

**Empirische Sonderpädagogik**, 2016, Nr. 2, S. 171-188  
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

## Viel bringt viel? – Effekte lernzeitverlängernder Maßnahmen auf die Lese- und Mathematikleistungen von Grundschülerinnen und Grundschulern mit ungünstiger Lernausgangslage

*Yvonne Blumenthal, Stefan Voß, Timo Tresp & Katja Koch*

*Universität Rostock*

### **Zusammenfassung**

Der vorliegende Beitrag untersucht die Effekte von lernzeitverlängernden Maßnahmen für Grundschülerinnen und Grundschüler in Mecklenburg-Vorpommern mit einer ungünstigen Lernausgangslage zum Zeitpunkt der Einschulung. Dazu wurde über die gesamte Grundschulzeit hinweg die Leistungsentwicklung von 67 Kindern in den Bereichen Lesen und Rechnen erfasst. Bei 19 Kindern wurde eine Lernzeitverlängerung durch Diagnoseförderklassen (DFK) realisiert, bei 18 durch eine Klassenwiederholung (KW) und 30 Kinder lernten in regulären Grundschulklassen (GSK) ohne eine Lernzeitverlängerung. Die Auswertungen der Daten mittels Hierarchy-linearer Modelle (HLM) weisen auf gleiche Entwicklungsverläufe der drei Untersuchungsgruppen in den Bereichen Lesen und Mathematik hin. Zum Ende der Klasse 4 erreichten die drei Gruppen ähnliche Leistungsniveaus. In allen drei Settings fiel auf, dass die Entwicklung mathematischer Kompetenzen über die Schulzeit hinweg verzögert erfolgte. Ungünstige Lernausgangslagen im Bereich Mathematik konnten den Analysen zufolge durch keine der untersuchten Beschulungsformen ausreichend kompensiert werden.

Schlüsselwörter: Lernzeitverlängerung, Rechnen, Lesen, Grundschule, Klassenwiederholung

### **The more the better? Effects of extended learning time on reading and mathematics outcomes of primary-school students at risk**

#### **Abstract**

This study analyses the effects of procedures to extend the period of learning for primary-school students from Mecklenburg-Western Pomerania that have a disadvantageous initial learning situation at the time of school enrollment. This survey includes 67 children which show low learning prerequisites in subjects such as reading and mathematics. As a method to prevent the necessity for special education 19 students were educated in special classes (Diagnoseförderklassen, DFK) with extended learning time, 18 children received an extended learning time due to grade repetition, and 30 children visited regular primary-school classes. The HLM analysis implies no differences in the developmental trajectories of the three study groups in the areas of reading and math. In addition, there were no differences in academic achievement at the end of fourth grade regarding reading and mathematics for any of the groups. The results showed a delay of math achievement development over time and supported the fact that a disadvanta-

geous initial learning situation in mathematics cannot be compensated sufficiently in any of the approaches.

Keywords: reading, mathematics, primary school, extended learning time, grade repetition

Ausgehend von der bereits überholten Annahme, dass leistungshomogene Gruppen ein günstiges Entwicklungsmilieu bieten (Klemm, 2009), werden derzeit im deutschen Schulsystem noch immer eine Reihe von Maßnahmen umgesetzt, die darauf abzielen, Leistungshomogenität herzustellen und zu erhalten. Dazu gehören z. B. altershomogene Jahrgangsklassen sowie die Gliederung des Schulsystems in leistungsabhängige Schulformen. Sowohl empirische Befunde als auch die pädagogische Praxis zeigen jedoch, dass sich Kinder bereits zum Schuleintritt in ihren Lernvoraussetzungen deutlich unterscheiden (Koch, Hartke & Blumenthal, 2009). Daraus resultieren – vor dem Hintergrund von Schulleistungsmodellen mit besonderer Akzentuierung der Lernzeit (u. a. Bloom, 1985; Helmke, 2004) – selbst unter optimalen Lernbedingungen interindividuelle Differenzen in der Lernzeit bis zum Erreichen eines Lernziels. Für Kinder, die aufgrund ihrer Lernausgangslage oder ihrer Leistungen die curricularen Mindestanforderungen nicht erreicht haben bzw. dies mit hoher Wahrscheinlichkeit nicht tun, werden (schul-) strukturell verschiedene Möglichkeiten eröffnet, die nominale Lernzeit zu verlängern. Zentrale Grundannahme dabei ist, dass die Bereitstellung zusätzlicher Lernzeit dem Lernerfolg zuträglich ist, wobei konzeptionell nicht festgelegt ist, wie effektiv diese Zeit tatsächlich zum Lernen genutzt wird. Während die Befundlage für die lernzeitverlängernde Klassenwiederholung relativ eindeutig ist, existieren für weitere Varianten von Lernzeitverlängerung noch wenige Befunde. Ebenso fehlen differenzierte Ergebnisse der Effekte lernzeitverlängernder Maßnahmen im Hinblick auf die Gruppe schulisch gefährdeter Kinder.

### **Forschungsstand**

*Effekte verlängerter Lernzeiten auf das schulische Lernen.* Zeit gilt als wichtige Rahmenbedingung für die Sicherung des Lernerfolgs. Die sogenannte nominale Unterrichtszeit ist jene Zeit, die in einem bestimmten Zeitraum zur Verfügung gestellt wird (Hesse, 1994). Die Forschungslage über den Zusammenhang von Unterrichtszeit und Lernerfolg ist wenig eindeutig. Walberg (1986) berichtete in einer Metaanalyse mit  $r = .40$  eine lediglich geringe Korrelation und zeigte auf, dass die Ergebnisse der 31 Einzeluntersuchungen stark variierten ( $r = .13$  -  $r = .71$ ). Ebenso verweisen Hosenfeld, Helmke, Ridder und Schrader (2002) auf einen relativ geringen Zusammenhang. Auch innerhalb der PISA-Studie des Jahres 2000 wurde die nominale Lernzeit als Prädiktor für Leistung untersucht und führte zu wenig eindeutigen Befunden (Schümer, 2001).

*Effekte von Klassenwiederholungen (KW).* Eine weitere Maßnahme zur Verlängerung der Lernzeit und zur Gewährleistung von Leistungshomogenität ist die Klassenwiederholung (Repetition). Durch zusätzliche Lernzeit und die wiederholte Beschäftigung mit den Lerninhalten sollen die betroffenen Kinder Defizite ausgleichen, Lernziele (doch noch) erreichen und damit eine erhöhte Chance auf Lernerfolge haben (Bless, Schüpach & Bonvin, 2005; Klemm, 2009). Für Klassenwiederholungen zieht die empirische Bildungsforschung ein nicht ganz eindeutiges, jedoch überwiegend kritisches Fazit: Kurzfristig, im Jahr der Wiederholung, werden dabei zum Teil positive, zum Teil ausbleibende Effekte berichtet, langfristig scheint sich die Klassenwiederholung eher negativ auf die Leistungen auszuwirken. So kamen auch Krohne und Tillmann (2006) in einer zusammenfassenden Darstellung der nationalen Forschungsbe-

funde zu dem Schluss, dass Klassenwiederholungen keine nachhaltig positiven Effekte haben. Dies geht einher mit dem bereits älteren Befund von Tietze und Rossbach (1998), die den damaligen internationalen Forschungsstand aus über 60 analysierten Studien zusammenfassten. In einer schweizer Längsschnittstudie wurden an einer Stichprobe von 4690 Kindern die kurz-, mittel- und langfristigen Effekte des Sitzenbleibens untersucht und kurz- und mittelfristig sowohl positive als auch negative Auswirkungen auf die Leistungsentwicklung festgestellt (Bless et al., 2005). Insbesondere im Wiederholungsjahr konnten Kinder von der Nichtversetzung profitieren. Dennoch standen die wenigen und kurzfristigen positiven Auswirkungen in keinem guten Verhältnis zu den Schwierigkeiten, die sich aus einer Wiederholung ergaben. So investierten die Kinder nicht nur ein Jahr mehr (Lebens-) Zeit in ihre Schullaufbahn, Klassenwiederholungen hatten zudem eine „vorstrukturierende negative Wirkung auf Selektionshürden im weiteren Verlauf der Schulkarriere“ (Bless, 2006, S. 18). Auch Silbergliß, Appleton, Burns und Jimerson (2006) konnten bei ihrer Untersuchung mit Kindern der ersten bis achten Klassenstufe keine Vorteile einer Klassenwiederholung im Vergleich zu einer kontinuierlichen Versetzung bei gleichen Lernvoraussetzungen festhalten.

Ein nationaler Forschungsbefund zur Effektivität einer Klassenwiederholung stammt von Demski und Liegmann (2014). Die Autorinnen greifen auf den Datenpool von über 11000 Personen zurück, die im Rahmen der National Education Panel Study (NEPS, Blossfeld, Maurice & Schneider, 2011) untersucht wurden und resümieren: „Auf Basis der Befunde unserer Studie scheinen Klassenwiederholungen nicht negativ auf die weitere Schul- und Berufsbiografie zu wirken. Dennoch lassen sich die Ergebnisse keinesfalls dahingehend interpretieren, dass Klassenwiederholungen die Leistungsentwicklung positiv beeinflussen.“ (Demski & Liegmann, 2014, S. 188).

Weiterhin wird jedoch auch von langfristig negativen Effekten berichtet: So wächst der Leistungsabstand der Kinder, die eine Klasse wiederholen, in den Folgeschuljahren (Tietze & Rossbach, 1998), sie gehörten am Ende ihrer Schullaufbahn oft zu den Leistungsschwächeren und erzielten niedrigere Schulabschlüsse (OECD, 2006). Über eine negative Befundlage wurde auch in neueren Veröffentlichungen auf nationaler sowie internationaler Ebene für höhere Klassenstufen berichtet (9. Klassen: z. B. Ehmke, Drechsel & Carstensen, 2008; 8. Klassen: z. B. Lamote, Pinxten, Van Den Noortgate & Van Damme, 2014). Auch Hattie (2013) beurteilte Repetitionen als eine für den Lernerfolg nicht zielführende Maßnahme ( $d = -0.16$ ).

*Effekte von Diagnoseförderklassen (DFK).* Eine weitere Variante für eine schulstrukturell organisierte Lernzeitverlängerung sind sogenannte Diagnoseförderklassen (DFK). DFK gibt es in verschiedenen Bundesländern, so z. B. in Bayern und in Mecklenburg-Vorpommern. Wesentlichster konzeptioneller Bestandteil dieser Beschulungsvariante ist, neben einer geringeren Schülerzahl (Klassengrößen:  $n < 15$ ), sowie zwei bis fünf zusätzliche Förderstunden durch Sonderschullehrkräfte, die Verlängerung der Lernzeit in der Schuleingangsphase von zwei auf drei Schuljahre (Ministerium für Bildung, Wissenschaft und Kultur, 2009). Damit einher geht eine curriculare Lernstoffstreckung, die vor allem einen kleinschrittigeren Vermittlungsprozess sowie einen höheren Individualisierungsgrad ermöglicht, um Entwicklungsverzögerungen auszugleichen. Wichtige konzeptionelle Bestandteile der DFK sind zudem eine umfangreiche Feststellung der individuellen Lern- und Entwicklungsausgangslage sowie die Erstellung eines individuellen Förderplanes und dessen ständige Kontrolle. Das System gestaltet sich insgesamt durchgängig, ein Wechsel in eine reguläre Grundschulklasse ist jederzeit möglich. Generelles Ziel ist eine Umschulung in eine reguläre dritte Grundschulklasse nach dem dritten Schul-

besuchsjahr. Anzumerken ist jedoch, dass das nichtstandardisierte Auswahlverfahren der Schülerinnen und Schüler für eine DFK zum Zeitpunkt der Datenerhebung zu sehr heterogenen Klassenkompositionen führt (Koch et al., 2009), die weder intra- noch interschulisch vergleichbar sind.

Deskriptive Untersuchungen, in denen als Erfolgskriterium die Rückführungsquoten nach Beendigung der DFK in eine reguläre Grundschulklasse (GSK) angesehen wurden, ermittelten geringe Quoten zwischen 22 % (für die Jahre 1986/87 bis 1994/95; Breitenbach & Lehner, 1999) und 29.4 % (für die Jahre 1994/95 - 2004/05; Ellinger & Koch, 2007). Eine Vergleichsstudie zu den Effekten der Beschulungsform DFK im Vergleich zu GSK zeigte auf, dass durch die Beschulung in DFK nur bedingt positive Effekte auf die Leistungsentwicklung erzielt wurden (Blumenthal, Hartke & Koch, 2010). Die Leseleistungen der DFK-Kinder waren nach drei Jahren Unterrichtszeit signifikant besser als die vergleichbarer Kinder aus regulären GSK nach zwei Schuljahren (same-grade-Vergleich, d. h. jeweils auf dem gleichen curricularen Niveau versetzt um ein Schuljahr). Bei den Rechenleistungen zeigte sich ein solcher Effekt nicht. Im same-age-Vergleich (jeweils nach gleicher Schulbesuchszeit) relativiert sich dieser zumindest teilweise positive Effekt. Gefährdete Grundschülerinnen und Grundschüler zeigen dann sowohl im Lesen als auch im Rechnen bessere Leistungen als gefährdete Schülerinnen und Schüler aus DFK. Trotz einer Verlängerung der Lernzeit um ein Jahr wurden die curricularen Lernziele der Klasse 2 nur im Lesen, nicht aber im Rechnen erreicht.

Zur Beantwortung der Frage, welche Effekte zeitlich begrenzte Lernzeitverlängerungsmaßnahmen auf die Schulleistungen von Kindern mit ungünstigen Lernvoraussetzungen haben, fehlt es sowohl auf nationaler als auch auf internationaler Ebene noch an Forschungsarbeiten, obgleich diese Maßnahmen zur derzeit gängigen Praxis zählen. Die Frage, ob gefährdete Kinder von einer Lernzeitverlängerung durch Klassenwieder-

holung im Vergleich eher profitieren, als von einer Lernzeitverlängerung in temporären Sonderklassen (wie DFK), ist bisher ebenso ungeklärt. Im vorliegenden Beitrag soll daher geprüft werden, welche Effekte lernzeitverlängernde Maßnahmen, in Form von Klassenwiederholung oder DFK, auf den Entwicklungsverlauf und das erreichte Niveau in den Lese- und Mathematikleistungen bis zum Ende der Grundschulzeit haben.

### *Hypothesen*

Die hier zu untersuchenden Settings unterscheiden sich in der nominalen Lernzeit. So erhielten die Kinder einer regulären GSK insgesamt vier Beschulungsjahre, die KW und die Kinder einer DFK ein zusätzliches Schuljahr. Während eine Klassenwiederholung auf reine Repetition und damit in der Regel lediglich auf eine Lernzeitstreckung, also einen rein quantitativen Unterschied zum regulären Unterricht setzt, weicht das Konzept der DFK durch die Streckung der Lernzeit nicht nur quantitativ, sondern darüber hinaus durch einen kleinschrittigeren und stärker individualisierten Vermittlungsprozess auch qualitativ von einer Klassenwiederholung sowie einer regulären Beschulung in einer GSK ab. Vor diesem Hintergrund und unter Einbezug des Forschungsstandes hinsichtlich der Leseentwicklung von Kindern mit einer ungünstigen Lernausgangslage (hier verstanden als ungünstige schulleistungsrelevante Voraussetzungen im Bereich der Wahrnehmungsleistungen und kognitiven Kompetenzen) ergeben sich für diesen Lernbereich folgende Hypothesen:

- 1) Kinder mit einer ungünstigen Lernausgangslage, die eine lernzeitverlängernde Maßnahme durch eine Beschulung in DFK erhielten, zeigen im same-grade-Vergleich im Lesen ein höheres Leistungsniveau zum Ende der vierten Klasse sowie eine höhere Leistungsentwicklung als Kinder, die eine Klasse wiederholten (KW), diese wiederum erzielen

ein günstigeres Leistungsniveau sowie eine günstigere Leistungsentwicklung als Kinder mit einer regulären Grundschullaufbahn (GSK) ( $DFK > KW > GSK$ ).

Zwar geben die zuvor benannten ersten Forschungsbefunde von Blumenthal et al. (2010) für den Vergleich der Settings DFK und GSK Hinweise dahingehend, dass die Gruppe GSK Vorteile im Bereich Mathematik erzielen könnte, da diese Befunde sich jedoch schwer theoretisch erklären lassen und wenig konform zu anderen Ergebnissen ausfallen, wird nachfolgend eine dem Lesen analoge Hypothesenformulierung für den Bereich Mathematik gewählt:

- 2) Kinder mit einer ungünstigen Lernausgangslage, die eine lernzeitverlängernde Maßnahme durch eine Beschulung in DFK erhielten, zeigen im same-grade-Vergleich in Mathematik ein höheres Leistungsniveau zum Ende der vierten Klasse sowie eine höhere Leistungsentwicklung als Kinder, die eine Klasse wiederholten (KW), diese wiederum erzielen ein günstigeres Leistungsniveau sowie eine günstigere Leistungsentwicklung als Kinder mit einer regulären Grundschullaufbahn (GSK) ( $DFK > KW > GSK$ ).

## Methode

### Design

Die vorliegende Studie evaluiert drei unterschiedliche Beschulungsformen hinsichtlich ihrer Wirksamkeit auf die Entwicklung der Lese- und Mathematikleistungen für in ihrer Entwicklung gefährdete Kinder über die Grundschulzeit.

Für den Vergleich der beiden Gruppen mit Lernzeitverlängerung (Diagnoseförderklassen - DFK, Klassenwiederholung - KW) mit der Gruppe regulärer Grundschulklassen (GSK) wurden Hierarchisch-lineare Modelle im same-grade-Vergleich gerechnet. Die abhängigen Variablen Mathematik- sowie Leseleistung wurden über vier Schuljahre zu je einem Messzeitpunkt pro Schuljahr erhoben. Aufgrund der Konzeption der DFK, deren Ziel es ist, den geforderten curricularen Inhalt zum Ende der zweiten Klasse in drei Schuljahren zu vermitteln, gingen für diese Schülerinnen und Schüler in die Modelle keine Werte vom Ende des ersten Schuljahres ein. Da Nachtstungen der Klassenwiederholerinnen und Klassenwiederholer erst ab Klasse 3 realisiert werden konnten, gingen von diesen Fällen keine Werte aus Klasse 2 in die Modelle ein. Somit weist die Gruppe GSK Daten zu vier,

Tabelle 1: Überblick über eingesetzte Verfahren

		<b>GSK</b>	<b>DFK</b>	<b>KW</b>
Schulbeginn	Lernausgangslage	EDI, KFT 1-2	EDI, KFT 1-2	EDI, KFT 1-2
1. Schulbesuchsjahr (2006/ 2007)	Mathematik	DEMAT 1+	-	DEMAT 1+
	Lesen	WLLP (1. Kl.)	-	WLLP (1. Kl.)
2. Schulbesuchsjahr (2007/ 2008)	Mathematik	DEMAT 2+	-	-
	Lesen	WLLP (2. Kl.)	-	-
3. Schulbesuchsjahr (2008/ 2009)	Mathematik	DEMAT 3+	DEMAT 2+	-
	Lesen	WLLP (3. Kl.)	WLLP (2. Kl.)	-
4. Schulbesuchsjahr (2009/ 2010)	Mathematik	DEMAT 4	DEMAT 3+	DEMAT 3+
	Lesen	WLLP (4. Kl.)	WLLP (3. Kl.)	WLLP (3. Kl.)
5. Schulbesuchsjahr (2010/ 2011)	Mathematik	-	DEMAT 4	DEMAT 4
	Lesen	-	WLLP (4. Kl.)	WLLP (4. Kl.)

die Vergleichsgruppen DFK und KW hingegen Daten zu drei Messzeitpunkten auf (vgl. Tabelle 1).

### Stichprobe

Die Stichprobe bildet ein Subsample aus der Mecklenburger Längsschnittstudie (MLS), die im Zeitraum von 2006 bis 2011 Daten von  $n = 1648$  Kindern in den Regionen Rostock und Rügen untersuchte. Von dieser Grundgesamtheit wurden  $n = 1505$  in Grundschulklassen (GSK),  $n = 110$  in DFK und  $n = 33$  in Langzeitklassen für Sprachförderung eingeschult. Letztere stehen nicht im Fokus der vorliegenden Untersuchung. Von den 1505 GSK wiederholten 60 Kinder zumindest ein Schuljahr und von den 110 DFK wurden  $n = 24$  nach den 3 Jahren in eine reguläre GSK innerhalb der beiden Untersuchungsregionen aufgenommen. Da im Fokus der vorliegenden Untersuchung Kinder mit einer ungünstigen Lernausgangslage stehen, wurden aus den Gruppen GSK, KW und DFK Kinder mit einem

niedrigen allgemeinen Entwicklungsstand ausgewählt. Ausgehend von einem im Schulleistungsbereich üblichen engen Durchschnittsbegriff wurde im Entwicklungsstand-Diagnostikum (EDI; Mlynek & Forster, 2005) dabei ein T-Wert  $\leq 43$  als Cutoff-Kriterium angelegt. Hinsichtlich der Intelligenzleistungen der Kinder wird als Einschlusskriterium im Kognitiven Fähigkeitstest für 1. und 2. Klassen (KFT 1-2; Kawthar & Perleth, 2005), unter Berücksichtigung des Konfidenzintervalls, der IQ-Wert  $\leq 90$  festgelegt. Dadurch ergab sich folgende Stichprobe:  $n_{\text{GSK}} = 228$ ,  $n_{\text{KW}} = 42$ ,  $n_{\text{DFK}} = 24$ . Von den nach dieser Selektion verbleibenden 42 Klassenwiederholerinnen und Klassenwiederholern entfielen 24 Kinder durch Wegzüge, schulstrukturelle Veränderungen, Wechsel an Schulen in privater Trägerschaft, Probebeschulungen in Sonderklassen, erneute Klassenwiederholungen, Verweigerung weiterer Testungen durch die Eltern oder längerfristige Erkrankungen. Von den 24 DFK-Schülerinnen und -Schülern entfielen aus gleichen Gründen 5

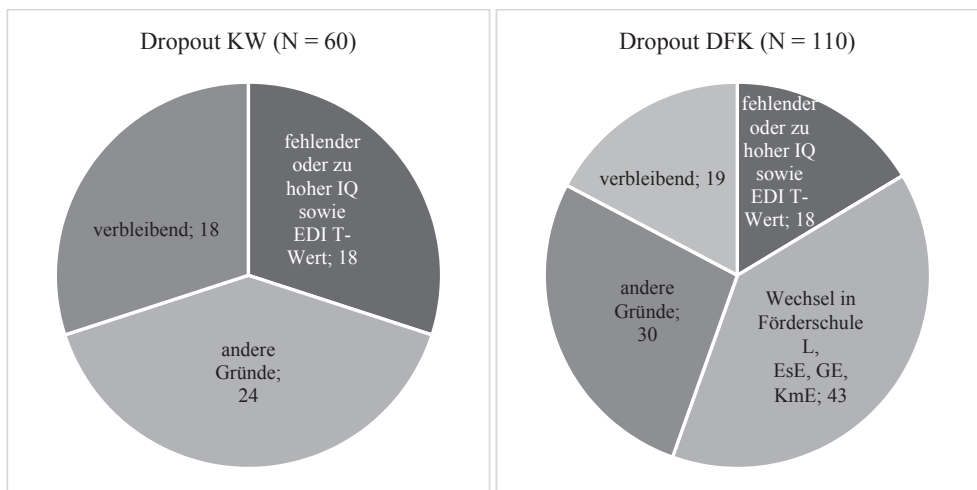


Abbildung 1: Überblick über den Daten-Dropout der Gruppe der Kinder, die eine Klasse wiederholten (KW) und der Gruppe der Kinder einer Diagnoseförderklasse (DFK)

*Erläuterungen.* Unter andere Gründe fallen: Wegzüge, schulstrukturelle Veränderungen, Wechsel an Privatschulen, Probebeschulungen in Sonderklassen, erneute Klassenwiederholungen, Verweigerung der Testungen durch Kinder oder Eltern, nicht auswertbare Testbögen oder längerfristige Erkrankungen; L – Förderschwerpunkt Lernen; EsE – Förderschwerpunkt emotional-soziale Entwicklung; GE – Förderschwerpunkt Geistige Entwicklung; KME – Förderschwerpunkt körperliche und motorische Entwicklung



Kinder. Aus der Gruppe der GSK konnten nur Kinder berücksichtigt werden, die die Klasse 4 in einer regulären Grundschulklasse beendeten ( $n = 116$ ). Neben dem Ausfall durch Testverweigerung, Krankheit, Umzüge etc. traten auch hier in erheblichem Umfang nicht beeinflussbare schulstrukturelle Umwandlungen auf, welche die Stichprobe reduzierten.

Die endgültigen Untersuchungsgruppen umfassen somit  $n_{GSK} = 30$  aus 22 Klassen,  $n_{KW} = 18$  aus 14 Klassen und  $n_{DFK} = 19$  aus 8 Klassen, welche sich in den schulisch relevanten Merkmalen (kognitive Fähigkeiten, Vorwissen und Wahrnehmungsleistungen) nicht signifikant voneinander unterscheiden (Haupteffekt MANOVA, Pillai-Spur:  $V = 0.09$ ,  $F(4, 128) = 1.45$ ,  $p = .223$ ). Zum Zeitpunkt der Einschulung waren die untersuchten Schülerinnen und Schüler im Durchschnitt etwa sieben Jahre alt ( $M_{GSK} = 6;8$  Jahre,  $SD_{GSK} = 0;3$  Jahre;  $M_{DFK} = 6;8$  Jahre,  $SD_{DFK} = 0;4$  Jahre;  $M_{KW} = 6;7$  Jahre,  $SD_{KW} = 0;3$  Jahre). Das Geschlechterverhältnis in der Gruppe der GSK betrug 63 % Jungen zu 37 % Mädchen, in den Gruppen der DFK bzw. KW war ein größerer Anteil Jungen zu verzeichnen (74 % Jungen vs. 26 % Mädchen bzw. 61 % Jungen vs. 39 % Mädchen). Tabelle 2 bietet einen Überblick über die Lernausgangslage der Vergleichsgruppen zu Beginn der ersten Klasse.

Um Einblick in den sozio-ökonomischen Status der Eltern der untersuchten Kinder zu erhalten, wurde ein Elternfragebogen zur Erfassung der sozialen Lage

(Koch, 2006) eingesetzt. Aufgrund unterschiedlicher Rücklaufquoten in den Untersuchungsgruppen können diese Daten in der vorliegenden Studie nicht berücksichtigt werden. Um jedoch einen Eindruck in die soziale Situation der Kinder zu bekommen, wurde das Einzugsgebiet der jeweiligen Schule herangezogen. Demnach wurden 50 % der GSK, 72 % der KW sowie 95 % der DFK an Schulen aus sozial-räumlich benachteiligten Gebieten unterrichtet. Es kann somit nicht ausgeschlossen werden, dass die Ergebnisse der dargestellten Analysen durch den sozioökonomischen Status der Eltern der untersuchten Kinder konfundiert sind.

### Erhebungsinstrumente

*Lernausgangslage.* Zu Beginn des ersten Schuljahres wurde das Entwicklungsstand-Diagnostikum (EDI; Mlynek & Forster, 2005) durchgeführt, um die Lernvoraussetzungen in den Bereichen Wahrnehmung (Ganzheitliche visuelle Wahrnehmung und Differenzierungsfähigkeit, Akustische Wahrnehmung und Auditives Gedächtnis, Optisches Gedächtnis, Visuomotorik und Händigkeit) und Denken (Rezeptive sprachliche Fähigkeiten, Pränumerische Fähigkeiten, Grundlegende Denkfähigkeiten) zu erfassen. Die Splithalf-Reliabilität der Testbatterie liegt nach Spearman-Brown bei  $r = .81 - .88$ . Zwischen der Skala Wahrnehmung und den Schreib- und Leseleistungen von Kindern ( $r = .93$ ) sowie zwischen dem Untertest Pränumerische Fähigkeiten und den Rechen-

Tabelle 2: Deskriptive Statistiken der Lernausgangslage

Messinstrument	Beschulungsform	gültige n	M	SD	Range
EDI (T-Werte)	GSK	30	35.82	6.03	20.43 - 43.27
	DFK	19	34.16	5.66	19.23 - 40.87
	KW	18	33.32	6.14	21.63 - 43.27
KFT 1 (IQ-Werte)	GSK	30	83.65	4.37	73 - 90
	DFK	19	80.57	5.54	70 - 90
	KW	18	82.29	5.31	72 - 90

leistungen ( $r = .96$ ) bestehen jeweils hohe Korrelationen (Mlynek & Forster, 2005).

Der Kognitive Fähigkeitstest für 1. und 2. Klassen (KFT 1-2) ist ein Verfahren zur Erfassung der Intelligenz (Kawthar & Perleth, 2005). Er gliedert sich in drei Testteile (Verbal, Nonverbal, Quantitativ), welche sich jeweils aus zwei Subtests à 24 Items (Sprachverständnis & Wortschatz, Schlussfolgerndes Denken & Matrizen, Beziehungen erkennen & Rechnerisches Denken) zusammensetzen. Eine Hauptkomponentenanalyse stützt die Faktorenstruktur (Kawthar, 2005). Für die Gesamtleistung weist das Verfahren eine hohe interne Konsistenz ( $r = .95$ ) auf (Kawthar, 2005). Die hohe interne Konsistenz legt eine eindimensionale Interpretation nahe, auch wenn die Items des KFT 1-2 modalitätsbezogen dreidimensional modelliert werden können.

*Erfassung der Leseleistung.* Jeweils am Ende jedes Schuljahres wurde zur Erfassung der Leseleistung die Würzburger Leise Leseprobe (WLLP; Küspert & Schneider, 1998) durchgeführt. Die WLLP ist mit fünf Minuten Bearbeitungszeit ein sehr zeitökonomisches Verfahren zur Erfassung der Leseleistung, die über die Dekodiergeschwindigkeit operationalisiert wird. Der Test weist im Vergleich zu Einzeltestverfahren (Diagnostischer Lesetest zur Frühdiagnose, Bremer Lesetest) akzeptable bis gute Korrelationen von  $r = .68$  bis  $r = .92$  auf (Küspert & Schneider, 1998). Als abhängige Variable wurde in der vorliegenden Untersuchung die Anzahl der richtig gelesenen Wörter herangezogen.

*Erfassung der Rechenleistung.* Zur Erfassung des Kriteriums der Rechenleistungen wurde am Ende des ersten Schuljahres der Deutsche Mathematiktest für erste Klassen (DEMAT 1+, Krajewski, Küspert, Schneider & Visé, 2002) eingesetzt. Das Verfahren gilt als Gruppentest zur ökonomischen Erfassung der Mathematikleistungen auf Basis der curricularen Inhalte aller Bundesländer. Für die Klassenstufen 2 bis 4 wurden analog dazu die Versionen DEMAT 2+ (Krajewski, Liehm & Schneider, 2004), DEMAT 3+

(Roick, Gölitz & Hasselhorn, 2004) und DEMAT 4 (Gölitz, Roick & Hasselhorn, 2006) jeweils am Ende jedes Schuljahres eingesetzt. Die internen Konsistenzen der Verfahren liegen im Bereich zwischen  $\alpha = .83$  und  $\alpha = .93$ . Für die vorliegenden Analysen wurden T-Wert-Äquivalente auf Grundlage einer Normierung an der Ausgangsstichprobe der MLS ( $N = 1648$ ) ermittelt.

### *Vorgehen in der Datenanalyse*

In Längsschnittstudien können die wiederholt erhobenen Daten als in den untersuchten Probanden geschachtelte Systeme aufgefasst werden, wodurch eine Mehrebenenstruktur gegeben ist. Die Überprüfung der Fragestellungen innerhalb der vorliegenden Untersuchung wurde auf Grundlage Hierarchisch-linearer Modellierungen (HLM, Bryk & Raudenbush, 1992) auf zwei Ebenen vorgenommen. Dazu wurden in einem ersten Schritt jeweils für die Bereiche Lesen und Mathematik unkontingente Modelle berechnet, um die Varianzanteile der schulischen Leistungen in den genannten Bereichen über die Gesamtgruppe zu überprüfen. In einem zweiten Schritt wurde die Angabe des Messzeitpunkts als Prädiktor auf Ebene 1 in das Modell aufgenommen. Dieser Prädiktor war derart kodiert, dass der letzte Messzeitpunkt als Nullpunkt diente, d. h. der Intercept kann als Maß des Leistungsniveaus zum Ende der Klasse vier (nach vier für GSK bzw. fünf Schuljahren für DFK und KW) interpretiert werden. Zudem wurde der Einfluss des jeweiligen Beschulungskonzeptes (GSK, DFK bzw. KW) auf die Regressionsparameter (Niveau und Anstieg) überprüft (intercept-and-slope-as-outcome-Modelle). Dazu wurde die Zugehörigkeit zur Gruppe von Kindern in DFK bzw. von Kindern, die eine Klasse wiederholten, durch die Dummy-Variablen DFK bzw. KW auf der zweiten Ebene berücksichtigt. Durch dieses Vorgehen wurde die Gruppenzugehörigkeit zu GSK zur Referenzkategorie. Um Vergleiche zwischen den Gruppen DFK und KW zu ermögli-



chen, wurden analoge Modelle gerechnet, jeweils mit der Gruppe KW als Referenzgruppe.

Entgegen der geläufigen Forschungspraxis, Matching-Verfahren (bspw. über propensity scores) zur Parallelisierung der Untersuchungsgruppen zu nutzen, wurde in der vorliegenden Studie ein anderer Zugang zur Kontrolle von Störvariablen gewählt, da es erfahrungsgemäß schwierig ist, statistische Zwillinge für die hochheterogenen Schülerprofile in DFK zu ermitteln (Voß et al., 2015). So wurden als Kontrollvariablen jeweils das Geschlecht, die kognitiven Fähigkeiten gemäß des KFT 1-2 sowie der Entwicklungsstand entsprechend des EDI in die Modelle auf Ebene 2 aufgenommen. Effektstärken bei signifikanten Gruppenunterschieden werden nach Tymms (2004) berechnet. In einem dritten Schritt wurde abschließend der jeweilige Zugewinn an erklärter Varianz des jeweils bedingten Modells im Vergleich zum zugehörigen unkontrollierten Modell zur Einschätzung des Model-Fit analysiert.

## Ergebnisse

### Deskriptive Statistiken

*Lesen.* Tabelle 3 zeigt die deskriptiven Statistiken für die Leseleistungen der drei Untersuchungsgruppen. Alle drei Untersuchungsgruppen können die Anzahl der richtig gelesenen Wörter im Mittel über die Schulbesuchszeit an der Grundschule konstant steigern. Am Ende der Grundschulzeit lesen die Gruppen im Mittel etwa zwischen 99 (entspricht einem Prozentrang zwischen 20 und 29) und 108 Wörtern (entspricht einem Prozentrang zwischen 33 und 48). Über alle Gruppen und Schuljahre war ein großer Range in den Leseleistungen zu beobachten.

*Mathematik.* Tabelle 4 zeigt die deskriptiven Statistiken für die Mathematikleistungen der drei Untersuchungsgruppen. Sowohl die Gruppe GSK ( $M = 51.31$ ) als auch die Gruppe KW ( $M = 47.73$ ) erzielen zum Ende der ersten Klasse im Mittel Mathematikleistungen im Normbereich. Über die weiteren Untersuchungsjahre sind die Mittelwerte aller drei Gruppen, bezogen auf die Gesamtstichprobe der MLS und auf das jeweilige curriculare Niveau, als unterdurchschnittlich einzustufen (T-Werte zwi-

Tabelle 3: Deskriptive Statistiken der Leseleistungen (Anzahl gelesener Wörter, WLLP)

Messzeitpunkt	Beschulungsform	gültige $n$	$M$ (Rohwert)	$SD$	Range
Ende Klasse 1	GSK (1. Schulbesuchsjahr)	30	49.03	23.33	19 - 116
	DFK	-	-	-	-
	KW (1. Schulbesuchsjahr)	18	38.33	25.12	5 - 86
Ende Klasse 2	GSK (2. Schulbesuchsjahr)	30	56.97	23.67	6 - 136
	DFK (3. Schulbesuchsjahr)	19	69.26	24.66	35 - 115
	KW	-	-	-	-
Ende Klasse 3	GSK (3. Schulbesuchsjahr)	30	78.60	20.63	37 - 133
	DFK (4. Schulbesuchsjahr)	19	87.63	17.97	54 - 121
	KW (4. Schulbesuchsjahr)	18	82.72	18.74	40 - 118
Ende Klasse 4	GSK (4. Schulbesuchsjahr)	30	102.37	18.91	56 - 138
	DFK (5. Schulbesuchsjahr)	19	107.79	19.97	78 - 139
	KW (5. Schulbesuchsjahr)	18	98.72	21.87	56 - 124

Tabelle 4: Deskriptive Statistiken der Mathematikleistungen (T-Werte, DEMAT 1+, DEMAT 2+, DEMAT 3+, DEMAT 4)

Messzeitpunkt	Beschulungsform	gültige n	M (T-Wert)	SD	Range
Ende Klasse 1	GSK (1. Schulbesuchsjahr)	30	51.31	8.63	27.89 - 61.96
	DFK	-	-	-	-
	KW (1. Schulbesuchsjahr)	18	47.73	10.28	26.72 - 60.79
Ende Klasse 2	GSK (2. Schulbesuchsjahr)	30	40.27	7.70	30.56 - 58.21
	DFK (3. Schulbesuchsjahr)	19	41.53	7.76	27.37 - 52.36
	KW	-	-	-	-
Ende Klasse 3	GSK (3. Schulbesuchsjahr)	30	41.46	7.73	29.99 - 59.65
	DFK (4. Schulbesuchsjahr)	19	36.42	7.63	21.00 - 47.97
	KW (4. Schulbesuchsjahr)	18	40.78	7.65	28.19 - 56.95
Ende Klasse 4	GSK (4. Schulbesuchsjahr)	30	39.78	7.45	21.96 - 56.29
	DFK (5. Schulbesuchsjahr)	19	40.93	9.63	27.24 - 60.25
	KW (5. Schulbesuchsjahr)	18	38.75	6.82	24.60 - 51.01

schen  $T = 39$  und  $T = 41$ ). Ähnlich der Leseleistungen ist ein großer Range in den Mathematikleistungen über alle Gruppen und Schuljahre zu beobachten.

### Unkonditionierte Modelle

In Tabelle 5 sind die Ergebnisse der unkonditionierten Modelle für die Bereiche Lesen und Mathematik zusammengefasst. Für den Bereich Lesen zeigte sich über die Gruppen hinweg ein durchschnittliches Leistungsniveau von  $\gamma_{00} = 76.23$  ( $p < .001$ ) korrekt gelöster Items zum Ende der vierten Klasse. Eine Analyse der geschätzten zufälligen Effekte  $u_{0i}$  ergab mit  $\chi^2 = 74.15$  ( $df = 66$ ,  $p = .230$ ) keine signifikanten schülerspezifischen Unterschiede im Leseleistungsniveau zum Ende der Klassenstufe 4 (nach vier bzw. fünf Schulbesuchsjahren). Eine Analyse der Varianzkomponenten deutet darauf hin, dass ein Anteil von etwa 3 % der erklärbaren Varianz zwischen den Kindern bedingt war ( $ICC = 0.03$ ).

Für den Bereich Mathematik zeigte sich zum Ende der Klassenstufe 4 (nach vier bzw. fünf Schulbesuchsjahren) über alle Gruppen hinweg ein gemittelttes Leistungsniveau von etwa 42 T-Wertpunkten

( $\gamma_{00} = 42.07$ ,  $p < .001$ ). Bei Betrachtung der geschätzten zufälligen Effekte zeigte sich, dass es mit  $\chi^2 = 122.24$  ( $df = 66$ ,  $p < .001$ ) signifikante schülerspezifische Unterschiede zwischen den Regressionskonstanten gab. Die Analyse der Varianzkomponenten zeigt, dass ein Anteil von 20 % der erklärbaren Varianz zwischen den Kindern bedingt war ( $ICC = 0.20$ ).

### Bedingte intercept-and-slope-as-outcome-Modelle

Die Ergebnisse der intercept-and-slope-as-outcome-Modelle (unter Berücksichtigung der Kontrollvariablen schulisches Vorwissen, kognitive Fähigkeiten sowie Geschlecht) für die Bereiche Lesen und Mathematik sind in Tabelle 6 dargestellt.

Es ergab sich im Lesen ein durchschnittliches Leistungsniveau von etwa 83 Rohwertpunkten ( $\gamma_{00} = 83.33$ ) für die Gruppe GSK zum Ende der Klasse 4 (nach vier Schulbesuchsjahren). Tendenziell erzielten Kinder der Gruppen DFK und KW zum Ende der Klasse 4, also nach fünf Schulbesuchsjahren, höhere Rohwerte ( $\gamma_{01} = 10.14$  bzw.  $\gamma_{02} = 2.69$ ). Diese Leistungsunterschiede fielen im Zusammenhang mit den

Tabelle 5: Ergebnisse der unkonditionierten Hierarchisch-linearen Modelle im Bereich Lesen und Mathematik

Feste Effekte	Lesen			Mathematik		
	Koeff (SE)	t(df = 66)	d	Koeff (SE)	t(df = 66)	d
Niveau, $\gamma_{00}$	76.23 (2.07)***	36.75	2.53 ± 0.07	42.07 (0.72)***	58.09	5.17 ± 0.09
Zufalls- effekte	Var	$\chi^2$ (df = 66)		Var	$\chi^2$ (df = 66)	
Niveau $u_{0i}$	30.62	74.15		16.26***	122.24	
Level-1- Fehler $e_{ij}$	907.21			66.20		
	Devianz: 2231.76, df = 2			Devianz: 1662.83, df = 2		

vergleichsweise hohen Standardfehlern jedoch nicht statistisch signifikant aus ( $p > .05$ ). Der mittlere Anstieg  $\gamma_{10} = 23.15$  deutete auf eine Leistungsverbesserung der Gruppe GSK im Lesen von ca. 23 Rohwertpunkten pro Schuljahr hin, aufgrund des hohen Standardfehlers fällt dieser Wert jedoch nicht signifikant aus ( $p > .05$ ). Tendenziell wiesen die DFK und KW höhere Leistungszuwächse über die Schuljahre auf ( $\gamma_{11} = 1.22$  bzw.  $\gamma_{12} = 2.75$ ), welche sich jedoch nicht signifikant von denen der GSK unterschieden ( $p > .05$ ). Die Zufallskomponenten zum Leistungsniveau  $u_{0i}$  bzw. zum Leistungsanstieg  $u_{1i}$  deuteten auf schüler-spezifische Unterschiede hin, die nicht durch das Modell erklärt werden konnten ( $\chi^2 = 178.74$ ,  $df = 61$ ,  $p < .001$  bzw.  $\chi^2 = 135.59$ ,  $df = 61$ ,  $p < .001$ ).

Um auf Unterschiede zwischen den Gruppen DFK und KW zu prüfen, wurde in einem analogen Modell (hier nicht dargestellt) die Zugehörigkeit zur Gruppe der KW als Referenzgruppe analysiert. Demnach unterscheiden sich DFK und KW nicht signifikant im Lesen hinsichtlich des Leistungsstandes nach fünf Schulbesuchsjahren ( $\gamma_{02} = 7.45$ ,  $SE = 6.49$ ,  $t(df = 61) = 1.15$ ,  $p = 0.256$ ) sowie hinsichtlich des Leistungsanstiegs über die Schuljahre ( $\gamma_{12} = -1.53$ ,  $SE = 2.78$ ,  $t(df = 61) = -0.55$ ,  $p = .584$ ).

Für die Gruppe GSK lag das mittlere Leistungsniveau im Lernbereich Mathematik unter Kontrolle des Geschlechts der Kinder bei  $\gamma_{00} = 13.68$  T-Wertpunkten, bei einem jährlichen negativen Anstieg um  $\gamma_{10} = -7.45$  T-Wertpunkten, was einen jährlichen Leistungsrückgang dieser Kinder um ca. 7 T-Wertpunkte anzeigt. Es zeigten sich keine signifikanten Unterschiede der Gruppe KW im Vergleich zu der Gruppe GSK, weder im Leistungsniveau zum Ende der vierten Klasse ( $\gamma_{02} = 1.32$ ,  $p > .05$ ), noch im Leistungsanstieg über die Schuljahre ( $\gamma_{12} = 0.53$ ,  $p > .05$ ). Für die Gruppe DFK ergaben sich keine signifikanten Unterschiede hinsichtlich des Leistungsniveaus zum Ende der vierten Klasse (nach fünf Schulbesuchsjahren;  $\gamma_{01} = 2.65$ ,  $p > .05$ ), wohl aber im jährlichen Leistungsanstieg ( $\gamma_{11} = 3.41$ ,  $p < .01$ ,  $d = 0.52 \pm 0.17$ ). Demnach fiel die mathematische Leistungsentwicklung dieser Kinder jährlich um 3 T-Wertpunkte höher aus. Die Zufallskomponenten des Leistungsniveaus  $u_{0i}$  sowie des Leistungsanstiegs  $u_{1i}$  deuteten auf weiterhin persistierende schülerspezifische Unterschiede hin, die nicht durch die Gruppenzugehörigkeit erklärt werden konnten ( $\chi^2 = 112.68$ ,  $df = 61$ ,  $p < .001$  bzw.  $\chi^2 = 85.00$ ,  $df = 61$ ,  $p < .05$ ).

Es zeigen sich keine signifikanten Unterschiede zwischen DFK und KW hinsichtlich des Leistungsstandes in Mathematik nach

fünf Schulbesuchsjahren ( $\gamma_{02} = 1.32$ ,  $SE = 2.23$ ,  $t(df = 61) = 0.59$ ,  $p = 0.555$ ), jedoch hinsichtlich des Leistungsanstiegs in diesem Bereich über die Schuljahre ( $\gamma_{12} = 2.88$ ,  $SE = 1.21$ ,  $t(df = 61) = 2.38$ ,  $p = .020$ ,  $d = 0.44 \pm 0.18$ ). Gemäß den Daten fällt der mathematische Entwicklungsverlauf für Kinder in DFK etwa drei T-Werte je Schuljahr höher aus als für Kinder der Gruppe KW.

Die hier berichteten Ergebnisse bleiben auch stabil, wenn Ausreißer, also Kinder, deren Mathematik- oder Leseleistungen zwei oder mehr Standardabweichungen unter dem Gruppenmittelwert der entsprechenden Messzeitpunkte lagen, aus den Analysen eliminiert werden.

Zusammenfassend ist für den Bereich Lesen festzuhalten, dass die untersuchten

Kindergruppen, entgegen der aufgestellten Hypothese, gleiche Leistungsniveaus am Ende der vierten Klasse (nach vier bzw. fünf Schulbesuchsjahren) sowie gleiche Leistungsentwicklungen über die Schuljahre zeigen (DFK = KW = GSK). Hinsichtlich der mathematischen Leistungen fallen die Daten ebenso nicht hypothesenkonform aus. Das erreichte Leistungsniveau zum Ende der vierten Klasse (nach vier bzw. fünf Schulbesuchsjahren) gleicht sich in allen drei Gruppen. Jedoch weisen Kinder in DFK über die Zeit eine signifikant höhere Leistungsentwicklung im Vergleich zu GSK ( $d = 0.52 \pm 0.17$ ) und KW ( $d = 0.44 \pm 0.18$ ) auf, bei Kindern der Gruppen GSK und KW zeigen sich hier keine statistisch abgesicherten Differenzen (DFK > GSK = KW).

Tabelle 6: Ergebnisse der bedingten Hierarchisch-linearen Modelle für die Bereiche Lesen und Mathematik

Feste Effekte	Lesen			Mathematik		
	Koeff (SE)	t(df = 61)	Effekt	Koeff (SE)	t(df = 61)	Effekt
Modell für den intercept $\beta_{0i}$						
Niveau, $\gamma_{00}$	83.33 (34.29)*	2.43	$5.79 \pm 2.38$	13.68 (17.30)	0.79	
DFK, $\gamma_{01}$	10.14 (6.28)	1.62		2.65 (2.24)	1.18	
KW, $\gamma_{02}$	2.69 (6.00)	0.45		1.32 (2.09)	0.63	
Vorwissen, $\gamma_{03}$	0.53 (0.38)	1.40		0.43 (0.16)*	2.67	$0.07 \pm 0.02$
Intelligenz, $\gamma_{04}$	0.01 (0.46)	0.03		0.14 (0.22)	0.61	
Geschlecht, $\gamma_{05}$	-6.96 (5.60)	-1.24		-3.69 (1.91)	-1.93	
Modell für den slope $\beta_{1i}$						
Anstieg, $\gamma_{10}$	23.15 (20.47)	1.13		-7.45 (8.66)	-0.86	
DFK, $\gamma_{11}$	1.22 (2.89)	0.42		3.41 (1.14)**	3.00	$0.52 \pm 0.17$
KW, $\gamma_{12}$	2.75 (3.04)	0.91		0.53 (1.16)	0.46	
Vorwissen, $\gamma_{13}$	0.31 (0.17)	1.76		0.10 (0.07)	1.39	
Intelligenz, $\gamma_{14}$	-0.18 (0.28)	-0.65		0.02 (0.11)	0.14	
Geschlecht, $\gamma_{15}$	-1.59 (2.76)	-0.58		-1.41 (0.95)	-1.49	
Zufallseffekte	Var	$\chi^2(df = 61)$		Var	$\chi^2(df = 61)$	
Niveau $u_{0i}$	312.63***	178.74		29.59***	112.68	
Anstieg $u_{1i}$	60.89***	135.59		4.98*	85.00	
Level-1-Fehler $e_{ij}$	207.13			43.68		
Devianz: 1993.72, df = 4			Devianz: 1592.71, df = 4			

Anmerkungen. \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

Betrachtet man den letzten Befund jedoch im Zusammenhang mit einer Analyse der deskriptiven Statistiken, lässt sich ein Artefakt in den Daten nicht ausschließen. So fällt auf, dass die Mathematikleistungen der Gruppen GSK und KW im Mittel zum ersten Messzeitpunkt deutlich höher ausfallen als in den Folgejahren. Das führt scheinbar zu den deutlich negativen Anstiegen der T-Werte (Leistungsrückgang) der GSK und KW. Für die DFK fehlen konzeptionsbedingt Angaben zu diesem Messzeitpunkt, was zu einer Überschätzung des slopes für diese Gruppe geführt haben könnte. Eine Reanalyse der Daten unter Ausschluss der Mathematik-Daten zum Ende der Klasse 1 der GSK und KW erhärtet diese Annahme. Der slope der GSK fällt in diesem Fall positiv, jedoch aufgrund des hohen Standardfehlers nicht signifikant verschieden von Null aus ( $\gamma_{10} = 12.00$ ,  $SE = 7.55$ ,  $t(df = 61) = 1.59$ ,  $p = .117$ ). Die Anstiege der DFK ( $\gamma_{11} = -0.23$ ,  $SE = 1.06$ ,  $t(df = 61) = -0.22$ ) und der KW ( $\gamma_{12} = -2.07$ ,  $SE = 1.86$ ,  $t(df = 61) = -1.12$ ) fallen nicht signifikant verschieden zu denen der GSK aus ( $p = .831$  bzw.  $p = .269$ ). Im Vergleich DFK zu KW (KW als Referenzgruppe) zeigen sich ebenfalls keine signifikanten Unterschiede hinsichtlich des Leistungsniveaus zum Ende der vierten Klasse ( $\gamma_{02} = 0.91$ ,  $SE = 2.25$ ,  $t(df = 61) = 0.41$ ,  $p = .686$ ) bzw. zum Leistungsanstieg über die Schuljahre ( $\gamma_{10} = 1.85$ ,  $SE = 1.88$ ,  $t(df = 61) = 0.98$ ,  $p = .329$ ). Demnach ist im Bereich der Mathematik eher von einem Gleichstand der drei untersuchten Gruppen hinsichtlich des Leistungsniveaus zum Ende der vierten Klasse und der Leistungsentwicklung über die Schuljahre auszugehen (DFK = GSK = KW).

### **Varianzvergleich und Beurteilung des Model-Fit**

Für die bedingten Modelle zum Lesen sowie zur Mathematik ergaben sich gegenüber den unkontrollierten Modellen geringere Model-Fit-Statistiken (Devianz Lesen:

2231.76 vs. 1993.72; Devianz Mathematik: 1662.83 vs. 1592.71). Aus der Hinzunahme der Prädiktoren DFK und KW und der Kontrollvariablen auf der Schülerebene resultierten jeweils signifikant bessere Model-Fit-Statistiken (Lesen:  $\chi^2 = 238.04$ ,  $df = 2$ ,  $p < .001$ ; Mathematik:  $\chi^2 = 70.11$ ,  $df = 2$ ,  $p < .001$ ). Dies zeigt sich auch durch die Analyse der durch die Modelle jeweils aufgeklärten Varianzanteile. Unter Hinzunahme der Gruppenzugehörigkeit zu den verschiedenen Beschulungskonzepten sowie unter Kontrolle des Vorwissens und der kognitiven Fähigkeiten zu Beginn der ersten Klasse als auch des Geschlechts konnten für den Bereich Lesen 77 % der Varianz aufgeklärt werden, für den Bereich Mathematik etwa 34 %.

### **Diskussion**

Die vorliegende Studie untersuchte die langfristigen Effekte lernzeitverlängernder Maßnahmen bei Kindern mit ungünstigen Lernausgangslagen im Vergleich zu einer regulären Beschulung in Grundschulklassen. Dazu wurden die Entwicklung und das erreichte Niveau der Lese- und Mathematikleistungen von Kindergruppen aus diesen Lernsettings über die gesamte Grundschulzeit analysiert. Die Gruppe DFK erzielte am Ende der vierten Klasse im Lesen das höchste Leistungsniveau, gefolgt von der Gruppe KW und abschließend der Gruppe GSK. Die Unterschiede fielen jedoch nicht signifikant aus ( $p > .05$ ). Ebenso zeigten alle drei Gruppen, entgegen der aufgestellten Hypothese, eine ähnliche positive Leistungsentwicklung im Lesen. Im Bereich der Mathematik wiesen die Kinder in GSK das geringste Leistungsniveau auf, gefolgt von den Gruppen KW und DFK. Zwar fällt der Vorteil der Gruppe der DFK im Vergleich zu den Kindern der Gruppen GSK ( $d = 0.52 \pm 0.17$ ) und KW ( $d = 0.44 \pm 0.18$ ) signifikant aus, jedoch ist die Verlässlichkeit dieser Daten zweifelhaft (Überschätzung der Gruppe DFK durch de-

signbedingt fehlende Daten am Ende der Klasse 1). Vermutlich ist hier eher von einem Gleichstand der Gruppen auszugehen. Rekurrend auf diese Vermutung kann insgesamt kein belastbarer positiver Effekt lernzeitverlängernder Maßnahmen auf die schulischen Leistungen im Lesen und in Mathematik nachgewiesen werden.

Aufgrund verschiedener methodischer Limitationen müssen die vorgestellten Befunde zur Wirksamkeit lernzeitverlängernder Maßnahmen vorsichtig interpretiert werden und bedürfen weiterer empirischer Prüfungen. So ist die untersuchte Stichprobe relativ klein, da über den gesamten Untersuchungszeitraum hinweg in merkbarem Ausmaß Datenausfälle hauptsächlich aufgrund von schulstrukturellen Umwandlungen hingenommen werden mussten. In diesem Zusammenhang ist ebenfalls darauf hinzuweisen, dass ein Teil des Dropouts der Kinder eine inhärente Logik aufweist. So gingen die Daten derjenigen Kinder, die im Laufe der Grundschulzeit an ein Sonderpädagogisches Förderzentrum umgeschult wurden, nicht mit in die Analysen ein. Da es sich hierbei um die Schülerinnen und Schüler mit den am schwächsten ausgeprägten Leistungsprofilen handelt, sind systematische Verzerrungen der Ergebnisse demnach nicht auszuschließen.

Eine Verzerrung der Leistungsvariablen mit dem Alter ist eine weitere mögliche Einschränkung der Ergebnisse, da durch die Lernzeitstreckung ein Jahr Altersunterschied zugunsten der DFK Kinder und der Klassenwiederholerinnen und Klassenwiederholer besteht. Erfahrungsgemäß führen bereits natürliche Reifungsprozesse zu einer Lernsteigerung bei einer Effektstärke von etwa  $d = 0.15$  (Hattie, 2013).

Weiterhin sollte die Operationalisierung der Leseleistungen um die Erfassung der Leseverständnisleistungen ergänzt werden. Neben weiteren akademischen Faktoren (z. B. Rechtschreibleistungen) sollten relevante motivationale und affektive Variablen sowie die Entwicklung sozial-emotionaler Kompetenzen multimodal einbezogen wer-

den. Methodenkritisch muss weiterhin angemerkt werden, dass die Leistungen im Bereich Mathematik nicht wie im Lesen mit einem, sondern mit den verschiedenen Verfahren der DEMAT Serie erhoben wurden. In der vorliegenden Studie wurden zur Analyse der Leistungsentwicklung die Ergebnisse der untersuchten Kinder auf T-Wert-Basis in die Modelle aufgenommen. Dabei wurde eine Messinvarianz dieser Tests angenommen, welche aber aufgrund der gegebenen Datenlage in der Form nicht belegbar ist. Empfehlenswert für Studien im Längsschnitt wäre der Einsatz von speziell für die Veränderungsmessung entwickelten Verfahren.

Zu beachten ist zudem die unterschiedliche soziale Situation der Kinder in den einzelnen Untersuchungsgruppen. Eine Konfundierung der dargelegten Ergebnisse hierhingehend kann nicht ausgeschlossen werden.

Die Datenanalysen zeigten, dass über die Gesamtstichprobe hinweg am Ende der vierten Klasse (nach vier bzw. fünf Schulbesuchsjahren) interindividuelle Unterschiede in den Lese- und Mathematikleistungen vorlagen, die sich nur zu einem Teil durch das jeweilige Beschulungskonzept bzw. die verwendeten Kontrollvariablen erklären ließen. In der vorliegenden Untersuchung könnte dies in Zusammenhang mit der Operationalisierung des Prädiktors „Beschulungskonzept“ stehen. Hierfür erfolgte lediglich eine nominale Zuordnung zu den Gruppen GSK, DFK bzw. KW. Für die DFK ist bekannt, dass konzeptbedingt kleinere Klassengrößen ( $n < 15$ ), eine curriculare Streckung sowie zusätzliche sonderpädagogische Förderung realisiert wurden. Informationen zu weiteren Bedingungsfaktoren wie Struktur und Qualität des Unterrichts (also Bedingungsfaktoren der Klassenebene) lagen für keine der drei Gruppen vor. Auf Grundlage aktuell vorliegender empirischer Forschungsbefunde ist jedoch davon auszugehen, dass gerade Unterrichtsvariablen, z. B. die Klassenführung (Effektstärke nach Marzano (2000)  $d = 0.52$  bis  $0.62$ ), die Lehrer-Schüler-Beziehung (Effektstärke nach



Cornelius-White (2007)  $d = 0.72$ ) oder der Einsatz effektiver Lehrstrategien (Effektstärke nach Hattie (2013)  $d = 0.60$ ), zur Aufklärung weiterer Varianz beitragen können. Es scheint also unabdingbar, die makroanalytische Perspektive durch mikroanalytische Daten zu ergänzen, da die nominale Unterrichtszeit selbst wenig über die tatsächliche Nutzung aussagt.

Trotz der methodenkritischen Einschränkungen geben die vorliegenden Daten erste Hinweise darauf, dass sich zusätzliche Lernzeit von einem Schuljahr, realisiert durch DFK bzw. durch Klassenwiederholung, im Vergleich zu einer regulären Beschulung in Grundschulklassen bei Kindern mit ungünstigen Lernausgangslagen tendenziell nicht in einem höheren Leistungsniveau in den Bereichen Lesen und Mathematik zum Ende der Grundschulzeit widerspiegelt. Dieser Befund reiht sich in die Befundlage unterschiedlicher Studien (für Klassenwiederholungen zusammenfassend Hattie, 2013) ein, die herausarbeiten, dass Lernzeitverlängerungen sich langfristig nicht deutlich positiv auswirken. Auch in der Beschulungsvariante DFK scheint sich die längere Verweildauer hinsichtlich der Lese- und Mathematikleistungen langfristig nicht eindeutig positiv auszuwirken.

Bedenklich ist, dass alle drei Untersuchungsgruppen zum Ende der Grundschulbesuchszeit im Vergleich zu ihren Klassenkameraden lediglich unterdurchschnittliche Mathematikleistungen erzielten (mittlere T-Werte zwischen 39 und 41; Range GSK T = 21.96 – 56.29; Range DFK T = 27.24 – 60.25; Range KW T = 24.60 – 51.01). Im Bereich des Lesens fallen diese Werte zwar günstiger aus (PR-Werte zwischen 20 und 48), dennoch werfen die hier dargelegten Befunde grundsätzliche Fragen nach der Effektivität des Unterrichts für schulisch gefährdete Kinder in allen betrachteten Beschulungskonzepten auf. Dieses Ergebnis sollte Anlass sein, auf der Suche nach Lösungen, Abschied zu nehmen vom Erfolgsglauben an schulstrukturelle Maßnahmen. Viel wesentlicher

scheinen hier der konkrete Unterricht und die Umsetzung gezielter Fördermaßnahmen für Kinder ungünstigen Ausgangsbedingungen zu sein.

Zentral erscheint in diesem Kontext die *Prävention* von Lernschwierigkeiten. Ein vieldiskutiertes Präventionskonzept ist das sog. Response to Intervention-Konzept (Rti; z. B. Blumenthal, Kuhlmann & Hartke, 2014; Fuchs & Fuchs, 2006; Huber & Grosche, 2012). Es zielt darauf ab, den Lernerfolg der Kinder zu sichern, indem Lernlücken frühzeitig erkannt und mit Hilfe besonders bewährter, d. h. evidenzbasierter Fördermaßnahmen geschlossen werden.

Vor dem Hintergrund der Forschungsergebnisse zur prädiktiven Aussagekraft von Vorläuferfähigkeiten des Schriftspracherwerbs sowie früher mathematischer Kompetenzen (z. B. Krajewski & Schneider, 2006; Mannhaupt, 2010) liegt es nahe, diese bereits vor Schulbeginn systematisch und gezielt zu fördern. Für den Lernerfolg im Grundschulalter scheint es demzufolge zielführend, die Potenziale vorschulischer Bildung adäquat zu nutzen und weiterzuentwickeln.

## Literaturverzeichnis

- Bless, G. (2006). Sitzenbleiben - eine vertretbare Maßnahme? Ergebnisse einer Schweizer Studie. *Schul-Management*, 37 (3), 16-18.
- Bless, G., Schüpbach, M. & Bonvin, P. (2005). Klassenwiederholung. Empirische Untersuchung zum Repetitionsentscheid und zu den Auswirkungen auf die Lernentwicklung sowie auf soziale und emotionale Faktoren. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 4, 297-311.
- Bloom, B. S. (1985). Learning for mastery. *Perspectives on instructional time*, 73-93.
- Blossfeld, H.-P., Maurice, J. v. & Schneider, T. (2011). *Nationales Bildungspanel NEPS. Grundidee, Konzeption und Design des Nationalen Bildungspanels für Deutsch-*

- land. Zugriff am 03.02.2016. Verfügbar unter: [https://www.uni-bamberg.de/file-admin/inbil/Publikationen/Working-Papers/WP\\_1.pdf](https://www.uni-bamberg.de/file-admin/inbil/Publikationen/Working-Papers/WP_1.pdf)
- Blumenthal, Y., Hartke, B. & Koch, K. (2010). Mecklenburger Längsschnittstudie: Wie effektiv sind Diagnoseförderklassen? *Zeitschrift für Heilpädagogik*, 61, 331-341.
- Blumenthal, Y., Kuhlmann, K. & Hartke, B. (2014). Diagnostik und Prävention von Lernschwierigkeiten im Aptitude Treatment Interaction- (ATI-) und Response to Intervention-(RTI-)Ansatz. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik (Tests und Trends N.F. Band 12, S. 61-81)*. Göttingen: Hogrefe.
- Breitenbach, E. & Lehner, A. (1999). Einmal lernbehindert, immer lernbehindert? *Behindertenpädagogik in Bayern*, 2, 101-107.
- Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods (Advanced quantitative techniques in the social sciences, Bd. 1)*. Newbury Park: Sage.
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 2007, Vol. 77, No. 1, 113-143.
- Demski, D. & Liegmann, A. B. (2014). Klassenwiederholungen im Kontext von Schul- und Berufsbiographien. In A. B. Liegmann, I. Mammes & K. Racherbäumer (Hrsg.), *Facetten von Übergängen im Bildungssystem. Nationale und internationale Ergebnisse empirischer Forschung* (S. 173-189). Münster: Waxmann.
- Ehmke, T., Drechsel, B. & Carstensen, C. H. (2008). Klassenwiederholen in PISA-I-Plus: Was lernen Sitzenbleiber in Mathematik dazu? *Zeitschrift für Erziehungswissenschaft*, 11 (3), 368-387.
- Ellinger, S. & Koch, K. (2007). Flexible Schulingangsstufe für Kinder mit sonderpädagogischem Förderbedarf. Eine kritische Bilanz zur Effektivität von Diagnose- und Förderklassen. *Zeitschrift für Heilpädagogik*, 3, 82-90.
- Fuchs, D. & Fuchs, L.S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41 (1), 93-99.
- Göltz, D., Roick, T. & Hasselhorn, M. (2006). *DEMAT 4+. Deutscher Mathematiktest für vierte Klassen*. Göttingen: Beltz.
- Hattie, J. A. (2013). *Lernen sichtbar machen. Überarbeitete deutschsprachige Ausgabe von Visible Learning*. (1. Nachdruck) Baltmannsweiler: Schneider Verlag Hohengehren.
- Helmke, A. (2004). *Unterrichtsqualität: Erfassen, Bewerten, Verbessern*. Seelze: Kallmeyersche Verlagsbuchhandlung.
- Hesse, H. G. (1994). Lehr-Lern-Zeit und Lernerfolg aus psychologischer Sicht. In W. Mitter & B. v. Kopp (Hrsg.), *Die Zeitdimension in der Schule als Gegenstand des Bildungsvergleichs* (S. 143-161). Köln, Weimar: Böhlau.
- Hosenfeld, I., Helmke, A., Ridder, A. & Schrader, F.-W. (2002). Die Rolle des Kontextes. In A. Helmke & R. S. Jäger (Hrsg.), *Die Studie MARKUS - Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext* (S. 155-256). Landau: Verlag Empirische Pädagogik.
- Huber, C. & Grosche, M. (2012). Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. *Zeitschrift für Heilpädagogik*, 8, 312-322.
- Kawthar, K. A. (2005). *Reliabilitäts- und Validitätsuntersuchungen zum neuen kognitiven Fähigkeitstest für die Primarstufe (KFT 1-2) unter besonderer Berücksichtigung von Kindern mit Hochbegabungen und Lernbehinderungen*. Unveröffentlichte Dissertation, Universität Rostock.
- Kawthar, K. A. & Perleth, C. (2005). *Kognitiver Fähigkeitstest für erste und zweite Klassen (KFT 1-2) - Versuchsversion*. Rostock: Universität Rostock.
- Klemm, K. (2009). *Klassenwiederholungen - teuer und unwirksam. Eine Studie zu den Ausgaben für Klassenwiederholungen in*

- Deutschland. Gütersloh: Bertelsmann-Stiftung.
- Koch, K. (2006). *Soziale Lage und Sozialisationsbedingungen von Schülern mit Lernbeeinträchtigungen vor dem Hintergrund klassischer und moderner Sozialstrukturmodelle*. Unveröffentlichte Habilitation, Universität Würzburg.
- Koch, K., Hartke, B. & Blumenthal, Y. (2009). *Merkmale von Kindern mit besonderem Förderbedarf im ersten Schuljahr. Erste Ergebnisse der Mecklenburger Längsschnittstudie*. Hamburg: Kovac.
- Krajewski, K., Küspert, P. & Schneider, W. & Visé, M. (2002). *DEMAT 1+*. Deutscher Mathematiktest für erste Klassen. Göttingen: Hogrefe.
- Krajewski, K., Liehm, S. & Schneider, W. (2004). *DEMAT 2+*. Deutscher Mathematiktest für zweite Klassen. Göttingen: Hogrefe.
- Krajewski, K. & Schneider, W. (2006). Mathematische Vorläuferfertigkeiten im Vorschulalter und ihre Vorhersagekraft für die Mathematikleistungen bis zum Ende der Grundschulzeit. *Psychologie in Erziehung und Unterricht*, 53, 246-262.
- Krohne, J. & Tillmann, K.-J. (2006). „Sitzenbleiben“ – eine tradierte Praxis auf dem Prüfstand. *Schulverwaltung Spezial*, 4, 6-9.
- Küspert, P. & Schneider, W. (1998). *Würzburger Leise Leseprobe (WLLP)*. Göttingen: Hogrefe.
- Lamote, C., Pinxten, M., Van Den Noortgate, W. & Van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic selfconcept. *Educational Studies*, 40 (5), 496-514.
- Mannhaupt, G. (2010). Phonologische Bewusstheit – eine notwendige Voraussetzung des Schriftspracherwerbs. In B. Hartke, K. Koch & K. Diehl (Hrsg.), *Förderung in der schulischen Eingangsphase* (S. 186-209). Stuttgart: Kohlhammer,
- Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us*. Aurora, CO: McREL.
- Ministerium für Bildung, Wissenschaft und Kultur (2009). *Verordnung zur Entscheidung und zum Verfahren über den Besuch von Diagnoseförderklassen an Grundschulen (Diagnoseförderklassenverordnung - DFKVO M-V)*. Zugriff am 22.06.2015. Verfügbar unter: <http://www.landesrecht-mv.de/jportal/portal/page/bsmvprod.phtml?showdoccase=1&st=nul1&doc.id=jlr-DF%C3%B6KIVMVrahmen&doc.part=X&doc.origin=bs>
- Mlynek, H. & Forster, B. (2005). *EDI - Entwicklungsstand - Diagnostikum. Förderdiagnostisches Verfahren zur Feststellung der Lernausgangslage (Lernfähigkeit) bei Schülern im Eingangsbereich und Unterstufenbereich*. Oberostendorf: FM-Verlag.
- OECD (2006). *Bildung auf einen Blick. OECD-Indikatoren*. Paris: OECD.
- Roick, T., Göllitz, D. & Hasselhorn, M. (2004). *DEMAT 3+*. Deutscher Mathematiktest für dritte Klassen. Göttingen: Hogrefe.
- Schümer, G. (2001). Institutionelle Bedingungen schulischen Lernens im internationalen Vergleich. In: PISA 2000 (Hrsg.), *Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 411-427). Opladen: Leske + Budrich.
- Silberglitt, B., Appleton, J. J., Burns, M. K. & Jirmerson, S. R. (2006). Examining the effects of grade retention on student reading performance: A longitudinal study. *Journal of School Psychology*, 44, 255-270.
- Tietze, W. & Rossbach, H. G. (1998). Sitzenbleiben. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 465-469). Weinheim: Beltz.
- Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55-66). Slough: National Foundation for Educational Research.
- Voß, S., Mahlau, K., Sikora, S., Blumenthal, Y., Diehl, K. & Hartke, B. (2015). *Evaluationsergebnisse des Projekts „Rügener In-*

*klusionsmodell (RIM) – Präventive und Integrative Schule auf Rügen (PISaR)“ nach vier Schuljahren zum Messzeitpunkt Juli 2014. Zugriff am 21.09.2015. Verfügbar unter: <http://www.rim.uni-rostock.de/uploads/media/RIM-Evaluationsbericht-2015-druck.pdf>.*

Walberg, H. J. (1986). Synthesis of research on teaching. In M.C. Wittrock (Ed.), *Handbook of research on reaching* (pp. 214-229). New York: Macmillan.

**Dr. Yvonne Blumenthal**

Dr. Stefan Voß  
Prof. Dr. Katja Koch  
Universität Rostock - PHF/Institut für  
Sonderpädagogische  
Entwicklungsförderung und  
Rehabilitation (ISER)  
August Bebel Str. 28  
18051 Rostock  
[yvonne.blumenthal@uni-rostock.de](mailto:yvonne.blumenthal@uni-rostock.de)

**Timo Tresp**

Leibniz-Institut für die Pädagogik der  
Naturwissenschaften und Mathematik  
Abteilung Didaktik der Chemie  
Olshausenstraße 62  
24118 Kiel  
[tresp@ipn.uni-kiel.de](mailto:tresp@ipn.uni-kiel.de)

Erstmalig eingereicht: 14.01.2015

Überarbeitung eingereicht: 16.10.2015

Angenommen: 22.12.2015