

Hartmut Ditton & Jan Krüsken

## **Denn wer hat, dem wird gegeben werden?**

### **Eine Längsschnittstudie zur Entwicklung schulischer Leistungen und den Effekten der sozialen Herkunft in der Grundschulzeit**

#### **Zusammenfassung**

*Jeweils zum Schuljahresende der zweiten bis vierten Jahrgangsstufe wurden Schulleistungen in Deutsch (Leseverständnis, Rechtschreibung) und Mathematik bei Grundschulern einer geschichteten Zufallsstichprobe aus Bayern und Sachsen (N = 77 Schulklassen mit N = 1 247 Schülern) erhoben. In allen Domänen zeigen sich erhebliche Leistungszuwächse, zudem ergibt sich eine reduzierte Leistungsstreuung im Leseverständnis und der Mathematik. In beiden Domänen kommt es durch den größeren Zugewinn der Schüler mit schwächeren Eingangsleistungen zu einem Ausgleich des Leistungsgefälles. Die schulischen Leistungen stehen durchgängig in deutlichem Zusammenhang mit Merkmalen der sozialen Herkunft, wobei die Beziehungen in Sachsen etwas geringer ausfallen als in Bayern. Entgegen dem Trend einer Angleichung der Schülerleistungen reduzieren sich die Herkunftseffekte im Zeitverlauf nicht, sondern nehmen bei gleichen Eingangsleistungen über die Zeit noch etwas zu.*

#### **Schlagworte**

*Leistungsentwicklung, Leseverständnis, Rechtschreibleistung, Mathematikleistung, Grundschule, soziale Herkunft, Matthäus-Effekt*

## **To Those Who Have, Will More be Given?**

### **A Longitudinal Study Concerning the Development of School Achievement and the Effects of Social Background during Primary School**

#### **Abstract**

*Students from a stratified random sample in Bavaria and Saxony (N = 1 247 in 77 classes) were tested in German (reading, comprehension, spelling) and mathematics at the end of the school year from second to fourth grade. Sizeable achievement gains were observed in every language domain and in mathematics, and there was a considerable reduction in variance of the reading and mathematics scores. A trend toward compensation was evident for both German and mathematics, a development resulting from a larger gain in achievement by the weaker pupils. Achievement consistently and substantially correlated with the social background of the pupils, although the relationships were weaker in Saxony than in Bavaria. Despite the general trend toward*

*compensation, the effects of social origin slightly increase over time, controlling for initial achievement.*

## **Keywords**

*development of achievement, reading comprehension, achievement in spelling, achievement in mathematics, primary school, socio-economic status, Matthew effect*

## **1. Einleitung**

Wie sich die Leistungsentwicklung insgesamt und im Vergleich sozialer Gruppen über die Grundschulzeit hinweg in Deutschland darstellt, ist eine offene Frage. Entsprechende Untersuchungen fehlen beziehungsweise liegen bereits lange zurück (Weinert & Helmke, 1997b). Wir untersuchen im vorliegenden Beitrag die Entwicklung der schulischen Leistungen in den Kernbereichen Mathematik und Deutsch (Leseverständnis, Rechtschreibung) im Verlauf der letzten beiden Grundschuljahre bis zum Zeitpunkt des Übertritts, d.h. vom Ende der zweiten bis zum Ende der vierten Jahrgangsstufe. Dabei stehen zwei Aspekte im Vordergrund: Bezüglich der Leistungsentwicklung über die Zeit interessiert zunächst die Stabilität beziehungsweise Variabilität der Schulleistungen in dieser Phase. Ein besonderes Augenmerk wird auf den Vergleich der Entwicklung bei leistungsstärkeren und leistungsschwächeren Schüler gelegt. Zum Zweiten geht es um die Frage, wie sich im Zeitverlauf die Leistungsunterschiede zwischen den sozialen Gruppen entwickeln. Bezüglich beider Aspekte (Leistungsniveau und soziale Herkunft) soll analysiert werden, ob anfänglich bestehende Leistungsunterschiede gleich bleiben oder über die Zeit zu- beziehungsweise abnehmen.

## **2. Fragestellung und Stand der Forschung**

Schule hat vielfältige Funktionen zu erfüllen (Fend, 1981). Grundlegend von Bedeutung ist ihr Bildungsauftrag resp. ihre Qualifikationsfunktion. Diesbezüglich soll die Grundschule „*durch die Vermittlung einer grundlegenden Bildung die Voraussetzungen für jede weitere schulische Bildung schaffen*“ (Bay. EUG, Art. 7, Abs. 4). Ganz ähnlich heißt es auch in der Vereinbarung der *Kultusministerkonferenz* (KMK) zu den 2003 verabschiedeten Bildungsstandards für die Primarstufe: „*Auftrag der Grundschule ist die Entfaltung grundlegender Bildung. Sie ist Basis für weiterführendes Lernen und für die Fähigkeit zur selbständigen Kulturaneignung*“ (Bildungsstandards Deutsch, KMK, 2004).

Die schulische Förderung in der Grundschulphase hat daher in erster Linie darauf abzielen, jedem einzelnen Kind eine sichere Basis für den weiteren Bildungsweg zu geben. Zu dieser Basis gehört auch die Vermittlung von grundlegenden und für die gesellschaftliche Teilhabe erforderlichen Kulturtechniken, wie die Beherrschung der Schriftsprache und elementare mathematische Fähigkeiten.

Diese sind in einem Mindestmaß *allen* Schülern, unabhängig vom Niveau ihrer Vorkenntnisse bei der Einschulung, zu vermitteln. Daneben ist auch zu gewährleisten, dass das Postulat der Chancengleichheit nicht verletzt wird, d.h. kein Kind soll bevorzugt oder benachteiligt werden. Versteht man unter Chancengleichheit, dass alle Schüler gewisse Mindeststandards während der Grundschulzeit erreichen müssen, wären vor allem die anfänglich schwachen Schüler mit geringen Vorkenntnissen zu fördern, wodurch eine Abnahme der anfänglichen Leistungsheterogenität zu erwarten wäre. Versteht man darunter allerdings die Ausrichtung auf den für jeden Schüler maximal möglichen Lernfortschritt während der Grundschulzeit, so wäre eine differenzielle Förderung aller Schüler geboten, mit der Konsequenz, dass bestehende Eingangsunterschiede zwischen den Schülern gleichbleiben würden, beziehungsweise sich sogar verstärken könnten, wenn die differentielle Förderung auf unterschiedliche Lernraten bei starken und schwachen Schülern trifft.

Neben dem Bildungsauftrag hat Schule auch eine Selektions- und Legitimationsfunktion zu erfüllen. Das betrifft im deutschen Schulsystem die Grundschule ganz besonders, da am Ende der Primarstufe die Entscheidung über den Besuch der weiterführenden Schulen ansteht (Ditton, 2007a, 2007b; Ditton & Krüskens, 2006; Ditton, Krüskens & Schauenberg, 2005). Trotz einer partiellen Entkoppelung von besuchter Schulform und letztlich erworbenem schulischem Abschluss (Baumert, Trautwein & Artelt, 2003) ist dieser Übergang eine wichtige Weichenstellung mit hoher Bedeutung. Die Aufteilung nach unterschiedlichen Schulformen mit voneinander abweichenden Anforderungsniveaus setzt voraus, dass nach Leistungsniveaus differenziert werden kann. Eine wesentliche Aufgabe der Lehrkräfte besteht darin, für ihre Schüler Erfolgsprognosen bezüglich der Bewährung auf unterschiedlich anspruchsvollen weiterführenden Schulen abzugeben. Dazu, inwieweit grundlegende Bildung für alle und Differenzierung nach Eignung und Neigung miteinander vereinbare Ziele sind, finden sich in den Vereinbarungen der KMK ebenso wenig konkrete Aussagen wie in den Gesetzen, Verordnungen oder Erlassen der Länder.

Auch wenn die Bildungsziele der Grundschule mit *allgemeiner Persönlichkeitsbildung* und *Anleitung zu selbständigem Lernen* umschrieben werden, dürfte dennoch die Vermittlung elementarer Grundfertigkeiten in Lesen, Mathematik sowie Heimat- und Sachkunde im Vordergrund stehen. Diesbezüglich starten Kinder mit sehr unterschiedlichen Voraussetzungen in ihre schulische Laufbahn. Eine Untersuchung zu Mathematik, Wortschatz und Lesen in der Schweiz zeigt, dass ein Teil der Kinder zum Beginn der Schulzeit bereits den Stoff der ersten und einige Kinder sogar einen erheblichen Teil des Lehrstoffes der zweiten Klassenstufe beherrschen, wohingegen andere Kinder nicht einmal mit Buchstaben und Zahlen vertraut sind (Moser, Stamm & Hollenweger, 2005). Im Hinblick auf die möglichen Lernzuwächse könnte erwartet werden, dass vom Grundschulunterricht diejenigen Schüler mehr profitieren, die anfangs mit einem niedrigen Kenntnisniveau starten, da hier die Distanz zu den erwarteten Standards und auch die Chancen für Lernzuwachs größer sind. Allerdings könnte das domänenspezifisch unterschiedlich sein. Für den Bereich des Lesens ist eher zu erwarten, dass im unteren Leistungsbereich ganz erhebliche Zuwächse erzielt werden, im oberen

Leistungsbereich könnten geringe Zuwächse zu verzeichnen sein: Wer schon frühzeitig gut lesen kann, wird in der Grundschulphase möglicherweise nicht mehr so erheblich dazu lernen können. In Rechtschreibung und Mathematik könnte es sich anders verhalten, hier könnten auch deutliche Zuwächse im oberen Leistungsbereich festzustellen sein. Angesichts fehlender aktueller Untersuchungen zur Leistungsentwicklung in der Grundschule bleibt man jedoch auf Spekulationen angewiesen.

Alles in allem ist jedenfalls davon auszugehen, dass die Grundschule einen schwierigen Auftrag zu erfüllen hat. Auf der einen Seite geht es darum, die anfangs schwachen Schüler auf ein solides Niveau zu heben, das eine erfolgreiche weitere schulische Laufbahn wahrscheinlich macht. Andererseits sollen auch die besonders begabten Schüler ihren Fähigkeiten gemäß gefördert werden. Wie man weiß, sind beide Ziele in aller Regel nicht leicht vereinbar (Baumert, Roeder, Sang & Schmitz, 1986; Beck et al., 1988; Treiber & Weinert, 1985). Zudem gelten inzwischen in der bildungspolitischen und öffentlichen Wahrnehmung allzu große Differenzen im Lernerfolg zum Ende einer Schulphase als problematisch. Zugleich steht aber nach der Primarstufe die Verteilung auf die weiterführenden Schulen abhängig von den erbrachten Leistungen an – zuviel an Leistungsausgleich wird es daher auch nicht geben dürfen.

Der Forschungsstand zur Leistungsentwicklung im Grundschulalter ist in Deutschland äußerst dürftig. Die bis heute bemerkenswerteste Studie in diesem Kontext ist ohne Zweifel LOGIK/SCHOLASTIK (Weinert & Helmke, 1997b). Auch diese Studie liefert jedoch nur wenig Anhaltspunkte dafür, welche Verläufe in den hier untersuchten Domänen zu erwarten sind. Das Leseverständnis wurde in Scholastik nur zu Beginn und am Ende der 2. Jahrgangsstufe erfasst. Einen breiteren Raum nahmen die Rechtschreibfähigkeiten ein, die im Verlauf der Grundschulzeit regelmäßig überprüft wurden. Es zeigte sich, dass die *Phonologische Bewusstheit* im Kindergarten einen hohen Prognosewert für die Leistungen in der Grundschule hatte. Ebenso bestand eine vergleichsweise enge Beziehung zwischen den Leistungen am Ende der Grundschulzeit und den Leistungen in der zweiten Klasse. Für Mathematik betrug die Korrelation zwischen den Leistungen in der zweiten und vierten Klasse  $r = .55$ , für Rechtschreibung war die Korrelation mit  $r = .65$  noch höher. In der zusammenfassenden Interpretation der Befunde zu Scholastik sprechen die Autoren deshalb von der Grundschulzeit als einer „... Zeit der Stabilisierung interindividueller Differenzen. Es tritt weder ein Schereneffekt auf, noch verringern sich die bestehenden Differenzen zwischen leistungsstärkeren und -schwächeren Schülern“ (Weinert & Helmke, 1997b, S. 467). Dies stellt allerdings eine vereinfachende Kennzeichnung der Befunde dar, die man in den im Einzelnen vorgestellten Teilberichten und Analysen nicht konsistent so wiederfindet.

Die von den Autoren festgestellte „Stabilisierung der interindividuellen Differenzen“ in den beobachteten Merkmalen bedeutet nicht zwangsläufig, dass die Leistungsvarianz im Zeitverlauf gleich bleiben muss. Aussagen über die Veränderung der Merkmalsvarianz über die Zeit erfordern den Einsatz vergleich-

barer Messinstrumente über die Zeit, was in der Scholastik-Studie nur in wenigen Bereichen gewährleistet war. Tatsächlich wird bei denjenigen Leistungsmaßen der Scholastik-Studie, bei denen vermutlich von vergleichbaren Messungen gesprochen werden kann, oft eine Abnahme der Leistungsvarianz im Zeitverlauf erkennbar (Weinert & Helmke, 1997b, S. 503), trotz der zunehmenden Jahresstabilitäten beziehungsweise Zweijahresstabilitäten. Überdies handelt es sich um inzwischen schon vergleichsweise alte Daten und differenzierte Auswertungen nach aus heutiger Sicht bedeutsamen Aspekten konnten nicht vorgenommen werden. So erfolgt z. B. keine differenziertere Darstellung der Entwicklungsverläufe in Abhängigkeit vom Niveau der Vorkenntnisse. Nach sozialen Gruppen wird in der Studie gar nicht differenziert, vermutlich lagen auch keine entsprechenden Daten vor. Auch hinsichtlich der Stichprobe, die auf den Raum München eingeschränkt war, lassen sich Bedenken bezüglich der Repräsentativität äußern.

Eindeutig auch aus anderen Studien bestätigt ist dennoch, dass das Vorwissen beziehungsweise Vorläuferfähigkeiten üblicherweise die besten Prädiktoren für den späteren Leistungsstand sind. Das entspricht auch dem, was Modelle zu Bedingungsfaktoren schulischer Leistungen erwarten lassen (Baumert, Watermann & Schümer, 2003; Heller, 1997; Sauer & Gattringer, 1985; Weinert & Helmke, 1997a). Alle Modelle gehen zwar gemeinsam von einer multiplen Determination aus, bei der zahlreiche Faktoren eine Rolle spielen (Kaufmann, 2008). Jeweils zeigt sich bzgl. der Leistungsentwicklung jedoch eine hohe bis sehr hohe Prognosekraft der Vor- auf die Nachtestleistung und insofern auch eine gewisse Stabilität von Leistungsrangreihen. Zudem ergibt sich durchweg (auch international) eine zwar unterschiedlich hohe, aber durchgängig bestehende Kopplung der erzielten Leistungen an Merkmale der sozialen Herkunft (sozialer Status, Bildungsstatus, Migrationsstatus). Zur Veränderung von anfangs bestehenden Leistungsdifferenzen im Verlauf der Schulzeit, ist die Forschungslage jedoch unklar. Bezogen auf Vergleiche der Ergebnisse aus den Studien IGLU (*Internationale Grundschul-Lese-Untersuchung*) und PISA (*Programme for International Student Assessment*) liegt zwar der Schluss nahe, dass sowohl die Leistungsstreuung im Verlauf der Schulzeit (erheblich) zunimmt als auch die Unterschiede zwischen den sozialen Gruppen deutlich größer werden. Insofern könnte allgemein von einem „Matthäus-Effekt“, also einer sich vergrößernden Leistungsschere ausgegangen werden. Bedeutsam für die Auseinanderentwicklung scheint hierbei allerdings zu sein, dass die unterschiedlichen Formen der Sekundarschule unterschiedliche „Entwicklungsmilieus“ darstellen, in denen die Leistungsentwicklung selbst bei gleichen Eingangsbedingungen unterschiedlich verläuft (Baumert, Stanat & Watermann, 2006).

Bezüglich der hier verfolgten Fragestellung ist am ehesten zu erwarten, dass in der Grundschulzeit eine gewisse Balance erreicht wird zwischen dem Anspruch, die schwächeren Schüler besonders zu fördern und der Forderung, dabei die leistungsstärkeren Schüler nicht aus dem Auge zu verlieren. Besonders viel an Ausgleich auf der einen Seite oder erhebliche Schereneffekte auf der anderen Seite

sind daher weniger wahrscheinlich. In diesem Sinne fallen auch die bisherigen Ergebnisse aus *Kompetenzaufbau und Laufbahnen im Schulsystem* (KOALA-S) zur Leistungsentwicklung während des vierten Grundschuljahres an einer Stichprobe bayrischer Grundschulklassen aus (Krüsken, 2007). Insgesamt finden sich deutliche Leistungszugewinne im Beobachtungszeitraum in allen Testbereichen. Die Herkunftseffekte waren in der dritten Klasse schon deutlich ausgeprägt und nahmen vom Ende der dritten zum Ende der vierten Jahrgangsstufe signifikant aber nicht stark zu. Hinweise auf einen Schereneffekt in Abhängigkeit der sozialen Herkunft während der Grundschulzeit finden sich dagegen in der Studie *Evaluation eines Vorschultrainings zur Prävention von Schriftsprachproblemen sowie Verlauf und Entwicklung des Schriftspracherwerbs in der Grundschule* (EVES) (Zöllner & Roos, 2009), einer Längsschnittstudie über den gesamten Grundschulzeitraum mit einer Stichprobe von 16 Heidelberger Grundschulen. Dies betrifft allerdings nur die Vorläuferfähigkeiten der Lesekompetenz (Wortdecodierung und -recodierung), im Bereich Rechtschreiben bleiben in dieser Studie die Herkunftseffekte stabil im Beobachtungszeitraum bestehen.

Ob sich im vorliegenden Längsschnitt, der mit zwei Grundschuljahren einen doppelt so langen Zeitabschnitt wie der erste Koala-Längsschnitt untersucht, ähnliche Befunde zeigen, ist allerdings damit nicht eindeutig abzuleiten. Wie sich die Leistungsentwicklung im Vergleich der beiden an der Untersuchung beteiligten Länder darstellt, ist ebenfalls eine offene Frage. Mit Blick auf die Ergebnisse aus den Ergänzungsuntersuchungen der internationalen Vergleichsstudien ist allerdings zu erwarten, dass in Sachsen die Herkunftseffekte geringer ausfallen werden als in Bayern. Dagegen dürften kaum beziehungsweise allenfalls geringe Unterschiede im Leistungsniveau zwischen beiden Ländern bestehen.

### **3. Anlage der Untersuchung und Erhebungsinstrumente**

In der Längsschnittstudie KOALA-S wurden jeweils zum Ende der zweiten, dritten und vierten Jahrgangsstufe die Fachleistungen der Schüler in Deutsch und Mathematik sowie die kognitiven Grundfähigkeiten erhoben. Eingesetzt wurden standardisierte und erprobte Schulleistungstests für den Primarbereich. Die Tests wurden von geschulten Testleitern administriert. Zu jedem Erhebungszeitpunkt wurden begleitende schriftliche Befragungen von Schülern, Eltern und Lehrkräften durchgeführt. Tabelle 1 gibt einen Überblick über die Erhebungsbereiche.

**Tabelle 1:** Erhebungszeitpunkte und Testverfahren in KOALA-S im Überblick

	Erhebung 2005 Ende Jahrgangsstufe 2 (T1)	Erhebung 2006 Ende Jahrgangsstufe 3 (T2)	Erhebung 2007 Ende Jahrgangsstufe 4 (T3)	Erläuterung
Erhebungstage	2	2	2	
Testbereiche Fachleistungstests (N Items*)	Leseverständnis (16) cr. $\alpha = .64$ Rechtschreiben (48) cr. $\alpha = .85$ Mathematik (14) cr. $\alpha = .75$	Leseverständnis (15) cr. $\alpha = .69$ Rechtschreiben (56) cr. $\alpha = .89$ Mathematik (15) cr. $\alpha = .78$	Leseverständnis (26) cr. $\alpha = .65$ Rechtschreiben (64) cronbachs $\alpha = 90$ Mathematik (15) cr. $\alpha = .71$	Summe Testzeit etwa zw. 120 bis 180 Min.
Kognitive Fähigkeiten	2 KFT Subskalen (30) cr. $\alpha = .68$	4 KFT Subskalen (60) cr. $\alpha = .80$	Teile von 2 KFT Subskalen (18) CFT 20 komplett cr. $\alpha = .74$	Verbindung über gemeinsame Items T1 – T3
Befragungen	schriftliche Schüler-, Eltern, und Lehrkräftebefragungen			Wdh. von T1 – T3

\*Anmerkung: In Mathematik und Lesen wurden auch mehrstufige Items eingesetzt. cr.  $\alpha =$  Cronbachs Alpha.

**Stichprobe.** In KOALA-S wurden geschichtete Zufallsstichproben öffentlicher Grundschulen in Bayern und Sachsen gezogen (Schichtungsmerkmale: Urbanisierung, Schulgröße, Schultyp, Region). Aus jeder Schule wurde eine zufällig ausgewählte Klasse in die Stichprobe aufgenommen. Von den für die erste Erhebung ausgewählten 84 Schulklassen nahmen 77 ( $N = 1\,453$  Schüler) an allen drei Erhebungen teil (Sachsen: 35 Schulklassen mit  $n = 582$  Schülern; Bayern: 42 Schulklassen mit  $n = 871$  Schülern). Dabei war für die Teilnahme an der Studie das Vorliegen schriftlicher Elterngenehmigungen für jede Erhebung erforderlich. Eine Zustimmung für alle Erhebungen haben in Bayern 89% ( $n = 778$ ) und in Sachsen 80% ( $n = 469$ ) der Eltern (zusammen:  $n = 1\,247$ ) gegeben. Über beide Länder hinweg liegen vollständige Testdatensätze von T1 bis T3 für 83% der Ursprungsstichprobe vor (Bayern:  $n = 753$ ; Sachsen:  $n = 448$ ). Fast alle Eltern (99%), die der ersten Teilnahme ihrer Kinder zugestimmt haben, haben im Verlauf der Erhebung mindestens einmal an den schriftlichen Befragungen teilgenommen. Damit liegen für die Angaben zur sozialen Herkunft, die in jede Befragungswelle aufgenommen wurden, nahezu vollständige Angaben vor.

**Soziale Herkunft.** Ein für die Analysen in KOALA-S zentrales Merkmal der sozialen Herkunft ist der *Bildungsstatus der Schülereltern*, der über die Angaben der Elternbefragungen ermittelt wurde. Die Abschlüsse wurden zu drei Kategorien zusammengefasst (nicht höher als ‚Hauptschule/bis zur 9. Klasse‘, ‚mittlerer Abschluss/bis Klasse 10‘, ‚Hochschulreife/Abitur‘). Für die Analysen verwendet wird der höchste erreichte Abschluss im Haushalt. Aufgrund der unterschiedlichen Sekundarschulsysteme in beiden Bundesländern unterscheidet sich der Anteil von Familien mit Hauptschul- und mittlerem Abschluss zwischen den Stichproben deut-

lich, nicht jedoch der Anteil von Eltern mit Hochschulreife. Während in Sachsen gut 60 % der Eltern über einen mittleren Abschluss verfügen und nur 5 % über einen Hauptschulabschluss, sind in Bayern die Haushalte mit Hauptschulabschluss mit 27 % in der Stichprobe deutlich stärker vertreten, die Gruppe der Eltern mit mittlerem Abschluss beträgt hier 36 %.

Die *berufliche Position* der Eltern wurde über eine offene Angabe zum ausgeübten Beruf und geschlossene Angaben zur Sozialversicherungsklasse, der Weisungsbefugnis und dem zeitlichen Umfang der Tätigkeit erfasst. Aus diesen Angaben wurde der *sozioökonomische Status* (ISEI) (Ganzeboom & Treiman, 1996) gebildet. Auch bezüglich der Verteilung des ISEI ergeben sich Länderunterschiede: Der mittlere ISEI liegt in Bayern über dem Wert in Sachsen, die Streuung ist dagegen gleich ( $M = 46.0$  vs.  $M = 43.3$ ;  $SD = 16.3$ ).

Das *Haushaltsnettoeinkommen* der Schülerfamilien wurde im Elternfragebogen über ein Item mit geschlossenen Antwortvorgaben zu neun aufsteigenden Einkommensgruppen erfasst. Die Verteilung auf die Einkommensgruppen unterscheidet sich zwischen beiden Ländern ebenfalls. Während in Sachsen 30 % der Familien angeben, ein monatliches Nettoeinkommen von unter 1 500 € zur Verfügung zu haben, sind dies in Bayern nur etwa 15 %. Über 3 000 € im Monat verdienen in Sachsen nur 17 %, in Bayern dagegen 29 % aller Eltern. Insgesamt sind damit in Bayern die Einkommen höher als in Sachsen.

Entsprechend der Konzeption des ISEI-Index zeigen sich deutliche Beziehungen zwischen dem sozioökonomischen Status der Eltern und deren Bildungsstatus ( $r = .51$ ) sowie dem Haushaltseinkommen ( $r = .48$ ). Die Beziehung zwischen dem Einkommen und dem Bildungsstatus der Eltern fällt etwas geringer aus ( $r = .37$ ). Wir verwenden in den nachfolgenden Analysen alle drei Indikatoren für das Konstrukt „soziale Herkunft“, da die drei Indikatoren neben einem gemeinsamen Varianzanteil jeweils noch spezifische Dimensionen sozialer Ungleichheit in der Stichprobe abbilden.

Schließlich bestehen auch bezüglich der erlernten *Muttersprache* der Kinder und den Anteilen an Familien mit *Migrationshintergrund* Unterschiede zwischen den Stichproben in beiden Bundesländern. Der Anteil an Schülern mit anderer Muttersprache als Deutsch liegt in Sachsen laut den Angaben der Lehrkräfte in einem Schüleranamnesebogen bei nur 1 %, gegenüber einem Anteil in Bayern von 5 %. Ebenso verhält es sich mit den Angaben zum Migrationsstatus durch die Eltern. In der sächsischen Stichprobe kommen 6 % der Schüler aus Familien mit Migrationshintergrund, in Bayern sind es etwa 16 %. In den folgenden Analysen wird auf die Bereiche Migration und Muttersprache nicht weiter eingegangen. Dies ist zum einen durch die geringen Fallzahlen in der sächsischen Stichprobe begründet. Zum anderen zeigt sich auch in der bayerischen Stichprobe, dass sich eigenständige Effekte der migrationsbezogenen Merkmale nicht mehr finden, wenn die oben genannten Merkmale der sozialen Herkunft kontrolliert werden.

**Skalierung und Verankerung der Tests über die Erhebungszeitpunkte.** Zu den drei Messzeitpunkten wurden über Ankeritems verbundene Fachleistungstests im Leseverständnis, Rechtschreiben und der Mathematik eingesetzt.



In Tabelle 1 sind die Anzahl der eingesetzten Items und die Reliabilitätskennwerte der Testzusammenstellungen im Einzelnen angegeben. Die Kennwerte der Rechtschreibtests sind dabei durchweg am höchsten ( $\alpha > .85$ ), gefolgt von den Tests in der Mathematik ( $\alpha > .71$ ). Die Leseverständnistests weisen dagegen geringere, aber noch akzeptable Werte auf (zwischen  $\alpha > .64$  und  $\alpha > .69$ ).

Für diese Fachtests liegen im Bereich Lesen und Mathematik Eichstichproben mehrerer Grundschuljahrgänge vom zweiten bis zum vierten Schuljahr vor, aus denen jahrgangsübergreifende Schwierigkeitsparameter für die verbindenden Ankeritems gebildet werden konnten. Damit können zuverlässige Aussagen zur Entwicklung der Fachleistungen über den Erhebungszeitraum getroffen werden. Da für die Erfassung der Rechtschreibleistungen und der kognitiven Fähigkeiten noch keine Referenzstichproben zur Verfügung standen, wurde hier ein abgeändertes Verfahren verwendet. Es wurde ebenfalls ein wiederholt eingesetzter Stamm von Ankeritems verwendet, der um jahrgangsspezifisches Material unterschiedlicher Schwierigkeit ergänzt wurde. Für die Skalierung über die Zeit wurden die Schwierigkeitsparameter der dreifach eingesetzten Ankeritems verwendet.

**Veranschaulichung der Leistungswerte.** Um die Leistungen im Längsschnitt zu veranschaulichen, wurden die Testwerte auf einer Skala mit einem Mittelwert von 250 und einer Standardabweichung von 50 Skalenpunkten zu T1 abgebildet. Bei dieser Transformation ergeben sich zu T2 und T3 dem Lernfortschritt entsprechende höhere Stichprobenmittelwerte, die dann im Vergleich zur Eingangsverteilung der Leistungen und deren Streuung (*SD*) interpretierbar sind.

**Testinhalte.** Zur Erfassung des *Leseverständnisses* wurden insgesamt sechs verschiedene Lesetexte eingesetzt, die von allen Schülern zu den Testzeitpunkten bearbeitet wurden. Es handelte sich um kurze literarische Texte beziehungsweise Geschichten und altersgemäße Sachtexte. Mit den Lesetests wird sinnverstehendes Lesen erfasst. Die Anforderungen reichen hierbei von einfacher, teilweise wörtlicher Informationsentnahme (< 250 Punkten) bis hin zum Ziehen komplexer Schlussfolgerungen beziehungsweise dem Erschließen impliziter Textinhalte, die ein Verständnis des gesamten Textes voraussetzen (> 300 Punkte).

Zur Ermittlung der *Rechtschreibleistungen* wurden standardisierte Wortdiktate unterschiedlicher Länge verwendet. Diktiert wurden zwischen 48 und 64 Wörter, wobei ein großer Teil der Wörter wiederholt verwendet und ein anderer Teil jahrgangs- und schwierigkeitsbezogen variiert wurde. Einige der wiederholt eingesetzten Wörter finden sich auch im Grundwortschatz für Bayern wieder. Neben der Erfassung verschiedener regelleiteter Rechtschreibleistungen (einschließlich der korrekten Groß- und Kleinschreibung) waren auch sog. Merkwörter, die unregelmäßige Schreibungen verlangten, enthalten. Testwerte über 300 Punkten verweisen auf eine hohe Sicherheit auch beim Schreiben von unbekanntem unregelmäßigen Wörtern, Testwerte von unter 250 Punkten deuten auf Schwierigkeiten auch bei einfachen, regelmäßigen Wörtern hin (bspw. lautgetreue Wörter).

Für die *Mathematik* wurde auf Aufgabensätze aus den Orientierungsarbeiten in Bayern zurückgegriffen, die noch nicht in einer Vollerhebung eingesetzt worden waren. Inhaltlich wurden die Bereiche Arithmetik und Zahlenverständnis, Geometrie,

Größen und Messen sowie sachbezogene Mathematik getestet. Testwerte unter 250 Punkten weisen darauf hin, dass nur einfachere mathematische Routinen in den vier Bereichen beherrscht werden, Testwerte über 300 Punkte verweisen auf die Fähigkeit, verschiedene Operationen auch in unbekannteren Aufgabenkontexten sicher anwenden und ggf. auch verknüpfen zu können.

Zu jedem Messzeitpunkt korrelieren die Fachleistungen der drei Testbereiche (zu T1 alle  $r > .49$ ; T2 alle  $r > .50$ ; T3 alle  $r > .49$ ). Daher wurde für eine zusammenfassende Darstellung aus den drei Einzeltests ein gemittelter *Leistungsindex* für die drei Erhebungszeitpunkte gebildet (Cronbachs  $\alpha$  T1/T2/T3  $> .75$ ).

Der Leistungsindex der ersten Erhebung wird auch verwendet, um nach drei Schülergruppen mit unterschiedlichem Eingangsniveau zu unterscheiden. Es wurden drei Leistungsgruppen gebildet (geringe/mittlere/hohe Schulleistung), wobei die Grenzwerte für die Gruppen bei einer halben Standardabweichung unter- und oberhalb des Mittelwertes gezogen wurden.

Die *kognitiven Fähigkeiten* der Schüler wurden mit dem *kognitiven Fertigkeitstest 1-3* (KFT 1-3) (Heller & Geisler, 1983) sowie dem *Grundintelligenztest Skala 2* (CFT 20) (Weiss, 1998) erhoben. Zu T1 wurden die Subskalen Beziehungserkennen und schlussfolgerndes Denken des KFT und zu T2 das komplette Verfahren mit vier Subskalen eingesetzt. Da ein Großteil der KFT-Items in der Version 1-3 für eine Erhebung in der vierten Jahrgangsstufe zu leicht war, wurden zu T3 nur noch die schweren Items verwendet und um eine Kurzversion des CFT 20 ergänzt. Um die drei Messungen über die Erhebungszeitpunkte zu verankern, wurde mit Hilfe der Ankeritems und einer Raschskalierung analog wie beim Testbereich Rechtschreiben vorgegangen.

## 4. Ergebnisse

Im ersten Teil des folgenden Abschnitts werden Ergebnisse zum Verlauf der Leistungsentwicklung berichtet. Darauf folgen in einem zweiten Teil die Analysen zu den Effekten der Merkmale der sozialen Herkunft.

### 4.1 Entwicklung der Schülerleistungen

In allen drei getesteten Domänen stehen die Schülerleistungen der späteren Erhebungen in engem Zusammenhang mit dem Vorwissen. Dieser Zusammenhang ist bei den Rechtschreibleistungen besonders stark ausgeprägt ( $r = .78$ ,  $r = .73$ ), ist aber auch beim Leseverständnis ( $r = .53$ ,  $r = .54$ ) und den Mathematikleistungen ( $r = .62$ ,  $r = .53$ ) noch deutlich. Für den zusammenfassenden Leistungsindex fallen die Korrelationen über die Zeit etwa ebenso hoch aus wie bei den Rechtschreibleistungen ( $r = .77$ ,  $r = .80$ ).

**Tabelle 2:** Korrelationen zwischen den Testleistungen der verschiedenen Erhebungen T1 bis T3

		L. T2	L. T3	RS T2	RS T3	M. T2	M. T3	L.-I. T1	L.-I. T2	L.-I. T3
Leseverständnis	<i>r</i>	<b>.53</b>	<b>.54</b>					<b>.81</b>	<b>.59</b>	<b>.59</b>
T1 (L.)	<i>N</i>	1159	1204					1178	1177	1175
Rechtschreiben	<i>r</i>			<b>.78</b>	<b>.73</b>			<b>.78</b>	<b>.63</b>	<b>.62</b>
T1 (RS)	<i>N</i>			1161	1194			1180	1180	1176
Mathematik T1	<i>r</i>					<b>.62</b>	<b>.53</b>	<b>.79</b>	<b>.77</b>	<b>.57</b>
(M.)	<i>N</i>					1163	1181	1178	1174	1177
Leistungsindex T2	<i>r</i>	<b>.84</b>	<b>.63</b>	<b>.80</b>	<b>.71</b>	<b>.82</b>	<b>.59</b>	<b>.77</b>		
(L.-I.)	<i>N</i>	1178	119	1181	1167	1186	1184	1201		
Leistungsindex T3	<i>r</i>	<b>.61</b>	<b>.80</b>	<b>.73</b>	<b>.84</b>	<b>.63</b>	<b>.79</b>	<b>.74</b>	<b>.80</b>	
	<i>N</i>	1181	1199	1177	1167	1186	1174	1203	1199	

Alle Korrelationen (Pearson) sind mit  $p < .01$  zuverlässig.

Die Korrelationen zwischen den Tests sind für die Gesamtstichprobe der Schüler aus Bayern und Sachsen angegeben, da die Beziehungen zwischen den Ländern nicht nennenswert differieren. Die nachfolgende spezifischere Darstellung der Leistungsentwicklung in den einzelnen Bereichen erfolgt jeweils dann für beide Länder getrennt, wenn erwähnenswerte Unterschiede bestehen.

**Leseverständnis.** In den mittleren Schülerleistungen im *Leseverständnis* finden sich zu keinem Messzeitpunkt zuverlässige Unterschiede zwischen beiden Bundesländern. In Sachsen liegen die mittleren Leseleistungen zu T1 bei 252 und in Bayern bei 255 Skalenpunkten. Bei den folgenden Erhebungen steigen die mittleren Leistungen deutlich an (Sachsen: 276, 291; Bayern 275, 288). Die mittleren Zuwächse über die beiden Schuljahre haben damit eine Größenordnung von etwa drei Viertel der Standardabweichung zu T1, wobei die Zuwächse im dritten Schuljahr größer ausfallen als im vierten Schuljahr (20.9 vs. 13.3 Punkte in der Gesamtstichprobe).

Daneben finden wir in beiden Ländern eine deutliche Abnahme der Standardabweichung über die Zeit um ca. 10 Skalenpunkte (Sachsen  $SD = 48/38$ ; Bayern  $SD = 49/40$ ). Noch differenzierter zeigt Abbildung 1 die Veränderung der Leistungsverteilung im Zeitverlauf anhand der Perzentilkennwerte. Bei der Eingangsmessung in der zweiten Jahrgangsstufe sind in der bayrischen Stichprobe die Leseleistungen heterogener als in Sachsen. Sowohl der Bereich zwischen dem 10. und 90. Perzentil als auch der Interquartilabstand fällt in Bayern breiter aus. Dies geht vor allem darauf zurück, dass die Leistungsspitze der bayerischen Schüler schon in der zweiten Jahrgangsstufe ein sehr hohes Leistungsniveau erreicht, was bei den besten sächsischen Schülern erst ein Jahr später der Fall ist. Bis zum Ende der vierten Jahrgangsstufe (T3) hat sich die Leistungsverteilung in beiden Ländern jedoch weitestgehend angeglichen und stellt sich nun fast identisch dar. Besonders

auffällig ist allerdings, dass in beiden Ländern vor allem die unteren Perzentile über die Zeit stark ansteigen.

**Abbildung 1:** Verteilung der Leseleistungen in Bayern (n = 886) und Sachsen (n = 525) von T1 bis T3

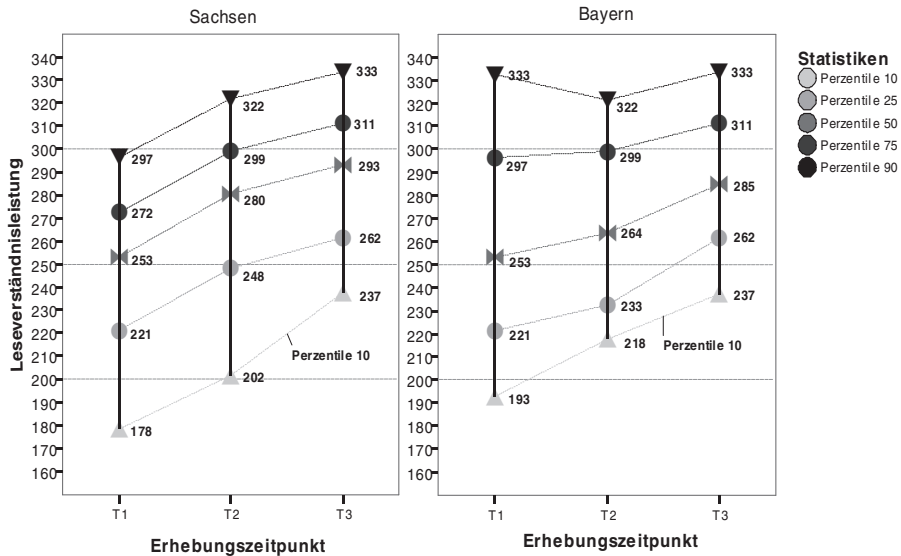
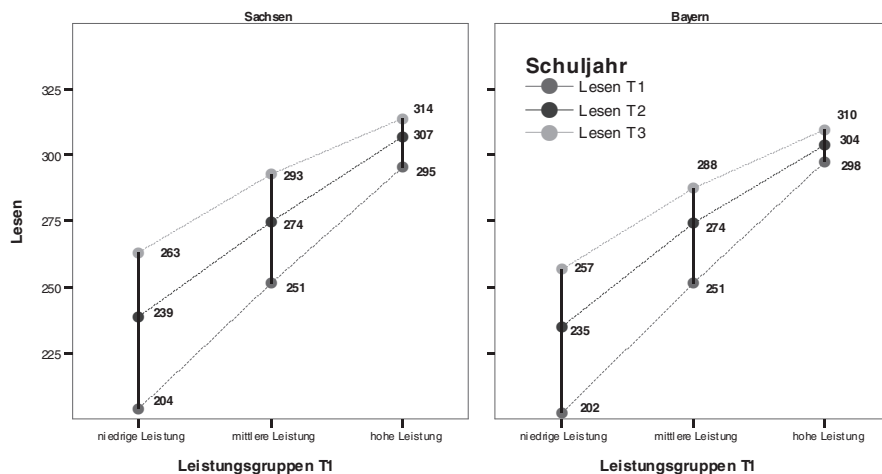


Abbildung 1 deutet bereits auf eine Gegenläufigkeit von Eingangsniveau der Leseleistungen und Leistungszuwachs hin. In der Tat weisen die Zuwächse in den Leseverständnisleistungen *negative* Korrelationen mit dem Eingangsniveau auf (bis T2:  $r = -.46$ ; bis T3:  $r = -.65$ ). Verdeutlicht wird diese Beziehung in Abbildung 2, in der die Entwicklung des Leseverständnisses über die Zeit für drei Gruppen von Schülern mit unterschiedlich hohen Eingangsleistungen (schwache, mittlere, hohe Leistungen) abgebildet ist. In Abbildung 2 ist besonders gut zu erkennen, dass in beiden Ländern die anfänglich schwachen Schüler die weitaus größten Lernfortschritte über die beiden Schuljahre erzielen (Sachsen: 59 Punkte; Bayern: 55 Punkte). Schüler mit hohen Eingangsleistungen haben sich dagegen vergleichsweise wenig in ihren Leseleistungen gesteigert (Sachsen: 19 Punkte; Bayern: 12 Punkte). Der Leistungsabstand zwischen unterer und oberer Leistungsgruppe halbiert sich damit von T1 zu T3: In Sachsen geht die Mittelwertdifferenz zwischen Schülern mit starker und schwacher Eingangsleistung von 91 Punkte auf 49 Punkte zurück, in Bayern reduziert sich der Leistungsabstand von 97 auf 53 Punkte. Gleichwohl bleibt allerdings die Rangfolge über die Zeit erhalten, die Leistungsunterschiede sind zwar sehr erheblich reduziert, aber nicht aufgehoben. Deutlich zu erkennen ist auch, dass die leistungsschwächsten Schüler in ihrem Leseverständnis am Ende der Grundschulzeit trotz des erheblichen Leistungszugewinns noch weit unter dem Niveau liegen, das die leistungsstärkste Gruppe bereits in der zweiten Jahrgangsstufe erreicht hatte; auch das

Leistungsniveau der mittleren Gruppe am Ende der dritten Jahrgangsstufe wird nicht ganz erreicht. Die beschriebenen Effekte der Leistungsentwicklung lassen sich durch gemischt-faktorielle Varianzanalysen mit dem Zwischensubjektfaktor Leistungsgruppe und dem Innersubjektfaktor Erhebungszeitpunkt absichern. Neben den Haupteffekten ist auch der Interaktionseffekt (Leistungsgruppe x Zeit) zuverlässig. Auch ein kovarianzanalytisches Vorgehen beziehungsweise Regressionsanalysen mit den Leistungsresiduen erbringen analoge Befunde.

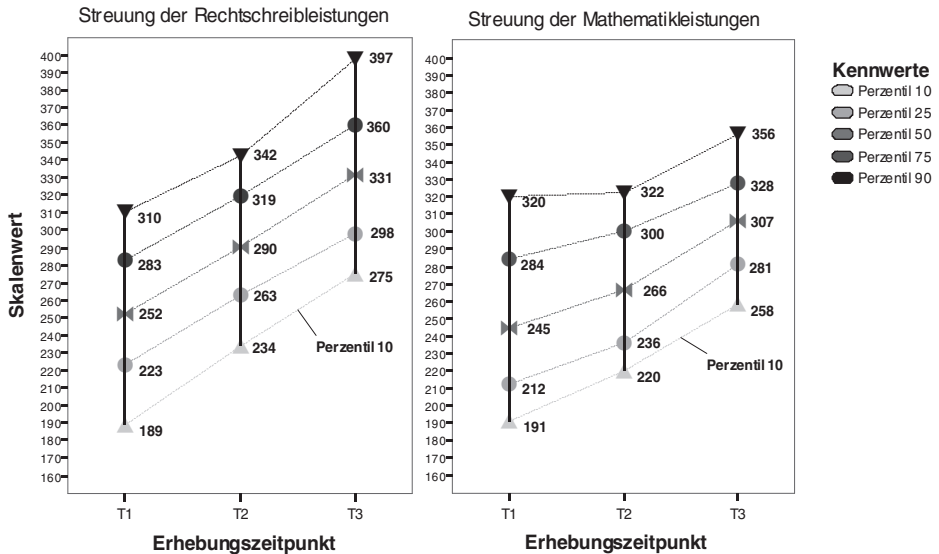
**Abbildung 2:** Veränderung der Leseleistungen in Abhängigkeit der Eingangsleistung in Bayern und in Sachsen



**Rechtschreibung und Mathematik.** Die Darstellung der Leistungsverteilungen über die drei Messzeitpunkte erfolgt für die Domänen Rechtschreibung und Mathematik für die Gesamtstichprobe der Schüler aus Bayern und Sachsen, da sich hier keine Länderunterschiede ergeben (Abbildung 3). Bei den *Rechtschreibleistungen* finden sich im Vergleich zu den anderen beiden Testbereichen die stärksten mittleren Zuwächse mit 79 Skalenpunkten im Zeitverlauf ( $M = 252/293/331$ ). Die Leistungszuwächse in der Rechtschreibleistung haben so die Größenordnung von mehr als einer 1,5-fachen Standardabweichung. Anders als beim Leseverständnis verändert sich die Leistungsstreuung im Untersuchungszeitraum nicht bedeutsam ( $SD = 50/45/48$ ). Zu erkennen ist dies auch am konstant bleibenden Interquartilabstand über die Zeit, der etwa 60 Skalenpunkte beträgt sowie am nahezu identischen Leistungszuwachs der Schüler des zehnten und neunzigsten Perzentils (86 bzw. 87 Punkte). In *Mathematik* verbessern sich die Schüler im Untersuchungszeitraum um durchschnittlich 51 Skalenpunkte ( $M = 251/269/303$ ). Der Lernzuwachs ist also auch hier mit einer Standardabweichung sehr erheblich. Gleichzeitig nimmt die Streuung der Leistungen, vergleichbar wie im Lesen und anders als im Rechtschreiben, über die Zeit deutlich ab ( $SD = 50/43/38$ ). Wie Abbildung 3 zu entnehmen ist, steigern sich die Schüler des 10. Perzentils im Zeitverlauf um 67 Punkte. Bei den Schülern des 90. Perzentils ist der Leis-

tungszuwachs zwar ebenfalls erheblich, mit 36 Punkten aber doch deutlich geringer.

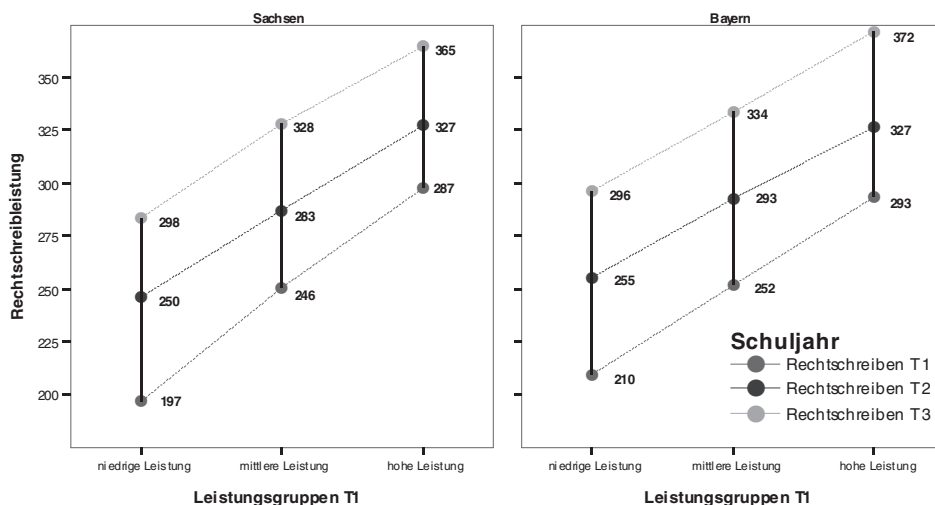
**Abbildung 3:** Verteilung der Rechtschreib- ( $N = 1\,373$ ) und Mathematikleistungen ( $N = 1\,384$ ) in der Gesamtstichprobe von T1 bis T3



Die Zuwächse in den Rechtschreibleistungen weisen ebenso wie die Zuwächse im Leseverständnis *negative* Korrelationen mit der Anfangsleistung auf (T2-Zuwachs  $r = -.42$ ; T3-Zuwachs  $r = -.37$ ). Allerdings verändern sich die Leistungsunterschiede zwischen Schülern mit niedrigen und hohen Eingangsleistungen von T1 nach T3 länderspezifisch unterschiedlich (Abbildung 4). Während sich die Leistungszuwächse für die leistungsschwachen und -starken Schüler in der bayerischen Stichprobe mit 86 vs. 79 Punkten nur wenig voneinander unterscheiden, ist der Zuwachs der leistungsschwachen Schüler in Sachsen mit 101 Punkten deutlich größer als der der leistungsstarken Schüler (78 Punkte). Von daher verringert sich in Bayern der Leistungsabstand zwischen den Gruppen über die Zeit nur wenig (von 83 auf 76 Punkte), in Sachsen geht er wesentlich stärker zurück (von 90 auf 67 Punkte). In einer Varianzanalyse mit dem Zwischensubjektfaktor Leistungsgruppe und dem Innersubjektfaktor Erhebungszeitpunkt lassen sich alle Effekte absichern, wobei die Interaktion (Leistungsgruppe x Zeitpunkt) nur knapp das Signifikanzniveau erreicht.

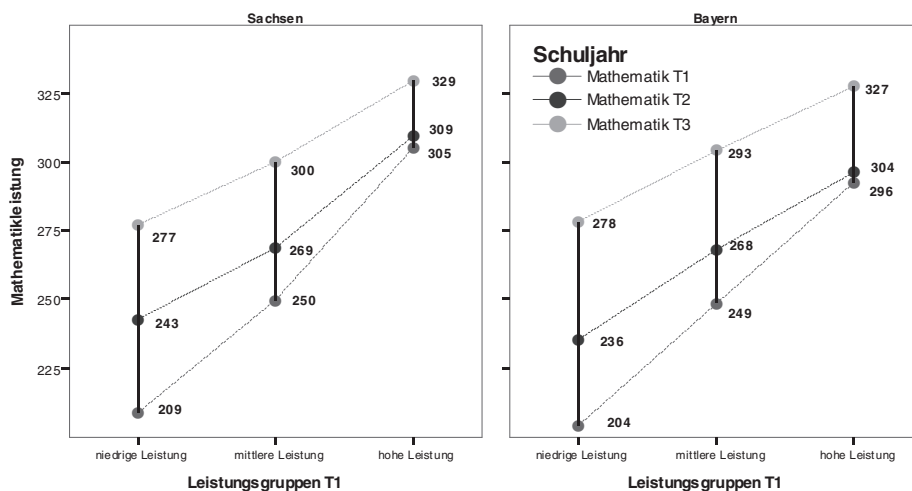
Auch der Zuwachs in den Mathematikleistungen korreliert negativ mit der Eingangsleistung (T2-Zuwachs  $r = -.54$ ; T3-Zuwachs  $r = -.66$ ). Zum Beginn der Erhebungen am Ende des zweiten Schuljahres beträgt der Leistungsabstand zwischen dem *unteren* und *oberen* Leistungsdrittel in Sachsen 96 und in Bayern 92 Skalenpunkte (Abbildung 5). In Sachsen steigert sich die anfänglich schwa-

**Abbildung 4:** Veränderung der Rechtschreibleistungen in Abhängigkeit der Eingangsleistung in Bayern und in Sachsen



che Schülergruppe um 68 Punkte, die Gruppe der leistungsstarken Schüler verzeichnet dagegen nur einen mittleren Zuwachs von 24 Punkten. Von daher reduziert sich der Leistungsabstand über die Zeit von 96 auf 52 Punkte. Ein nahezu identisches Muster findet sich für die bayerische Stichprobe. Hier beträgt der Leistungszugewinn in der unteren Leistungsgruppe 74 Punkte gegenüber 31 Punkten in der oberen Leistungsgruppe. Über die Zeit geht der Leistungsabstand von 92 Punkten auf 49 Punkte zurück. Wiederum lassen sich in einer Varianzanalyse (Leistungsgruppe x Erhebungszeitpunkt) alle Terme einschließlich der Interaktion statistisch absichern.

**Abbildung 5:** Veränderung der Mathematikleistungen in Abhängigkeit der Eingangsleistung in Bayern und in Sachsen



**Kognitive Fähigkeiten.** Nicht nur bezüglich der fachlichen Leistungen zeigen sich erhebliche Zuwächse im Erhebungszeitraum, sondern auch in den parallel erhobenen allgemeinen kognitiven Fähigkeiten. Die Zuwächse betragen hier bezogen auf die Gesamtstichprobe in etwa eine Standardabweichung ( $M = 250/282/299$ ). Eine Reduzierung der Streuung findet sich jedoch nicht ( $SD = 50/45/53$ ), ebenso wenig bestehen Länderunterschiede. Die kognitiven Fähigkeiten der Folgemessungen korrelieren hoch mit dem Eingangsniveau (T2:  $r = .55$ ; T3:  $r = .44$ ) und wiederum weisen die Zuwachswerte negative Zusammenhänge mit der Eingangsleistung auf (T2:  $r = -.57$ ; T3:  $r = -.47$ ). Auch bezüglich der kognitiven Fähigkeiten verbessern sich also die anfänglich schwachen Schüler mehr als die stärkeren Schüler. Die Korrelationen zwischen den kognitiven Fähigkeiten und den schulischen Leistungen fallen ähnlich hoch aus wie die Korrelationen der untersuchten Domänen untereinander (um  $r = .50$ ). Mit Cross-Lagged-Panel Analysen für die drei Meßzeitpunkte lässt sich diesbezüglich auch zeigen, dass zwischen den kognitiven Fähigkeiten und den schulischen Leistungen wechselseitige Beziehungen in mäßiger bis mittlerer Höhe bestehen (Pfadkoeffizienten zwischen  $\beta = .11$  und  $.29$ ). Hierbei sind die Effekte der schulischen Leistungen auf die kognitiven Fähigkeiten etwas stärker ausgeprägt als die umgekehrten Effekte.

Zusammenfassend ergeben sich in allen Testbereichen in beiden Ländern sehr erhebliche Leistungszuwächse in der zweiten Hälfte der Grundschulzeit. Die größten Zuwächse ergeben sich für die Rechtschreibleistungen. Die Leistungen am Ende der vierten Jahrgangsstufe korrelieren in allen Testbereichen deutlich mit der Eingangsleistung, auch hierbei zeigen sich stärkere Zusammenhänge im Rechtschreiben als in Mathematik und im Leseverständnis. Im Zeitverlauf nimmt die Heterogenität der Schülerleistungen bezogen auf das Leseverständnis und die Mathematik ab. Insbesondere anfänglich schwache Schüler zeigen hier höhere Zuwächse als anfänglich starke Schüler. Mit Blick auf die Rechtschreibleistungen zeigt sich eine solche ausgleichende Tendenz nur in Sachsen.

## 4.2 Schulische Leistungen und soziale Herkunft

Um die Effekte der sozialen Herkunft auf die Schülerleistungen zu untersuchen, wurden in einem ersten Schritt in multiplen Regressionen die Wirkungen der elterlichen Schulbildung, des Sozioökonomischen Status (ISEI) und der Höhe des Familieneinkommens auf die Fachleistungen getrennt nach den Erhebungszeitpunkten analysiert. In Tabelle 3 sind die Ergebnisse für die zweite und vierte Jahrgangsstufe dargestellt, im Anhang finden sich die Analysen für die dritte.

In *Bayern* lassen sich am Ende des zweiten Schuljahres 11 beziehungsweise 9% der Varianz der Lese- und Mathematikleistungen auf die soziale Herkunft der Schüler zurückführen. Fast identische Werte finden sich für das Ende der vierten



**Tabelle 3:** Regressionen der Leistungen zu T1 und T3 auf die Herkunftsmerkmale (Sachsen:  $n = 403$ ; Bayern:  $n = 648$ )

	Leseverständnis				Rechtschreiben				Mathematik									
	Sachsen		Bayern		Sachsen		Bayern		Sachsen		Bayern							
	B	beta	p <	B	beta	p <	B	beta	p <	B	beta	p <						
<b>Leistungen T1</b>																		
Konstante	209.5	.001	220.0	.001	205.3	.001	228.8	.001	190.8	.001	219.8	.001						
Bildungsstatus	10.6	.12	.05	14.8	.22	.001	16.1	.16	.08	9.8	.17	.001	16.2	.17	.01	10.3	.17	.001
Soziök. Status	0.2	.00	n.s.	8.0	.16	.001	-0.9	-0.2	n.s.	5.8	.12	.01	-3.8	-0.7	n.s.	7.1	.15	.005
Familien-einkommen	3.0	.10	.08	0.6	.02	n.s.	0.7	.02	n.s.	0.7	.03	n.s.	5.0	.16	.01	1.6	.04	ns
$R^2$	.028		.112		.020		.069		.056		.090							
<b>Leistungen T3</b>																		
Konstante	258.8	.001	254.8	.001	284.3	.001	304.2	.001	255.3	.001	267.3	.001						
Bildungsstatus	9.3	.14	.05	11.2	.22	.001	14.7	.17	.005	10.0	.17	.001	12.9	.19	.001	8.9	.18	.001
Soziök. Status	2.7	.08	n.s.	4.9	.12	.01	1.8	.04	n.s.	7.8	.16	.001	-0.7	-0.2	n.s.	1.4	.04	n.s.
Familien-einkommen	2.0	.09	n.s.	1.4	.06	n.s.	1.7	.06	n.s.	1.8	.07	n.s.	3.4	.15	.01	3.4	.15	.001
$R^2$	.049		.105		.042		.100		.074		.086							

\*Bildungsstatus: 1 = nicht höher als HS, 2 = RS; 3 = GY; Soziök. Status (ISEI, z-standardisiert), Familieneinkommen von 1 = unter 500€ bis 9 = 6000€ und mehr, s. Methode; hervorgehobene Koeffizienten sind mit  $p < .05$  zuverlässig.

Jahrgangstufe. Bei den Rechtschreibleistungen finden wir einen leichten Anstieg in der Varianzaufklärung durch die drei Herkunftsmerkmale über die Zeit von zunächst 7% auf schließlich 10%. In allen drei Domänen erweist sich der höchste Schulabschluss der Eltern, gefolgt vom sozioökonomischen Status (ISEI) als der bedeutsamste Prädiktor. Schüler aus Familien mit höherem Bildungs- und sozialem Status erzielen die höheren Fachleistungen. Zu T<sub>3</sub> ergibt sich noch ein eigenständiger Effekt des Familieneinkommens auf die Mathematikleistungen. In *Sachsen* ist die Varianzaufklärung zu T<sub>1</sub> in allen Domänen geringer als in Bayern. Zwischen 2% der Leistungsvarianz im Rechtschreiben und 5,6% der Leistungsvarianz in Mathematik sind auf die Herkunftsmerkmale rückführbar. Der Effekt des elterlichen Bildungsstatus ist in allen drei Domänen zuverlässig, bei den Mathematikleistungen ergibt sich zusätzlich ein Effekt des Einkommens. Der Anteil der durch die Herkunftsmerkmale erklärten Varianz steigt über die Schulzeit in Sachsen etwas an (beim Rechtschreiben auf 4,2%, in der Mathematik auf 7,4%), liegt aber jeweils unter den Werten, die in der bayerischen Stichprobe ermittelt wurden. Auch in Sachsen ist von den Herkunftsmerkmalen der Bildungsstatus der bedeutsamste Prädiktor der Leistungen. In der Mathematik ergibt sich wiederum ein zusätzlicher Effekt in Form von Leistungsvorteilen von Schülern aus Familien mit höherem Familieneinkommen. Die querschnittlichen Analysen zu Beginn und zum Ende des Erhebungszeitraums ergeben damit in beiden Ländern strukturell ähnliche Beziehungen zwischen sozialer Herkunft und schulischer Leistung, die Stärke der Zusammenhänge ist in Sachsen allerdings geringer als in Bayern.

Die Ergebnisse aus längsschnittlichen Analysen zum Einfluss der Herkunftsmerkmale auf die schulischen Leistungen sind in Tabelle 3 wiedergegeben. Die untere Hälfte der Tabelle beinhaltet die Ergebnisse aus Regressionsanalysen, in denen nur die Eingangsleistungen als Prädiktoren der am Ende der Grundschulzeit erreichten Leistungen berücksichtigt wurden. Der oberen Tabellenhälfte können die Effekte der Herkunftsmerkmale unter Kontrolle der Eingangsleistungen entnommen werden. Wie die Ergebnisse zeigen, sind die Eingangsleistungen erwartungsgemäß die mit deutlichem Abstand besten Prädiktoren für den letztlich erreichten Leistungsstand. Die Leistungsunterschiede zwischen Schülern, die zu T<sub>1</sub> der oberen und der unteren Leistungsgruppe angehört haben, betragen zu T<sub>3</sub> unter Kontrolle der sozialen Herkunft im Lesen in beiden Ländern etwa 40 Skalenpunkte.<sup>1</sup> Für Rechtschreibung ergibt sich eine Differenz von 62 Punkten in Sachsen und 72 Punkten in Bayern. In Mathematik schließlich resultieren etwa 40 Punkte in Sachsen und 36 Punkte in Bayern. In beiden Ländern finden sich über die Effekte der Eingangsleistungen hinausgehende Effekte der Herkunftsmerkmale. Die zusätzliche Berücksichtigung der Herkunftsmerkmale verbessert die Regressionsmodelle

---

1 Da die T<sub>1</sub>-Leistungen in der Regression als z-standardisierte Werte eingegeben wurden, entspricht das Regressionsgewicht  $B$  der Veränderung der Leistung zu T<sub>3</sub>, wenn sich der T<sub>1</sub>-Leistungswert um eine Standardabweichung ändert. Um die zu erwartenden Leistungen für T<sub>3</sub> für die obere vs. untere Leistungsgruppe zu ermitteln, ergibt sich ein Abstand von  $2x B$ , da deren mittlere Leistungen zu T<sub>1</sub> um zwei Standardabweichungen auseinander liegen (vgl. Text bzw. Abbildung 2).

in allen drei Domänen in beiden Ländern zwar nur geringfügig, aber zuverlässig. In Sachsen lassen sich zwischen 1.0 % und 1.5 % zusätzlicher Leistungsvarianz erklären, in Bayern zwischen 1.5 % und 2.7 %. Obwohl diese Effekte nicht sehr groß sind, sind sie bemerkenswert, denn Wirkungen der sozialen Herkunft werden ja bereits durch die Berücksichtigung der unterschiedlichen Eingangsleistungen gebunden. Betrachtet man die relevanten Herkunftsmerkmale im Einzelnen, so lässt sich in Sachsen nur für Mathematik der Effekt des Bildungsstatus statistisch absichern. In Bayern zeigt sich für die Leseleistungen ein zuverlässiger Effekt des Bildungsstatus der Eltern ( $\beta = .11$ ), im Rechtschreiben ein Effekt des sozioökonomischen Status ( $\beta = .08$ ). In Mathematik ergeben sich für Bayern Effekte des Bildungsstatus und der Einkommenshöhe ( $\beta = .11$  bzw.  $\beta = .13$ ).

Auf den ersten Blick vermitteln die Ergebnisse der Regressionsanalysen im Querschnitt (Tabelle 3) und Längsschnitt (Tabelle 4) ein unterschiedliches Bild: Nach den Ergebnissen in Tabelle 3 nehmen die auf die Herkunftsmerkmale rückführbaren Varianzanteile über die Zeit nicht in allen Testbereichen zu. Trotzdem zeigen die Ergebnisse in Tabelle 4, dass Herkunftseffekte unter Kontrolle der Eingangsleistungen bestehen, sich also die schulischen Leistungen in Abhängigkeit vom sozialen Status im Verlauf der beiden untersuchten Schuljahre auseinander entwickeln. Wie diese zunächst paradox anmutenden Ergebnisse zu Stande kommen, soll beispielhaft für die Effekte des Bildungsstatus differenzierter aufgezeigt werden.<sup>2</sup> Die Entwicklungsverläufe werden für den gemittelten Leistungsindex dargestellt, der die Verläufe in den drei Domänen gut repräsentiert. Auf eine Differenzierung nach Ländern wird verzichtet, da sich für beide Länder übereinstimmende Muster ergeben.

Der Mittelwert für den Leistungsindex liegt zu T1 für die Gesamtstichprobe bei 251 und zu T3 bei 307 Punkten. Wie bei den einzelnen Testbereichen nimmt auch beim Leistungsindex die Streuung im Zeitverlauf ab (von 40.5 Punkten auf 34.3 Punkte), die Reduktion der Streuung beträgt etwa 15 %. Differenziert nach *Eingangsleistung* (unteres, mittleres, hohes Niveau) und elterlichem *Bildungsstatus* sind die Entwicklungsverläufe von T1 nach T3 in Abbildung 6 dargestellt. Wie die Abbildung 6 zeigt, laufen vor allem in der oberen und unteren Leistungsgruppe im Zeitverlauf die schulischen Leistungen auseinander, in der mittleren Leistungsgruppe ist das nur geringfügig der Fall. Bei (annähernd) gleicher Anfangsleistung erzielen sowohl in der oberen als auch in der unteren Leistungsgruppe Schüler aus Familien mit höheren schulischen Abschlüssen die höheren Leistungszuwächse. Zu T3 sind in der unteren Leistungsgruppe die Mittelwertsdifferenzen zwischen oberer und mittlerer beziehungsweise oberer und unterer Bildungsgruppe statistisch abzusichern. In der oberen Bildungsgruppe

2 Wir verwenden für die nachfolgende exemplarische deskriptive Analyse nur einen Indikator der sozialen Herkunft und den zusammenfassenden Leistungsindex, um die Darstellung einfach zu halten. Dafür wurde der Bildungsstatus der Eltern ausgewählt, weil dieser nach den Analysen in Tabelle 3 den (relativ) stärksten Einfluss der drei Indikatoren beinhaltet. In der weiterführenden Analyse in Tabelle 4 sind wieder alle drei Indikatoren berücksichtigt.

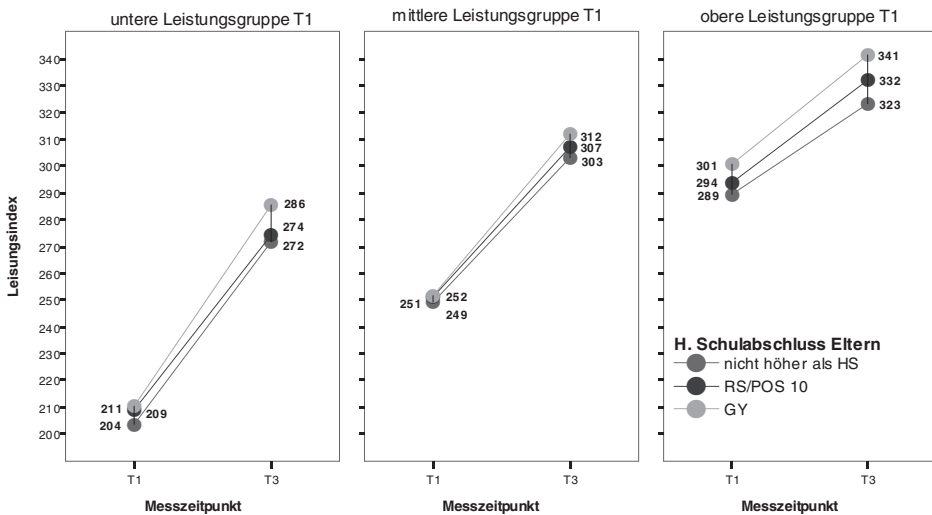
**Tabelle 4:** Regressionen der Leistungen der vierten Klasse auf die Herkunftsmerkmale unter Kontrolle der Eingangsleistungen (Sachsen: N = 403; Bayern: N = 648)

	Leseverständnis				Rechtschreiben				Mathematik			
	Sachsen		Bayern		Sachsen		Bayern		Sachsen		Bayern	
	B	beta	p <	B	beta	p <	B	beta	p <	B	beta	p <
<i>Modell mit der Eingangsleistung T1 und Herkunftsm.</i>												
Konstante	<b>274.7</b>	.001	<b>267.1</b>	.001	<b>312.8</b>	.001	<b>319.5</b>	.001	<b>279.5</b>	.001	<b>278.0</b>	.001
Leistung T1*	19.1	<b>.50</b>	.001	20.4	<b>.51</b>	.001	31.7	<b>.73</b>	.001	36.1	<b>.69</b>	.001
Bildungsstatus	5.3	.08	n.s.	5.5	<b>.11</b>	.05	4.4	.05	n.s.	3.0	.05	n.s.
Soziodök. Status	2.7	.07	n.s.	1.6	.04	n.s.	2.3	.05	n.s.	3.7	<b>.08</b>	.05
Familieneinkommen	0.9	.04	n.s.	1.1	.05	n.s.	1.3	.04	n.s.	1.2	.05	n.s.
R <sup>2</sup> Gesamt (L+H)		<b>.291</b>			<b>.330</b>			<b>.556</b>			<b>.548</b>	
R <sup>2</sup> -Zugewinn Herkunft über T1-Leistung hinaus		<b>+0.016</b>			<b>+0.020</b>			<b>+0.010</b>			<b>+0.015</b>	
												<b>+0.027</b>
<i>Modell nur mit der Eingangsleistung T1</i>												
Konstante	<b>290.0</b>	.001	<b>284.7</b>	.001	<b>328.3</b>	.001	<b>332.6</b>	.001	<b>299.8</b>	.001	<b>304.5</b>	.001
Leistung T1*	20.1	<b>.53</b>	.001	22.5	<b>.56</b>	.001	37.9	<b>.74</b>	.001	38.0	<b>.70</b>	.001
R <sup>2</sup> nur T1 Leistung		<b>.275</b>			<b>.310</b>			<b>.546</b>			<b>.533</b>	
												<b>.233</b>

\* Jeweils entsprechende Fachtestleistung zu T1 als z-Wert; hervorgehobene Koeffizienten mit p < .05 zuverlässig.  
 \*\* aus Modellen in denen nur die Leistung bzw. nur die soz. Herkunft zur Erklärung der T3-Leistungen berücksichtigt wurden; s. Anhang.

werden alle drei möglichen Mittelwertvergleiche signifikant. Man könnte vermuten, dass zumindest in der unteren Leistungsgruppe unterschiedliche Nachhilfektivitäten in Abhängigkeit der Schicht den höheren Leistungszugewinn der Schüler aus der oberen Bildungsschicht vermitteln. Zusätzliche Analysen erbringen dafür allerdings keine eindeutigen Anhaltspunkte.

**Abbildung 6:** Schülerleistungen zu T1 und T3 in Abhängigkeit von der Eingangsleistung und dem Bildungsstatus der Eltern ( $N = 1\ 081$ )



Um die Entwicklungsverläufe insgesamt und innerhalb der drei Leistungsgruppen besser nachvollziehbar zu machen, sind in Tabelle 5 die Zuwachswerte, also die Differenzen zwischen dem Leistungsindex T3 und T1, wiedergegeben. Innerhalb jeder der drei Niveaugruppen ist für Kinder aus Familien mit höherem Bildungsstatus ein *größerer* Leistungszuwachs zu verzeichnen als für Kinder aus Familien mit einem niedrigeren Bildungsstatus. Dagegen ergeben sich für die Zuwachswerte über die Niveaustufen hinweg summiert keine Unterschiede zwischen den drei Statusgruppen (Spalte Gesamt: 55.7; 54.2; 53.8). Dies geht darauf zurück, dass die Leistungszuwächse der Schüler aus der Gruppe mit niedriger Eingangsleistung höher ausfallen als die der Schüler mit hoher Eingangsleistung (Zeile Gesamt: 68.1; 57.3; 39.8) und zugleich Schüler aus Familien mit niedrigerem Bildungsstatus in dieser Gruppe anteilmäßig weitaus häufiger vertreten sind (40 %) als Schüler aus Familien mit mittlerem (29 %) oder hohem Bildungsstatus (15 %). In der Gruppe mit guten Eingangsleistungen, bei der geringere Leistungszuwächse vorliegen, verhält es sich umgekehrt, hier sind Kinder aus Familien mit niedrigem Bildungsstatus seltener vertreten (22 %) als Kinder aus Familien mit hohem Bildungsstatus (45 %). Die anfangs bestehende unterschiedliche Verteilung der Statusgruppen auf die Leistungsgruppen führt in Verbindung mit den differenziellen Zuwachswerten also dazu, dass über die Leistungsgruppen hinweg gerechnet gleiche Zuwachswerte resultieren.

**Tabelle 5:** Zuwachs im Leistungsindex nach drei Leistungsniveaus zu T1 und dem Bildungsstatus der Eltern

Bildungsstatus		Leistungsgruppe zu T1			Gesamt
		untere	mittlere	obere	
HS	<i>M</i>	67.7	54.3	35.9	55.7
	<i>SD</i>	26.7	19.7	20.4	26.1
	<i>N (%)</i>	74 (40%)	71 (38%)	40 (22%)	185 (100%)
RS	<i>M</i>	65.4	56.0	39.4	54.2
	<i>SD</i>	23.8	21.9	24.4	25.3
	<i>N (%)</i>	142 (29%)	217 (44%)	135 (27%)	494 (100%)
GY	<i>M</i>	74.9	60.3	40.9	53.8
	<i>SD</i>	25.4	21.9	26.1	27.5
	<i>N (%)</i>	62 (15%)	158 (39%)	182 (45%)	402 (100%)
Gesamt	<i>M</i>	68.1	57.3	39.8	54.3
	<i>SD</i>	25.2	21.6	24.9	26.3
	<i>N (%)</i>	278 (26%)	446 (41%)	357 (33%)	1081

\* Differenzwerte T3- T1 des Leistungsindex, ebenfalls auf der 250-Metrik.

Dieser Zusammenhang lässt sich nachvollziehen, indem man aufschlüsselt, wie die Mittelwerte für die drei Statusgruppen zu Stande kommen. Auf den Gesamtmittelwert der unteren Statusgruppe wirkt sich stark der hohe Anteil an Kindern mit schwachen Eingangsleistungen aus (40%), die im Verlauf der beiden Schuljahre erhebliche Lernzuwächse erzielen (67.7 Punkte auf dem Leistungsindex). Von dieser anfangs leistungsschwachen Gruppe gehen in den Gesamtmittelwert der unteren Statusgruppe (55.7 Punkte) rechnerisch 27.1 Punkte ein ( $67.7 \times 0.4$ ). Wenig Bedeutung für den Mittelwert dieser Gruppe (7.8 Punkte) hat die Tatsache, dass der Lernzuwachs in der anfangs höchsten Leistungsgruppe vergleichsweise gering ist (35.9 Punkte), da ein recht geringer Teil der Kinder dieser Gruppe zuzurechnen ist (22%). Für die obere Statusgruppe stellen sich diese Beziehungen spiegelbildlich dar. Da wenige Kinder der Gruppe mit schwachen Eingangsleistungen angehören (15%), wirkt sich der hohe Lernzuwachs in dieser Gruppe (74.9 Punkte) nur wenig auf den Gesamtmittelwert von 53.8 aus (11.6 Punkte). Bedeutsamer wirkt sich hier aus (18.5 Punkte), dass sehr viele Kinder der Gruppe mit hohen Eingangsleistungen angehören (45%), auch wenn in dieser Gruppe der Leistungszuwachs unterdurchschnittlich ist (40.9). Wären die Statusgruppen gleich auf die Leistungsgruppen verteilt gewesen, würden die differentiellen Zuwächse auch in den Gesamtmittelwerten aufscheinen.

Zur Vervollständigung dieser Analysen sind in Tabelle 6 die Ergebnisse aus zwei längsschnittlichen Regressionsanalysen für den Leistungsindex wiedergegeben. In Modell 1 werden nur die Eingangsleistungen als Prädiktor der Endleistungen berücksichtigt. Dadurch lassen sich 55% der Leistungsvarianz zu T3 erklären. In Modell 2 werden zusätzlich die Merkmale der sozialen Herkunft einbezogen. Auf diese gehen weitere 1.9% der Varianz zurück. Neben dem Effekt des elterlichen

Bildungsstatus ( $\beta = .07$ ) ist der Effekt der Einkommenshöhe zuverlässig ( $\beta = .10$ ). Die zusammenfassende Regressionsanalyse bestätigt damit (wie Abbildung 6 bzw. Tabelle 5), dass die untersuchten Herkunftsmerkmale über die Eingangsleistungen hinaus Einfluss auf das erreichte Leistungsniveau am Ende der Grundschulzeit haben.

**Tabelle 6:** Regression des Leistungsindex T3 auf die Eingangsleistung und Herkunftsmerkmale

	Modell 1 (Eingangsleistung)			Modell 2 (Herkunft hinzu)		
	B	beta	p <	B	beta	p <
Konstante	<b>305.7</b>		.001	<b>288.1</b>		.001
Leistungsindex T1	32.5	<b>.74</b>	.001	30.8	<b>.70</b>	.001
Bildungsstatus				3.2	<b>.07</b>	.01
Sozioök. Status				0.5	.02	n.s.
Familieneinkommen				2.1	<b>.10</b>	.001
$R^2$ Gesamt		<b>.549</b>			<b>.568</b>	
$R^2$ -Zugewinn Modell2					<b>(+.019)</b>	

\* Leistungsindex zu T1 als z-Wert; hervorgehobene Koeffizienten mit  $p < .05$  zuverlässig.

## 5. Zusammenfassung und Diskussion

Die Leistungen und Leistungsreihen weisen über den Untersuchungszeitraum von zwei Jahren eine vergleichsweise hohe Stabilität auf. Für den zusammenfassenden Leistungsindex liegt die Korrelation zwischen Vor- und Nachtest bei  $r = .74$ , ebenso hoch ist sie für die Rechtschreibleistungen. Mathematik und Lesen liegen mit  $r = .53$  beziehungsweise  $r = .54$  etwas unter diesem Wert. Mit den vorliegenden Daten lassen sich Aussagen zur Stabilität der Leistungsdifferenzen über die letzten zwei Grundschuljahre machen, wobei man davon ausgehen kann, dass ebenfalls bereits deutliche Unterschiede zwischen den Schülern zum Zeitpunkt der Einschulung vorliegen. Nach den Befunden der Scholastik-Studie wäre zu erwarten, dass zum Zeitpunkt der Einschulung erhobene Leistungsindikatoren ebenfalls deutliche Zusammenhänge mit den Leistungen am Ende der Grundschulzeit erkennen lassen sollten, diese aber doch geringer ausfallen würden als die hier berichteten Stabilitätskoeffizienten für die kürzeren Zeiträume. Anders formuliert: Die Leistungen am Ende der Grundschulzeit lassen sich durch vorangegangene, zeitlich nähere Messungen etwas besser erklären als durch zeitlich frühere Indikatoren.

Der Fokus auf die zeitliche Stabilität von interindividuellen Leistungsdifferenzen zu verschiedenen Messzeitpunkten und die Formulierung „zunehmende Stabilisierung interindividueller Unterschiede“ während der Grundschulzeit legt (oberflächlich betrachtet) ein eher deterministisches Entwicklungskonzept schulischer Kompetenzen nahe, in dem die Beschulung selbst eher eine untergeordnete Rolle zu spielen scheint. Dass diese „Stabilisierung“ zum Ende der Primarstufe

aber auch in Teilen ein Ergebnis der Beschulung sein könnte, beziehungsweise zu welchem Anteil die Grundschule die Leistungsheterogenität selbst erst generiert oder auch reduziert, gerät damit aus leicht aus dem Blick. Auch bei den hier berichteten Zweijahresstabilitätskoeffizienten ist darauf hinzuweisen, dass damit die letztlich zu T<sub>3</sub> erzielten Schulleistungen zwar nicht unerheblich, aber doch nicht annähernd vollständig determiniert sind. Außerdem sagt eine hohe Stabilität der Leistungsrangreihen nicht viel über das am Ende erreichte Ausmaß der Leistungsheterogenität aus.

Für die Rechtschreibleistungen wurde bezüglich der Streuung keine Veränderung festgestellt. Die Abstände zwischen anfangs leistungsschwächeren und -stärkeren Schülern nehmen in der sächsischen Stichprobe etwas ab, in Bayern bleiben sie nahezu unverändert bestehen. Anders verhält es sich in Lesen und Mathematik. In beiden Domänen ist in beiden Ländern eine Reduzierung der Streuung und sind geringere Interquartilabstände festzustellen. In beiden Domänen gewinnen die leistungsschwächeren Schüler erheblich mehr hinzu als die leistungsstärkeren. Bezüglich dieser beiden grundlegenden Basisfähigkeiten wird über die Grundschulzeit ein Ausgleich der anfangs bestehenden gravierenden Leistungsunterschiede erreicht. Weitgehend stabil bleiben allerdings die Leistungsrangreihen, zudem verbleiben erhebliche Niveauunterschiede: Die Schüler der anfangs unteren Leistungsgruppe erzielen im Lesen, trotz des größeren Lernfortschrittes über die zwei Schuljahre hinweg, am Ende der Grundschulzeit gerade das Leistungsniveau, das die mittlere Leistungsgruppe ein Schuljahr zuvor erreicht hatte, und ihre Leistungen liegen immer noch deutlich unter den Eingangsleistungen der leistungsstärksten Gruppe. In Mathematik sieht es ganz ähnlich aus. Abgesehen vom Bereich Rechtschreibung hat die Grundschule somit erheblich für Ausgleich gesorgt, ohne allerdings das in der zweiten Jahrgangsstufe bestehende Gefälle in den Kernleistungsbereichen vollständig abzubauen. Es bleibt offen, ob das angestrebt ist und überhaupt erreichbar wäre.

Inwieweit kann der Effekt der Regression zur Mitte bei wiederholten Messungen die hier beobachteten Effekte erklären? Zunächst ist anzumerken, dass wir – wie bei vielen anderen wiederholten Messungen – tatsächlich eine Regression zur Mitte in allen Leistungsbereichen vorgefundenen haben. Wie Nachtigall und Suhl (2002) anschaulich darlegen, handelt es sich dabei aber entgegen der verbreiteten Meinung nicht um ein Artefakt in dem Sinne, dass eine Verfälschung der Daten oder ein Messfehler vorliegt, sondern um *ein tatsächlich beobachtbares* Datenmuster, dass sich daraus ergibt, dass zwei positiv linear, *aber nicht perfekt* korrelierende Maße im Zeitverlauf wiederholt gemessen werden. Bei mittlerer Eingangsleistung besteht für den Anteil der Testergebnisse, der nicht durch die Vortestleistung determiniert ist (der Residualanteil), eine ausgewogene Wahrscheinlichkeit etwas über oder unter dem ersten Testergebnis zu liegen (so sind Residuen definiert). In den extremen Leistungsbereichen des Vortests (sehr gute Leistung zu T<sub>1</sub> oder sehr schlechte Leistung zu T<sub>1</sub>) ist dagegen die Wahrscheinlichkeit, dass die residualen



Abweichungen beim Nachtest unter beziehungsweise über dem ersten Testergebnis liegen etwas höher als die Wahrscheinlichkeit beim Nachtest ein noch extremeres Ergebnis als beim Vortest zu erzielen. Die Folge ist eine leichte Reduktion der Varianz der Nachtestleistung. Insofern haben wir tatsächlich eine Regression zur Mitte beobachtet. Allerdings kann das relativ allgemeine Phänomen der Regression zur Mitte nicht die vorgefundene Form (Ausgleich vor allem im unteren Leistungsbereich) und die Stärke der Varianzreduktion im Leseverständnis und der Mathematik erklären. Wir haben die starke Vermutung, dass die Abnahme der Leistungsheterogenität bei gleichzeitigem Anstieg des Leistungsniveaus auch aufgrund des Unterrichts in der Grundschule in der zuvor berichteten Form ausfällt.

Die geringeren Zuwächse der anfangs leistungsstarken Schüler auf „Deckeneffekte“ aufgrund zu leichter Testverfahren zu T<sub>3</sub> zurückzuführen, erscheint ebenfalls nicht überzeugend. Die Leseverständnistests, die zu T<sub>2</sub> und T<sub>3</sub> eingesetzt wurden, hatten ebenso wie die Mathematiktests und die Verfahren zur Überprüfung der Rechtschreibleistungen eine ausreichende Anzahl schwerer Items. Zu T<sub>3</sub> ergibt sich bei beiden Verfahren zwar eine etwas „rechtssteile“ Verteilung, die schwersten Items der Verfahren wurden aber in beiden Testbereichen nur von sehr wenigen Schülern gelöst (in der Mathematik von 3 %, im Leseverständnis von 8 %).

Für die unterschiedlichen Verläufe im Leseverständnis und der Rechtschreibung im Fach Deutsch bietet sich auch eine inhaltlich plausible Erklärung an: Es ist durchaus möglich, dass die anfänglich guten Schüler sehr früh ein Leseniveau erreicht haben, das aus Sicht der Lehrkräfte keiner weiteren Förderung mehr bedarf. Wenn diese Schüler bereits hinreichend lesen können, erscheint es möglicherweise sinnvoller, sie stattdessen in anderen Teilbereichen des Faches zu fördern, wie dem Aufsatzschreiben, der formalen Sprachbeherrschung oder eben der Rechtschreibung.

Eingangs ist dargelegt worden, dass es unterschiedliche Vorstellungen zum Verlauf der Leistungsentwicklung im Grundschulbereich gibt: Die Schule könnte anfangs bestehende Leistungsunterschiede vergrößern (Matthäus-Prinzip), sie könnte auch keinerlei Effekte auf die eingangs bestehenden Unterschiede haben, so dass diese im vollen Umfang bestehen bleiben, oder sie könnte egalisierend wirken und zur Verringerung von Leistungsdifferenzen beitragen. In den vorgestellten Analysen finden wir überwiegend Hinweise auf egalisierende Tendenzen, trotz der hohen zeitlichen Stabilität der interindividuellen Unterschiede. Die Tatsache, dass sich aus dem Niveau der Vortestleistung die Leistungen am Ende der Grundschulzeit am besten vorhersagen lässt, schließt nicht aus, dass sich die Bandbreite der zu T<sub>3</sub> beobachteten Leistungen in der Population im Vergleich zu T<sub>1</sub> deutlich verkleinert und sich alle Schüler auf einem höherem Leistungsniveau als zu Beginn der Erhebung befinden.

Sowohl in Sachsen als auch in Bayern erzielen Schüler der höheren Statusgruppen in allen Leistungsbereichen die besseren Ergebnisse. Die Herkunftseffekte sind dabei in Bayern stärker ausgeprägt als in Sachsen. Während auf die Herkunftsmerkmale in Bayern am Ende der zweiten Jahrgangsstufe ca. 9–10 % der

Varianz zurückgeführt werden können, betragen die Anteile in Sachsen ca. 5–7%. Allerdings nähern sich die Effekte der sozialen Herkunft in Sachsen bis zur vierten Jahrgangsstufe den Werten, die für Bayern gefunden wurden, etwas an.

Überdies finden sich in beiden Ländern Effekte der sozialen Herkunft unter Kontrolle der Vortestleistungen. Die Herkunftsmerkmale erklären zwischen 1–2% an Varianz über die Vortestleistungen hinaus. Diese Effekte werden auf Grund der negativen Korrelation zwischen Eingangsleistung und Leistungszuwachs erst bei einer Betrachtung der Leistungszuwächse differenziert nach Eingangsleistung und sozialem Status deutlich. Ein größerer Teil der Kinder der unteren sozialen Gruppen gehört anfangs zur Gruppe der leistungsschwächeren Schüler, die im Zeitverlauf deutlich höhere Zuwachswerte erzielen als die leistungsstärkeren. Dagegen gehört der größere Teil der Kinder aus Familien mit hohem Status der oberen Leistungsgruppe an, die geringere Leistungszuwächse erzielt. Der zu erwartende Leistungsausgleich zwischen den sozialen Gruppen tritt nun gerade deswegen nicht ein, weil *innerhalb* aller Leistungsgruppen die Kinder aus den Familien mit höherem sozialem Status größere Leistungszuwächse erreichen. Besonders die Kinder aus den Familien mit dem höchsten Bildungsstatus setzen sich damit über die Zeit von den anderen ab.

Im Gesamtergebnis, d. h. bezogen auf die anfänglich beobachtete Heterogenität der Leistungen, zeigt sich damit ein durchaus ausgleichender Effekt in den letzten zwei Grundschuljahren. Allerdings reicht das nicht aus, um die Leistungsunterschiede völlig auszugleichen. Obwohl die Schülerleistungen sich angleichen, bleiben die Leistungsrangreihen im Wesentlichen erhalten. Die Herkunftseffekte werden dabei nicht reduziert, sie nehmen bei gleichen Eingangsleistungen sogar etwas zu, und es ist zu vermuten, dass der erhebliche Leistungsrückstand der Schüler aus den unteren Schichten am Ende der Grundschulzeit sich auf die weitere Leistungsentwicklung in der Sekundarstufe auswirken wird.

Diesbezüglich kann nochmals daran erinnert werden, dass die Grundschule zwei widersprüchliche Aufgaben zu bewältigen hat. Sie soll allen Kindern eine solide Basis für den weiteren Bildungsweg verschaffen und doch auch eine Verteilung auf unterschiedlich anspruchsvolle Schullaufbahnen erreichen. Darüber, wie die hier berichteten Ergebnisse vor diesem Hintergrund zu werten sind, lässt sich empirisch nicht entscheiden. Zumindest wird man aber sagen dürfen, dass die Grundschule ihre Aufgaben jedenfalls nicht schlecht erfüllt.

## Literatur

- Baumert, J., Roeder, P.M., Sang, F. & Schmitz, B. (1986). Leistungsentwicklung und Ausgleich von Leistungsunterschieden in Gymnasialklassen. *Zeitschrift für Pädagogik*, 32, 639–660.
- Baumert, J., Stanat, P. & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In J. Baumert & P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen*:

- Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (S. 95–188). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baumert, J., Trautwein, U. & Artelt, C. (2003). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 261–333). Opladen: Leske+Budrich.
- Baumert, J., Watermann, R. & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. *Zeitschrift für Erziehungswissenschaft*, 6(1), 46–72.
- Bayerisches Gesetz über das Erziehungs- und Unterrichtswesen (BayEUG) in der Fassung der Bekanntmachung vom 31. Mai 2000 (GVBl S. 414, ber. S. 632, BayRS 2230-1-1-UK), zuletzt geändert durch Gesetz vom 22. Juli 2008 (GVBl S.467) Zugriff am 17.08.2009 unter: [http://by.juris.de/by/gesatmmt/EUG\\_BY\\_2000.h](http://by.juris.de/by/gesatmmt/EUG_BY_2000.h)
- Beck, M., Bromme, R., Heymann, H.W., Mannhaupt, G., Skowronek, H. & Treumann, K. (1988). Gefangen im Datenlabyrinth. Kritische Sichtung eines Forschungsberichts zum schulischen Chancenausgleich. *Zeitschrift für Pädagogische Psychologie*, 2, 91–111.
- Ditton, H. (2007a). Kompetenzdiagnostik bei Übergangentscheidungen. In M. Prenzel & I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik. Zeitschrift für Erziehungswissenschaft. Sonderheft 7 / 2007*. (S. 187–199). Wiesbaden: Verlag für Sozialwissenschaften.
- Ditton, H. (Hrsg.). (2007b). *Kompetenzaufbau und Laufbahnen im Schulsystem. Eine Längsschnittuntersuchung an Grundschulen*. Münster: Waxmann.
- Ditton, H. & Krüskens, J. (2006). Der Übergang von der Grundschule in die Sekundarstufe. *Zeitschrift für Erziehungswissenschaft*, 9(3), 348–372.
- Ditton, H., Krüskens, J. & Schauenberg, M. (2005). Bildungsungleichheit – der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft*, 8(2), 285–304.
- Fend, H. (1981). *Theorie der Schule*. München: Urban & Schwarzenberg.
- Ganzeboom, H.B.G. & Treiman, D.J. (1996). Internationally Comparable Measures of Occupational Status fort he 988 International Standard Classification of Occupations. *Social Science Research*, 25(1), 201–239.
- Heller, K.A. (1997). Individuelle Bedingungsfaktoren der Schulleistung: Literaturüberblick. In F.E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 181–221). Weinheim: Psychologie Verlags Union.
- Heller, K.A. & Geisler, H.-J. (1983). *Kognitiver Fähigkeitstest (Grundschulform)*. KFT 1-3. Weinheim: Beltz.
- Kaufmann, A. (2008). *Die Rolle motivationaler Schülermerkmale bei der Entstehung sozialer Disparitäten des Schulerfolgs. Eine Längsschnittuntersuchung an Grundschulen in Bayern und Sachsen*. Berlin: Mensch und Buch.
- Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2004): Bildungsstandards im Fach Mathematik für den Primarbereich. Beschluss vom 15.10.2004. *Beschlüsse der Kultusministerkonferenz*. Bonn: KMK. Zugriff am 25.07.2009 unter: [http://www.kmk.org/fileadmin/veroeffentlichungen/\\_beschluesse/](http://www.kmk.org/fileadmin/veroeffentlichungen/_beschluesse/)
- Krüskens, J. (2007). Entwicklung von Schülerleistungen und Zensuren in der Grundschule. In H. Ditton (Hrsg.), *Kompetenzaufbau und Laufbahnen im Schulsystem* (S. 41–61). Münster: Waxmann.
- Moser, U., Stamm, M. & Hollenweger, J. (2005). *Für die Schule bereit? Lesen, Wortschatz, Mathematik und soziale Kompetenzen beim Schuleintritt*. Aarau: Sauerländer.

- Nachtigall, C. & Suhl, U. (2002). Der Regressionseffekt. Mythos und Wirklichkeit. *Metheval Report*, 4(2). Zugriff am 25.07.2009 unter: [www.uni-jena.de/svw/metheval/report/](http://www.uni-jena.de/svw/metheval/report/)
- Sauer, J. & Gattringer, H. (1985). Soziale, familiale, kognitive und motivationale Determinanten der Schulleistung. Ein Beitrag zu einem Strukturmodell der Bedingungen des Schulerfolgs bei Grundschulern. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 37(2), 288–309.
- Treiber, B. & Weinert, F.E. (1985). *Gute Schulleistungen für alle? Psychologische Studien zu einer pädagogischen Hoffnung*. Münster: Aschendorff.
- Weinert, F.E. & Helmke, A. (Hrsg.). (1997a). *Bedingungsfaktoren schulischer Leistungen*. Göttingen: Hogrefe Verlag.
- Weinert, F.E. & Helmke, A. (Hrsg.). (1997b). *Entwicklung im Grundschulalter*. Weinheim: Psychologie Verlags Union.
- Weiss, R.H. (1998). *CFT 20. Grundintelligenztest Skala 2* (4. Auflage). Göttingen: Hogrefe.
- Zöller, I. & Roos, J. (2009). Einfluss individueller Merkmale und familiärer Faktoren auf den Schriftspracherwerb. In J. Roos, & H. Schöler (Hrsg.), *Entwicklung des Schriftspracherwerbs in der Grundschule*. (S. 47–108). Wiesbaden: VS Verlag für Sozialwissenschaften.

*Hartmut Ditton*, Prof. Dr., Inhaber des Lehrstuhls für Allgemeine Pädagogik/Erziehungs- und Sozialisationsforschung, Ludwig-Maximilians-Universität München  
E-Mail: [ditton@lmu.de](mailto:ditton@lmu.de)

*Jan Krüsken*, Dr., Wissenschaftlicher Mitarbeiter am Lehrstuhl für Allgemeine Pädagogik/Erziehungs- und Sozialisationsforschung, Ludwig-Maximilians-Universität München  
E-Mail: [kruesken@edu.uni-muenchen.de](mailto:kruesken@edu.uni-muenchen.de)

**Tabelle A1:** Regressionen der Leistungen in der dritten Klasse (T2) auf die Herkunftsmerkmale

Schritt	Leseverständnis						Rechtschreiben						Mathematik							
	Sachsen		Bayern		Sachsen		Bayern		Sachsen		Bayern		Sachsen		Bayern		Sachsen		Bayern	
	B	p <	B	p <	B	p <	B	p <	B	p <	B	p <	B	p <	B	p <	B	p <	B	p <
Konstante	238.2	.001	229.6	.001	257.2	.001	264.1	.001	226.6	.001	234.5	.001	234.5	.001	234.5	.001	234.5	.001	234.5	.001
Bildungsstatus	7.7	.09	n.s.	14.8	.23	.001	11.55	.14	.05	10.9	.20	.001	12.3	.16	.01	9.5	.17	.01	9.5	.17
Soziök. Status	2.2	.05	n.s.	3.6	.07	n.s.	2.7	.04	n.s.	6.3	.15	.005	-3.2	-.07	n.s.	4.1	.10	.05	4.1	.10
Familieneinkommen	3.7	.12	.05	2.0	.07	n.s.	0.5	.02	n.s.	1.0	.04	n.s.	3.5	.13	.05	2.5	.10	.05	2.5	.10
R <sup>2</sup>	<b>.035</b>		<b>.094</b>		<b>.022</b>		<b>.101</b>		<b>.040</b>		<b>.086</b>		<b>.086</b>		<b>.086</b>		<b>.086</b>		<b>.086</b>	

\* Soziökonomischer Status (ISEI, z-standardisiert); hervorgehobene Koeffizienten mit p < .05 zuverlässig.