwicklung im Bildungswesen (IQB) an der Humboldt-Universität zu Berlin

Unter den Linden 6

10099 Berlin, Germany

Sarah.Lenz@iqb.hu-berlin.de

Zusammenfassung

Inwieweit Leistungsdifferenzierung (bzw. *tracking*) schulische Leistungen und Bildungsungleichheit beeinflusst, wird kontrovers diskutiert. Ziel der vorliegenden Studie ist es, die Ergebnisse einer Arbeit von Hartmut Esser und Julian Seuring zu dieser Fragestellung anhand von Daten der IQB-Bildungstrendstudien zu replizieren. Theoretisch werden weiterführend nicht-intendierte Folgen von Leistungsdifferenzierung diskutiert. Empirisch bestätigt sich die Grundannahme einer stärkeren kognitiven Homogenität von Schulklassen in strikt differenzierenden Bundesländern nur teilweise. Außerdem lassen sich positive Zusammenhänge zwischen kognitiver Homogenität und schulischen Leistungen nicht replizieren. Vielmehr scheinen differenzielle Kompositionseffekte und Peerprozesse den positiven Effekten der Unterrichtstrukturierung in kognitiv homogenen Schulkontexten entgegenzuwirken. Wie in der Originalstudie variiert soziale Ungleichheit jedoch nicht zwischen den Systemen. Die Ergebnisse deuten darauf hin, dass Leistungsunterschiede zwischen den Sekundarschulsystemen teilweise auf unbeobachtete Bundesländerunterschiede zurückgehen könnten.

1 Einleitung

Welche Rolle die institutionelle Ausgestaltung des Bildungssystems für die Reproduktion von sozialer Ungleichheit spielt, ist eine zentrale soziologische Fragestellung. Ihr widmet sich eine im Band 49 (Heft 5-6) dieser Zeitschrift erschienene Studie von Hartmut Esser und Julian Seuring (2020). Die Autoren untersuchen die Folgen der frühen leistungsbezogenen Differenzierung der Schülerschaft innerhalb des Sekundarschulsystems in Deutschland für schulische Leistungen von Schüler*innen und für soziale Bildungsungleichheit. Die Aufteilung der Schülerschaft auf verschiedene Sekundarschularten gemäß ihrer in der Primarstufe erbrachten Leistungen wird in der internationalen Literatur auch als (*between-school ability*) *tracking* bezeichnet. Ihr liegt die Annahme zugrunde, dass die kognitive Homogenisierung der Schülerschaft durch die Anpassung des Unterrichts an die kognitiven Fähigkeiten der Schüler*innen deren individuelle Förderung erleichtert (vgl. Hallinan 1994; Sørensen 1970).

Esser und Seuring beschreiben ihre Arbeit als „korrigierende Replikation“ auf eine sogenannte „Standardposition“, der zufolge leistungsbezogene Differenzierung nach meritokratischen Prinzipien soziale Ungleichheiten in der Bildungsbeteiligung und in den Leistungen eher verstärkt (z. B. Gamoran 2009). Im Gegensatz zu dieser Position gehen Esser und Seuring davon aus, dass Übergänge von der Primar- in die Sekundarstufe in strikt leistungsdifferenzierten Sekundarschulsystemen zum Beispiel aufgrund von verbindlichen Übergangsempfehlungen stärker an das in der Grundschule erworbene Wissen und die kognitiven Fähigkeiten und dafür weniger stark an die soziale Herkunft geknüpft sind. Dadurch sollte Leistungsdifferenzierung mit einer stärkeren kognitiven Homogenisierung der Schülerschaft einhergehen, während die soziale Durchmischung unverändert bleibt.

Aufbauend auf diesen grundlegenden Argumenten zu den Folgen von *tracking* für die Zusammensetzung bzw. *Komposition* von Schulklassen¹ testen die Autoren zwei theoretische Kernannahmen bzw. Mechanismen. Diese beziehen sich primär auf die Auswirkungen der kognitiven Homogenität von Schulklassen auf den Erwerb schulischer Leistungen (also auf

¹ Als Kompositionsmerkmale von Schulklassen werden im Folgenden immer sowohl das Niveau in einem Merkmal (z. B. den kognitiven Fähigkeiten) als auch die Homogenität in diesem Merkmal verstanden.

sog. *Kompositionseffekte*)²: 1) Kognitive Homogenität auf der Klassenebene hat positive Effekte auf Lernerfolge, wodurch die Leistungen in strikt leistungsdifferenzierten Bildungssystemen im Schnitt höher sind als in liberaleren Systemen. 2) Kognitive Homogenität in Schulklassen wirkt bei strikter leistungsbezogener Differenzierung noch effizienter auf die Leistungsentwicklung als in weniger (strikt) differenzierten Systemen.

Neben einer theoretischen Systematisierung erweitern die Autoren die Debatte auch auf der empirisch-methodischen Ebene. So nutzen sie vorgelagerte Messungen der kognitiven Fähigkeiten der Schüler*innen als Indikatoren von Vorleistungen einerseits und als Indikatoren der leistungsbezogenen Zusammensetzung von Schulklassen andererseits. Mithilfe dieser Variablen lässt sich prüfen, ob stärkere Zusammenhänge zwischen der sozialen Herkunft und schulischen Leistungen in stärker leistungsdifferenzierten Systemen tatsächlich eine höhere soziale Ungleichheit abbilden (dies entspräche der „Standardposition“; vgl. z. B. Domina et al. 2019; Heisig et al. 2020; Skopek et al. 2016) oder ob schulische Leistungen in stärker leistungsdifferenzierten Systemen lediglich enger mit dem Vorwissen assoziiert sind. Der letztgenannte Befund würde der Intention von *tracking* entsprechen. Um das Ausmaß an Leistungsdifferenzierung in den Bildungssystemen der deutschen Bundesländer zu messen, berücksichtigen Esser und Seuring einen von Helbig und Nikolai (2015) entwickelten Index. Ausgehend von dessen Ausprägungen wurden die deutschen Bundesländer in drei Kategorien unterteilt, die einerseits die Verbindlichkeit von Übergangsempfehlungen und andererseits das Ausmaß der Kontrolle schulischer Abläufe in einem Schulsystem abbilden.

Empirische Analysen von Daten zu Siebtklässler*innen (Jahrgangsstufe 2012/13) aus der National Educational Panel Study (NEPS) bestätigen die zentralen Vermutungen der Autoren. So zeigt sich, dass Leistungen im Lesen und in Mathematik in leistungsdifferenzierenden Bundesländern höher ausfallen als in den übrigen Bundesländern. Dieser Vorteil scheint teilweise über das Ausmaß an kognitiver Homogenität in den Schulklassen vermittelt zu sein. Darüber hinaus deuten die Befunde darauf hin, dass sich durch den Besuch von leistungshomogeneren Klassen Vorteile für den Kompetenzerwerb im Vergleich zu leistungsheterogeneren Klassen ergeben. Dieser positive Effekt von Leistungshomogenität

² Zusammenhänge zwischen Kompositionsmerkmalen (also z. B. das kognitive Niveau und die kognitive Homogenität) und schulischen Leistungen werden im Folgenden als Kompositionseffekte bezeichnet.

fällt in den strikt leistungsbezogen differenzierten Schulsystemen sogar noch größer aus als in den liberaleren Systemen. Insgesamt zeichnen die Ergebnisse ein Bild, dem zufolge *tracking* die Leistungen im Mittel verbessert und soziale Ungleichheit nicht vergrößert.

Die geschilderten Ergebnisse sind von außerordentlicher politischer Brisanz. Aus der traditionellen Dreigliedrigkeit (Hauptschule, Realschule und Gymnasium) haben sich über die Weiterentwicklung bestehender sowie die Einrichtung neuer Sekundarschularten in der Mehrzahl der Bundesländer neue Strukturen herausgebildet, die einen gesamtdeutschen Trend zu zweigliedrigen Schulsystemen erkennen lassen (vgl. Neumann et al. 2013). Insbesondere die Bundesländer Berlin, Bremen, Hamburg, Schleswig-Holstein und das Saarland haben ab dem Schuljahr 2009/10 auf sogenannte „Zwei-Wege-Modelle“ umgestellt, um leistungsbezogene Differenzierung abzubauen.³ Die Studie von Esser und Seuring deutet an, dass derartige Reformmaßnahmen ein Schritt in die falsche Richtung gewesen sein könnten, sollten doch die Schulartzusammenlegungen zumindest im nichtgymnasialen Bereich leistungsheterogenere Klassenzusammensetzungen mit sich gebracht haben. Die Resonanz auf die Ergebnisse von Esser und Seuring spiegelt sich u. a. in einem Beitrag im Deutschen Schulportal wider, demzufolge die strikte Differenzierung nach Leistung beim Übergang an die weiterführende Schule Bildungsungleichheit „[a]nders als bisher angenommen“ nicht verstärkt, aber zu einem höheren Leistungsniveau führen kann (Anders 2021).

Allerdings weist die Studie von Esser und Seuring einige theoretische und methodische Lücken auf. Auf theoretischer Ebene ist hier in erster Linie die Reduktion der Argumentation auf die Effizienz von Unterricht zu nennen. Zahlreiche andere Argumente, die sich auf nicht-intendierte Folgen von Leistungsdifferenzierung wie soziale Schließung und Peer-Effekte beziehen und in der Literatur umfassend diskutiert werden (z. B. Domina et al. 2019; Van Houtte 2016), lassen die Autoren unerwähnt. Fraglich ist außerdem, ob die Ergebnisse für das deutsche Bildungssystem repräsentativ sind. Trotz der Verwendung einer Stichprobe, in der 13 von 16 Bundesländern mit ihren verschiedenen Ausmaßen an Leistungsdifferenzierung abgebildet sind, müssen die Autoren mit einer hohen Zahl an Fallausschlüssen umgehen.⁴ In

³ In Zwei-Wege-Modellen gibt es neben Gymnasien nur noch Gesamtschulen, die sich an Schüler*innen aller Leistungsgruppen richten und diese in heterogenen Lerngruppen zu den verschiedenen allgemeinbildenden Schulabschlüssen, einschließlich dem Abitur, führen.

⁴ Die drei spät sortierenden Bundesländer Berlin, Brandenburg und Mecklenburg-Vorpommern wurden nicht in die Analysen aufgenommen, weil die erforderlichen Daten für den Übergang und die Leistungen noch nicht

der Stichprobe wird jede Schulklasse im Schnitt von nur 7,6 Schüler*innen repräsentiert. Merkmale der Klassenkomposition könnten somit nicht korrekt abgebildet sein. Geht man davon aus, dass Schüler*innen aus sozial weniger privilegierten Familien eine höhere Ausschlusswahrscheinlichkeit haben (vgl. Meinck et al. 2017), so sind hier durchaus Verzerrungen in der Bestimmung der Komposition und dessen Effekten denkbar. Auch die vergleichsweise hohe Gymnasialquote in der verwendeten Stichprobe ist auffällig. Im Vergleich zur tatsächlichen Gymnasialquote der zugrundeliegenden Gesamtpopulation, die der amtlichen Schulstatistik zufolge im Schuljahr 2012/13 bei 39 Prozent lag (Statistisches Bundesamt 2013, eigene Berechnung), betrug die Quote in der verwendeten NEPS-Stichprobe 58 Prozent. Eine verzerrte Abbildung der Prozesse in den nicht-gymnasialen Schulen ist daher nicht auszuschließen.

Das Anliegen der vorliegenden Studie besteht darin, die äußerst wichtigen Ergebnisse von Esser und Seuring mit repräsentativen Daten der IQB-Bildungstrendstudien 2015 und 2018 zu replizieren und um weiterführende Analysen zu ergänzen. Dieses Anliegen ist auch durch die Debatte um die sogenannte Replikationskrise und die zahlreichen Aufrufe und Ermunterungen, sich vermehrt der Replikation von zentralen Forschungsergebnissen zu widmen, begründet (z. B. Freese & Peterson 2017). Replikationsstudien bilden einen Grundbaustein von Wissenschaft (Camerer et al. 2018). Entsprechend ist die vorliegende Studie aus soziologischer Sicht nicht nur wichtig, um die empirischen Analysen von Esser und Seuring auf ihre Belastbarkeit zu überprüfen. Vielmehr soll die Replikation auch auf nicht berücksichtigte theoretische Zusammenhänge hinweisen und damit zur Weiterentwicklung (soziologischer) Theorieansätze zur Erklärung der Folgen von Leistungsdifferenzierung beitragen. Die vorliegende Replikation geht über bisherige kritische Anmerkungen (vgl. Lohmann 2021), die mit deskriptiven Analysen untermauert wurden (vgl. Klemm 2021), hinaus, indem sie die Analysen der Originalstudie mit Daten aus den IQB-Bildungstrendstudien direkt repliziert. Ziel ist es, die Kernaussage der Originalstudie zu überprüfen: Führt die mit strikter Leistungsdifferenzierung verknüpfte kognitive Homogenisierung der Schülerschaft zu höheren Schulleistungen, ohne dabei Bildungsungleichheit zu verstärken?

vorlagen. Von den verbleibenden 4719 Schüler*innen mussten zudem knapp 44 Prozent aufgrund von fehlenden Leistungsdaten oder unvollständigen Angaben in den Modellvariablen ausgeschlossen werden.

2 Theorie

2.1 Das MoAbiT und die Operationalisierung der daraus abgeleiteten Mechanismen

Ihre Hypothesen leiten Esser und Seuring aus dem MoAbiT („Model of Ability Tracking“) ab. Dabei handelt es sich um eine theoretische Modellierung der Mechanismen, über die eine meritokratische Leistungsdifferenzierung und die daraus resultierende kognitive Homogenisierung der Schülerschaft innerhalb von Sekundarschulen auf den Erwerb schulischer Kompetenzen und auf soziale Bildungsungleichheit wirken könnte (vgl. Esser 2016). Das Modell geht über die Betrachtung des sozialen Niveaus von Schulen und Schulklassen und dessen mögliche Effekte auf schulische Leistungen hinaus, indem es explizit auf die Rolle weiterer Kompositionsmerkmale eingeht. Dazu zählen neben der sozialen Homogenität in den Klassen und deren Zusammenspiel mit dem sozialen Niveau auch die mittleren kognitiven Fähigkeiten und die Homogenität in diesem Merkmal. Eine ausführliche Beschreibung des MoAbiT und seiner Implikationen für die vorliegenden Forschungsfrage findet sich in der Originalstudie auf den Seiten 281 bis 289. Der folgende Abschnitt fasst die Kernimplikationen für die Frage zusammen, welche Folgen *tracking* für den Bildungserwerb und für soziale Bildungsungleichheit haben könnte. Außerdem wird für jeden Mechanismus kurz seine Operationalisierung in der Originalstudie erklärt.

Effekte kognitiver Fähigkeiten auf Individual- und Klassenebene: Mediation und Moderation

Aus dem MoAbiT leiten Esser und Seuring zunächst einige *Grundannahmen* ab, auf denen dann in einem nächsten Schritt die Herleitung und empirische Überprüfung zentraler *Mechanismen* aufbauen. Die Grundannahmen beziehen sich auf die Folgen von Leistungsdifferenzierung für die *Bildungsbeteiligung* und für die *Komposition* von Schulklassen (Esser & Seuring 2020: 286ff.). Erwartet wird, dass die Übergänge von der Primar- in die Sekundarstufe in leistungsdifferenzierten Sekundarschulsystemen stärker an die kognitiven Fähigkeiten und weniger stark an die soziale Herkunft gekoppelt sind. Dadurch, so die Autoren, sollten Schulklassen in strikt leistungsdifferenzierten Systemen kognitiv homogener sein, während die Leistungsdifferenzierung keinen (direkten) Effekt auf die soziale Homogenität der Klassen habe.

Die Beschreibung von *Mechanismen*, welche dem Zusammenhang zwischen Leistungsdifferenzierung und schulischen Leistungen sowie sozialer Ungleichheit zugrunde liegen könnten, fokussiert explizit auf die Beziehungen zwischen *tracking* und der Effizienz

von Unterricht. Zwei zentrale Argumente stehen dabei im Vordergrund: *Erstens* ist „schulische[s] Lernen [...] stark auch von der Passung des Unterrichts nach Inhalt und Schwierigkeitsgrad zu den jeweiligen kognitiven Fähigkeiten der Kinder“ abhängig, wobei kognitive Homogenität „eine bessere individuelle Betreuung bei Schwierigkeiten“ erlaubt (S. 283). Lehrkräfte in leistungshomogenen Klassen sollten ihren Unterricht also angemessener an ihre Schülerschaft anpassen können als in leistungsheterogenen Klassen (Eccles & Roeser 2013). *Zweitens* wird argumentiert, dass Leistungsdifferenzierung mit einer passgenaueren Ausrichtung von Curricula auf die unterschiedlichen kognitiven Fähigkeiten verbunden ist, welche die Fokussierung des Unterrichts noch weiter erleichtert.

Den erstgenannten Mechanismus, d. h. systematische Änderungen der strukturellen Bedingungen von Unterricht durch eine leistungsbezogene Sortierung der Schülerschaft, bezeichnen die Autoren als *Mediation*. Sie wird in den empirischen Analysen untersucht, indem zunächst der Haupteffekt von Leistungsdifferenzierung auf die Leistungen von Schüler*innen betrachtet wird (*Systemeffekt*). In einem darauffolgenden Schritt überprüfen die Autoren, inwieweit sich der Systemeffekt durch den (erwarteten) positiven Effekt kognitiver Homogenität auf Leistungen erklären lässt.

Den zweitgenannten Mechanismus bezeichnen die Autoren als *Moderation*. Hier geht es um eine zusätzliche Effektivitätssteigerung, die durch eine Anpassung des Curriculums an die kognitive Homogenität der Schülerschaft ermöglicht wird. Diese Annahme untersuchen die Autoren, indem sie die Effekte des kognitiven Niveaus und der kognitiven Homogenität von Schulklassen in Schulsystemen mit strikter Leistungsdifferenzierung mit den entsprechenden Effekten in weniger strikt differenzierten bzw. *liberalen* Systemen vergleichen. Sie erwarten, dass die positiven Effekte des kognitiven Niveaus und der kognitiven Homogenität in den strikt leistungsdifferenzierten Systemen stärker ausfallen als in den liberalen Systemen. Getestet werden diese Annahmen in der Originalstudie zunächst anhand von Interaktionseffekten der Systemeigenschaften mit der kognitiven Klassenkomposition. Zusätzlich schätzen Esser und Seuring getrennte Modelle für die verschiedenen Schulsysteme und vergleichen die Kompositionseffekte zwischen den Modellen.

Weiterhin erwarten die Autoren, dass sich der Zusammenhang zwischen der sozialen Herkunft und den schulischen Leistungen nicht zwischen mehr und weniger strikt differenzierten Schulsystemen unterscheidet. Gleiches wird auch für die Wirkung der individuellen kognitiven Fähigkeiten erwartet (Esser & Seuring 2020: 288).

2.2 Nicht überprüfte Grundannahme: Soziale Segregation in mehr und weniger leistungsdifferenzierenden Schulsystemen

Eine wichtige Grundannahme der Autoren lautet, dass strikte Leistungsdifferenzierung mit höherer sozialer Heterogenität einhergehen sollte und dass „es damit die sozialen Segregationen weniger geben würde als bei der freien Wahl der Schulen durch die Eltern“ (Esser & Seuring 2020: 285). Dies begründen die Autoren damit, dass sekundäre Herkunftseffekte, also leistungsunabhängige Einflüsse der sozialen Herkunft auf Bildungsentscheidungen (vgl. Boudon 1974), in strikt differenzierten Systemen schwächer ausfallen sollten, weil die Eltern aufgrund von verbindlichen Übergangsempfehlungen dort weniger Einfluss auf den Bildungsübergang ihrer Kinder haben. Die Autoren unterstellen dabei eine Nullkorrelation zwischen den kognitiven Fähigkeiten und der sozialen Herkunft (Esser & Seuring 2020: 285). Dies ist aus analytischer Sicht sinnvoll; aus theoretischer Sicht ist die Annahme allerdings nicht plausibel, da Leistungen und Fähigkeiten sowohl zu Beginn als auch zum Ende der Grundschulzeit stark mit der sozialen Herkunft korrelieren (z. B. Helbling et al. 2019; Jackson 2013). Die Korrelation zwischen Grundschulleistungen und sozialer Herkunft wird auf primäre Herkunftseffekte (d. h. direkte Einflüsse der sozialen Herkunft auf Lernerfolge) zurückgeführt (vgl. Boudon 1974) und führt dazu, dass kognitive Homogenität infolge von *tracking* faktisch auch immer mit sozialer Segregation zwischen den verschiedenen Schularten verbunden ist (Jenkins et al. 2008; Lenz et al. 2019).

Die Autoren überprüfen ihre Annahme einer höheren sozialen Homogenität in weniger strikt differenzierten Schulsystemen jedoch nicht. Damit bleibt unklar, ob meritokratisches *tracking* tatsächlich sekundäre Herkunftseffekte abfedert oder ob primäre Herkunftseffekte der erwarteten Abnahme von sozialer Segregation entgegenstehen. Die gleichzeitige Wirkung dieser zwei Effekte von Leistungsdifferenzierung (also erstens abnehmende sekundäre Effekte und zweitens soziale Segregation aufgrund von primären Effekten) würde einem Szenario entsprechen, das in der Originalstudie zwar angedeutet wird (s. Abbildung 2b, S. 284), auf das die Autoren aber inhaltlich keinen Bezug nehmen. Studien zu den Auswirkungen der aktuellen Schulstrukturreformen weisen jedenfalls darauf hin, dass eine Reduktion von Leistungsdifferenzierung weder zwingend mit einer stärkeren sozialen Segregation zwischen Schularten noch mit stärkeren sekundären Herkunftseffekten einhergeht (Lenz et al. 2019, 2021).

2.3 Nicht-intendierte Nebenerscheinungen von Leistungsdifferenzierung I: Soziale Grenzziehungsprozesse und die Polarisierung von psychologischen Leistungsprädiktoren

Das theoretische Modell der Originalstudie und die sich daraus ergebenden analytischen Möglichkeiten bilden einen wichtigen Beitrag zu der auch international kontrovers geführten Debatte um *tracking* und seine Folgen für schulisches Lernen und Bildungsungleichheit. Als besonders wertvoll ist zu erachten, dass das Modell die Ableitung von Annahmen erlaubt, die eigentlich den Kern der Idee von *tracking* widerspiegeln – also die Folgen einer kognitiven Homogenisierung der Schülerschaft. Kritisch kann dagegen gesehen werden, dass *nicht-intendierte Nebenerscheinungen*, die mit einer kognitiven Homogenisierung der Schülerschaft einhergehen könnten, außer Acht gelassen werden.

In der Literatur wird beispielsweise diskutiert, dass Leistungshomogenität in Schulklassen Grenzziehungsprozesse hervorrufen kann (vgl. Lamont & Molnár 2002), die den positiven Effekten einer optimalen Anpassung des Unterrichts an das Fähigkeitsniveau der Schülerschaft entgegenstehen. Dies könnte sich in einer Polarisierung von Einstellungen und Verhaltensweisen zwischen Klassen (und Schularten) niederschlagen, die durch ein unterschiedliches mittleres Leistungs- und Anspruchsniveau geprägt sind (Berends 1995). Belegt ist beispielsweise, dass Schüler*innen in Klassen mit einem niedrigeren Leistungsniveau negativere Einstellungen zu Schule und Lernen haben als Schüler*innen in Klassen mit einem höheren Leistungsniveau (Müller & Hofmann 2016). Darüber hinaus kann *between-school tracking* zu einer Desillusionierung der Schülerschaft in Schulklassen mit einem niedrigen Leistungsniveau und darüber zu geringerem schulischen Engagement und schlechterem Betragen führen (Van Houtte 2016). Für Deutschland liegen Hinweise für eine Stigmatisierung von Schüler*innen an den weniger anspruchsvolleren Schularten (insbesondere Hauptschulen) vor, die über die Ausbildung distinkter kollektiver Identitäten negativ auf die Motivation wirken kann (Knigge & Hannover 2011).

Der Besuch einer bestimmten Schulart könnte in leistungsdifferenzierten Schulsystemen demnach auf (nicht beobachtete) psychologische Merkmale wie Lernfreude und Anstrengungsbereitschaft wirken. Dabei sind negative Effekte für die am wenigsten anspruchsvollen nicht-gymnasialen Schularten (sprich Hauptschulen und Hauptschulklassen in Schularten mit mehreren Bildungsgängen) und positive Effekte für die anspruchsvolleren Schularten (sprich Realschulen und noch stärker Gymnasien und Gymnasialzweige an

Gesamtschulen) zu erwarten. Derartige Effekte könnten unabhängig vom Vorwissen auftreten und sich scherenartig auf die Leistungsentwicklung auswirken. Ein solcher Mechanismus wäre ggf. in einer geringeren Wirkung des Vorwissens auf die Leistungen in stärker leistungsdifferenzierten Systemen erkennbar. In der Originalstudie zeichnet sich eine solche Tendenz ab, ist doch der Effekt der individuellen kognitiven Fähigkeiten in den liberalen Systemen am stärksten.

2.4 Nicht-intendierte Nebenerscheinungen von Leistungsdifferenzierung II:

Differenzielle Kompositionseffekte und Peer-Effekte

In der Literatur zu *tracking* wird häufig darauf verwiesen, dass leistungsschwächere Schüler*innen in leistungsheterogenen Klassen davon profitieren von Peers mit hohen Leistungen und Fähigkeiten umgeben zu sein (Chiu & Chow 2015; Domina et al. 2019). Insbesondere könnte *Leistungsheterogenität* deshalb effektiv sein, weil Kompositionseffekte *je nach den Vorleistungen* der Schüler*innen unterschiedlich ausfallen. So ist bekannt, dass Schüler*innen mit geringen Vorleistungen am meisten darunter leiden von Peers mit geringen Leistungen umgeben zu sein (Lavy et al. 2012), während umgekehrt auch positive Effekte durch leistungsstarke Peers vor allem für Schüler*innen mit geringen Vorleistungen zu finden sind (Imberman et al. 2012; Kiss 2013). Im Umkehrschluss würde dies bedeuten, dass Effekte kognitiver Homogenität je nach den individuellen Vorleistungen der Schüler*innen variieren. In heterogenen Kontexten sollten es die Schüler*innen mit geringeren kognitiven Fähigkeiten sein, die von Peer-Effekten profitieren, während potentielle Nachteile für Schüler*innen mit höheren kognitiven Fähigkeiten in diesen Kontexten weniger bedeutsam sein sollten. Im Durchschnitt sollten die Leistungen in heterogeneren Kontexten durch diesen Mechanismus im Vergleich zu homogenen Kontexten steigen. Empirische Studien unterstützen diese Vermutung indirekt, indem sie zeigen, dass leistungsschwächere Schüler*innen stärker von Heterogenität profitieren als leistungsstärkere Schüler*innen darunter leiden (Kuzmina & Ivanova 2018; Matthewes 2020). Entsprechende Prozesse diskutieren Esser und Seuring jedoch nicht. Auch einen Interaktionseffekt zwischen individuellen kognitiven Fähigkeiten und kognitiver Homogenität testen die Autoren nicht.

Außerdem könnte die soziale Schließung in leistungshomogenen Kontexten dazu führen, dass *Opportunitäten für positive Kompositionseffekte* ausschließlich in Schulklassen mit einem hohen kognitiven Niveau bestehen, während homogene Klassen mit einem niedrigen Niveau ein vermehrtes Potential für negative Kompositionseffekte bieten. Diese Argumentation

knüpft an die Forschung zu *Peer-Effekten*, also der Wirkung von sozialen Normen, sozialen Vergleichsprozessen, sozialem Lernen und sozialer Nachahmung auf die Einstellungen und das Verhalten von Schüler*innen, an (Jansen et al. 2021; Kretschmer et al. 2018; Lorenz et al. 2020). So könnten in leistungshomogenen Klassen mit einem niedrigen Niveau seltener positive Vorbilder zu finden sein (Kumpulainen & Wray 2002) und die Möglichkeiten der Unterstützung durch leistungsstärkere Mitschüler*innen begrenzt sein (vgl. DiMaggio & Garip 2012). Deshalb sollten Schüler*innen mit höheren kognitiven Fähigkeiten, die in homogenen Kontexten überwiegend von Peers mit hohen Fähigkeiten und einer vorteilhaften sozialen Herkunft umgeben sind, von Homogenität tendenziell profitieren, während der Kompetenzerwerb bei Schüler*innen mit geringeren kognitiven Fähigkeiten, die in homogenen Kontexten überwiegend von Peers mit geringen Fähigkeiten und einer nachteiligen sozialen Herkunft umgeben sind, in homogenen Kontexten tendenziell eher leiden sollte. Entsprechend dieser Annahmen sollte kognitive Homogenität nur dann positiv auf den Kompetenzerwerb wirken, wenn das kognitive Niveau im Mittel hoch ist. Auch wenn sich dieses Muster in der Originalstudie nicht abzeichnet (der Interaktionseffekt zwischen dem kognitiven Niveau und kognitiver Homogenität ist durchweg nicht signifikant und sogar negativ), finden sich an anderer Stelle Belege dafür (Chiu et al. 2017).

2.5 Unbeobachtete Systemunterschiede

Abschließend sei auf ein methodisches Problem verwiesen, dem jeder Systemvergleich mit dem Ziel der Identifikation eines Institutions- oder Systemeffekts ausgesetzt ist. Heterogenität zwischen Systemen besteht meistens nicht nur in den interessierenden institutionellen Regelungen, sondern auch in vielen anderen Merkmalen. Im Anwendungsfall könnte nicht nur die Leistungsdifferenzierung zwischen den Ländergruppen variieren, sondern auch andere Merkmale von Schulsystemen wie die Schulausstattung oder die Ausbildung des Lehrpersonals. Darüber hinaus könnten Systemunterschiede in strukturellen Eigenschaften wie der Sozialstruktur oder des ökonomischen Wohlstands Systemunterschiede in den mittleren Kompetenzen der Schülerschaft begründen, weil diese mit Bildungserfolg zusammenhängen. Da solche Systemunterschiede meistens nicht umfassend kontrolliert werden können, sind entsprechende Vergleichsstudien immer der Gefahr eines *omitted-variable bias* ausgesetzt (vgl. Hanushek & Wößmann 2006; Wößmann 2016).

Dieses Argument wird von den Autoren der Originalstudie nicht aufgegriffen, obwohl sich die Ländergruppen deutlich in ihrer Sozialstruktur und ihrem Bruttoinlandsprodukt unterscheiden

(vgl. Statista Research Department 2021).⁵ Derartige Systemunterschiede, die nicht kausal mit Leistungsdifferenzierung zusammenhängen, könnten die Leistungen von Schüler*innen bereits vor dem Übergang in die Sekundarstufe beeinflusst und zu entsprechenden Systemunterschieden in den schulischen Leistungen geführt haben. Unter der Annahme, dass Kompositionseffekte in der Originalstudie aufgrund der oben genannten Einschränkungen der verwendeten Stichprobe nicht korrekt abgebildet wurden, könnte der Systemunterschied ggf. nicht auf den positiven Effekt kognitiver Homogenität, sondern auf unbeobachtete Länderunterschiede zurückgehen.

3 Replikationskonzept: Daten und Methode

Die Replikation basiert auf Daten der IQB-Bildungstrendstudien. Diese querschnittlich angelegten Schulleistungstudien erheben alternierend alle sechs Jahre sprachliche und mathematisch-naturwissenschaftliche Kompetenzen von Schüler*innen der neunten Jahrgangsstufe in repräsentativen Stichproben für alle 16 Bundesländer. Daten aus dem Jahr 2015 ermöglichen die Analyse der Deutsch-Lesekompetenzen und Daten aus dem Jahr 2018 Analysen der mathematischen Kompetenzen (Stanat et al. 2018; Stanat et al. 2021). Um die Ergebnisse in einem weiteren Kompetenzbereich abzusichern, werden zusätzlich die im Jahr 2015 erhobenen Englisch-Lesekompetenzen berücksichtigt. Die in der Originalstudie untersuchten Schüler*innen besuchten im Schuljahr 2010/11 die fünfte Klasse; die am IQB-Bildungstrend 2015 teilnehmenden Jugendlichen stammten also aus derselben Schülerkohorte.

Die Repräsentativität ist ein entscheidender Vorteil der IQB-Bildungstrenddaten gegenüber den NEPS-Daten. Zudem sollten sich durch die spätere Erhebung in den IQB-Daten deutlichere Systemeffekte bzw. Mediationseffekte der Kompositionsmerkmale nachweisen lassen, da sich die unterschiedlich günstigen Bedingungen in den Schulklassen eine längere Zeit auf Lehr- und Lernprozesse auswirken konnten. Nachteilig ist dagegen das querschnittliche Design. So wurden die kognitiven Fähigkeiten wie auch die Kompetenzen in der neunten Jahrgangsstufe und nicht – wie in der Originalstudie – zu Beginn der Sekundarstufe erhoben. Daher ist zu erwarten, dass die Effekte der kognitiven Fähigkeiten auf individueller Ebene und des kognitiven Niveaus auf Klassenebene in der Replikation größer

⁵ Die in der Originalstudie vorgenommene Kontrolle des Migrationshintergrunds und der sozialen Herkunft der Schüler*innen ist nicht mit einer Berücksichtigung differenzieller Sozialstrukturen gleichzusetzen.

ausfallen als in der Originalstudie, da sich die lernrelevanten strukturellen Eigenschaften der Schularten, Schulen und Schulklassen nicht nur auf die Kompetenzentwicklung, sondern auch auf die Entwicklung der kognitiven Fähigkeiten auswirken (Becker et al. 2012). Anfängliche Fähigkeitsunterschiede zwischen den Klassen sollten demnach bis zum Ende der Sekundarstufe I noch etwas zugenommen haben und stärker mit den zeitgleich erhobenen Kompetenzen korrelieren. Es finden sich aber auch Hinweise darauf, dass die verwendeten Messwerte für kognitive Fähigkeiten von inhaltlichen Lernprozessen relativ unabhängig und damit weitgehend zeitstabil sind (vgl. Weinert 2001).

3.1 Stichproben

Nach Ausschluss von Förderschulen umfassen die Stichproben 31.350 Schüler*innen in 1.425 Klassen (IQB-Bildungstrend 2015) bzw. 43.179 Schüler*innen in 2.009 Klassen (IQB-Bildungstrend 2018). Anhand dieser Daten wurden die Kompositionsmerkmale geschätzt. Da in den verschiedenen Kompetenzbereichen jeweils nur ein Teil der Stichprobe getestet und Schüler*innen mit fehlenden Angaben zum Migrationshintergrund ausgeschlossen wurden, gingen in die Analysen zu Deutsch- bzw. Englisch-Lesekompetenzen Daten zu 28.625 Schüler*innen in 1.421 Klassen bzw. 20.834 Schüler*innen in 1.406 Klassen ein. Die Effekte auf mathematische Kompetenzen wurden anhand von Daten zu 20.772 Schüler*innen in 1.953 Klassen untersucht.

3.2 Variablen

Als Indikatoren für schulische *Leistungen in der Sekundarstufe* dienen die Kompetenzwerte in den Bereichen Deutsch-Leseverstehen, Englisch-Leseverstehen und im Fach Mathematik, die im Rahmen der IQB-Bildungstrendstudien basierend auf den Ergebnissen standardisierter Kompetenztests als *Plausible Values* geschätzt wurden (vgl. Stanat et al. 2016: 515 ff., 2019: 419 ff.). Die *Plausible-Values* ermöglichen verlässliche Schätzungen von Kompetenzverteilungen in Populationen und Subpopulationen sowie von Zusammenhängen zwischen Kompetenzwerten und weiteren Merkmalen der Schüler*innen. Die Leistungsindikatoren wurden analog zum Vorgehen der Originalstudie z-standardisiert ($M = 0$, $SD = 1$). Von einer Mittelwertbildung für mehrere Kompetenzbereiche wird in der vorliegenden Studie u. a. aufgrund der verschiedenen Datengrundlagen abgesehen.

Der *soziale Hintergrund* (SES) wird analog zur Originalstudie über den jeweils höchsten Wert des *International Socio-Economic Index of Occupational Status* der Eltern operationalisiert

(HISEI; Ganzeboom et al. 1992). Die HISEI-Werte wurden im Rahmen der IQB-Bildungstrendstudien aus Angaben der Schüler*innen zu den Berufen ihrer Eltern abgeleitet. Die *kognitiven Fähigkeiten* (ABL) sind durch einen Untertest zum schlussfolgernden Denken des Berliner Tests zur Erfassung fluider und kristalliner Intelligenz für die 8. bis 10. Jahrgangsstufe gemessen, der aus figuralen (nonverbalen) Aufgaben besteht (BEFKI 8-10; Wilhelm et al. 2014). Dieser Test bildet dasselbe Konstrukt ab wie der für das NEPS entwickelte Matrizen-Test des schlussfolgernden Denkens (NEPS-MAT). Fehlende SES- bzw. ABL-Werte werden durch multiple Imputationen ersetzt (Rubin 1987). Wie in der Originalstudie wird das soziale und kognitive *Niveau* der Schulklassen über die ABL- und SES-Mittelwerte abgebildet (NABL bzw. NSES). Die jeweiligen invertierten Standardabweichungen dienen als Indikatoren für die kognitive und soziale *Homogenität* (HABL bzw. HSES) – höhere Ausprägungen entsprechen einer höheren Homogenität. Die besuchte *Schulart* geht als Dummy-Variable in die Analysen ein, die zwischen Gymnasien und nichtgymnasialen Schularten unterscheidet (1=Gymnasium).

Als zusätzliche Kontrollvariablen gehen auf Individualebene Dummies für das *Geschlecht* der Schüler*innen (FEM; 1=Mädchen) und den *Migrationshintergrund* (MHG; 1=mit Migrationshintergrund, d. h. mind. ein Elternteil im Ausland geboren) in die Analysen ein. Die *Dauer des Vorschulbesuchs* wurde in den IQB-Bildungstrendstudien nicht erfasst. Da für diese Variable in der Originalstudie aber über alle Modelle hinweg kein Effekt festzustellen war, ist keine nennenswerte Auswirkung auf die Ergebnisse der Replikation zu erwarten.

3.3 Analytisches Vorgehen

Im ersten Schritt werden auf deskriptiver Ebene Mittelwerte und Standardabweichungen der verwendeten Konstrukte berechnet. Um zentrale Annahmen zu Systemunterschieden in diesen Variablen zu testen, werden die Verteilungsparameter nicht – wie in der Originalstudie – nur für die Gesamtstichproben, sondern auch für die drei Ländergruppen berechnet und auf statistisch signifikante Unterschiede geprüft (*t*-Test, $p < 0,05$).

Zur Identifizierung von Systemeffekten der Leistungsdifferenzierung werden wie in der Originalstudie drei Ländergruppen gebildet. Die Einteilung der Bundesländer nach der *Stringenz von tracking* (T) geht auf Überlegungen von Helbig und Nikolai (2015) zurück und berücksichtigt die Verbindlichkeit von Übergangsempfehlungen und das Ausmaß der Kontrolle von schulischen Abläufen. Die Referenzgruppe (T1 – keine Verbindlichkeit, wenig

Kontrolle) umfasst die Bundesländer Berlin, Bremen, Hessen, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz und Schleswig-Holstein. Zur zweiten Gruppe (T2 – entweder Verbindlichkeit oder stärkere Kontrolle) zählen die Bundesländer Brandenburg, Hamburg, Niedersachsen, Saarland, Sachsen-Anhalt und Thüringen. In Baden-Württemberg, Bayern und Sachsen ist die Kontrolle schulischer Abläufe vergleichsweise hoch, und für die erste Untersuchungskohorte gab es in allen drei Ländern verbindliche Übergangsempfehlungen, sodass sie die dritte Gruppe (T3) mit der stringentesten Leistungsdifferenzierung bilden. Da die Verbindlichkeit der Übergangsempfehlung zum Schuljahr 2012/13 in Baden-Württemberg abgeschafft wurde, wird das Land für die Analysen der im IQB-Bildungstrend 2018 erhobenen Mathematikleistungen der zweiten Gruppe (T2) zugeordnet.

Das multivariate Vorgehen folgt dem der Originalstudie. Geschätzt werden hierarchische lineare Regressionsmodelle, die die geclusterte Datenstruktur (Schüler*innen in Schulklassen) berücksichtigen und eine verlässliche Beurteilung von Effekten der Klassenkomposition ermöglichen. In die Modelle, anhand derer die Annahmen zur *Mediation* der Systemeffekte geprüft werden, werden schrittweise die Stringenz, die soziale Herkunft auf Individual- und Klassenebene sowie die kognitiven Fähigkeiten auf Individual- und Klassenebene, jeweils inkl. der Interaktion von Niveau und Homogenität, und die Schulart kontrolliert. Um die Annahmen zur *Moderation* – d. h. Veränderungen der Kompositionseffekte je nach Stringenz der Differenzierung – zu untersuchen, werden zusätzlich Interaktionseffekte der Stringenz mit den Merkmalen der Klassenzusammensetzung, ihrer Interaktion und der sozialen Herkunft und kognitiven Fähigkeiten auf Individualebene geschätzt. Weiterhin werden die Regressionsmodelle (ohne und mit Kontrolle der Schulart) für die drei Ländergruppen getrennt geschätzt, um spezielle Unterschiede zwischen den Systemen zu veranschaulichen. Über die direkte Replikation der Modelle aus der Originalstudie hinaus werden Modelle geschätzt, die eine mögliche Interaktion zwischen der individuellen kognitiven Fähigkeit und der kognitiven Homogenität in den Klassen prüfen.

Durch die z-Standardisierung der abhängigen Variablen, lassen sich die Regressionskoeffizienten als Leistungsunterschiede in Standardabweichungen interpretieren. Wie in der Originalstudie wurden alle kontinuierlichen unabhängigen Variablen auf ein [0;1]-Intervall transformiert. Die Regressionskoeffizienten weisen somit die maximalen Effekte der Indikatoren aus. Für die Interaktionsterme ist zu beachten, dass die konditionalen Haupteffekte jeweils den Effekt von x_1 (z. B. dem kognitiven Niveau) bei minimaler

Ausprägung von x_2 (minimale kognitive Homogenität bzw. maximale Heterogenität) repräsentieren und umgekehrt, während die Interaktionseffekte die Veränderung zu den konditionalen Haupteffekten abbilden, wenn x_1 bzw. x_2 die höchste empirische Ausprägung annehmen.

Um anhand der Daten der IQB-Bildungstrendstudien repräsentative Aussagen über die zugrundeliegenden Gesamtpopulationen zu treffen, werden in allen Analysen Fallgewichte auf Individual- und Klassenebene verwendet (Stanat et al. 2016: 120, 2019: 126).

4 Ergebnisse und Replikationserfolg

4.1 Deskriptive Befunde

Die Tabellen 1 und 2 bilden die Mittelwerte und Streuungen der analysierten Variablen in den Daten der IQB-Bildungstrends 2015 und 2018 ab. Im Gegensatz zur Originalstudie sind die Angaben nicht nur für die Gesamtstichproben, sondern auch getrennt nach Ländergruppen dargestellt.

- Tabelle 1 etwa hier -

- Tabelle 2 etwa hier -

Der Anteil von Schüler*innen, die ein Gymnasium besuchen, liegt in den Gesamtstichproben bei 38,3 (2015) bzw. 39,6 Prozent (2018) und damit deutlich unter dem für die NEPS-Daten berichteten Wert von 58,4 Prozent. Dies bestätigt die Vermutung, dass die Ergebnisse von Esser und Seuring nicht für Deutschland repräsentativ sind (und wahrscheinlich auch nicht für die Ländergruppen). Die potentiell höhere Aussagekraft der Bildungstrenddaten spiegelt sich außerdem darin wider, dass die Angaben zur Komposition von Schulklassen auf durchschnittlich 22 (2015) bzw. 21,5 Schüler*innen (2018) basieren. In der Originalstudie basierten die Angaben im Schnitt nur auf 7,6 Schüler*innen pro Schulklasse.

Die deskriptiven Analysen lassen außerdem auf die oben genannten Grundannahmen zur Bildungsbeteiligung und Komposition von Schulklassen im Vergleich zwischen den Ländergruppen schließen. Entsprechend der in Tabelle 3 zusammengefassten Befunde, bestätigen sich diese Annahmen nur teilweise. Repliziert werden kann die Annahme, dass sich die soziale Segregation nicht zwischen den Systemen unterscheidet (s. Tab. 3, Zeile 1). Dies

entspricht den Vermutungen von Esser und Seuring sowie den Ergebnissen vorheriger Studien (z. B. Lenz et al. 2019). Außerdem fällt die kognitive Homogenität in der Stichprobe aus dem Jahr 2015 in den T3-Ländern signifikant höher aus als in den T1-Ländern (s. Tab. 3, Zeile 2, Spalte „BT15“). Für das Jahr 2018 ist der Systemunterschied in der kognitiven Homogenität allerdings nicht signifikant (s. Tab. 3, Zeile 2, Spalte „BT18“). Gleichfalls liefern die deskriptiven Ergebnisse keine Hinweise auf die Annahme, dass die kognitive Zusammensetzung von Schulklassen in leistungsdifferenzierten Systemen stärker variiert als in liberaleren Systemen (s. Tab. 3, Zeile 3). Dies sind äußerst wichtige Befunde, da sie die Grundlage für die in der Originalstudie untersuchten Mechanismen des MoAbiT bilden.

- Tabelle 3 etwa hier -

Insgesamt ist die deskriptive Befundlage nicht eindeutig genug, um die Grundannahmen zu stützen, dass Klassen in strikt leistungsdifferenzierten Schulsystemen kognitiv homogener sind und sich im Niveau der mittleren kognitiven Fähigkeiten stärker unterscheiden als in liberaleren Schulsystemen. Damit erscheint die Schlussfolgerung, dass Systemunterschiede zwischen den Ländergruppen in Bildungserfolgen und Bildungsungleichheit tatsächlich die Folge von Unterschieden in der kognitiven Homogenität von Schulklassen sind, bereits auf der Grundlage deskriptiver Analysen zweifelhaft.

4.2 Replikation einer Mediation

Die Annahmen der Originalstudie zur Mediation von Systemeffekten durch Unterschiede in den kognitiven Klassenzusammensetzungen sind in der linken Spalte von Tabelle 4 zusammengetragen. Die weiteren Spalten erklären ihre jeweilige Operationalisierung in den Regressionsmodellen (welche überall der Operationalisierung in der Originalstudie entspricht) sowie den Replikationserfolg. Die vollständigen Regressionsergebnisse finden sich in Appendix A. Für die sprachlichen Kompetenzen (Lesen Deutsch und Englisch, IQB-Bildungstrend 2015) finden sich unter Kontrolle von Geschlecht und Migrationshintergrund keine signifikanten Systemeffekte (s. Tab. 4, Zeile 1). Signifikante Leistungsvorteile in Ländern mit strikter Leistungsdifferenzierung (T3) können nur für mathematische Kompetenzen repliziert werden (IQB-Bildungstrenddaten 2018) (s. Tab. 4, Zeile 1). Im Gegensatz zur Originalstudie fällt der Unterschied zwischen T1- und T2-Ländern jedoch nicht signifikant aus.

- Tabelle 4 etwa hier -

Die folgenden Analysen gehen analog zur Originalstudie den Fragen nach, durch welche Mechanismen dieser Systemeffekt vermittelt wird und ob er tatsächlich das Resultat einer kognitiven Homogenisierung der Schulklassen ist. Da ein Systemeffekt auf die Lesekompetenzen in Deutsch und Englisch nicht replizierbar ist, fokussiert der Text im Folgenden in erster Linie auf die Ergebnisse für Mathematik. Wie in der Originalstudie bleibt der Systemeffekt auf die Mathematikleistungen auch nach Kontrolle der individuellen sozialen Herkunft bestehen, vermindert sich aber mit Kontrolle der sozialen Komposition der Schulklassen (etwas weniger deutlich als im Original). Wie in der Originalstudie zeigt sich außerdem, dass sich die Effekte der sozialen Klassenkomposition schrittweise auf die Effekte der kognitiven Fähigkeiten verlagern. Sie verringern sich unter Kontrolle der individuellen Fähigkeiten (s. Tab. 4, Zeile 2) und weiter unter Kontrolle der kognitiven Klassenzusammensetzung (s. Tab. 4, Zeile 3). Diese Tendenz ist den IQB-Bildungstrenddaten zufolge allerdings weniger deutlich ausgeprägt. Insbesondere bleiben im Gegensatz zur Originalstudie signifikante Effekte der sozialen Homogenität bestehen.

Im Gegensatz zur Originalstudie verringert sich durch die Kontrolle der kognitiven Fähigkeiten aber auch der Systemeffekt weiter deutlich, während er in der Originalstudie konstant bleibt (vgl. Tab. A3 in Appendix A). Den Ergebnissen zufolge ist also ein großer Teil der Leistungsvorteile in Ländern mit strikter Differenzierung den kognitiven Fähigkeiten der Jugendlichen sowie dem kognitiven und sozialen Niveau der Klassen zuzurechnen. Dies ist nicht überraschend, sind doch alle drei Variable in den T3-Ländern signifikant höher ausgeprägt als in den anderen beiden Gruppen (vgl. Tab. 2).

Zentral ist außerdem, dass sich der positive Effekt kognitiver Homogenität nicht replizieren lässt (s. Tab. 4, Zeile 4). Für die Mathematikleistungen der Jugendlichen scheint neben den individuellen Eingangsvoraussetzungen und der sozialen Zusammensetzung der Schülerschaft also nur das kognitive Niveau in den Klassen entscheidend, während für die kognitive Homogenität und deren Interaktion mit dem kognitiven Niveau keine signifikanten Einflüsse festzustellen sind (s. Tab. 4, Zeile 5).

Die Replikation einer Vermittlung von Systemeffekten durch positive Effekte von kognitiver Homogenisierung gelingt somit nicht. Vielmehr scheint der Systemeffekt über die im Mittel günstigere soziale Klassenzusammensetzung und die höheren kognitiven Fähigkeiten (sowohl

auf Individual- als auch auf Klassenebene) in den Ländern mit strikter Differenzierung vermittelt zu sein.⁶

4.3 Replikation der Moderation

Die Annahmen zur Moderation, also einer differenziellen Wirkung der kognitiven Klassenzusammensetzung zwischen den Systemen, sowie ihre regressionsanalytische Überprüfung und der Replikationserfolg sind in Tabelle 5 abgetragen. Signifikant unterschiedliche Effekte des kognitiven Niveaus und vor allem der kognitiven Homogenität finden sich durchweg nicht (s. Tab. 5, Zeilen 1 und 2 und Tab. B1-B3 in Appendix B). Der Effekt der kognitiven Klassenkomposition variiert also nicht zwischen den Systemen. Auch in den nach System getrennten Modellen leistet die kognitive Homogenität mit einer Ausnahme (signifikant negativer Einfluss auf Mathematik-Kompetenzen in T2-Ländern ohne Kontrolle der Schulart) keinen signifikanten Beitrag zur Erklärung von Schülerleistungen (vgl. Tab. C1-C3 in Appendix C). Der Einfluss des kognitiven Niveaus auf Deutsch- und Englisch-Lesekompetenzen fällt in den strikt leistungsdifferenzierenden Ländern sogar geringer aus und ist in dieser Gruppe nicht statistisch bedeutsam.

- Tabelle 5 etwa hier -

Repliziert werden kann hingegen der Befund, dass die Effekte der sozialen Herkunft in den strikt leistungsbezogen differenzierenden Ländern nicht stärker ausfallen als in den anderen beiden Ländergruppen (s. Tab. 5, Zeile 3). Die Ergebnisse der Replikation weisen sogar darauf hin, dass die Effekte der individuellen sozialen Herkunft auf Deutsch- und Mathematikleistungen in der Gruppe mit der höchsten Stringenz signifikant geringer ausfallen. Dies würde der These einer reduzierten sozialen Ungleichheit durch *tracking* entsprechen. Da der Effekt der kognitiven Fähigkeiten – wie oben beschrieben – in den strikt differenzierenden Ländern aber ebenfalls signifikant geringer ausfällt, stützen die Ergebnisse der Replikation allerdings nur bedingt die Interpretation eines „Bildungsklima[s]“, das sich an den „Prinzipien von Effizienz und Leistungsgerechtigkeit“ orientiert (Esser und Seuring 2020: S. 286).

⁶ Einschränkung muss erwähnt werden, dass Unterschiede in den individuellen kognitiven Fähigkeiten und im kognitiven Niveau in der Originalstudie noch etwas weniger bedeutsam gewesen sein sollten als in den Daten der IQB-Bildungstrendstudien (vgl. Abschnitt 3).

4.4 Schülerleistungen im Systemvergleich

Die Abbildungen 1 bis 3 verdeutlichen graphisch die vorhergesagten Leistungsunterschiede zwischen Klassen mit einer unterschiedlichen kognitiven Zusammensetzung. Während Esser und Seuring berichten, dass kognitive Homogenität vor allem bei einem niedrigen kognitiven Niveau in der Schulklasse vorteilhaft für individuelle Leistungen ist (vgl. Abb. 3 in der Originalstudie), zeichnet sich anhand der Daten des IQB-Bildungstrend 2015 das umgekehrte Bild ab: Kognitive Homogenität ist sowohl in Schulklassen, die durch ein geringes kognitives Niveau gekennzeichnet sind, als auch in Klassen mit einem hohen Niveau mit geringeren schulischen Leistungen assoziiert. Die schwächsten Leseleistungen sind sowohl in Deutsch (s. Abb. 1) als auch in Englisch (s. Abb. 2) in kognitiv homogenen Schulklassen mit einem niedrigen kognitiven Fähigkeitsniveau zu erwarten; bedeutsame Systemunterschiede sind dabei nicht zu beobachten

- Abbildung 1 etwa hier -

- Abbildung 2 etwa hier -

- Abbildung 3 etwa hier -

Im mathematischen Bereich sind für die T1- und T2-Länder kaum Unterschiede zwischen heterogeneren und homogeneren Klassen erkennbar (s. Abb. 3). Lediglich für die strikt differenzierenden (T3-)Bundesländer deutet sich für Klassen mit niedrigem Fähigkeitsniveau ein Vorteil durch kognitive Homogenität an. Die Analysen deuten aber nicht auf einen signifikanten Unterschied hin. Ausschlaggebend für die mathematischen Kompetenzen von Schüler*innen scheint dem gegenüber vielmehr zu sein, in welcher Ländergruppe sie unterrichtet wurden und ob sie eine Klasse mit einem niedrigeren oder höheren kognitiven Fähigkeitsniveau besuchen.

4.5 Weiterführende Ergebnisse

Im Folgenden werden die Ergebnisse von weiterführenden Analysen berichtet. Hier geht es zunächst um die Frage, ob Leistungsdifferenzierung zu einer Polarisierung von psychologischen Merkmalen wie Motivation führen kann (vgl. Abschnitt 2.3). Diese Annahme lässt sich indirekt über den Systemvergleich des Effektes von individuellen kognitiven Fähigkeiten auf Leistungen prüfen (s. Tab. 6, Zeile 1 sowie Appendizes B und C).

Der signifikant geringere Beitrag, den kognitive Fähigkeiten in strikt leistungsdifferenzierten Systemen zur Vorhersage der Mathematikkompetenz leisten, deutet darauf hin, dass Vorwissen in strikt leistungsdifferenzierten Systemen eine geringere Rolle spielt als in den liberalen Systemen. Ob dieser Unterschied in der Wirkung des Vorwissens tatsächlich auf soziale Grenzziehungsprozesse und eine Stigmatisierung von leistungsschwächeren Schüler*innen in den weniger anspruchsvollen Schularten der leistungsdifferenzierten Bildungssysteme zurückgeht, lässt sich zwar nicht direkt prüfen; dieses Ergebnis steht jedoch im Kontrast zu der Idee eines meritokratischen Sekundarschulsystems einerseits und der positiven Effekte von kognitiver Homogenität andererseits.

- Tabelle 6 etwa hier -

In einem nächsten Schritt wird die Annahme geprüft, dass die mit Leistungsdifferenzierung einhergehende soziale Schließung Leistungsvorsprünge von leistungsstärkeren Schüler*innen (mit hohen kognitiven Fähigkeiten) begründet, während leistungsschwächere Schüler*innen in kognitiv homogenen Kontexten geringere Leistungszuwächse erzielen als in heterogeneren Kontexten (vgl. Abschnitt 2.4). Hierzu wird (systemübergreifend) eine Interaktion zwischen der individuellen kognitiven Fähigkeit und kognitiver Homogenität geprüft (s. Tabelle 6, Zeile 2 sowie Appendix D). Für die Deutsch- und Englisch-Lesekompetenzen finden sich keine signifikanten Interaktionseffekte. Für die mathematischen Kompetenzen weist ein deutlicher Interaktionseffekt aber durchaus auf differenzielle Kompositionseffekte und einen Schereneffekt durch soziale Schließung hin. Das Ergebnis ist in Abbildung 4 abgetragen: Während Schüler*innen mit hohen kognitiven Fähigkeiten in homogenen Kontexten höhere Leistungen erzielen, wirkt sich kognitive Homogenität für Schüler*innen mit geringeren kognitiven Fähigkeiten nachteilig aus.

- Abbildung 4 etwa hier -

In einem letzten Schritt soll geprüft werden, ob kognitive Homogenität – systemübergreifend – nur in solchen Klassen effektiv ist, die sich auch durch ein höheres kognitives Fähigkeitsniveau auszeichnen (s. Tabelle 6, Zeile 3 sowie Appendix D). Die dahinterstehende Annahme baut auf Argumenten zu Peer-Effekten auf (vgl. Abschnitt 2.4). In der Tendenz zeichnet sich dieses Bild allenfalls sehr schwach für die Englisch-Lesekompetenz und die Mathematikleistung ab. Der Effekt von kognitiver Homogenität auf die Leistungen ist in den Klassen mit dem niedrigsten Niveau negativ; und die negative Wirkung von Homogenität

nimmt mit einem zunehmenden Niveau ab. Allerdings ist der Interaktionseffekt nicht signifikant.

5 Zusammenfassung und Diskussion

Die vorliegende Studie repliziert die Analysen einer von Esser und Seuring (2020) publizierten Studie zu den Effekten von Leistungsdifferenzierung auf die Leistungen von Schüler*innen und auf soziale Ungleichheit in den Leistungen. Die Replikation basiert auf repräsentativen Daten zweier IQB-Bildungstrendstudien aus den Jahren 2015 und 2018, die sprachliche und mathematisch-naturwissenschaftliche Kompetenzen in der Sekundarstufe erhoben haben. Im Gegensatz zur Originalstudie wurden verschiedene Kompetenzbereiche nicht zusammengefasst, sondern getrennt betrachtet.

Deskriptive Analysen zeigten zunächst, dass sich die soziale Segregation nicht bedeutsam zwischen den Systemen unterschieden hatte. Allerdings unterschied sich die kognitive Homogenität nur in der Stichprobe aus dem Jahr 2015 signifikant zwischen den Systemen. Ein (scheinbar systembedingter) Leistungsvorteil in den Bundesländern mit einer strikten Leistungsdifferenzierung konnte wiederum nur für die Stichprobe aus dem Jahr 2018 und den Kompetenzbereich Mathematik festgestellt werden.

Auf der Basis multivariater Analysen ließ sich der Befund replizieren, dass sich Effekte der sozialen Klassenkomposition (sowohl in Bezug auf das Niveau als auch in Bezug auf die Homogenität) *teilweise* auf Effekte der kognitiven Klassenzusammensetzung zurückführen lassen. Auch zeigte sich, dass die von Esser und Seuring vorgeschlagene erweiterte Berücksichtigung von Kompositionsmerkmalen einen Teil des Systemeffekts, also des Ausgangsunterschieds in den mittleren Mathematikkompetenzen zwischen mehr und weniger leistungsdifferenzierenden Ländergruppen, erklärt.

Können diese Ergebnisse so gedeutet werden, dass Leistungsdifferenzierung ein effektiveres Lernen ermöglicht und damit einen Leistungsvorsprung in Mathematik begründet? Einer Antwort auf diese Frage ließe sich annähern, indem die von Esser und Seuring vorgeschlagenen Mechanismen überprüft werden. Deren Wirkung ließ sich aber nicht replizieren. Dies betrifft insbesondere die Annahme, dass sich kognitive Homogenität positiv auf Lernerfolge auswirkt und zwar auch dann, wenn das Fähigkeitsniveau in der Schulklasse gering ist. Die Schlussfolgerung der Originalstudie, dass strikte Leistungsdifferenzierung wegen den positiven Effekten von kognitiver Homogenität mit höheren schulischen

Leistungen einhergeht, muss daher auf der Grundlage der Replikationsergebnisse zurückgewiesen werden.

Weiterführende Analysen der IQB-Bildungstrenddaten weisen im Wesentlichen darauf hin, dass eine Homogenisierung gerade bei einem niedrigen kognitiven Niveau negative Konsequenzen haben kann, während Schüler*innen eher von einer homogenen Klassenzusammensetzung profitieren, wenn sie selbst höhere kognitive Fähigkeiten aufweisen und wenn auch die mittleren kognitiven Fähigkeiten in der Klasse höher sind. Demnach scheinen differenzielle Kompositionseffekte (und ggf. Peer-Effekte) den positiven Effekten der Strukturierung von Unterricht in homogenen Schulkontexten entgegenzuwirken.

Schließlich sei darauf verwiesen, dass die Befunde der vorliegenden Replikationen auf einen geringeren Effekt der sozialen Herkunft in strikt leistungsdifferenzierten im Vergleich zu den liberaleren Systemen hindeuten. Dies als geringeren sozialen Gradienten in leistungsdifferenzierteren Schulsystemen zu deuten, würde allerdings einer Reihe der zuvor genannten Ergebnisse widersprechen. Zusätzlich sollte berücksichtigt werden, dass der Effekt der individuellen kognitiven Fähigkeiten in den leistungsdifferenzierenden Bundesländern geringer ausfiel als in den liberaleren Ländern. Auch erhebliche Unterschiede in der Sozialstruktur sowohl innerhalb als auch zwischen den Ländergruppen sollten bei der Interpretation dieses Befunds berücksichtigt werden.

5.1 Mögliche Alternativerklärungen für den Systemunterschied

Mit den Daten der IQB-Bildungstrendstudien ließ sich für dieselbe Kohorte, für die Esser und Seuring mit NEPS-Daten einen Systemeffekt auf Deutsch- und Mathematikleistungen beobachten konnten, kein Systemeffekt im Bereich der Deutsch-Lesekompetenz replizieren. Für die Schülerkohorte, die drei Jahre später die neunte Klasse besuchte, ließ sich aber zumindest ein Systemunterschied in den Mathematikkompetenzen belegen. So schnitten die Schüler*innen in Bayern, Baden-Württemberg und Sachsen im Schnitt besser in den Kompetenztests ab als die Schüler*innen in den übrigen Bundesländern. In diesen am striktesten nach Leistung differenzierenden Bundesländern erzielten Schüler*innen sowohl in homogen als auch in heterogen zusammengesetzten Klassen höhere Leistungen, und dies unabhängig vom Fähigkeitsniveau der Klassen. Kognitive Homogenität innerhalb von Schulklassen scheint also nicht ursächlich hinter dem sogenannten Systemeffekt zu stehen. Somit stellt sich die Frage nach möglichen Alternativerklärungen.

Diesbezüglich lohnt sich ein Blick auf den IQB-Ländervergleich im Jahr 2011. Zum Zeitpunkt dieser Erhebung besuchten die im Rahmen des IQB-Bildungstrends 2018 getesteten Schüler*innen die zweite Klassenstufe. Nimmt man ihre Mitschüler*innen aus der vierten Klassenstufe und vergleicht deren Mathematikkompetenzen zwischen den Ländergruppen, so ergibt sich folgendes Bild: In Bayern und Sachsen (sowie Baden-Württemberg), also den Bundesländern mit strikter Leistungsdifferenzierung in der Sekundarstufe, erzielten die Grundschüler*innen bereits in der vierten Klassenstufe im Schnitt die höchsten Mathematikkompetenzen (vgl. Stanat et al. 2012: 119). Dagegen lagen die mittleren Kompetenzen von Grundschüler*innen in den Bundesländern mit den liberalsten Sekundarschulsystemen allesamt unter dem bundesdeutschen Durchschnitt. Der Frage, ob dieser Systemunterschied tatsächlich die Folge einer höheren Kontrolle schulischer Abläufe und der dadurch vergleichsweise geringeren tertiären Herkunftseffekte war (z. B. durch verzerrte Lehrkrafterwartungen; vgl. Lorenz 2021; Lorenz et al. 2016), sollten sich weiterführende empirische Studie widmen. Die Folge verbindlicher Übergangsempfehlungen und einer kognitiven Homogenisierung der Schülerschaft kann der Ländergruppenunterschied aber nicht sein, sollten doch die Folgen sekundärer Effekte sowie die positiven Effekte der Homogenisierung vor allem während des Sekundarschulbesuchs greifen und sichtbar sein.⁷ Hieraus könnte geschlussfolgert werden, dass ein Teil der von Esser und Seuring berichteten Systemeffekte nicht hauptsächlich durch die vermuteten Mechanismen getrieben wurden, sondern durch andere, nicht beobachtete Unterschiede zwischen den Ländergruppen. Diese Vermutung wird auch durch den deskriptiven Befund gestützt, dass das kognitive und soziale Niveau in den Klassen der Bundesländer mit strikter Leistungsdifferenzierung signifikant höher ausfiel als in den übrigen Bundesländern. Dem Problem eines möglichen *omitted-variable bias* sind im Prinzip alle nationalen und internationalen Vergleichsarbeiten ausgesetzt, die den Folgen von institutionellen Unterschieden zwischen Schularten bzw. Schulsystemen auf den Grund zu gehen versuchen (vgl. Abschnitt 2.5).

⁷ Die Leistungsunterschiede könnten jedoch eine indirekte Folge von Leistungsdifferenzierung sein, wenn Schüler*innen und ihre Familien sich aufgrund der antizipierten Leistungsdifferenzierung in der Grundschule stärker anstrengen als in weniger leistungsdifferenzierenden Systemen (vgl. Skopek & Passaretta 2021). Auch diesem Mechanismus sollten weiterführende Studien auf den Grund gehen.

5.2 Einteilung von Ländergruppen

Neben diesem allgemeinen Problem der Analyse von Systemunterschieden auf Basis von Ländergruppen, ist auch die in der Originalstudie vorgenommene und in der Replikation übernommene Einteilung der Ländergruppen kritisch zu betrachten (vgl. auch Lohmann 2021). Esser und Seuring beurteilen die Stringenz der Differenzierung anhand der Merkmale Kontrolle und Verbindlichkeit. Außer Acht gelassen wird, dass sich weitere Merkmale von *tracking* teils deutlich zwischen den Ländern innerhalb der betrachteten Ländergruppen unterscheiden.

Besonders offensichtlich ist dies für die Anzahl verschiedener Schularten in der Sekundarstufe, die z. B. von Lenz et al. (2019) als Unterscheidungsmerkmal von Schulsystemen berücksichtigt wurde. Dieser Unterscheidung liegt die Annahme zugrunde, dass eine höhere Anzahl an Schularten mit größeren curricularen Unterschieden sowie Unterschieden im Anspruchsniveau und den Abschlusszielen zwischen den Schularten einhergeht und daher eine leistungsbezogene Differenzierung zur Folge haben sollte – auch ohne Kontrolle schulischer Abläufe und verbindlichen Übergangsempfehlungen. So gibt es in Hessen, Niedersachsen und Nordrhein-Westfalen wie in Baden-Württemberg und Bayern Hauptschulen, Realschulen und Gymnasien zwischen denen bedeutsame leistungsbezogene Unterschiede bestehen sollten – lediglich die höheren Anteile an Gesamtschule mit leistungsheterogenen Klassen als Alternative zu den traditionellen Schularten sprechen für eine geringere Differenzierung. Dagegen gibt es in Sachsen neben Gymnasien mit der *Oberschule* nur noch eine weitere Schulart, die sowohl auf den Hauptschulabschluss als auch auf den Mittleren Schulabschluss vorbereitet. Das sog. *within-school-tracking*, d. h. die abschlussbezogene Differenzierung innerhalb der Oberschulen, beginnt aber erst in der 7. Jahrgangsstufe. Es ist also fraglich, wie sinnvoll die Zuordnung von Sachsen zu den Ländern mit strikt differenzierten Schulsystemen ist – insbesondere, da in der Originalstudie Leistungen in der 7. Klasse als abhängige Variable verwendet wurden, auf die sich die mit der leistungsbezogenen Differenzierung einhergehende kognitive Homogenisierung erst seit Beginn des Schuljahres auswirken konnte (vgl. auch Matthewes 2020, der Sachsen zu einer Gruppe spät bzw. weniger differenzierender Bundesländer zusammengefasst hat).

5.3 Ausblick

Ein wichtiges Resultat der empirischen Analysen ist, dass Esser und Seuring im Gegensatz zu vielen anderen Studien Kompositionseffekte nicht nur in Hinblick auf die Mittelwerte in individuellen Schülereigenschaften untersuchen, sondern darüber hinaus auch die Heterogenität bzw. Homogenität in diesen Merkmalen als theoretisch bedeutsame Variablen in Betracht ziehen. Die Wirkrichtungen der untersuchten Variablen ließ sich anhand von Daten der IQB-Bildungstrendstudien allerdings nicht replizieren.

Eine Ursache könnte sein, dass konkurrierende Prozesse, die im MoAbiT nicht berücksichtigt sind, dem erwarteten positiven Effekt von Leistungshomogenität entgegenstehen. Dazu zählen u.a. *Peer-Effekte*, also die Folgen von Interaktionsprozessen zwischen den Schüler*innen. Peer-Effekte haben in homogenen Kontexten eine geringere Bedeutung als in heterogenen Kontexten, denn heterogene Kontexte weisen eine höhere Varianz in den Eigenschaften der Peers und damit auch ein erhöhtes Potential für die Schüler*innen auf, sich in die eine („nach unten“) oder andere Richtung („nach oben“) anzupassen (vgl. Buchmann & Dalton 2002; Dollmann & Rudolphi 2020; Lorenz et al. 2020). Bislang nicht direkt untersucht wurde jedoch, ob Peer-Prozesse in leistungsheterogenen Schulklassen in höheren oder niedrigeren mittleren Kompetenzen der Schülerschaft münden. Angenommen werden kann, dass leistungsschwächere Schüler*innen in heterogenen Kontexten von positiven Peer-Effekten profitieren könnten, während dieser Vorteil von negativen Peer-Effekten für leistungsstärkere Schüler*innen wieder untergraben werden könnte. Auf der Aggregatebene sollten Peer-Effekten dann in heterogenen Kontexten zu ähnlichen Kompetenzständen führen wie in homogenen Kontexten. Diese Annahme wird indirekt durch Befunde gestützt, denen zufolge Kompositionseffekte in heterogenen Kontexten keine negativen Auswirkungen auf die mittleren Leistungen haben (Duflo et al. 2011; Hanushek et al. 2003; Kiss 2013). Nicht geklärt ist aber, inwieweit Schüler*innen mit unterschiedlichen Lernständen in heterogenen Kontexten überhaupt miteinander interagieren und ob sich daraus positive Effekte für schulische Kompetenzen ergeben. Homophilie, also die Tendenz sich mit Personen zu umgeben, die ähnliche Eigenschaften aufweisen, könnte derartigen positiven Effekten entgegenstehen (vgl. Lorenz et al. 2021).

Weiterführende Studien sollten diesen Fragen nicht nur empirisch nachgehen, sondern auch die zugrunde liegenden *theoretischen Argumente* systematisieren. Dies betrifft auch die Frage, inwieweit die in den Abschnitten 2.3 und 2.4 beschriebenen Mechanismen überhaupt auf

Peer-Prozesse zurückgehen oder ob beispielsweise Kompositionseffekte des kognitiven Niveaus auf andere Mechanismen, z. B. im Zusammenhang mit Lehrkraft- und Unterrichtsmerkmalen, zurückgehen.

Idealerweise sollten in zukünftigen Studien Aspekte von Leistungsdifferenzierung *separat* berücksichtigt werden (vgl. Domina et al. 2019), da einzelne institutionelle Rahmenbedingungen, wie die Übergangsregelungen von der Primar- in die Sekundarstufe, differenzielle Auswirkungen haben können als andere Rahmenbedingungen, wie die Art der leistungsbezogenen Aufteilung von Schüler*innen (z. B. *between-school tracking* im Vergleich zu *within-school tracking*). Darüber hinaus können verschiedene institutionelle Rahmenbedingungen in einem komplexen interagierenden Wirkungszusammenhang stehen (Domina et al. 2019).

Um die mit Systemvergleichen einhergehenden methodischen Probleme zu umgehen (insbesondere die Gefahr unbeobachteter, konfundierender Systemunterschiede), eignen sich *längsschnittliche* und *experimentelle* oder quasi-experimentelle Analysedesigns (z. B. Kiss 2013). Über *Trendanalysen* lassen sich außerdem die Auswirkungen von Schulstrukturen als Systemveränderungen untersuchen. Diese können überraschende Ergebnisse hervorbringen, wie einen nicht-intendierten Effekt der Verstärkung sekundärer Effekte der ethnischen Herkunft durch die Reduktion leistungsbezogener Differenzierung (Lenz et al. 2021). Zukünftige Studien sollten außerdem Einstellungen und Erwartungen von Lehrkräften zwischen den Systemen vergleichen, um der Annahme nachzugehen, dass leistungsverzerrende Einflüsse durch Lehrkräfte in stärker leistungsdifferenzierten Schulsystemen gegenüber liberaleren Systemen reduziert werden können. Die vorliegende Arbeit und ihre Erweiterung der theoretischen Argumente um die Folgen von Leistungsdifferenzierung sollten hierfür eine wichtige Grundlage liefern.

6 Literatur

- Anders, F., 2021 (27. Januar): Verstärkt das gegliederte Schulsystem soziale Ungleichheiten? Das Deutsche Schulportal.
- Becker, M., O. Lüdtke, U. Trautwein, O. Köller & J. Baumert, 2012: The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of educational psychology* 104: 682.
- Berends, M., 1995: Educational stratification and students' social bonding to school. *British Journal of Sociology of Education* 16: 327–351.
- Boudon, R., 1974: *Education, opportunity, and social inequality: Changing prospects in western society*. New York: John Wiley & Sons.
- Buchmann, C. & B. Dalton, 2002: Interpersonal influences and educational aspirations in 12 countries: the importance of institutional context. *Sociology of education* 75: 99–122.
- Camerer, C.F., A. Dreber, F. Holzmeister, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, G. Nave, B.A. Nosek & T. Pfeiffer, 2018: Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nature Human Behaviour* 2: 637–644.
- Chiu, M.M. & B.W.-Y. Chow, 2015: Classmate characteristics and student achievement in 33 countries: Classmates' past achievement, family socioeconomic status, educational resources, and attitudes toward reading. *Journal of Educational Psychology* 107: 152.
- Chiu, M.M., B.W.-Y. Chow & S.W. Joh, 2017: Streaming, tracking and reading achievement: A multilevel analysis of students in 40 countries. *Journal of Educational Psychology* 109: 915.
- DiMaggio, P. & F. Garip, 2012: Network Effects and Social Inequality. *Annual Review of Sociology* 38: 93–118.
- Dollmann, J. & F. Rudolphi, 2020: Classroom composition and language skills: the role of school class and friend characteristics. *British Journal of Sociology of Education* 41: 1200–1217.
- Domina, T., A. McEachin, P. Hanselman, P. Agarwal, N. Hwang & R.W. Lewis, 2019: Beyond tracking and detracking: The dimensions of organizational differentiation in schools. *Sociology of Education* 92: 293–322.

- Duflo, E., P. Dupas & M. Kremer, 2011: Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review* 101: 1739–74.
- Eccles, J.S. & R.W. Roeser, 2013: Schools as developmental contexts during adolescence. Bd. 321–337, S. 225–241 in: B.I. Weiner (Hrsg.), *Handbook of Psychology*. Hoboken: John Wiley & Sons.
- Esser, H., 2016: Bildungssysteme und ethnische Bildungsungleichheit. S. 331–396 in: C. Diehl, C. Hunkler & C. Kristen (Hrsg.), *Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten*. Wiesbaden: Springer VS.
- Esser, H. & J. Seuring, 2020: Kognitive Homogenisierung, schulische Leistungen und soziale Bildungsungleichheit: Theoretische Modellierung und empirische Analyse der Effekte einer strikten Differenzierung nach den kognitiven Fähigkeiten auf die Leistungen in der Sekundarstufe und den Einfluss der sozialen Herkunft in den deutschen Bundesländern mit den Daten der „National Educational Panel Study“ (NEPS). *Zeitschrift für Soziologie* 49: 277–301.
- Freese, J. & D. Peterson, 2017: Replication in social science. *Annual Review of Sociology* 43: 147–165.
- Gamoran, A., 2009: Tracking and inequality: New directions for research and practice. S. 213–228 in: M.W. Apple, S.J. Ball & L.A. Gandin (Hrsg.), *The Routledge International Handbook of the Sociology of Education*. London: Routledge.
- Ganzeboom, H.B.G., P.M. De Graaf & D.J. Treiman, 1992: A standard international socio-economic index of occupational status. *Social Science Research* 21: 1–56.
- Hallinan, M.T., 1994: Tracking: From theory to practice. *Sociology of education* 67: 79–84.
- Hanushek, E.A., J.F. Kain, J.M. Markman & S.G. Rivkin, 2003: Does peer ability affect student achievement? *Journal of applied econometrics* 18: 527–544.
- Hanushek, E.A. & L. Wößmann, 2006: Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal* 116: C63–C76.
- Heisig, J.P., B. Elbers & H. Solga, 2020: Cross-national differences in social background effects on educational attainment and achievement. *Compare* 50: 165–184.
- Helbig, M. & R. Nikolai, 2015: *Die Unvergleichbaren: Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Bad Heilbrunn: Klinkhardt.

- Helbling, L.A., M.J. Tomasik & U. Moser, 2019: Long-term trajectories of academic performance in the context of social disparities: Longitudinal findings from Switzerland. *Journal of Educational Psychology* 111: 1284.
- Imberman, S.A., A.D. Kugler & B.I. Sacerdote, 2012: Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review* 102: 2048–82.
- Jackson, M. (Hrsg.), 2013: *Determined to Succeed? Performance versus Choice in Educational Attainment*. Stanford: Stanford University Press.
- Jansen, M., Z. Boda & G. Lorenz, 2021: Social comparison effects on academic self-concepts—Which peers matter most? *PsyArXiv preprint*: .
- Jenkins, S.P., J. Micklewright & S.V. Schnepf, 2008: Social segregation in secondary schools: how does England compare with other countries? *Oxford Review of Education* 34: 21–37.
- Kiss, D., 2013: The impact of peer achievement and peer heterogeneity on own achievement growth: Evidence from school transitions. *Economics of Education Review* 37: 58–65.
- Klemm, K., 2021: Zu den Effekten kognitiver Homogenisierung. *Kritische Bemerkungen zu den Befunden von Hartmut Esser und Julian Seuring*. S. 11. Essen: URN: urn:nbn:de:0111-pedocs-219944 - <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-219944>.
- Knigge, M. & B. Hannover, 2011: Collective school-type identity: Predicting students' motivation beyond academic self-concept. *International Journal of Psychology* 46: 191–205.
- Kretschmer, D., L. Leszczensky & S. Pink, 2018: Selection and influence processes in academic achievement—more pronounced for girls? *Social Networks* 52: 251–260.
- Kumpulainen, K. & D. Wray, 2002: *Classroom interactions and social learning: From theory to practice*. New York: Routledge.
- Kuzmina, Y. & A. Ivanova, 2018: The effects of academic class composition on academic progress in elementary school for students with different levels of initial academic abilities. *Learning and Individual Differences* 64: 43–53.
- Lamont, M. & V. Molnár, 2002: The study of boundaries in the social sciences. *Annual review of sociology* 28: 167–195.

- Lavy, V., M.D. Paserman & A. Schlosser, 2012: Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal* 122: 208–237.
- Lenz, S., M. Holtmann, C. Rjosk & P. Stanat, 2019: Soziokulturelle Segregation an weiterführenden Schulen–Analysen zur Rolle der Gliederung des deutschen Schulsystems und schulstruktureller Reformmaßnahmen. *Zeitschrift für Erziehungswissenschaft* 22: 1333–1358.
- Lenz, S., C. Rjosk, G. Lorenz & P. Stanat, 2021: Ethnische Segregation zwischen Schularten in mehrgliedrigen Schulsystemen und im „Zwei-Wege-Modell“. Analysen im Kontext von schulstrukturellen Reformmaßnahmen in Berlin, Bremen und Hamburg. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 73: 59–84.
- Lohmann, J., 2021 (20. März): Mit Ignoranz die bildungspolitische Restauration ausrufen. Eine Schule für Alle.
- Lorenz, G., 2021: Subtle discrimination: Do stereotypes among teachers trigger bias in their expectations and widen ethnic achievement gaps? *Social Psychology of Education* 24: 537–571.
- Lorenz, G., Z. Boda & Z. Salikutluk, 2021: Oppositional culture revisited. Friendship dynamics and the creation of social capital among Turkish minority adolescents in Germany. *Journal of Ethnic and Migration Studies* ahead of print: 1–20.
- Lorenz, G., Z. Boda, Z. Salikutluk & M. Jansen, 2020: Social influence or selection? Peer effects on the development of adolescents' educational expectations in Germany. *British Journal of Sociology of Education* 41: 643–669.
- Lorenz, G., S. Gentrup, C. Kristen, P. Stanat & I. Kogan, 2016: Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 68: 89–111.
- Matthewes, S.H., 2020: Better Together? Heterogeneous Effects of Tracking on Student Achievement (CEP Discussion Paper No. No. 1706). London: Centre for Economic Performance (CEP).
- Meinck, S., D. Cortes & S. Tieck, 2017: Evaluating the risk of nonresponse bias in educational large-scale assessments with school nonresponse questionnaires: a theoretical study. *Large-scale Assessments in Education* 5: 1–21.
- Müller, C.M. & V. Hofmann, 2016: Does being assigned to a low school track negatively affect psychological adjustment? A longitudinal study in the first year of secondary school. *School effectiveness and school improvement* 27: 95–115.

- Neumann, M., K. Maaz & M. Becker, 2013: Die Abkehr von der traditionellen Dreigliedrigkeit im Sekundarschulsystem: Auf unterschiedlichen Wegen zum gleichen Ziel? RdJB Recht der Jugend und des Bildungswesens 61: 274–292.
- Rubin, D.B., 1987: Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.
- Skopek, J. & G. Passaretta, 2021: Socioeconomic inequality in children's achievement from infancy to adolescence: The case of Germany. Social Forces 100: 86–112.
- Skopek, J., M. Triventi & S. Buchholz, 2016: How do educational systems affect social inequality of educational opportunities? The role of tracking in comparative perspective. S. 214–232 in: R. Becker (Hrsg.), Research handbook on the sociology of education. Cheltenham: Edward Elgar Publishing.
- Sørensen, A.B., 1970: Organizational differentiation of students and educational opportunity. Sociology of Education 43: 355–376.
- Stanat, P., K. Böhme, S. Schipolowski & N. Haag, 2016: IQB-Bildungstrend 2015. Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich. Münster: Waxmann.
- Stanat, P., K. Böhme, S. Schipolowski, N. Haag, S. Weirich, K. Sachse, L. Hoffmann & F. Federlein, 2018: IQB-Bildungstrend Sprachen 2015 (IQB-BT 2015) (Version 5) [Datensatz]. . Berlin: Institut zur Qualitätsentwicklung im Bildungswesen.
- Stanat, P., H.A. Pant, K. Böhme & D. Richter, 2012: Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik: Ergebnisse des IQB-Ländervergleichs 2011. Münster: Waxmann.
- Stanat, P., S. Schipolowski, N. Mahler, S. Weirich & S. Henschel, 2019: IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich. Münster, New York: Waxmann.
- Statista Research Department, 2021 (19. Mai): Bruttoinlandsprodukt (BIP) in Deutschland nach Bundesländern im Jahr 2020(in Millionen Euro). .
- Statistisches Bundesamt, 2013: Fachserie 11: Bildung und Kultur, Reihe 1: Allgemeinbildende Schulen, Schuljahr 2012/2013. Wiesbaden: Statistisches Bundesamt.
- Van Houtte, M., 2016: Lower-track students' sense of academic futility: Selection or effect? Journal of Sociology 52: 874–889.
- Weinert, F.E., 2001: Der Einfluss der Schule auf die kognitive Entwicklung. Beiträge zur Lehrerinnen-und Lehrerbildung 19: 93–102.

- Wilhelm, O., U. Schroeders & S. Schipolowski, 2014: Berliner Test zur Erfassung fluiden und kristallinen Intelligenz für die 8. bis 10. Jahrgangsstufe (BEFKI 8–10). Göttingen: Hogrefe.
- Wößmann, L., 2016: The importance of school systems: Evidence from international differences in student achievement. *Journal of Economic Perspectives* 30: 3–32.

Tabellen

Tab. 1: Wertebereich, Mittelwerte und Streuungen der verwendeten Konstrukte in den Ländergruppen und in der Gesamtstichprobe für die Daten des IQB-Bildungstrend 2015

Konstrukte	Werte- bereich	T1		T2		T3		Gesamtstichprobe	
		MW	(SD)	MW	(SD)	MW	(SD)	MW	(SD)
Deutsch Lesekom.	[-6,6;3,8]	-0,06	0,99	0,00	0,92	0,05	0,96	-0,01	0,97
Englisch Lesekom.	[-7,3;3,7]	0,01	0,97	-0,05	0,96	0,10	0,95	0,03	0,96
SES	[0;1]	0,51	(0,26)	0,51	(0,25)	0,52	(0,26)	0,51	(0,26)
ABL	[0;1]	0,58	(0,12)	0,58	(0,12)	0,59	(0,12)	0,58	(0,12)
NSES	[0;1]	0,46	(0,18)	0,47	(0,16)	0,47	(0,19)	0,47	(0,18)
HSES	[0;1]	0,37	(0,12)	0,39	(0,12)	0,39	(0,12)	0,38	(0,12)
NABL	[0;1]	0,53	(0,14)	0,54	(0,13)	0,57	(0,14)	0,55	(0,14)
HABL	[0;1]	0,53	(0,11)	0,55	(0,12)	0,56	(0,10)	0,55	(0,11)
GYM (in %)	{0;1}	39,6		39,3		36,0		38,3	
T: Stringenz (in %)	{0;1}	46,0		35,5		18,5			
FEM (in %)	{0;1}	50,1		49,7		49,6		49,8	
MHG (in %)	{0;1}	34,5		20,0		27,6		29,0	
N Schüler*innen		13175		10152		5298		28625	
N Klassen		658		506		257		1421	

Anmerkungen : SES=soziale Herkunft, ABL=kognitive Fähigkeiten, NSES=Niveau SES, HSES=Homogenität SES, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulform, T=Stringenz der Leistungsdifferenzierung, FEM=Geschlecht, MHG=Migrationshintergrund, N=Fallzahl

Tab. 2: Wertebereich, Mittelwerte und Streuungen der verwendeten Konstrukte in den Ländergruppen und in der Gesamtstichprobe für die Daten des IQB-Bildungstrend 2018

Konstrukte	Werte- bereich	T1		T2		T3		Gesamtstichprobe	
		MW	(SD)	MW	(SD)	MW	(SD)	MW	(SD)
Mathe Kompetenz	[-3,7;3,9]	0,07	0,99	0,11	0,94	0,47	0,96	0,16	0,98
SES	[0;1]	0,52	(0,27)	0,51	(0,26)	0,55	(0,26)	0,52	(0,26)
ABL	[0;1]	0,59	(0,12)	0,59	(0,12)	0,61	(0,12)	0,59	(0,12)
NSES	[0;1]	0,51	(0,20)	0,50	(0,19)	0,54	(0,20)	0,51	(0,20)
HSES	[0;1]	0,50	(0,11)	0,50	(0,10)	0,51	(0,10)	0,50	(0,10)
NABL	[0;1]	0,47	(0,13)	0,48	(0,13)	0,53	(0,14)	0,48	(0,13)
HABL	[0;1]	0,60	(0,13)	0,60	(0,13)	0,61	(0,13)	0,60	(0,13)
GYM (in %)	{0;1}	41,7		38,4		37,0		39,6	
T: Stringenz (in %)	{0;1}	46,5		40,2		13,3			
FEM (in %)	{0;1}	51,5		47,9		51,7		50,2	
MHG (in %)	{0;1}	37,0		32,5		25,1		33,0	
N Schüler*innen		9656		8362		2754		20772	
N Klassen		904		785		264		1953	

Anmerkungen : SES=soziale Herkunft, ABL=kognitive Fähigkeiten, NSES=Niveau SES, HSES=Homogenität SES, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulform, T=Stringenz der Leistungs differenzierung, FEM=Geschlecht, MHG=Migrationshintergrund, N=Fallzahl

Tab. 3: Befunde zu Grundannahmen, die den MoAbiT-Mechanismen vorausgehen: Die Komposition von Schulklassen im Vergleich zwischen den Systemen.

	Annahme in der Originalstudie	Operationalisierung	BT15	BT18
1	Die soziale Segregation unterscheidet sich nicht zwischen den Systemen.	Mittlere soziale Homogenität von Schulklassen im Vergleich zwischen den Systemen	JA	JA
2	Die kognitive Homogenität von Schulklassen ist in den strikt leistungsdifferenzierten Systemen höher.	Mittlere kognitive Homogenität von Schulklassen im Vergleich zwischen den Systemen	JA	NEIN
3	Schulklassen unterscheiden sich stärker in ihrer kognitiven Komposition in stärker leistungsdifferenzierten Systemen.	Varianz im mittleren kognitiven Niveau von Schulklassen im Vergleich zwischen den Systemen	NEIN	NEIN

Tab. 4: Übersicht und Replikationserfolg in Bezug auf die erwarteten Mechanismen, welche dem Zusammenhang zwischen Leistungsdifferenzierung und schulischen Leistungen zugrunde liegen („Mediation“).

	Annahme in der Originalstudie	Regressionsmodell	Replikation BT15 (Deutsch Lesekompetenz)	Replikation BT15 (Englisch Lesekompetenz)	Replikation BT18 (Mathematische Kompetenz)
1	Der Ausgangseffekt von Differenzierung (Systemeffekt) ist positiv.	Tab. A1-A3, Modell 1	NEIN	NEIN	JA
2	Soziale Kompositionseffekte reduzieren sich unter der Kontrolle von kognitiven Fähigkeiten.	Tab. A1-A3, Modell 4	JA	JA	JA
3	Soziale Kompositionseffekte reduzieren sich weiter unter der Kontrolle von kognitiver Homogenität.	Tab. A1-A3, Modell 5	JA	JA	JA
4	Kognitive Homogenität hat einen positiven Effekt auf schulische Leistungen.	Tab. A1-A3, Modell 5	NEIN	NEIN	NEIN
5	Der Systemeffekts reduziert sich unter der Kontrolle der kognitiven Klassenzusammensetzung.	Tab. A1-A3, Modell 5	-	-	JA (aber: Niveau entscheidend)

Anmerkung: Die vollständigen Regressionsergebnisse finden sich in Appendix A. Die entsprechenden Ergebnisse von Esser und Seuring (2020) finden sich in Tab. 4 auf S. 292.

Tab. 5: Übersicht und Replikationserfolg in Bezug auf die Annahme einer systematischen Variation von Zusammenhängen zwischen Leistungsdifferenzierung und schulischen Leistungen („Moderation“).

	Annahme in der Originalstudie	Regressionsmodell	Replikation BT15 (Deutsch Lesekompetenz)	Replikation BT15 (Englisch Lesekompetenz)	Replikation BT18 (Mathematische Kompetenz)
1	Der positive Effekt eines höheren Niveaus in der kognitiven Klassenzusammensetzung fällt in den stärker leistungsdifferenzierten Systemen stärker aus.	In Tab. B1-B3: Interaktion zwischen kognitiver Klassenzusammensetzung und System In Tab. C1-C3: Getrennte Effekte der kognitiver Klassenzusammensetzung für jedes System	NEIN	NEIN	NEIN
2	Der positive Effekt von kognitiver Homogenität fällt in den stärker leistungsdifferenzierten Systemen stärker aus.	In Tab. B1-B3: Interaktion zwischen kognitiver Homogenität und System In Tab. C1-C3: Getrennte Effekte der kognitiven Homogenität für jedes System	NEIN	NEIN	NEIN
3	Der Effekt der sozialen Herkunft verstärkt sich nicht durch Leistungsdifferenzierung.	In Tab. B5: Interaktion zwischen sozialer Herkunft und System In Tab. C1-C3: Getrennte Effekte der sozialen Herkunft für jedes System	JA	JA	JA

Anmerkung: Die vollständigen Regressionsergebnisse finden sich in Appendix B und Appendix C. Die entsprechenden Ergebnisse von Esser und Seuring (2020) finden sich in Tab. 5 auf S. 294 sowie Tab. 6 auf S. 295.

Tab. 6: Befunde zu weiterführenden Annahmen.

		Operationalisierung	BT15 (Deutsch Lesekompetenz)	BT15 (Englisch Lesekompetenz)	BT18 (Mathematische Kompetenz)
1	Aufgrund der Polarisierung von psychologischen Merkmalen zwischen Schularten sollte der Effekt des Vorwissens auf Leistungen in stärker leistungsdifferenzierten Systemen geringer sein als in liberaleren Systemen (vgl. Abschnitt 2.3).	Interaktionseffekt zwischen individuellen kognitiven Fähigkeiten und Differenzierung bzw. Vergleich des Effekts von individuellen kognitiven Fähigkeiten zwischen den Systemen	JA	JA	JA
2	Aufgrund von sozialer Schließung und differenziellen Kompositionseffekten in leistungsdifferenzierten Systemen profitieren Schüler*innen mit geringeren kognitiven Fähigkeiten weniger von kognitiver Homogenität als Schüler*innen mit höheren kognitiven Fähigkeiten (vgl. Abschnitt 2.4).	Interaktionseffekt zwischen individueller kognitiver Fähigkeit und kognitiver Homogenität [über alle Systeme hinweg]	NEIN	NEIN	JA
3	Aufgrund von Peer-Effekten sollte kognitive Homogenität nur in solchen Klassen effektiv sein, die auch durch ein höheres kognitives Fähigkeitsniveau geprägt sind (vgl. Abschnitt 2.4).	Interaktionseffekt zwischen kognitiver Homogenität und kognitiven Niveau [über alle Systeme hinweg]	NEIN	(JA) Nicht signifikant	(JA) Nicht signifikant

Anmerkung: Die vollständigen Regressionsergebnisse finden sich in Appendix D.

Abbildungen

Abb. 1: Effekte der Interaktion von kognitivem Niveau, kognitiver Homogenität und der Stringenz der Differenzierung auf die Deutsch-Lesekompetenz in der Sekundarstufe

Anmerkung: Daten des IQB-Bildungstrends 2015.

Abb. 2: Effekte der Interaktion von kognitivem Niveau, kognitiver Homogenität und der Stringenz der Differenzierung auf die Englisch-Lesekompetenz in der Sekundarstufe

Anmerkung: Daten des IQB-Bildungstrends 2015.

Abb. 3: Effekte der Interaktion von kognitivem Niveau, kognitiver Homogenität und der Stringenz der Differenzierung auf die Mathematik-Kompetenz in der Sekundarstufe

Anmerkung: Daten des IQB-Bildungstrends 2018.

Abb. 4: Systemübergreifender Interaktionseffekt von kognitiven Fähigkeiten (ABL) und kognitiver Homogenität (HABL) auf die Mathematikleistung in der Sekundarstufe

Anmerkung: Daten des IQB-Bildungstrends 2018.

Appendix A: Mediation der Systemeffekte

Tab. A1: Systemeffekte als Mediation veränderter Bedingungen auf die Deutsch-Lesekompetenz in der Sekundarstufe

	MoAbiT	Modell 1		Modell 2		Modell 3		Modell 4		Modell 5		Modell 6	
AV = Deutsch Lesekomp.		<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	+			0,33	(0,03)	0,25	(0,03)	0,21	(0,03)	0,22	(0,03)	0,21	(0,03)
NSES	+					4,06	(0,43)	3,47	(0,38)	1,73	(0,30)	1,17	(0,29)
HSES	+					1,19	(0,40)	1,02	(0,37)	0,56	(0,31)	0,29	(0,31)
NSES*HSES	0					-2,91	(0,71)	-2,56	(0,64)	-1,63	(0,52)	-1,07	(0,52)
ABL	+							1,82	(0,07)	1,67	(0,07)	1,67	(0,07)
NABL	+									2,21	(0,30)	2,05	(0,31)
HABL	+									-0,08	(0,31)	0,12	(0,33)
NABL*HABL	+									0,26	(0,54)	-0,04	(0,57)
GYM	+											0,24	(0,04)
T: Stringenz (Ref. T1)													
T2	≥0	0,04	(0,06)	0,04	(0,06)	0,00	(0,04)	0,00	(0,03)	-0,01	(0,03)	-0,01	(0,02)
T3	+	0,08	(0,07)	0,08	(0,06)	0,05	(0,04)	0,03	(0,04)	-0,05	(0,03)	-0,02	(0,03)
FEM		0,20	(0,02)	0,20	(0,02)	0,20	(0,02)	0,22	(0,02)	0,22	(0,02)	0,22	(0,02)
MHG		-0,28	(0,02)	-0,26	(0,02)	-0,25	(0,02)	-0,23	(0,02)	-0,23	(0,02)	-0,23	(0,02)
Konstante		-0,13	(0,04)	-0,31	(0,04)	-2,03	(0,22)	-2,81	(0,20)	-3,11	(0,21)	-2,88	(0,22)
R ² (Individualebene)		0,04		0,06		0,05		0,14		0,13		0,13	
R ² (Klassenebene)		0,00		0,00		0,64		0,63		0,80		0,82	
N Schüler*innen						28625							
N Klassen						1421							

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, T=Stringenz der Leistungsdifferenzierung, FEM=Geschlecht, MHG=Migrationshintergrund, N=Fallzahl; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.

Tab. A2: Systemeffekte als Mediation veränderter Bedingungen auf die Englisch-Lesekompetenz in der Sekundarstufe

	MoAbiT	Modell 1		Modell 2		Modell 3		Modell 4		Modell 5		Modell 6	
AV = Englisch Lesekomp.		<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	+			0,23	(0,03)	0,16	(0,03)	0,12	(0,03)	0,13	(0,03)	0,13	(0,03)
NSES	+					4,40	(0,49)	3,81	(0,44)	1,85	(0,33)	1,02	(0,29)
HSES	+					1,36	(0,44)	1,18	(0,40)	0,64	(0,32)	0,25	(0,31)
NSES*HSES	0					-2,89	(0,79)	-2,53	(0,72)	-1,48	(0,58)	-0,64	(0,54)
ABL	+							1,83	(0,06)	1,71	(0,06)	1,71	(0,06)
NABL	+									1,88	(0,33)	1,65	(0,30)
HABL	+									-0,73	(0,34)	-0,43	(0,33)
NABL*HABL	+									1,35	(0,59)	0,91	(0,56)
GYM	+											0,36	(0,04)
T: Stringenz (Ref. T1)													
T2	≥0	-0,04	(0,07)	-0,03	(0,06)	-0,08	(0,04)	-0,08	(0,04)	-0,10	(0,03)	-0,09	(0,03)
T3	+	0,08	(0,07)	0,08	(0,07)	0,06	(0,04)	0,03	(0,04)	-0,06	(0,03)	-0,02	(0,03)
FEM		0,11	(0,02)	0,12	(0,02)	0,12	(0,02)	0,13	(0,01)	0,14	(0,01)	0,13	(0,01)
MHG		0,00	(0,02)	0,02	(0,02)	0,03	(0,02)	0,06	(0,02)	0,06	(0,02)	0,06	(0,02)
Konstante		-0,16	(0,05)	-0,28	(0,05)	-2,24	(0,26)	-3,02	(0,24)	-3,01	(0,25)	-2,68	(0,22)
R ² (Individualebene)		0,01		0,01		0,01		0,10		0,09		0,09	
R ² (Klassenebene)		0,00		0,00		0,63		0,62		0,79		0,81	
N Schüler*innen						28034							
N Klassen						1406							

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, T=Stringenz der Leistungsdifferenzierung, FEM=Geschlecht, MHG=Migrationshintergrund, N=Fallzahl; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.

Tab. A3: Systemeffekte als Mediation veränderter Bedingungen auf die Mathematik-Kompetenz in der Sekundarstufe

AV = Mathe Globalskala	MoAbiT	Modell 1		Modell 2		Modell 3		Modell 4		Modell 5		Modell 6	
		<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	+			0,43	(0,03)	0,30	(0,03)	0,24	(0,03)	0,24	(0,03)	0,24	(0,03)
NSES	+					4,54	(0,38)	3,46	(0,32)	1,71	(0,29)	0,66	(0,29)
HSES	+					1,63	(0,41)	1,30	(0,33)	0,76	(0,26)	0,16	(0,26)
NSES*HSES	0					-3,01	(0,67)	-2,38	(0,56)	-1,38	(0,45)	-0,34	(0,45)
ABL	+							3,49	(0,06)	3,30	(0,06)	3,31	(0,06)
NABL	+									2,11	(0,31)	1,84	(0,26)
HABL	+									-0,38	(0,26)	-0,02	(0,22)
NABL*HABL	+									0,52	(0,52)	-0,15	(0,45)
GYM	+											0,42	(0,03)
T: Stringenz (Ref. T1)													
T2	≥0	0,04	(0,05)	0,04	(0,05)	0,06	(0,03)	0,04	(0,03)	0,00	(0,02)	0,01	(0,02)
T3	+	0,35	(0,08)	0,35	(0,07)	0,26	(0,05)	0,22	(0,03)	0,15	(0,03)	0,22	(0,02)
FEM		-0,20	(0,02)	-0,19	(0,02)	-0,20	(0,02)	-0,19	(0,01)	-0,19	(0,01)	-0,19	(0,01)
MHG		-0,30	(0,02)	-0,26	(0,02)	-0,26	(0,02)	-0,20	(0,02)	-0,20	(0,02)	-0,21	(0,02)
Konstante		0,02	(0,04)	-0,21	(0,04)	-2,44	(0,23)	-3,92	(0,19)	-3,85	(0,22)	-3,34	(0,22)
R ² (Individualebene)		0,06		0,09		0,07		0,39		0,37		0,37	
R ² (Klassenebene)		0,04		0,04		0,69		0,69		0,85		0,89	
N Schüler*innen						20772							
N Klassen						1953							

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, T=Stringenz der Leistungsdifferenzierung, FEM=Geschlecht, MHG=Migrationshintergrund, N=Fallzahl; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.

Appendix B: Moderation der Systemeffekte

Tab. B1: Systemeffekte als Moderation veränderter Bedingungen auf die Deutsch-Lesekompetenz in der Sekundarstufe

	MoAbiT	Modell 1		Modell 2		Modell 3	
AV = Deutsch Lesekomp.		<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	+	0,21	(0,03)	0,30	(0,05)	0,30	(0,05)
NSES	+	1,17	(0,29)	1,17	(0,46)	1,20	(0,48)
HSES	+	0,29	(0,31)	0,73	(0,48)	0,75	(0,49)
NSES*HSES	0	-1,07	(0,52)	-1,73	(0,81)	-1,77	(0,83)
ABL	+	1,67	(0,07)	1,94	(0,11)	1,94	(0,11)
NABL	+	2,05	(0,31)	2,07	(0,49)	2,08	(0,49)
HABL	+	0,12	(0,33)	0,02	(0,52)	0,00	(0,53)
NABL*HABL	+	-0,04	(0,57)	0,06	(0,94)	0,09	(0,95)
GYM	+	0,24	(0,04)	0,25	(0,04)	0,23	(0,06)
T: Stringenz (Ref. T1)							
T2	≥0	-0,01	(0,02)	-0,03	(0,46)	0,03	(0,47)
T3	+	-0,02	(0,03)	1,15	(0,65)	1,14	(0,66)
T2*SES	0			-0,07	(0,08)	-0,07	(0,08)
T3*SES	0			-0,21	(0,08)	-0,21	(0,08)
T2*NSES	0			-0,07	(0,64)	-0,21	(0,66)
T3*NSES	0			-0,06	(0,61)	-0,06	(0,68)
T2*HSES	0			-0,60	(0,68)	-0,66	(0,68)
T3*HSES	0			-1,02	(0,65)	-1,02	(0,67)
T2*NSES*HSES	0			1,34	(1,18)	1,50	(1,20)
T3*NSES*HSES	0			1,30	(1,12)	1,30	(1,17)
T2*ABL	0			-0,38	(0,16)	-0,37	(0,16)
T3*ABL	0			-0,60	(0,16)	-0,59	(0,16)
T2*NABL	≥0			0,55	(0,73)	0,49	(0,74)
T3*NABL	+			-0,93	(0,93)	-0,93	(0,93)
T2*HABL	≥0			0,98	(0,76)	1,05	(0,77)
T3*HABL	+			-0,94	(0,98)	-0,93	(0,99)
T2*NABL*HABL	0			-1,76	(1,35)	-1,87	(1,37)
T3*NABL*HABL	0			1,68	(1,62)	1,66	(1,63)
T2*GYM	≥0					0,07	(0,08)
T3*GYM	≥0					0,00	(0,10)
Konstante		-2,88	(0,22)	-3,12	(0,29)	-3,12	(0,30)
R ² (Individualebene)		0,13		0,18		0,18	
R ² (Klassenebene)		0,82		0,85		0,85	
N Schüler*innen				28625			
N Klassen				1421			

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, T=Stringenz der Leistungsdifferenzierung; Geschlecht und Migrationshintergrund kontrolliert; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.

Tab. B2: Systemeffekte als Moderation veränderter Bedingungen auf die Englisch-Lesekompetenz in der Sekundarstufe

AV = Englisch Lesekomp.	MoAbiT	Modell 1		Modell 2		Modell 3	
		<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	+	0,13	(0,03)	0,10	(0,05)	0,10	(0,05)
NSES	+	1,02	(0,29)	1,24	(0,46)	1,15	(0,46)
HSES	+	0,25	(0,31)	0,70	(0,48)	0,67	(0,48)
NSES*HSES	0	-0,64	(0,54)	-1,39	(0,82)	-1,30	(0,83)
ABL	+	1,71	(0,06)	1,87	(0,10)	1,87	(0,10)
NABL	+	1,65	(0,30)	1,75	(0,42)	1,73	(0,42)
HABL	+	-0,43	(0,33)	-0,64	(0,51)	-0,58	(0,51)
NABL*HABL	+	0,91	(0,56)	0,98	(0,89)	0,90	(0,91)
GYM	+	0,36	(0,04)	0,35	(0,04)	0,39	(0,06)
T: Stringenz (Ref. T1)							
T2	≥0	-0,09	(0,03)	-0,30	(0,49)	-0,27	(0,50)
T3	+	-0,02	(0,03)	0,76	(0,63)	0,65	(0,63)
T2*SES	0			0,10	(0,07)	0,10	(0,07)
T3*SES	0			0,01	(0,07)	0,01	(0,07)
T2*NSES	0			0,01	(0,65)	-0,03	(0,68)
T3*NSES	0			-0,35	(0,59)	-0,06	(0,64)
T2*HSES	0			0,10	(0,73)	0,08	(0,74)
T3*HSES	0			-1,12	(0,61)	-0,99	(0,62)
T2*NSES*HSES	0			0,12	(1,26)	0,17	(1,29)
T3*NSES*HSES	0			1,58	(1,06)	1,31	(1,09)
T2*ABL	0			-0,23	(0,14)	-0,23	(0,14)
T3*ABL	0			-0,33	(0,15)	-0,33	(0,15)
T2*NABL	≥0			0,30	(0,70)	0,27	(0,70)
T3*NABL	+			-0,80	(0,83)	-0,75	(0,83)
T2*HABL	≥0			0,43	(0,77)	0,43	(0,78)
T3*HABL	+			-0,10	(0,97)	-0,20	(0,97)
T2*NABL*HABL	0			-0,56	(1,35)	-0,57	(1,38)
T3*NABL*HABL	0			0,75	(1,53)	0,89	(1,54)
T2*GYM	≥0					0,02	(0,09)
T3*GYM	≥0					-0,12	(0,10)
Konstante		-2,68	(0,22)	-2,86	(0,33)	-2,84	(0,33)
R ² (Individualebene)		0,13		0,10		0,10	
R ² (Klassenebene)		0,82		0,82		0,82	
N Schüler*innen				28034			
N Klassen				1406			

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, T=Stringenz der Leistungsdifferenzierung; Geschlecht und Migrationshintergrund kontrolliert; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.

Tab. B3: Systemeffekte als Moderation veränderter Bedingungen auf die Mathematik-Kompetenz in der Sekundarstufe

AV = Mathe Global-Skala	MoAbiT	Modell 1		Modell 2		Modell 3	
		<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	+	0,24	(0,03)	0,30	(0,04)	0,30	(0,04)
NSES	+	0,66	(0,29)	0,64	(0,35)	0,62	(0,36)
HSES	+	0,16	(0,26)	0,08	(0,29)	0,07	(0,30)
NSES*HSES	0	-0,34	(0,45)	-0,29	(0,51)	-0,27	(0,52)
ABL	+	3,31	(0,06)	3,43	(0,09)	3,43	(0,09)
NABL	+	1,84	(0,26)	1,98	(0,39)	1,97	(0,39)
HABL	+	-0,02	(0,22)	0,07	(0,30)	0,07	(0,30)
NABL*HABL	+	-0,15	(0,45)	-0,48	(0,63)	-0,48	(0,63)
GYM	+	0,42	(0,03)	0,42	(0,03)	0,43	(0,04)
T: Stringenz (Ref. T1)							
T2	≥0	0,01	(0,02)	0,41	(0,39)	0,39	(0,40)
T3	+	0,22	(0,02)	0,01	(0,49)	-0,01	(0,51)
T2*SES	0			-0,07	(0,06)	-0,07	(0,06)
T3*SES	0			-0,19	(0,08)	-0,19	(0,08)
T2*NSES	0			0,11	(0,52)	0,15	(0,55)
T3*NSES	0			-0,30	(0,76)	-0,27	(0,80)
T2*HSES	0			0,08	(0,47)	0,11	(0,48)
T3*HSES	0			0,02	(0,69)	0,03	(0,71)
T2*NSES*HSES	0			-0,24	(0,82)	-0,28	(0,85)
T3*NSES*HSES	0			0,56	(1,24)	0,53	(1,27)
T2*ABL	0			-0,16	(0,14)	-0,16	(0,14)
T3*ABL	0			-0,34	(0,17)	-0,34	(0,17)
T2*NABL	≥0			-0,66	(0,58)	-0,65	(0,58)
T3*NABL	+			0,61	(0,60)	0,62	(0,60)
T2*HABL	≥0			-0,40	(0,48)	-0,41	(0,48)
T3*HABL	+			0,54	(0,53)	0,53	(0,53)
T2*NABL*HABL	0			0,87	(0,97)	0,89	(0,98)
T3*NABL*HABL	0			-0,39	(1,09)	-0,38	(1,08)
T2*GYM	≥0					-0,02	(0,06)
T3*GYM	≥0					-0,01	(0,09)
Konstante		-3,34	(0,22)	-3,43	(0,26)	-3,42	(0,26)
R ² (Individualebene)		0,37		0,38		0,38	
R ² (Klassenebene)		0,89		0,91		0,91	
N Schüler*innen				20772			
N Klassen				1953			

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, T=Stringenz der Leistungsdifferenzierung; Geschlecht und Migrationshintergrund kontrolliert; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.

Appendix C: Systemvergleich

Tab. C1: Änderung der allgemeinen Zusammenhänge für die Deutsch-Lesekompetenz im System-Vergleich

		MoAbiT		ohne Kontrolle Schulart				mit Kontrolle Schulart					
				T1		T2		T3		T1		T2	
AV = Deutsch Lesekomp.		<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	0	0,28	(0,05)	0,24	(0,06)	0,12	(0,06)	0,28	(0,05)	0,24	(0,06)	0,11	(0,06)
NSES	0	1,66	(0,47)	1,68	(0,43)	1,72	(0,45)	1,14	(0,47)	1,02	(0,43)	1,15	(0,48)
HSES	0	0,93	(0,47)	0,38	(0,46)	0,00	(0,47)	0,70	(0,47)	0,09	(0,47)	-0,28	(0,48)
NSES*HSES	0	-2,24	(0,79)	-1,03	(0,85)	-0,93	(0,83)	-1,68	(0,80)	-0,26	(0,86)	-0,45	(0,83)
ABL	0	1,88	(0,11)	1,65	(0,12)	1,36	(0,12)	1,87	(0,11)	1,65	(0,12)	1,36	(0,12)
NABL	+	2,13	(0,46)	2,99	(0,56)	1,32	(0,76)	2,02	(0,48)	2,67	(0,61)	1,23	(0,77)
HABL	+	-0,36	(0,47)	0,79	(0,55)	-0,99	(0,83)	-0,02	(0,51)	1,08	(0,60)	-0,88	(0,84)
NABL*HABL	0	0,63	(0,87)	-1,34	(0,99)	1,87	(1,28)	0,12	(0,93)	-1,83	(1,07)	1,71	(1,29)
GYM	+							0,23	(0,1)	0,32	(0,05)	0,23	(0,08)
FEM		0,17	(0,03)	0,28	(0,03)	0,25	(0,03)	0,17	(0,03)	0,28	(0,03)	0,24	(0,03)
MHG		-0,26	(0,03)	-0,25	(0,03)	-0,16	(0,03)	-0,27	(0,03)	-0,26	(0,03)	-0,17	(0,03)
Konstante		-3,09	(0,28)	-3,59	(0,37)	-2,41	(0,59)	-2,93	(0,28)	-3,26	(0,39)	-2,15	(0,60)
R ² (Individualebene)		0,16		0,13		0,10		0,16		0,13		0,10	
R ² (Klassenebene)		0,78		0,84		0,83		0,79		0,87		0,84	
N Schüler*innen		13175		10152		5298		13175		10152		5298	
N Klassen		658		506		257		658		506		257	

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, T=Stringenz der Leistungsdifferenzierung, FEM=Geschlecht, MHG=Migrationshintergrund, N=Fallzahl; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.

Tab. C2: Änderung der allgemeinen Zusammenhänge für die Englisch-Lesekompetenz im System-Vergleich

		MoAbiT		ohne Kontrolle Schulart						mit Kontrolle Schulart					
				T1		T2		T3		T1		T2		T3	
AV = Englisch Lesekomp.				<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	0			0,10	(0,05)	0,19	(0,05)	0,13	(0,05)	0,09	(0,05)	0,19	(0,05)	0,13	(0,05)
NSES	0			1,98	(0,52)	1,97	(0,54)	1,78	(0,45)	1,12	(0,46)	1,10	(0,50)	1,12	(0,45)
HSES	0			1,02	(0,49)	1,12	(0,54)	0,00	(0,40)	0,65	(0,47)	0,73	(0,55)	-0,32	(0,40)
NSES*HSES	0			-2,20	(0,86)	-2,11	(0,98)	-0,54	(0,76)	-1,27	(0,82)	-1,11	(0,97)	0,02	(0,72)
ABL	0			1,83	(0,10)	1,63	(0,10)	1,57	(0,11)	1,83	(0,10)	1,63	(0,10)	1,57	(0,11)
NABL	+			1,90	(0,43)	2,38	(0,59)	1,14	(0,73)	1,71	(0,42)	1,99	(0,57)	1,06	(0,71)
HABL	+			-1,12	(0,49)	-0,56	(0,59)	-0,88	(0,82)	-0,58	(0,51)	-0,16	(0,59)	-0,74	(0,82)
NABL*HABL	0			1,72	(0,90)	1,01	(1,06)	1,99	(1,25)	0,89	(0,90)	0,35	(1,04)	1,77	(1,24)
GYM	+									0,39	(0,1)	0,41	(0,07)	0,26	(0,08)
FEM				0,11	(0,02)	0,16	(0,02)	0,16	(0,03)	0,11	(0,02)	0,16	(0,02)	0,15	(0,03)
MHG				0,04	(0,02)	-0,01	(0,03)	0,13	(0,03)	0,04	(0,02)	-0,01	(0,03)	0,13	(0,03)
Konstante				-3,02	(0,35)	-3,41	(0,40)	-2,67	(0,55)	-2,75	(0,33)	-2,99	(0,38)	-2,39	(0,54)
R ² (Individualebene)				0,10		0,09		0,09		0,10		0,09		0,09	
R ² (Klassenebene)				0,80		0,79		0,79		0,83		0,83		0,81	
N Schüler*innen				12929		9949		5156		12929		9949		5156	
N Klassen				654		497		255		654		497		255	

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, T=Stringenz der Leistungs differenzierung, FEM=Geschlecht, MHG=Migrationshintergrund, N=Fallzahl; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.

Tab. C3: Änderung der allgemeinen Zusammenhänge für die Mathematik-Kompetenz im System-Vergleich

		MoAbiT		ohne Kontrolle Schulart						mit Kontrolle Schulart					
				T1		T2		T3		T1		T2		T3	
AV = Mathe Global-Skala				<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	0			0,30	(0,04)	0,25	(0,05)	0,11	(0,06)	0,30	(0,04)	0,24	(0,05)	0,11	(0,06)
NSES	0			1,75	(0,37)	1,92	(0,47)	1,31	(0,73)	0,62	(0,35)	0,81	(0,47)	0,34	(0,72)
HSES	0			0,81	(0,31)	0,78	(0,41)	0,54	(0,62)	0,07	(0,29)	0,18	(0,42)	0,08	(0,65)
NSES*HSES	0			-1,46	(0,54)	-1,71	(0,74)	-0,54	(1,19)	-0,26	(0,52)	-0,58	(0,73)	0,30	(1,15)
ABL	0			3,39	(0,09)	3,40	(0,11)	3,12	(0,14)	3,39	(0,09)	3,41	(0,11)	3,13	(0,14)
NABL	+			2,52	(0,48)	1,56	(0,46)	2,52	(0,57)	1,95	(0,38)	1,38	(0,44)	2,61	(0,49)
HABL	+			-0,02	(0,38)	-0,92	(0,41)	-0,04	(0,47)	0,07	(0,30)	-0,35	(0,40)	0,63	(0,45)
NABL*HABL	0			-0,27	(0,79)	1,43	(0,81)	0,30	(1,03)	-0,47	(0,62)	0,43	(0,78)	-0,92	(0,91)
GYM	+									0,42	(0,04)	0,43	(0,05)	0,43	(0,08)
FEM				-0,21	(0,02)	-0,19	(0,02)	-0,16	(0,03)	-0,21	(0,02)	-0,19	(0,02)	-0,15	(0,03)
MHG				-0,18	(0,02)	-0,20	(0,02)	-0,22	(0,04)	-0,19	(0,02)	-0,21	(0,02)	-0,24	(0,04)
Konstante				-4,04	(0,28)	-3,55	(0,35)	-4,09	(0,47)	-3,29	(0,26)	-3,09	(0,34)	-3,78	(0,47)
R ² (Individualebene)				0,38		0,37		0,33		0,39		0,37		0,33	
R ² (Klassenebene)				0,83		0,83		0,88		0,89		0,88		0,92	
N Schüler*innen				9656		8362		2754		9656		8362		2754	
N Klassen				904		785		264		904		785		264	

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, T=Stringenz der Leistungs differenzierung, FEM=Geschlecht, MHG=Migrationshintergrund, N=Fallzahl; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.

Appendix D: Zusatzanalysen systemübergreifender Effekte

	Deutsch Lesekomp.				Englisch Lesekomp.				Mathematik-Komp.			
	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)	<i>b</i>	(<i>se</i>)
SES	0,21	(0,03)	0,21	(0,03)	0,13	(0,03)	0,13	(0,03)	0,24	(0,03)	0,24	(0,03)
NSES	1,16	(0,29)	1,16	(0,29)	0,98	(0,28)	0,98	(0,28)	0,90	(0,29)	0,91	(0,29)
HSES	0,29	(0,31)	0,29	(0,31)	0,22	(0,31)	0,21	(0,30)	0,36	(0,25)	0,37	(0,26)
NSES*HSES	-1,06	(0,52)	-1,06	(0,52)	-0,58	(0,53)	-0,57	(0,53)	-0,71	(0,45)	-0,73	(0,46)
ABL	1,85	(0,29)	1,66	(0,07)	1,72	(0,25)	1,71	(0,06)	2,68	(0,28)	3,30	(0,06)
NABL	2,01	(0,14)	2,05	(0,31)	2,12	(0,14)	1,72	(0,31)	1,99	(0,14)	1,87	(0,27)
HABL	0,30	(0,35)	0,13	(0,33)	0,05	(0,30)	-0,37	(0,33)	-0,74	(0,30)	-0,17	(0,23)
ABL*HABL	-0,36	(0,56)			-0,03	(0,48)			1,15	(0,50)		
NABL*HABL			-0,07	(0,57)			0,78	(0,57)			0,21	(0,46)
GYM	0,25	(0,04)	0,25	(0,04)	0,36	(0,04)	0,36	(0,04)	0,35	(0,03)	0,35	(0,03)
FEM	0,22	(0,02)	0,22	(0,02)	0,13	(0,01)	0,13	(0,01)	-0,19	(0,01)	-0,19	(0,01)
MHG	-0,23	(0,02)	-0,23	(0,02)	0,06	(0,02)	0,06	(0,02)	-0,21	(0,02)	-0,21	(0,02)
Konstante	-2,96	(0,23)	-2,88	(0,04)	-2,93	(0,22)	-2,71	(0,22)	-3,11	(0,24)	-3,42	(0,22)
R ² (Individualebene)	0,13		0,13		0,09		0,09		0,39		0,37	
R ² (Klassenebene)	0,82		0,82		0,81		0,81		0,87		0,87	
N Schüler*innen		28625				28034				20772		
N Klassen		1421				1406				1953		

Anmerkungen : SES=soziale Herkunft, NSES=Niveau SES, HSES=Homogenität SES, ABL=kognitive Fähigkeiten, NABL=Niveau ABL, HABL=Homogenität ABL, GYM=Schulart, FEM=Geschlecht, MHG=Migrationshintergrund, N=Fallzahl; signifikante Koeffizienten ($p < 0,05$) fett hervorgehoben; Standardfehler in Klammern.