

# **Modell Schriftsprach-Moderatoren (MSM)**

## **Abschlussbericht der wissenschaftlichen Begleitung nach vier Jahren**

**Wolfgang Deimel  
Gerd Schulte-Körne**

Universität Marburg, Dezember 2006

## Inhalt

1	Einleitung .....	1
2	Methode .....	3
2.1	Zeitplan und durchgeführte Untersuchungen 2004 – 2006.....	3
2.2	Messinstrumente .....	3
2.2.1	Rechtschreib- und Lesetests.....	3
2.2.2	FEESS (Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen).....	4
2.2.3	Elternfragebogen.....	5
2.2.4	Lehrkräfte.....	6
2.2.5	Unterrichtsbeobachtung .....	6
2.3	Fragestellung, Zielkriterien, Hypothesen, Signifikanzniveau, Analyseplan.....	6
2.3.1	Fragestellung .....	6
2.3.2	Zielkriterien.....	6
2.3.3	Hypothesen.....	7
2.3.4	Signifikanzniveau.....	7
2.3.5	Statistischer Analyseplan .....	7
2.4	Stichprobenbeschreibung .....	9
3	Ergebnisse.....	10
3.1	Gruppenunterschiede Rechtschreibung / Lesen .....	10
3.1.1	Deskriptive Ergebnisse .....	10
3.1.2	Primäre Analysen.....	10
3.2	Längsschnittliche Verläufe Rechtschreibung und Lesen .....	12
3.2.1	Gruppenvergleiche im Längsschnitt .....	12
3.2.2	Darstellung der Verläufe für die einzelnen Klassen .....	17
3.3	Analyse der potentiellen Prädiktoren für die Schriftsprachleistung .....	20
3.3.1	Potentielle Prädiktoren.....	20
3.3.2	Linearität und Richtung der Zusammenhänge zwischen Prädiktoren und Zielvariablen .....	21
3.3.3	Regressionen .....	22
3.4	Die Nutzung von Korrekturmöglichkeiten .....	24
3.5	Die Bedeutung individueller Voraussetzungen.....	25
3.6	Wie entwickeln sich Risikokinder?.....	26
3.7	Der Fragebogen FEESS .....	28
3.8	Unterrichtsbeobachtung .....	30
3.9	Zeugnisnoten.....	33

3.10	Lehrerfragebogen.....	35
3.11	Elternfragebogen .....	36
4	Zusammenfassung.....	40
5	Literatur.....	45
6	Anhang: Elternbogen Ende Klasse 4 .....	46



# 1 Einleitung

Das „Modell Schriftsprachmoderatoren“ (MSM) war ein Schulversuch im Aufsichtsbe-  
reich des Staatlichen Schulamts für den Schwalm-Eder-Kreis und den Kreis Waldeck-  
Frankenberg. Er wurde im Auftrag des hessischen Kultusministeriums durchgeführt. Mit  
der wissenschaftlichen Begleitung wurde Herr PD Dr. Gerd Schulte-Körne, Klinik für  
Kinder- und Jugendpsychiatrie und -psychotherapie des Universitätsklinikums Gießen und  
Marburg beauftragt.

Das Hauptziel des Modellprojektes war, durch verbesserten Unterricht im Schriftsprach-  
erwerb die Zahl von Schülerinnen und Schülern mit besonderen Schwierigkeiten im  
Rechtschreiben und Lesen so gering wie möglich zu halten. Es handelt sich also um die  
Prävention von Lese-Rechtschreibschwäche (LRS). Dieses Ziel sollte durch die Umset-  
zung zweier Unterrichts- und Förderkonzepte erreicht werden, die dank binnendifferen-  
zierender und individualisierender Methoden eine Förderung im regulären Klassenunter-  
richt ermöglichen: die „Rechtschreibwerkstatt“ von Herrn Norbert Sommer-Stumpfenhorst  
und „Lollipop“ von Herrn Wilfried Metze, Frau Gisela Dorst und Co-Autoren.

Das Leitungsteam des Projekts bestand zuletzt aus Frau Dipl.-Psych. Carmen Grebe  
(Schulpsychologin), Herrn Dipl.-Psych. Peter Kühne (Schulpsychologe, Projektleiter) und  
Frau Uta Opper-Fiedler (Schulfachliche Dezernentin).

Vor ca. zwei Jahren haben wir einen Zwischenbericht zum MSM vorgelegt. Dieser berück-  
sichtigte die Daten bis Ende der zweiten Klasse in der Kernstichprobe der 15 teilneh-  
menden Klassen (je fünf Rechtschreibwerkstatt, Lollipop und Kontrollgruppe). Da die Ergeb-  
nisse überaus spannend waren, gelang es uns, eine Verlängerung des Projekts zu erreichen.  
Diese Verlängerung bezog sich hauptsächlich auf die Hauptfragestellung nach Konzeptun-  
terschieden in den Schriftsprachleistungen der Kinder.

Das ursprüngliche Projekt hatte zwei Nebenfragestellungen. Zum einen wurde anhand von  
10 weiteren Lehrkräften, deren Daten der Kernstichprobe zugefügt wurden, Fragen zur  
Lehrerrolle und zur Konzeptumsetzung untersucht. Zum anderen wurde anhand von sechs  
Klassen (je drei Kontrollgruppe und Rechtschreibwerkstatt) orientierend die Schriftsprach-  
leistungen am Beginn der Sekundarstufe I untersucht. Diese beiden Nebenfragestellungen  
wurden in der Verlängerung nicht weiter verfolgt. Vielmehr konzentrierten wir uns auf die  
Fortführung der Untersuchung der Schriftsprachleistungen in der Grundschule. Leider gel-  
ang es uns nicht, mit der kompletten Stichprobe von 15 Klassen in die Verlängerung zu  
gehen. Zwei Lehrkräfte der Kontrollgruppe konnten nicht zur Fortführung motiviert wer-  
den. Auch in der Lollipop-Gruppe schieden zwei Klassen aus. Dies war jeweils durch  
Wechsel der Lehrkräfte bedingt, wodurch keine Kontinuität in der Konzeptumsetzung  
mehr gewährleistet war. In der Rechtschreibwerkstattgruppe wurden alle fünf Klassen wei-  
ter untersucht, für die Auswertung wurde aber eine Klasse ausgeschlossen, in der seit eini-  
ger Zeit die Rechtschreibwerkstatt nur noch im Förderunterricht und nicht mehr im regulä-  
ren Deutschunterricht stattfand.

Von den verbleibenden Klassen gab es bei zweien auch einen Lehrkraftwechsel. In einer  
RSW-Klasse wechselte die Lehrkraft Anfang der dritten Klasse. Die neue Lehrerin hatte  
aber die Fortbildung und die zweijährige Begleitung absolviert, so dass hier von einer Kon-  
tinuität der Konzeptumsetzung ausgegangen werden kann. In einer Kontrollklasse gab es  
mehrfache Wechsel, inklusive einer Referendarin hatte die Klasse im Laufe der vier Jahre  
vier verschiedene Lehrkräfte. Da bei den Kontrollklassen aber keine Vorgaben bezüglich  
der verwendeten Konzepte gemacht wurden, bleibt die Klasse für die Auswertung erhalten.

Dieser Abschlussbericht versteht sich als eine Aktualisierung des ausführlichen Zwischenberichts. Hauptanliegen ist die Darstellung der Daten von Ende der Jahrgangstufe 4. Auf eine erneute Literaturübersicht wird an dieser Stelle verzichtet; im Methodenteil werden hauptsächlich die Inhalte dargestellt, die sich seit dem Zwischenbericht verändert haben. Für eine umfassende Einführung in den Kontext und die Methoden des MSM seien die Leserinnen und Leser daher auf den Zwischenbericht verwiesen.

## 2 Methode

### 2.1 Zeitplan und durchgeführte Untersuchungen 2004 – 2006

Die regelmäßigen (halbjährlichen) Erfassungen der Rechtschreib- und Leseleistungen wurden fortgeführt, zusätzlich wurden punktuell weitere Informationen erhoben. Tabelle 1 zeigt die schematische Darstellung der Datenerhebungen zu den einzelnen Messzeitpunkten. Die Instrumente werden in den folgenden Abschnitten erläutert.

	<b>Zeitp. 7</b> (Mitte Klasse 3)	<b>Zeitp. 8</b> (Ende Klasse 3)	<b>Zeitp. 9</b> (Mitte Klasse 4)	<b>Zeitp. 10</b> (Ende Klasse 4)
<b>Kinder</b>	Rechtschreibtest Lesetest	Rechtschreibtest Korrektur RST Lesetest	Rechtschreibtest Lesetest Fragebogen FEES	Rechtschreibtest Korrektur RST Lesetest IGLU- Rechtschreibtest
<b>Eltern</b>	-	-	Fragebogen FB-E4	-
<b>Lehrer</b>	Fragebogen FB-7+	Fragebogen FB-7+	Fragebogen FB-7+	Fragebogen FB-7+ HKS-FBB
<b>Unterrichts- beobachtung</b>	Allgemein	Allgemein Konzeptspezifisch	Allgemein	Allgemein Konzeptspezifisch

Tabelle 1: Messinstrumente zu den verschiedenen Zeitpunkten

### 2.2 Messinstrumente

Bei der Beschreibung der verwendeten Instrumente sei auf die Darstellung im Zwischenbericht verwiesen. An dieser Stelle werden nur die Instrumente dargestellt, die im Laufe der letzten zwei Jahre dazu gekommen sind.

#### 2.2.1 Rechtschreib- und Lesetests

Zu allen Zeitpunkten wurde wieder die Würzburger Leise-Leseprobe (WLLP, Küspert & Schneider 1998) durchgeführt. Für die Messzeitpunkte Mitte Klasse 3 und 4 wurden die Normwerte wieder interpoliert, da diese Zeitpunkte nicht über eine eigene Normierung verfügen.

Die verwendeten Rechtschreibtests sind:

- Zeitpunkt 7: WRT2+ (Birkel 1994)
- Zeitpunkt 8: DRT 3 (Müller 2003)
- Zeitpunkt 9: DRT 4 (Grund, Haug & Naumann 2003)
- Zeitpunkt 10: DERET 3-4 (Stock, in Vorbereitung) und Gutschrift|diagnose 3-6 (Löffler & Meyer-Schepers 2001)

Der DERET 3-4 ist noch nicht erschienen, zur Zeit der Testdurchführung war die Normierung aber bereits abgeschlossen und die Materialien wurden uns von der Autorin Frau Stock freundlicherweise zur Verfügung gestellt.

Ende der vierten Klasse wurde zusätzlich zum regulären Rechtschreibtest noch „Gutschrift|Diagnose 3-6“ durchgeführt. Dieser Test ist identisch mit DoSE (Dortmunder Schriftkompetenzermittlung) von denselben Autorinnen; der Test wurde lediglich umbenannt. DoSE ist der Test, der für die Ermittlung der orthografischen Kompetenz im Rahmen der IGLU-Studie verwendet wurde. In Bos et al (2004) sind bundeslandspezifische Kennwerte (Mittelwerte und Streuungen) aufgeführt, so dass wir mit Gutschrift|Diagnose bzw. DoSE unsere Gruppen direkt mit den hessischen IGLU-Ergebnissen vergleichen können. Hierzu werden unsere Rohwerte unter Verwendung vom hessischen Mittelwert und der Standardabweichung in T-Werte transformiert.

Jeweils am Ende eines Schuljahres erhielten die Kinder Gelegenheit, ihre Rechtschreibtests (DRT 3 bzw. DERET 3-4) zu korrigieren. Die Erfassung der korrigierten Werte ist sinnvoll, da das Korrigieren nach einem Diktat in Grundschulklassen üblich ist. Die Bestimmung der T-Werte der Tests erfolgte aufgrund der unkorrigierten Testhefte, da nur dann die Normwerte anwendbar sind. Die Korrektur erfolgte deshalb auf ein getrenntes Blatt und die Kinder durften nicht im Original-Testheft korrigieren. Die korrigierten Tests sind auch Grundlage für die Kategoriale Fehleranalyse, bei der die Tests „qualitativ“ nach bestimmten Fehlertypen ausgewertet werden.

### **2.2.2 FEESS (Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen)**

Es handelt sich beim FEESS 3-4 (Rauer & Schuck 2003) um einen Fragebogen für Grundschüler dritter und vierter Klassen. Er besteht aus zwei Teilen mit 37 bzw. 53 Items, die zu insgesamt sieben Skalen zusammengefasst werden. Die folgenden Beschreibungen sind dem Manual (S. 9 und 10) entnommen.

- Soziale Integration (SI): Ausmaß, in dem ein Kind sich durch die Mitschüler und Mitschülerinnen angenommen fühlt und sich selbst als vollwertiges Gruppenmitglied betrachtet.
- Klassenklima (KK): Ausmaß, in dem die Kinder der Klasse sozial angemessen und freundschaftlich miteinander umgehen und ein gutes Verhältnis zueinander haben.
- Selbstkonzept der Schulfähigkeit (SK): Ausmaß, in dem ein Kind sich den schulischen Aufgaben gewachsen fühlt und seine schulischen Fähigkeiten positiv bewertet.
- Schuleinstellung (SE): Ausmaß, in dem ein Kind sich in der Schule insgesamt wohl fühlt.
- Anstrengungsbereitschaft (AB): Ausmaß, in dem ein Kind bereit ist, sich auf Neues einzulassen und Anforderungen in der Schule zu bewältigen, auch wenn dazu besondere Bemühungen erforderlich sind.
- Lernfreude (LF): Ausmaß, in dem ein Kind Freude an seiner alltäglichen schulischen Arbeit hat und mit froher Erwartungshaltung an seine Arbeit geht.
- Gefühl des Angenommenseins (GA): Ausmaß, in dem ein Kind sich von seinen Lehrern und Lehrerinnen angenommen, verstanden und unterstützt fühlt

Abbildung 1 veranschaulicht die Struktur des FEESS.



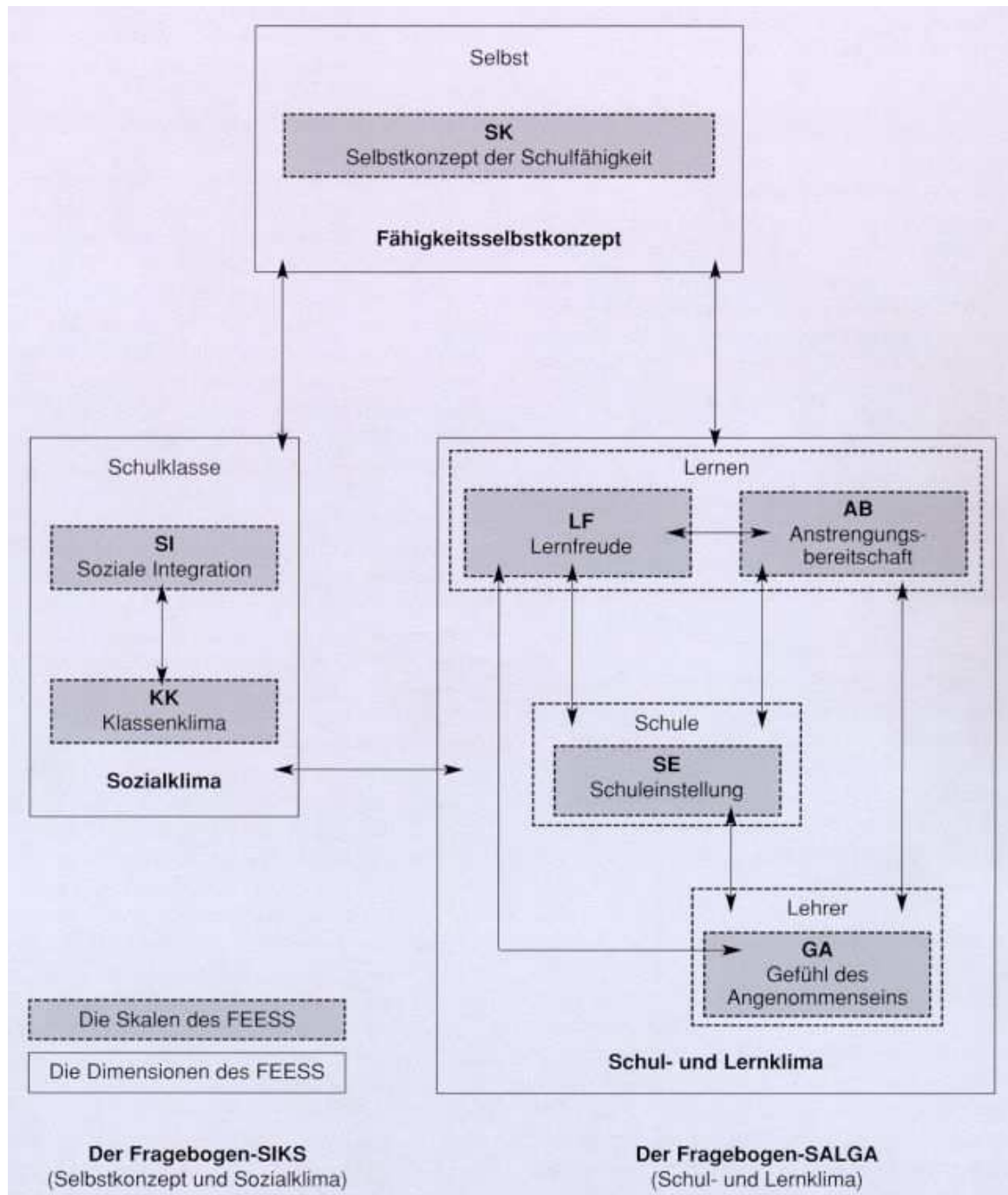


Abbildung 1: Struktur des FEESS (entnommen dem Manual S. 11)

### 2.2.3 Elternfragebogen

Gegen Ende der vierten Klasse wurde noch einmal ein Elternfragebogen verteilt (dokumentiert im Anhang). Wie bereits in der zweiten Klasse wurde detailliert außerschulische Förderung abgefragt. Außerdem wurde erfragt, ob die Eltern über die Hausaufgaben hinaus Lesen und Schreiben üben, ob bzw. wie lange sich das Kind am Rechner mit Lernprogrammen bzw. Spielen beschäftigt und wie lange es fernsieht. Schließlich fragten wir, ob die Bekanntgabe der Zwischenergebnisse Konsequenzen hatte und wie die Eltern das Modellprojekt allgemein bewerten.

## **2.2.4 Lehrkräfte**

Der bisher verwendete Fragebogen zu Einstellungen und Unterrichtspraxis wurde in gekürzter Fassung weiterverwendet. Da Verhaltensauffälligkeiten der Kinder aus dem HKS-Spektrum bei der Zwischenauswertung eine wichtige Rolle gespielt hatten, wurde dieses Konstrukt erneut erfasst, um aktuelle Daten zu erhalten. Verwendet wurde dazu der „Fremdbeurteilungsbogen HKS“ (FBB-HKS) aus dem DISYPS-KJ (Diagnostik-System für psychische Störungen im Kindes- und Jugendalter nach ICD-10 und DSM-IV, Döpfner & Lehmkuhl 2000). Das DISYPS-KJ orientiert sich, wie der Name nahelegt, an den beiden großen Diagnosemanualen für psychische Störungen. Der FBB-HKS besteht aus 20 Verhaltensbeschreibungen, denen für das jeweilige Kind auf einer vierstufigen Skala abgestuft zugestimmt werden kann („gar nicht“ bis „besonders“). Im Original wird jedes Item zweifach beurteilt, zum einen, wie zutreffend die Beschreibung ist, zum anderen, als wie problematisch das Verhalten erlebt wird. Wir verwendeten nur den ersten Teil, also die Ausprägung des Verhaltens.

## **2.2.5 Unterrichtsbeobachtung**

Zu jedem Zeitpunkt wurde wieder die allgemeine Unterrichtsbeobachtung, bestehend aus Fragebogen und MAI (s. Zwischenbericht) durchgeführt. Die konzeptspezifische Unterrichtsbeobachtung wurde aus ökonomischen Gründen nur zweimal, jeweils am Ende der dritten und vierten Klasse, durchgeführt. Die Instrumente wurden nicht verändert.

## **2.3 Fragestellung, Zielkriterien, Hypothesen, Signifikanzniveau, Analyseplan**

### **2.3.1 Fragestellung**

Die Hauptfragestellung unserer Studie hat sich nicht verändert, sie wird hier nun auf die Daten Ende der vierten Klasse angewendet.

Die Frage ist, ob sich die drei Gruppen, also Rechtschreibwerkstatt (RSW), Lollipop (LP) und Kontrollgruppe (KG) am Ende der vierten Klasse hinsichtlich der Schriftsprachleistungen voneinander unterscheiden, wenn die relevanten Prädiktoren berücksichtigt werden.

Diese Prädiktoren wurden für den Zwischenbericht ausgewertet und zusammengefasst, anschließend wurden sie auf Relevanz geprüft. Für die Endauswertung verwenden wir dieselben Prädiktoren, werden sie aber erneut komplett auf Relevanz prüfen, da sich der Einfluss der Prädiktoren durchaus geändert haben kann.

Wir werden dann die selektierten Einflussfaktoren auf Schriftsprachleistungen als Kovariate im Modell berücksichtigen, um zu vermeiden, dass Unterschiede zwischen den Gruppen in irgendeiner der Kovariaten fälschlicherweise als Gruppenunterschied interpretiert werden. Diese Fragestellung kann inferenzstatistisch als Hypothese behandelt werden.

### **2.3.2 Zielkriterien**

Die Zielkriterien sind dieselben wie im Zwischenbericht, sollen hier aber der Verständlichkeit halber erneut aufgeführt werden.

### **2.3.2.1 Primäre Zielkriterien**

Primäre Zielkriterien für die Schriftsprachleistung sind

1. die Rechtschreibleistung (RST10),
2. der Anteil der Kinder mit einer Rechtschreibschwäche (RST10\_pr10),
3. die Leseleistung (LES10) und
4. der Anteil der Kinder mit einer Leseschwäche (LES10\_pr10),

jeweils erhoben am Ende der vierten Klasse.

Rechtschreibschwäche und Leseschwäche wird angenommen, wenn ein individueller Testwert kleiner oder gleich  $T=37$  ist. Das entspricht einem Prozentrang von 10. Dieser Wert stellt einen Kompromiss dar; oft wird 5% als Prävalenzschätzung für Legasthenie angegeben, aber auch der Wert von einer Standardabweichung unter dem Mittelwert (entspricht Prozentrang 16) ist als Maß für eine unterdurchschnittliche Leistung gebräuchlich. Zudem führt eine strengere Definition, also z.B. Prozentrang 5, dazu, dass auf Klassenebene die Quoten schwacher Kinder sehr stark schwanken, da bereits ein Kind mehr oder weniger in kleinen Klassen die Quote stark verändert.

### **2.3.2.2 Sekundäre Zielkriterien**

Es gibt keine sekundären Zielkriterien

### **2.3.3 Hypothesen**

Die primäre Hypothese lautet, dass sich die drei Gruppen RSW, LP und KG hinsichtlich der primären Zielkriterien voneinander unterscheiden.

Sollte diese Hypothese bewiesen werden können, wird überprüft, ob sich RSW von LP, RSW von KG und LP von KG hinsichtlich der primären Zielkriterien unterscheiden.

Die Hypothesen werden in Abschnitt 2.3.5 weiter formalisiert.

Alle anderen Fragestellungen sind deskriptiver bzw. explorativer Art.

### **2.3.4 Signifikanzniveau**

Das globale Signifikanzniveau dieses Zwischenberichts wird auf  $\alpha = 0.05$  fixiert.

### **2.3.5 Statistischer Analyseplan**

#### **2.3.5.1 Primäre Analysen**

##### **2.3.5.1.1 A-priori-Ordnung der Hypothesen**

Die Studie umfasst die vier primären Zielkriterien, die in Abschnitt 2.3.2.1 formuliert wurden. Die Konzepte RSW und LP zielen primär auf eine Verbesserung der Rechtschreibleistung ab. Eine Verbesserung der Leseleistung wird ebenfalls angestrebt. Aus diesem Grund werden die Zielkriterien nicht als zusammengefasster Endpunkt analysiert sondern hierarchisch. Das bedeutet, dass Rechtschreibung vor Lesen getestet wird.

Da es nicht nur das Ziel ist, die Rechtschreibleistung bzw. die Leseleistung zu verbessern, sondern mit den Programmen gleichzeitig der Anteil förderbedürftiger Kinder reduziert

werden soll, werden RST10 und RST10\_pr10 sowie LES10 und LES10\_pr10 gleichberechtigt behandelt. Entsprechend wird hier für die primäre Hypothese für multiples Testen adjustiert.

Die a-priori-Ordnung der Zielkriterien nach RST und LES führt zu folgender hierarchischer Testprozedur:

H1. RSW, LP und KG unterscheiden sich auf RST10 ( $\alpha = 0,025$ ).

Wenn H1. auf dem 2,5% Testniveau signifikant ist, werden folgende Hypothesen nachgeschaltet:

H1.a) RSW und LP unterscheiden sich auf RST10 ( $\alpha = 0,025$ ).

H1.b) RSW und KG unterscheiden sich auf RST10 ( $\alpha = 0,025$ ).

H1.c) LP und KG unterscheiden sich auf RST10 ( $\alpha = 0,025$ ).

H2. RSW, LP und KG unterscheiden sich auf RST10\_pr10 ( $\alpha = 0,025$ ).

Wenn H2. auf dem 2,5% Testniveau signifikant ist, werden folgende Hypothesen nachgeschaltet:

H2.a) RSW und LP unterscheiden sich auf RST10\_pr10 ( $\alpha = 0,025$ ).

H2.b) RSW und KG unterscheiden sich auf RST10\_pr10 ( $\alpha = 0,025$ ).

H2.c) LP und KG unterscheiden sich auf RST10\_pr10 ( $\alpha = 0,025$ ).

Sollten H1., H1.a), H1.b), H1.c), H2., H2.a), H2.b) und H2.c) jeweils auf dem 2,5% Fehlerniveau signifikant sein, wird folgendes Hypothesensystem weiter getestet:

H3. RSW, LP und KG unterscheiden sich auf LES10

Wenn H3. auf dem 2,5% Testniveau signifikant ist, werden folgende Hypothesen nachgeschaltet:

H3.a) RSW und LP unterscheiden sich auf LES10 ( $\alpha = 0,025$ ).

H3.b) RSW und KG unterscheiden sich auf LES10 ( $\alpha = 0,025$ ).

H3.c) LP und KG unterscheiden sich auf LES10 ( $\alpha = 0,025$ ).

H4. RSW, LP und KG unterscheiden sich auf LES10\_pr10 ( $\alpha = 0,025$ ).

Wenn H4. auf dem 2,5% Testniveau signifikant ist, werden folgende Hypothesen nachgeschaltet:

H4.a) RSW und LP unterscheiden sich auf LES10\_pr10 ( $\alpha = 0,025$ ).

H4.b) RSW und KG unterscheiden sich auf LES10\_pr10 ( $\alpha = 0,025$ ).

H4.c) LP und KG unterscheiden sich auf LES10\_pr10 ( $\alpha = 0,025$ ).

Es ist nicht zu erwarten, dass das Testsystem bis zu den Hypothesen H3 und H4 kommt. Es ist eher anzunehmen, dass die Prozedur vorher abbricht. Da jedoch die Rechtschreibleistung der Leseleistung zu bevorzugen ist und zugunsten einer höheren statistischen Macht nicht für weiteres multiples Testen adjustiert werden soll, wird das Signifikanzniveau nicht auf die Hypothesen H1 bis H4 gleichmäßig aufgeteilt.

### 2.3.5.2 Sekundäre Analysen

Explorativ werden folgende detaillierte Analysen durchgeführt:

- die Fortschritte in der Rechtschreibung und dem Lesen im Zeitverlauf,
- die Bedeutung individueller Voraussetzungen,
- die Entwicklung von Risiko-Kindern,
- elterliches Verhalten,
- Unterrichtsbeobachtung,
- Rückmeldungen über die Konzeptumsetzungen,

sowie einige weitere Fragestellungen.

## 2.4 Stichprobenbeschreibung

Durch die Reduktion der Stichprobe verbleiben zusammen zehn Klassen in der Auswertung, jeweils drei in LP und KG sowie vier in RSW.

Für die statistischen Analysen werden nur die Kinder berücksichtigt, die seit Einschulung dabei sind. Ab- und Zugänger werden nicht berücksichtigt, da durch fehlende Daten die Analysen – je nach Fragestellung – auf stark unterschiedlichen Stichprobengrößen beruhen würden.

Ebenfalls von der weiteren Auswertung ausgeschlossen wurden zwei Integrationskinder und ein Kind, welches längere Zeit stationär in einer Klinik untergebracht war und deshalb nicht am normalen Unterricht teilnehmen konnte. Daraus resultiert eine Gesamtstichprobe von n=159. Tabelle 2 zeigt die Stichprobenzusammensetzung.

	<b>RSW</b>	<b>LP</b>	<b>KG</b>	<b>Gesamt</b>
<b>n</b>	63	51	45	159

Tabelle 2: Stichprobengröße

Die Tests (Rechtschreibung, Lesen, IQ) liegen von allen Kindern vor. Wenn Kinder bei einem der Testtermine fehlten, wurden sie nachuntersucht. Auch die Elternbögen 1 und 2 liegen für alle Kinder vor. Lediglich vom letzten Elternbogen (s. 2.2.3) fehlen vier Exemplare. Da dieser Bogen aber nur für ausgewählte Fragestellungen benötigt wird, wurde die Stichprobe nicht weiter reduziert.

## 3 Ergebnisse

### 3.1 Gruppenunterschiede Rechtschreibung / Lesen

#### 3.1.1 Deskriptive Ergebnisse

Zunächst erfolgt die Beschreibung der gruppenspezifischen Ergebnisse für die vier Hauptzielkriterien und den zusätzlich durchgeführten Rechtschreibtest DoSE (IGLU). Tabelle 3 zeigt die Gruppenmittelwerte am Ende der vierten Klasse.

	<b>RSW</b>	<b>LP</b>	<b>KG</b>
RST (T-Werte)	47,87±8,96	49,75±10,76	44,91±7,47
RST (Anteil Rechtschreibschwache %)	15,87±36,84	9,80±30,03	13,33±34,38
IGLU-Hessen (T-Werte)	50,91±9,68	53,25±9,68	47,10±9,27
IGLU-Hessen (Anteil Rechtschreibschwache %)	11,11±31,68	5,88±23,76	17,78±38,66
Lesetest (T-Werte)	56,79±10,50	59,88±9,86	56,04±10,58
Lesetest (Anteil Leseschwache %)	4,76±21,47	1,96±14,00	4,44±20,84

Tabelle 3: Gruppenmittelwerte für die vier Zielvariablen und den IGLU-Test am Ende der vierten Klasse

Bei den RST-Mittelwerten liegt Lollipop praktisch genau auf dem Erwartungswert 50. Auch der Anteil Rechtschreibschwacher liegt mit aufgerundet 10% beim Erwartungswert. Die Lollipop-Klassen entsprechen somit genau den deutschlandweiten Normen des Tests. Im Vergleich dazu liegen die Werte der RSW-Gruppe etwas niedriger, bzw. der Anteil Rechtschreibschwacher ist mit 16% höher als der Erwartungswert. Noch etwas ungünstiger fallen die Werte der Kontrollgruppe aus. Im Vergleich zu den Zwischenergebnissen hat sich die RSW verbessert und weist nun günstigere Werte als die KG aus.

Die IGLU-Werte liegen generell höher als die DERET-Werte. Dies liegt daran, dass Hessen im IGLU-E-Ländervergleich etwas schlechter abschnitt als der Durchschnitt. Bezogen auf die hessischen IGLU-Ergebnisse liegt die RSW-Gruppe näherungsweise an den Erwartungswerten T=50 bzw. Anteil 10%. Unsere Kontrollgruppe weist auch im Verhältnis zum hessischen IGLU-Durchschnitt etwas ungünstigere Werte auf, während die Lollipop-Werte im Vergleich zu IGLU leicht überdurchschnittlich ausfallen.

Wenn wir diese Daten mit denen nach zwei Jahren vergleichen, dann hat sich LP leicht verschlechtert. Insbesondere die Rate Rechtschreibschwacher ist im Vergleich zur Deutschland-Norm jetzt nicht mehr erniedrigt, sondern entspricht genau der Norm.

Die Lesetest-Werte fallen weiterhin ziemlich hoch aus. Alle drei Gruppen liegen bei den T-Werten deutlich über und bei den Anteilen Leseschwacher deutlich unter der Norm, wobei Lollipop hier die besten Zahlen erreicht.

#### 3.1.2 Primäre Analysen

Zur statistischen Überprüfung der Gruppenunterschiede wurden Kovarianzanalysen berechnet, bei der der Einfluss der Prädiktoren und der Gruppenzugehörigkeit auf die Rechtschreib- bzw. Lesetestwerte geprüft wurde. Die Vorarbeiten zu diesen Analysen, z.B. Selektion der relevanten Prädiktoren, werden im Anschluss dargestellt (siehe 3.3).

Mit der Variable „Gruppe“ und den Variablen, die durch die Regression als Prädiktoren selektiert wurden (bzw. Vorwissen und TRF für LES10\_pr10) (siehe 3.3.3), wurde je eine (pro abhängiger Variable) Kovarianzanalyse (ANCOVA) durchgeführt. Dabei wurden im Modell nur die Haupteffekte spezifiziert. In Tabelle 4 sind zusätzlich die Ergebnisse der nachgeordneten Paarvergleiche gegeben.

	<b>RST10</b>	<b>LES10</b>	<b>RST10_pr10</b>	<b>LES10_pr10</b>
Haupteffekt Gruppe	0,28	0,079	0,81	0,71
Haupteffekt Vorwissen	<0,0001	<0,0001	<0,0001	-
Haupteffekt IQ	0,0001	0,0007	-	-
Haupteffekt HKS	0,0009	-	0,0007	-
Haupteffekt Unterrichtsmenge	0,0028	-	0,06	-
Haupteffekt Geschlecht	0,01	-	-	-
Vergleich RSW – LP	0,62	0,025	0,56	0,44
Vergleich RSW - KG	0,28	0,35	0,59	0,93
Vergleich LP – KG	0,12	0,25	0,94	0,53

Tabelle 4: Ergebnisse der Kovarianzanalysen; Tests zu p-Werten, die fett gedruckt sind, sind entsprechend der verwendeten hierarchischen Testprozedur signifikant

Gemäß unseres hierarchischen Hypothesensystems müssen wir die Testung nach dem nicht-signifikanten Ergebnis des Gruppeneffekts für RST10 abbrechen. Dies bedeutet, dass wir im gesamten System keine signifikanten Unterschiede feststellen können. Die übrigen p-Werte sollen aber unter diesem Vorbehalt trotzdem dargestellt werden. Wir sehen, dass – unabhängig vom Hypothesensystem – bezüglich des Effekts „Gruppe“ und der zugehörigen Paarvergleiche – auch auf Einzeltestebene keine signifikanten Effekte auftreten (mit Ausnahme des Paarvergleichs RSW-LP im Lesen).

Im Vergleich zu den Zwischenergebnissen haben sich die Gruppen also global, d.h. bezüglich Rechtschreibung und Lesen, soweit angenähert, dass es keine signifikanten Gruppenunterschiede mehr gibt.

Da die Testung nach zwei Jahren auf der Basis einer anderen, größeren Stichprobe beruhte, könnte man vermuten, dass die Stichprobenveränderung eine Rolle gespielt haben könnte. Deshalb haben wir für die aktuelle Stichprobe noch einmal die Testwerte Ende der zweiten Klasse getestet. Wir kommen dabei zu denselben Schlussfolgerungen wie mit der größeren Stichprobe des Zwischenberichts. Daher können wir folgern, dass die nicht-signifikanten Ergebnisse Ende der vierten Klasse kein Artefakt der Stichprobenveränderung sind, sondern tatsächlich auf der unterschiedlichen Entwicklung der Schriftsprachfertigkeiten in den Gruppen beruhen.

## 3.2 Längsschnittliche Verläufe Rechtschreibung und Lesen

### 3.2.1 Gruppenvergleiche im Längsschnitt

Im Folgenden werden die vier abhängigen Variablen der statistischen Testung auf Gruppenunterschiede im Verlauf dargestellt. Die Mittelwerte des letzten Messzeitpunkts sind also die Mittelwerte, die Grundlage der Testung (s. 3.1.2) waren. Die HSP, die zum Zeitpunkt 3 (Januar 03) durchgeführt wurde, ist bekanntlich zu leicht, wie auch diese Daten eindrucksvoll zeigen.

		Zeitp.	RST		Lesen	
			Mittelwert	Standard-abw.	Mittelwert	Standard-abw.
<b>Kontrollgr.</b>	Januar 03	3	52,44	8,30	-	-
	Juli 03	4	42,38	5,36	45,80	9,83
	Januar 04	5	46,09	7,38	50,80	9,56
	Juli 04	6	43,51	7,09	50,31	10,41
	Januar 04	7	44,76	8,22	50,42	9,55
	Juli 05	8	45,47	8,60	52,16	10,50
	Januar 06	9	45,02	9,99	54,44	11,95
	Juli 06	10	44,91	7,47	56,04	10,58
<b>Lollipop</b>	Januar 03	3	50,51	9,23	-	-
	Juli 03	4	45,39	6,71	51,57	7,14
	Januar 04	5	49,12	8,64	55,88	8,22
	Juli 04	6	51,59	8,71	54,47	9,35
	Januar 04	7	51,57	8,88	54,57	9,13
	Juli 05	8	51,45	8,41	55,53	9,90
	Januar 06	9	50,55	10,73	58,57	10,92
	Juli 06	10	49,75	10,76	59,88	9,86
<b>RSW</b>	Januar 03	3	52,40	7,35	-	-
	Juli 03	4	42,32	5,22	49,03	10,90
	Januar 04	5	44,22	8,96	52,54	10,31
	Juli 04	6	44,10	7,69	52,19	11,79
	Januar 04	7	44,57	8,76	50,60	11,02
	Juli 05	8	47,95	8,72	53,05	12,12
	Januar 06	9	47,56	8,75	54,89	12,49
	Juli 06	10	47,87	8,96	56,79	10,50

Tabelle 5: Rechtschreib- und Lesetestwerte (T-Werte) in den drei Gruppen



Wenn man für die Kontrollgruppe einen normgemäßen Verlauf (Wissenszuwachs) annimmt, dann beträgt der Schwierigkeitsunterschied zwischen HSP1+ (Januar 03) und WRT1+ (Juli 03) 10 T-Punkte, also eine ganze Standardabweichung. Für die grafische Darstellung der Verläufe soll dieser Niveauunterschied angepasst werden, da die Verlaufskurven sonst sehr irreführend wären. Daher werden die HSP-Werte für die Darstellung um 10 T-Punkte nach unten korrigiert. Die BISC-Werte werden wieder als erster Messzeitpunkt in die Grafiken aufgenommen. Da es sich nicht um auf eine Normgruppe bezogene T-Werte handelt, werden die Werte so adjustiert, dass sich für die Kontrollgruppe derselbe Wert wie für die Messzeitpunkte 3 und 4 ergibt. Die ersten drei Messzeitpunkte verlaufen aufgrund dieser Korrekturen für die Kontrollgruppe also genau waagrecht.

Beim Lesetest (WLLP) gibt es die Normproblematik nicht, da immer dasselbe Verfahren durchgeführt wurde. Hier sind nur die BISC-Werte wieder so zu justieren, dass sich für die KG ein waagerechter Verlauf ergibt. Die WLLP ist erst ab Ende Kl. 1 durchführbar, daher entfällt hier Messzeitpunkt 3.

Tabelle 5 zeigt zunächst die Werte für Rechtschreib- und Lesetests in den drei Gruppen, Tabelle 6 entsprechend die Anteile Rechtschreib- bzw. Leseschwacher. In Abbildung 2 - Abbildung 5 werden die Verläufe grafisch dargestellt.

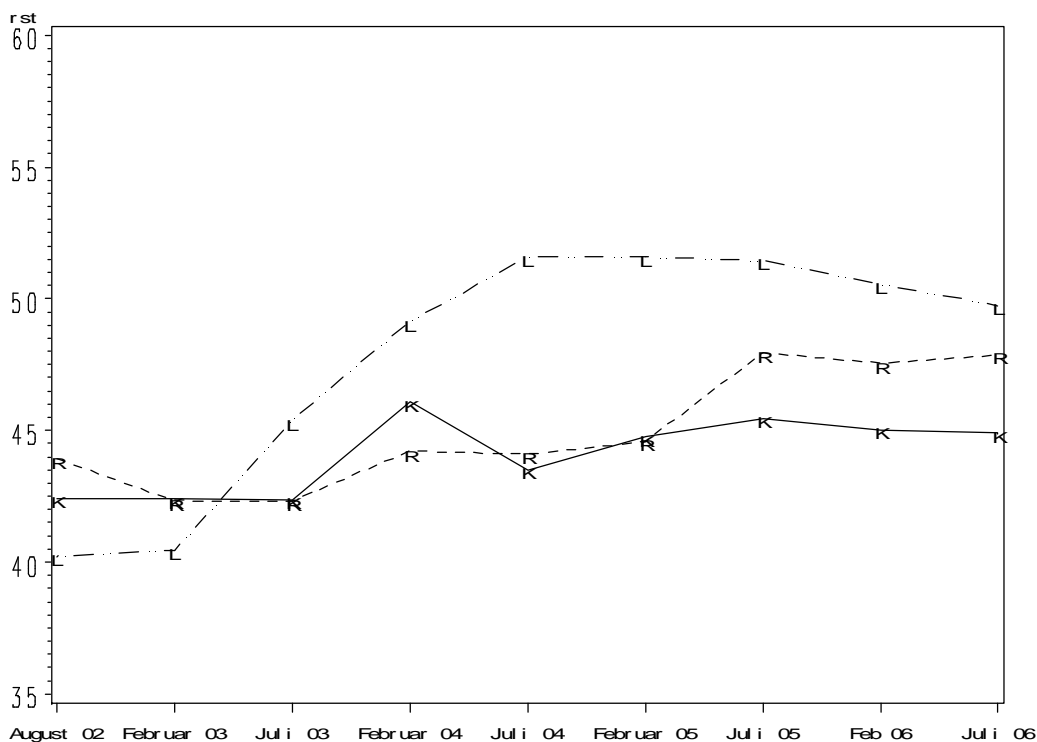


Abbildung 2: Verläufe für die Rechtschreib-T-Werte (bzw. BISC-Werte für den ersten Messzeitpunkt) in den drei Gruppen. R=RSW, L=Lollipop und K=Kontrollgruppe

		Zeitp.	RST		Lesen	
			Mittelwert	Standard-abw.	Mittelwert	Standard-abw.
<b>KG</b>	Juli 03	4	17,78	38,66	22,22	42,04
	Januar 04	5	11,11	31,78	11,11	31,78
	Juli 04	6	24,44	43,46	13,33	34,38
	Januar 04	7	13,33	34,38	11,11	31,78
	Juli 05	8	20,00	40,45	11,11	31,78
	Januar 06	9	20,00	40,45	6,67	25,23
	Juli 06	10	13,33	34,38	4,44	20,84
<b>Lollipop</b>	Juli 03	4	3,92	19,60	1,96	14,00
	Januar 04	5	3,92	19,60	1,96	14,00
	Juli 04	6	3,92	19,60	1,96	14,00
	Januar 04	7	5,88	23,76	1,96	14,00
	Juli 05	8	1,96	14,00	3,92	19,60
	Januar 06	9	5,88	23,76	1,96	14,00
	Juli 06	10	9,80	30,03	1,96	14,00
<b>RSW</b>	Juli 03	4	14,29	35,27	15,87	36,84
	Januar 04	5	20,63	40,79	4,76	21,47
	Juli 04	6	23,81	42,93	4,76	21,47
	Januar 04	7	20,63	40,79	15,87	36,84
	Juli 05	8	9,52	29,59	7,94	27,25
	Januar 06	9	12,70	33,56	7,94	27,25
	Juli 06	10	15,87	36,84	4,76	21,47

Tabelle 6: Anteile (%) Rechtschreib- und Leseschwacher in den drei Gruppen

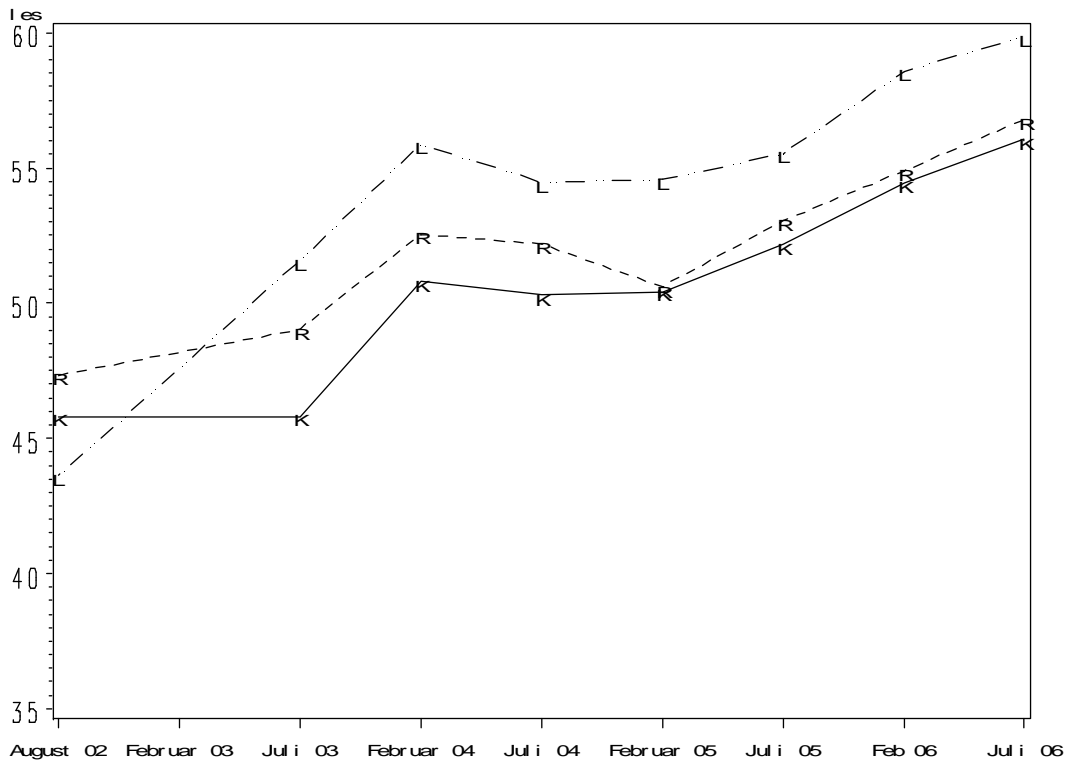


Abbildung 3: Verläufe für die Lese-T-Werte (bzw. BISC-Werte für den ersten Messzeitpunkt) in den drei Gruppen. R=RSW, L=Lollipop und K=Kontrollgruppe

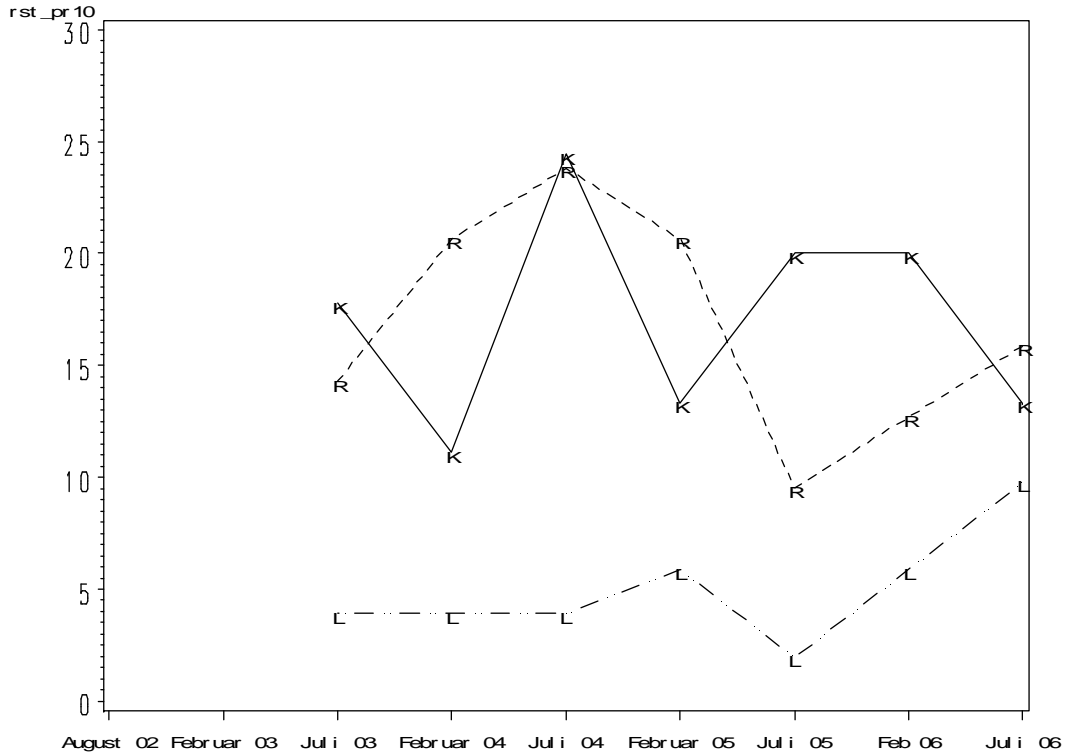


Abbildung 4: Anteile von Kindern mit Prozentrang < 10 im Rechtschreibtest in den drei Gruppen. R=RSW, L=Lollipop und K=Kontrollgruppe

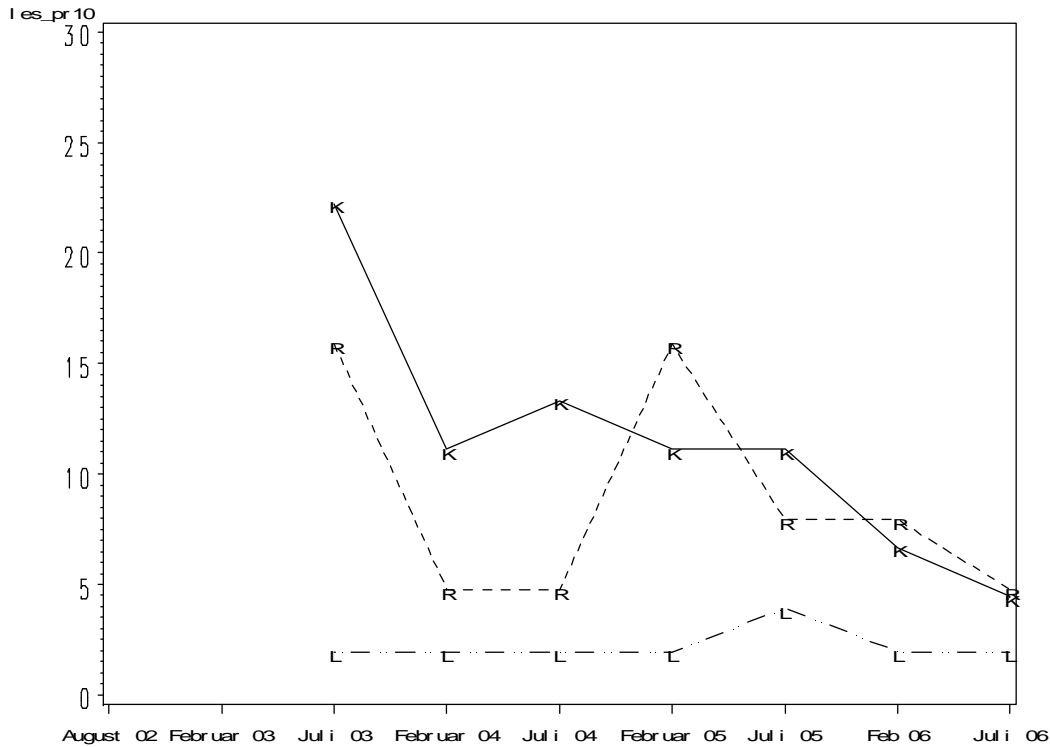


Abbildung 5: Anteile von Kindern mit Prozentrang < 10 im Lesetest in den drei Gruppen. R=RSW, L=Lollipop und K=Kontrollgruppe

Bei den Rechtschreib-T-Werten (Abbildung 2) sieht man, dass sich die Gruppen gegen Ende der vier Jahre in ihren Mittelwerten annähern. Wichtig ist der Vergleich mit der entsprechenden Abbildung im Zwischenbericht, da sich die Stichprobe ja verkleinert hat. Wir sehen, dass durch die Stichprobenveränderung die Kurve der Kontrollgruppe etwas abgesackt ist. Das führt zu einer Annäherung von KG und RSW.

Bei den Anteilen Rechtschreibschwacher (Abbildung 4) sieht man die Anfälligkeit dieser Darstellung gegenüber Zufallsschwankungen, da die Verläufe von RSW und KG deutliche Auf- und Abwärtsbewegungen vollführen, die kaum inhaltlich erklärbar sind. Vielmehr kann angenommen werden, dass Pendelbewegungen einzelner Kinder um die Schwelle für Rechtschreibschwäche ( $T=37$ ) herum zu diesen Sprüngen führen. Als Tendenz erkennt man in dieser Abbildung aber, dass es bei Lollipop in der vierten Klasse eine deutliche Zunahme der Zahl Rechtschreibschwacher gibt.

Bei den Lese-T-Werten setzt sich im Wesentlichen die bekannte Entwicklung fort. Alle drei Gruppen verzeichnen weitere relative Fortschritte gegenüber der Norm. Es sollte in Betracht gezogen werden, dass die WLLP mit zunehmendem Grundschulalter auch zu leicht wird. Weiterhin ist zu berücksichtigen, dass wir dieses Verfahren sehr oft durchgeführt haben. Es ist zwar nicht anzunehmen, dass die Kinder in nennenswertem Umfang Wörter und Bilder der Testhefte erinnern konnten, aber sicher ist die zunehmende Vertrautheit mit der Testsituation und den Durchführungsmodalitäten der WLLP ein Punkt, der vielleicht zum Teil die guten Ergebnisse erklärt. Die Anteile Leseschwacher (Abbildung 5) liefert keine zusätzlichen Erkenntnisse.

### 3.2.2 Darstellung der Verläufe für die einzelnen Klassen

Nach der Betrachtung der Werte auf Gruppenebene sollen nun die Verläufe der einzelnen Klassen angeschaut werden. Hier geht es in erster Linie darum, ob die Gruppenmittelwerte durch ähnliche Entwicklungen in allen Klassen zustande kommen oder ob einzelne Ausreißer die Mittelwerte stark beeinflussen. Abbildung 6 und Abbildung 7 zeigen die Verläufe für Rechtschreibung und Anteil Rechtschreibschwacher sowie Lesen und Anteil Leseschwacher.

Durch die Stichprobenverkleinerung werden die Kurvenverläufe insgesamt unruhiger, besonders trifft das für die Anteile Rechtschreib- bzw. Leseschwacher zu.

Hinsichtlich der Verlaufstendenzen bei der Rechtschreibung finden wir für die je drei Klassen von KG und LP einigermaßen homogene Verläufe, in der RSW-Gruppe laufen die Linien am Ende auseinander. Bei den Anteilen Rechtschreibschwacher sind KG und RSW zumindest bei der letzten Messung relativ dicht beieinander, während die drei LP-Klassen zu deutlich unterschiedlichen Werten kommen.

Beim Lesen und bei den Anteilen Leseschwacher sind die Werte insgesamt homogener, lediglich die drei Kontrollklassen liegen bei den Lesewerten deutlich auseinander.

Folgende Tendenzen lassen sich ablesen:

- Rechtschreibleistung: Bei der Kontrollgruppe ein im Mittel fast waagerechter Verlauf mit ganz leichtem Anstieg über die Zeit; bei Lollipop ein deutlicher Anstieg bis zum Sommer 2004, danach eine minimale Verschlechterung; bei RSW ein kollektiver Anstieg von Februar 05 zu Juli 05 bei insgesamt heterogenen Verläufen.
- Anteile Rechtschreibschwacher: Bei insgesamt sehr heterogenen Verläufen eine leichte Reduktion (= Verbesserung) der Werte für KG und RSW, die drei LP-Klassen lassen sich nicht zu einer gemeinsamen Tendenz zusammenfassen.
- Leseleistung: die Werte aller drei Gruppen steigen an.
- Anteile Leseschwacher: Verringerung der Raten für KG und RSW, LP bleibt bei konstant niedrigen Werten.

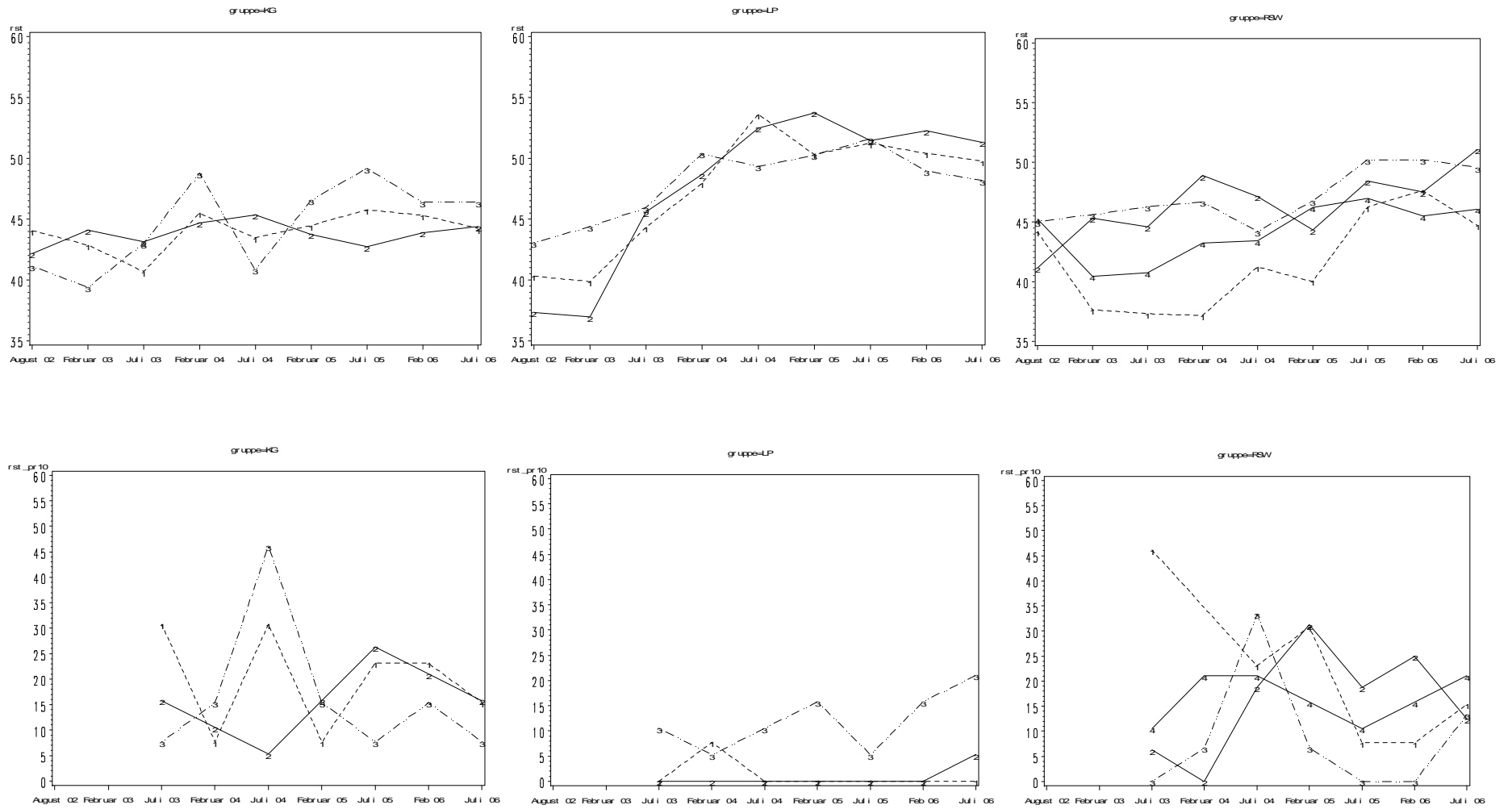


Abbildung 6: Verläufe der Rechtschreibtests für die einzelnen Klassen. Obere Reihe: T-Werte, untere Reihe: Anteile Rechtschreibschwacher in %

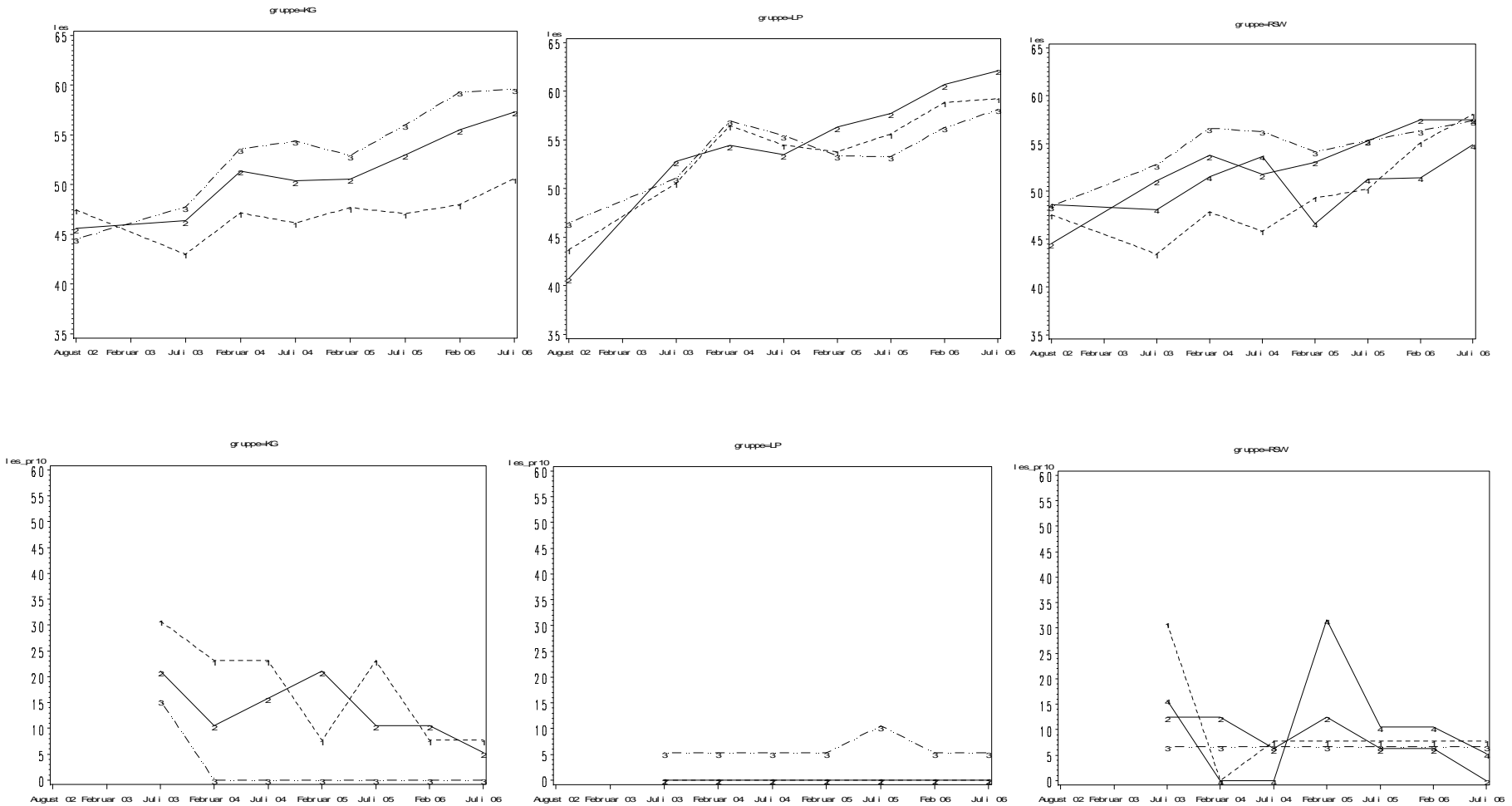


Abbildung 7: Verläufe der Lesetests für die einzelnen Klassen. Obere Reihe: T-Werte, untere Reihe: Anteile Leseschwacher in %

### 3.3 Analyse der potentiellen Prädiktoren für die Schriftsprachleistung

#### 3.3.1 Potentielle Prädiktoren

Sollte es Unterschiede in dem Erfolg der Konzepte auf die Schriftsprachleistung geben, ließe sich dieses möglicherweise auf Unterschiede in den Gruppen zu Beginn der Untersuchung zurückführen. Daher ist es wichtig zu analysieren, ob sich die Gruppen hinsichtlich der vermuteten Prädiktoren voneinander unterscheiden. Tabelle 7 zeigt die potentiellen Prädiktoren.

Beschreibung	Wertebereich
Vorwissen	Index
IQ	T-Werte
HKS-Summe	0 - 60
TV-Konsum pro Tag	1 = 0-30min. 2 = 30-60min. 3 = >60min
Schicht Soziostatus	Index 0 – 1
Schicht Wohnbedingungen	Index 0 - 1
Geburtsrang (Stellung in Geschwisterreihe)	1-x
Zweisprachig?	0=einsprachig 1=zweisprachig
Geschlecht	0=weiblich 1=männlich
Klassengröße	n
Ausländeranteil	Anteil %
Jungenanteil	Anteil %
Unterrichtsmenge Deutsch	Stunden pro Woche <sup>1</sup>

Tabelle 7: Potentielle Prädiktoren für die Schriftsprachleistung

In Tabelle 8 werden die Gruppenmittelwerte der potentiellen Prädiktoren dargestellt. Dabei wird auf dieselben Prädiktoren wie im Zwischenbericht zurückgegriffen. Vorwissen wurde wieder aus den Werten vom BISC und Buchstaben- und Zahlenkenntnis gebildet. HKS ist der Summenwert des Fragebogens FBB-HKS, der den entsprechenden Skalenwert des TRF-Fragebogens (Zwischenbericht) ersetzt. Fernsehkonsum wird durch den entsprechenden

<sup>1</sup> Die Fehlzeiten wurden monatlich, die Anzahl der wöchentlichen Deutschstunden einmal pro Halbjahr erfasst. Anschließend wurde die durchschnittliche wöchentliche Unterrichtszeit (Deutsch) berechnet, indem die durchschnittlichen wöchentlichen Fehlzeiten (Lehrkraft und Kinder) von der durchschnittlichen nominellen wöchentlichen Deutschstundenzahl abgezogen wurden. Diese Unterrichtsmenge wurde für den Zeitraum der zwei Schuljahre (Klasse 3 und 4) gemittelt.



den Wert des zuletzt erhobenen Elternbogens repräsentiert und „Beaufsichtigung der Hausaufgaben“ wurde herausgenommen, da es hierzu keinen aktuellen Wert gibt.

Die letzte Spalte enthält die durchschnittliche Effektstärke aus den drei Gruppenvergleichen.

	<b>RSW</b>	<b>LP</b>	<b>KG</b>	<b>d<sub>mean</sub></b>
Vorwissen	8,72±2,93	9,16±3,30	9,52±2,62	0,18
IQ (T-Werte)	50,13±6,87	54,02±9,46	55,46±10,0	0,41
HKS-Index	9,09±9,87	10,73±10,5	7,51±8,77	0,22
Fernsehkonsum	2,22±0,52	2,23±1,17	2,00±0,48	0,24
Soziale Schicht 1 (Soziostatus)	0,50±0,17	0,51±0,14	0,54±0,16	0,16
Soziale Schicht 2 (Wohnbedingungen)	0,82±0,16	0,87±0,16	0,91±0,17	0,35
Geburtsrang	1,80±1,16	1,82±0,82	1,59±0,71	0,18
Klassengröße	24,69±2,14	23,96±3,52	24,92±5,90	0,17
Anteil Ausländer % <sup>2</sup>	14,67±9,84	7,16±4,55	12,39±7,86	0,68
Anteil Jungen %	46,78±11,8	52,83±8,30	43,97±9,36	0,62
Unterrichtsmenge (Stunden pro Woche)	5,13 ± 0,40	5,66 ± 0,18	5,42 ± 0,54	0,98

Tabelle 8: Gruppenmittelwerte der potentiellen Prädiktoren

Die meisten mittleren Effekte sind klein. Es finden sich zwei mittelgroße Effekte: Der Jungenanteil ist bei LP am größten und in der KG am kleinsten, der Anteil ausländischer Kinder ist in der RSW am größten und in LP am geringsten. Einen großen Effekt finden wir bei der Unterrichtsmenge. Die durchschnittliche Zahl der effektiven Wochenstunden Deutsch (inklusive Förderunterricht; Stundenausfall und Krankheit der Kinder wurden abgezogen): LP hat die größte, RSW die kleinste Zahl von effektiven Wochenstunden.

### 3.3.2 Linearität und Richtung der Zusammenhänge zwischen Prädiktoren und Zielvariablen

Im nächsten Schritt wurden Streudiagramme der Prädiktoren mit den Zielvariablen gebildet (aus Platzgründen hier nicht dargestellt). Das Ziel dieses Schrittes war es, eventuelle nicht-lineare Zusammenhänge zu finden, die dann durch Korrelationen nicht adäquat abgebildet wären.

<sup>2</sup> Diese Variable geht in die spätere Auswertung in zweifacher Art ein. Zum einen wird der Ausländeranteil der Klasse verwendet (der damit für alle Kinder einer Klasse gleich ist), zum anderen wird die individuelle Sprachigkeit eines Kindes (einsprachig deutsch vs. mehrsprachig) verwendet. Es sind dies zwei Aspekte derselben Einflussgröße, nämlich Mehrsprachigkeit als individueller Einflussfaktor auf die Schriftsprachleistungen und ein hoher Ausländeranteil in der Klasse als möglicher Einflussfaktor für alle in der Klasse. Der Anteil an Jungen wird demgegenüber bei den weiteren Analysen nur als individuelle Variable Geschlecht berücksichtigt.

Durch visuelle Inspektion der Plots entstand der Verdacht auf einen nicht-linearen Zusammenhang bei dem Maß für HKS. Mittels folgender Transformation wurde annähernde Linearität hergestellt:  $hks2 = \ln(hks) * (1 / hks^{**2})$

Die neue Variable wurden anschließend mit 10000 multipliziert, um sie in den gleichen Wertebereich wie die Originalvariable zu bringen. Die Polung ist allerdings nach der Transformation anders, d.h. hohe Ausprägungen stehen jetzt nicht mehr für hohe, sondern für niedrige Auffälligkeit. Neben der Linearität der Zusammenhänge muss außerdem deren Richtung (positiv oder negativ) plausibel sein. Die Korrelationen zwischen Prädiktoren und Zielvariablen sind in Tabelle 9 dargestellt. Werte > 0,2 sind in Einzeltestung auf dem 0,01-%-Niveau signifikant und werden fett dargestellt.

Die Richtung der Zusammenhänge ist bei den signifikanten Korrelationen in allen Fällen theoriekonform.

	<b>RST10<sup>3</sup></b>	<b>RST10_pr10</b>	<b>LES10</b>	<b>LES10_pr10</b>
Vorwissen	<b>0,41</b>	<b>0,32</b>	<b>0,40</b>	0,10
IQ	<b>0,40</b>	0,16	<b>0,35</b>	0,05
HKS	<b>-0,30</b>	<b>-0,25</b>	-0,07	-0,16
HKS2	-0,19	<b>-0,21</b>	-0,02	-0,03
TV-Konsum	-0,04	-0,07	0,06	0,03
Schicht 1	<b>0,30</b>	0,19	0,16	0,04
Schicht 2	<b>0,24</b>	<b>0,26</b>	0,19	0,13
Geburtsrang	-0,18	-0,20	-0,06	-0,14
Klassengröße	-0,05	-0,05	-0,02	0,07
Sprachigkeit	-0,06	-0,10	-0,03	0,00
Ausländeranteil	-0,13	-0,08	-0,02	-0,02
Geschlecht	<b>-0,26</b>	-0,20	-0,07	-0,06
Unterrichtsmenge	0,14	0,11	0,10	0,12

Tabelle 9: Korrelationen zwischen Prädiktoren und Zielvariablen. Koeffizienten ab  $r=0,21$  entsprechen p-Werten von 1% oder kleiner und sind fett dargestellt. Die Variablen des Anteils von Rechtschreib- bzw. Leseschwachen wurden zur einheitlichen Darstellung umgepolt, d.h. eigentlich wird hier der Anteil von Kindern mit Prozentrang > 10 dargestellt.

### 3.3.3 Regressionen

Als nächster Schritt der Auswertung werden mit den bisher selektierten Variablen Regressionen (abhängige Variable auf Prädiktoren) ohne Berücksichtigung des Konzeptes durch-

<sup>3</sup> RST10: Rechtschreibtest-T-Wert Ende der vierten Klasse

RST10\_pr10: Anteil Rechtschreibschwacher mit Prozentrang < 10

LES10: Lesetest-T-Wert Ende der vierten Klasse

LES10\_pr10: Anteil Leseschwacher mit Prozentrang < 10

geführt. Das dient dazu, statistisch redundante Variablen zu entfernen, da die gleichzeitige Verwendung von hoch korrelierenden Variablen im Modell nicht sinnvoll ist. Außerdem werden Variablen mit geringem Einfluss auf die Zielvariablen durch diesen Schritt ausgeschlossen.

Es wird pro abhängiger Variable je eine schrittweise multiple Regression (abhängige Variable auf Prädiktoren) mit p-Wert Selektion auf dem 1%-Niveau durchgeführt. Tabelle 10 zeigt die selektierten Variablen pro Kriteriumsvariable. Die Zahlen unter den Kriteriumsvariablen geben die aufgeklärten Varianzen an. Für das Kriterium Anteil Leseschwacher wurde kein Prädiktor selektiert. Dies überrascht nicht, da es ja auch keine signifikanten Korrelationen gab (s. Tabelle 9).

	<b>RST10</b>	<b>RST10_pr10</b>	<b>LES10</b>	<b>LES10_pr10</b>
Gesamtmodell	0,41	0,21	0,21	-
Vorwissen	0,17	0,11	0,15	-
IQ	0,09	-	0,06	-
HKS-Index	0,06	0,06	-	-
Unterrichtsmenge	0,06	0,04	-	-
Geschlecht	0,03	-	-	-

Tabelle 10: Selektierte Variablen pro Kriteriumsvariable mit den dazugehörigen aufgeklärten Varianzen ( $r^2$ )

Wie bei den Korrelationen mit den Kriteriumsvariablen sieht man auch bei den Ergebnissen der Regressionen, dass die Zusammenhänge zwischen Prädiktoren und Kriteriumsvariablen für die T-Werte (Schreiben und Lesen) deutlich höher ausfallen als mit den jeweiligen dichotomisierten Variablen (Anteil Rechtschreib- bzw. Leseschwacher). Dies kann methodisch interpretiert werden: Durch die Dichotomisierung wird viel echte Varianz vernichtet, wodurch die Zusammenhänge zu anderen Maßen geringer ausfallen. Beim Lesen führt dieser Effekt gar dazu, dass für den Anteil Leseschwacher gar keine Variable mehr selektiert wird.

Die Gruppen unterscheiden sich leicht hinsichtlich der Prädiktoren. Dies wird deutlich an den korrigierten Gruppenmittelwerten, die bei den Kovarianzanalysen berechnet wurden („LSMEANS“). Die Unterschiede zwischen tatsächlichen und korrigierten Werten drücken aus, inwieweit eine Gruppe hinsichtlich der Prädiktoren gegenüber den anderen benachteiligt oder im Vorteil ist. Tabelle 11 zeigt exemplarisch für RST10 den Vergleich zwischen tatsächlichen und korrigierten Mittelwerten.

Man sieht, dass RSW praktisch unverändert bleibt und somit quasi als Referenz dient. Im Vergleich dazu wird der LP-Wert um genau 1 T-Punkt gesenkt und der Wert der KG um 1,3 T-Punkte angehoben. Dies bedeutet, dass Lollipop im Vergleich zu RSW etwas günstigere Bedingungen hatte (hauptsächlich mehr Deutschstunden), während die KG etwas benachteiligt war (hauptsächlich niedrigerer mittlerer IQ).

	<b>KG</b>	<b>LP</b>	<b>RSW</b>
Mittelwerte RST10	44,9	49,7	47,9
Korrigierte Mittelwerte RST10	46,2	48,7	47,8

Tabelle 11: Vergleich zwischen tatsächlichen und korrigierten Mittelwerten

### 3.4 Die Nutzung von Korrekturmöglichkeiten

Jeweils Ende der 3. und 4. Klasse wurden die Rechtschreibtests (DRT3 bzw. DERET 3-4) mit Korrekturphase durchgeführt. Die Kinder konnten dabei 15 Minuten lang mit üblichen Hilfsmitteln (Wörterbücher) die geschriebenen Wörter korrigieren. Die Ermittlung der Testnormwerte erfolgte natürlich auf Basis der unkorrigierten Testhefte, da sonst von der vorgegebenen Durchführung abgewichen worden wäre. Getrennt von den Normwerten wurde daher für die Rohwerte bestimmt, wie stark und in welche Richtung sie sich durch die Korrekturen veränderten.

Tabelle 12 zeigt die Verteilung der Verbesserungswerte für den DRT 3. Es wurden einzeln bis zu 7 Wörter verbessert, am anderen Ende der Verteilung stehen insgesamt 4 Kinder, die ihren Rohwert durch die Korrektur verschlechterten. Ein Fünftel der Stichprobe verändert den Rohwert nicht und ca. ein Viertel verbessert sich um ein Wort.

Verbesserung des Rohwerts	Anzahl Kinder	Prozent
7	1	0,6
6	2	1,3
5	6	3,4
4	16	10,1
3	16	10,1
2	39	24,5
1	44	27,7
0	31	19,5
-1	4	2,5

Tabelle 12: Verbesserungen des DRT3-Rohwerts durch die Korrekturphase

Auch beim DERET (Tabelle 13) wurden bis zu 7 Wörter verbessert, am anderen Ende der Verteilung stehen insgesamt 9 Kinder, die ihren Rohwert durch die Korrektur verschlechterten. Je ca. ein Viertel der Stichprobe verändert den Rohwert nicht bzw. verbessert sich um ein Wort.

Korrelationen zwischen dem Ausmaß der Verbesserung und dem T-Wert des Tests zeigen, dass es keinen Zusammenhang gibt (DRT3:  $r=0,07$ ; DERET:  $r=0,02$ ). Es ist also nicht so, dass z.B. besonders gute Schüler dann auch noch deutlichere Verbesserungen erreichen als schwache Schüler. Es scheint vielmehr so, dass alle gleich viel von der Korrekturphase profitieren.

Tabelle 14 zeigt die Verbesserungen aufgeschlüsselt nach Gruppenzugehörigkeit. Die mittleren Verbesserungen liegen jeweils etwa zwischen einem und zwei Wörtern. Tendenziell schneiden die Kinder der Rechtschreibwerkstatt am besten ab, jedoch sind die zugehörigen Effektstärken nur gering – was daran liegt, dass die Streuungen innerhalb der Gruppen recht groß sind.

Verbesserung des Rohwerts	Anzahl Kinder	Prozent
7	2	1,3
6	3	1,9
5	3	1,9
4	6	3,8
3	16	10,1
2	35	22,0
1	42	26,4
0	43	27,0
-1	7	4,4
-2	2	1,3

Tabelle 13: Verbesserungen des „DERET 3-4“-Rohwerts durch die Korrekturphase

	RSW	LP	KG	Effektstärke
Ende Klasse 3	-1,94±1,65	-1,45±1,42	-1,84±1,57	0,21
Ende Klasse 4	-1,73±1,78	-0,90±1,36	-1,44±1,55	0,36

Tabelle 14: Verbesserung beim Rohwert (Zahl richtiger Wörter) durch Nutzung der Korrekturphase und Gruppenvergleich mit mittlerer Effektstärke

### 3.5 Die Bedeutung individueller Voraussetzungen

Die Rolle dieser Prädiktoren wurde bereits unter 3.1 betrachtet; sie sollen hier noch einmal unter ausgewählten Fragestellungen analysiert werden.

Generell geht es darum, inwieweit demographische, soziale und biologische Merkmale die Schriftsprachleistung bestimmen. Es handelt sich hierbei also um Einflüsse, die sich einer Modifizierung durch Unterrichtung weitgehend entziehen.

Tabelle 15 zeigt den Einfluss des Sprachhintergrundes auf die Schriftsprachleistungen Ende der vierten Klasse. Einsprachig deutsche Kinder erzielen erwartungsgemäß bessere Rechtschreib- und Lesetestwerte als zweisprachige Kinder. Die Effektstärken sind gering ausgeprägt.

	Zweisprachige (n=16)	Einsprachige (n=143)	Effektstärke
<b>RST</b>	46,4 ± 8,2	47,8 ± 9,5	0,16
<b>LES</b>	56,9 ± 9,9	57,7 ± 10,5	0,08

Tabelle 15: Einfluss des Sprachhintergrundes auf die Schriftsprachleistungen

Tabelle 16 zeigt den Geschlechtseffekt auf die Schriftsprachleistungen Ende der vierten Klasse. Wie erwartet schreiben und lesen Mädchen etwas besser als Jungen. Der Effekt auf die Rechtschreibung ist mittelgroß, auf das Lesen klein.

	Jungen (n=81)	Mädchen (n=78)	Effektstärke
<b>RST</b>	45,2 ± 8,9	50,1 ± 9,2	0,54
<b>LES</b>	56,9 ± 10,5	58,3 ± 10,3	0,08

Tabelle 16: Geschlechtseffekt auf die Schriftsprachleistungen

	<b>RST</b>	<b>LES</b>
IQ	0,40	0,35
BISC	0,32	0,26
Buchstabenkenntnis	0,36	0,35
Zahlenkenntnis	0,30	0,30

Tabelle 17: Korrelationen individueller Voraussetzungen mit Rechtschreibung und Lesen Ende Klasse 4

Tabelle 17 zeigt die Korrelationen der individuellen Voraussetzungen mit Rechtschreibung und Lesen Ende der vierten Klasse. Trotz des zeitlichen Abstandes bei der Ermittlung dieser Maße (der IQ wurde im Laufe der ersten Klasse und die anderen Maße bei Einschulung erhoben) finden sich noch recht hohe Zusammenhänge. Diese fallen auch kaum schwächer aus als Ende der zweiten Klasse. Die langfristige Bedeutung dieser Maße wird damit unterstrichen.

### 3.6 Wie entwickeln sich Risikokinder?

Wir waren dieser Frage bereits im Zwischenbericht nachgegangen. Nun liegen Daten von zwei weiteren Jahren vor.

Als Maß für das Risiko zur Entwicklung von Lese-Rechtschreibschwierigkeiten verwenden wir wieder die Kombination aus BISC-Testwert, Buchstabenkenntnis und Zahlenkenntnis, jeweils bei Einschulung erhoben. Die Gesamtgruppe wird dann anhand der Verteilung dieser Variable aufgeteilt: das Viertel mit den schlechtesten Werten wird als Gruppe mit hohem Risiko (H) aufgefasst, das Viertel mit den besten Werten als Gruppe mit niedrigem Risiko (N). Die dazwischen liegende Hälfte der Stichprobe dient als Vergleich (M). Abbildung 8 zeigt die Verläufe für den Rechtschreibtest, Abbildung 9 die Verläufe für den Lesetest.

Wie bereits vor zwei Jahren verlaufen die Linien weitgehend parallel. Das heißt, dass die Teilgruppe mit den schlechtesten Startbedingungen auch noch Ende der 4. Klasse einen etwa gleich hohen Rückstand im Rechtschreiben und Lesen aufweist. Am Ende der Klasse 4 liegen die Teilgruppen mit hohem bzw. niedrigem Risiko im Rechtschreiben um 10 T-Punkte und im Lesen um 12 T-Punkte auseinander.

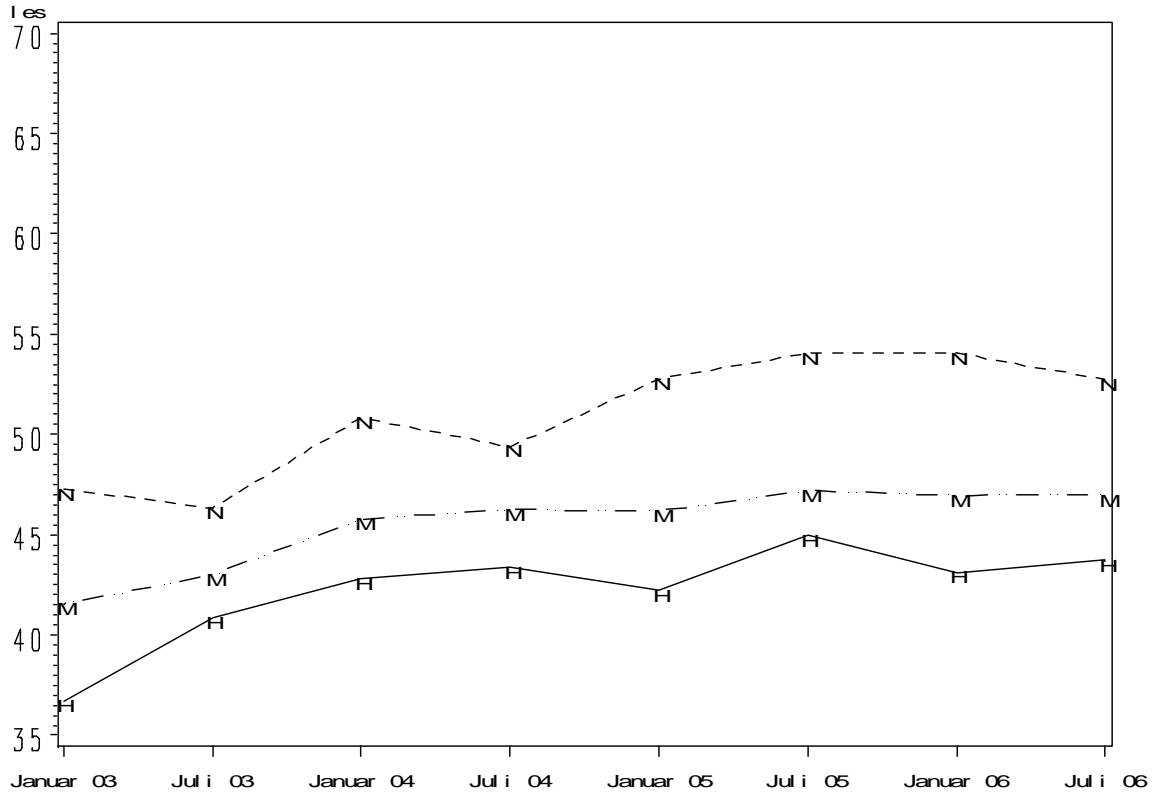


Abbildung 8: Verlauf der Rechtschreib-T-Werte. H=Hohes Risiko, M=Mittleres Risiko, N=Niedriges Risiko.

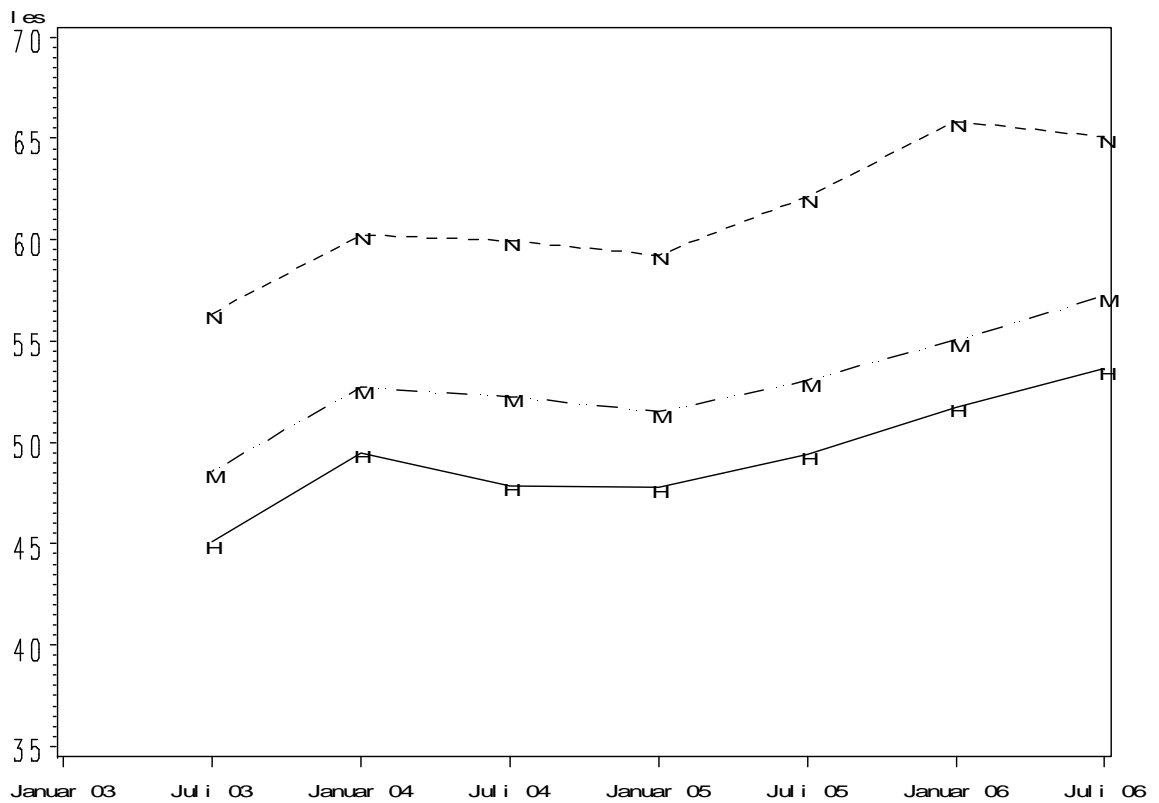


Abbildung 9: Verlauf der Lesetest-T-Werte. H=Hohes Risiko, M=Mittleres Risiko, N=Niedriges Risiko.

### 3.7 Der Fragebogen FEES

Wie bereits unter 2.2.2 ausgeführt, besteht der „Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen“ aus sieben Skalen, die als T-Werte (Mittelwert 50, STA 10) ausgegeben werden. Es handelt sich dabei um Normwerte, die auf dem Vergleich mit einer großen Normstichprobe beruhen. Tabelle 18 zeigt die Kennwerte und Effektstärken der Skalen. Die Mittelwerte streuen ziemlich eng um den Erwartungswert 50. Im Mittel entsprechen die Einschätzungen unserer Schüler also denen der Normstichprobe. Es fällt auf, dass die RSW-Werte durchgängig die höchsten sind. Da die Skalen alle gleich gepolt sind, bedeutet dies, dass die emotionalen und sozialen Erfahrungen in der RSW-Gruppe am positivsten ausgeprägt sind. Zur Einschätzung des Ausmaßes dieser Unterschiede wurden zweierlei Effektstärken berechnet.  $ES_m$  ist die mittlere Effektstärke aller drei Paarvergleiche (RSW-LP, RSW-KG, LP-KG). Die Effekte sind durchgängig klein oder unbedeutend, der Grenzwert für einen mittleren Effekt (0,5) wird nirgends erreicht. Dies unterschätzt jedoch die Unterschiede zwischen einzelnen Paaren, wenn z.B. KG und LP dicht beieinander liegen, was etwa bei SI der Fall ist. Deshalb wurden zusätzlich Effektstärken des Paarvergleichs RSW-LP ( $ES_{RL}$ ) angegeben. Die höchsten ES finden sich auf den Skalen SI, KK und GA. Aber auch diese Werte bleiben unterhalb der Schwelle eines mittelgroßen Effekts. Wir finden zwar somit durchgängig positivere Werte für die RSW, allerdings ist das Ausmaß dieser Differenzen nur gering.

Skala	Beschreibung	KG	LP	RSW	$ES_m$	$ES_{RL}$
SI	Soziale Integration	47,47±8,74	47,16±10,75	51,62±8,13	0,33	0,47
KK	Klassenklima	48,93±9,56	49,06±12,91	53,27±8,81	0,29	0,38
SK	Selbstkonzept d. Schulfähigkeit	49,20±9,88	50,86±11,35	52,22±8,79	0,20	0,13
SE	Schuleinstellung	46,53±9,10	48,37±10,31	49,38±8,32	0,21	0,11
AB	Anstrengungsbereitschaft	49,93±10,35	50,55±10,24	52,40±9,42	0,17	0,19
LF	Lernfreude	47,98±9,42	50,12±10,59	50,71±7,56	0,20	0,06
GA	Gefühl des Angenommenseins	50,64±9,56	47,41±9,52	51,52±8,13	0,30	0,46

Tabelle 18: Mittelwerte, Standardabweichungen und Effektstärken des FEES im Gruppenvergleich.  $ES_m$ : mittlere Effektstärke des Vergleichs aller drei Gruppen,  $ES_{RL}$ : Effektstärke des Vergleichs RSW – LP

Abbildung 10 zeigt die grafische Veranschaulichung der Gruppenunterschiede. Es ist zu beachten, dass die Effekte hier dramatischer aussehen, als sie in Wirklichkeit sind, da die vertikale Achse nur den Bereich zwischen 40 und 60 darstellt.



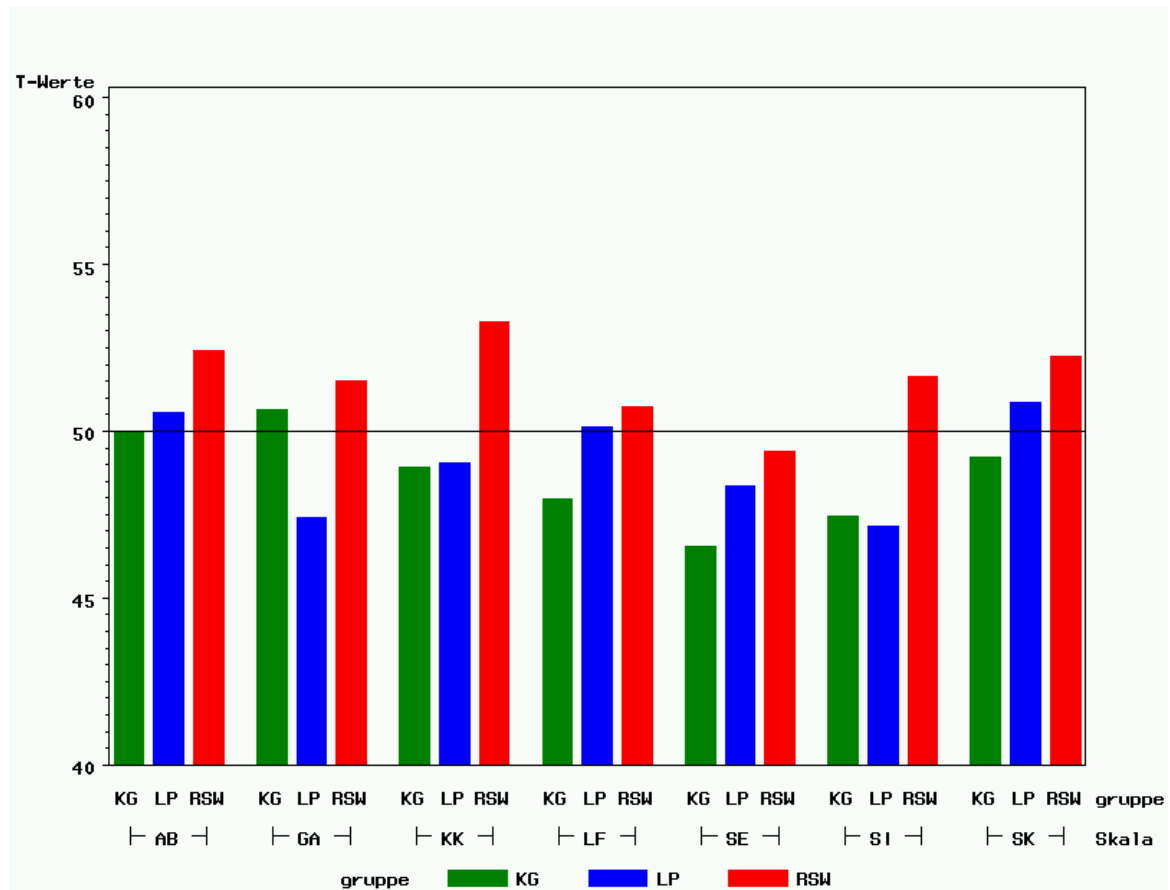


Abbildung 10: Gruppenunterschiede auf den FEESS-Skalen

Skala	Beschreibung	RST	LES
SI	Soziale Integration	0,10	0,05
KK	Klassenklima	0,08	-0,08
SK	Selbstkonzept d. Schulfähigkeit	<b>0,38</b>	<b>0,35</b>
SE	Schuleinstellung	0,17	0,05
AB	Anstrengungsbereitschaft	0,14	0,08
LF	Lernfreude	<b>0,22</b>	0,13
GA	Gefühl des Angenommenseins	0,14	0,03

Tabelle 19: Korrelationen der FEESS-Skalen mit Rechtschreibung und Lesen Ende der vierten Klasse. Fett gedruckte Werte sind auf dem 1%-Niveau signifikant.

In Tabelle 19 sind die Korrelationen der FEESS-Skalen mit Rechtschreib- und Leseleistung dargestellt. Das eingeschätzte Selbstkonzept der Schulfähigkeit hat tatsächlich einen deutlichen Zusammenhang sowohl zur Rechtschreibung als auch zum Lesen. Bei den übrigen Skalen finden wir (mit Ausnahme des Koeffizienten Lernfreude-RST) keine Zusammenhänge.

### 3.8 Unterrichtsbeobachtung

Im Folgenden werden Verläufe der wichtigsten Variablen der Unterrichtsbeobachtung dargestellt. Die Aussagekraft ist jedoch durch die Stichprobenverkleinerung deutlich eingeschränkt; die Kurven von KG und RSW stellen lediglich die Werte je dreier Klassen dar. Eine weitere Einschränkung betrifft die Beobachter. Bis Ende Klasse zwei gab es vier verschiedene Beobachter, die bei allen Messzeitpunkten dieselben Klassen besuchten. Die Beobachter waren so gut wie möglich über die drei Gruppen verteilt, um systematische Beobachtereffekte zu minimieren. In der zweiten Projekthälfte konnten wir diese Bedingungen aus personellen Gründen nicht mehr einhalten und es kam zu Beobachterwechseln.

Abbildung 11 zeigt zunächst den Verlauf der durchschnittlichen Aufmerksamkeit der Kinder im Unterricht. Die Erfassung erfolgte mit einer computerisierten und modifizierten Fassung des Münchner Aufmerksamkeits-Inventars (MAI), die auf unmittelbarer Beobachtung und Kategorisierung des Verhaltens der Kinder beruht. Es handelt sich somit um eine vergleichsweise objektive und wenig interpretierende („niedrig-inferentiell“) Erfassung des Konstrukts. Die Verläufe bis zum Ende Klasse 2 entsprechen gut denen des Zwischenberichts, d.h., dass die Stichprobenveränderung sich nicht auf die Mittelwerte ausgewirkt hat. In den ersten beiden Jahren waren die Kinder in den RSW-Klassen aufmerksamer als in den beiden anderen Gruppen, aber gegen Ende der Grundschulzeit nähern sich die Werte aller drei Gruppen an. Es gibt zudem einen allgemeinen Trend über die Zeit, da die Werte gruppenunabhängig größer werden.

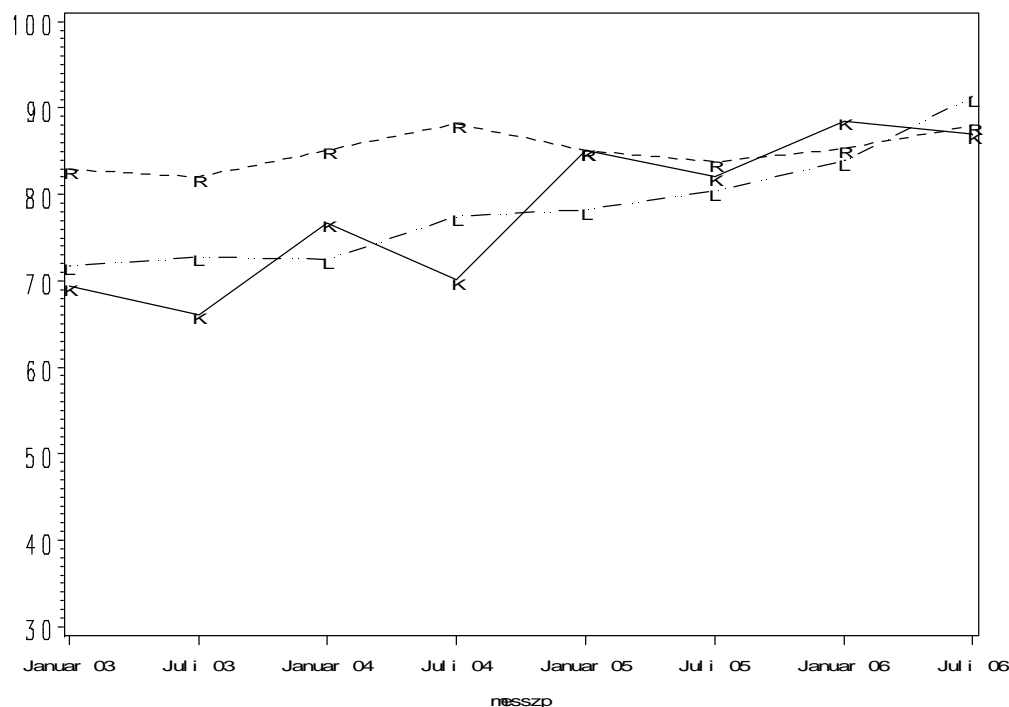


Abbildung 11: Prozentuale Aufmerksamkeit der Schüler während der Unterrichtsbeobachtung (Münchner Aufmerksamkeits-Inventar MAI)

In Abbildung 12 und Abbildung 13 sind die Ergebnisse des Beobachtungsbogens der allgemeinen Unterrichtsbeobachtung dargestellt. Durch Faktorenanalysen dieses hoch-inferentiellen (d.h.: interpretierenden) Instruments waren mit den Daten bis Ende Klasse 2 zwei Faktoren extrahiert worden, Klassenmanagement und Klassenklima. Die Werte auf der Ordinate sind Itemmittelwerte; ein Wert von „5“ bedeutet, dass die Ausprägung des Faktors im Sinne des Konstrukts maximal ist.

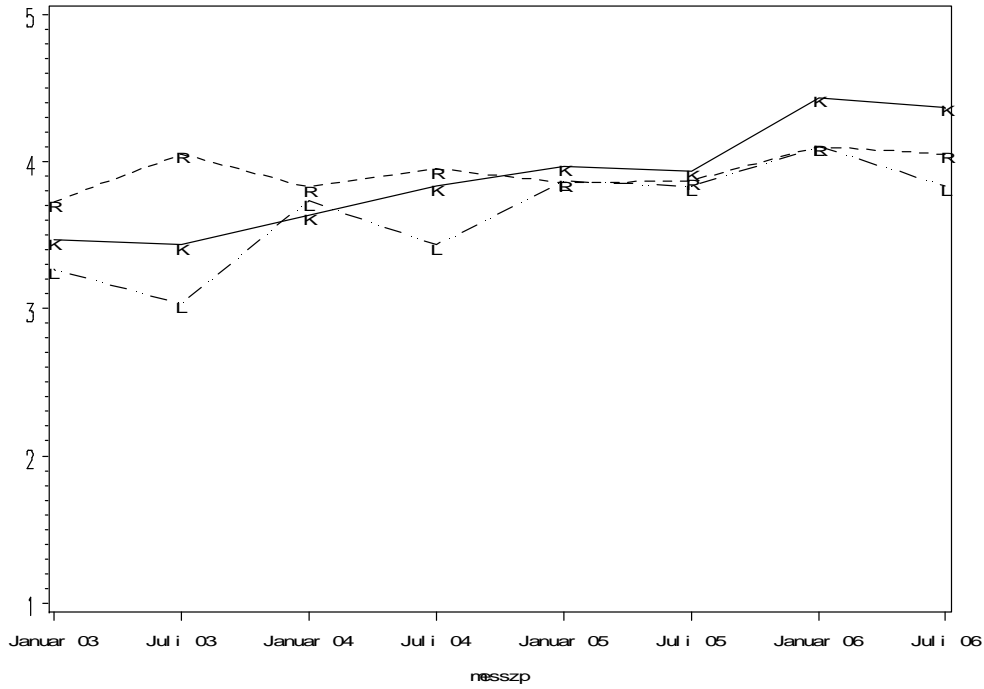


Abbildung 12: Klassenmanagement gemäß Einschätzungsbogen bei der Unterrichtsbeobachtung

Klassenmanagement (Abbildung 12): Auch hier steigen die Werte über die Zeit etwas an, die Gruppen liegen aber relativ dicht beieinander. Es gibt hier also kaum Gruppenunterschiede.

Klassenklima (Abbildung 13): Die LP-Klassen erreichen durchgängig die relativ geringsten Werte, aber die Gruppen nähern sich im Laufe der Zeit deutlich an.

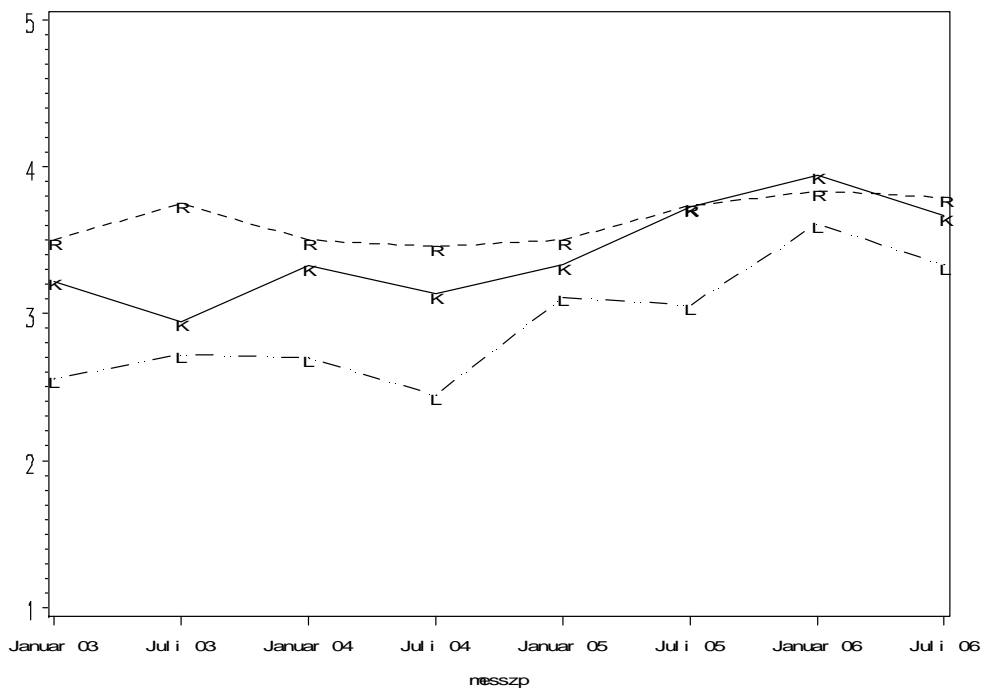


Abbildung 13: Klassenklima gemäß Einschätzungsbogen bei der Unterrichtsbeobachtung

Orientierend wurden wieder Korrelationen der Werte zum letzten Zeitpunkt mit den entsprechenden Klassenmittelwerten im Rechtschreiben und Lesen berechnet. Die Koeffizienten waren, wie im Zwischenbericht, entweder unbedeutend oder negativ, was unsinnig ist, da dies Klassen mit hohem Klassenmanagement-Werten die relativ schlechtesten Schriftsprach-Mittelwerte bescheinigt. Auch hier gilt wieder die Einschränkung der geringen Stichprobengröße, jedoch waren diese Tendenzen bereits im Zwischenbericht erkennbar. Es bleibt also weiterhin unklar, warum unsere Unterrichtsbeobachtung nicht sinnvoll mit den Schriftsprachergebnissen in Beziehung steht.

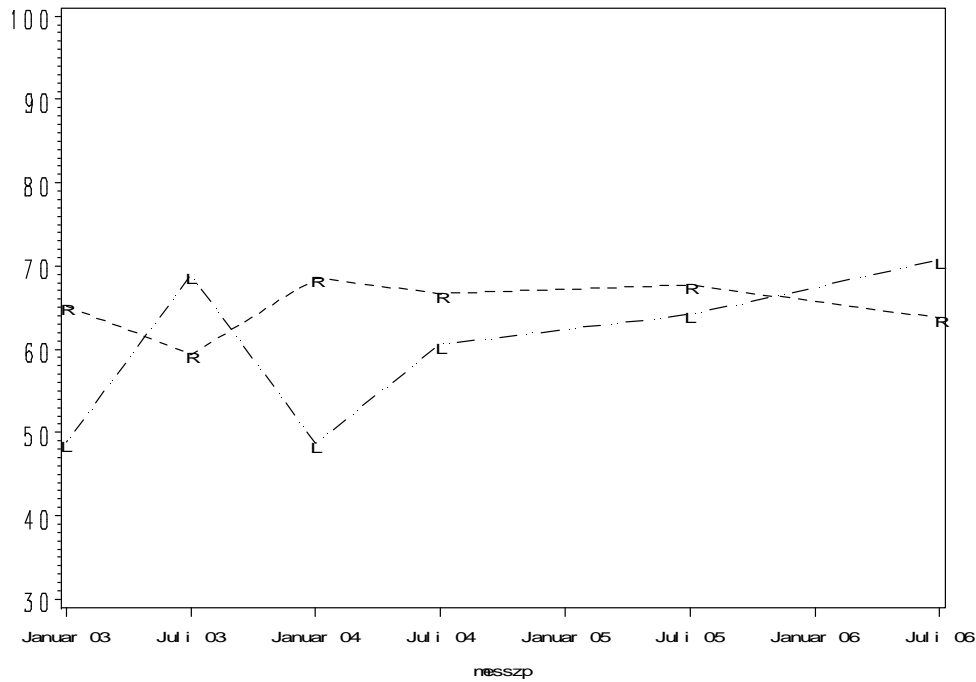


Abbildung 14: Konzeptspezifische Unterrichtsbeobachtung, Verhalten der Lehrkraft in Prozent des Optimum

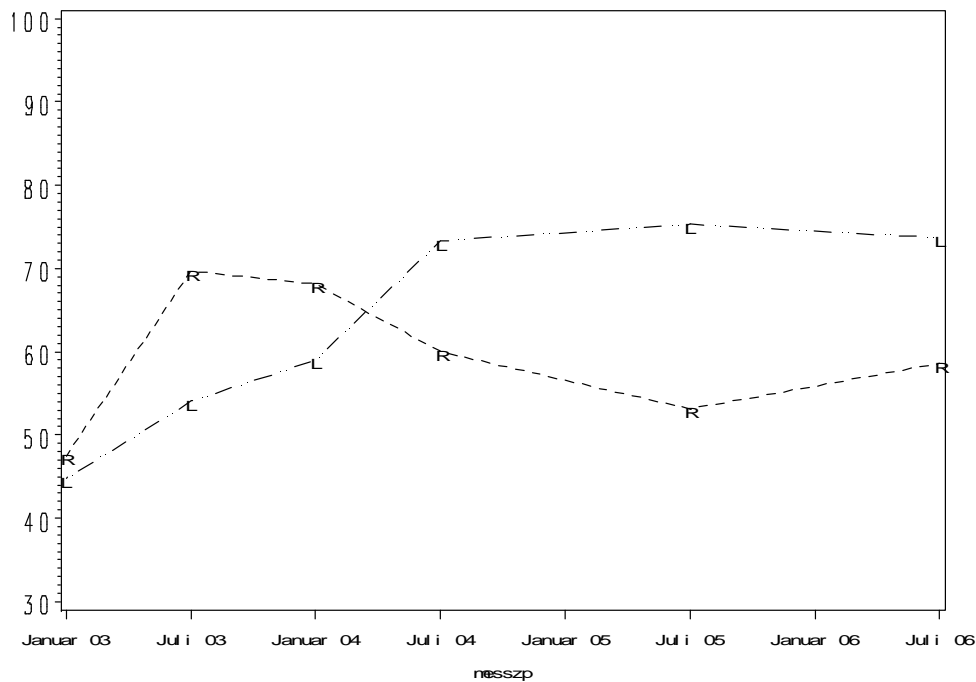


Abbildung 15: Konzeptspezifische Unterrichtsbeobachtung, Verhalten der Schüler in Prozent des Optimum

Die konzeptspezifische Unterrichtsbeobachtung wurde in der zweiten Projekthälfte noch zweimal, jeweils am Ende einer Klasse, durchgeführt. Zu diesem Zweck besuchten Fachkräfte des jeweiligen Konzepts die Klassen und schätzten ein, wie gut die Vorgaben des jeweiligen Konzepts erfüllt wurden. Die Werte wurden dann in zwei Kategorien, Verhalten der Lehrkraft und Verhalten der Schüler, zusammengefasst. Die Werte sind als Prozente dargestellt, wobei 100% erreicht würden, wenn auf allen Items optimale Werte erreicht würden.

Verhalten der Lehrkraft (Abbildung 14): Es findet auch hier eine Konzeptannäherung bei den Werten statt, am Ende weist Lollipop sogar etwas höhere Werte auf.

Verhalten der Schüler (Abbildung 15): Die Kreuzung der Kurven, die sich bereits im Zwischenbericht ansatzweise erkennen lässt, setzt sich fort. Das heißt, dass in den ersten zwei Jahren das Schülerverhalten in den RSW-Klassen konzeptkonformer war, während in der zweiten Projekthälfte die LP-Klassen bessere Werte erreichten.

### 3.9 Zeugnisnoten

Ende der vierten Klasse wurden die Zeugnisnoten in Deutsch und Mathematik erhoben, und zwar jeweils die Noten Ende der Klassenstufen 2, 3 und 4. Tabelle 20 zeigt die Kennwerte für die Gesamtstichprobe. Für die Berechnungen wurden die Klassenmittelwerte der Noten zugrunde gelegt, d.h. die Spalten *Minimum* und *Maximum* beziehen sich auch auf Klassenmittelwerte. Wie man an diesen beiden Kennwerten sieht, streuen die Klassenmittelwerte erheblich. Für Deutsch Ende der vierten Klasse hat die „beste“ Klasse einen Notenschnitt von 2,25 und die schlechteste Klasse einen Schnitt von nur 3,38. Deutsch- und Mathenoten unterscheiden sich im Niveau kaum. Wir sehen bei beiden den allgemeinen Trend von Klasse 2 zu Klasse 4, dass das Notenniveau etwas sinkt.

	Klasse	Mittelwert	STA	Minimum	Maximum
<b>Deutsch</b>	2	2,33	0,31	1,89	2,77
	3	2,63	0,26	2,13	2,92
	4	2,72	0,37	2,25	3,38
<b>Mathe</b>	2	2,33	0,30	1,88	2,68
	3	2,44	0,29	1,94	2,79
	4	2,62	0,41	2,00	3,31

Tabelle 20: Kennwerte der Zeugnisnoten auf Basis der Klassenmittelwerte

Es stellt sich die Frage, wie sich das unterschiedliche Klassenniveau erklärt. Eine naheliegende Hypothese ist, dass das Notenniveau mit dem Rechtschreib- und Leseniveau der Klasse korrespondiert. Hierzu wurden Korrelationen berechnet (s. Tabelle 21). Hierbei ist zunächst zu beachten, dass die Datenbasis sehr klein ist (n=10 Klassen) und damit Zufallsschwankungen sehr hoch ausfallen.

Es gelten folgende Erwartungen:

- Werte zwischen Maßen derselben Klassenstufe sollten höher ausfallen als zwischen Maßen unterschiedlicher Klassenstufen. Beispiel: Deutschnote Klasse 4 sollte höher mit Rechtschreibtest Klasse 4 korrelieren als mit Rechtschreibtest Klasse 3.

- Rechtschreib- und Lesetests sollten höher mit Deutsch- als mit Mathenoten korrelieren.
- Die Koeffizienten sollten negativ sein, da hohe Werte bei Noten schlechten Leistungen entsprechen, aber hohe Werte bei Tests guten Leistungen.

Die Koeffizienten erreichen zwar zum Teil hohe Werte, jedoch sind diese meist von der Richtung her unsinnig: der Wert von 0,68 für den Zusammenhang zwischen der Mathe-Note und dem Lesetest in Klasse 4 besagt, dass Klassen mit hohem Leseniveau im Durchschnitt schlechte Mathenoten aufweisen. Es tauchen zwar vereinzelt negative Werte auf, diese sind dann aber nicht Werten der gleichen Klassenstufe zugeordnet, z.B. gibt es eine Korrelation zwischen der Deutschnote in Klasse 4 und dem Rechtschreibtest in Klasse 3, aber nicht mit dem Rechtschreibtest Klasse 4. Zusammenfassend können wir sagen, dass es keine sinnvollen bzw. systematischen Zusammenhänge zwischen den Klassenniveaus im Rechtschreiben und Lesen und den Notenniveaus gibt.

		Rechtschreibtest			Lesetest		
		Klasse 2	Klasse 3	Klasse 4	Klasse 2	Klasse 3	Klasse 4
Note Deutsch	Kl. 2	0,39	0,11	0,17	0,32	0,48	0,31
	Kl. 3	0,14	-0,05	0,14	0,00	-0,04	0,40
	Kl.4	0,00	-0,41	-0,12	-0,21	-0,18	0,20
Note Mathe	Kl 2	0,13	0,04	0,38	-0,06	-0,02	0,60
	Kl. 3	0,37	-0,10	0,35	0,04	-0,10	0,65
	Kl. 4	0,51	0,02	0,37	0,17	0,02	0,68

Tabelle 21: Korrelationen von Zeugnisnoten und Rechtschreib- bzw. Lesetests auf Klassenebene

Es stellt sich dann also die Frage nach den individuellen Zusammenhängen zwischen Rechtschreibung und Noten (s. Tabelle 22). Die Basis dieser Berechnungen sind die jeweiligen Einzelwerte der Gesamtstichprobe (n=159).

		Rechtschreibtest			Lesetest		
		Klasse 2	Klasse 3	Klasse 4	Klasse 2	Klasse 3	Klasse 4
Note Deutsch	Kl. 2	-0,41	-0,45	-0,45	-0,19	-0,28	-0,28
	Kl. 3	-0,52	-0,58	-0,59	-0,29	-0,44	-0,37
	Kl.4	-0,58	-0,64	-0,64	-0,33	-0,42	-0,42
Note Mathe	Kl 2	-0,53	-0,48	-0,40	-0,29	-0,42	-0,24
	Kl. 3	-0,50	-0,50	-0,43	-0,31	-0,46	-0,31
	Kl. 4	-0,44	-0,45	-0,35	-0,28	-0,38	-0,23

Tabelle 22: Korrelationen von Zeugnisnoten und Rechtschreib- bzw. Lesetests auf Schülerebene

Alle Koeffizienten fallen erwartungskonform negativ aus. Die anderen beiden Erwartungen werden aber von den Daten nicht erfüllt. Zum einen sind die Koeffizienten gleicher Klassenstufe nicht konsistent höher als Koeffizienten ungleicher Klassenstufen, zum an-

deren sind die Zusammenhänge von Lesen und Rechtschreiben mit den Deutschnoten nicht durchgängig höher als mit den Mathenoten. Für den Zusammenhang Deutschnote – Rechtschreibtest Ende Klasse 4 findet sich aber immerhin ein Wert von 0,64. Dies ist angesichts der Tatsache, dass Zeugnisnoten recht grob gerastert sind und dort auch andere Aspekte der Leistung als nur die Rechtschreibung eingehen, kein schlechter Wert.

Die Korrekturphase zum Rechtschreibtest war u.a. durchgeführt worden, da dies eher der Unterrichtswirklichkeit entspricht als das Diktat ohne Korrekturmöglichkeit. Die Korrelation zwischen Deutschnote Ende Klasse 4 und dem korrigierten Rohwert Ende Klasse 4 fällt geringfügig höher aus ( $r=0,68$ ) als zum Standard-T-Wert (0,64).

Eine Systematik finden wir immerhin noch: der Zusammenhang der Noten ist zur Rechtschreibung konsistent höher als zum Lesen.

### 3.10 Lehrerfragebogen

Der bisherige Lehrerfragebogen wurde verkürzt weiterverwendet. Leider wird die Aussagekraft durch die Stichprobenverkleinerung sehr verringert. Exemplarisch soll daher hier eine Skala dargestellt werden. Seit Ende der Fortbildung wurden die Lehrkräfte der beiden Konzeptgruppen gefragt, zu wieviel Prozent sie die Inhalte der Fortbildung in die Praxis umsetzen.

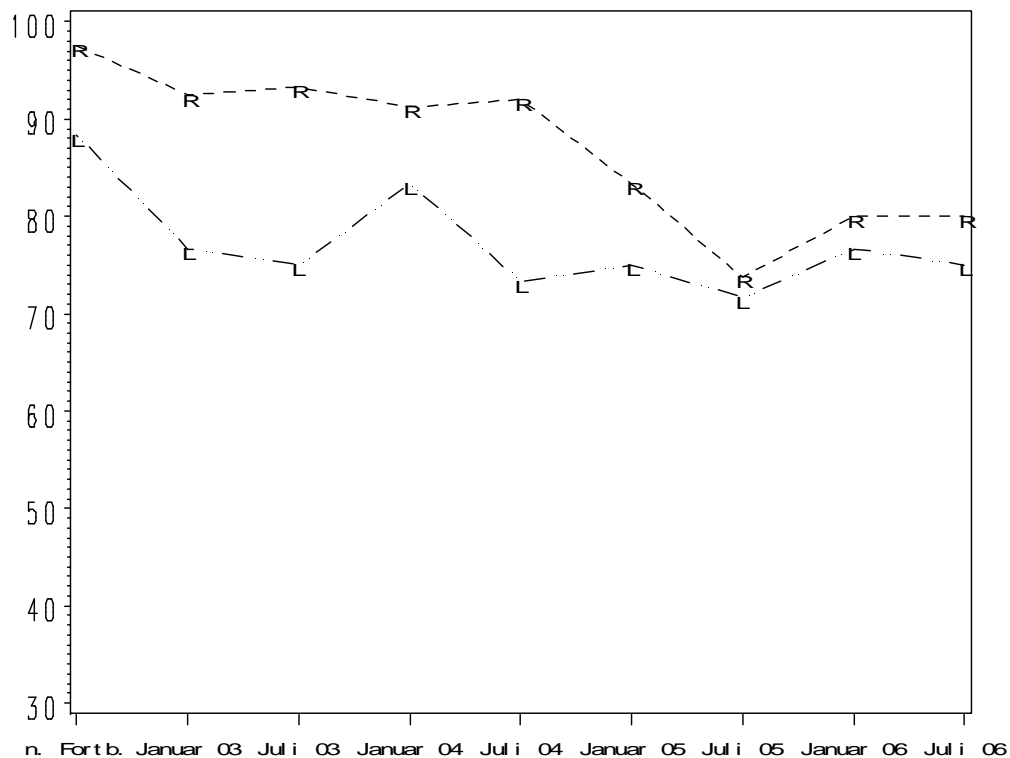


Abbildung 16: Eingeschätzte Konzeptumsetzung (in Prozent) bei Lollipop und RSW

Abbildung 16 zeigt die Verläufe für die beiden Konzeptgruppen. Zu beachten ist, dass sich die Kurve bei Lollipop nur aus den Angaben von drei Lehrkräften, bei RSW von vier Lehrkräften speist.

Wie bei den Schriftsprach-Testwerten findet auch hier eine Annäherung der Konzepte statt. Die höhere relative Konzepttreue in der RSW-Gruppe sinkt im Verlauf um ca. 10%

ab. Wenn man aber berücksichtigt, dass zum letzten Zeitpunkt die Fortbildung bereits über vier Jahre zurückliegt und außerdem in den letzten beiden Jahren keine systematische Betreuung mehr stattfand, dann können die Werte von ca. 80% doch als durchaus bemerkenswert eingestuft werden.

### 3.11 Elternfragebogen

Die erste Frage im Elternfragebogen galt der außerschulischen Förderung. Zunächst war anzukreuzen, ob das Kind außerschulische Förderung aktuell bekommt oder in den letzten zwei Jahren bekommen hat.

	<b>RSW</b>	<b>Lollipop</b>	<b>KG</b>
n	8	4	8
%	13,1%	8,7%	17,8%

Tabelle 23: Anzahl und Anteil von Kindern mit außerschulischer Förderung

Außerschulische Förderung wird am häufigsten in der Kontrollgruppe wahrgenommen, (s. Tabelle 23) gefolgt von der RSW. Am seltensten wird in der Lollipop-Gruppe außerschulisch gefördert. Da unter dem Begriff „außerschulische Förderung“ sehr verschiedene Dinge subsumiert sind, ist der Blick auf die Aufschlüsselung nach Art der Förderung sinnvoll (s. Tabelle 24).

	<b>RSW</b>	<b>Lollipop</b>	<b>KG</b>
Nachhilfe	2 (25%)	2 (50%)	3 (37,5%)
Legasthenietherapie	4 (50%)	0	5 (62,5%)
Lerntherapie	0	0	0
Sonstiges (z.B. Ergotherapie)	4 (50%)	4 (100%)	0

Tabelle 24: Außerschulische Förderung aufgeteilt nach Art der Förderung. Die Prozentzahlen beziehen sich nicht auf die Gesamtstichprobe, sondern nur auf die Bögen, bei denen außerschulische Förderung angekreuzt worden war. Da Mehrfachnennungen möglich waren, ergeben die Werte zusammen z.T. mehr als 100%

Da die Art der Förderung in einem eigenen Textfeld zu spezifizieren war, können wir die acht Fälle, bei denen „Sonstiges“ angekreuzt worden war, näher beschreiben. Bei sieben davon handelt es sich um Ergotherapie, der achte Fall scheint eine Art Lerntherapie oder Nachhilfe zu sein.

Da die Zahlen recht klein sind, wirken Unterschiede bei den Prozentwerten größer, als sie in Wirklichkeit sind. Orientierend wurden Signifikanztests (Fishers Exakter Test) durchgeführt. Lediglich für die Kategorie „Sonstiges“ gibt es eine signifikante Ungleichverteilung zwischen den Gruppen.

Es wurde auch abgefragt, ob Eltern zur „Selbsthilfe“ greifen und mit den Kindern – über die Hausaufgaben hinaus – Lesen bzw. Rechtschreiben üben (Tabelle 25 u. Tabelle 26).



<b>Schreiben üben</b>	<b>RSW</b>	<b>Lollipop</b>	<b>KG</b>	<b>Gesamt</b>
selten oder nie	63,9 %	53,2 %	43,2 %	54,6 %
bis 1 St. pro Woche	21,3 %	31,9 %	36,4 %	29,0 %
1 – 2 Stunden pro Woche	8,2 %	12,8 %	18,2 %	12,5 %
mehr als 2 Stunden pro Woche	6,6 %	2,1 %	2,3 %	4,0 %

Tabelle 25: Schreiben üben mit den Eltern über die Hausaufgaben hinaus

<b>Lesen üben</b>	<b>RSW</b>	<b>Lollipop</b>	<b>KG</b>	<b>Gesamt</b>
selten oder nie	55,7 %	46,8 %	51,1 %	51,6 %
bis 1 St. pro Woche	18,0 %	25,5 %	22,2 %	21,6 %
1 – 2 Stunden pro Woche	16,4 %	19,2 %	22,2 %	19,0 %
mehr als 2 Stunden pro Woche	9,8 %	8,5 %	4,4 %	7,8 %

Tabelle 26: Lesen üben mit den Eltern über die Hausaufgaben hinaus

Für beide Fragen gilt, dass jeweils etwa die Hälfte der Eltern nicht mit ihren Kindern über die Hausaufgaben hinaus üben. Es gibt nur geringe Verteilungsunterschiede zwischen den Gruppen, mit einer leichten Tendenz, dass in der RSW-Gruppe jeweils etwas weniger geübt wird als in den beiden anderen Gruppen.

<b>Lernprogramme PC</b>	<b>RSW</b>	<b>Lollipop</b>	<b>KG</b>	<b>Gesamt</b>
selten oder nie	74,2 %	61,7 %	60,0 %	66,2 %
bis 1 St. pro Woche	11,3 %	21,3 %	24,4 %	18,2 %
1 – 2 Stunden pro Woche	9,7 %	12,8 %	13,3 %	11,7 %
mehr als 2 Stunden pro Woche	4,8 %	4,3 %	2,2 %	3,9 %

Tabelle 27: Nutzung von Lernprogrammen am PC

<b>Computerspiele</b>	<b>RSW</b>	<b>Lollipop</b>	<b>KG</b>	<b>Gesamt</b>
weniger als eine halbe Stunde pro Tag	65,6 %	57,8 %	61,4 %	62,0 %
bis 1 Stunde pro Tag	27,9 %	31,1 %	34,1 %	30,7 %
Mehr als 2 Stunden pro Tag	6,6 %	11,1 %	4,6 %	7,3 %

Tabelle 28: Nutzung von PC-Spielen und Gameboy etc.

Zwei weitere Fragen galten der Computernutzung zu Hause. Zunächst wurde gefragt, wie lange sich die Kinder mit Lernprogrammen zum Lesen bzw. Schreiben beschäftigen (s. Tabelle 27). Etwa ein Drittel der Stichprobe nutzt demnach überhaupt Lernprogramme. In der RSW-Gruppe sind diese Hilfsmittel etwas weniger gebräuchlich als in den beiden anderen Gruppen. Bei der Frage nach Nutzung von Computerspielen und Spielekonsolen

(Tabelle 28) bezeichneten immerhin gut sieben Prozent der Eltern ihre Kinder als Vielspieler (mehr als zwei Stunden pro Tag).

<b>Fernsehen</b>	<b>RSW</b>	<b>Lollipop</b>	<b>KG</b>	<b>Gesamt</b>
weniger als eine halbe Stunde pro Tag	11,3 %	14,9 %	4,4 %	10,4 %
bis 1 Stunde pro Tag	77,4 %	61,2 %	68,9 %	70,1 %
Mehr als 2 Stunden pro Tag	11,3 %	23,4 %	26,7 %	19,5 %

Tabelle 29: Fernsehkonsum pro Tag

Der Fernsehkonsum (Tabelle 29) schlägt demgegenüber mit höheren Nutzungszeiten zu Buche, insgesamt knapp 20% der Kinder sehen pro Tag mehr als zwei Stunden fern. In der RSW-Gruppe war dieser Anteil relativ niedrig (11,3%), in der Kontrollgruppe relativ hoch (26,7%).

Nun stellt sich die Frage, ob diese Verhaltensweisen in einem Zusammenhang mit den Rechtschreib- und Leseleistungen stehen. Dazu wurden Korrelationen berechnet (Tabelle 30). Lediglich für das Üben von Rechtschreiben und Lesen ergibt sich ein mittlerer Zusammenhang mit der Rechtschreibung: Eltern von Kindern mit schlechter Rechtschreibung üben relativ viel. Die Nutzung von Lernprogrammen oder Spielen sowie der Fernsehkonsum haben dagegen keine Auswirkungen auf die Schriftsprachleistungen. Bereits im Zwischenbericht fanden wir keine Auswirkungen des TV-Konsums.

	<b>Rechtschreibtest</b>	<b>Lesetest</b>
Üben Rechtschreiben	-0,33	-0,14
Üben Lesen	-0,33	-0,21
Computer Lernprogramme	0,05	0,16
Computer Spiele	-0,04	0,14
Fernsehkonsum	-0,04	0,06

Tabelle 30: Korrelationen der Fragebogenitems mit Rechtschreib- und Lesetest Ende Klasse 4

Abschließend wurde nach der Bewertung des Projekts gefragt (Tabelle 31). Der genaue Wortlaut war: „Dass die Klasse meines Kindes an einem Modellprojekt zur Verbesserung des Schriftsprachwerberbs teilnimmt, finde ich ...“.

	<b>sehr gut</b>	<b>gut</b>	<b>ist mir egal</b>	<b>schlecht</b>	<b>sehr schlecht</b>
<b>KG</b>	51,1 %	48,9 %	-	-	-
<b>LP</b>	48,9 %	51,1 %	-	-	-
<b>RSW</b>	39,3 %	45,9 %	1,6 %	8,2 %	4,9 %
<b>Gesamt</b>	45,8 %	48,4 %	0,7 %	3,3 %	2,0 %

Tabelle 31: Bewertung des Projekts durch die Eltern

Außerdem wurden die Eltern mittels offenem Textfeld auch gefragt, ob die Bekanntgabe der Zwischenergebnisse Ende 2004 Konsequenzen hatte. Speziell wurde gefragt, ob die Eltern z.B. Nachhilfe organisiert hätten, einen Elternabend gefordert oder Ähnliches. (Die folgenden Angaben beziehen sich auf alle abgegebenen Elternbögen, also auch auf solche Fälle, die nicht in der Auswertungsstichprobe sind.)

Aus der RSW-Gruppe wurde die Frage 16 mal beantwortet („nein“ wird hier natürlich nicht mitgezählt), aus der LP-Gruppe 4 mal und aus der KG 3 mal.

Inhaltlich passierte in der RSW-Gruppe Folgendes (es gab Mehrfachnennungen): 5 mal wurde „Elternabend“ angegeben, 4 mal Gespräche mit der Lehrkraft, 2 mal wurde eine Diagnostik der Lese-Rechtschreibfertigkeiten eingeleitet und 1 mal wurde eine Ergotherapie begonnen. Diese 12 Nennungen zielen also nicht direkt auf eine Verbesserung der Leistungen.

3 mal wurde Nachhilfe begonnen, 4 mal übten die Eltern selbst mehr mit dem Kind und 1 mal wurde eine Legastheniebehandlung begonnen. Wir sehen also bei insgesamt 8 Kindern, dass die Eltern Maßnahmen eingeleitet haben, die direkt auf die Verbesserung der Schriftsprachleistungen abzielen.

Bei Lollipop wurde 2 mal Üben und 1 mal Nachhilfe angegeben. 1 mal schrieben die Eltern, dass sie das Lollipop-Programm einem noch nicht eingeschulerten Geschwister zum Spielen gegeben haben. In dieser Gruppe sehen wir also 3 direkte Fördermaßnahmen. In der KG wurde je 1 mal Nachhilfe, Legastheniebehandlung und vermehrtes Üben angegeben.

Wir stellen also fest, dass in der RSW-Gruppe öfter Maßnahmen zur direkten Schriftsprachverbesserung eingeleitet wurden als in den beiden anderen Gruppen. Die absoluten Zahlen sind allerdings recht klein und es scheint sehr unwahrscheinlich, dass dadurch die Gruppenmittelwerte systematisch beeinflusst worden sein könnten. Trotzdem soll an dieser Stelle exemplarisch ein Blick auf die Rechtschreibleistungen dieser Gruppe geworfen werden. Von den 8 Fällen mit direkten Maßnahmen sind 5 in der Auswertungsstichprobe enthalten. Wenn man die RSW-Gruppe entsprechend aufteilt, d.h., die 5 Kinder mit direkten Maßnahmen mit dem Rest der Gruppe vergleicht, so ergibt sich folgendes Bild: die 5 Kinder mit Nachhilfe bzw. Legastheniebehandlung erreichten am Ende der zweiten Klasse einen Rechtschreib-T-Wert von 37,2 und am Ende der vierten Klasse einen T-Wert von 35,6. Der Rest der RSW-Gruppe erreichte am Ende der zweiten Klasse einen Rechtschreib-T-Wert von 44,7 und am Ende der vierten Klasse einen T-Wert von 48,9.

Die Kinder, deren Eltern direkte Maßnahmen eingeleitet hatten, erreichten also tatsächlich deutlich unterdurchschnittliche Leistungen. Allerdings sind die T-Werte im Laufe der Jahrgangsstufen 3 und 4 nicht besser, sondern sogar noch etwas schlechter geworden. Der Rest der Gruppe hatte eine positive Entwicklung und konnte die Rechtschreibleistung steigern. Dies entkräftet das mögliche Argument, dass die vermehrten Anstrengungen seitens der Eltern in der RSW-Gruppe für die positive Gruppenentwicklung in den Klassen 3 und 4 (mit-)verantwortlich sein könnten.

## 4 Zusammenfassung

### Vergleich der Unterrichtskonzepte

Zentrale Fragestellung des Modells Schriftsprachmoderatoren war, ob durch spezifisch veränderten Unterricht die Lese- und Rechtschreibleistung, insbesondere bei den Lese- und Rechtschreibschwachen, verbessert werden kann. Zur Beantwortung dieser Frage wurden in dieser Längsschnittstudie von der ersten bis zur vierten Klasse drei Gruppen (Rechtschreibwerkstatt, Lollipop, Kontrollgruppe), die sich hinsichtlich der Methoden zur Vermittlung der Schriftsprachkompetenz unterscheiden, verglichen.

An der wissenschaftlichen Begleitung des Modellprojekts nahmen zunächst insgesamt 305 Kinder und 25 Lehrkräfte der Primarstufe, sowie 149 Kinder und sechs Lehrkräfte der Sekundarstufe aus den Schulamtsbereichen Schwalm-Eder-Kreis und Waldeck-Frankenberg teil. Nach zwei Jahren wurde der Teil des Gesamtprojekts, der sich mit den Schriftsprachleistungen in der Grundschule befasst, fortgeführt. In die Endauswertung gingen 10 Klassen mit zusammen 159 Kindern ein.

Aus der Forschungsliteratur wurden relevante Prädiktoren für den Schriftspracherwerb ermittelt und in diesem Projekt erfasst. Der mögliche Einfluss dieser Faktoren auf die Hauptfragestellung des Projektes wurde untersucht, um zu vermeiden, dass die vorhandenen Unterschiede in der Schriftsprachleistung nicht eventuell falsch als Konzeptunterschiede interpretiert werden.

Die Hauptfragestellung, ob sich die drei Gruppen, Rechtschreibwerkstatt (RSW), Lollipop (LP) und Kontrollgruppe (KG) am Ende der vierten Klasse hinsichtlich Rechtschreiben und Lesen voneinander unterscheiden, wurde für vier verschiedene abhängige Variablen geprüft: Rechtschreibung (T-Werte, und Anteil von Rechtschreibschwachen in Prozent) und Lesefähigkeit (T-Werte und Anteil von Leseschwachen in Prozent)<sup>4</sup>.

Nachdem die Gruppenunterschiede nach zwei Jahren beeindruckend groß waren, gab es bei allen relevanten Variablen nach vier Jahren eine Annäherung der Gruppen. Bezogen auf die deutschlandweiten Normen des Rechtschreibtests DERET 3-4 lagen die Gruppenmittelwerte jeweils knapp unter dem Erwartungswert 50, wobei die Lollipop-Kinder die höchsten und die Kontrollkinder die geringsten Leistungen erzielten. Bei den Anteilen Rechtschreibschwacher liegt Lollipop zwar wieder vorn, erreicht aber nur noch den Erwartungswert von 10%. Der Anteil Rechtschreibschwacher ist hier also genau so hoch, wie aufgrund der deutschlandweiten Normen zu erwarten.

Ergänzend zum Vergleich mit unserer Kontrollgruppe wurde auch der Rechtschreibtest DoSE (heißt jetzt: gut|schrift) durchgeführt, der zur Ermittlung der Schriftsprachkompetenz beim Ländervergleich der IGLU-Untersuchung benutzt worden war. Dies erlaubt uns, die Werte unserer Gruppen direkt mit den hessischen Normwerten von DoSE zu vergleichen. Die auf die hessischen Normen bezogenen Werte fallen erwartungsgemäß höher aus, da Hessen beim Ländervergleich im unteren Mittelfeld rangierte. Die Gruppenwerte streuen um den Erwartungswert 50. Bei den Anteilen Rechtschreibschwacher erreicht Lollipop mit 5,9% niedrigere Werte, als aufgrund der Normen zu erwarten (10%). Der Anteil Rechtschreibschwacher in der RSW-Gruppe entspricht fast genau dem Erwartungswert, in der KG gibt es etwas mehr Rechtschreibschwache, als zu erwarten.

---

<sup>4</sup> Rechtschreibschwäche und Leseschwäche wurde angenommen, wenn ein individueller Testwert kleiner oder gleich T=37 ist. Das entspricht einem Prozentrang von 10.

Die Lesetest-Werte fallen sämtlich ziemlich hoch aus. Wir müssen uns hier fragen, ob der Test WLLP in den Klassenstufen 3 und insbesondere 4 noch zuverlässig misst, oder ob die Kinder hier etwas unterfordert sind.

Die **statistische Prüfung** bei den vier Zielvariablen (T-Werte im Rechtschreib- und Lesetest, sowie Anteile Rechtschreibschwacher und Leseschwacher) unter Berücksichtigung aller relevanten Prädiktoren ergibt keinerlei signifikante Gruppenunterschiede mehr.

Neben der Betrachtung der Schriftsprachleistung am Ende der vierten Klasse ist der **Verlauf der Entwicklung im Lesen und Rechtschreiben** von großem Interesse. Es liegen die kompletten halbjährlichen Rechtschreib- und Lesemessungen über vier Jahre sowie die Erfassung der Vorläuferfertigkeiten bei Einschulung vor.

Nach zwei Jahren hatten sich die Gruppen bei den mittleren Rechtschreib-Werten deutlich auseinanderentwickelt, in den Klassenstufen drei und vier fand eine Annäherung der Gruppen statt, die insbesondere auf der Verbesserung der Rechtschreibleistungen in der RSW-Gruppe beruht.

Die LP-Gruppe bewegte sich bis zum Ende der zweiten Klasse klar nach oben<sup>5</sup>, konnte dieses Niveau aber bis zum Ende der vierten Klasse nur knapp halten. Die RSW-Gruppe zeigt nach anfänglichem Absinken der Werte zunächst nur eine minimale Tendenz nach oben; erst in der zweiten Hälfte der dritten Klasse gibt es eine Verbesserung des Gruppenmittelwerts. Die Kontrollgruppe bewegt sich im Mittel nur leicht nach oben und erreicht den relativ schlechtesten Platz.

Bei den Anteilen von rechtschreibschwachen Kindern sieht es erwartungsgemäß ganz ähnlich aus. Die Entwicklung erscheint allerdings dramatischer als bei den T-Werten; in der RSW-Gruppe und KG steigt der Anteil von Rechtschreibschwachen zwischenzeitlich (Mitte Klasse 2) auf über 20%. Die Lollipop-Gruppe hatte hier konstant gute Werte unterhalb des Erwartungswertes von 10, bei der letzten Messung steigt der Anteil Rechtschreibschwacher aber auf ca. 10%.

Die Entwicklung der Lesetest-Werte zeigt alle Gruppen viel näher beieinander. Bei den Leseschwachen erreicht Lollipop konstant sehr gute Werte, während die beiden anderen Gruppen anfangs eher ungünstige Werte aufweisen und erst im Verlauf bis zum Ende der vierten Klasse zu sehr guten Werten kommen.

Aus organisatorischen Gründen war das Projekt zunächst auf zwei Jahre angelegt. Da wir die Beteiligten u.a. dadurch zur Mitarbeit gewonnen hatten, dass wir ihnen die Rückmeldung detaillierter Ergebnisse nach zwei Jahren in Aussicht stellten, ergab sich mit der Verlängerung des Projekts auf vier Jahre Laufzeit ein mögliches Problem. Es ist nicht von vornherein auszuschließen, dass die Rückmeldung der Gruppenergebnisse zu unterschiedlichen Reaktionen führte, da die Ergebnisse eben für die Gruppen sehr unterschiedlich ausfielen. Speziell die relative Verbesserung der Rechtschreibleistungen in der RSW-Gruppe im dritten Schuljahr könnte von kritischen Kommentatoren mindestens teilweise der Überraschung und den möglicherweise daraus resultierenden Aktivitäten zur Verbesserung der Schriftsprachleistungen zugerechnet werden. Wir sind und waren uns dieser Problematik bewusst; letztendlich war diese Situation aber unvermeidlich. Zum einen wäre eine vierjährige Projektzeit finanziell und organisatorisch nicht von Anfang an leist- und planbar gewesen, zum anderen waren wir nach zwei Jahren mit den Beteiligten im Wort, ihnen Rückmeldungen über die Ergebnisse zu geben.

---

<sup>5</sup> Da die Werte jeweils die Position relativ zu den Klassennormen des Tests angeben, heißt eine Bewegung nach oben also, dass ein überproportionaler Wissenszuwachs stattfindet

Wir können daher nur retrospektiv so gut wie möglich analysieren, ob die Mitteilung der Zwischenergebnisse unterschiedliche Auswirkungen auf die Gruppen hatte. Mit den Lehrkräften wurde im November 2006 je ein semi-strukturiertes Interview zum Projektende durchgeführt, bei dem die Konsequenzen der Zwischenergebnisse für Lehrkraft und Schule differenziert abgefragt wurden. Die Darstellung dieser Erkenntnisse erfolgt an anderem Orte. Neben vermehrten Anstrengungen von Lehrkräften und Schulen ist es auch denkbar, dass einzelne Eltern auf Ergebnisse ihres Kindes mit außerschulischen oder häuslichen Aktivitäten reagieren. Wir haben dies in dem letzten Elternbogen erfasst. Zwar findet sich in der RSW-Gruppe ein relativ höherer Anteil von Kindern, bei denen in den Klassenstufen 3 und 4 Maßnahmen zur Verbesserung der Schriftsprachleistungen (wie z.B. Nachhilfe oder vermehrtes Üben mit den Eltern) initiiert wurden. Die absoluten Zahlen sind aber gering (8 in RSW und 3 in LP) und zudem konnten wir zeigen, dass die Sub-Gruppe der RSW-Kinder mit direkten Maßnahmen zur Verbesserung der Schriftsprachleistungen den Verlauf der Rechtschreib-Werte in der RSW-Gruppe nicht etwa verbessert, sondern sogar geringfügig verschlechtert. Damit können wir darlegen, dass es seitens der Eltern keine effektiven Maßnahmen zur Verbesserung der Schriftsprachleistungen gegeben hat, die sich positiv auf den Verlauf in der RSW-Gruppe ausgewirkt haben.

Eine weitere Fragestellung des Projektes bezieht sich auf die Entwicklung von **Kindern mit einem Risiko zur Ausbildung von Lese-Rechtschreibschwierigkeiten**. Dazu wurden drei Gruppen gebildet: Kinder mit niedrigem, Kinder mit einem mittleren und eine Gruppe Kinder mit einem hohen Risiko. Kinder mit **hohem Risiko** erzielten erwartungsgemäß die schlechtesten Rechtschreib-Werte, Kinder mit **geringem Risiko** die besten. Nachdem wir dies bereits im Zwischenbericht feststellen konnten, so überrascht uns nun doch, dass die Effekte sich sogar über alle vier Jahre fortsetzen. Im Detail: der stärkste Einzelprädiktor ist die Buchstabenkenntnis bei Einschulung. Wenn man bei Einschulung erfasst, wieviele Buchstaben die Kinder bereits benennen können, so hat diese Variable noch nach vier Schuljahren einen deutlichen Effekt auf die Rechtschreibleistung – die Korrelation beträgt 0,36.

Wir können auch festhalten, dass die Spreizung der Gruppen bei der ersten Messung nach einem halben Jahr Unterrichtung über den gesamten Zeitraum erhalten bleibt. D. h., **wer zu Beginn schlecht schreibt, der hat auch am Ende der vierten Klasse die größten Probleme bei der Rechtschreibung**. Die Verläufe für das Lesen sehen ganz ähnlich aus; auch hier haben die Kinder mit hohem Risiko die schlechtesten Testwerte.

Die Einführung einer **Korrekturphase** nach regulärer Durchführung der Rechtschreibtests DRT 3 und DERET 3-4 diente der Erhöhung der „ökologischen Validität“ der Rechtschreiberfassung. Korrekturmöglichkeiten nach freien Texten oder Diktaten gehören in Grundschulen zum Standard, dabei sind auch Hilfsmittel wie Wörterbücher erlaubt. Zudem ist die Vermittlung von Korrekturstrategien Teil der von uns untersuchten Konzepte. Parallel zur Ermittlung des regulären Test-T-Werts anhand der unkorrigierten Testhefte verglichen wir die unkorrigierten mit den korrigierten Rohwerten. Die meisten Kinder (70 – 80%) konnten ihren Rohwert durch die Korrekturphase verbessern. Das Ausmaß der Verbesserung erreichte dabei vereinzelt bis zu sieben Rohwertpunkte. Im Mittel wurden Verbesserungen um ca. 1,5 bis 1,7 Punkte erreicht, wobei im Gruppenvergleich RSW leicht – bei geringen Effekten - im Vorteil war. Wir konnten also zeigen, dass die Korrekturphasen tatsächlich zu leichten, konzeptunspezifischen Verbesserungen der Testleistungen führte. Die Bedeutung dieser Verbesserung hängt vom Niveau der Leistung ab: im mittleren Leistungsbereich führt eine Rohwertverbesserung von 2 etwa zu einem um 2 T-Punkte besseren Normwert, am unteren Rand der Verteilung kann die Verbesserung des Normwertes aber deutlich höher sein.

## **Bedeutung von Einflussfaktoren auf die Schriftsprachentwicklung**

Unabhängig von der Bedeutung der Prädiktoren auf den Konzeptvergleich wurden der **Einfluss der Prädiktoren** auf die Schriftsprachleistungen (Lesen und Rechtschreiben) untersucht. Die Ergebnisse lassen sich wie folgt zusammenfassen:

- Das Vorwissen der Kinder (gemessen bei Einschulung) hat auch Ende der vierten Klasse noch einen deutlichen Zusammenhang zu den Schriftsprachleistungen. Die Korrelation zwischen Vorwissen und Rechtschreibung bzw. Lesen Ende der vierten Klasse beträgt jeweils ca. 0,4.
- Einsprachig deutsche Kinder erzielen Ende der vierten Klasse nur noch marginal bessere Rechtschreib- und Lesetestwerte als zweisprachige Kinder.
- Der Geschlechtseffekt hat sich für die Rechtschreibung gegenüber dem Zwischenbericht vergrößert. Mädchen schreiben deutlich besser als Jungen (mittelgroßer Effekt), beim Lesen finden wir jedoch nur einen unbedeutenden Effekt.
- Das Ausmaß des TV-Konsums spielt in unserer Stichprobe auch Ende der vierten Klasse keine Rolle. Dieses Ergebnis bestätigt nicht die Befunde der Studie von Ennemoser et al. (2003) die zeigten, dass sich der längere Fernsehkonsum leistungsmindernd auf die Leseleistung von Drittklässlern auswirkt.

**Emotionale und soziale Schulerfahrungen** wurden mit dem Fragebogen FEESS erfasst, der von den Kindern ausgefüllt wurde. Die mittleren T-Werte aller sieben Skalen liegen recht eng um den Erwartungswert 50. In der RSW-Gruppe werden durchgängig die höchsten Werte erzielt, die Effekte im Vergleich zu den anderen Gruppen sind aber nur klein oder unbedeutend. Einen Zusammenhang zu den Schriftsprachleistungen finden wir in erster Linie mit der Skala „Selbstkonzept der Schulfähigkeit“, die 0,35 - 0,38 mit Rechtschreibung bzw. Lesen korreliert.

Wir haben drei Bausteine verwendet, um für die Schriftsprachleistungen relevante Informationen aus der **Beobachtung von Unterrichtssituationen** zu gewinnen. Dies waren die Erfassung der Aufmerksamkeit der Kinder im Unterricht durch ein niedrig-inferentes Verfahren (MAI), die Erfassung komplexer Zusammenhänge (Unterrichtsmangement der Lehrkraft und Unterrichtsklima) durch einen hoch-inferenten Beobachtungsbogen sowie die Erfassung spezifischer Aspekte der Konzeptumsetzung in den beiden Konzeptgruppen. Die ersten beiden Bausteine sind nicht konzeptspezifisch und wurden ausgewählt, weil sie in der Literatur als bedeutsame Faktoren für gute Schriftsprachleistungen beschrieben worden sind.

*Entgegen unseren Erwartungen ergaben sich für alle drei Bausteine enttäuschende Ergebnisse:* Die Zusammenhänge mit Schriftsprachleistungen sind nur sehr niedrig, z.T. sogar erwartungswidrig. Dieses Resultat fanden wir bereits nach zwei Jahren und können es nun nach vier Jahren bestätigen. Wir finden keinen einleuchtenden Grund dafür.

Bei den erfassten **Zeugnisnoten** in Deutsch und Mathematik fanden wir zunächst Niveauunterschiede zwischen den einzelnen Klassen, d.h., dass der Noten-Mittelwert in den Klassen jeweils deutlich unterschiedlich ausfällt. Diese Niveauunterschiede korrespondieren nicht mit den Niveauunterschieden in der Schriftsprachleistung. Wir finden aber deutliche Zusammenhänge zwischen individuellen Noten und Schriftsprachleistungen, d.h., dass ein Kind mit guten Schriftsprachleistungen auch gute Deutschnoten bekommt.

In den meisten **Familien** (84% für Schreiben und 77% für Lesen) wird weniger als eine Stunde pro Woche über die Hausaufgaben hinaus geübt. Die Konzeptgruppen unterschei-

den sich nicht nennenswert in diesen Verteilungen. Diese Zahlen entsprechen denen nach zwei Jahren. Es gibt einen Zusammenhang ( $r=-0,33$ ) zwischen Ausmaß des Übens und Schriftsprachleistungen: Eltern von Kindern mit geringen Schriftsprachleistungen üben mehr mit ihren Kindern.

Nach Eltern-Angaben werden zwischen 13 (RSW) und 18 (KG) Prozent der Kinder außerschulisch gefördert, wobei die Art der Förderung (z.B. Nachhilfe, Legastheniebehandlung, Ergotherapie) aber variiert. Hinweise auf bedeutsame Gruppenunterschiede fanden wir dennoch nicht.

Explizit wurde nach Konsequenzen aus den Zwischenergebnissen gefragt. In der RSW-Gruppe gab es etwas öfter Maßnahmen zur direkten Schriftsprachverbesserung (wie z.B. Nachhilfe), die Untergruppe dieser Kinder konnte ihre Leistungen bis Ende vierter Klasse aber keineswegs verbessern. Diese Art der Elterninitiative blieb also ohne positive Folgen auf die Schriftsprachleistungen.

Erneut erfasst wurde der Fernsehkonsum der Kinder und außerdem erstmalig die Nutzung von Computern für Spiele bzw. Lernprogramme. Für diese Maße fanden wir keine Zusammenhänge mit den Schriftsprachleistungen.

Bei der **Bewertung des Projekts** finden wir negative Einschätzungen – wie nach zwei Jahren – nur in der RSW-Gruppe. 13% dieser Eltern fanden es „schlecht“ oder „sehr schlecht“, dass die Klasse des Kindes am Modellprojekt teilnahm. In LP und KG gibt es ausschließlich „gut“ oder „sehr gut“ als Antwort auf die Frage.

## Fazit

Die Bedeutung zweier Unterrichtskonzepte für den Schriftspracherwerb wurde unter Berücksichtigung zentraler, aus der Forschungsliteratur abgeleiteter Faktoren untersucht. Zu diesen zählen *Unterrichtsfaktoren* (Klassenmanagement, Unterrichtsmenge, Aufmerksamkeit der Kinder), *Klassenfaktoren* (Klassengröße, Ausländeranteil), *Individuumsfaktoren* (Vor- und Sprachkenntnisse des Kindes, Intelligenz, Verhalten, Fernsehen), und *familiäre Faktoren* (Bildungsstand der Eltern, Geburtsrang, häusliches Lernumfeld). Durch das gewählte Untersuchungsdesign und die statistische Auswertung wurden diese Einflussfaktoren für den Gruppenvergleich adäquat berücksichtigt.

Nachdem die Gruppenunterschiede bei den Schriftsprachleistungen und insbesondere bei der Rechtschreibung nach zwei Jahren noch sehr groß waren, haben sich die Gruppen nach vier Jahren deutlich angenähert und die Unterschiede sind nicht mehr statistisch signifikant. Dies beruht in erster Linie auf der Verbesserung der Schriftsprachleistungen in der RSW-Gruppe in den Jahrgangsstufen 3 und 4, während die LP-Gruppe ihr gutes Niveau nicht ganz halten konnte. Die Konzeptgruppen unterscheiden sich also am Ende der Grundschulzeit nicht mehr sehr hinsichtlich der Schriftsprachleistungen. Wir können aber festhalten, dass Lollipop deutlich früher als RSW ein gutes Schriftsprachniveau erreicht hat, was durch den sehr deutlichen Gruppenunterschied am Ende der zweiten Klasse dokumentiert wird.

Neben dieser Hauptfragestellung wurden noch weitere Fragen untersucht. Ein wichtiger Befund ist, dass Risikokinder (gemessenes Risiko bei Einschulung) auch nach vier Jahren noch schlechtere Schriftsprachleistungen aufweisen als Nicht-Risikokinder. Schlechte Startbedingungen sind also ein hartnäckiges Handikap; hier bieten sich Ansatzpunkte für weitergehende Forschung, nämlich inwieweit gezielte Interventionen an dieser Stelle die Bedeutung des Risikos für die spätere Laufbahn mindern können.



Mit dem Fragebogen FEESSE wurde die persönliche Erlebniswelt der Kinder im Unterricht in den Blick genommen. Wir fanden hier leichte Vorteile für RSW, die aber von der Ausprägung her unbedeutend waren. Die Konzepte unterscheiden sich hierin also nicht nennenswert.

Wir blicken nun auf eine Gesamt-Projektlaufzeit von fast fünf Jahren zurück. Eine der unseren ähnliche Fragestellung wurde bisher in Deutschland nicht mit vergleichbarem Aufwand untersucht. Dem entsprechend waren viele Hürden zu nehmen und Hartnäckigkeit bei der Durchführung unserer Studie war gefragt. Es wäre wünschenswert, wenn durch weitere Projekte dieser Art die empirische Basis für schulpolitische Entscheidungen erweitert würde.

## 5 Literatur

- Birkel, P.: Weingartener Grundwortschatz-Rechtschreib-Test für zweite und dritte Klassen (WRT2+). Hogrefe, Göttingen 1994.
- Döpfner, M., Lehmkuhl, G.: DISYPS-KJ. Diagnostik-System für psychische Störungen im Kindes- und Jugendalter nach ICD-10 und DSM IV. Verlag Hans Huber, Bern 2000.
- Ennemoser, M., Schiffer, K., Reinsch, C., & Schneider, W.: Fernsehkonsum und die Entwicklung von Sprach- und Lesekompetenzen im frühen Grundschulalter. Zeitschrift fuer Entwicklungspsychologie und Pädagogische Psychologie 2003, 35(1), 12-26.
- Grund, M., Haug, G., Naumann, C.L.: DRT 4. Diagnostischer Rechtschreibtest für 4. Klassen. Hogrefe, Göttingen 2003.
- Löffler, I., Meyer-Schepers, U.: DoSE – Dortmunder-Schriftkompetenz-Ermittlung. Unveröffentlichtes Manuskript, 2001.
- Müller, R.: DRT 3. Diagnostischer Rechtschreibtest für 3. Klassen. Hogrefe, Göttingen 2003.
- Rauer, W., Schuck, K.-D.: FEESSE 3-4. Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern dritter und vierter Klassen. Beltz, Göttingen 2003.
- Stock, C.: DERET 3-4. Deutscher Rechtschreibtest für das dritte und vierte Schuljahr. Hogrefe, Göttingen, in Vorbereitung.

## 6 Anhang: Elternbogen Ende Klasse 4

1. Erhält Ihr Kind das am Modellprojekt Schriftsprachmoderatoren (MSM) teilnimmt aktuell (oder erhielt es in den letzten zwei Jahren) außerschulische Förderung, z.B. Nachhilfe, Ergotherapie oder Lese-Rechtschreibförderung bzw. Legastheniebehandlung außerhalb der Schule?.....  
.....ja  nein

Die Fragen **in diesem Kasten** brauchen Sie nur zu bearbeiten, wenn Ihr Kind außerschulische Förderung bekommt oder bekommen hat.

Seit wann bekommt bzw. wie lange bekam Ihr Kind außerschulische Förderung?

Wieviele Stunden pro Woche umfasst(e) die Förderung bzw. Nachhilfe? \_\_\_\_

Mein Kind bekommt / bekam . . .

Nachhilfe .....

Lese-Rechtschreibförderung oder Legasthenietherapie .....

Lerntherapie .....

Sonstiges (z.B. Ergotherapie).....

Bitte erläutern Sie bitte hier, welche Art von außerschulischer Förderung Ihr Kind erhält und bei wem (Name, Institut) es die Förderung bekommt oder bekam:

---

---

2. Ich übe mit meinem Kind Lesen über die Hausaufgaben hinaus . . .
- selten oder nie .....
- bis 1 Stunde pro Woche.....
- 1 – 2 Stunden pro Woche .....
- mehr als 2 Stunden pro Woche.....

3. Ich übe mit meinem Kind Schreiben über die Hausaufgaben hinaus . . .
- selten oder nie .....
  - bis 1 Stunde pro Woche .....
  - 1 – 2 Stunden pro Woche.....
  - mehr als 2 Stunden pro Woche.....
4. Mein Kind beschäftigt sich am Computer mit Lernprogrammen zum Lesen bzw. Schreiben. . .
- selten oder nie .....
  - bis 1 Stunde pro Woche .....
  - 1 – 2 Stunden pro Woche.....
  - mehr als 2 Stunden pro Woche.....
5. Mein Kind spielt am PC oder mit Playstation / Gameboy etc. pro Tag:
- weniger als eine halbe Stunde .....
  - 30 – 60 Minuten .....
  - mehr als eine Stunde .....
6. Durchschnittlicher Fernsehkonsum des Kindes pro Tag:
- weniger als eine halbe Stunde .....
  - 30 – 60 Minuten .....
  - mehr als eine Stunde .....
7. Dass die Klasse meines Kindes an einem Modellprojekt zur Verbesserung des Schriftspracherwerbs teilnimmt finde ich . . .
- sehr gut.....
  - gut .....
  - ist mir egal .....
  - schlecht .....
  - sehr schlecht .....

8. Hat das Bekanntwerden der Zwischenergebnisse im Lesen und Rechtschreiben Ende 2004 Konsequenzen für Sie bzw. Ihr Kind gehabt? Z.B., haben Sie wegen eventuell schlechter Ergebnisse Nachhilfe organisiert, einen Elternabend eingefordert oder etwas in der Art unternommen?

---

9. Wenn Sie uns noch etwas mitteilen wollen:

---