



Prof. Dr. Markus Linnemann, Dr. Simone Jambor-Fahlen, Dr. Stefanie Bredthauer,  
Dr. Kathrin Hippmann, Dr. Fabian Schmidt, Prof. Dr. Michael Becker-Mrotzek,  
Prof. Dr. Olaf Köller

## Lesen macht stark – Grundschule

### Teil A: Ergebnisse der wissenschaftlichen Begleitung

## Inhalt

0. Einleitung .....	3
I. Eignung des Materials zur Identifizierung von Schwierigkeiten beim Schriftspracherwerb	
5	
II. Eignung des Materials zur Unterstützung bei Schwierigkeiten im Schriftspracherwerb .....	6
Anlage der Studie .....	6
Stichprobe .....	8
Ergebnisse der standardisierten Tests .....	9
TEPHOBE .....	9
CFT 1-R .....	9
ELFE .....	10
SLRT .....	13
Zusammenfassung der Ergebnisse aus der Erhebung der standardisierten Tests .....	14
Lesen macht stark – Die Meilensteine .....	16
Meilenstein 1 .....	16
Meilenstein 2 .....	17
Meilenstein 3 .....	18
Meilenstein 4 .....	19
Meilenstein 5 .....	20
Meilenstein 6 .....	21
Meilenstein 7 .....	23
Zusammenfassung .....	24
Fazit zu den Aufgaben der Meilensteine .....	24
Strukturelle Unterschiede zwischen den an der Evaluation beteiligten Schulen .....	25
III. Implementation: Ergebnisse der Lehrerbefragung zur Tauglichkeit des Materials .....	26
Nutzung und Akzeptanz des Materials .....	26
Nutzung der zusätzlichen Lehrerwochenstunden .....	28
Fazit .....	28

## 0. Einleitung

Im Schuljahr 2013/14 startete das Projekt zur Diagnose und Förderung schriftsprachlicher Kompetenzen im Anfangsunterricht „Niemanden zurücklassen – Lesen macht stark Grundschule“ (LMS). Die Zielsetzungen des Projekts „LMS – Grundschule“ sind

- das frühzeitige Erkennen der Kinder mit Schwierigkeiten beim Schriftspracherwerb
- die Ableitung individueller Förderempfehlungen, basierend auf den diagnostischen Erkenntnissen
- die Dokumentation der Lernentwicklung für Elternarbeit und Förderplanarbeit mit Fachkräften
- die Prävention von Lese-Rechtschreibschwierigkeiten
- Professionalisierung der teilnehmenden Lehrkräfte in Bezug auf Diagnose und Förderung

In Zusammenarbeit mit dem Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache der Universität zu Köln, dem Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik an der Christian-Albrechts-Universität zu Kiel und mit dem Cornelsen-Verlag wurde ein lehrwerksunabhängiges, zweigliedriges Arbeitsmaterial aus Schüler- und Lehrerheft entwickelt, das Diagnoseaufgaben und Förderhinweise enthält. Die Diagnoseaufgaben gliedern sich in sieben Meilensteine, die jeweils mehrere Aufgaben zu verschiedenen Fertigkeiten und Fähigkeiten der 1. und 2. Jahrgangsstufe beinhalten (Becker-Mrotzek et al., 2014).

Nach einer ersten Pilotierungsphase mit zehn Schulen wurde mit Beginn des Schuljahres 2014/15 das vom *Institut für Qualitätsentwicklung an Schulen Schleswig-Holstein* koordinierte Projekt in 100 Schulen im ersten Schuljahr gestartet (Abbildung 1). Bei 30 Schulen bestand das Design zusätzlich aus dem Einsatz der entwickelten Diagnoseaufgaben und aus der Erhebung der standardisierten Tests *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler* (Lenhard & Schneider, 2006), *SLRT-II: Lese- und Rechtschreibtest* (Moll & Landerl, 2010), *TEPHOBE - Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit* (Mayer, 2013) und dem *Grundintelligenztest Skala 1 – Revision (CFT 1 – R)* (Weiß & Osterland, 2012).

Während der wissenschaftlichen Begleitung des Projektes mit Fokus auf die Erstellung der Diagnoseaufgaben und der Förderhinweise wurden zusätzlich Fragestellungen zur Evaluation verschiedener Teilaspekte des Projektes verfolgt, die zwar z.T. anhand der gleichen Stichprobe, jedoch in je eigenen Forschungsdesigns bearbeitet wurden. Die Fragestellungen beziehen sich auf die Diagnosefunktion der Materialien, auf ihre Förderfunktion sowie auf die Implementation des Diagnose- und Fördertools. Im Einzelnen stellten sich folgende Fragestellungen:

1. Ist das Material geeignet, um Kinder mit Schwierigkeiten beim Schriftspracherwerb zu identifizieren? (Diagnosefunktion)
2. Ist das Material geeignet, um Kinder mit Schwierigkeiten beim Schriftspracherwerb in ihrem Lernprozess zu unterstützen? (Förderfunktion)
3. Gelingt die Implementation des Materials in der schulischen Praxis? (Implementation)

Diese Teilaspekte werden im Folgenden getrennt dargelegt.



## I. Eignung des Materials zur Identifizierung von Schwierigkeiten beim Schriftspracherwerb

*Ist das Material geeignet, um Kinder mit Schwierigkeiten beim Schriftspracherwerb zu **identifizieren**? (Diagnosefunktion)*

Zur Untersuchung der Diagnosefunktion des Instruments „Lesen macht stark“ werden die Schnittmengen der Ergebnisse zur Feststellung von Förderbedarf des entwickelten Instruments und externer standardisierter Tests miteinander verglichen, d.h. es wurde untersucht, ob das Material dieselben Kinder identifiziert wie ein standardisierter Test, der die Messung ähnlicher Kompetenzen zum Ziel hat. Hiermit soll überprüft werden, ob die Diagnoseaufgaben als Screening-Instrument geeignet sind.

Als Grundlage für diese Berechnung dienten die jeweils von den Testverfahren vorgegebenen Schwellenwerte. Bei Kindern, die einen Punktwert unterhalb dieses jeweiligen Schwellenwerts erzielten, wurde ein Förderbedarf angenommen. Für den Abgleich wurden die Ergebnisse des Testverfahrens SLRT-Orthografie zum Erhebungszeitpunkt VI und die Ergebnisse der Screeningaufgabe des „Lesen macht stark“-Materials zur Rechtschreibkompetenz zum Erhebungszeitpunkt V herangezogen (vgl. Abbildung 1). Denn bei diesem Vergleich konnte von einer hohen Ähnlichkeit der getesteten Kompetenzkonstrukte (durch die Aufgabenformate) sowie einen geringen zeitlichen Abstand zwischen den Erhebungszeitpunkten (beide zum Ende der ersten Klasse) ausgegangen werden.

Das Ergebnis ist in Tabelle 1 dargestellt und zeigte mit 78.5% eine große Schnittmenge der Zuweisungen beider Verfahren zu den Gruppen *kein Förderbedarf* und *Förderbedarf*. Das heißt, für die meisten Kinder ergab die Screeningaufgabe des Materials und das externe Testverfahren dieselbe Gruppenzuordnung. Wenn jedoch ausschließlich die Gruppe der Kinder betrachtet wird, für die das Testverfahren SLRT-Orthografie einen Förderbedarf ergab, so zeigt sich, dass von diesen Kindern 27% nicht durch die „Lesen macht stark“-Aufgabe identifiziert werden konnten. Dieses Ergebnis kann durch die unterschiedlichen Zielsetzungen der beiden Verfahren erklärt werden: Im Gegensatz zu dem standardisierten Testverfahren soll das „Lesen macht stark“-Material ein ökonomisch einsetzbares Screening für den Regelunterricht liefern, das durch Bearbeitungen von Einzelaufgaben abgesichert wird. Dieses Screening wird nicht einmalig, sondern kontinuierlich und in unterschiedlichen Teilbereichen des Schriftspracherwerbs durchgeführt. Dass diese Konzeption des Materials sinnvoll gewählt ist, zeigen die Ergebnisse zur Dynamik der Lernverläufe der Schülerinnen und Schüler (Hippmann et al., 2016). Darüber hinaus soll das Material ein Heranführen von Lehrkräften an Diagnostikverfahren und gezielte Förderung in ihren Klassen ermöglichen.

Tabelle 1: Vergleich der Ergebnisse von SLRT-Orthografie und „Lesen macht stark“-Rechtschreibkompetenz

n=750		MS4-Rechtschreibkompetenz	
		Förderbedarf	Kein Förderbedarf
SLRT-Orthografie	Förderbedarf	26	9
	Kein Förderbedarf	152	563

## II. Eignung des Materials zur Unterstützung bei Schwierigkeiten im Schriftspracherwerb

*Ist das Material geeignet, um Kinder mit Schwierigkeiten beim Schriftspracherwerb in ihrem Lernprozess zu **unterstützen**? (Förderfunktion)*

Ziel war es, zu untersuchen, ob das diagnostische Instrument „Lesen macht stark“ mit zwei zusätzlichen Lehrerwochenstunden einen positiveren Einfluss auf die Entwicklung schriftsprachlicher Kompetenzen von Schülerinnen und Schülern hat als der Einsatz des Instrumentes ohne diese zusätzlichen Stunden. Hierzu wird verglichen, ob die Schülerinnen und Schüler in den Aufgaben aus „Lesen macht stark“ besser abschneiden, wenn die Lehrerinnen und Lehrer Zusatzstunden erhalten haben (Experimentalgruppe 1 = EG1) als wenn die Lehrerinnen und Lehrer diese Stunden nicht erhalten haben (Experimentalgruppe 2 = EG2). Des Weiteren sollte evaluiert werden, ob das diagnostische Material mit und ohne Zugabe von Stunden insgesamt das schriftsprachliche Lernen besser fördert als das Lernen ohne die Anwendung des diagnostischen Instrumentes (Kontrollgruppe = KG).

### Anlage der Studie

Die Datenerhebung fand in einem Längsschnittdesign mit unterschiedlichen zeitlichen Abständen in zehn Erhebungswellen statt (Abbildung 1). Drei dieser Erhebungswellen beinhalteten die Durchführung standardisierter Tests, zu sieben Messzeitpunkten wurden neu entwickelte Instrumente zur Datenerhebung (Diagnoseaufgaben aus „Lesen macht stark – Grundschule“) genutzt.

Zur Erhebung der Lesekompetenz, Rechtschreibkompetenz, phonologischen Bewusstheit und kognitiven Grundfertigkeiten wurden folgende standardisierten Tests eingesetzt:

- Lesekompetenz: Um die Lesekompetenzen der Kinder zu erfassen, wurde mit dem *ELFE 1-6 Ein Leseverständnistest für Erst- bis Sechstklässler* (Lenhard et al., 2006) ein standardisiertes Testverfahren eingesetzt. Der Test besteht aus den Untertests Wort-, Satz- und Textverständnis. Zum Messzeitpunkt 1, am Ende der ersten Klasse, wurde der ELFE-Test erstmals durchgeführt. Da die Lesefähigkeiten zu diesem frühen Zeitpunkt noch sehr basal sind, wurde auf die Durchführung des Untertests Textverständnis verzichtet, während zum Messzeitpunkt 2, am Ende des zweiten Schuljahrs, alle drei Untertests durchgeführt wurden.
- Rechtschreibkompetenz: Die Rechtschreibfähigkeiten der Schülerinnen und Schüler wurden durch den Rechtschreibteil des *SLRT II: Lese- und Rechtschreibtest* (Moll & Landerl, 2010) erfasst. Der SLRT erlaubt eine differenzierte Diagnose im Bereich des lautierenden und des orthographischen Schreibens.
- Phonologische Bewusstheit: Aus dem standardisierten Testinstrument *Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit TEPHOBE* (Mayer, 2013) wurden die Testteile zu Reimen und Anlauten verwendet, um die phonologische Bewusstheit der Kinder zu erfassen. Dies geschah zu Beginn des ersten Schuljahrs, da es sich um eine sogenannte Vorläuferfähigkeit des Schriftspracherwerbs handelt.

- Kognitive Grundfähigkeiten: Mit dem *Grundintelligenztest Skala 1 – Revision CFT 1-R* (Weiß & Osterland, 2012) wurden die intellektuellen Grundfähigkeiten der Kinder erfasst. Der Test wurde am Ende des zweiten Schuljahrs durchgeführt. Die Reliabilität (Retest nach vier Wochen Gesamt  $r_{tt} = .95$ ) ist als hoch zu bewerten.

Erhebungszeitpunkt	Schulstufe 1					Schulstufe 2				
	I. 09/14	II. 10/14	III. 11/14	IV. 01/15	V. 05/15	VI. 7/2015	VII. 11/15	VIII. 01/16	IX. 05/16	X. 07/16
EG 1 / EG 2	MS 1	THE- PHO- BE	MS 2	MS 3	MS 4	ELFE Wort- und Satz- ver- ständ- nis; SLRT	MS 5	MS 6	MS 7	CFT-R, ELFE Wort-, Satz- und Text- ver- ständ- nis; SLRT
Kontrollgruppe KG		Rei- me und An- laute								

Anmerkung: MS=Meilenstein; hellblau unterlegt=Einsatz standardisierter Tests

Abbildung 1: Test- und Evaluationsdesign

Die teilnehmenden Probandinnen und Probanden wurden in Klassenverbänden drei Untersuchungsgruppen zugeteilt:

- EG1: Diese Klassen arbeiteten mit dem Instrument „Lesen macht stark“ und erhielten dafür zusätzliche Lehrerwochenstunden zur Unterstützung des Schriftspracherwerbs (im Folgenden EG1 oder „Material + Stunden“ genannt).
- EG2: Diese Klassen arbeiteten ebenfalls mit dem Instrument „Lesen macht stark“, erhielten hierfür jedoch keine zusätzlichen Lehrerwochenstunden (im Folgenden EG2 bzw. „nur Material“ genannt).
- KG: Diese Klassen arbeiteten ohne das Instrument „Lesen macht stark“ (im Folgenden Kontrollgruppe KG genannt).

Im Rahmen des „Lesen macht stark“-Projektes erhielten einige der beteiligten Schulen (EG1) zusätzliche Lehrerwochenstunden zur Unterstützung des Schriftspracherwerbs. Ziel der Evaluation war es, zu untersuchen, ob das diagnostische Instrument „Lesen macht stark“ in Verbindung mit zusätzlichen Lehrerstunden einen positiveren Einfluss auf die Entwicklung schriftsprachlicher Kompetenzen von Schülerinnen und Schülern hat als der Einsatz des Instrumentes ohne diese zusätzlichen Stunden (EG2). Hierzu wurde verglichen, ob die Schülerinnen und Schüler in den Aufgaben aus „Lesen macht stark“ besser abschneiden, wenn die Lehrerinnen und Lehrer Zusatzstunden erhalten haben, als wenn die Lehrerinnen und Lehrer diese Stunden nicht erhalten haben.

Des Weiteren sollte evaluiert werden, ob das diagnostische Material mit und ohne Zugabe von Stunden insgesamt das schriftsprachliche Lernen besser fördert als das Lernen ohne die Anwendung des diagnostischen Instrumentes.

Da es sich bei der Anlage der Evaluation hinsichtlich der Förderfunktion im weiteren Sinne um eine Interventionsstudie handelt, wurden hierfür vorwiegend Varianzanalysen zur zufallskritischen Absicherung der deskriptiven Werte herangezogen. Zur Beantwortung der Frage nach

der Diagnosefunktion des Materials wurden anhand von Kreuztabellen die Schnittmengen der externen Tests und des „Lesen macht stark“-Instruments bestimmt. Die Implementation des Instruments wurde anhand deskriptiv-statistischer Auswertungen des Lehrkräfte-Fragebogens untersucht.

## Stichprobe

Die Stichprobe bestand aus 1555 Schülerinnen und Schülern aus 30 Grundschulen Schleswig-Holsteins. Je zehn Schulen bildeten eine Einheit – eine Versuchsbedingung –, die Gruppengrößen waren in etwa gleich. In EG1 befanden sich 496 Probandinnen und Probanden (31.9%), in EG2 512 (32.9%) und in der Kontrollgruppe 547 Schülerinnen und Schüler (35.2%). Aufgrund fehlender Werte schwankt die Zahl der Probandinnen und Probanden leicht je nach Analyse, da nicht alle Schülerinnen und Schüler an allen Tests teilgenommen haben.

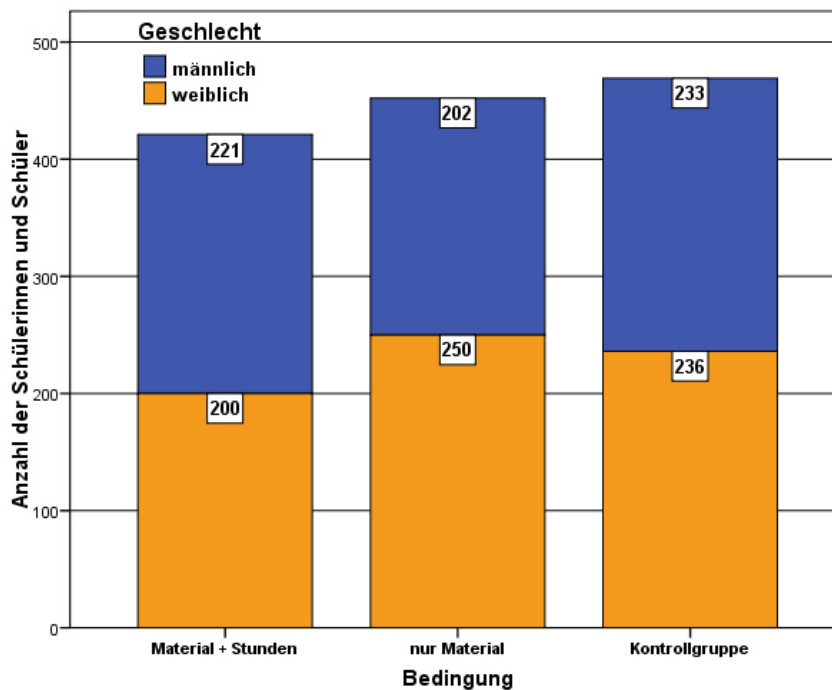


Abbildung 2: Verteilung der Schülerinnen und Schüler auf die Bedingungen und nach Geschlecht

Die Bereinigung der Datei erfolgte konservativ. Es wurden lediglich sechs Schülerinnen und Schüler bei verschiedenen Analysen ausgeschlossen, die unplausible (d.h. höhere als eigentlich zu erreichende) Werte hatten.

Bezogen auf die „Lesen macht stark“-Aufgaben und die standardisierten Tests THEPHOBE, ELFE, CFT, SLRT lagen von 274 Schülerinnen und Schülern komplette Datensätze vor. Diese verteilten sich auf 115 Schülerinnen und Schüler der EG1 und 156 der EG2. Wertet man nur die Instrumente aus „Lesen macht stark“ ohne die standardisierten Tests, liegen komplette Datensätze von 415 Schülerinnen und Schülern vor. Hier sind Kontextvariablen wie z.B. das



Alter nicht mit einbezogen. Dies würde die Anzahl der kompletten Datensätze weiter verringern, da diese Daten nicht für alle vorliegen.

Das mittlere Alter der Schülerinnen und Schüler betrug zum Zeitpunkt der ersten Datenerhebung im September 2014 6,8 Jahre (SD=0.41; Min=5.75; Max=8.33). Die Geschlechter waren insgesamt etwa gleichverteilt (s. Abbildung 2).

Als Daten der Schülerinnen und Schüler für die Aufgaben der Meilensteine lagen nicht die Rohwerte, sondern pro Aufgabe aggregierte Daten, also Punktsummen, vor. Die Punktsummen wurden von den Lehrkräften selbst gebildet. Reliabilitätswerte ließen sich so nicht berechnen.

## Ergebnisse der standardisierten Tests

### TEPHOBE

Zur ersten Kontrolle der Voraussetzungen der Schülerinnen und Schüler und zur ersten Einschätzung nach sechs Wochen wurden die Untertests „Anlaute“ und „Reimen“ des standardisierten Tests TEPHOBE eingesetzt. Es zeigten sich keine signifikanten Unterschiede zwischen den Gruppen (Anlaute:  $F(2, 1220)=1.47$ ;  $p=.230$ ; partielles  $\eta^2=.002$ ; Reimen:  $F(2, 1220)=0.41$ ;  $p=.666$ ; partielles  $\eta^2=.001$ ). Abbildung 3 zeigt die Mittelwerte.

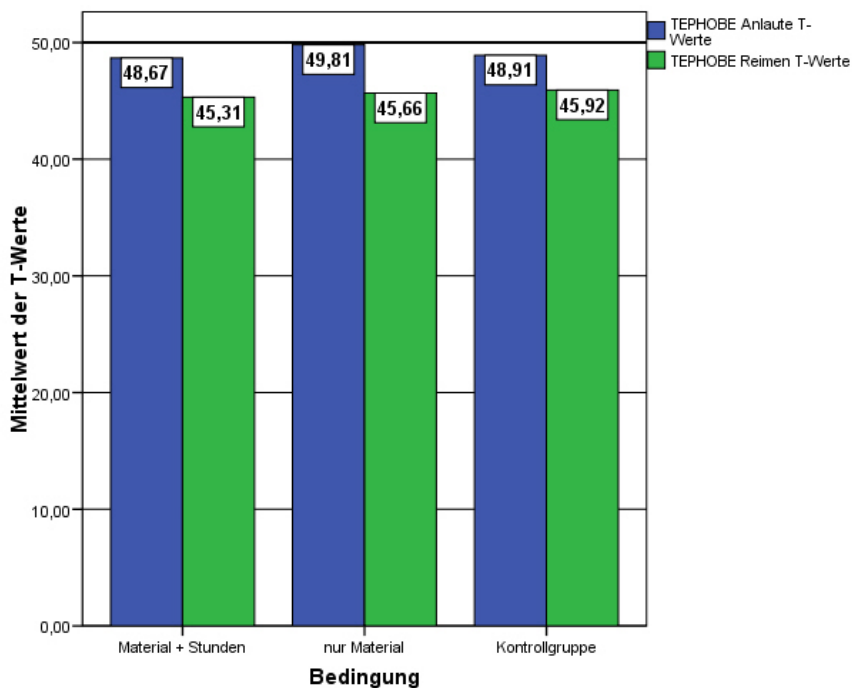


Abbildung 3: Mittelwerte im TEPHOBE Anlaute und Reime für die verschiedenen Gruppen

### CFT 1-R

Als Kontrollvariable und zur Beschreibung der Stichprobe wurde der CFT-R zum letzten Erhebungsmesszeitpunkt eingesetzt. Es zeigte sich, dass die EG2 die geringsten Werte erhielt und sich signifikant von der Kontrollgruppe unterschied ( $t(720)=-3.31$ ;  $p=0.001$ ). Der Unterschied zur EG1 betrug 1.3 T-Werte (nicht signifikant), der Unterschied zur KG 2.3 T-Wert-Punkte (signifikant), was in der Stichprobe etwa 1/4 Standardabweichung, in der Normierungsstichprobe

etwa 1/3 Standardabweichung entspricht. Die Reihenfolge war also  $EG2 < EG1 < KG$ . Der Gesamtmittelwert der Stichprobe betrug 50.1 T-Wert-Punkte ( $SD=9.2$ ). T-Werte sind normierte Werte mit dem Mittelwert 50 und der Standardabweichung 10 (s. Abbildung 4).<sup>1</sup>

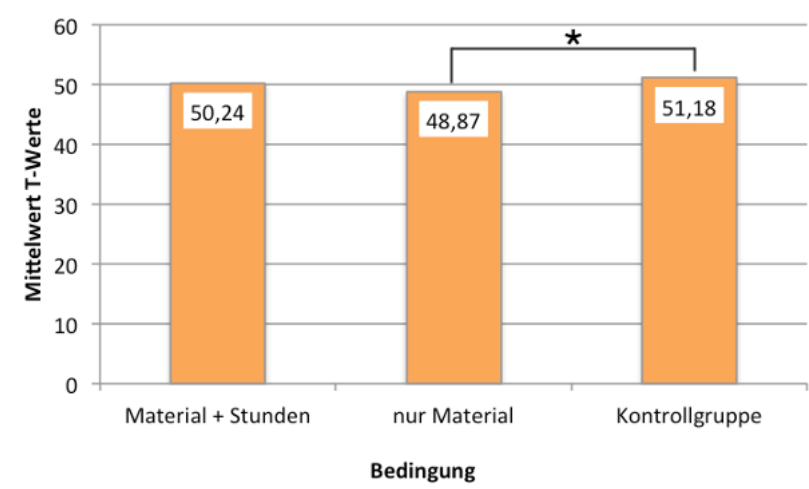


Abbildung 4: Mittelwerte im CFT für die verschiedenen Gruppen

### ELFE

Der ELFE-Lesekompetenztest besteht aus den Untertests Wort-, Satz- und Textverständnis. Da nicht zu beiden Zeitpunkten alle Untertests durchgeführt wurden, konnte keine Veränderung des Gesamtestwertes errechnet werden. Die Untertests wurden daher getrennt betrachtet. Durchgeführt wurde eine Varianzanalyse (ANOVA) mit dem Messwiederholungsfaktor ZEIT und der unabhängigen Variable GRUPPE (EG1, EG2, KG). Abhängige Variablen waren die erreichten Punkte in den entsprechenden Untertests<sup>2</sup>.

#### Wortverständnis

Es zeigte sich ein signifikanter Haupteffekt ZEIT, d.h. über alle Gruppen hinweg zeigte sich ein Anstieg der Werte von Messzeitpunkt 1 des ELFE (= Erhebungszeitpunkt VI) zum Messzeitpunkt 2 (= Erhebungszeitpunkt X) ( $F(1, 1027)=3884.25$ ;  $p<.001$ ; partielles  $\eta^2=.791$ ). Ein unterschiedlicher Anstieg je nach Bedingung konnte nicht gefunden werden ( $F(2, 1027)=1.89$ ;  $p=.151$ ; partielles  $\eta^2=.004$ ). Auch ein signifikanter Haupteffekt GRUPPE stellte sich nicht ein, d.h. über die Messzeitpunkte hinweg gab es keine Unterschiede zwischen den Gruppen ( $F(2, 1027)=1.45$ ;  $p=.236$ ; partielles  $\eta^2=.003$ ). Abbildung 5 verdeutlicht die Werte grafisch.

<sup>1</sup> Ein Einbezug der CFT 1-R-Werte in die folgenden Berechnungen als Kovariate veränderte an keiner Stelle die Ergebnisse substantiell, so dass diese Ergebnisse im Folgenden nicht berichtet werden, um den Text nicht überzustrapazieren.

<sup>2</sup> Verwendet wurden bei den Veränderungsmessungen nicht die normierten (standardisierten) Werte (z- oder T-Werte), denn diese maskieren individuelle Veränderungen. So kann ein Schüler während eines Jahres durchaus dazugelernt haben, trotzdem aber aufgrund seiner stets gleichen Position in einer Normierungsgruppe den gleichen T-Wert bzw. Prozentrang erlangen (jeweils z.B. T-Wert von 50, obwohl seine Anzahl richtig erkannter Wörter stetig steigt). Innerhalb eines Messzeitpunkt spielt diese Problematik keine Rolle.

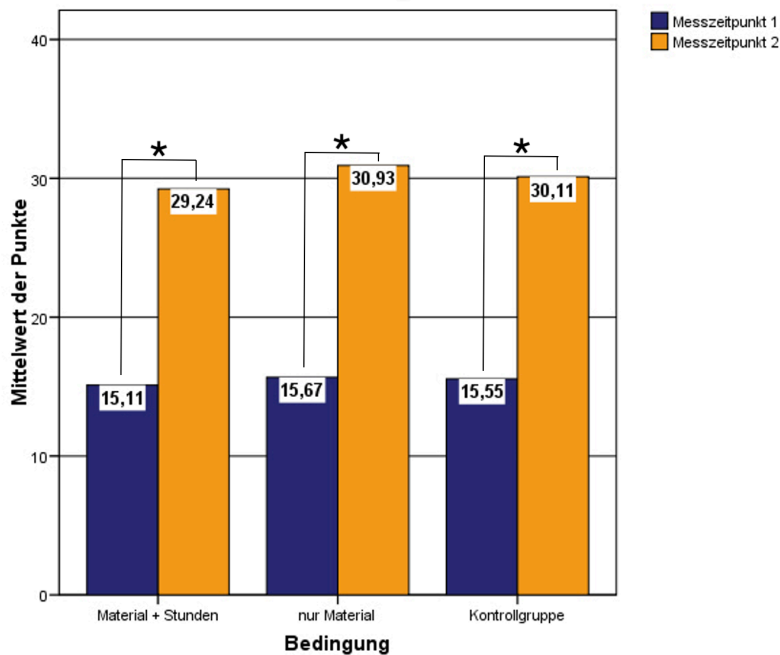


Abbildung 5: Mittelwerte im Wortverstehen für die verschiedenen Gruppen zum ersten und zweiten Messzeitpunkt

#### Satzverständnis

Es zeigte sich ein signifikanter Haupteffekt ZEIT ( $F(1, 1027) = 9525.63$ ;  $p < .001$ ; partielles  $\eta^2 = .903$ ), d.h. die Gruppen verbesserten sich. Ein unterschiedlicher Anstieg je nach Bedingung konnte genauso wenig gefunden werden ( $F(2, 1027) = 2.97$ ;  $p = .052$ ; partielles  $\eta^2 = .006$ ), wie ein signifikanter Haupteffekt GRUPPE ( $F(2, 1027) = 1.71$ ;  $p = .181$ ; partielles  $\eta^2 = .003$ ). Abbildung 6 verdeutlicht die Werte.

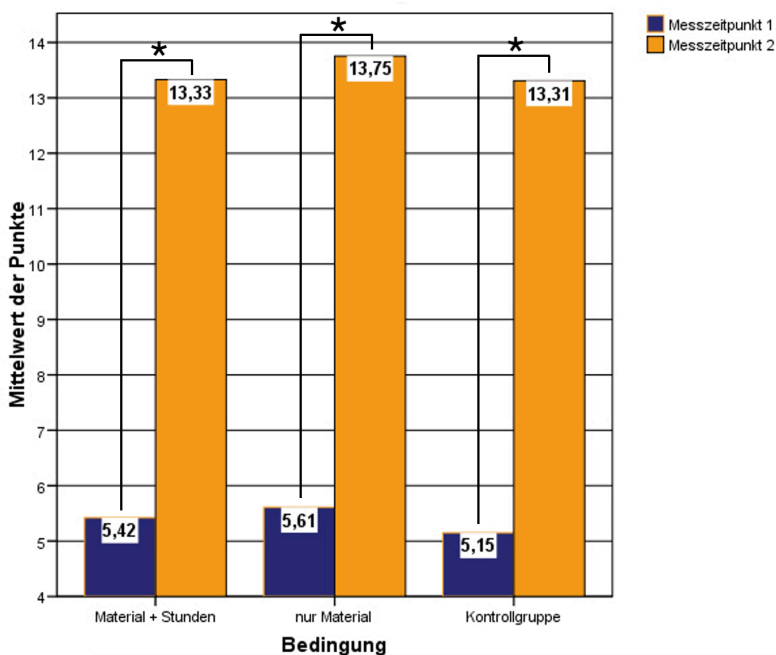


Abbildung 6: Mittelwerte im Satzverstehen für die verschiedenen Gruppen zum ersten und zweiten Messzeitpunkt

### Textverständnis

Textverständnis wurde nur beim zweiten Messzeitpunkt erhoben. Hier wurde statt der Veränderungsmessung eine einzelne Messung als einfaktorielle Varianzanalyse durchgeführt. Es zeigte sich kein signifikanter Unterschied zwischen den Gruppen ( $F(2, 1203)=.00$ ;  $p=1.00$ ;  $\text{Eta}^2=.00$ ). Im Textverstehen unterschieden sich die Schülerinnen und Schüler am Ende der 2. Klasse also nicht (Abbildung 7), in allen Gruppen wurden also etwa 8 Fragen zum Textverständnis richtig beantwortet.

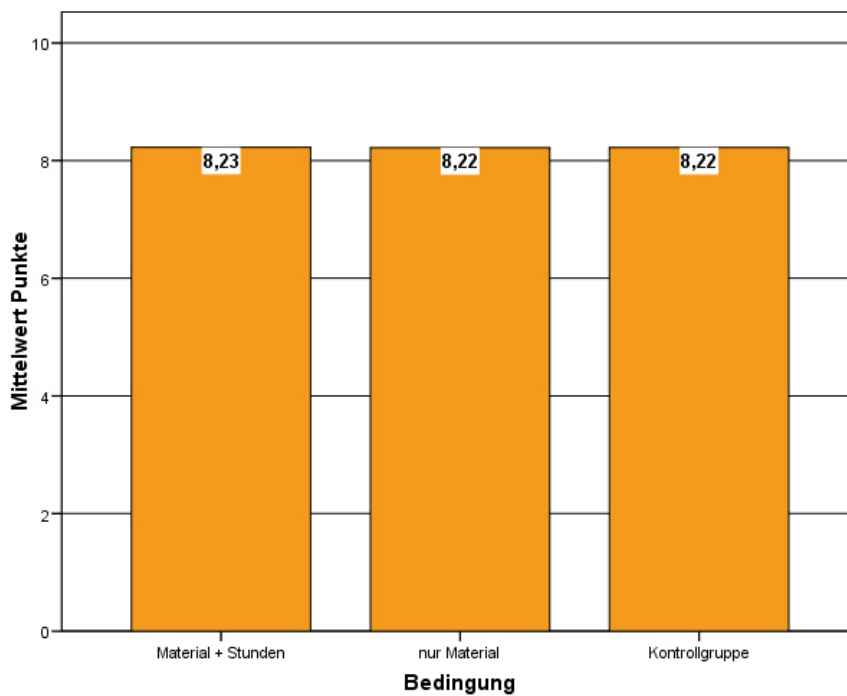


Abbildung 7: Mittelwerte im Textverstehen für die verschiedenen Gruppen zum zweiten Messzeitpunkt

### Gesamttestwert

Der ELFE-Lesetest bietet die Möglichkeit, alle Subtests (Wortverständnis, Satzverständnis, Textverständnis) zu einem Gesamtergebnis zusammenzufassen. Hierzu stellt der Test einen normierten Wert bereit, der sich als z-Wert deuten lässt. Auch im Gesamttestwert ließ sich kein Unterschied zwischen den Gruppen finden ( $F(2, 942)=1.91$ ;  $p=.149$ ; partielles  $\text{Eta}^2=.004$ ). Abbildung 8 zeigt die Unterschiede. Im Unterschied zur Normierungsstichprobe des ELFE lagen die Gruppen also jeweils etwa eine halbe Standardabweichung unter dem Mittelwert.

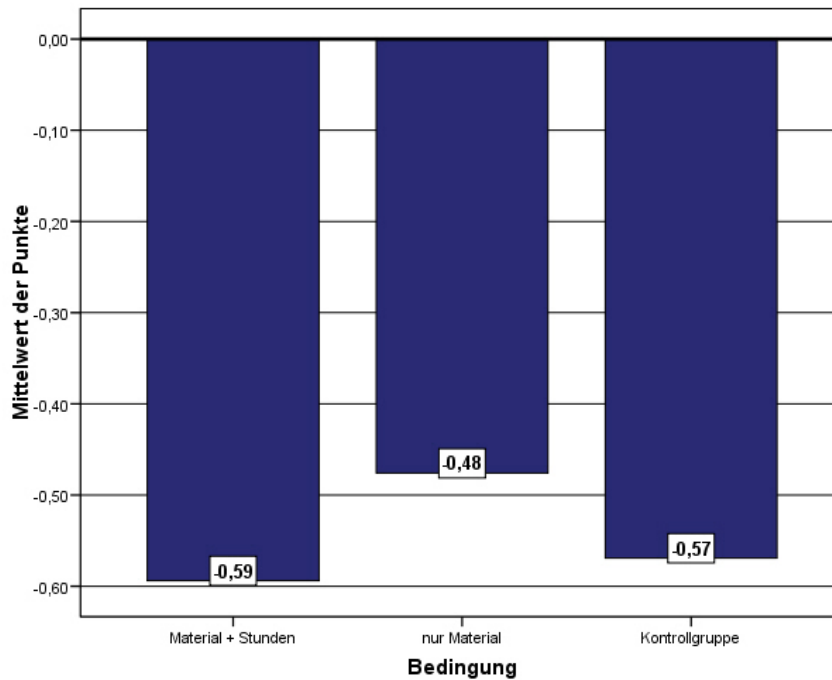
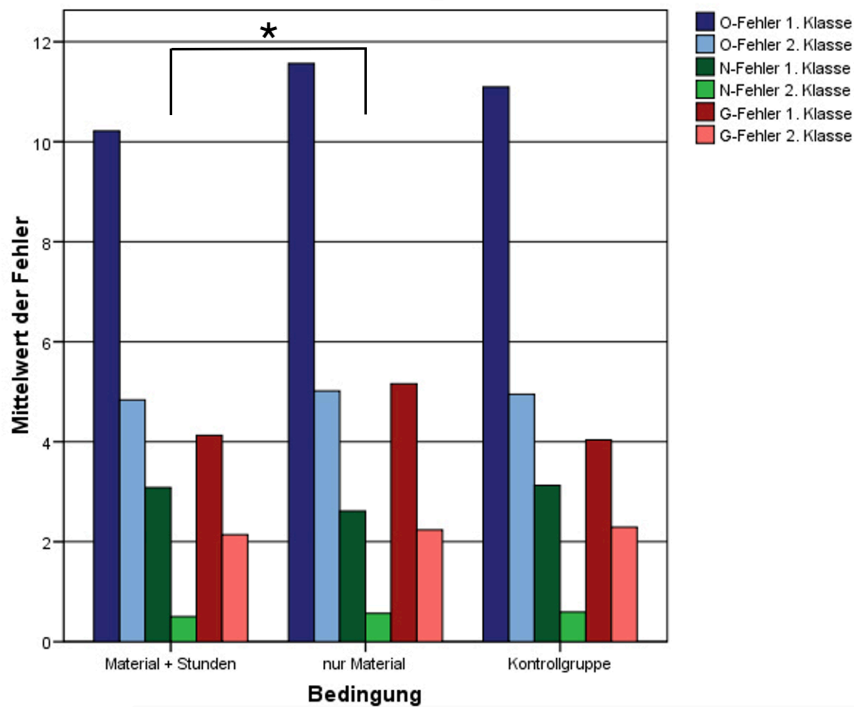


Abbildung 8: Mittelwerte im Gesamttest für die verschiedenen Gruppen zum zweiten Messzeitpunkt

### SLRT

Ähnlich wie beim ELFE-Lesetest wurde eine Varianzanalyse mit Messwiederholungsfaktor ZEIT (zweifach abgestuft) durchgeführt. Unabhängige Variable war wiederum die Gruppenzuordnung (GRUPPE: EG1, EG2, KG). Der SLRT bietet drei Werte an, die als abhängige Variablen in die Analyse eingegangen sind: nicht lautgetreue Fehler (N-Fehler), lautgetreue Schreibungen, die jedoch orthografisch falsch sind (O-Fehler) sowie Fehler der Groß- und Kleinschreibung (G-Fehler). Der signifikante Haupteffekt ZEIT zeigt, dass die Fehlerzahl abnimmt, der signifikante Haupteffekt GRUPPE zeigt, dass die Gruppen über die Fehlerarten und Zeitpunkte hinweg unterschiedlich viele Fehler machen. Die EG2 machte signifikant mehr Fehler als die EG1, die Kontrollgruppe lag zwischen den beiden Gruppen. Interessant sind die Interaktionseffekte GRUPPE x ZEIT bzw. GRUPPE x ZEIT x FEHLERART, denn diese zeigen, ob eine Gruppe sich (möglicherweise bei einer bestimmten Fehlerart) stärker verbessert hat als eine andere. Die nicht signifikante Interaktion ZEIT x GRUPPE ( $p=.083$ ) zeigte keine Unterschiede in der Veränderung zwischen den Gruppen vom ersten zum zweiten Zeitpunkt. Die signifikante Interaktion GRUPPE x ZEIT x FEHLERART fordert jedoch ein genaueres Betrachten: Während die EG2 sich bei den O- und G-Fehlern stärker verbesserte als die anderen Gruppen, verbesserte sie sich bei den N-Fehlern weniger stark als die beiden Gruppen EG1 und KG (Abbildung 9), allerdings ist das Ausgangsniveau bereits geringer, d.h. die N-Fehler der EG 2 sind bereits zum ersten Zeitpunkt geringer als bei den beiden anderen Gruppen.



Anmerkung: Die dunklen Balken geben die Werte am Ende der erste Klasse an, die hellen Balken diejenigen am Ende der zweiten Klasse.

Abbildung 9: Mittlere Fehleranzahl als Funktion der Bedingung für drei Fehlerarten: nicht lautgetreue Fehler (O-Fehler, blau), lautgetreue Fehler (N-Fehler, grün) und Fehler der Groß-Kleinschreibung (G-Fehler, rot)

#### *Zusammenfassung der Ergebnisse aus der Erhebung der standardisierten Tests*

Während sich zu Beginn der Datenerhebung im standardisierten Verfahren TEPHOBE keine Unterschiede zwischen den drei Gruppen ergaben, zeigten sich in zu den späteren Messzeitpunkten erhobenen Daten einige Unterschiede zwischen den Gruppen, die jedoch zumeist marginal waren. Der Fragestellung entsprechend wurde erwartet, dass die Reihenfolge bezüglich der Fähigkeiten stets  $KG < EG2 < EG1$  ist. Dies konnte so nicht bestätigt werden. Vielmehr zeigte sich, mitunter nur deskriptiv und nicht zufallskritisch abgesichert, dass meistens die EG2 höhere Werte erzielte, so beim ELFE Wortlesen am Ende der ersten Klasse sowie beim SLRT bei den Fehlertypen O und G. Die standardisierten Verfahren konnten so zusammenfassend keine positiven Effekte des Projekts LMS auf die Leistungsentwicklung der Schülerinnen und Schüler zeigen.

Zur Interpretation dieser Ergebnisse wurde zwei Erklärungsansätzen nachgegangen:

#### 1. Ausschluss leistungsschwacher Kinder bei den Erhebungen:

Auf Grundlage von Beobachtungen durch die Testleitenden kam die Hypothese auf, dass Lehrkräfte dafür gesorgt haben könnten, dass leistungsschwache Schülerinnen und Schüler zum Teil nicht an den Erhebungen mit den standardisierten Tests teilnahmen, z.B. um diese vor Drucksituationen zu schützen. Zur Prüfung dieser Hypothese wurden zum einen die Teilnahmequoten der verschiedenen Schulen an den drei Erhebungszeitpunkten mit den standardisierten Tests verglichen (Dropout-Analyse).

Hierbei zeigten sich keine Auffälligkeiten, die Teilnahmequoten sind vergleichbar. Zum anderen wurde überprüft, ob ein solcher Ausschluss leistungsschwacher Kinder ggf. im Verlauf der Studie erfolgt sein könnte. Hierfür wurde am Beispiel des SLRT untersucht, ob ein Zusammenhang zwischen den Testergebnissen der Kinder bei der erstmaligen Testdurchführung (Erhebungszeitpunkt VI) und bei der zweiten Testdurchführung (Erhebungszeitpunkt X) besteht. Die Ergebnisse zeigen, dass die Mehrzahl der Kinder, für die der SLRT bei der ersten Durchführung einen Förderbedarf ergab, bei der zweiten Testdurchführung nicht mehr teilnahmen. Die Korrelation dieses Zusammenhangs liegt mit  $-0.27$  im mittleren Bereich und ist signifikant. Auch der Vergleich der Leistungen beim erstmaligen SLRT zwischen den Kindern, die ebenfalls am zweiten SLRT teilnahmen, und denjenigen, die das nicht taten, anhand eines t-Tests ergab einen signifikanten Mittelwertunterschied zwischen diesen Gruppen. Diese Ergebnisse deuten darauf hin, dass für die Intervention relevante Gruppe von Schülerinnen und Schülern keine Daten zur Leistungsentwicklung zu dem zweiten Messzeitpunkt vorliegen. Die Annahme liegt nahe, dass die Ergebnisse der Evaluation so eine konservative Schätzung der Leistungsentwicklung darstellen.

2. Strukturelle Leistungsunterschiede auf Schulebene zwischen EG1, EG2 und KG:  
Die Annahme, dass sich die Gruppen EG1, EG2 und KG in der Leistung strukturell auf Schulebene unterscheiden, wurde anhand der Vera-Daten<sup>3</sup> aus den Jahren 2015, 2016 und 2017 geprüft. Hierzu wurden die an dem Projekt LMS beteiligten Schulen in den Leistungen im Lesetest und dem Mathematiktest (Muster und Strukturen) zu den drei Zeitpunkten miteinander verglichen. Signifikante Unterschiede zwischen den Leistungsniveaus der Schulen würden unterschiedliche Grundvoraussetzungen in den Leistungen der Schülerinnen und Schüler implizieren, die in die Auswertung der Evaluation miteinbezogen werden müssten. Zu keinem der drei Zeitpunkte zeigte sich ein signifikanter Unterschied zwischen den Schulen der EG1, EG2 und KG (2015: Lesetest,  $F(2, 26)=.52$ ,  $p > .05$ , Mathematiktest  $F(2, 26)=.68$ ,  $p > .05$ ; 2016: Lesetest,  $F(2, 26)=.44$ ,  $p > .05$ , Mathematiktest  $F(2, 26)=.51$ ,  $p > .05$ ; 2017: Lesetest,  $F(2, 26)=.70$ ,  $p > .05$ , Mathematiktest  $F(2, 26)=1.40$ ,  $p > .05$ ). Die Ergebnisse weisen damit darauf hin, dass sich die Schulen nicht anhand der Vera-Daten (Mathematiktest und Lesetest) unterscheiden und daher keine strukturellen Unterschiede im Ausgangsniveau angenommen werden müssen. Auch in den hier nicht im Detail berichteten Anteilen der Risikoschülerinnen und Risikoschüler liegen keine signifikanten Unterschiede zwischen den Schulen vor.

---

<sup>3</sup> Dies sind vergleichende Lernstandserhebungen in der 3. und 8. Jahrgangsstufe, die flächendeckend in allen allgemeinbildenden Schulen in Deutschland durchgeführt werden.

## Lesen macht stark – Die Meilensteine

Da die „Lesen macht stark“-Aufgaben sich von Messzeitpunkt (ab hier MS für „Meilenstein“ genannt) zu Messzeitpunkt unterscheiden, wurde zur Evaluation des Programms keine Veränderungsmessung modelliert. Einzelne Aufgaben wurden jedoch pro MS mittels Faktorenanalyse (Hauptachsenanalyse) zu Konstrukten (z.B. Schriftkompetenz produktiv und rezeptiv, Textkompetenz) zusammengefasst. Dies diente der einfacheren Handhabung der Auswertungen. Die (jeweils gleich oder ähnlich klingenden) Konstruktbenennungen beschreiben möglicherweise tatsächlich das gleiche Konstrukt, es kann jedoch aufgrund der unterschiedlichen und unterschiedlich schwierigen Items nicht sichergestellt werden, dass gleiche Konstruktwerte die gleiche Ausprägung einer Kompetenz bedeuten. Daher wurde jeweils innerhalb eines Meilensteins untersucht, inwieweit sich die Gruppen unterscheiden, nicht über die Meilensteine hinweg. Da die Kontrollgruppe kein Lesen macht stark-Material erhalten hat, werden jeweils die Gruppen EG1 und EG2 miteinander verglichen.

### Meilenstein 1

Da der Meilenstein 1 Aufschluss über den schriftsprachlichen Stand der Schülerinnen und Schüler zu Beginn des „Lesen macht stark“-Programmes gibt und zudem die Aufgaben nicht zu Kompetenzen zusammengefasst sind, werden hier auf Aufgabenebene Unterschiede zwischen den Gruppen verdeutlicht. Tabelle 2 zeigt deskriptive und inferenzstatistische Werte für die beiden Gruppen. Von den acht durchgeführten Untertests zeigten vier signifikante Unterschiede zwischen den Gruppen. Die Richtung ist dabei stets die gleiche: EG2>EG1 mit einer kleinen Effektstärke, d.h. dass die Eingangsvoraussetzungen insgesamt für EG2 leicht besser sind.

Tabelle 2: Deskriptive und inferenzstatistische Werte für die Gruppen EG1 und EG2.

	<i>M</i>		<i>SD</i>		<i>d</i>	<i>t</i>
	EG1	EG2	EG1	EG2		
Graphomotorik	1.55	1.56	.61	.62	-0.02	0.43
Namen schreiben	1.64	1.76	.64	.49	-0.21**	3.12
Buchstaben erkennen	6.44	6.33	1.76	1.86	0.06	0.83
Laut-Buchstabenzuordnung	4.92	5.17	1.61	1.49	-0.16*	2.23
Reimwörter	4.46	4.50	1.05	1.06	-0.04	0.47
Silbengliederung	5.94	5.92	1.36	1.57	0.01	0.19
Anlaute	6.59	7.32	2.87	2.59	-0.27***	3.84
Texte verstehen	.72	.81	.45	.39	-0.21**	2.93

Anmerkungen: \*= $p < .05$ , \*\*= $p < .01$ , \*\*\*= $p < .001$



### Meilenstein 2

*Schriftkompetenz rezeptiv.* Einzige rezeptive Aufgabe war die Aufgabe „Wort-Bild-Zuordnung“. Hier zeigten sich keine signifikanten Unterschiede zwischen den Gruppen hinsichtlich des Mittelwerts ( $t(751.3^4)=1.21$ ;  $p=.224$ ). Deskriptiv lagen die Werte der EG1 über denen der EG2 mit einem fast nicht wahrnehmbaren Effekt von etwa  $d=0.09$  (Tabelle 3). Die Streuung der Werte der EG1 lag signifikant über der der EG2 ( $p=.001$ ), die Leistungen der Schülerinnen und Schüler der EG1 variierten also etwas stärker.

Tabelle 3: Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven Schriftkompetenz.

Gruppe	N	M	SD	SE
EG1	373	5.42	2.70	.14
EG2	419	5.64	2.42	.12

*Schriftkompetenz produktiv.* Die Aufgaben „Pseudowörter“ und „Wörter schreiben“ dienten als Ausgangspunkt für eine zusammenfassende Faktorenanalyse (Hauptachsenfaktorenanalyse). Da die Aufgaben unterschiedliche Skalierungen besitzen, können die Werte der Schülerleistungen nicht einfach addiert werden. Mit Hilfe einer Faktorenanalyse lässt sich ein Kennwert erstellen, der mehrere Variablen zusammenfasst, vorausgesetzt es macht inhaltlich Sinn, was hier gegeben ist, da beide Aufgaben produktive Schriftkompetenz messen, und vorausgesetzt der resultierende Faktorwert stellt die originalen Werte ausreichend gut dar. Der hier erstellte Faktor erklärte 72.0 % der Gesamtvarianz. Der Wert ist ausreichend hoch, da sich ein erheblicher Anteil der Unterschiede im Bearbeiten der Aufgaben „Pseudowörter“ und „Wörter schreiben“ durch die Unterschiede der Faktorwerte zwischen den Schülerinnen und Schülern darstellen lassen. Ein t-Test zur Untersuchung der Unterschiede zwischen den Schülerinnen und Schülern der drei Gruppen mit den berechneten Faktorwerten zeigte signifikante Unterschiede ( $t(744.3)=4.18$ ;  $p<.001$ ), die Mittelwerte der EG2 lagen über den Mittelwerten der EG1 (Tabelle 4). Die Streuung der Werte der EG1 lag signifikant über der der EG2 ( $p=.001$ ), die Schülerinnen und Schüler der EG1 sind also etwas unterschiedlicher.

Tabelle 4. Deskriptive Werte für die Gruppen EG1 und EG2 in der produktiven Schriftkompetenz.

Gruppe	N	M	SD	SE
EG1	369	-.14	.96	.05
EG2	416	.13 <sup>*</sup>	.86	.04

*Textkompetenz rezeptiv und produktiv.* Als rezeptive Textkompetenz wurde „Globales Textverstehen“ erhoben. Hier zeigten sich kein signifikanter Unterschied ( $t(774.8)=1.00$ ;  $p=.314$  (Tabelle 5)). Eine produktive Aufgabe wurde nicht gestellt.

<sup>4</sup> Korrigierte Freiheitsgrade wegen Varianzungleichheit.

Tabelle 5: Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven und produktiven Textkompetenz.

Gruppe	N	M	SD	SE
EG1	368	.62	.49	.03
EG2	419	.59	.49	.02

### Meilenstein 3

*Schriftkompetenz rezeptiv.* Zu den zu analysierenden Aufgaben gehörten „Richtige Wörter finden, genaues Lesen“, „Wort-Bild-Zuordnung“ und „Sätze beurteilen, genaues und schnelles Lesen (mit Zeitvorgabe)“. Eine Faktorenanalyse, deren Ausgangsbasis die Korrelationsmatrix aus Tabelle 6 war, konnte mit einem Faktor 55.4 % der Gesamtvarianz erklären<sup>5</sup>.

Tabelle 6: Korrelationsmatrix

	Wort-Bild-Zuordnung	Sätze beurteilen, genaues und schnelles Lesen M3
richtige Wörter finden	.65	.47
Wort-Bild-Zuordnung		.50

Ein t-Test der Faktorwerte der beiden Gruppen EG1 und EG2 zeigte signifikante Unterschiede ( $t(704.7)=2.22$ ;  $p<.05$ ): Die Werte der EG 2 lagen über den Werten der EG1 (s. Tabelle 7).

Tabelle 7: Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven Schriftkompetenz.

Gruppe	N	M	SD	SE
EG1	348	-.08	.95	.05
EG2	399	.07 *	.85	.04

*Schriftkompetenz produktiv.* Die Korrelation zwischen den beiden Aufgaben betrug  $r=.78$ , was eine Zusammenfassung zu einem Faktor rechtfertigte. Dieserklärte 77.9% der Varianz auf. Ein t-Test zeigte signifikante Unterschiede zwischen den beiden Gruppen ( $t(675.1)=3.80$ ;  $p<.001$ ): Die EG2 erreichte höhere Werte als die EG1 (Tabelle 8). Die Varianzen der beiden Gruppen waren unterschiedlich, die Schülerinnen und Schüler der EG2 waren sich ähnlicher als die der EG1.

<sup>5</sup> Die erklärte Varianz ist bei Hauptachsen-Faktorenanalysen meist geringer als bei Hauptkomponentenanalysen. Ein weiterer Faktor hätte bei dieser Analyse keinen substantiellen zusätzlichen Beitrag geliefert.

Tabelle 8: Deskriptive Werte für die Gruppen EG1 und EG2 in der produktiven Schriftkompetenz.

Gruppe	N	M	SD	SE
EG1	349	-.14 *	1.02	.05
EG2	401	.12	.84	.04

*Textkompetenz rezeptiv und produktiv.* Bei den Aufgaben zur Textkompetenz wird zwar in produktive und rezeptive Aufgaben unterschieden, die in die Berechnungen aufzunehmende Punktzahl lag zur Auswertung jedoch nur in Kombination (Summe) vor. Ein leichter signifikanter Unterschied zeigte sich bei dieser Summe der Textkompetenz ( $t(762)=2.14$ ;  $p<.05$ ). Die deskriptiven Werte zeigt Tabelle 9. Im Gegensatz zu den meisten anderen Analysen erhielt hier die EG1 höhere Werte als die EG2.

Tabelle 9. Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven und produktiven Textkompetenz.

Gruppe	N	M	SD	SE
EG1	357	2.29 *	.83	.04
EG2	407	2.16	.85	.04

#### Meilenstein 4

*Schriftkompetenz rezeptiv.* Die Faktorenanalyse, in die die gleichen Aufgabenformate wie in MS 3 gingen, erreichte mit einem einzigen Faktor eine Varianzaufklärung von 47.4%. Dies ist gegenüber dem MS 3 eine etwas geringere Anpassung, was an leicht geringeren Interkorrelationen (Tabelle 10) liegt.

Tabelle 10: Korrelationsmatrix

	Wort-Bild-Zuordnung	Sätze beurteilen, genaues und schnelles Lesen
richtige Wörter finden, genaues Lesen	.60	.38
Wort-Bild-Zuordnung		.36

Unterschiede zwischen den Gruppen zeigte ein signifikanter t-Test ( $t(766)=2.01$ ;  $p<.05$ ). Die Gruppe EG2 erzielte höhere Werte als die EG1 (Tabelle 11)

Tabelle 11: Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven Schriftkompetenz.

Gruppe	N	M	SD	SE
EG1	356	-.07 *	.87	.05
EG2	412	.06	.87	.04

*Schriftkompetenz produktiv.* Die produktive Schriftkompetenz wurde durch die gleichen Aufgabenformate gemessen wie in MS 3. Die Korrelation der beiden Aufgaben lag bei  $r=.64$ . Ein Faktor der durchgeführten Faktorenanalyse erklärte 63.5 % der Gesamtvarianz. Unterschiede zwischen den Gruppen hinsichtlich der Mittelwerte der Faktorwerte zeigten sich keine ( $t(770)=1.10$ ;  $p=.271$ ). Tabelle 12 zeigt die deskriptiven Werte.

Tabelle 12: Deskriptive Werte für die Gruppen EG1 und EG2 in der produktiven Schriftkompetenz.

Gruppe	N	M	SD	SE
EG1	359	-.04	.92	.05
EG2	413	.03	.84	.04

*Textkompetenz rezeptiv und produktiv.* Wie in MS 3 wurden die rezeptiven und produktiven Textkompetenzen zu einem frühen Zeitpunkt zu einem Wert zusammengefasst, der sich nicht mehr differenzieren lässt. In MS 4 waren dies die Aufgaben „Globales und detailliertes Textverstehen“ und „Überschrift und Postkarte schreiben“. Tabelle 13 zeigt die deskriptiven Werte, die sich signifikant unterscheiden ( $t(707.7)=2.69$ ;  $p<.01$ ): Die EG2 erzielt höhere Werte als die EG1.

Tabelle 13: Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven und produktiven Textkompetenz.

Gruppe	N	M	SD	SE
EG1	358	3.01	1.13	.06
EG2	410	3.22 *	.97	.05

### Meilenstein 5

*Schriftkompetenz rezeptiv.* Für die rezeptiven Fähigkeiten in MS 5 wurden die Aufgaben „Richtige Wörter finden, genaues Lesen“ und „Sätze beurteilen, genaues und schnelles Lesen (mit Zeitvorgabe)“ einer Faktorenanalyse unterzogen. Eine Varianzaufklärung von nur 35.8 % und eine Interkorrelation von  $r=.36$  schien jedoch zu gering, um die Variablen zu einem Faktorwert zusammenzufassen. Einzelanalysen der beiden Aufgaben hinsichtlich der beiden Gruppen EG1 und EG2 zeigten jeweils keine signifikanten Unterschiede ( $t(771)=1.15$ ;  $p=.245$  bzw.  $t(622.5)=1.77$ ;  $p=.077$ ). Lediglich die Varianzen zeigten sich unterschiedlich zugunsten der EG1 bei der Aufgabe „Sätze beurteilen“. Deskriptive Werte zeigt Tabelle 14.

Tabelle 14: Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven Schriftkompetenz.

Aufgabe	Gruppe	N	M	SD	SE
Richtige Wörter finden - genaues Lesen M5	EG1	347	7.03	1.39	.08
	EG2	426	7.14	1.39	.07
Sätze beurteilen - genaues und schnelles Lesen (mit Zeitvorgabe) M5	EG1	345	7.29	1.43	.08
	EG2	424	7.46	1.07	.05

*Schriftkompetenz produktiv.* Die produktive Schriftkompetenz wurde durch die Aufgabenformate „Wörter schreiben“ und „Lückendiktat, richtig schreiben“ gemessen. Die Korrelation der beiden Aufgaben lag bei  $r=.64$ . Ein Faktor der durchgeführten Faktorenanalyse erklärte 63.5 % der Gesamtvarianz. Unterschiede zwischen den Gruppen hinsichtlich der Mittelwerte der Faktorwerte zeigten sich keine ( $t(759)=0.72$ ;  $p=.474$ ). Tabelle 15 zeigt die deskriptiven Werte.

Tabelle 15: Deskriptive Werte für die Gruppen EG1 und EG2 in der produktiven Schriftkompetenz.

Gruppe	N	M	SD	SE
EG1	344	-.03	.92	.05
EG2	417	.02	.85	.04

*Textkompetenz rezeptiv und produktiv.* Wie in MS 4 wurden die rezeptiven und produktiven Textkompetenzen mit Hilfe der gleichen Aufgabenformate gemessen und zusammengefasst. Tabelle 16 zeigt die deskriptiven Werte, die sich signifikant unterscheiden ( $t(695.8)=3.21$ ;  $p<.01$ ): Die EG2 erzielte höhere Werte als die EG1.

Tabelle 16: Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven und produktiven Textkompetenz.

Gruppe	N	M	SD	SE
EG1	340	4.89	2.20	.12
EG2	422	5.39 *	2.01	.10

### Meilenstein 6

*Schriftkompetenz rezeptiv.* Für die rezeptiven Fähigkeiten in MS 6 wurden die Aufgaben „Richtige Wörter finden, schnelles Lesen“ und „Sätze beurteilen, genaues und schnelles Lesen (mit Zeitvorgabe)“ und „Fehler erkennen, Wissen über Orthografie“ einer Faktorenanalyse unterzogen. Eine Varianzaufklärung von nur 39.4 % und eine mittlere Interkorrelation von  $r=.38$  (s. Tabelle 17) schien jedoch zu gering, um die Variablen zu einem Faktorwert zusammenzufassen. Einzelanalysen der Aufgaben hinsichtlich der beiden Gruppen EG1 und EG2 zeigten

signifikanten Unterschiede für die Aufgabe Sätze beurteilen, bei der die EG2 besser abschnitt (s. Tabelle 18).

Tabelle 17: Korrelationsmatrix

	Schreiben: Fehler erkennen M6	Lesen: Sätze beurteilen M6
Lesen: Richtige Wörter finden, schnelles Lesen	.32	.48
Schreiben: Fehler erkennen		.34

Tabelle 18: Deskriptive und inferenzstatistische Werte für die Gruppen EG1 und EG2 in der rezeptiven Schriftkompetenz.

	Gruppe	N	M	SD	SE	t	p
Lesen: Richtige Wörter finden, schnelles Lesen M6	EG1	329	9.64	2.20	.12	1.81	.071
	EG2	416	9.91	1.78	.09		
Schreiben: Fehler erkennen M6	EG1	333	3.44	2.02	.11	.30	.768
	EG2	411	3.48	1.96	.10		
Lesen: Sätze beurteilen M6	EG1	330	6.65	1.90	.11	2.37*	.018
	EG2	414	6.95	1.48	.07		

Anmerkung: \* =  $p < .01$

*Schriftkompetenz produktiv.* Die Aufgabenformate in MS 6 entsprechen denen des MSs 5, korrelierten mit  $r = .64$  und wurden mit Hilfe einer Faktorenanalyse zu einem Faktor „produktive Schriftkompetenz“ zusammengefasst (Varianzaufklärung 63.5%). Der Faktorwert der Gruppen EG1 und EG2 unterschied sich nicht signifikant ( $t(719) = 1.53$ ;  $p = .126$ ). Tabelle 19 zeigt die deskriptiven Werte des Faktors.

Tabelle 19. Deskriptive Werte für die Gruppen EG1 und EG2 in der produktiven Schriftkompetenz.

Gruppe	N	M	SD	SE
EG1	313	.06	.85	.05
EG2	408	-.04	.91	.05

*Textkompetenz rezeptiv und produktiv.* Wie in MS 3 wurden die rezeptiven und produktiven Textkompetenzen zusammengefasst. Signifikante Unterschiede zeigten sich nicht ( $t(720) = 0.89$ ;  $p = .372$ ). Tabelle 20 zeigt die deskriptiven Werte.

Tabelle 20: Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven und produktiven Textkompetenz.

Gruppe	N	M	SD	SE
EG1	316	6.06	1.83	.10
EG2	406	6.18	1.77	.09

### Meilenstein 7

*Schriftkompetenz rezeptiv.* Für die rezeptiven Fähigkeiten in MS 7 wurden die Aufgaben „Richtige Wörter finden, schnelles Lesen“ und „Fehler erkennen, Wissen über Orthografie“ einer Faktorenanalyse unterzogen. Eine Varianzaufklärung von nur 30.6 % und eine Korrelation von  $r=.31$  erlaubte kein Zusammenfassen zu einem Faktor. Einzelne t-Tests konnten keine signifikanten Unterschiede ausmachen (s. Tabelle 21).

Tabelle 21: Deskriptive und inferenzstatistische Werte für die Gruppen EG1 und EG2 in der rezeptiven Schriftkompetenz.

						t	df	p
	Gruppe	N	M	SD	SE			
Lesen: Richtige Wörter finden, schnelles Lesen M7	EG1	304	10.71	1.00	.06	1.55	540.2	.122
	EG2	408	10.82	.75	.04			
Schreiben: Fehler erkennen M7	EG1	304	4.21	2.32	.13	.19	708	.850
	EG2	406	4.17	2.21	.11			

*Schriftkompetenz produktiv.* Die Aufgabenformate in MS 7 entsprechen denen in MS 5 und korrelierten mit  $r=.67$  und wurden mit Hilfe einer Faktorenanalyse zu einem Faktor „produktive Schriftkompetenz“ zusammengefasst, der 66.7% der Varianz aufklärt. Der Faktorwert der Gruppen EG1 und EG2 unterschied sich nicht signifikant ( $t(706)=0.51$ ;  $p=.610$ ). Tabelle 22 zeigt die deskriptiven Werte des Faktors.

Tabelle 22: Deskriptive Werte für die Gruppen EG1 und EG2 in der produktiven Schriftkompetenz.

	Gruppe	N	M	SD	SE
Schriftkompetenz produktiv M7 Faktor	Material + Stunden	302	-.02	.94	.05
	nur Material	406	.01	.86	.04

*Textkompetenz rezeptiv und produktiv.* Kein signifikanter Unterschied zeigte sich beim Textverstehen ( $t(707)=0.84$ ;  $p=.402$ ). Die deskriptiven Werte zeigt Tabelle 23.

Tabelle 23: Deskriptive Werte für die Gruppen EG1 und EG2 in der rezeptiven und produktiven Textkompetenz.

Gruppe	N	M	SD	SE
Material + Stunden	EG1	4.72	1.88	.11
nur Material	EG2	4.84	1.80	.09

## Zusammenfassung

Insgesamt zeigten einige Aufgaben signifikante Unterschiede der Gruppen EG1 und EG2. Bei signifikanten Unterschieden erreichte die EG2 bis auf einmal die höheren Werte. So erreichten die Schülerinnen und Schüler der EG2 bereits zum ersten Zeitpunkt in einigen Aufgaben höhere Werte, so in „Namen schreiben“, „Laut-Buchstabenzuordnung“, „Anlaute“ und „Texte verstehen“. Mit fortschreitender Zeit (hin zu MS 7) werden für zunehmend weniger Aufgaben die Unterschiede zwischen den Gruppen signifikant (s. Tabelle 24). In MS 7 sind keine Unterschiede mehr zu verzeichnen. Anzumerken ist jedoch, dass die Effekte in den meisten Fällen marginal sind.

Tabelle 24: Übersicht über signifikante und nicht signifikante Unterschiede der einzelnen Aufgabenblöcke in den verschiedenen Meilensteinen.

	MS 1	MS 2	MS 3	MS 4	MS 5	MS 6	MS 7
Schriftkompetenz rezeptiv	Namen schreiben; Laut-Buchstabenzuordnung;	n. sig.	*; EG2	*; EG2	n. sig.	Sätze beurteilen: *; EG2; andere n. sig.	n. sig.
Schriftkompetenz produktiv	Anlaute; Texte verstehen	***; EG2	***; EG2	n. sig.	n. sig.	n. sig.	n. sig.
Textkompetenz rezeptiv und produktiv		n. sig	*; EG1	**; EG2	**; EG2	n. sig.	n. sig.

## Fazit zu den Aufgaben der Meilensteine

Ziel der Untersuchung war es, zu evaluieren, ob das diagnostische Instrument Lesen macht stark mit zusätzlichen Lehrerstunden einen positiveren Einfluss auf schriftsprachliche Kompetenzen hat als der Einsatz des Instrumentes ohne diese Stunden. Hierzu wurde verglichen, ob die Schülerinnen und Schüler in den Aufgaben aus Lesen macht stark besser abschneiden, wenn die Lehrerinnen und Lehrer Zusatzstunden erhalten haben, als wenn die Lehrerinnen und Lehrer diese Stunden nicht erhalten haben.

Insgesamt lässt sich der vermutete Effekt nicht zeigen. Eine längsschnittliche Untersuchung der Leistungszuwächse war durch die Konstitution der Testverfahren nicht möglich, da



beispielsweise keine Ankeritems verwendet wurden. Dies stellt eine Hürde für die Evaluation dar, ist aber keine zu bemängelnde Eigenschaft des LMS Materials. Über die Meilensteine hinweg werden immer weniger Unterschiede signifikant. Das lässt darauf schließen, dass die Schulen mit zusätzlichen Stunden nicht besser abschneiden als die Schulen ohne zusätzliche Stunden. Die Aussage, dass sich die Schulen damit angleichen ist auf Basis der Datenlage nicht möglich. Diese Befunde werden durch die Befunde aus dem standardisierten Test TEPHOBE bestärkt, der ebenfalls keine Unterschiede zwischen den Gruppen zeigen konnte.

### Strukturelle Unterschiede zwischen den an der Evaluation beteiligten Schulen

Um weitere relevante Daten zu den Schulen zu erhalten, wurden am Ende der Evaluation leitfadengestützte, telefonische Interviews mit den beteiligten Lehrkräften durchgeführt. Zentrales Ziel dieser Befragung war es, detaillierte Informationen zu den Förderkonzepten der Schulen zu erhalten. Daneben wurde bspw. erfragt, ob die Lehrkräfte die Fakultas für den Deutschunterricht haben oder welche Lehrwerke für den Schriftspracherwerb genutzt wurden.

Die Auswertung der Telefoninterviews mit den Schulen zeigen eine Vergleichbarkeit der Evaluations- und Kontrollgruppe hinsichtlich vieler Aspekte. So werden beispielsweise ähnliche Lehrmaterialien für die Schriftsprachvermittlung verwendet, der Deutschunterricht wird nur in wenigen Fällen von Lehrkräften ohne entsprechende Fakultas erteilt, etwa an der Hälfte der Schulen kann im Deutschunterricht zur Umsetzung von Förderung regelmäßig mit einer doppelten Lehrkraftbesetzung gearbeitet werden sowie fast alle Schulen erstellen und setzen eigene Förderpläne um.

Im Rahmen der Telefoninterviews wurden allerdings auch zwei Unterschiede zwischen der Evaluations- und der Kontrollgruppe deutlich. Zum einen ist die Größe der VERA-Risikostufe an den Schulen der Kontrollgruppe homogener als an den Evaluationsschulen. Das heißt, der Anteil der Schülerinnen und Schüler, der anhand der VERA-Daten einer Risikogruppe zugeordnet wird, variiert bei den Evaluationsschulen stark, während dieser Wert bei den Schulen der Kontrollgruppe kaum Auffälligkeiten zeigt. Zum anderen geben deutlich mehr Schulen der Kontrollgruppe (90%) als der Evaluationsgruppe (50%) an, dass eine systematische Vernetzung aller an der Förderung beteiligten Lehr- und Fachkräfte an der jeweiligen Schule stattfindet. Dies könnte ein Indikator dafür sein, dass an den Schulen der Kontrollgruppe der Förderung ein höherer Stellenwert beigemessen wird und dies zu einer stärkeren Professionalisierung der Lehrkräfte der Schulen sowie der Schulen als Institutionen in diesem Bereich führt. Diese Faktoren könnten potenziellen Interventionseffekten entgegengewirkt haben.

### III. Implementation: Ergebnisse der Lehrerbefragung zur Tauglichkeit des Materials

#### *Gelingt die **Implementation** des Materials in die schulische Praxis? (Implementation)*

Im Bereich der Implementation wird die Praxistauglichkeit des Instruments „Lesen macht stark“ sowie seine Akzeptanz bei den Lehrkräften untersucht. Um zu erfahren, wie das Projekt an den einzelnen Schulen gestaltet wird, wurde nach zwei Jahren eine schriftliche Befragung an den beteiligten Schulen durchgeführt. Aus den 20 Schulen, die die „Lesen macht stark“-Materialien bearbeitet haben, beantworteten 24 Lehrkräfte – ausschließlich Lehrerinnen – den Fragebogen. Vorgestellt werden im Folgenden einige zentrale Ergebnisse.

#### Nutzung und Akzeptanz des Materials

Etwa 71% (17 Personen) gaben an, das Material hilfreich bis sehr hilfreich zu finden. 15 Personen (63%) gaben an, das Material hilfreich für die Identifizierung von Kindern, die Schwächen im Lesen und Schreiben haben, für 16 Personen (67%) ist das Material hilfreich bei der Einschätzung der Leistung einzelner Schülerinnen und Schüler. 11 Personen gaben an, das Material sei hilfreich für die Einschätzung des Niveaus der Klasse. Nur einer Lehrerin half das Material beim Vergleich zweier Schulklassen. Die Ergebnisse sind positiv zu bewerten, da der Zweck des Materials, bestimmte Schülerinnen und Schüler zu identifizieren, mit den Ansichten der Lehrkräfte übereinstimmt und das Material insgesamt als sehr hilfreich gesehen wird.

Über die Frage, ob das Material hilfreich war hinaus, wurde nach dem konkreten Nutzen gefragt. 83% der Befragten gaben an, die Ergebnisse mehrerer Meilensteine des Programms für die Einschätzung des individuellen Lernverlaufs der Schülerinnen und Schüler zu nutzen. Vorwiegend genutzt, so die Befragten, wurden die Ergebnisse zudem für Elterngespräche und für die Erstellung eines Lernplans. Die Ergebnisse dienten auch zum Austausch mit den Kolleginnen, nicht jedoch zur Argumentation bei der Beantragung von Unterstützung. Hier gaben nur 4 Lehrerinnen an, dies sehr häufig bzw. immer zu tun.

Ziel des „Lesen macht stark-Materials“ ist es, die Lehrkräfte dabei zu unterstützen, Schwierigkeiten der Kinder beim Schriftspracherwerb frühzeitig zu erkennen, damit Sie die Kinder im Erwerbsprozess gezielt unterstützen können. Daher wurde erfragt, welche Informationen die Lehrkräfte für eine Förderentscheidung nutzen. Alle Lehrerinnen nutzen ihre eigenen Beobachtungen. Bis auf eine Lehrerin geben alle Lehrerinnen an, die Informationen aus den Klas-senaufgaben von „Lesen macht stark“ zu nutzen, bis auf 4 Lehrerinnen gaben alle an, Informationen aus den Einzelaufgaben zu nutzen. Die Werte liegen höher als bei den Fragen nach standardisierten Tests: Hier gaben nur 2/3 der Lehrerinnen an, Tests zu benutzen. Allerdings geben auch nahezu alle Lehrerinnen an, Schreibproben und eigene Arbeitsblätter als Informationsquelle zu nutzen.

Schließlich wurden die Befragten nach ihrer Meinung zu verschiedenen Aspekten befragt. Tabelle 25 zeigt die Minima und Maxima sowie die Mediane für die einzelnen Fragen im Überblick. Die Pole der 5-stufigen Skala lagen bei (1)=trifft überhaupt nicht zu bis (5) = trifft voll und

ganz zu. Die Zwischenwerte wurden nur durch die Werte 2, 3 und 4 angegeben, nicht jedoch sprachlich. Die Fragen wurden von 23 der 24 Lehrerinnen beantwortet.

Nur 35% der Befragten (8 Personen) gaben an, wieder mit „Lesen macht stark“ arbeiten zu wollen. Die gleiche Anzahl lehnte dies ab. 30% waren sich nicht einig darüber. Dennoch gaben 61% an, durch die Materialien in der Schülerleistung unterstützt zu werden. Offensichtlich halten einige Lehrkräfte diese Unterstützung nicht für notwendig.

Die beiden Fragen, die die geringsten Werte zeigten, beziehen sich einerseits auf die Akzeptanz im Kollegium und andererseits auf die eigenen fachlichen Erkenntnisse. Hier lag der Median nur bei 2. Die Frage nach dem Wunsch nach Material für weitere Klassenstufen polarisierte am meisten: 52% waren eher dagegen, 44% eher dafür. Eine Mitte gab es kaum.

Während die Akzeptanz bei den Lehrerinnen eher mittelmäßig war, wurde sie von den Lehrerinnen bei den Schülerinnen und Schülern als hoch eingeschätzt: 61% gaben an, die Schülerinnen und Schüler arbeiten gerne mit dem Material. Zieht man die Skalenmitte hinzu, ergibt sich sogar ein Wert von 91%. Auch können die Schülerinnen und Schüler sich nach Meinung der Lehrkräfte mit den Figuren identifizieren und sie verstehen, was sie zu tun haben.

Die Lehrerinnen gaben mehrheitlich an, dass die das Material übersichtlich, optisch ansprechend und in der Handhabung praktisch gestaltet fanden. 4 Lehrerinnen fanden das Material zu schwierig.

Tabelle 25: Deskriptive Werte der Bewertungen verschiedener Aussagen durch die Lehrkräfte

Aussage	Min	Max	Median
Ich würde wieder mit LMS arbeiten.	1	5	3
Ich fühle mich durch die Materialien in meiner Einschätzung der Schülerleistung unterstützt.	1	5	5
Ich wünsche mir LMS auch für weitere Klassenstufen.	1	5	2
Ich würde LMS Kollegen empfehlen.	1	5	3
Ich würde das Material für meine Schule bestellen.	1	5	3
Die Akzeptanz im Kollegium ist hoch.	1	5	2
Die häufige Erfassung der Lese- und Schreibfähigkeit der SuS finde ich sinnvoll.	1	5	4
Kinder mit Schwächen im Lesen und Schreiben können durch den Einsatz von LMS schneller erkannt und somit früher gefördert werden.	1	5	3
LMS sollte auch Material enthalten, mit dem ein Lernverlauf eines Kindes und der ganzen Klasse einfach festgehalten werden kann.	1	5	3
Durch die häufige Erfassung der Lese- und Schreibfähigkeiten wird mein Unterricht sehr gestört.	1	5	3
Die SuS arbeiten gern mit dem LMS Material.	1	5	4
Die SuS können sich mit den Figuren identifizieren.	1	5	4
Die SuS empfinden die Arbeit als Test.	1	5	4
Die Aufgaben sind verständlich erklärt.	3	5	4
Die SuS verstehen was sie zu tun haben.	3	5	5
Das LMS-Material ist übersichtlich gestaltet.	3	5	5
Das LMS-Material ist optisch ansprechend gestaltet.	3	5	5
Die Handhabung des LMS-Materials ist praktisch.	2	5	5
Die Aufgaben sind zu schwierig.	1	4	3
Das Niveau der Aufgaben schwankt innerhalb eines Meilensteins.	1	5	3

Aussage	Min	Max	Median
Ich habe durch den Einsatz des LMS-Materials neue Erkenntnisse über meine Schülerinnen und Schüler gewonnen.	1	5	3
Ich habe durch den Einsatz des LMS-Materials neue Erkenntnisse über die Inhalte (Lesen, Schreiben & Textverständnis) gewonnen.	1	5	2

## Nutzung der zusätzlichen Lehrerwochenstunden

Zehn Schulen wurden zusätzliche Lehrerstunden zur Umsetzung des „Lesen macht stark“ Programms zur Verfügung gestellt. In einer Befragung dieser Lehrkräfte wurde ermittelt, wie die Stunden eingesetzt wurden. Konkret gefragt wurde „Wie werden die zusätzlichen Lehrerwochenstunden an Ihrer Schule genutzt?“ Tabelle 26 zeigt das Antwortmuster von 10 der insgesamt 12 Lehrerinnen, die den Fragebogen beantwortet haben. 9 der 10 Probandinnen und Probanden gaben an, die Stunden als Förderstunden für diejenigen Schülerinnen und Schüler zu nutzen, die eine schwache Leistung in den Einzelaufgaben zeigten. Nur eine Lehrerin gab an, dass die Zusatzstunden für das Projekt nicht notwendig sind.

Tabelle 26: Angaben zur Nutzung der Lehrerwochenstunden (n=10 Lehrkräfte)

Antwort	Anzahl der Lehrkräfte
Nutzung der zusätzlichen Lehrerwochenstunden: Zur Durchführung der Einzelaufgaben	8
Nutzung der zusätzlichen Lehrerwochenstunden: Als zusätzliche Förderstunden	4
Nutzung der zusätzlichen Lehrerwochenstunden: Zum Unterrichten in Tandems	6
Nutzung der zusätzlichen Lehrerwochenstunden: Zur Durchführung der Klassenaufgaben	4
Nutzung der zusätzlichen Lehrerwochenstunden: Zur Förderung der Schülerinnen und Schüler, die in der EA eine schwache Leistung zeigten.	9
Nutzung der zusätzlichen Lehrerwochenstunden: Zur personell verstärkten Differenzierung des Unterrichts	4
Diese zwei zusätzlichen Lehrerwochenstunden sind für das LMS-Projekt nicht unbedingt notwendig	1

## Fazit

Der vorgestellten Evaluationsstudie zu „Lesen macht stark – Grundschule“ lagen drei Fragen zugrunde: Zum einen die Frage danach, ob das Material geeignet ist, um Kinder mit Schwierigkeiten beim Schriftspracherwerb zu identifizieren. Zum anderen die Frage nach der Eignung des Materials, um diese Kinder in ihrem Lernprozess zu unterstützen. Und schließlich die Frage, ob die Implementation des Materials in der schulischen Praxis gelingt.

Die Ergebnisse zur Frage danach, ob das Material geeignet ist, um Kinder mit Schwierigkeiten beim Schriftspracherwerb zu identifizieren, lassen sich wie folgt zusammenfassen:

Der Abgleich der Ergebnisse des externen standardisierten Tests mit den Ergebnissen der Screeningaufgaben aus dem LMS-Material zeigte eine große Schnittmenge. Das heißt, mehrheitlich wurde von beiden Verfahren für dieselben Kinder ein Förderbedarf ermittelt. Allerdings gab es auch Kinder, für die das externe Testverfahren einen Förderbedarf feststellte, die LMS-Aufgaben hingegen nicht. Das zeigt, dass das LMS-Material seine Screeningfunktion erfüllt, die Ergebnisse allerdings, wie zu erwarten, nicht mit denen eines standardisierten Testverfahrens gleichgestellt werden können. Dafür ermöglicht es jedoch ein ökonomisch einsetzbares Screening im Regelunterricht, das durch Bearbeitung von Einzelaufgaben abgesichert und kontinuierlich und in unterschiedlichen Teilbereichen des Schriftspracherwerbs durchgeführt wird. Diese Aspekte der Ökonomie und der Abbildung von dynamischen Lernverläufen in verschiedenen Teilbereichen des Schriftspracherwerbs, bieten einen Mehrwert gegenüber dem standardisierten Testverfahren.

Hinsichtlich der Frage, ob das Material geeignet ist, um Kinder mit Schwierigkeiten beim Schriftspracherwerb in ihrem Lernprozess zu unterstützen, lässt sich Folgendes festhalten:

Für das Leseverstehen gemessen mit Hilfe eines standardisierten Tests gilt, dass beim Verstehen auf der Wort- und Satzebene sich zwar zeigte, dass die EG2 und die Kontrollgruppe besser abschnitt als die EG1, die Veränderung vom ersten zum zweiten Zeitpunkt jedoch gleich war, so dass das Material (mit und ohne Lehrerstunden) nicht besser gewirkt hat als in der Kontrollgruppe ohne Material. Das gleiche gilt für das Textverstehen, das nur beim zweiten Zeitpunkt getestet wurde. Auch hier zeigten sich keine Unterschiede zwischen den drei Gruppen.

Für das Rechtschreiben zeigte sich, dass sich bei der Auswertung des SLRT je nach Fehlerart unterschiedliche Verbesserungen ergaben. Bei den lautgetreuen Fehlern sind die Unterschiede zwischen den Gruppen marginal. Die insgesamt größte Steigerung zeigte sich bei der Gruppe EG1 bei den nicht lautgetreuen Fehlern.

Insgesamt sind die Ergebnisse bislang uneindeutig und bieten die Möglichkeit weitgehender Interpretation.

Für die diagnostischen Aufgaben des Materials zeigten sich signifikante Unterschiede der Gruppen EG1 und EG2. Bei signifikanten Unterschieden erreichte die EG2 bis auf ein Mal die höheren Werte. So erreichten die Schülerinnen und Schüler der EG2 bereits zum ersten Zeitpunkt in einigen Aufgaben höhere Werte. Mit fortschreitender Zeit (hin zum letzten Meilenstein) wurden für zunehmend weniger Aufgaben die Unterschiede zwischen den Gruppen signifikant. Am Ende der Intervention schnitten die Schulen der EG1 nicht besser ab als die Schulen der EG2, ein Effekt der zusätzlichen Lehrkraftstunden konnte nicht festgestellt werden.

Die Ergebnisse hinsichtlich der Frage, ob die Implementation des Materials in der schulischen Praxis gelingt, lassen sich wie folgt zusammenfassen:

Aussagen zur Akzeptanz und zur Nutzung des Materials zeigten widersprüchliche Aussagen. Während etwa 2/3 der Lehrkräfte angaben, das Material hilfreich zu finden, um schwache

Schülerinnen und Schüler zu identifizieren, würden es nur 1/3 gerne weiterhin einsetzen und nur die Hälfte würde es gerne für weitere Klassenstufen weiterentwickelt sehen. Daneben geben jedoch jeweils etwa 2/3 der Lehrkräfte an, dass sie das Material zur Feststellung von Förderbedarf der Kinder einsetzen, es zur Beobachtung und Bestimmung des Lernstands der Kinder hilfreich bis sehr hilfreich finden. Darüber hinaus geben ebenfalls 2/3 der Lehrkräfte an, dass sie die Förderhinweise des Materials hilfreich bis sehr hilfreich finden. Insgesamt kann also festgestellt werden, dass das Material sowohl als Diagnoseinstrument als auch als Fördermaterial demnach vom Großteil der Lehrkräfte als hilfreich bewertet wird. Auf Basis der erhobenen Daten kann nicht nachvollzogen werden, ob die Nachfrage nach LMS mit dem durchschnittlichen Leistungsniveau der Schülerinnen und Schüler in Verbindung steht.

Hinsichtlich der Optik und der Praktikabilität gefällt das Material den Lehrkräften.

Die Lehrkräfte, die zusätzliche Wochenstunden erhalten haben, nutzen sie weitgehend zur Förderung der durch die Lesen macht stark-Aufgaben ermittelten Kinder. Es lagen jedoch keine Daten vor, inwieweit die Förderhinweise aus LMS in den jeweiligen Stunden Einsatz gefunden haben. Es ist daher nicht möglich, die Zusatzstunden einzig als Interventionszeit im Rahmen des Projekts zu bewerten.

Um die verbliebenen offenen Fragen aus den Ergebnissen dieser Evaluationsstudie auflösen zu können sowie den Förderaspekt des Instruments zu stärken, wurde ein ergänzendes systematisches Trainingsprogramm entwickelt und eine zweite Evaluationsstudie („Lesen macht stark – Training“) angeschlossen. Ihre Ergebnisse werden im nun folgenden zweiten Teil dieses Berichts vorgestellt.



**Lesen**  
macht stark

Dr. Simone Jambor-Fahlen, Dr. Stefanie Bredthauer, Dr. Kathrin Hippmann, Dr. Steffani Saß, Prof. Dr. Michael Becker-Mrotzek, Prof. Dr. Olaf Köller

Teil B: Ergebnisse der wissenschaftlichen Begleitung  
zu „Lesen macht stark – Training“

## Inhaltsverzeichnis

<b>1. Präambel</b> .....	<b>4</b>
<b>2. Kurzbeschreibung des Trainings und der Erhebungsinstrumente</b> .....	<b>5</b>
2.1 Kurzbeschreibung Trainingsablauf.....	5
2.2 Kurzbeschreibung der Evaluationsstudie.....	6
2.3 Instrumente.....	8
2.3.1 Instrumente zur Erfassung der Kontrollvariablen und der Leistungsentwicklung.....	8
2.3.2 Instrumente zur Erfassung der Implementation Fidelity.....	10
<b>3. Zusammenfassung der Ergebnisse der Interventions- und Implementationsstudie „Lesen macht stark – Training“</b> .....	<b>12</b>
3.1 Ergebnisse der Interventionsstudie.....	12
3.2 Ergebnisse der Implementationsstudie.....	12
3.2.1 Durchführung der Diagnostik und Förderung.....	12
3.2.2 Beobachtung des Unterrichts.....	12
3.2.3 Akzeptanz der Lehrkräfte.....	13
<b>4. Die Ergebnisse im Detail</b> .....	<b>13</b>
4.1 Teil A: Ergebnisse der Leistungsentwicklung.....	14
4.1.1 Stichprobenbeschreibung und Auswahl der Gruppen für die Analysen.....	14
4.1.2 Vergleich der beiden Gruppen hinsichtlich der Kontrollvariablen.....	15
4.1.3 Vergleich der beiden Gruppen hinsichtlich der Fördermaßnahmen.....	18
4.2 Teil B: Umsetzung der Diagnostik und des Trainings.....	22
4.2.1 Wiedergabetreue des Konzeptes.....	22
4.2.2 Bewertung der Fortbildung.....	41
4.2.3 Bewertung des Konzeptes und des Materials durch die Lehrkräfte.....	45
<b>5. Diskussion</b> .....	<b>62</b>
<b>6. Fazit</b> .....	<b>70</b>
<b>7. Literatur</b> .....	<b>70</b>
<b>8. Anhang</b> .....	<b>72</b>
8.1 Darstellung der Ergebnisse des Vergleichs der Experimentalgruppe mit der Kontrollgruppe.....	72
8.2 Darstellung der Ergebnisse des Vergleichs der Trainingsgruppe mit der Experimentalgruppe ohne Training und der Kontrollgruppe.....	77
8.3 Vergleich der Trainingsgruppe mit der Kontrollgruppe und der Experimentalgruppe ohne Training hinsichtlich der Fördermaßnahmen.....	80
8.4 Kurzdarstellung weiterer Analysen.....	83
8.5 Zusammenfassung der zusätzlichen Analysen.....	84



8.6 Interview Schulen..... 86

## 1. Präambel

Die drei Kooperationspartner – das Institut für Qualitätsentwicklung an Schulen Schleswig-Holstein (IQSH), das Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache der Universität zu Köln (MI) und das Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik an der Universität Kiel (IPN) – haben mit „Lesen macht stark – Grundschule“ (Lms-Diagnostik) ein innovatives Instrument zur Diagnose und Förderung der Lese- und Schreibkompetenz in der Grundschule für die Jahrgänge 1 bis 4 entwickelt. Das Instrument wurde zunächst pilotiert, um seine grundsätzliche Eignung zu ermitteln. Ab den Schuljahren 2014/15 und 2015/16 wurde es in den Jahrgängen 1 und 2 an 110 Schulen in Schleswig-Holstein eingeführt und in einer großen Studie (N = 1555) evaluiert, um seine Wirkung zu erheben. Die Darlegung der Ergebnisse aus dieser Studie sind in Teil 1 dieses Berichts zu finden.

Da die Evaluation des Instruments in der Schulpraxis durchgeführt wurde, kam es zu unvermeidbaren Einschränkungen bei der Durchführung. Zum Beispiel war eine zufällige Stichprobenziehung nicht möglich.

Die Leistungsdaten zeigten keinerlei signifikante Vorteile für die Kinder der Experimentalgruppe. Gründe hierfür werden in einer unsystematischen Förderung vermutet. Die Lehrkräfte, die zusätzliche Wochenstunden erhalten haben, nutzten sie weitgehend zur Förderung der durch die Lms-Diagnostik ermittelten Kinder. Es lagen jedoch keine Daten vor, inwieweit die Förderhinweise aus der Lms-Diagnostik in den jeweiligen Stunden Einsatz gefunden haben oder andere Fördermittel genutzt wurden. Auch die Förderzeiten und Inhalte der Förderung wurden nicht erhoben. Insofern bleibt allgemein unklar, welchen Einfluss die Förderung auf die Leistungen der Kinder hatte. Es ist zu vermuten, dass sich dadurch die marginalen Unterschiede in den jeweiligen Gruppen erklären lassen. In der Folge wurde ergänzend ein systematisches Trainingsprogramm entwickelt: Da das Diagnose- und Förderinstrument „Lesen macht stark – Grundschule“ in erster Linie ein Screeningverfahren der Lese- und Schreibkompetenzen darstellt, enthalten die Hefte zwar Förderhinweise, aber kein systematisches Trainingsprogramm. Diese Lücke soll das Fördertraining „Lesen macht stark – Training“ (Lms-Training) schließen. Durch häufige Wiederholungen und viele kurze Trainingseinheiten soll die Automatisierung des Lesens und Schreibens von Buchstaben, Silben und Wörtern systematisch gefördert werden. Die Wirkung des Trainingsprogramms sollte ebenfalls empirisch belegt werden. Aus diesem Grund wurde eine Folgestudie initiiert, die einerseits die Konzepttreue (Implementation Fidelity) und andererseits die Wirksamkeit des Trainings untersuchte.

Die Studie startete im Schuljahr 2018/19 mit insgesamt 24 Klassen und endete coronabedingt im Schuljahr 2019/20 nach dem ersten Halbjahr (20.08.18, Beginn 1. Klasse im Schuljahr

18/19) – 31.01.20, Ende 1. Halbjahr, 2. Klasse im Schuljahr 19/20). Nach diesem Zeitraum führten die Lehrkräfte keine Trainings mehr durch.

Der Ablauf orientierte sich an der ersten Evaluation von „Lesen macht stark – Grundschule“ (s.o.). Jedoch gibt es keine dritte Gruppe mit zusätzlichen Stunden für die Schulen, da bei dieser Studie seitens des Schulministeriums keine zusätzlichen Stunden zur Verfügung gestellt wurden. Diese Studie dient als Nacherhebung und hat zudem das Ziel, die Implementation Fidelity (Konzepttreue) zu untersuchen. Um eine Vergleichbarkeit zu der vorangegangenen Evaluation herzustellen, wurden die Testungen, Fragebögen und Meilensteine analog eingesetzt und um weitere Erhebungsinstrumente, die Implementation Fidelity betreffend, ergänzt. Die Implementation vom Lms-Training wurde so fortwährend überprüft.

## 2. Kurzbeschreibung des Trainings und der Erhebungsinstrumente

### 2.1 Kurzbeschreibung Trainingsablauf

Bei dem Förderkonzept handelt es sich um ein strukturiertes, sequenziertes und individualisiertes Training. Im Fokus stehen die basalen Lese- und Schreibfertigkeiten. Alle Trainingseinheiten gliedern sich in fünf Phasen. In jeder Trainingseinheit durchlaufen die Schülerinnen und Schüler die Phasen Herstellung der Zieltransparenz (1), Aktivierung des Vorwissens (2), Instruktion (3), Automatisierung (4) und Formatives Feedback (5). Das Lernen im Tandem bildet ein zentrales Element des gemeinsamen Trainierens. Die Konzeption des Fördertrainings erfolgte auf Basis von Forschungserkenntnissen zu den o.g. Aspekten des Lernens.

Basierend auf den Ergebnissen von der Lms-Diagnostik werden jeweils sechs bis acht Kinder einer Trainingsgruppe zugeteilt. Insgesamt wurden fünf Trainingsbausteine durchgeführt: die Bausteine 1, 2 und 5 zum Lesen und die Bausteine 3 und 4 zum Schreiben. Mehrheitlich hat dieselbe Lehrkraft die Lms-Diagnostik und das Lms-Training in ihrer Klasse durchgeführt, vereinzelt hat die Klassenlehrkraft die Lms-Diagnostik und die Förderlehrkraft das Lms-Training durchgeführt.

Der Ablauf der Fördertrainings ist stets gleichbleibend:

1. Herstellung der Zieltransparenz: Die Lehrkraft stellt den Kindern in der Fördergruppe das zu erreichende Ziel explizit vor (Hattie & Beywl, 2013).
2. Aktivierung des Vorwissens: Die Lehrkraft aktiviert spielerisch die bereits gelernten Wörter (Simons, 1992).
3. Instruktion: Die Lehrkraft vermittelt explizit durch Demonstration der Handlungsschritte (*Modelling*, Philipp, 2012) die Lerninhalte.

4. Automatisierung: Durch mehrmalige kurze Übungseinheiten wird eine Automatisierung der basalen Lese- und Schreibfähigkeiten erreicht (Hasselhorn & Gold, 2017).
5. Formatives Feedback: Die Lehrkraft überprüft gemeinsam mit den Kindern inwieweit die jeweiligen Lernziele erreicht wurden und gibt ihnen eine Rückmeldung dazu (Black & William, 1998)

Für das Lesetraining gilt, dass mit dem chorischen Lesen im Tandem eine unterstützende Methodik gewählt wurde, bei der Schülerinnen und Schüler im Tandem (homogenes Leistungslevel) abwechselnd lesen, sich gegenseitig unterstützen sowie abschließend auch korrigieren. Die Trainingsergebnisse werden über einen Levelcheck überprüft und durch das formative Feedback der Lehrkraft begleitet. Das Schreibtraining folgt dem gleichen Muster (Zieltransparenz herstellen, Aktivierung von Vorwissen, Instruktion, Trainingsphase, Feedback) wie das Lesetraining. In der Trainingsphase wird ebenfalls im Tandem das automatisierte Schreiben von Buchstaben, Silben und Wörtern geübt.

Das Lms-Training ist unterteilt in Leistungslevels. Das jeweilige Tandem absolviert die einzelnen Level und meldet sich dann zum Levelcheck an. Ist der Check für beide Tandempartner bestanden, folgt das nächste Leistungslevel. Nach einer Einführungswoche in den Ablauf des Lms-Trainings agieren die Tandems autark, so dass sie bei jedem neuen Trainingssetting immer wieder dort starten, wo beim letzten Setting gestoppt wurde. Die Kinder trainieren viermal die Woche für 15 bis 20 Minuten.

## 2.2 Kurzbeschreibung der Evaluationsstudie

Ziel der Evaluation war es, zu untersuchen, ob das Lese- und Schreibtraining „Lesen macht stark“ einen positiven Einfluss auf die Entwicklung schriftsprachlicher Kompetenzen von Schülerinnen und Schülern hat. Dafür wurden Schülerinnen und Schüler, die am Lms-Trainingsprogramm teilnahmen (im Folgenden Trainingsgruppe genannt), mit Schülerinnen und Schülern, die an keinem Lms-Training teilnahmen (Kontrollgruppe und Experimentalgruppe ohne Lms-Training) hinsichtlich der Entwicklung schriftsprachlichen Kompetenzen verglichen. Die Trainingsgruppe wurde aus dem Klassenverbund extrahiert. Der Klassenverbund stellte die Experimentalgruppe dar. In der Experimentalgruppe, also mit der gesamten Klasse, wurde zunächst die Lms-Diagnostik durchgeführt. Mit den so identifizierten Kindern (Trainingsgruppe) wurde anschließend das neu entwickelte Trainingsprogramm durchgeführt. Zur Überprüfung der Wirksamkeit des Lese- und Schreibtrainings wurde ein Längsschnittdesign mit verschiedenen Erhebungen von standardisierten Testinstrumenten und einem Elternfragebogen geplant. Zum einen wurden relevante Kontrollvariablen zur Überprüfung der Vergleichbarkeit der Kontroll- und der Experimentalgruppe erfasst und zum anderen wurden standardisierte Testinstrumente verwendet, welche die Lese- und

Schreibkompetenz der Schülerinnen und Schüler erfassen sollten. Aufgrund der Corona Pandemie konnten nicht alle Erhebungen planmäßig durchgeführt werden. Realisiert wurden vier Messzeitpunkte. Der erste Messzeitpunkt fand am Anfang des 1. Schuljahres (s. Abb. 1) statt. Es wurden relevante Hintergrundvariablen mittels eines Elternfragebogens sowie die phonologische Bewusstheit der Kinder erfasst. Die phonologische Bewusstheit ist eine sogenannte Vorläuferfähigkeit des Schriftspracherwerbs und wurde mit den Untertests *Reimwörter* und *Anlaute* des TEPHOBE (Mayer, 2011) erhoben. Am Ende des 1. Schuljahres fand die Erfassung der Schreib- und Lesekompetenz statt. Dafür wurden der Salzburger Lese- und Rechtschreibtest (SLRT-II; Moll & Landerl, 2010) und der Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6; Lenhard & Schneider, 2006) eingesetzt. Zur Überprüfung der Entwicklung der Lesefähigkeiten wurde in Klassenstufe 2 am Ende des 1. Schulhalbjahres erneut der ELFE 1-6 eingesetzt. Aufgrund der Corona Pandemie konnten keine weiteren Messzeitpunkte zur Erfassung der längsschnittlichen Schreib- und Lesekompetenz realisiert werden. Es war allerdings möglich, zu Beginn der 3. Klassenstufe die Kontrollvariable *Intelligenz* mittels des Grundintelligenztest Skala 1 – Revision (CFT 1-R; Weiß & Osterland, 2012) zu erfassen sowie erneut den Elternfragebogen auszuteilen, um Eltern von Kindern, die am 1. Messzeitpunkt keinen Elternfragebogen ausgefüllt haben, nach relevanten Hintergrundvariablen zu befragen.

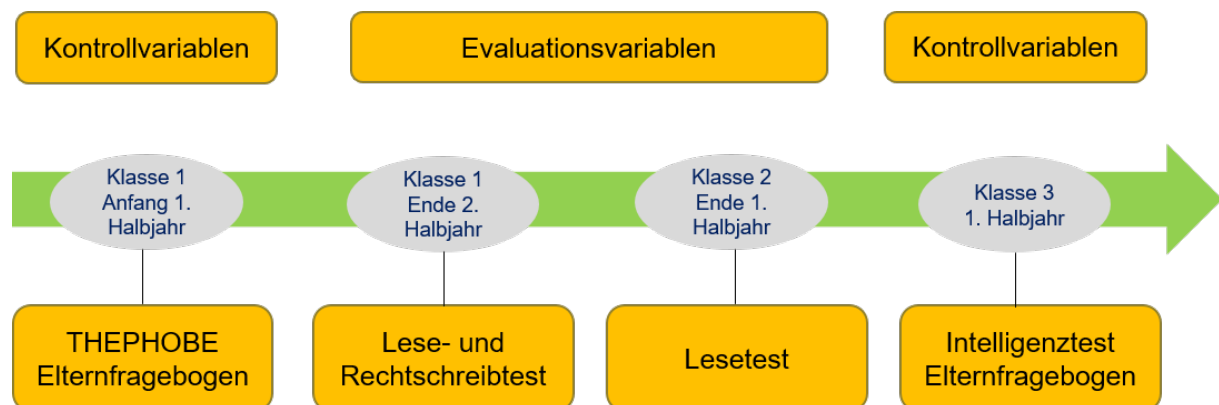


Abb. 1: Zeitplan und Design der Evaluationsstudie

Abbildung 2 zeigt als Ergänzung die Durchführung der Diagnostik mit dem Lms-Diagnosematerial (Meilensteine) und die sich anschließende Förderung mit den Trainingsbausteinen exemplarisch für die erste Klasse.

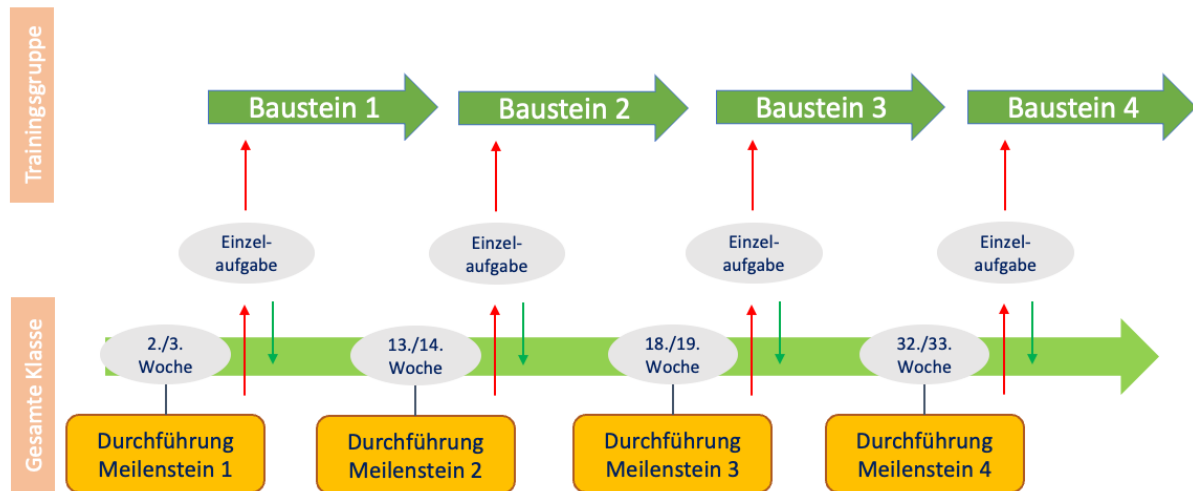


Abb. 2: Durchführung der Meilensteine und der Trainingsbausteine

## 2.3 Instrumente

### 2.3.1 Instrumente zur Erfassung der Kontrollvariablen und der Leistungsentwicklung

Es wurde ein umfangreicher Fragebogen entwickelt, der über die Schulen an die Eltern weitergegeben wurde. Der Elternfragebogen umfasst verschiedene Themenkomplexe zu den sprachlichen Kompetenzen der Eltern und der Kinder sowie zum Bildungshintergrund der Eltern. Zur Erfassung der Lese- und Schreibkompetenz, sowie der Intelligenz wurden standardisierte Testinstrumente ausgewählt.

#### Sozioökonomischer Status

Um den sozioökonomischen Status (SES) zu erfassen, wurden die Eltern u.a. nach der Anzahl der Bücher Zuhause gefragt. Hierbei werden fünf Ausprägungen berücksichtigt: (1) 0-25 Bücher, (2) 26-100 Bücher, (3) 101-200, (4) 201-500, (5) mehr als 500 Bücher. Der Buchbestand der Eltern kann als Indikator für das soziokulturelle Milieu gesehen werden und wurde in zahlreichen Studien verwendet (Heppt, Olczyk, & Volodina, 2022).

#### Migrationshintergrund

Der Migrationsstatus der Schülerinnen und Schüler wurde über die Angaben zu den Geburtsländern beider Eltern im Elternfragebogen bestimmt. Demnach liegt kein Migrationshintergrund vor, wenn beide Eltern in Deutschland geboren sind. Von einem partiellen Migrationshintergrund wird gesprochen, wenn ein Elternteil im Ausland geboren wurde, während ein vollständiger Migrationshintergrund vorliegt, wenn beide Elternteile aus dem Ausland stammen. Für die Datenanalyse wurde der Faktor binär kodiert (0 = kein Migrationshintergrund; 1 = partieller oder vollständiger Migrationshintergrund).

---

## Sprachlicher Hintergrund

Der sprachliche Hintergrund wurde unabhängig von der Frage des Geburtslandes und somit getrennt von einem möglichen Migrationshintergrund über Fragen zur Sprachpraxis innerhalb der Familie, zur Einschätzung der sprachlichen Kompetenzen der Kinder im Deutschen sowie ggf. in ihrer Familiensprache und über Fragen zu den Sprachkenntnissen beider Elternteile erhoben. Der sprachliche Hintergrund wurde durch drei Variablen erfasst. Der familiäre sprachliche Hintergrund (1) wurde durch die Frage nach den in der Familie gesprochenen Sprachen erfasst und resultiert in den Ausprägungen „nur Deutsch“, „Deutsch und mindestens eine weitere Sprache“ sowie „eine nichtdeutsche Sprache“ (vgl. hierzu auch Rüßmann et al. 2016). Für die Datenanalyse wurde der Faktor binär kodiert (0 = es wird nur Deutsch gesprochen; 1 = es wird Deutsch und mind. eine weitere Sprache gesprochen).

Die sprachlichen Kompetenzen des Kindes (2) wurden ebenfalls von den Eltern erfragt. Der Faktor umfasst die Kategorien „Das Kind spricht nur Deutsch“ sowie „Das Kind spricht Deutsch und eine weitere Sprache“.

Darüber hinaus wurde die Familiensprache (3) erfasst als die Sprache, die hauptsächlich in der Familie gesprochen wird. Hier kommen folgende Ausprägungen vor „Familiensprache Deutsch“, „Familiensprache nicht Deutsch“ sowie „Keine Angabe, bzw. mehrere Sprachen ausgewählt“. Auch dieser Faktor ging binär kodiert in die weitere Analyse ein (0 = Familiensprache Deutsch; 1 = Familiensprache nicht Deutsch).

Zusätzlich wurden die Eltern nach den Deutschkenntnissen ihrer Kinder befragt. Die Eltern gaben auf einer fünfstufigen Skala an, wie gut die Kinder deutsch sprechen und verstehen (1 = sehr gut, 2 = gut, 3 = mittelmäßig, 4 = schlecht, 5 = sehr schlecht).

## Lesefähigkeit

Um die Lesekompetenzen der Kinder zu erfassen, wurde mit dem ELFE-Lesekompetenztest (Lenhard & Schneider, 2006) ein standardisiertes Testverfahren eingesetzt. Der Test besteht aus den Untertests Wort-, Satz- und Textverständnis. Der ELFE-Test wurde zu zwei Messzeitpunkten durchgeführt. Am Ende der 1. Klasse wurden die beiden Untertests Wort- und Satzverständnis verwendet. Zum Ende des 1. Schulhalbjahres in der 2. Klasse wurden alle drei Untertests durchgeführt.

## Schreibfähigkeit

Um die Schreibkompetenzen der Schülerinnen und Schüler zu erfassen, wurde der Rechtschreibtest des SLRT (Moll & Landerl, 2010) genutzt. Der Rechtschreibtest testet die Fähigkeiten der Kinder in den Bereichen lautorientiertes und orthographisch korrektes

Schreiben. Der Test wurde aufgrund des coronabedingten Abbruchs der Studie nur einmalig am Ende des 1. Schuljahres durchgeführt.

#### Kognitive Grundfähigkeiten

Mit dem Grundintelligenztest Skala 1 CFT 1-R (Weiß & Osterland, 2012) wurden die intellektuellen Grundfähigkeiten der Kinder erfasst. Der Test wurde zu Beginn des 3. Schuljahrs durchgeführt.

#### Phonologische Bewusstheit

Aus dem standardisierten Testinstrument TEPHOBE (Mayer, 2013) wurden die Testteile zu Reimen und Anlauten verwendet, um die phonologische Bewusstheit der Kinder zu erfassen. Dies geschah zu Beginn des ersten Schuljahrs, da es sich um eine sogenannte Vorläuferfähigkeit des Schriftspracherwerbs handelt.

Die standardisierten Testverfahren wurden als Gruppentestung durchgeführt.

#### 2.3.2 Instrumente zur Erfassung der Implementation Fidelity

Zur Überprüfung der Implementation Fidelity wurden verschiedenen Instrumente eingesetzt:

**Die Dokumentation der Diagnostik** erfasst, zu welchen Zeitpunkten die Lms-Diagnostik durchgeführt wurde und welche Kinder Auffälligkeiten bei der Bearbeitung der Aufgaben zeigen.

**Die Dokumentation der Förderung** zeigt auf, welche Schülerinnen und Schüler an welchen Fördereinheiten teilgenommen haben. Aus der Dokumentation der Bögen wird weiterhin ersichtlich, wie viele der vorgesehenen Fördereinheiten stattgefunden haben.

**Die Bögen zur Beobachtung der Förderung** verfolgen das Ziel, die Durchführungsmodalitäten der Trainingseinheiten zu erfassen und aufzuzeigen, ob die Umsetzung in der Fördergruppe konzepttreu erfolgte.

**Der Fragebogen zur flankierenden Fortbildung** wurde von teilnehmenden Lehrkräften im Anschluss an das Fortbildungsangebot zu Lms-Training ausgefüllt. Dabei sollten die Lehrkräfte Aussagen zu verschiedenen Aspekten des Fortbildungsangebots auf einer vierstufigen Skala bewerten. Die Fragen unterteilten sich in zwei Abschnitte: Allgemeine Fragen zur Fortbildung und spezifische Fragen zu Lms-Training.

**Der Lehrerfragebogen** richtete sich an die Lehrkräfte, die das Lms-Training angewendet haben und an die der Kontrollgruppe. Neben persönlichen Angaben zur Person und zum beruflichen



Hintergrund erfasst der Fragebogen, wie Lehrkräfte den Einsatz und die Nutzung des Trainings bewerten.

Die folgende Tabelle 1 gibt eine Übersicht darüber, welche Instrumente mit welchem Ziel durchgeführt und wie oft sie eingesetzt wurden.

Dimension	Ziel	Instrument	Durchführung	Messzeitpunkte (MZP)
<b>Dokumentation der Diagnostik durch das Lms-Diagnostik Material</b>	Wann wurde der Meilenstein durchgeführt? Wer wurde identifiziert? Wurden die Ergebnisse durch die Einzelaufgabe überprüft? Wurde das Kind in die Trainingsgruppe aufgenommen? Wenn nein, warum nicht?	Dokumentationsbogen (Papier, digital)	4-mal im 1. Sj. 2-mal im 2. Sj.	6
<b>Dokumentation der Fördermaßnahmen</b>	Dokumentation der Förderung: Wer fördert wen wie oft und wie lange nach den Inhalten aus dem Trainingsbaustein?	Dokumentationsbogen (Papier, digital)	4-mal wöchentlich	5
<b>Beobachtung der Förderung</b>	Erfassung der Durchführungsmodalitäten	Beobachtungsbogen (Papier)	3-mal/ Sj.1/ Schule = 15-mal 1mal Sj.2/Schule = 5-mal	Beobachtung von 5 Schulen in Baustein 2,3,4 und 5
<b>Akzeptanz/ Bewertung der Lk-Fortbildung</b>	Evaluation der Fortbildung als Teil der Implementation	Dokumentationsbogen (Papier)	3-mal/ 1. Sj. 2-mal/ 2. Sj.	5
<b>Lehrerfragebogen</b>				
Persönlicher Hintergrund, Bewertung des Trainings, Bewertung der Diagnostik	s. Konstrukte des Lehrerfragebogens in Tabelle 5	Lehrerfragebogen (Papier)	IG: 4-mal KG: 1-mal	5

Tab. 1: Übersicht Instrumente der Implementation Fidelity

### 3. Zusammenfassung der Ergebnisse der Interventions- und Implementationsstudie „Lesen macht stark – Training“

#### 3.1 Ergebnisse der Interventionsstudie

Ergebnisse zur Wirkung der eingesetzten Trainings deuten auf positive Effekte bzgl. Der Lesefähigkeit hin. So steigern die Kinder, die an mindesten einem Baustein des Lesetrainings teilgenommen haben, ihre Wortverständnisfähigkeit im Laufe des Schuljahres deutlich stärker als die Kinder, die kein Fördertraining erhalten haben.

#### 3.2 Ergebnisse der Implementationsstudie

##### 3.2.1 Durchführung der Diagnostik und Förderung

Die Ergebnisse zur Überprüfung der Konzepttreue zeigen differenzierte Einblicke zum Einsatz des Lms-Trainingsprogramms:

Die Einteilung bzw. Zuordnung der Kinder in die Fördergruppen sollten auf Basis ihrer Leistungen in der Lms-Diagnostik erfolgen. Auffällige Leistungen von Kindern in einem Meilenstein im Rahmen der Klassenaufgabe sollten durch die erneute Durchführung der Aufgabe in einer Einzelsituation überprüft werden. Bestätigt sich hier die Schwäche, werden die Kinder in die Fördergruppe aufgenommen.

Es wird deutlich, dass von allen Kindern, die in der Lms-Diagnostik auffällig schwache Leistungen zeigten, nur 52% tatsächlich der Fördergruppe zugeordnet wurden. Etwa die Hälfte der Kinder, bei denen ein Förderbedarf festgestellt wurde, erhielt somit keine zusätzliche Förderung. Demgegenüber stehen 39% von Kindern mit unauffälligen Leistungen in der Lms-Diagnostik, die in die Fördergruppe aufgenommen wurden.

Vorgesehen ist pro Förderbaustein die Förderung an vier Tagen in der Woche für jeweils etwa 20 Minuten. Es wird deutlich, dass in den ersten drei Förderbausteinen zumindest 70-80% der vorgesehenen Fördereinheiten angeboten wurden. Ab dem vierten Baustein sinkt jedoch das Förderangebot. In den letzten beiden Förderbausteinen werden im Durchschnitt nur noch 60% der bestimmten Fördereinheiten durchgeführt. An einer Schule lag der Anteil der durchgeführten Förderung sogar nur noch bei rund 20%. Wurde die Förderung angeboten, nahmen die Kinder regelmäßig am Training im Rahmen der ersten vier Bausteine teil. Die Teilnahme an der angebotenen Förderung im Baustein 5 lag jedoch an zwei Schulen bei nur noch rund 50%.

##### 3.2.2 Beobachtung des Unterrichts

In allen Fördergruppen wurden zu zwei Zeitpunkten nicht-teilnehmende Beobachtungen durchgeführt und dokumentiert. Diese ergeben, dass die Umsetzung der Fördereinheiten in einigen Aspekten sehr konzepttreu erfolgt. Beispielsweise werden die Levelteilungen, die der Binnendifferenzierung dienen, und die Arbeit in Tandems sinnvoll eingesetzt. Darüber

hinaus wirken sowohl die Schülerinnen und Schüler als auch die Lehrkräfte sehr routiniert und motiviert in der Nutzung des Förderkonzepts. Allerdings zeigt die Beobachtung auch, dass die Umsetzung nicht durchgängig konzepttreu erfolgt. Bei einzelnen Aspekten sind immer wieder Abweichungen vom Konzept zu beobachten. Betroffen sind hiervon u.a. das Eingehen auf den spezifischen Förderbedarf einzelner Schülerinnen und Schüler sowie das Ineinandergreifen der verschiedenen Trainingsphasen.

### 3.2.3 Akzeptanz der Lehrkräfte

Mittels eines Fragebogens wurde erhoben, wie das Trainingsprogramm durch die beteiligten Lehrkräfte (N = 9) bewertet wurde.

Als positiv wird bewertet, dass das Trainingsprogramm auf die Automatisierung der basalen Lese- und Schreibfähigkeiten fokussiert. Die Lehrkräfte geben weiterhin an, dass sie sich durch das Training in ihrer Förderarbeit unterstützt fühlen und eine bedarfsgerechte Förderung durch die Nutzung des Trainingsprogramms erfolgt. Auch die Passung zwischen der Lms-Diagnostik und dem Lms-Training bewerten die Lehrkräfte als gut. Weniger positiv wird der Aufwand bei der Vorbereitung auf die Arbeit mit dem Trainingsprogramm bewertet. Diesen schätzen die Lehrkräfte als eher hoch ein. Weiterhin schätzen einige Lehrkräfte die Fördergruppen als zu groß ein, obwohl die Fördergruppen auf eine Größe von sechs bis acht Kindern beschränkt wurden.

Die beteiligten Lehrkräfte bewerten das Material insgesamt jedoch eher positiv bis sehr positiv.

## 4. Die Ergebnisse im Detail

Die Studie zum Trainingsprogramm umfasst die Intervention auf der einen (Teil A) und die Erfassung der Konzepttreue auf der anderen Seite (Teil B). Folgende Forschungsfragen sollen mit der Studie beantwortet werden:

### Teil A

Wie wirkt sich das Training auf die Lese- und Schreibkompetenzen der Schülerinnen und Schüler aus?

### Teil B

1. Ist die Konzepttreue (Förderqualität) gewährleistet?
  - a) Werden die Kinder, die in der Diagnostik auffällig waren, in die Fördergruppen aufgenommen?
  - b) Findet die Förderung regelmäßig statt?
  - c) Nehmen die Kinder regelmäßig an der Förderung teil?
  - d) Inwieweit wird die Förderung konzepttreu umgesetzt?

2. Inwieweit wird die Fortbildung als Teil der Implementation von den Lehrkräften akzeptiert (Evaluation der Fortbildung)?
3. Inwieweit wird das Lms-Training von den Lehrkräften akzeptiert?

Im Folgenden werden die Ergebnisse detailliert berichtet. Teil A umfasst die Leistungsentwicklung der Schülerinnen und Schüler.

#### 4.1 Teil A: Ergebnisse der Leistungsentwicklung

##### 4.1.1 Stichprobenbeschreibung und Auswahl der Gruppen für die Analysen

Es nahmen insgesamt 533 Schülerinnen und Schüler aus zehn Schulen in Schleswig-Holstein an der Untersuchung teil. Von drei Kindern wurden die Testhefte nicht ausgewertet, da keine Elterngenehmigung vorlag. Die Stichprobe umfasst somit 530 Lernende. Die Kontrollgruppe besteht aus 282 Schülerinnen und Schülern (14 Schulklassen), in dieser Gruppe wurde weder die Lms-Diagnostik noch das Lms-Training durchgeführt. An fünf der Schulen wurde in zehn Schulklassen die Lms-Diagnostik durchgeführt. In dieser Experimentalgruppe befinden sich insgesamt 248 Schülerinnen und Schüler. Aus der Gesamtheit der Experimentalgruppe werden verschiedene Differentialgruppen gebildet, die dann in die Analysen aufgenommen werden (vgl. Abb. 3). Anhand der Lms-Diagnostik wurden in der Experimentalgruppe 93 Schülerinnen und Schüler identifiziert, die dann nachfolgend an mindestens einem Trainingsbaustein teilnahmen (Trainingsgruppe, TG). Aufgrund der Maßnahmen im Zusammenhang mit der Corona-Pandemie konnten nicht alle Trainingsbausteine planmäßig durchgeführt werden. Eine Analyse der Durchführungshäufigkeit der Trainingsbausteine zeigt, dass vor allem die Schreibtrainings nur unregelmäßig durchgeführt wurden (s. Abschnitt 4.2.1). Weiterhin konnten nicht alle Leistungstests im Längsschnitt erhoben werden (siehe Abschnitt 2.2). Da der 5. Trainingsbaustein zur Leseförderung nach der zweiten Leistungserhebung stattfand, werden Schülerinnen und Schüler, die nur am 5. Trainingsbaustein teilnahmen, nicht in den Analysen berücksichtigt. Somit besteht die Trainingsgruppe aus 72 Schülerinnen und Schülern, die begleitend zur Leistungserhebung an mindestens einem Baustein (1 bis 4, also mindestens einem Baustein des Lese- und/oder Schreibtrainings) teilgenommen haben. In dieser Trainingsgruppe haben 53 Lernende an mindestens einem Baustein des Lesetrainings teilgenommen. Lesen wurde in den Trainingsbausteinen 1, 2 und 5 trainiert. An mindestens einem Baustein Schreiben haben 47 Kinder teilgenommen (Baustein 3 und 4). Folglich haben 156 Kinder aus der Experimentalgruppe an keinem Baustein teilgenommen (EG ohne Training).

Aufgrund der genannten Besonderheiten konnten die ursprünglich beabsichtigten Analysen nicht wie geplant durchgeführt werden. Zur Überprüfung der Wirksamkeit der Bausteine in Bezug auf die Lesefähigkeit wurden deshalb nur die Kinder ausgewählt, die am Baustein 1 oder 2 teilgenommen haben (Lesegruppe, LG). Diese Lesegruppe wurde zum einen mit der

gesamten Kontrollgruppe sowie mit den Kindern der Experimentalgruppe, die an keinem Training teilgenommen haben, verglichen. Die Vergleiche der anderen Gruppen (Trainingsgruppe vs. Kontrollgruppe und EG ohne Training) sind im Anhang zu finden. Eine Überprüfung der Wirksamkeit des Schreibtrainings war nicht möglich, da die Schreibfähigkeit nur zu einem Messzeitpunkt erhoben werden konnte. Dennoch ist der Vergleich der Schreibgruppe (SG) mit der Kontrollgruppe (KG) sowie der Experimentalgruppe ohne Training (EGoT) der Vollständigkeit halber aufgeführt.

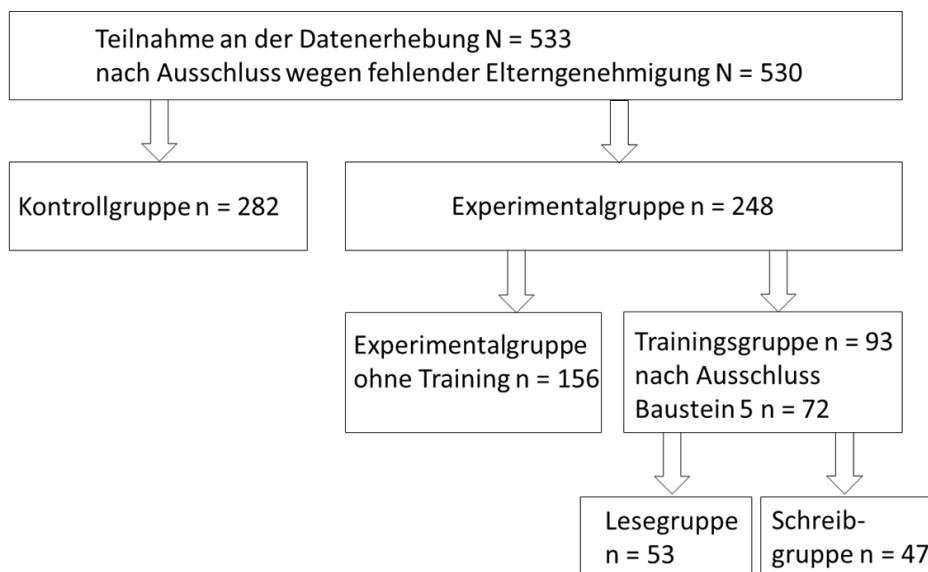


Abb. 3: Auswahl der Differentialgruppen

#### 4.1.2 Vergleich der beiden Gruppen hinsichtlich der Kontrollvariablen

Die Schulen der Experimental- und Kontrollgruppe wurden im Vorfeld so ausgewählt, dass sie sich hinsichtlich sozioökonomischer Merkmale und des Leistungsstands ähnelten. Dazu wurden schulstatistische Daten sowie Leistungsdaten der VERA-Erhebungen gesichtet und die Schulen hinsichtlich dieser Daten ausgewählt. Um zu überprüfen, ob die Schulauswahl gelungen ist und die beiden Gruppen hinsichtlich spezifischer demographischer und sozioökonomischer Merkmale vergleichbar sind, wurden zum einen verschiedene Kontrollvariablen mittels Fragebogen erhoben und zum anderen weitere Testverfahren appliziert. Da nicht alle Eltern einen Fragebogen ausgefüllt haben (N = 422), konnten nicht die Daten aller Kinder in die Analysen einbezogen werden. Die ausführlichen Analysen sind im Anhang (Abschnitt I) dargestellt. Die Ergebnisse zeigen, dass die Kontrollgruppe und Experimentalgruppe vergleichbare Voraussetzungen haben. Weiterhin wurden die Differentialgruppen, die in die Analysen zur Überprüfung der Wirksamkeit eingehen, mit der

Kontrollgruppe bzw. der Experimentalgruppe ohne Training bzgl. der Kontrollvariablen verglichen. Nachfolgend wird der Vergleich zwischen der Lesegruppe mit der Experimentalgruppe ohne Training und der Kontrollgruppe dargestellt. Im Anhang finden sich weitere Analysen.

### Demographische Variablen

Der deskriptive Vergleich zwischen den Gruppen zeigt, dass einige Merkmale in den Gruppen annähernd gleich verteilt sind. So beträgt das mittlere Alter der Schülerinnen und Schüler zum Zeitpunkt der ersten Datenerhebung im September 2018 in allen drei Gruppen 6,8 Jahre (LG:  $SD = 0,41$ ;  $Min = 6,08$ ;  $Max = 8,08$ ; EGoT:  $SD = 0,49$ ;  $Min = 6,0$ ;  $Max = 8,4$ ; KG:  $SD = 0,46$ ;  $Min = 5,8$ ;  $Max = 8,3$ ). Ein partieller oder vollständiger Migrationshintergrund liegt bei 15 % der Lernenden der Lesegruppe (54% kein Migrationshintergrund, 31% fehlende Angaben) und bei 22% der Schülerinnen und Schüler in der Experimentalgruppe ohne Training vor (61% kein Migrationshintergrund, 17% fehlende Angaben) sowie bei 29% der Kinder in der Kontrollgruppe (55% kein Migrationshintergrund, 16 % fehlende Angaben). Zusätzlich wurde der sprachliche Hintergrund mit drei Variablen erfasst. Die Ergebnisse zeigen, dass in 50% der Familien in der Lesegruppe nur Deutsch gesprochen wird, in 16% der Familien werden Deutsch und eine andere Sprache gesprochen und von 35% der Familien fehlen die Angaben. In der Experimentalgruppe ohne Training sprechen 58% der Familien ausschließlich Deutsch zuhause, in 22% der Familien werden Deutsch und eine andere Sprache gesprochen und von 20% der Familien fehlen die Angaben. In der Kontrollgruppe sprechen 54% der Familien ausschließlich Deutsch zuhause, in 27% der Familien werden Deutsch und eine andere Sprache gesprochen und von 19% der Familien fehlen die Angaben. Die Familien, in denen auch eine andere Sprache als Deutsch gesprochen wurde, gaben zusätzlich an, welche Sprache das Kind zuhause spricht und welche Sprache hauptsächlich in der Familie gesprochen wird. Dabei zeigt sich, dass diese Variablen nicht ganz gleich in den Gruppen verteilt sind. In der Lesegruppe sprechen 13% der Kinder nur Deutsch zuhause und 87% der Kinder Deutsch und eine andere Sprache. In der Experimentalgruppe ohne Training sprechen 16% der Kinder nur Deutsch zuhause und 84% der Kinder Deutsch und eine andere Sprache. In der Kontrollgruppe sprechen hingegen 26% der Kinder nur Deutsch und 74% der Kinder sprechen Deutsch und eine andere Sprache. In Bezug auf die Familiensprache zeigt sich, dass in der Experimentalgruppe ohne Training etwas weniger Familien nur Deutsch sprechen (40%) als in der Lesegruppe (50%) und der Kontrollgruppe (50%). Befragt nach den Deutschkenntnissen der Kinder, geben mehr Eltern in der Experimentalgruppe ohne Training (85%) und in der Kontrollgruppe (96%) an, dass ihre Kinder über sehr gute bis gute Kenntnisse verfügen als in der Lesegruppe (75%). Da die letzten drei erhobenen Variablen nur von Eltern beantwortet wurden, deren Familiensprache nicht ausschließlich Deutsch war, können diese Variablen aufgrund der geringen Stichprobengröße nicht als Kovariaten in den Analysen verwendet werden.

Der sozioökonomische Status der Eltern wurde mittels des Bücherbesitzes erfasst. Die Gruppen unterscheiden sich nicht signifikant hinsichtlich der Anzahl des Besitzes von Büchern zuhause ( $F(2, 390) = 1,209$ ;  $p = ,30$ ; partielles  $\eta^2 = ,006$ ).

Bei der Betrachtung des Geschlechts zeigt sich, dass mehr Jungen in der Lesegruppe (LG: Jungen = 62%) sind als in der Experimentalgruppe ohne Training bzw. der Kontrollgruppe (EGoT: Jungen = 42%; KG: Jungen = 50%). Diese Variable wurde deshalb als Kovariate mit in die Analysen aufgenommen.

#### TEPHOBE

Zur ersten Kontrolle der Voraussetzungen der Schülerinnen und Schüler wurden die Untertests „Anlaute“ und „Reimen“ des standardisierten Tests TEPHOBE zu Beginn der 1. Klassenstufe eingesetzt. Es liegen valide Daten von 381 Schülerinnen und Schülern im Untertest „Reimen“ und  $n = 338$  für den Untertest „Anlaute“ vor. Die Ergebnisse zeigen signifikante Unterschiede zwischen den Gruppen (Reimen:  $F(2, 378) = 15,855$ ;  $p < ,001$ ; partielles  $\eta^2 = ,077$ ; Anlaute:  $F(2, 335) = 9,343$ ;  $p < ,001$ ; partielles  $\eta^2 = ,053$ ). Abbildung 4 zeigt die Mittelwerte der Untertests der drei Gruppen. Folglich unterscheiden sich die Gruppen hinsichtlich ihrer Vorläuferfähigkeiten des Schriftspracherwerbs und die beiden T-Werte der Untertests des TEPHOBE werden als Kovariaten mit in die Analysen aufgenommen.

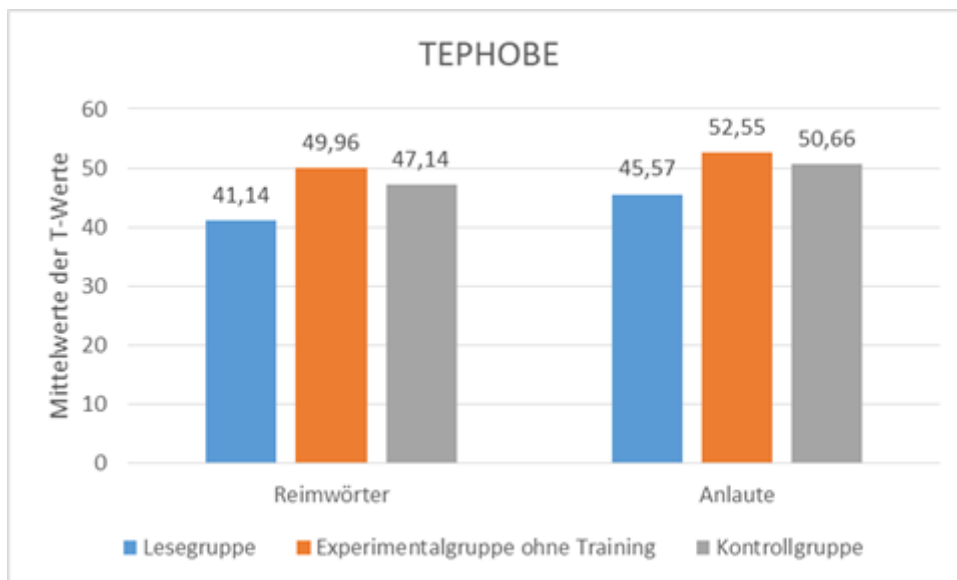


Abb. 4: Mittelwerte der T-Werte in den beiden TEPHOBE-Untertests „Anlaute“ und „Reime“ für die drei Gruppen (Lesegruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe)

## CFT 1-R

Als Kontrollvariable und zur Beschreibung der Stichprobe wurde der CFT 1-R zum letzten Erhebungsmesszeitpunkt eingesetzt. Es liegen Daten von 392 Schülerinnen und Schülern vor. Die Analyse zeigt, dass sich die drei Gruppen signifikant voneinander unterscheiden ( $F(2, 389) = 6,668$ ;  $p < ,001$ ; partielles  $\eta^2 = ,033$ ). Daher wird der Intelligenzwert als Kovariate in die nachfolgenden Analysen aufgenommen. Die Mittelwerte der drei Gruppen sind in Abbildung 5 dargestellt.

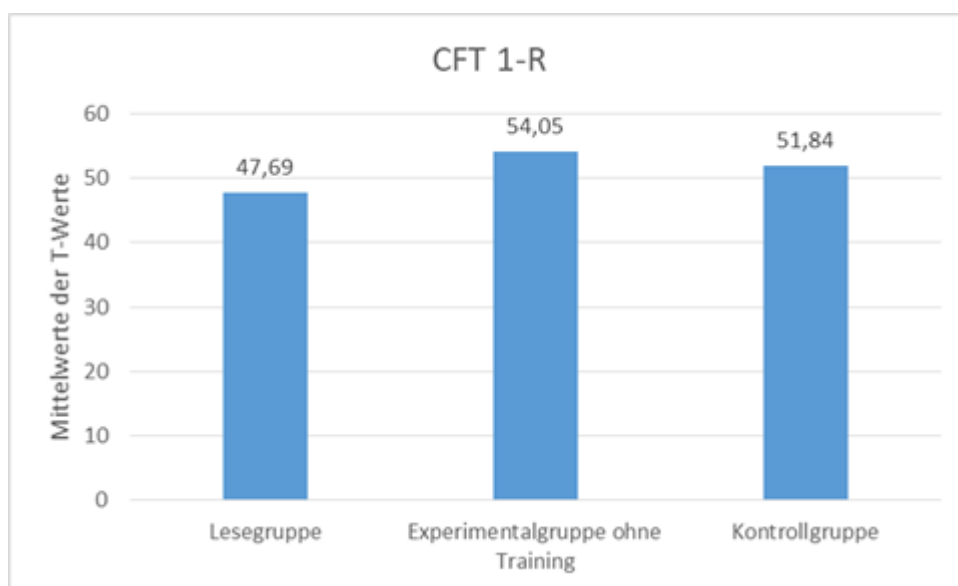


Abb. 5: Mittelwerte der T-Werte im CFT 1-R für die drei Gruppen (Lesegruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe)

### 4.1.3 Vergleich der beiden Gruppen hinsichtlich der Fördermaßnahmen Entwicklung der Lesekompetenz

Zur Überprüfung der Wirksamkeit des Lesetrainings wurden nur die Schülerinnen und Schüler aus der Experimentalgruppe ausgewählt, die an mindestens einem Trainingsbaustein Lesen (Baustein 1 und 2) teilgenommen haben. Diese Lesegruppe werden zum einem mit der gesamten Kontrollgruppe verglichen und zum anderen mit den Schülerinnen und Schülern der Experimentalgruppe, die kein Training erhalten haben. Der ELFE-Lesekompetenztest besteht aus den Untertests Wort-, Satz- und Textverständnis. Da nicht zu beiden Zeitpunkten alle Untertests durchgeführt wurden, kann keine Veränderung des Gesamtwertes errechnet werden. Die Untertests werden daher getrennt betrachtet. Durchgeführt wird eine Varianzanalyse mit Kovariaten (ANCOVA) mit dem Messwiederholungsfaktor Zeit und der



unabhängigen Variable Gruppe (LG, KG, EGoT). Abhängige Variablen sind die T-Werte in den beiden entsprechenden Untertests (Wort- und Satzverständnis).

### Wortverständnis

Für den Untertest „Wortverständnis“ liegen insgesamt Daten für 261 Schülerinnen und Schüler vor. Die Ergebnisse zeigen einen signifikanten Haupteffekt des Faktors Gruppe ( $F(2, 255) = 8,314$ ;  $p < ,001$ ; partielles  $\text{Eta}^2 = ,061$ ) sowie einen signifikanten Haupteffekt des Faktors Zeit ( $F(1, 255) = 25,23$ ;  $p < ,001$ ; partielles  $\text{Eta}^2 = ,09$ ). Das bedeutet, dass sich zum einen die Gruppen über die beiden Messzeitpunkte hinweg hinsichtlich der Ergebnisse im Wortverständnis unterscheiden und zum anderen, dass sich über die Gruppen hinweg das Wortverständnis über die Zeit verbessert hat. Dabei zeigt sich, dass sich der Zuwachs der Wortverständnisfähigkeit von Messzeitpunkt 1 zu Messzeitpunkt 2 zwischen den drei Gruppen signifikant unterscheidet ( $F(2, 255) = 10,839$ ;  $p < ,001$ ; partielles  $\text{Eta}^2 = ,078$ ). Ein differenzierter Vergleich der Gruppen zeigt, dass die Schülerinnen und Schüler, die am Lesetraining teilnahmen, innerhalb eines Schulhalbjahres sowohl einen größeren Leistungszuwachs als die Kontrollgruppe ( $F(1, 144) = 5,061$ ;  $p = ,026$ ) als auch einen größeren Leistungszuwachs als die Experimentalgruppe ohne Training ( $F(1, 94) = 11,149$ ;  $p < ,001$ ) erzielen. Dabei zeigt sich ein größerer Effekt im Vergleich der Lesegruppe mit der Experimentalgruppe ohne Trainings (Effektstärke: partielles  $\text{Eta}^2 = ,106$ ) als im Vergleich der Lesegruppe mit der Kontrollgruppe (Effektstärke: partielles  $\text{Eta}^2 = ,034$ ). In Abbildung 6 sind die Werte graphisch verdeutlicht.

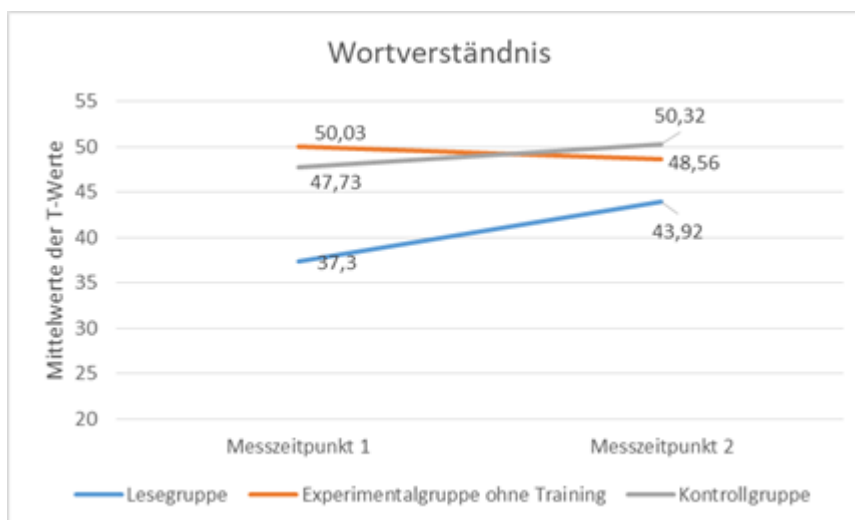


Abb. 6: Mittelwerte im Wortverstehen für die drei Gruppen (Lesegruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe) zum ersten und zweiten Messzeitpunkt

## Satzverstehen

Für den Untertest „Satzverständnis“ liegen Daten für 268 Schülerinnen und Schüler vor. Die Ergebnisse zeigen einen signifikanten Haupteffekt des Faktors Gruppe ( $F(2, 263) = 10,442$ ;  $p < ,001$ ; partielles  $\text{Eta}^2 = ,074$ ) sowie einen signifikanten Haupteffekt des Faktors Zeit ( $F(1, 263) = 20,420$ ;  $p < ,001$ ; partielles  $\text{Eta}^2 = ,072$ ). Das bedeutet, dass sich zum einen die beiden Gruppen über die beiden Messzeitpunkte hinweg hinsichtlich der Ergebnisse im Satzverständnis unterscheiden und zum anderen, dass sich über beide Gruppen hinweg das Satzverständnis über die Zeit verbessert hat. Allerdings kann kein unterschiedlich starker Anstieg der Satzverständnissfähigkeit vom ersten zum zweiten Messzeitpunkt in Abhängigkeit der Gruppe gefunden werden ( $F(2, 263) = 1,202$ ;  $p = ,30$ ; partielles  $\text{Eta}^2 = ,009$ ). In Abbildung 7 sind die Werte graphisch verdeutlicht.

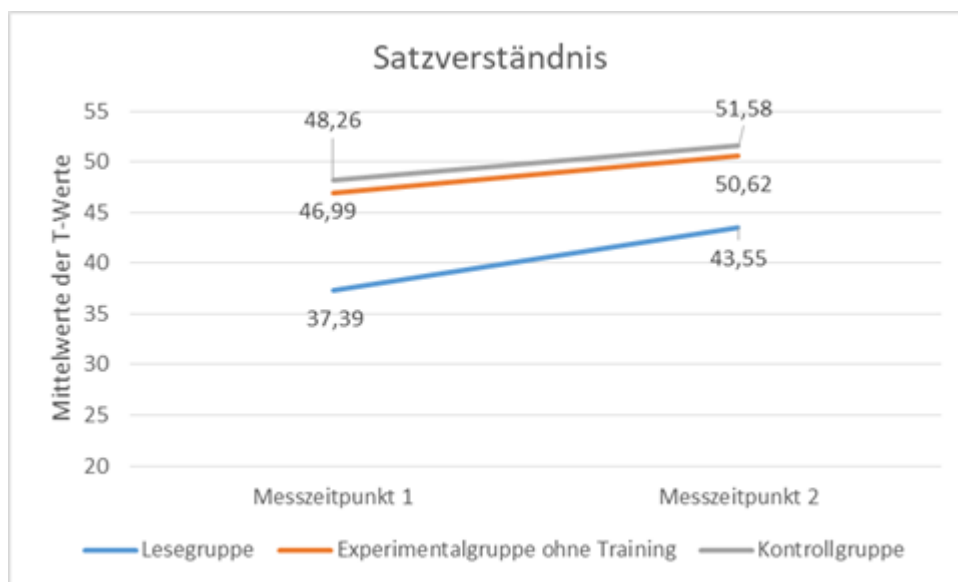


Abb. 7: Mittelwerte im Satzverstehen für die drei Gruppen (Lesegruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe) zum ersten und zweiten Messzeitpunkt

## Textverständnis

Da der Untertest „Textverständnis“ des ELFE nur zu einem Messzeitpunkt erhoben (siehe Abbildung 1) wurde, liegen hierfür nur querschnittliche Daten vor. Die Messung fand nach den Bausteinen 1 und 2 des Lesetrainings statt. Da keine Ausgangsmessung vorliegt, ist eine Interpretation der Wirksamkeit des Trainings in Bezug auf die Entwicklung des Textverständnisses nicht möglich. Die einfaktorielle ANOVA ergibt, dass sich das Textverständnis signifikant zwischen den Gruppen unterscheidet ( $F(2,382) = 9,171$ ;  $p < ,001$ ; partielles  $\text{Eta}^2 = ,028$ ). Die Mittelwerte der T-Werte des Textverständnisses sind für die drei

Gruppen in Abbildung 8 dargestellt. Eine differenzierte Analyse mittels Kontrastanalysen zeigt, dass das Textverständnis der Lesegruppe signifikant schlechter ist als das Textverständnis der Experimentalgruppe ohne Training ( $t(382) = 3,619$ ;  $p < ,001$ ; Cohens  $d = 0,69$ ) sowie der Kontrollgruppe ( $t(382) = 4,605$ ;  $p < ,001$ ; Cohens  $d = 0,78$ ).

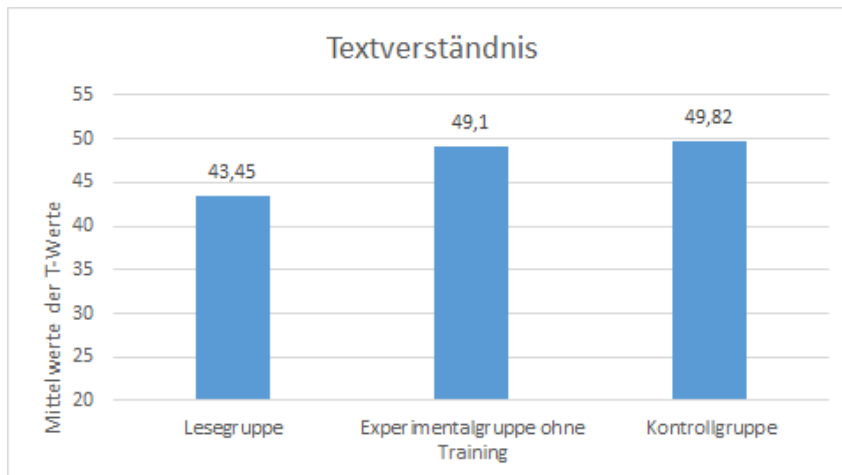


Abb. 8: Mittelwerte im Textverstehen für die drei Gruppen (Lesegruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe)

### Schreibkompetenz

Im Folgenden werden die Schülerinnen und Schüler, die an mindestens einem Baustein des Schreibtrainings (Schreibgruppe) teilgenommen haben, mit der Experimentalgruppe ohne Training und der Kontrollgruppe verglichen. Da der SLRT nur zu einem Messzeitpunkt erhoben (siehe Abbildung 1) wurde, lagen nur querschnittliche Daten vor und eine Überprüfung der Wirksamkeit des Schreibtrainings (Baustein 3 und 4) ist somit nicht möglich. Der SLRT liefert drei verschiedene abhängige Variablen. Die einfaktorielle ANOVA zeigt, dass sich die Gruppen in Bezug auf alle Fehlerarten signifikant voneinander unterscheiden (O-Fehler:  $F(2,404) = 4,613$ ;  $p < ,001$ ; partielles  $\eta^2 = ,022$ ; N-Fehler:  $F(2,404) = 71,233$ ;  $p < ,001$ ; partielles  $\eta^2 = ,261$ ; G-Fehler:  $F(2,404) = 8,518$ ;  $p < ,001$ ; partielles  $\eta^2 = ,040$ ). Eine differenzierte Betrachtung der Gruppen mittels geplanter Kontrastanalysen zeigt, dass es hinsichtlich der O-Fehler (orthographische Fehler) keinen signifikanten Unterschied zwischen der Schreibgruppe und der Kontrollgruppe ( $t(402) = 1,392$ ;  $p = ,165$ ; Cohens  $d = 0,22$ ) gibt. Die Schülerinnen und Schüler beider Gruppen machen somit ähnlich viele orthographische Fehler (siehe Abbildung 9). Hingegen zeigt sich, dass die Schreibgruppe mehr orthographische Fehler macht als die Experimentalgruppe ohne Training ( $t(402) = -2,255$ ;  $p = ,025$ ; Cohens  $d = 0,39$ ). Ein sehr bedeutsamer Unterschied zeigt sich bei den lautgetreuen Fehlern (N-Fehler). Die

Schreibgruppe macht signifikant mehr N-Fehler als die Kontrollgruppe ( $t(402) = -10,481; p < ,001$ ; Cohens  $d = 1,70$ ) sowie mehr Fehler als die Experimentalgruppe ohne Training ( $t(402) = 9,586; p < ,001$ ; Cohens  $d = 1,66$ ). Die Effektstärke zeigt einen sehr starken Effekt. Bezüglich der Fehler in der Groß- und Kleinschreibung (G-Fehler) zeigt sich, dass die Schreibgruppe ebenfalls signifikant mehr Fehler macht als die Kontrollgruppe ( $t(402) = -2,482; p = ,013$ ) sowie die Experimentalgruppe ohne Training ( $t(402) = 2,15; p = ,037$ ). Der Effekt ist als mittlerer Effekt einzuordnen (Cohens  $d = 0,40$  bzw. Cohens  $d = 0,43$ ).

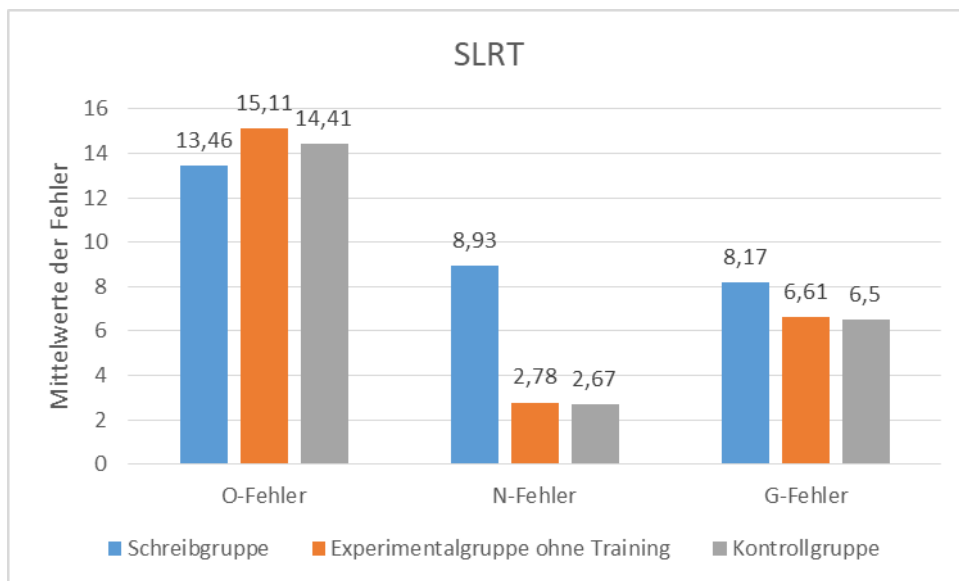


Abb. 9: Mittlere Fehleranzahl für drei Fehlerarten: orthographische Fehler (O-Fehler), lautgetreue Fehler (N-Fehler) und Fehler der Groß-Kleinschreibung (G-Fehler) für die drei Gruppen (Schreibgruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe)

Teil B der Studie befasst sich mit Fragen zur konzepttreuen Umsetzung der Diagnostik und des Trainings.

## 4.2 Teil B: Umsetzung der Diagnostik und des Trainings

### 4.2.1 Wiedergabetreue des Konzeptes

Im Folgenden werden die Ergebnisse der Implementationsstudie mit dem Fokus auf die Wiedergabetreue berichtet. Die Wiedergabetreue umfasst dabei zum einen die Nutzung der Diagnostik für die Zuteilung der Kinder in eine Fördergruppe und zum anderen die regelmäßige Durchführung der Förderung in den einzelnen Förderbausteinen.

## Aufnahme in die Trainingsgruppe auf Basis der Lms-Diagnostik

Das Interventionskonzept sieht vor, dass Kinder, deren Leistungen in der Klassenaufgabe des jeweiligen Meilensteins von der Lms-Diagnostik unter einem vorher definierten Cut-off-Wert liegen, diese Aufgabe in einer Eins-zu-Eins-Situation (Einzelaufgabe in Lms-Diagnostik) mit der Lehrkraft noch einmal bearbeiten. Bestätigt sich hier das schwache Abschneiden wird von einem Förderbedarf ausgegangen und das Kind der Trainingsgruppe zugeordnet.

Zunächst wurde für jede Aufgabe des Meilensteins geprüft, ob ein Kind eine schwache Leistung in einer der Klassenaufgabe zeigte und wenn ja, ob diese Leistung von der Lehrkraft dann in der Einzelaufgabe geprüft und gegebenenfalls ein Risiko bestätigt wurde. Wenn ein Kind in der Klassenaufgabe auffällige Leistungen zeigte, die Leistung in der Einzelaufgabe jedoch nicht bekannt ist, weil diese entweder nicht durchgeführt wurde oder nicht angegeben wurde, wird diese Aufgabe nicht in die Datenanalyse aufgenommen. Das heißt, es werden nur Leistungen als auffällig bewertet, zu denen auch Informationen aus der Einzelaufgabe vorliegen. Im nächsten Schritt wurde geprüft, ob die Kinder, bei denen das Risiko durch die Einzelaufgabe bestätigt wurde, in die Trainingsgruppe aufgenommen wurden. Kinder, zu denen keine Angabe zur Aufnahme in die Trainingsgruppe vorliegt, wurden nicht in die Auswertung einbezogen.

Somit ergeben sich vier mögliche Gruppenzuordnungen:

- Gruppe 1: Kinder, die in einem Meilenstein in keiner Aufgabe auffällige Leistungen zeigen und nicht der Trainingsgruppe zugeteilt werden.
- Gruppe 2: Kinder, die in keiner Aufgabe eines Meilensteins auffällige Leistungen zeigen, trotzdem jedoch der Trainingsgruppe zugeteilt werden.
- Gruppe 3: Kinder, die in mindestens einer Aufgabe (Klassen- & Einzelaufgabe) eines Meilensteins Leistungen unterhalb des Cut-off-Werts zeigen und die in die Trainingsgruppe aufgenommen werden.
- Gruppe 4: Kinder, die in mindestens einer Aufgabe (Klassen- & Einzelaufgabe) eines Meilensteins Leistungen unterhalb des Cut-off-Werts zeigen und die nicht in die Trainingsgruppe aufgenommen werden.

Im Folgenden werden die Ergebnisse aller sechs Meilensteine berichtet.

### **Meilenstein 1 (n = 204)**

Bei 77% der Kinder wurde kein Förderbedarf diagnostiziert und sie wurden nicht der Trainingsgruppe zugeteilt. Neun Kinder, die laut Diagnostik keinen Förderbedarf aufweisen, wurden trotzdem in die Trainingsgruppe aufgenommen. Von den insgesamt 38 Kindern, die in diesem Meilenstein Leistungen in mindestens einer der Klassen- und der dazugehörigen Einzelaufgabe unterhalb des festgelegten Cut-Off-Werts zeigten, wurden 27 in die

Trainingsgruppe aufgenommen. Elf Kinder erhielten keine zusätzliche Förderung. Das heißt, 29% der identifizierten Kinder mit Förderbedarf konnten nicht am Training teilnehmen, während 25% der Kinder, die in die Fördergruppe aufgenommen wurden, keinen festgestellten Förderbedarf aufweisen (siehe Abbildung 10).

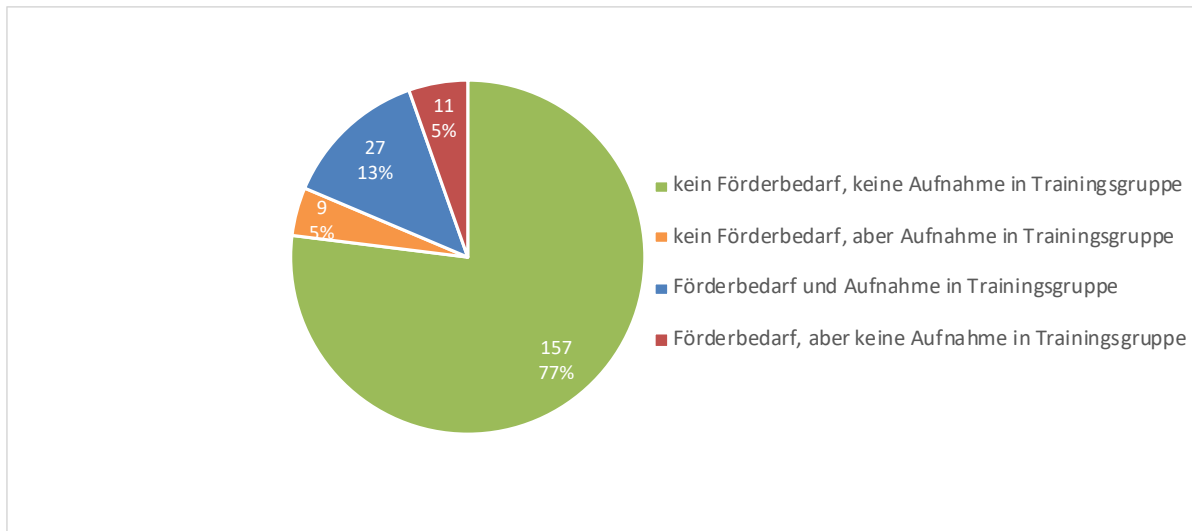


Abb. 10: Meilenstein 1 - Diagnostik und Förderung

### Meilenstein 2 (n = 183)

Bei 82% der Kinder wurde kein Förderbedarf diagnostiziert und sie wurden nicht der Trainingsgruppe zugeteilt. Elf Kinder, die laut Diagnostik keinen Förderbedarf aufweisen, wurden trotzdem in die Trainingsgruppe aufgenommen. Von den insgesamt 23 Kindern, die in diesem Meilenstein Leistungen in mindestens einer der Klassen- und der dazugehörigen Einzelaufgabe unterhalb des festgelegten Cut-Off-Werts zeigten, wurden 15 in die Trainingsgruppe aufgenommen. Acht Kinder erhielten keine zusätzliche Förderung. Das heißt, 35% der identifizierten Kinder mit Förderbedarf konnten nicht am Training teilnehmen, während 42% der Kinder, die in die Fördergruppe aufgenommen wurden, keinen festgestellten Förderbedarf aufweisen (siehe Abbildung 11).

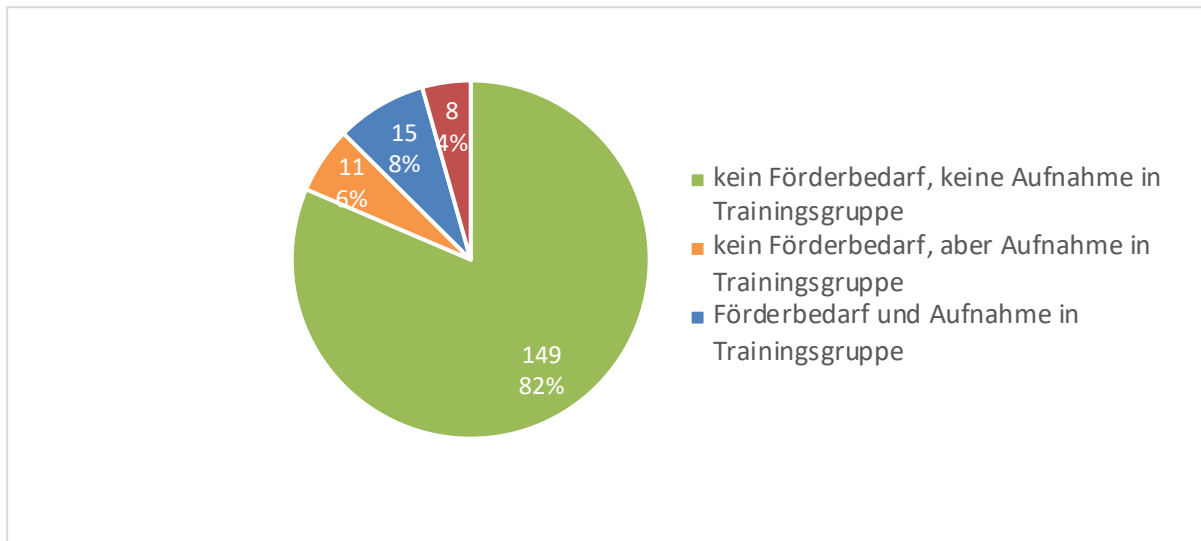


Abb. 11: Meilenstein 2 - Diagnostik und Förderung

### Meilenstein 3 (n = 198)

Bei 74% der Kinder wurde kein Förderbedarf diagnostiziert und sie wurden nicht der Trainingsgruppe zugeteilt. Zwölf Kinder, die laut Diagnostik keinen Förderbedarf aufweisen, wurden trotzdem in die Trainingsgruppe aufgenommen. Von den insgesamt 39 Kindern, die in diesem Meilenstein Leistungen in mindestens einer der Klassen- und der dazugehörigen Einzelaufgabe unterhalb des festgelegten Cut-Off-Werts zeigten, wurden 20 in die Trainingsgruppe aufgenommen. 19 Kinder erhielten keine zusätzliche Förderung. Das heißt, 49% der identifizierten Kinder mit Förderbedarf konnten nicht am Training teilnehmen, während 37% der Kinder, die in die Fördergruppe aufgenommen wurden, keinen festgestellten Förderbedarf aufweisen (siehe Abbildung 12).

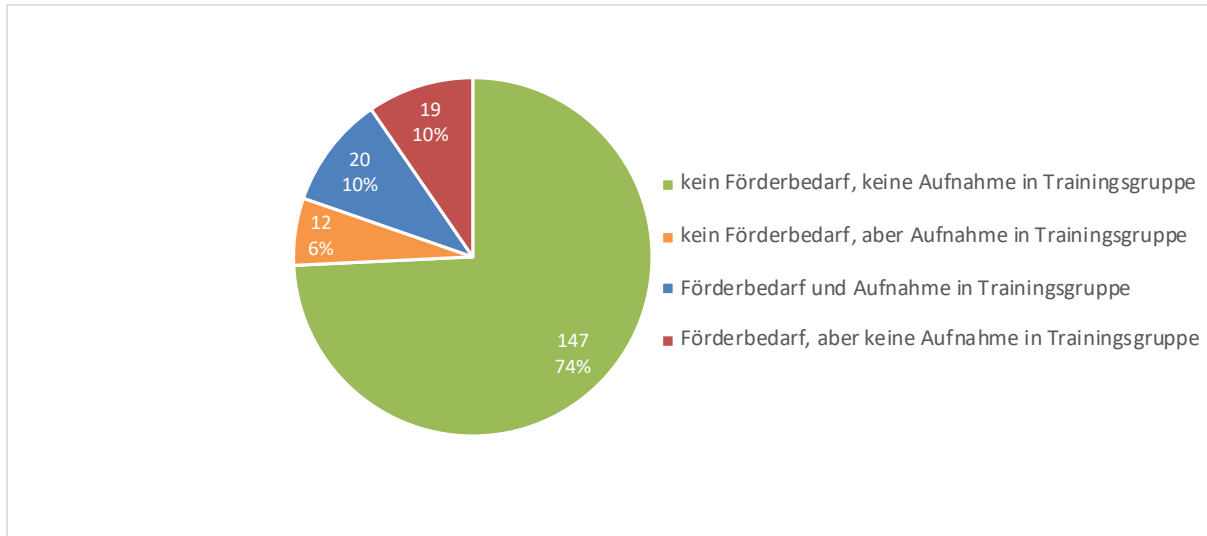


Abb. 12: Meilenstein 3 - Diagnostik und Förderung

#### Meilenstein 4 (n = 191)

Bei 72% der Kinder wurde kein Förderbedarf diagnostiziert und sie wurden nicht der Trainingsgruppe zugeteilt. 19 Kinder, die laut Diagnostik keinen Förderbedarf aufweisen, wurden trotzdem in die Trainingsgruppe aufgenommen. Von den insgesamt 35 Kindern, die in diesem Meilenstein Leistungen in mindestens einer der Klassen- und der dazugehörigen Einzelaufgabe unterhalb des festgelegten Cut-Off-Werts zeigten, wurden elf in die Trainingsgruppe aufgenommen. 24 Kinder erhielten keine zusätzliche Förderung. Das heißt, 69% der identifizierten Kinder mit Förderbedarf konnten nicht am Training teilnehmen, während 63% der Kinder, die in die Fördergruppe aufgenommen wurden, keinen festgestellten Förderbedarf aufweisen (siehe Abbildung 13).



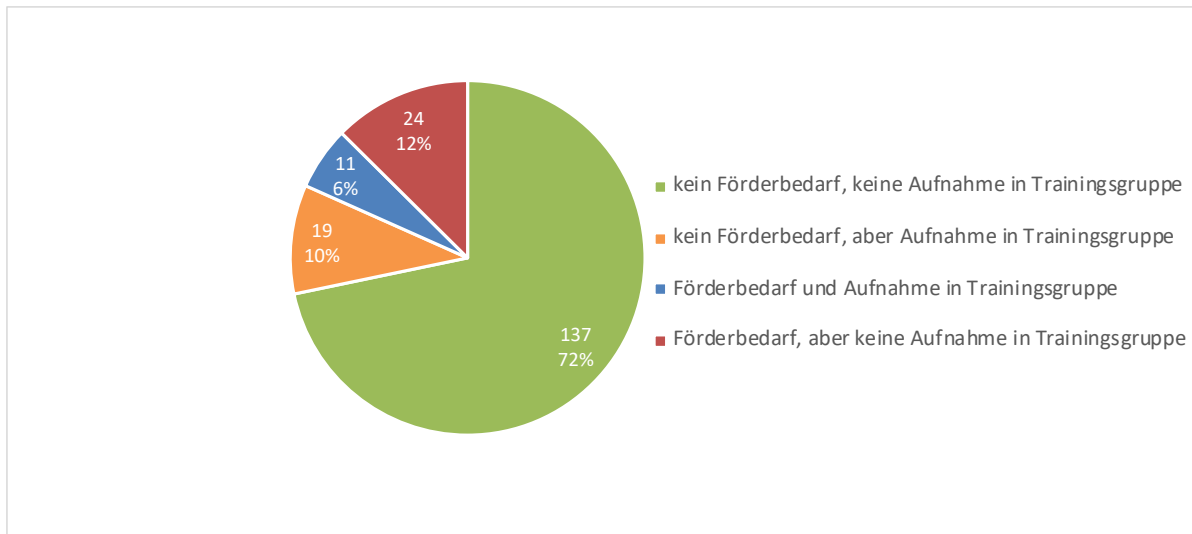


Abb. 13: Meilenstein 4 - Diagnostik und Förderung

### Meilenstein 5 (n = 208)

Bei 70% der Kinder wurde kein Förderbedarf diagnostiziert und sie wurden nicht der Trainingsgruppe zugeteilt. Sieben Kinder, die laut Diagnostik keinen Förderbedarf aufweisen, wurden trotzdem in die Trainingsgruppe aufgenommen. Von den insgesamt 56 Kindern, die in diesem Meilenstein Leistungen in mindestens einer der Klassen- und der dazugehörigen Einzelaufgabe unterhalb des festgelegten Cut-Off-Werts zeigten, wurden 23 in die Trainingsgruppe aufgenommen. 33 Kinder erhielten keine zusätzliche Förderung. Das heißt, 59% der identifizierten Kinder mit Förderbedarf konnten nicht am Training teilnehmen, während 23% der Kinder, die in die Fördergruppe aufgenommen wurden, keinen festgestellten Förderbedarf aufweisen (siehe Abbildung 14).

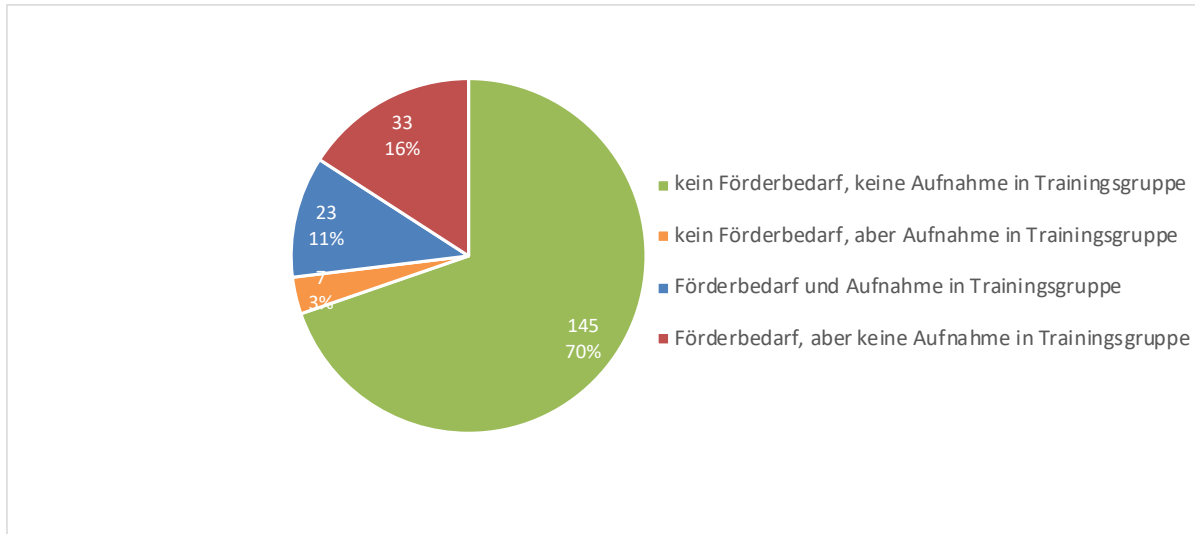


Abb. 14: Meilenstein 5 - Diagnostik und Förderung

### Meilenstein 6 (n = 198)

Bei 78% der Kinder wurde kein Förderbedarf diagnostiziert und sie wurden nicht der Trainingsgruppe zugeteilt. 14 Kinder, die laut Diagnostik keinen Förderbedarf aufweisen, wurden trotzdem in die Trainingsgruppe aufgenommen. Von den insgesamt 30 Kindern, die in diesem Meilenstein Leistungen in mindestens einer der Klassen- und der dazugehörigen Einzelaufgabe unterhalb des festgelegten Cut-Off-Werts zeigten, wurden 19 in die Trainingsgruppe aufgenommen. Elf Kinder erhielten keine zusätzliche Förderung. Das heißt, 37% der identifizierten Kinder mit Förderbedarf konnten nicht am Training teilnehmen, während 42% der Kinder, die in die Fördergruppe aufgenommen wurden, keinen festgestellten Förderbedarf aufweisen (Abbildung 15).

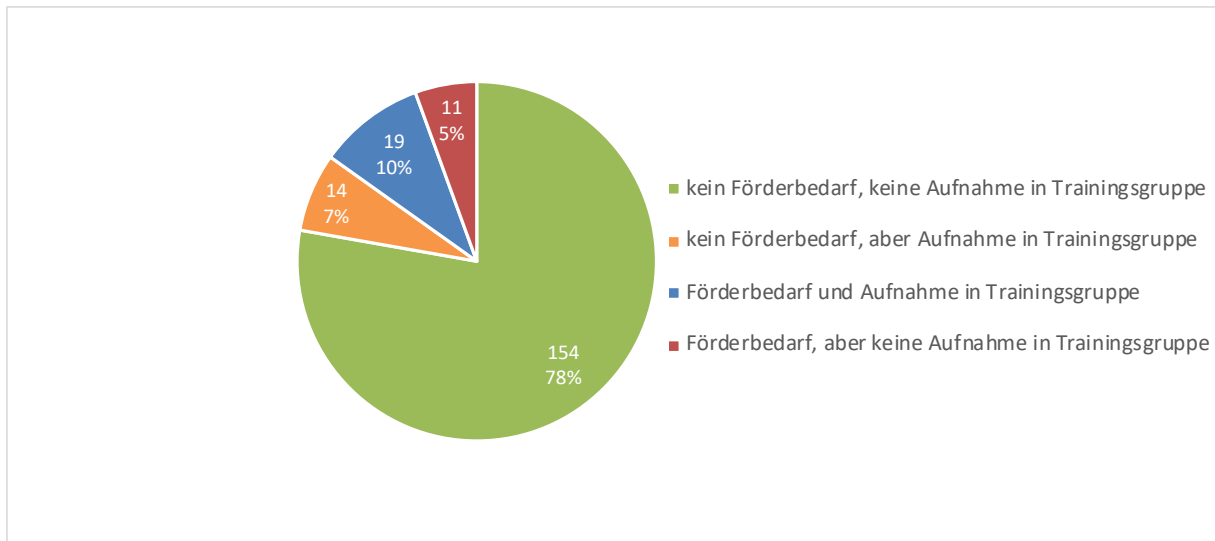


Abb. 15: Meilenstein 6 - Diagnostik und Förderung

### Meilensteine 1-6 (n= 1182)

Über alle Meilensteine hinweg zeigt sich das folgende Bild:

In 81% der Fälle wurden unauffällige Leistungen in den Aufgaben der Meilensteine beobachtet. Trotzdem wurden in 6% der Fälle Kinder, deren Leistungen nicht unter dem Cut-off-Wert lagen, einer Trainingsgruppe zugeordnet. Auf der anderen Seite wurde in 19% der Fälle ein Förderbedarf festgestellt, der jedoch nur in 10% der Fälle zu einer zusätzlichen Förderung der Kinder führte. Insgesamt wurden also in 48% der Fälle, in denen eine unterdurchschnittliche Leistung beobachtet wurde, darauf verzichtet, diese Kinder der Trainingsgruppe zuzuteilen (siehe Abbildung 16).

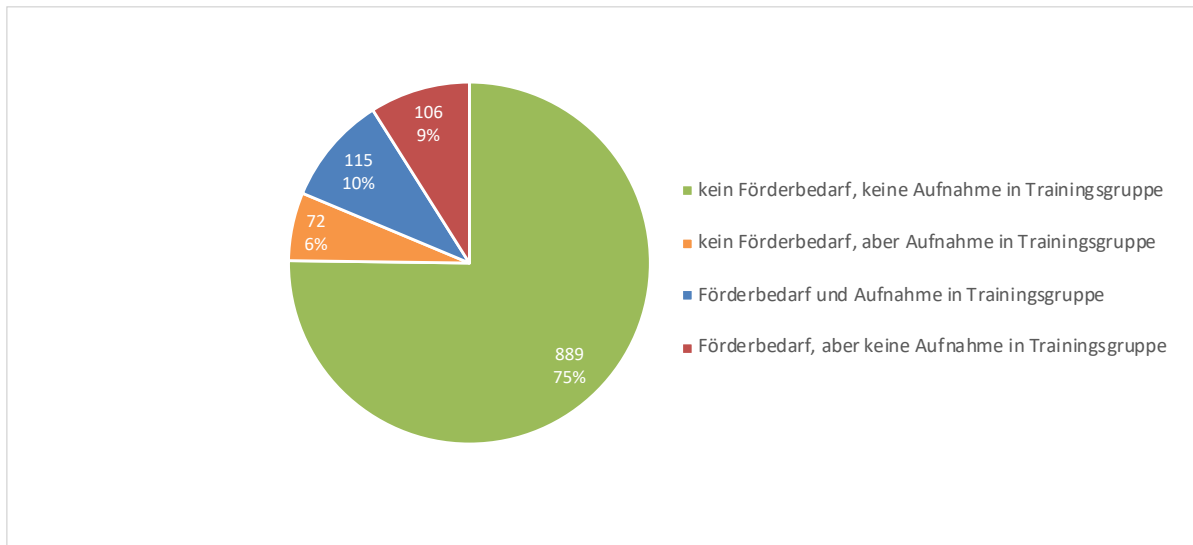


Abb. 16: Meilensteine 1-6: Diagnostik und Förderung

Über alle beobachteten Fälle hinweg, setzten sich die Trainingsgruppen zwar zum Großteil aus Kindern zusammen, bei denen eine Leistung unter dem Cut-off-Wert festgestellt wurde (61%), aber in 39% der Fälle erfolgte die Zuordnung zum Training, obwohl die Leistung der Kinder nicht auffällig war (siehe Abbildung 17).

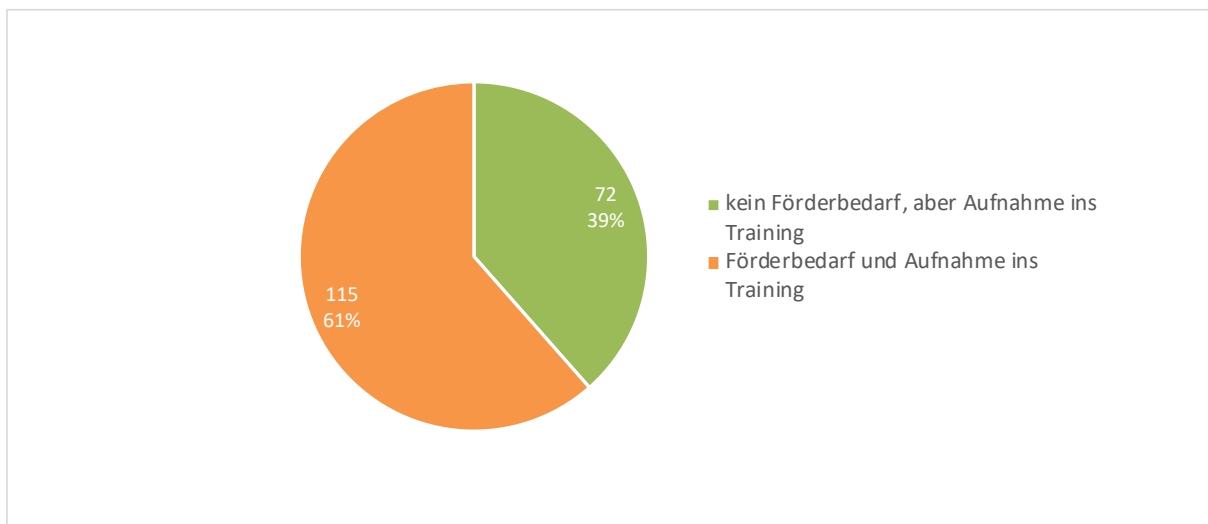


Abb. 17: Meilensteine 1-6 Zusammensetzung der Trainingsgruppe

### Durchführung der Förderung an den Schulen

Das Konzept sieht insgesamt 5 Förderbausteine mit unterschiedlichem Förderumfang vor:

- Baustein 1: 10 Wochen Training, 40 Fördereinheiten

- Baustein 2: 4 Wochen Training, 16 Fördereinheiten
- Baustein 3: 10 Wochen Training, 40 Fördereinheiten
- Baustein 4: 3 Wochen Training, 12 Fördereinheiten
- Baustein 5: 15 Wochen Training, 60 Fördereinheiten

Es liegen Daten von insgesamt 93 Schülerinnen und Schülern vor, die an mindestens einer Trainingsgruppe teilgenommen haben. Die Kinder teilen sich wie folgt auf die Schulen auf:

- Schule 201: 12 Kinder (13%) haben an der Förderung in mindestens einem Baustein teilgenommen.
- Schule 202: 16 Kinder (17%) haben an der Förderung in mindestens einem Baustein teilgenommen.
- Schule 203: 15 Kinder (16%) haben an der Förderung in mindestens einem Baustein teilgenommen.
- Schule 204: 27 Kinder (29%) haben an der Förderung in mindestens einem Baustein teilgenommen.
- Schule 205: 23 Kinder (25%) haben an der Förderung in mindestens einem Baustein teilgenommen.

Die Tabelle 2 stellt die Verteilung der Kinder nach der Anzahl der von ihnen durchlaufenen Förderbausteine pro Schule dar.

Schule	Anzahl teilgenommener Förderbausteine					Gesamt
	1	2	3	4	5	
<b>201</b>	4	2	3	2	1	12
<b>202</b>	5	4	5	1	1	16
<b>203</b>	8	1	2	3	1	15
<b>204</b>	15	6	4	2	0	27
<b>205</b>	11	8	3	1	0	23
<b>Gesamt</b>	43 (46%)	21 (23%)	17 (18%)	9 (10%)	3 (3%)	93

Tab. 2: Verteilung der Kinder in Fördergruppen

So haben beispielsweise an der Schule 204 insgesamt 27 Kinder eine Förderung erhalten. Von diesen 27 Kindern haben vier an drei Förderbausteinen teilgenommen, keines der Kinder erhielt in fünf Bausteinen Förderung.

Betrachtet man das Bild der Förderung von Kindern über alle Schulen hinweg, haben insgesamt 43 der 93 Kinder, also 46%, an nur einem Baustein teilgenommen, drei Kinder (3%) haben die Förderung aller fünf Bausteine durchlaufen.

Im Folgenden wird dargestellt, wie häufig die Förderung tatsächlich stattgefunden hat und ob die Schülerinnen und Schüler auch an der angebotenen Förderung teilgenommen haben. Die Zahlen basieren auf den Angaben der Lehrkräfte. Hinsichtlich der Teilnahme der Schülerinnen und Schüler an der Förderung ist zu beachten, dass jeweils die Teilnahme an den Fördereinheiten berichtet wird. Werden also in einer Schule statt der anvisierten 40 Fördereinheiten nur 26 angeboten (Schule 205), dann bezieht sich der prozentuale Anteil auf 26 Fördereinheiten und nicht auf die vorgesehenen 40.

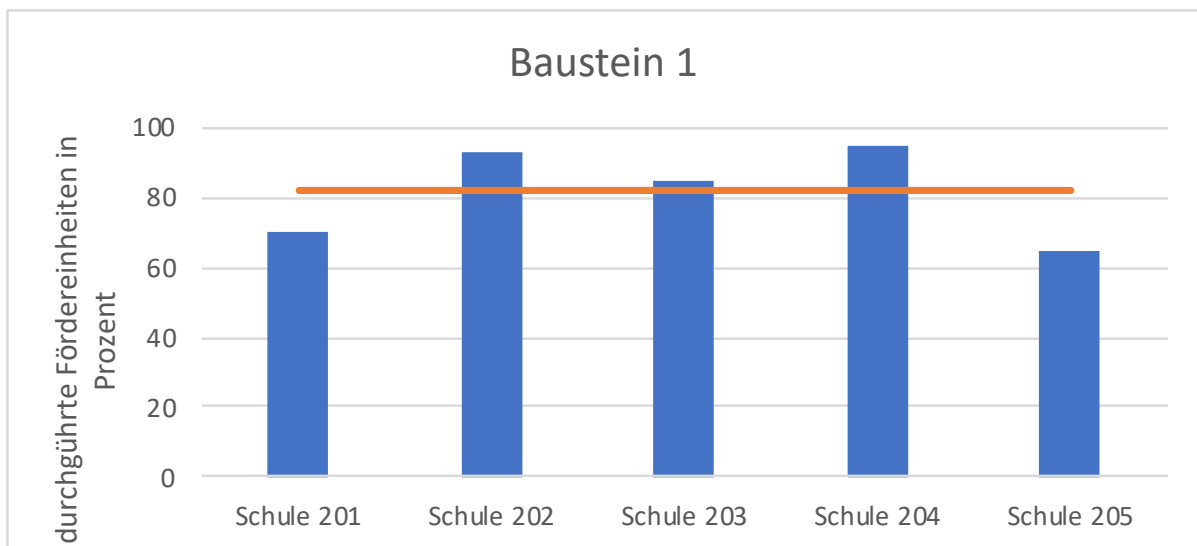


Abb. 18: Durchgeführte Fördereinheiten je Schule in Baustein 1 (Angaben in Prozent)

Im Rahmen des Baustein 1 sollten insgesamt 40 Fördereinheiten verteilt über 10 Wochen stattfinden. In einer Schule wurde ein Maximum von 38 Fördereinheiten (95%) durchgeführt, während in der Schule 205 nur 26 Fördereinheiten (64%) angeboten wurden.

Im Durchschnitt wurden 33 Fördereinheiten (82%) durchgeführt (siehe Abbildung 18).

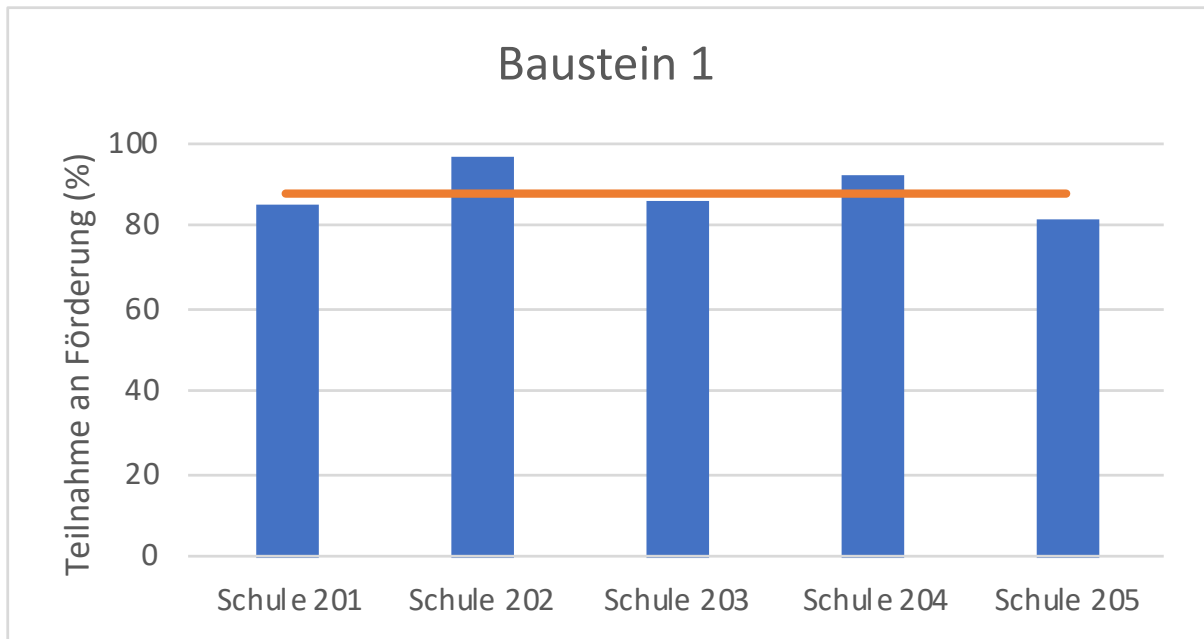


Abb. 19: durchschnittliche Teilnahme der Schülerinnen und Schüler an den angebotenen Fördereinheiten

Die Schülerinnen und Schüler nahmen an allen Schulen an über 80% der jeweils angebotenen Fördereinheiten teil. Im Durchschnitt über alle teilnehmenden Schülerinnen und Schüler konnten 87% an der Förderung teilnehmen (siehe Abbildung 19).

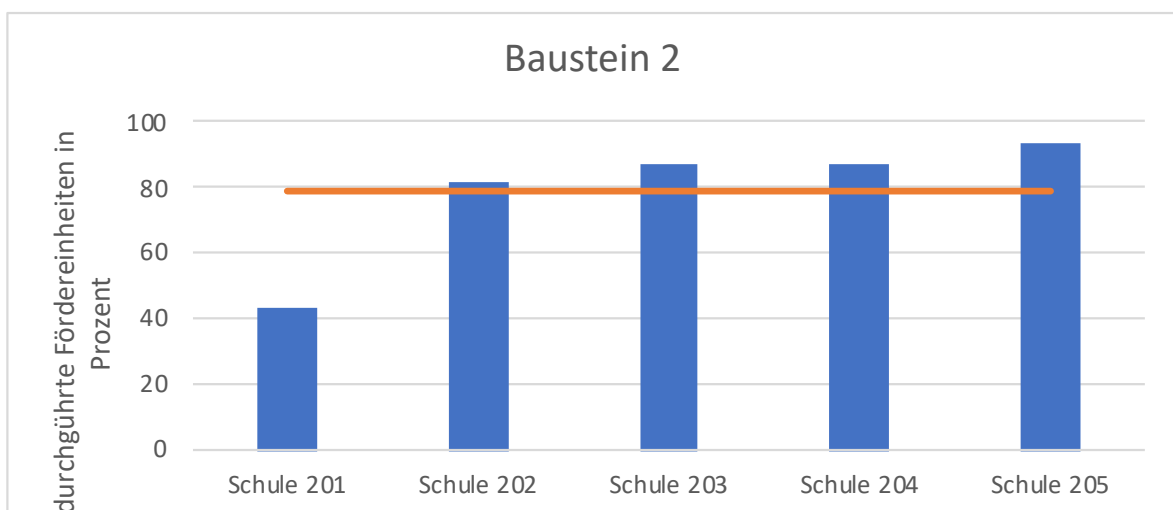


Abb. 20: Durchgeführte Fördereinheiten je Schule in Baustein 2 (Angaben in Prozent)

Für den Förderbaustein 2 sind 16 Fördereinheiten verteilt über vier Wochen vorgesehen. Die maximale Anzahl durchgeführter Fördereinheiten war 15 (84%), gegenüber der minimalen Anzahl von durchgeführten Fördereinheiten von 7 (44%). Im Durchschnitt wurden 13 Fördereinheiten (79%) durchgeführt (siehe Abbildung 20).

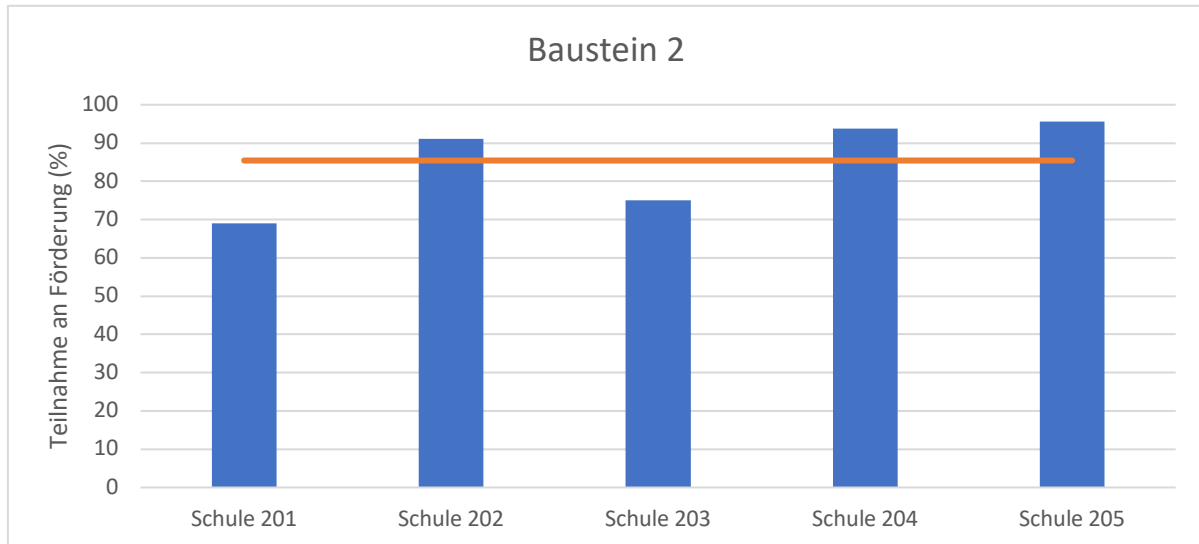


Abb. 21: durchschnittliche Teilnahme der Schülerinnen und Schüler an den angebotenen Fördereinheiten

Die Teilnahme an der Förderung variiert in Baustein 2 stärker als in Baustein 1. In Schule 201 wurden bspw. nur 7 der eigentlich vorgesehenen 16 Fördereinheiten durchgeführt, teilgenommen haben die Kinder in dieser Schule jedoch durchschnittlich nur an 5 Fördereinheiten. Im Durchschnitt besuchten alle teilnehmenden Schülerinnen und Schüler 85% der angebotenen Förderung (siehe Abbildung 21).



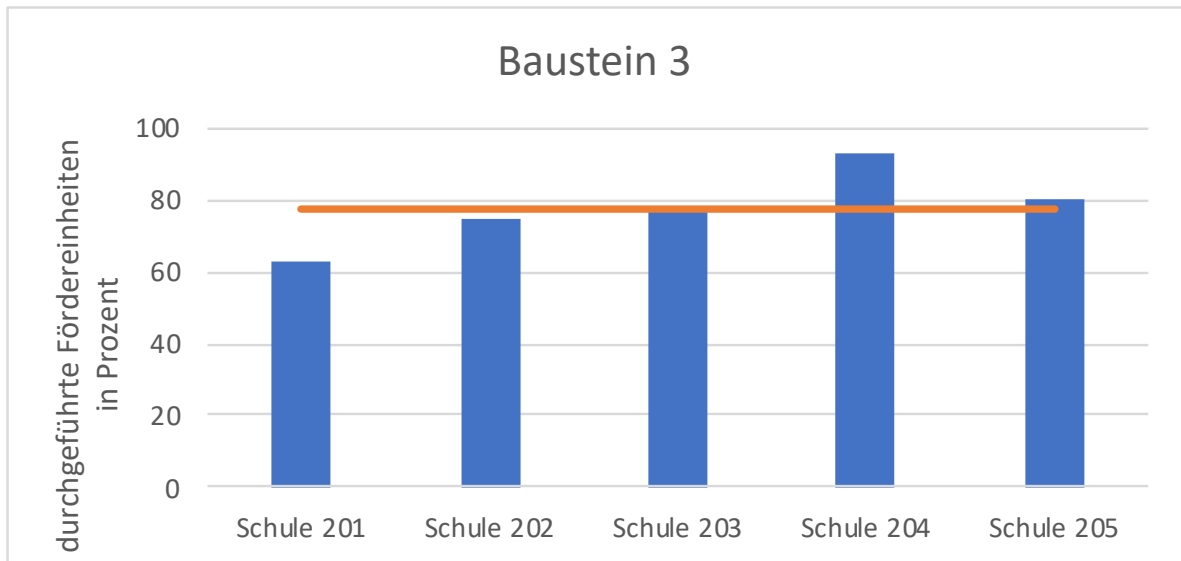


Abb. 22: Durchgeführte Fördereinheiten je Schule in Baustein 3 (Angaben in Prozent)

Für den Förderbaustein 3 sind 40 Fördereinheiten verteilt über zehn Wochen vorgesehen. Die maximale Anzahl durchgeführter Fördereinheiten war 37 (93%), gegenüber der minimalen Anzahl von durchgeführten Fördereinheiten von 25 (63%). Im Durchschnitt wurden 31 Fördereinheiten (78%) durchgeführt (siehe Abbildung 22).

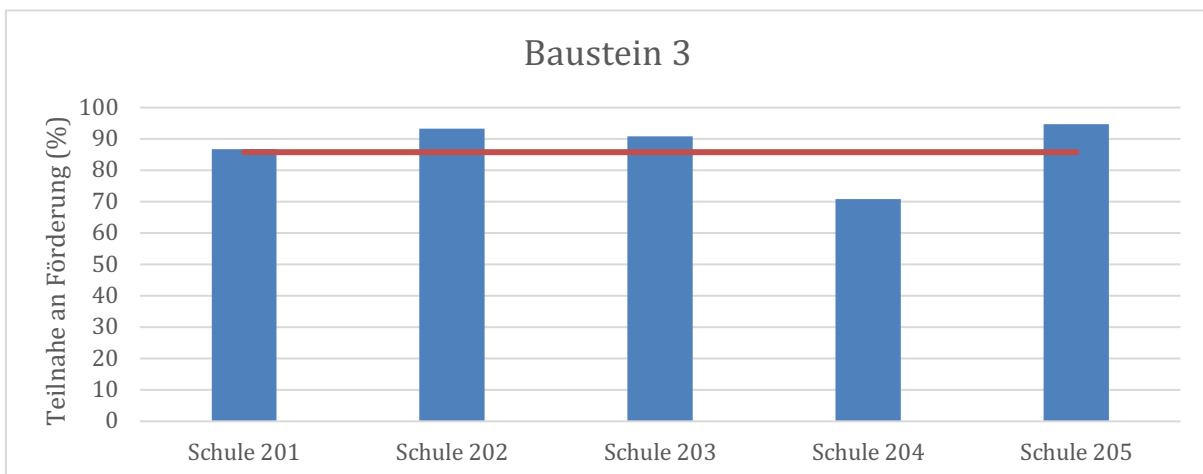


Abb. 23: durchschnittliche Teilnahme der Schülerinnen und Schüler an den angebotenen Fördereinheiten

Der Großteil der Schülerinnen und Schüler nahm regelmäßig an der angebotenen Förderung teil. Im Schnitt wurden 85% der angebotenen Förderung von den Kindern auch besucht (siehe Abbildung 23).

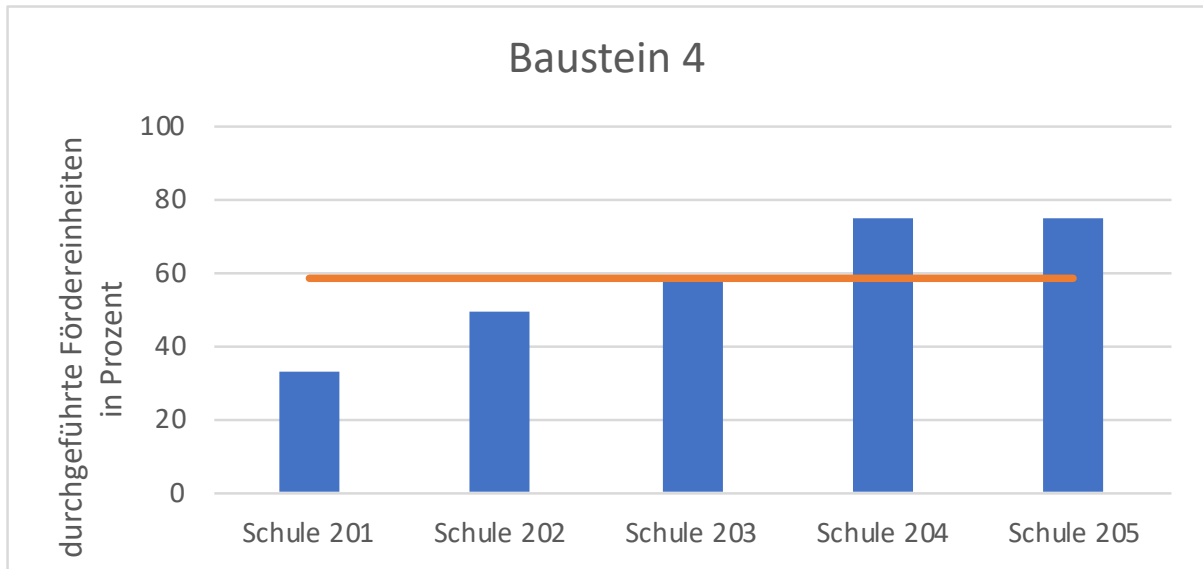


Abb. 24: Durchgeführte Fördereinheiten je Schule in Baustein 4 (Angaben in Prozent)

Für den Förderbaustein 4 sind 12 Fördereinheiten verteilt über drei Wochen vorgesehen. Die maximale Anzahl durchgeführter Fördereinheiten war 9 (75%), gegenüber der minimalen Anzahl von durchgeführten Fördereinheiten von 4 (33%). Im Durchschnitt wurden 7 Fördereinheiten (58%) durchgeführt (siehe Abbildung 24).

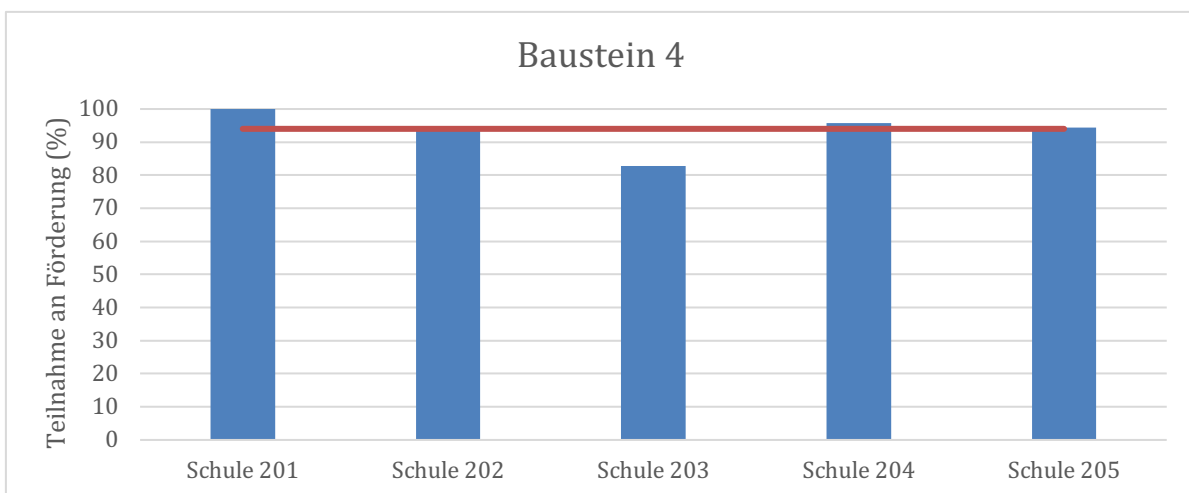


Abb. 25: durchschnittliche Teilnahme der Schülerinnen und Schüler an den angebotenen Fördereinheiten

Der Großteil der Schülerinnen und Schüler nahm regelmäßig an der angebotenen Förderung teil. Im Schnitt wurden 94% der angebotenen Förderung von den Kindern auch besucht (siehe Abbildung 25).

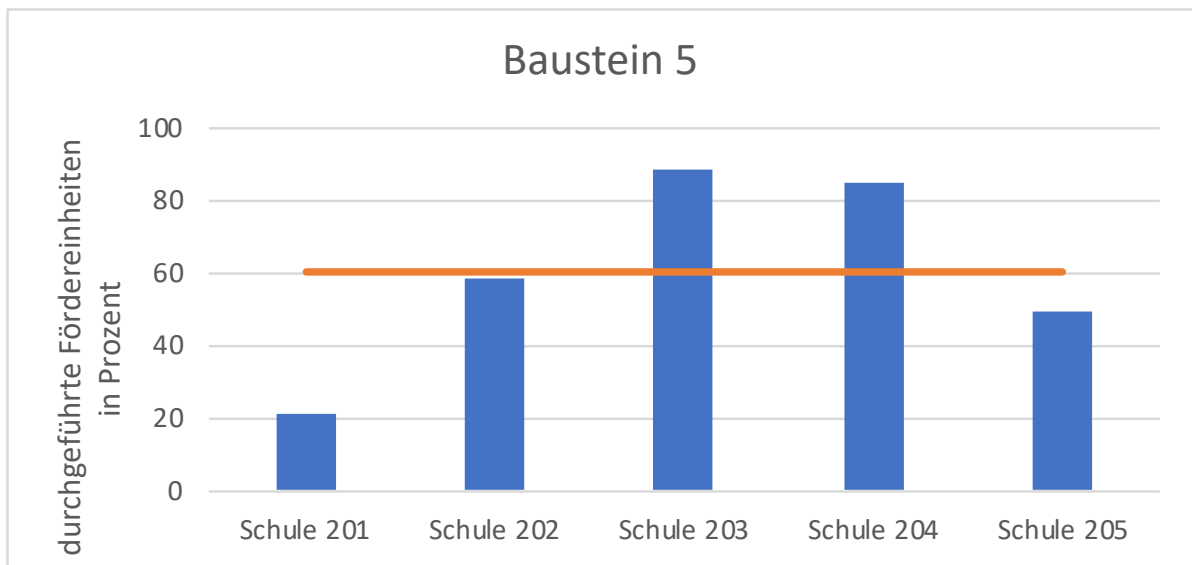


Abb. 26: Durchgeführte Fördereinheiten je Schule in Baustein 5 (Angaben in Prozent)

Für den Förderbaustein 5 sind 60 Fördereinheiten verteilt über 15 Wochen vorgesehen. Die maximale Anzahl durchgeführter Fördereinheiten war 53 (88%), gegenüber der minimalen Anzahl von durchgeführten Fördereinheiten von 13 (22%). Im Durchschnitt wurden 36 Fördereinheiten (61%) durchgeführt (siehe Abbildung 26).

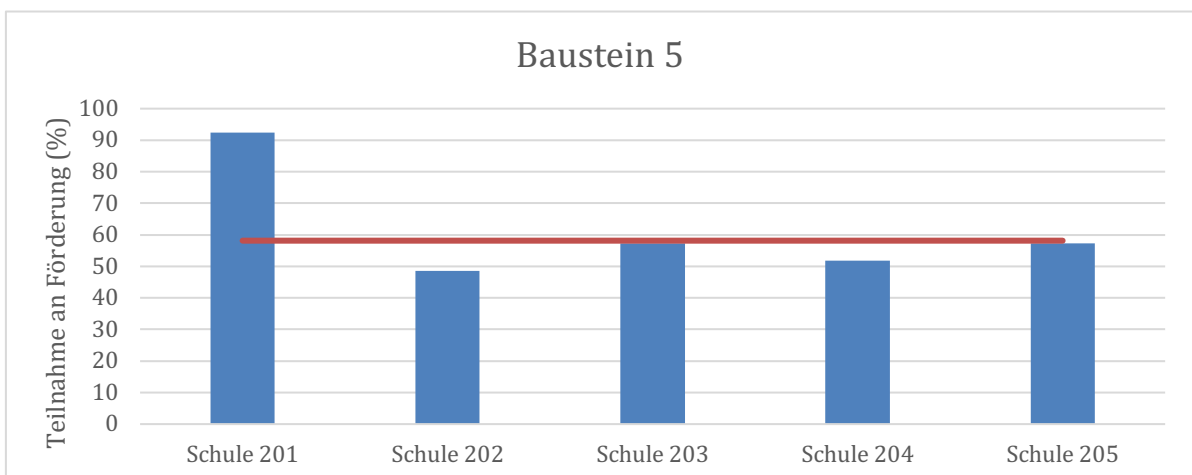


Abb. 27: durchschnittliche Teilnahme der Schülerinnen und Schüler an den angebotenen Fördereinheiten

Viele Schülerinnen und Schüler verpassten einen großen Teil der angebotenen Förderung. Durchschnittlich wurden in Baustein 5 nur 58% der angebotenen Förderung von den Kindern auch wahrgenommen (siehe Abbildung 27). Einzig die Schule 201 sticht hier hervor. Die Kinder dieser Schule nahmen 92% des Förderangebots war, was jedoch auch gegenüber den anderen Schulen deutlich reduziert war (13 statt der anvisierten 60 Fördereinheiten).

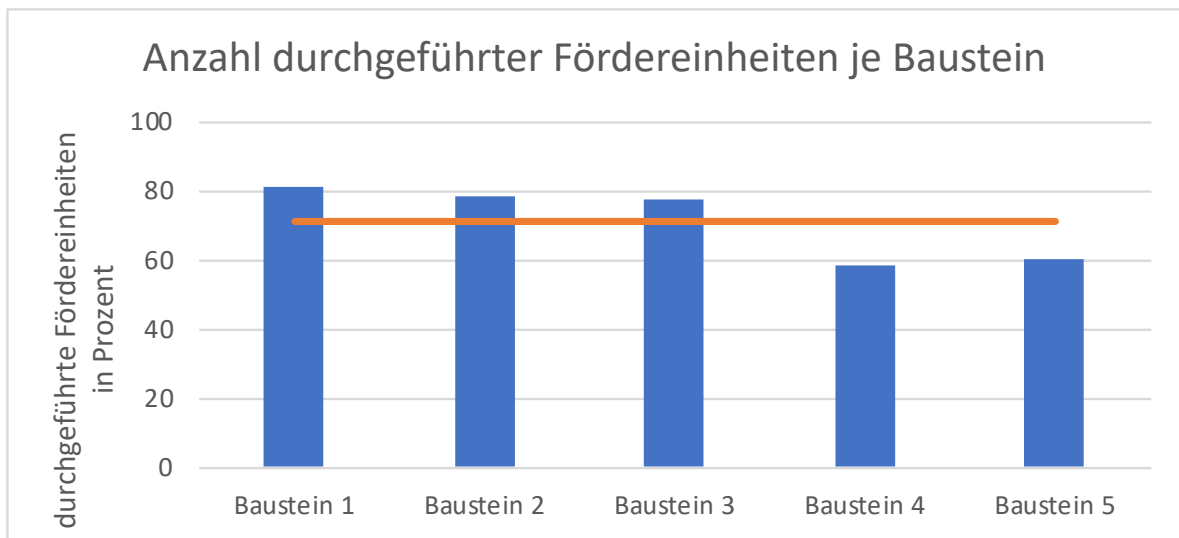


Abb. 28: Anzahl durchgeführter Fördereinheiten je Baustein (Angaben in Prozent)

In Abbildung 28 ist zu erkennen, dass die Anzahl der durchgeführten Fördereinheiten je Baustein im Laufe der Zeit abnimmt. Waren es zu Beginn der Fördermaßnahme noch durchschnittlich 82% der geplanten Einheiten, die im Schnitt auch durchgeführt wurden, sinkt die Anzahl der durchgeführten Fördereinheiten insbesondere bei den Bausteinen 4 und 5 auf etwa 60%. Im Durchschnitt wurden über alle Bausteine hinweg 71% der anvisierten Fördereinheiten auch tatsächlich angeboten.

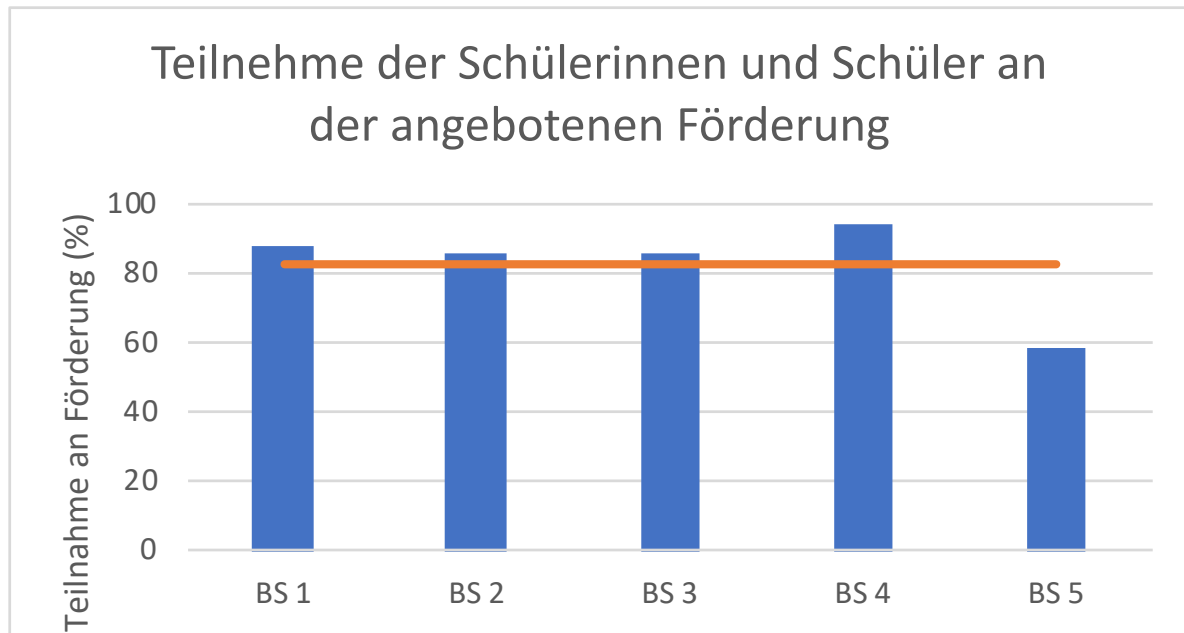


Abb. 29: durchschnittliche Teilnahme der Schülerinnen und Schüler an den angebotenen Fördereinheiten

Aus Abbildung 29 wird ersichtlich, dass die Schülerinnen und Schüler zumeist regelmäßig an der angebotenen Förderung teilnahmen. Allein im Baustein 5 zeigt sich eine deutlich verminderte Teilnahme an nur 58% der angebotenen Förderung. Über alle Bausteine hinweg besuchten die Kinder 82% der angebotenen Förderung.

Insgesamt lässt sich feststellen, dass viele Daten fehlen, weil die Lehrkräfte ihr Vorgehen nicht komplett dokumentiert haben. Die Auswertung der vorhandenen Daten zeigt, dass die Diagnostik nicht konsequent von den Lehrkräften als Basis für die Zuteilung der Kinder in die Fördergruppen genutzt wurde. Viele Kinder mit festgestelltem Förderbedarf erhielten keine zusätzliche Förderung; ein Großteil der Kinder nahm an nur einem Förderbaustein teil. Es lässt sich zudem beobachten, dass die Durchführung der Förderung an den Schulen stark zwischen den Fördergruppen und zwischen den Bausteinen variierte. Dabei wurde zu Beginn der Studie die Förderung noch relativ konzepttreu angeboten, im Verlauf nahm das Angebot an Förderung jedoch ab. Die Schülerinnen und Schüler, die einer Fördergruppe zugeteilt waren, nahmen aber regelmäßig an der angebotenen Förderung teil. Im Baustein 5 konnte keine regelmäßige Teilnahme der Kinder mehr stattfinden, da die Pandemie zu diesem Zeitpunkt ihren Anfang nahm.

Zu 1d) Inwieweit wird die Förderung konzepttreu umgesetzt?

Im Rahmen der Untersuchung der Wiedergabetreue wurde der Teilfrage nachgegangen, inwieweit die Förderung konzepttreu umgesetzt wurde. Als Erhebungsinstrument dienten nicht-teilnehmende Beobachtungen von Fördereinheiten, die mittels eines strukturierten Beobachtungsbogens dokumentiert wurden. Insgesamt fanden 19 dieser Beobachtungen statt, die sich auf die Bausteine 2 bis 5 und die Umsetzung der Fördereinheiten an fünf verschiedenen Schulen erstreckten. Durchgeführt wurden diese Erhebungen von einer geschulten Beobachterin, um Qualität und Vergleichbarkeit der Beobachtungen sicherzustellen.

Der strukturierte Beobachtungsbogen umfasste folgende Teile:

Teil A – Allgemeine Informationen (z.B. Datum, Schulstunde, Trainingseinheit)

Teil B – Offene Beobachtungen zu den einzelnen vorgegebenen Phasen der Fördereinheit (z.B. Dauer, Übereinstimmung mit den Inhalten und Methoden im Manual, verwendete Medien) (siehe Abbildung 30)

Teil C – Geschlossene Beobachtungen zu Schlüsselmerkmalen des Trainingsprogramms anhand von Likert-Skalen (z.B. Arbeit in Tandems, Differenzierung nach Lernstand, Nutzung von Lernkarteien) (siehe Abbildung 31)

<i>Phase</i>	<i>Dauer</i> <i>(Beginn &amp; Ende)</i>	<i>Inhalte &amp; Beobachtungen</i>	<i>Materialien &amp; Medien</i>
<i>Herstellung von Zieltransparenz</i> <small>(Inhaltliche Vorbereitung der SuS; präzise Benennung des Trainingsziels zur Herstellung von Motivation und Transparenz)</small>			
<i>Aktivierung des Vorwissens</i> <small>(Spielerische Wiederholung bisheriger Inhalte, um mit neuen Inhalten an bestehendes Wissen anzuknüpfen)</small>			
<i>Instruktion</i> <small>(Einführung neuer Informationen und Inhalte; direkte und explizite Vermittlung)</small>			

Abb. 30: Ausschnitt von Teil B des Beobachtungsbogens

	Trifft nicht zu (1)	(2)	(3)	(4)	Trifft voll zu (5)
<i>Alle SuS haben einen aufgeräumten und v.a. ungestörten Arbeitsplatz.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Alle SuS arbeiten während der Phase der Automatisierung in Tandems.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Während der Automatisierungsphase arbeiten Partner mit ähnlichem Lern-/ Entwicklungsstand zusammen.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 31: Ausschnitt von Teil C des Beobachtungsbogens

Die Auswertung der durch die Beobachtung erfassten Daten ergibt, dass die Umsetzung der Trainingseinheiten hinsichtlich wesentlicher Aspekte konzepttreu erfolgte: So kann die Umsetzung der vorgegebenen Gruppengröße bestätigt werden (4-8 Schülerinnen und Schüler in den beobachteten Einheiten). Außerdem wurde die Einteilung in unterschiedliche Levels in den Heften durchgängig genutzt ( $\emptyset$  5 von 5 Punkten) und die Automatisierungsphase fand stets in Tandems statt ( $\emptyset$  4 von 5 Punkten), die möglichst leistungshomogen waren ( $\emptyset$  5 von 5 Punkten).

Allerdings wurden auch einige Abweichungen beobachtet, die eine Gefährdung der Konzepttreue darstellen: In einem Drittel der beobachteten Fördereinheiten wurde die erste Trainingsphase („Herstellung von Zieltransparenz“) nicht durchgeführt. Darüber hinaus griffen in etwa der Hälfte der Einheiten die verschiedenen Phasen des Trainingsprogramms nicht ineinander und in ca. jeder fünften Einheit gingen die Lehrkräfte nicht auf den spezifischen Unterstützungsbedarf der einzelnen Schülerinnen und Schüler ein. Neben diesen allgemeinen Auffälligkeiten ergaben sich außerdem Abweichungen, die ausschließlich bei Baustein 4 auftraten. Hier hielten sich die Lehrkräfte durchweg nicht genau an die im Manual Vorgehensweisen, sondern brachten eigene Ideen ein. Des Weiteren wirkten die Kinder in diesem Baustein häufig durch die Arbeit mit dem Trainingsprogramm überfordert und die Trainingsintervalle von 15 bis 20 Minuten wirkten in fast allen Fällen vom Umfang her nicht angemessen.

#### 4.2.2 Bewertung der Fortbildung

Der Fragebogen wurden insgesamt an vier Messzeitpunkten (n = 7, n = 5, n = 6, n = 6) jeweils nach Durchführung der Fortbildung eingesetzt und vor Ort von den beteiligten Lehrkräften ausgefüllt.

Der Fragebogen umfasst zwei Teile: Im ersten Teil werden zehn allgemeine Fragen zur Bewertung der Fortbildung erfasst. Dort finden sich Items, wie z.B. „Die Fortbildung war klar und gut strukturiert.“, „Die Ziele der Fortbildung waren klar und nachvollziehbar.“, „Die Fortbildungsleiterin wirkte während der Fortbildung engagiert.“ oder „In dieser Fortbildung habe ich viel dazu gelernt.“ bzw. „Insgesamt bin ich mit der Fortbildung zufrieden.“

Die Lehrkräfte bewerteten die Aussagen auf einer vierstufigen Likert-Skala von „Stimme überhaupt nicht zu“ über „Stimme eher nicht zu“ und „Stimme eher zu“ zu „Stimme voll zu“. Für die Auswertung wurden die Werte summiert und dann zur besseren Vergleichbarkeit Prozentwerte gebildet. Abbildung 32 zeigt die Entwicklung der Bewertungen über die Messzeitpunkte hinweg. Es wird deutlich, dass die Fortbildungen von den teilnehmenden Lehrkräften durchweg sehr positiv bewertet werden. Die Zustimmung liegt in einem Bereich von 98,6 % zu MZP 1 bis 95,4 % zu MZP 4.

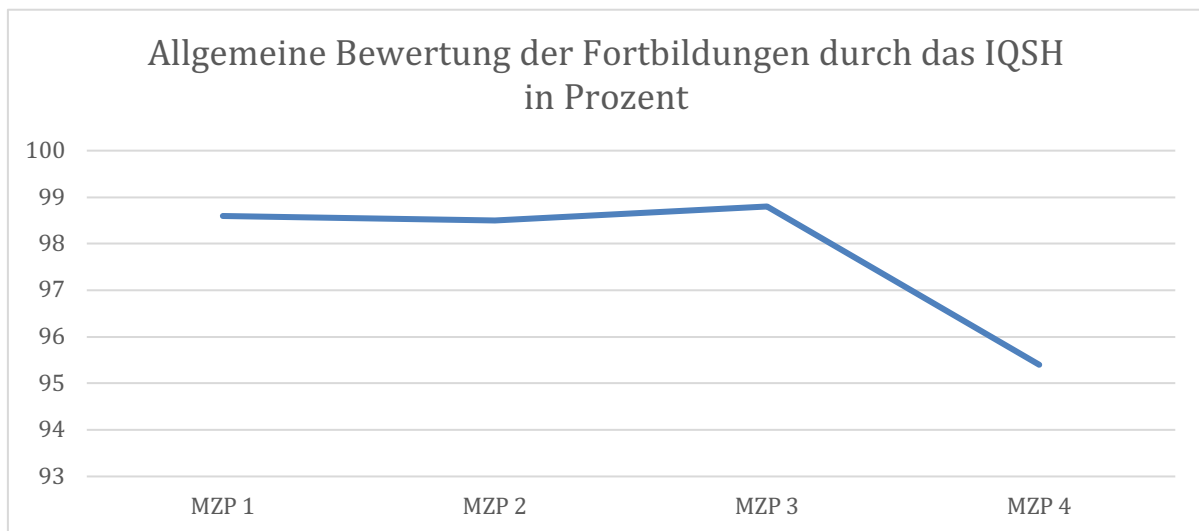


Abb. 32: Allgemeine Bewertung der Fortbildungen

Im zweiten Teil des Fragebogens finden sich insgesamt 13 Items, die die Fortbildungen hinsichtlich der Vermittlung fachlicher Inhalte bewerten. Die folgende Tabelle 3 zeigt die Items, die positiv bzw. negativ formuliert wurden. Die Lehrkräfte bewerteten die Aussagen auf einer vierstufigen Skala von „Stimme überhaupt nicht zu“ über „Stimme eher nicht zu“ über „Stimme eher zu“ zu „Stimme voll zu“.



Positiv formulierte Items	Negativ formulierte Items
Die Fortbildungsinhalte zu „Lms-Training“ schließen gut an mein bisheriges Verständnis von Lese- und Schreibförderung an.	Ich fand die Fortbildungsinhalte zu komplex.
Wenn ich mit „Lms-Training“ so arbeite wie in der Fortbildung vermittelt wurde, passt das gut zu meinem Verständnis von regelmäßigen und klar strukturierten Übungsreihen.	Die Fortbildung hat bei mir mehr Fragen ausgelöst als sie beantwortet hat.
Die heutige Fortbildung lässt mich „Lms-Training“ besser verstehen als vorher.	Die Umsetzung der Fortbildungsinhalte in meiner Fördergruppe kommt mir sehr schwierig vor.
Durch die Fortbildungsinhalte habe ich ein besseres Verständnis davon, wie ich mit „Lms-Training“ arbeiten kann.	Ich kann mir nicht vorstellen, wie ich die Fortbildungsinhalte in meiner Fördergruppe mit „Lms-Training“ umsetzen soll.
Ich denke, dass die Fortbildung dazu führt, dass ich Kinder mit Lernschwierigkeiten besser unterstützen kann als vorher.	
Die heutige Fortbildung erleichtert mir die Arbeit mit „Lms-Training“ in meiner Fördergruppe.	
Die Fortbildung hat mir Inhalte vermittelt, die mir vorher nicht so bekannt waren.	
Die Fortbildung hat mir Inhalte vermittelt, die mir für meine Arbeit in der Schule weiterhelfen.	

Tab. 3: Items zur Bewertung der fachlichen Inhalte in der Fortbildung

Tabelle 3 zeigt differenziert solche Items auf, die positiv formuliert wurden und solche, die negativ formuliert wurden. Die Auswertung trennt nach diesen beiden Kategorien. Für die Auswertung wurden die Bewertungen summiert und zur Vergleichbarkeit in Prozentangaben umgewandelt.

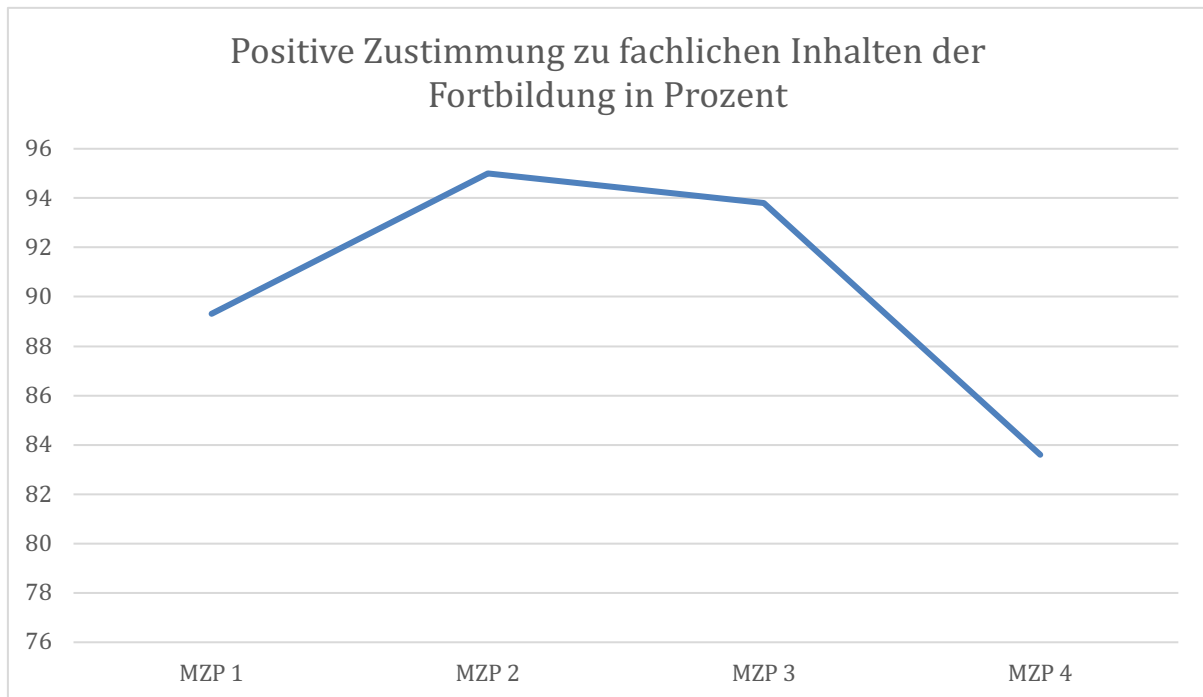


Abb. 33: Positive Zustimmung zu fachlichen Inhalten der Fortbildung

Abbildung 33 zeigt die Entwicklung der Bewertung der teilnehmenden Lehrkräfte über die vier Messzeitpunkte hinweg bezogen auf die fachlichen Inhalte. Es ist zu erkennen, dass auch die Vermittlung fachlicher Inhalte durchgehend sehr positiv, mit Werten von 83,6 % bis hin zu 95 % bewertet wurden. Zwar ist hier von Messzeitpunkt (MZP) 3 zu MZP 4 eine Verschlechterung der Bewertungen zu sehen, trotzdem ist der Wert mit 83,6 % immer noch sehr hoch.

Neben den positiv formulierten Items finden sich auch solche, die negative Aussagen beinhalten (s. Tabelle 3).

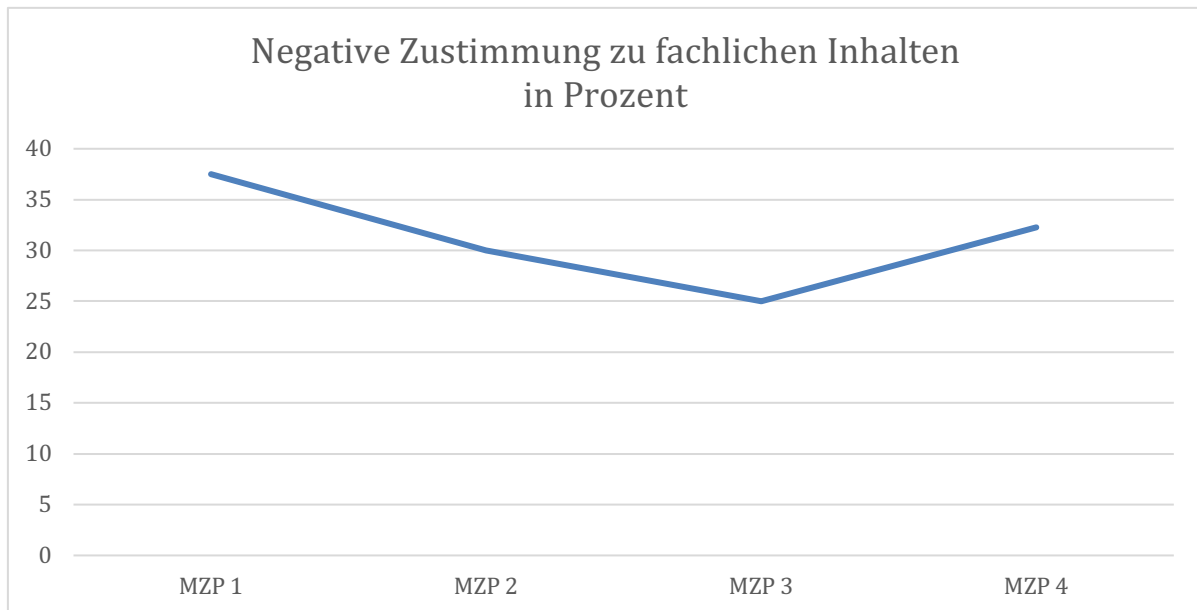


Abb. 34: Negative Zustimmung zu fachlichen Inhalten der Fortbildung

Abbildung 34 zeigt die Zustimmung zu negativ formulierten Items über die Messzeitpunkte hinweg auf. Es ist zu erkennen, dass die teilnehmenden Lehrkräfte den Aussagen mit Werten zwischen 25 % und 37,5 % eher nicht zustimmen. Im Umkehrschluss bedeutet dies, dass sie die Vermittlung fachlicher Inhalte bei den Fortbildungen eher positiv bewerten.

Insgesamt lässt sich feststellen, dass die Fortbildungen, die durch das IQSH durchgeführt wurden von den teilnehmenden Lehrkräften sowohl bezogen auf die allgemeinen Aspekte der Durchführung, wie z.B. die Expertise der Fortbildnerinnen, als auch bezogen auf die spezifischen fachlichen Inhalte als sehr positiv bewertet wurden.

#### 4.2.3 Bewertung des Konzeptes und des Materials durch die Lehrkräfte

Der Lehrkraftfragebogen richtet sich an Lehrkräfte, die das Lms-Training angewendet haben, und an die Lehrkräfte der Kontrollgruppe. Neben persönlichen Angaben zur Person und beruflichem Hintergrund ermittelt der Fragebogen, wie Lehrkräfte den Einsatz und die Nutzung des Trainings bewerten. Am Ende dieses Fragebogens erhielten die Lehrkräfte außerdem die Möglichkeit, ihre eigenen Gedanken und Anregungen niederzuschreiben. Der Fragebogen wurde zu insgesamt fünf Messzeitpunkten an die Experimental- und an die Kontrollgruppe ausgegeben (EG = 4 MZP; KG = 1 MZP). Der Grund für die mehrmalige Ausgabe des Fragebogens liegt in der Unterschiedlichkeit der Trainingsbausteine: Die Bausteine 1, 2 und 5 fördern das Lesen; die Bausteine 3 und 4 fördern das Schreiben. Die Befragung wurde jeweils nach den einzelnen Bausteinen durchgeführt. Die Bausteine 3 und 4 wurden dabei zusammengefasst, weil beide Bausteine inhaltlich und im Ablauf gleich waren. Daher ergeben

sich vier Messzeitpunkte in der Experimentalgruppe und ein Messzeitpunkt in der Kontrollgruppe.

Die folgende Tabelle 4 führt die einzelnen Konstrukte des Lehrerfragebogens auf.

<b>Teil A: Personenbezogene Hintergrunddaten</b>		
<b>Konstrukt</b>	<b>Items</b>	<b>Variablen (Items)</b>
(1) Persönlicher und beruflicher Hintergrund	6	Geschlecht Alter Qualifikation
(2) Schulpraxis	6 (KG)	Diagnose Fördermaterialien Förderung
<b>Teil B: Akzeptanz der Intervention (EG)</b>		
<b>Konstrukt</b>	<b>Items</b>	<b>Variablen</b>
(3) Lms-Diagnostik	4	Zuweisung Förderung Bezug Förderung/ Diagnose
(4) Training/ Nutzung	8	Verwendung Material Ergänzung durch andere Materialien Vollständigkeit & Präzision
(5) Akzeptanz/ Bewertung	25	Bewertung von Lms-Training:

		Angemessenheit: Umfang/ Dauer/ Gruppengröße  Akzeptanz & Nutzung im Kollegium  Verzahnung, Diagnose & Förderung  Auswahl inhaltlicher Schwerpunkte
--	--	---

Tab. 4: Konstrukte des Lehrerfragebogens

Insgesamt haben 23 Lehrkräfte an der Studie teilgenommen. Davon haben 12 Lehrkräfte mit dem Lms-Training gearbeitet, elf Lehrkräfte waren Teil der Kontrollgruppe. Die einzelnen Konstrukte werden im Folgenden ausgewertet. **Konstrukt (1) umfasst Hintergrundmerkmale der teilnehmenden Lehrkräfte.** Die folgende Tabelle 5 zeigt das Alter zu Beginn der Studie und das Geschlecht der teilnehmenden Lehrkräfte.

	Alter			Geschlecht		
	Jahre	Mittelwert	fehlend	weiblich	männlich	fehlend
IG (N = 12)	26 – 50	37,8	3	10	0	2
KG (N = 11)	26 – 64	43,9	0	9	0	2

Tab. 5: Alter und Geschlecht der Lehrkräfte

Die Qualifikation der Lehrkräfte wurde mit Fragen nach ihrem universitären Abschluss, nach etwaigen Zusatzqualifikationen sowie über die Dauer ihrer beruflichen Tätigkeit als Lehrkraft erfasst. In der Experimentalgruppe haben alle Lehrkräfte (N = 9) das Fach Deutsch für das Lehramt der Primarstufe studiert. Von den teilnehmenden Lehrkräften haben vier eine Zusatzqualifikation zum Thema „Deutsch als Zweitsprache“ (8-mal keine Angabe). Auf die Frage „Seit wie vielen Jahren unterrichten Sie an einer Grund- oder Gemeinschaftsschule?“ antworten die Lehrkräfte im Mittel mit 10,6 Jahren (Min. = 2; Max. = 20).

In der Kontrollgruppe haben die Lehrkräfte zwischen zwei und 30 Jahren an einer Grund- oder Gemeinschaftsschule unterrichtet, im Mittel 14,8 Jahre.

**Das Konstrukt (2) erfasst den Umgang mit Diagnose- und Fördermaterialien** und die Maßnahmen zur Lese- und Schreibförderung der Lehrkräfte in den Kontrollschulen. Es wurde erfasst, aufgrund welcher Informationen die Lehrkräfte eine Entscheidung über den Förderbedarf der Kinder treffen, in welcher Form und wie umfangreich Förderstunden eingesetzt werden und ob die Lernfortschritte der Schülerinnen und Schüler strukturiert und formal (z.B. mit Hilfe von Testmaterialien) erfasst werden.

Abbildung 35 zeigt die Ergebnisse hinsichtlich der Verwendung von Diagnoseinstrumenten. Die Lehrkräfte der Kontrollgruppe geben mit 72,7 % an, dass sie den Lernverlauf ihrer Lernenden strukturiert mit Hilfe von Testmaterialien beobachten; 18,2 % hingegen verneinen dies. Während 63,3 % der Lehrkräfte angeben, dass sie diese Ergebnisse für Schülerfeedbacks nutzen, verneinen dies 27,3 % (siehe Abbildung 36).

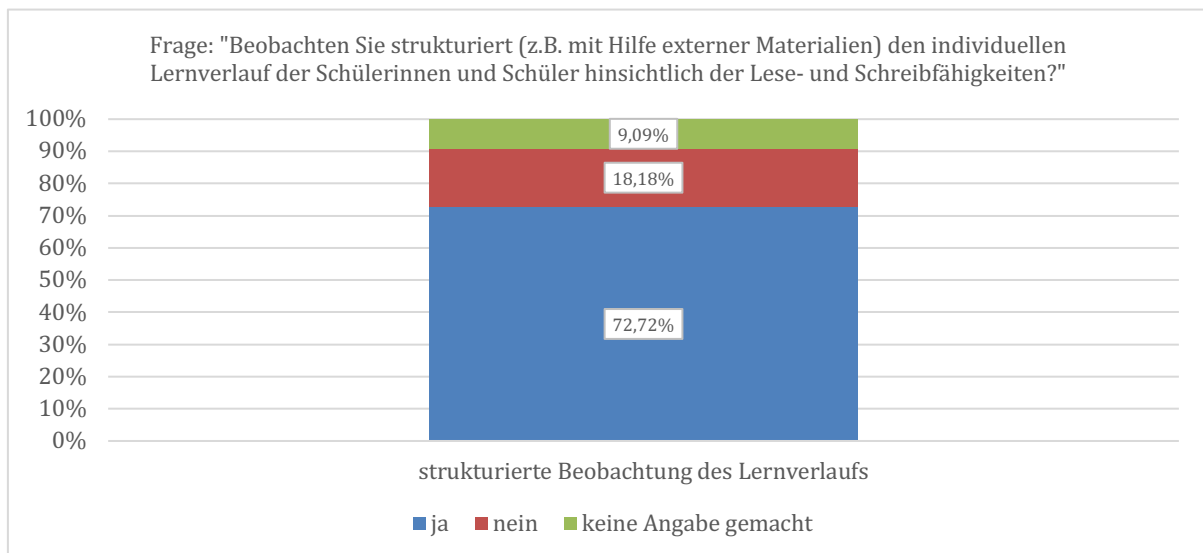


Abb. 35: Strukturierte Beobachtung des individuellen Lernverlaufs der Schülerinnen und Schüler hinsichtlich Lese- und Schreibfähigkeiten (N = 11)

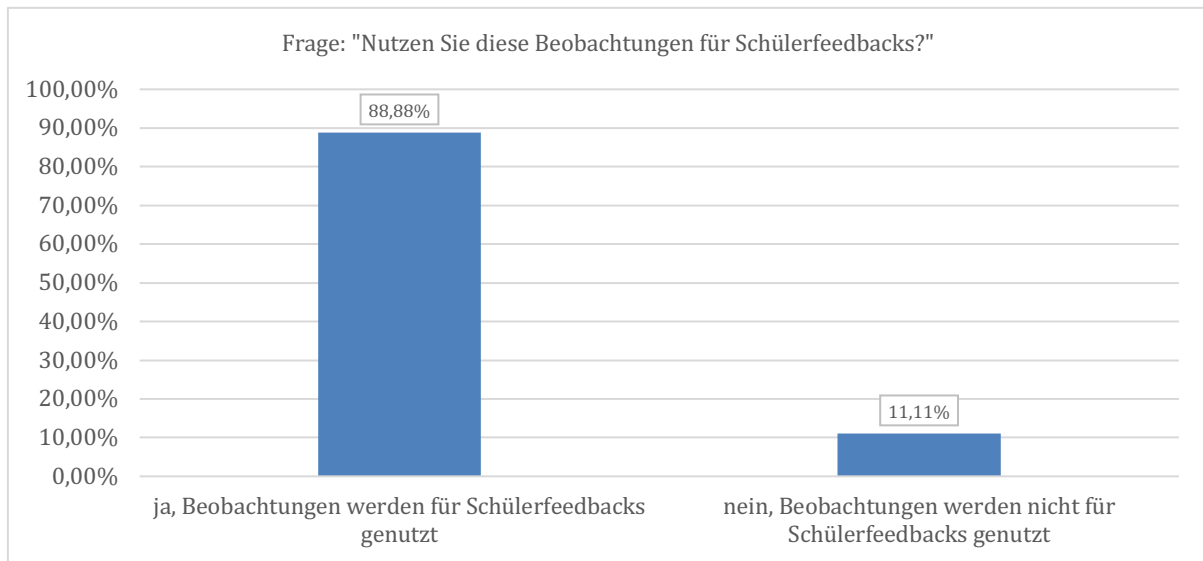


Abb. 36: Einsatz der Beobachtungen für Schülerfeedbacks (N = 11)

Bei der Frage „Aufgrund welcher Informationen treffen Sie eine Entscheidung über den Förderbedarf eines Kindes?“, hatten die Lehrkräfte verschiedene Auswahlmöglichkeiten, die in Tabelle 6 dargestellt sind:

	in Prozent		
	ja	nein	fehlende Angabe
eigene Beobachtungen aus dem Unterricht	90,9	0	9,1
Testverfahren (z.B. Stolperwörtertest, Hamburger Schreibprobe)	72,7	9,1	18,2
Schreibproben (Texte der Kinder)	90,9	9,1	0
eigens für diesen Zweck erstellte Arbeitsblätter	18,2	27,3	45,5
Einschätzungen von Kollegen	72,7	0	18,2

Tab. 6: Feststellung des Förderbedarfs eines Kindes

Die folgende Abbildung 37 zeigt, welche Maßnahmen zur Lese- und Schreibförderung an den Schulen umgesetzt werden: Ein Großteil der Lehrkräfte (91 %) an den Kontrollschulen geben an, dass sie Lern- bzw. Förderpläne schreiben und diese auch umsetzen. Mehrheitlich wird die Deutschförderung nicht durch die reguläre Klassenlehrkraft, sondern durch zusätzliche Fachlehrkräfte durchgeführt (54,5 %). Die Lehrkräfte geben weiterhin mehrheitlich an, dass sie bei Bedarf Kontakt zu externen Fachkräften, wie z.B. Logopäden und Logopädinnen oder speziellen LRS-Förderkräften, aufnehmen. Auffällig ist, dass knapp 82 % der befragten Lehrkräfte angeben, dass eine Vernetzung aller an der Förderung beteiligten Personen stattfindet.

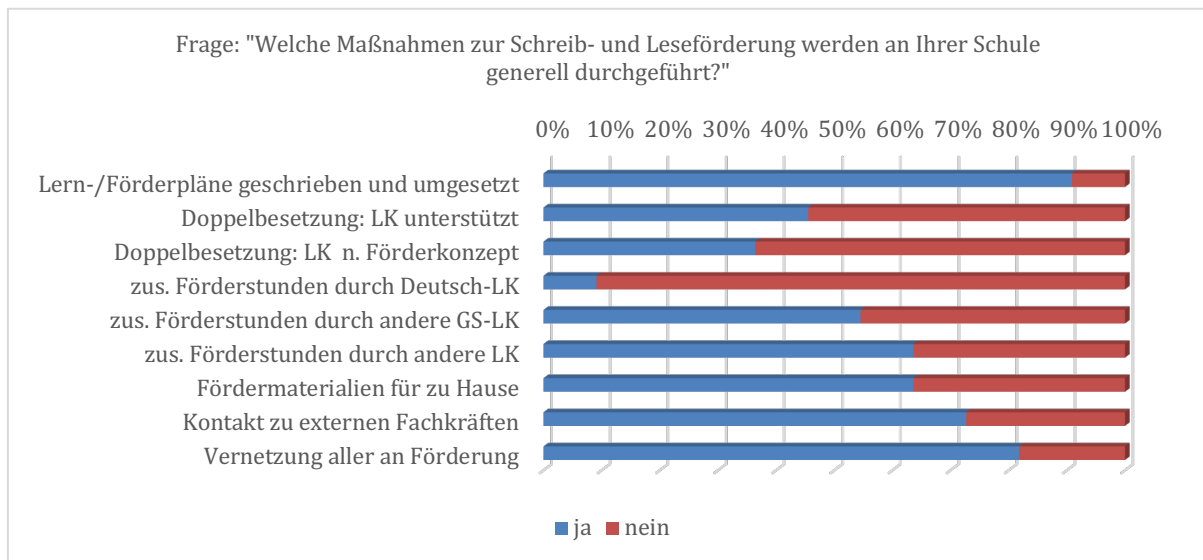


Abb. 37: Maßnahmen zur Schreib- und Leseförderung (N = 11)

Auf die Frage hin, in welcher Form die Förderung der entsprechenden Schülerinnen und Schüler umgesetzt wird, gibt die Mehrheit an, dass die Förderung in den regulären Deutschunterricht integriert wird (81,8 %). Nur 36,4 % der Lehrkräfte geben an, dass die Förderung an ihrer Schule in Kleingruppen durchgeführt wird; in Einzelsituationen wird an keiner Schule Förderunterricht durchgeführt (siehe Abbildungen 38).



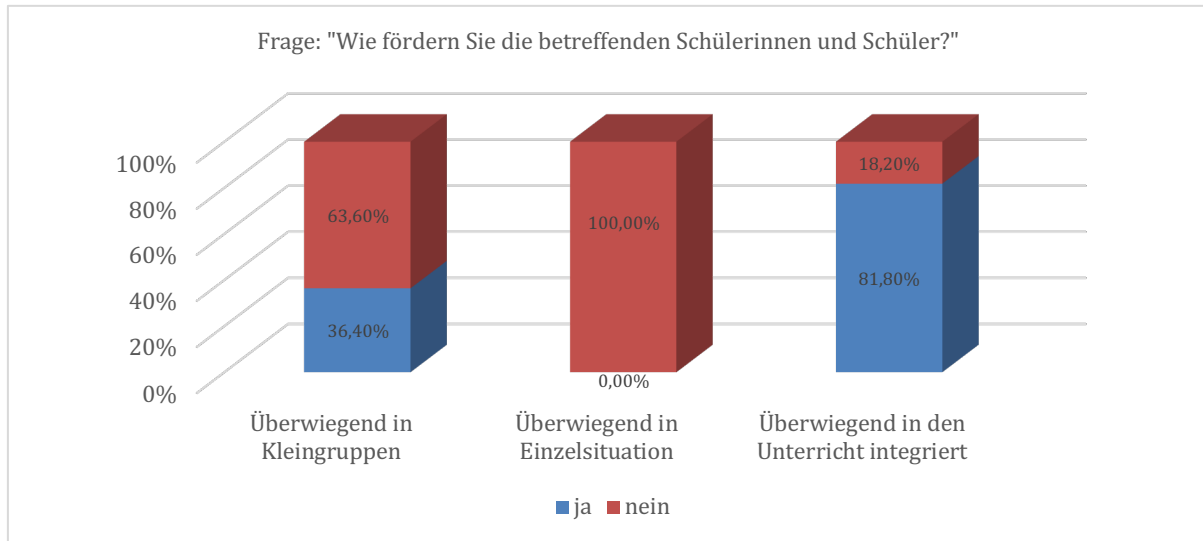


Abb. 38: Sozialformen des Lms-Trainings (N = 11)

Inhaltlich fokussiert die Förderung der Schülerinnen und Schüler mehrheitlich (81,8 %) auf Teilbereiche im Lesen, Schreiben und im Textverständnis, jedoch geben auch 63,6 % der Lehrkräfte an, dass sie die schriftsprachlichen Fähigkeiten eher breit fördern (siehe Abbildung 39).

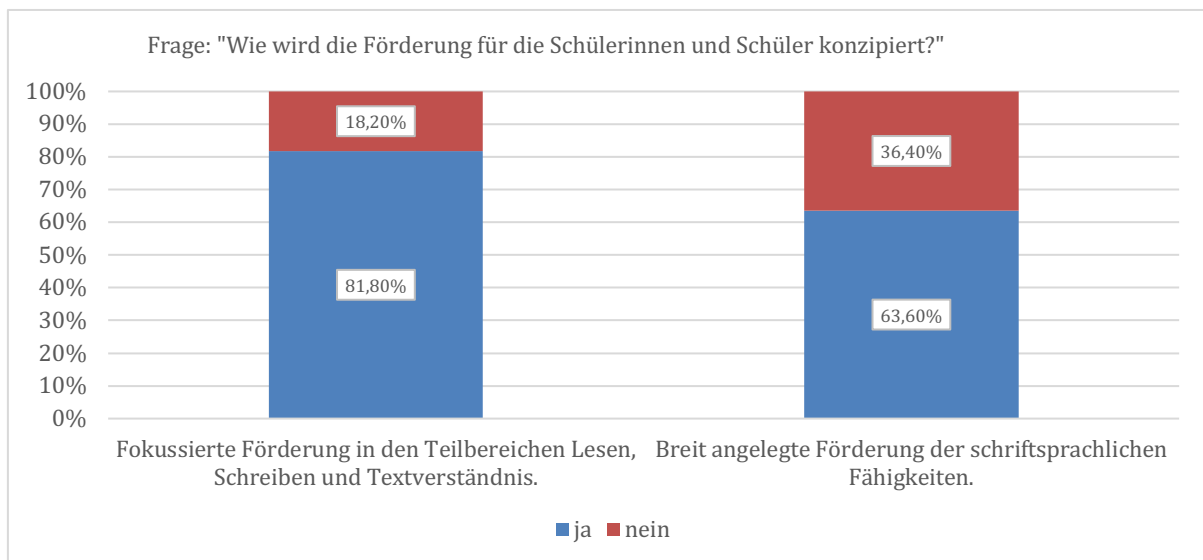


Abb. 39: Fokussierte Förderung vs. breit angelegte Förderung (N = 11)

Die folgenden Abbildungen 40 und 41 zeigen, dass mehr als die Hälfte der Kontrollschulen (55 %) bestimmte, der Schule zur Verfügung stehende Stunden (z.B. Präventionsstunden), für die gezielte Förderung der schriftsprachlichen Kompetenzen nutzen. Dabei geben diese

Lehrkräfte weiterhin an, dass die Dauer der Förderung zwischen einer Stunde pro Woche und 5 Stunden pro Woche schwankt (s. Abb. 40). Es wird also deutlich, dass die Kontrollschulen insgesamt keine auffällig intensive Lese- und Schreibförderung für die Schülerinnen und Schüler und auch kein spezielles Förderprogramm oder Training anbieten.

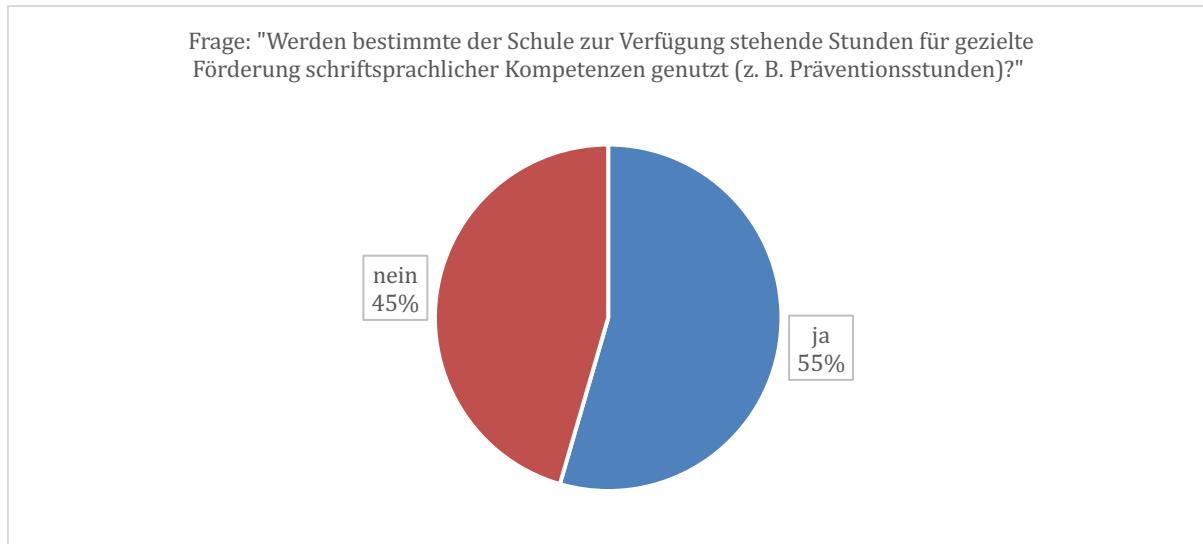


Abb. 40: Gezielte Förderung an Schulen (N = 11)

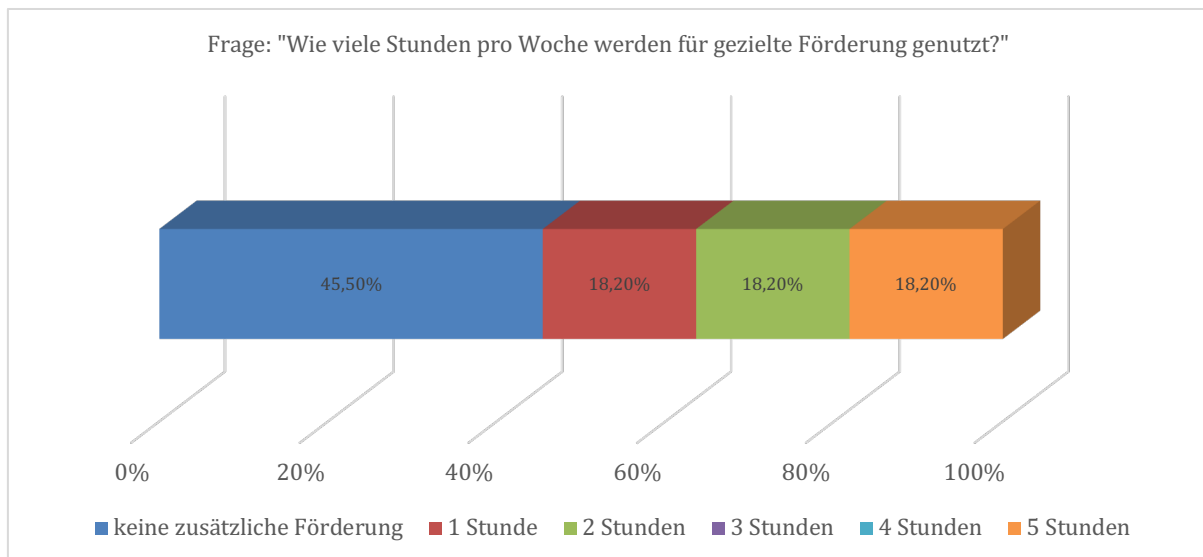


Abb. 41: Übersicht Höhe zusätzlicher Förderung

**Das Konstrukt (3) umfasst Fragen, die die Zuweisung der Kinder in die Fördergruppen bzw. die zugrundeliegende Diagnostik, umfassen. Die Abbildungen 42 und 43 geben die Ergebnisse**

der Befragung auf einer fünfstufigen Likert-Skala von 1 (nicht sinnvoll) bis 5 (sehr sinnvoll) wieder. Die Lehrkräfte der Experimentalgruppe halten die Bestimmung des Förderbedarfs aufgrund der Lms-Diagnostik über alle Bausteine hinweg für sehr zuverlässig (4,13 – 4,5). Auf die Frage hin, inwieweit die Lehrkräfte es für sinnvoll erachten, nach jedem Baustein erneut eine Diagnostik durchzuführen (s. Abb. 43), antworteten die Lehrkräfte ebenso mit im Mittel mit hoher Zustimmung (4,75 – 5,0).

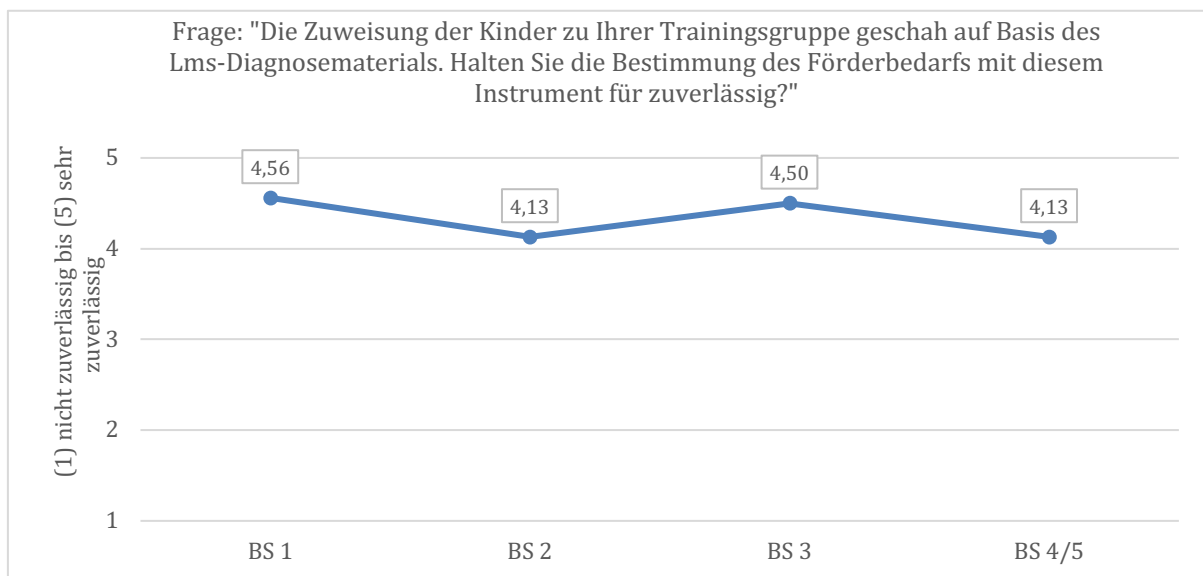


Abb. 42: Einschätzung von Zuweisung der Kinder zu Trainingsgruppe auf Basis Lms-Diagnosematerial

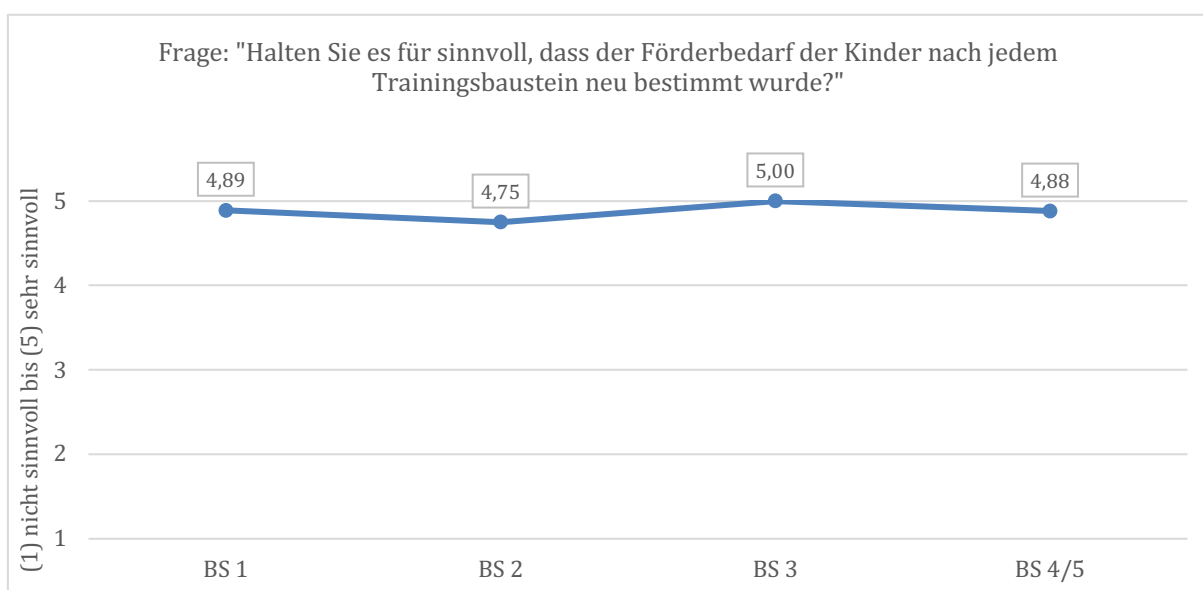


Abb. 43: Einschätzung von Neubestimmung des Förderbedarfs von Kindern nach jedem Lms-Trainingsbaustein

Die Lehrkräfte gaben des Weiteren mehrheitlich an, dass sie Einblicke in die Ergebnisse des Diagnoseverfahrens erhalten haben (in dem Fall, dass eine andere Lehrkraft die Diagnostik durchgeführt hat) und diese auch für die Gestaltung der Förderung genutzt haben (s. Abb. 44 – 46). In Abb. 44 wird jedoch auch deutlich, dass es einige Lehrkräfte gibt (35 – 37,5%), die die Ergebnisse der Diagnostik nicht für ihre Förderung genutzt haben. Abb. 44 zeigt, dass jeweils in den Bausteinen 1, 2 und 4/5 ca. 80 – 90 % der Lehrkräfte selbst die Diagnostik und die Förderung durchgeführt haben, während in Baustein 3 eine andere Verteilung sichtbar ist: Hier haben ca. 43 % der Lehrkräfte die Diagnostik durchgeführt und etwa genauso viele haben diese nicht durchgeführt.

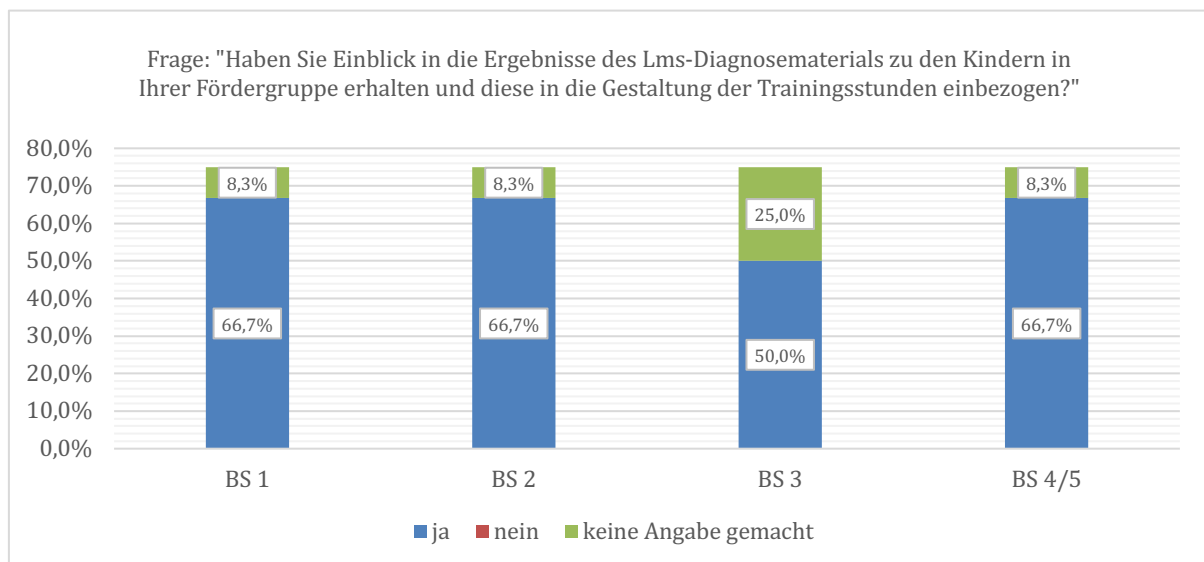


Abb. 44: Einbezug von Ergebnissen der Lms-Diagnostik in Gestaltung der Trainingsstunden (1) (N = 12)

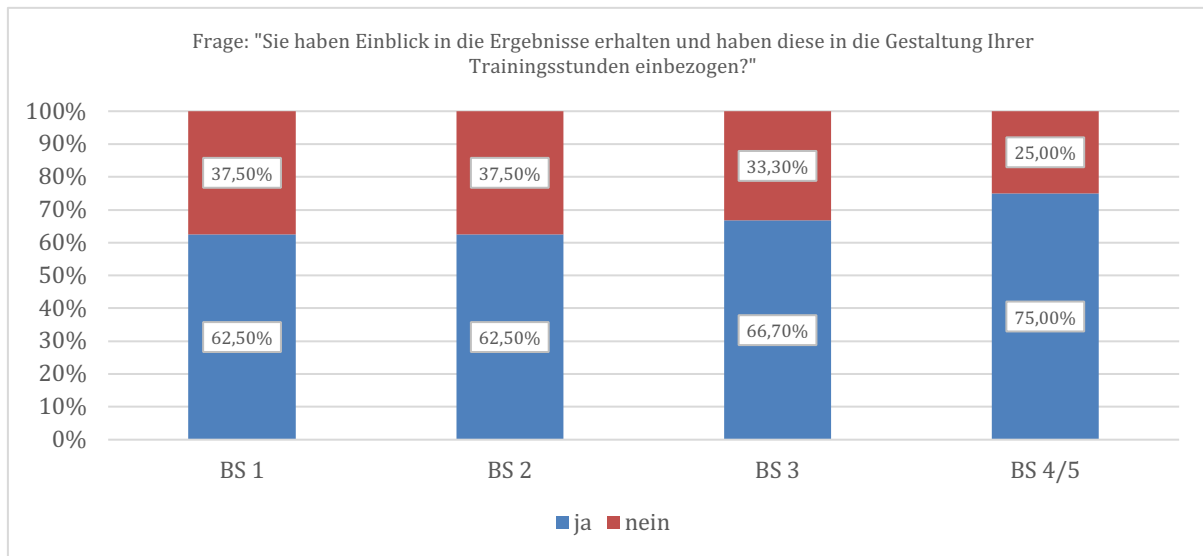


Abb. 45: Einbezug von Ergebnissen der Lms-Diagnostik in Gestaltung der Trainingsstunden (2) (N = 8)

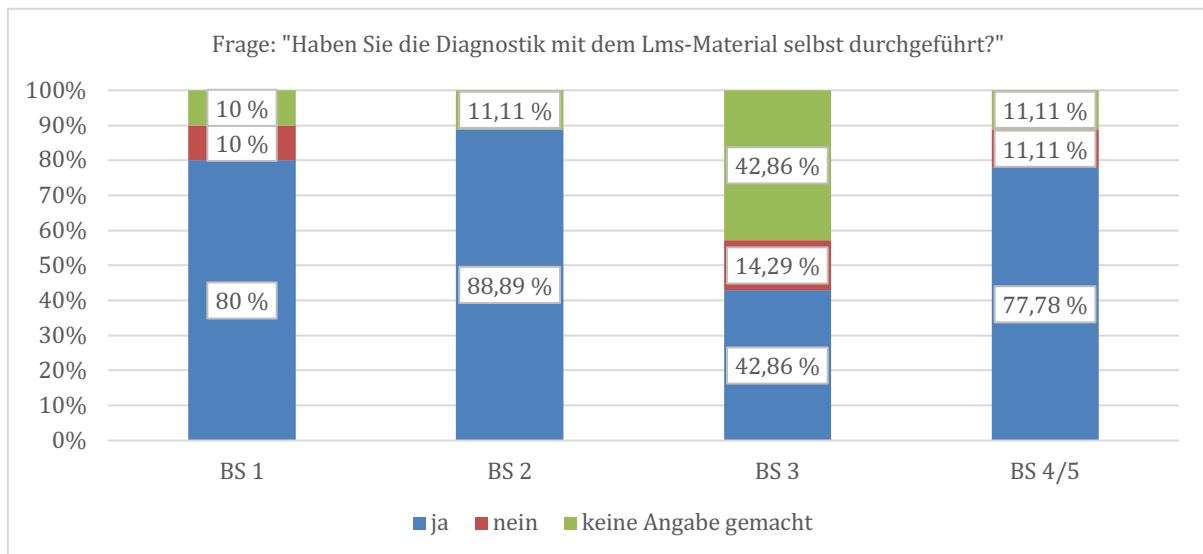


Abb. 46: Durchführung der Diagnostik durch Lehrkraft (BS 1: N = 10, BS 2: N = 9, BS 3: N = 7, BS 4: N = 9)

**Das Konstrukt (4) umfasst Items, die die Nutzung des Trainingsprogramms und die Modalitäten der Durchführung erfassen sollen.** Abbildung 47 zeigt, dass nur ein geringer Anteil der Schulen ein Förderband in ihrer Studentafel umgesetzt hat (11,1%). Mehrheitlich findet die Förderung mit dem Lms-Training an den Schulen integrativ im regulären Deutschunterricht statt (44,4% – 66,7%). An einigen Schulen wurde die Förderung additiv, also zusätzlich zum regulären Deutschunterricht durchgeführt.

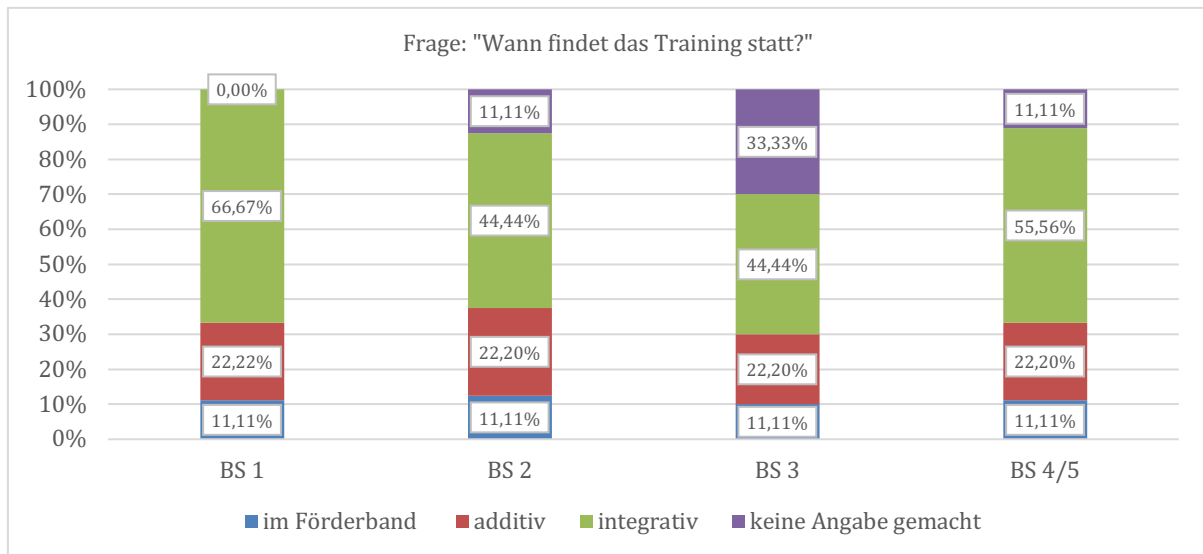


Abb. 47: Zeitlicher Rahmen des LMS-Trainings

Die Abbildungen 48 und 49 zeigen, dass die Lehrkräfte alle Elemente des Trainings genutzt haben, jedoch wurde teilweise auch darüber hinaus weiteres Fördermaterial genutzt.

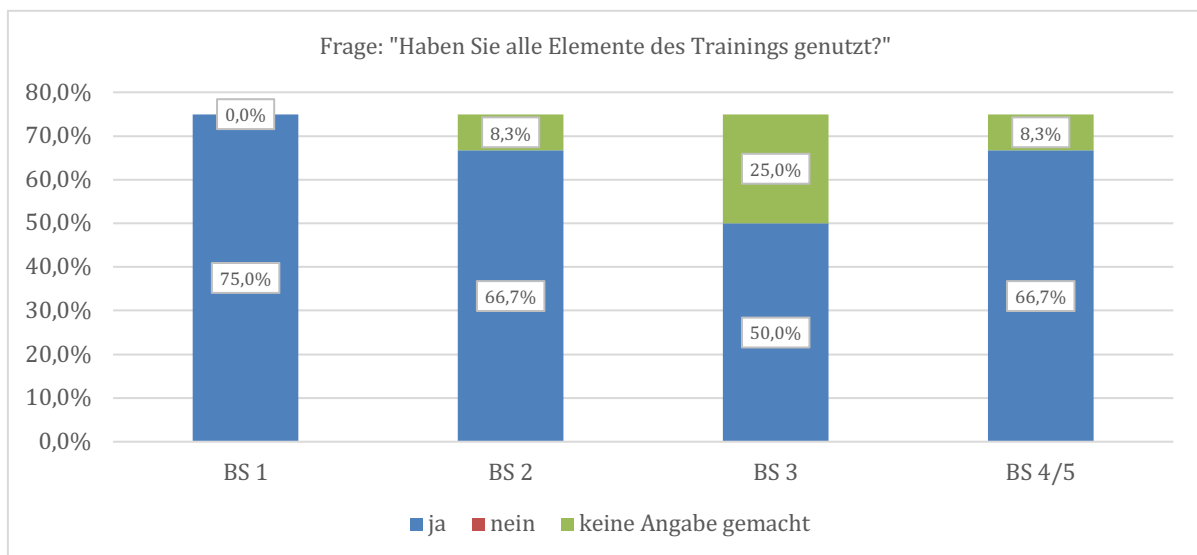


Abb. 48: Nutzung aller Elemente des Lms-Trainings (N = 12)

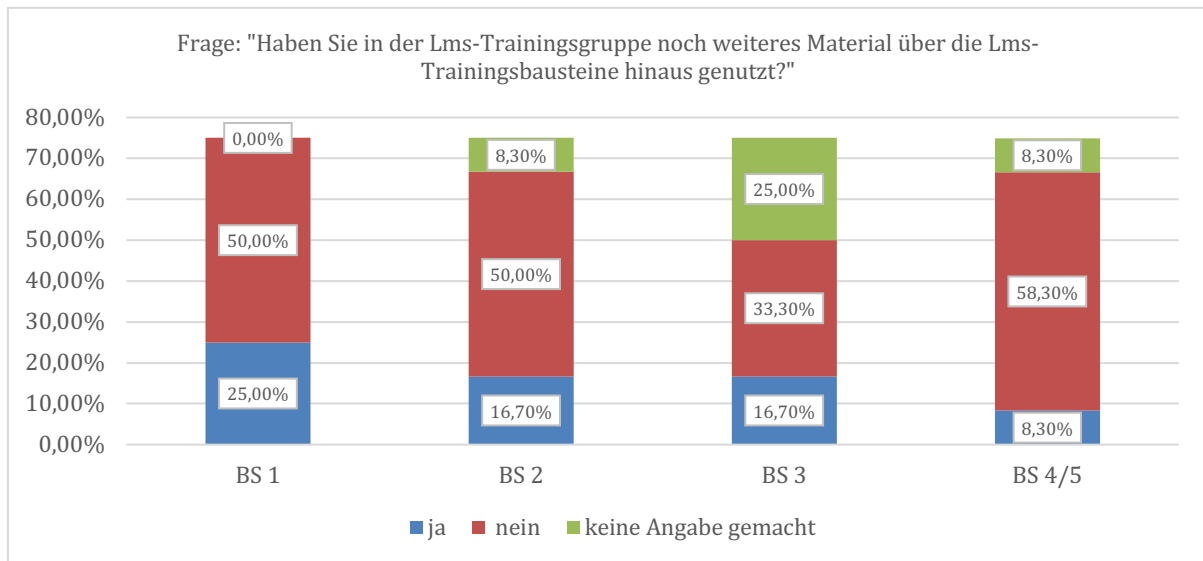


Abb. 49: Nutzung von weiterem Material über Lms-Trainingsbausteine hinaus (N = 12)

Die Lehrkräfte geben weiterhin an, dass sie die Formulierungshilfen aus den Lehrermanualen zur Umsetzung des Trainings im Mittel recht genau übernommen haben. Hier finden sich über die verschiedenen Bausteine hinweg auf einer fünfstufigen Likert-Skala hohe Punktzahlen (4,0 – 4,67) (siehe Abbildung 50).

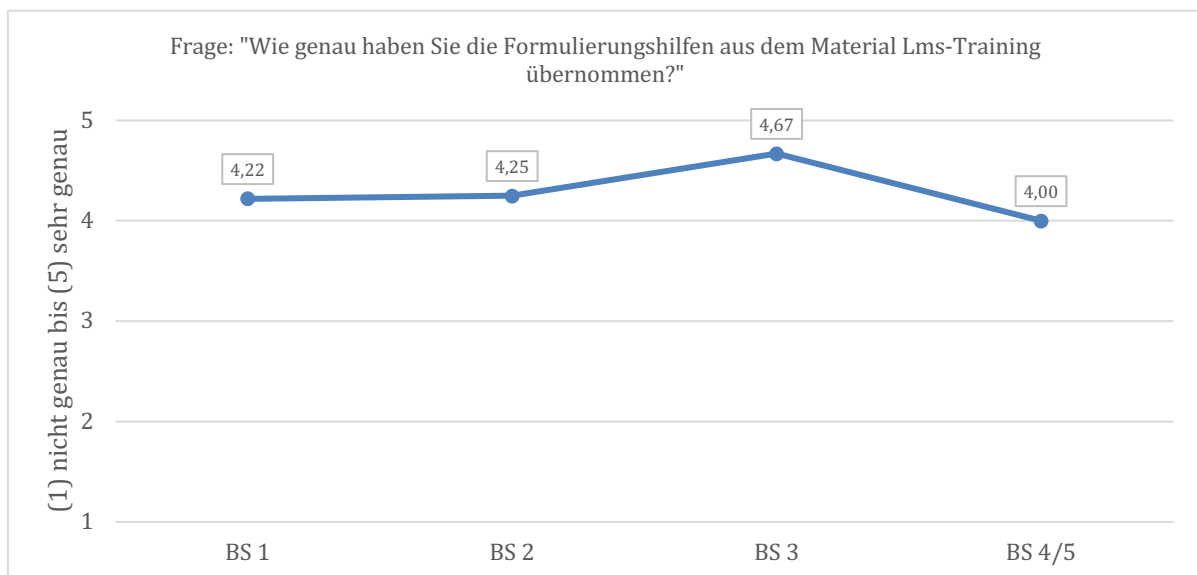


Abb. 50: Übernahme der Formulierungshilfen aus Lms-Trainingsprogramm

Abbildung 51 zeigt in Bezug auf die Frage, ob die Lehrkräfte das Lms-Trainingsprogramm auch außerhalb der Trainingsgruppen zur Förderung eingesetzt haben, ein heterogenes Bild. In Baustein 1 geben noch die Hälfte der Lehrkräfte an, dass sie das Material auch in anderen

Kontexten einsetzen. Ab Baustein 2 jedoch verneinen die Lehrkräfte diese Frage größtenteils (50% – 66,7 %).

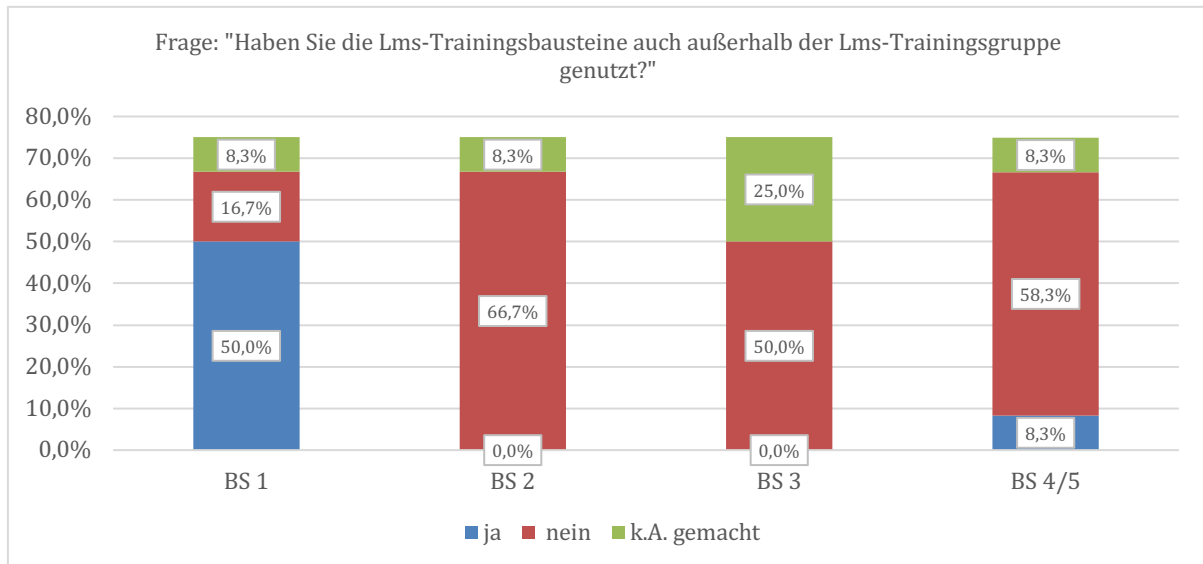


Abb. 51: Nutzung von Lms-Trainingsbausteinen außerhalb der Trainingsgruppe (N = 12)

Im Fragebogen sind zu diesem Konstrukt auch drei offene Fragen zu finden, die sich auf positive bzw. negative Elemente des Trainings sowie auf die persönliche Einschätzung der Lehrkräfte beziehen und was ihnen am Material besonders gut oder nicht gut gefallen hat. Die folgenden Tabellen geben die Antworten der Lehrkräfte bezogen auf die verschiedenen Bausteine wieder. Insgesamt wird deutlich, dass der strukturierte Aufbau und die Zielsetzung der Automatisierung positiv durch die Lehrkräfte beurteilt wurde. Die Komplexität des Trainings und der damit verbundene Zeitaufwand wurden negativ bewertet. Die einzelnen Antworten finden sich in den nachfolgenden Tabellen 7 – 11.

Positive Elemente	Negative Elemente
Spielvarianten: Sticker, Karteikärtchen, Monsterkarten, Buchstabenrennen, Dirigentenspiel, Trainer-Sportler	Buchstabenkarten legen ist zeitintensiv sehr komplexer Aufbau – Kinder haben zu viel zu bedenken (hohe Selbstkompetenz erforderlich); Konzentration + Disziplin gefragt
Lob und Bestätigung => Motivation	4x Training/Woche zu umfangreich und einseitig bei integrativer Förderung; Ankommenszeit im Klassenverband für die Kinder wichtig



Herausnehmen aus Klassenverband	Trainingsgruppe zu groß für Differenzierung,
Tandem-Arbeit	manchmal das "Rausnehmen" aus dem regulären Unterricht => Ängste, Inhalte aus dem regulären Deutschunterricht zu verpassen
Automatisierte, wiederkehrende Abläufe => Sicherheit, Ansprache und Übung in der Kleingruppe	Tandem-Partner muss nicht nur dem Leistungsniveau entsprechen, sondern auch vom Typ her passen
Transparenz über Lernfortschritte => selbstständiges Ankreuzen	wenig Abwechslung (Bemerkung der SuS: „Schon wieder?!“) => Monotonie
	Enttäuschung des Kindes beim Nichterreichen des nächsten Levels, auch wiederholtes Nichterreichen => Tandem-Partner hängt in der Warteschleife
	zu viele Inhalte bis zum Level-Check, lange Wartezeiten bei der Einführung des Level-Checks

Tab. 7: Positive und Negative Elemente des Trainings auf Motivation der Kinder

Positive Aspekte	Negative Aspekte
Wiederholung/Automatisierung	Zeitaufwand der Dokumentation
Lernfortschritt und Lernerfolg der Kinder + Transparenz	Integrative Form des Trainings => Kinder aus Klasse entnehmen, Fehlen in regulärem Unterricht
strukturierter Aufbau	
Arbeit im Tandem und Team	

Tab. 8: Positive und negative Aspekte des Trainings in Baustein 1

Positive Aspekte	Negative Aspekte
Wiederholung/Automatisierung	Zeitaufwand der Dokumentation
Lernfortschritt und Lernerfolg der Kinder + Transparenz	Integrative Form des Trainings => Kinder aus Klasse entnehmen, Fehlen in regulärem Unterricht
Optimierung der Selbstkompetenz der SuS	Größe der Trainingsgruppe
strukturierter Aufbau: ausführliche Erläuterungen im Begleitmaterial	Zusätzliche Lehrerwochenstunden für Training als Band
Arbeit im Tandem und Team	

Ansprechende Materialien	
--------------------------	--

Tab. 9: Positive und negative Aspekte des Trainings in Baustein 2

Positive Aspekte	Negative Aspekte
Wiederholung/Automatisierung	Zeitaufwand der Dokumentation => Belastungsgrenze
Lernfortschritt und Lernerfolg der Kinder + Transparenz	Komplexität (schwer für weniger selbständige Kinder)
Optimierung der Selbstkompetenz der SuS, Lerneffekte	Homogene Teams
strukturierter Aufbau: ausführliche Erläuterungen im Begleitmaterial, guter und sinnvoller Aufbau	Zusätzliche Lehrerwochenstunden für Training als Band
Arbeit im Tandem und Team	
Ansprechende Materialien,	
Erkenntnisse über Leistungsstand und -Entwicklung	

Tab. 10: Positive und negative Aspekte des Trainings in Baustein 3

Positive Aspekte	Negative Aspekte
Wiederholung/Automatisierung	Zeitaufwand der Dokumentation
Lernfortschritt und Lernerfolg der Kinder + Transparenz	Zeitaufwand des Trainings
Optimierung der Selbstkompetenz der SuS	
strukturierter Aufbau: ausführliche Erläuterungen im Begleitmaterial	
Motivation durch Teamarbeit und Levelcheck	
Ansprechende Materialien	
Regelmäßige und kurze Trainingszeit	
Abwechslung zwischen Lesen und Schreiben	
Sicherheit durch Konstanz (gleiche Tandempartner)	

Tab. 11: Positive und negative Aspekte des Trainings in Baustein 4

**Das Konstrukt (5) erfasst die Bewertung des Lms-Trainings durch die Lehrkräfte**, dargestellt in Abbildung 52. Die insgesamt 25 Items wurden zwecks Auswertung in vier Kategorien eingeordnet: *Allgemeine Aussagen* zu Lms-Training, wie z.B. der Wunsch, dieses auch für weitere Klassenstufen erhalten zu können (blau); Aussagen zur *Nützlichkeit* des Materials, wie z.B. eine sinnvolle Fokussierung des Automatisierungsaspekts im Material (rot); Aussagen zur *Angemessenheit*, wie z.B. zum zeitlichen Umfang der Förderung (grün); Aussagen, die *negativ* formuliert wurden, z.B. hinsichtlich einer möglichen Überforderung der Schülerinnen und Schüler (lila). Die Aussagen wurden von den Lehrkräften auf einer fünfstufigen Likert-Skala von „trifft überhaupt nicht zu“ (1) bis hin zu „trifft voll und ganz zu“ (5) bewertet.

Abbildung 52 zeigt, dass die Lehrkräfte Aussagen der Kategorie *Nützlichkeit* am höchsten bewertet haben: über die Bausteine hinweg im Mittel von 4,26 bis hin zu 4,63 Punkten. Auch die Aussagen zur Kategorie *Allgemeine Bewertung* erhalten eine hohe Zustimmung. Hier ist auffällig, dass die Bewertungen im dritten Baustein nicht ganz so viel Zustimmung erhalten haben, aber immer noch gut ausgefallen sind (3,89). Eher im durchschnittlichen Bereich befinden sich die Bewertungen zur Kategorie *Angemessenheit* (3,44 – 4,08). Hier wird deutlich, dass das Material als umfangreich und voraussetzungsreich erlebt wurde. Die *negativ formulierten Aussagen* erhalten durchgehend wenig Zustimmung (1,65 - ,07). Auffällig ist, dass die Bewertungen zum Ende des Trainings hin besser werden, also eine klare positive Tendenz erkennbar ist.

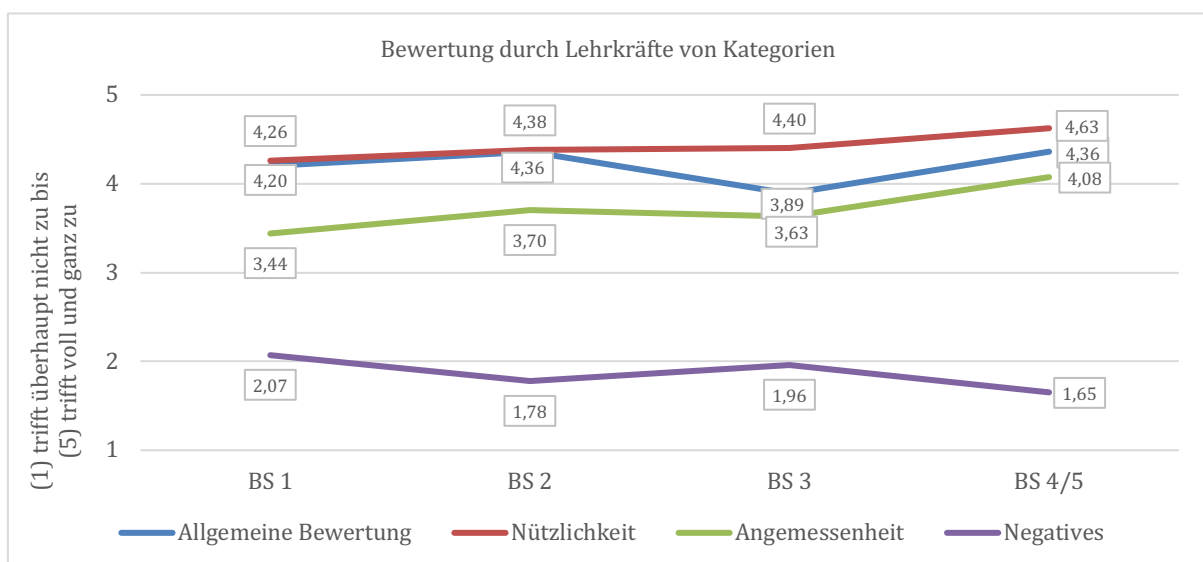


Abb. 52: Bewertung verschiedener Kategorien durch Lehrkräfte über zeitlichen Verlauf der Trainingsbausteine

## 5. Diskussion

Die Studie zum Trainingsprogramm umfasst die Intervention auf der einen (Teil A) und die Erfassung der Konzepttreue auf der anderen Seite (Teil B). Mit der Studie sollten folgende Forschungsfragen beantwortet werden:

### Teil A

Wie wirkt sich das Training auf die Lese- und Schreibkompetenzen der Schülerinnen und Schüler aus?

### Teil B

1. Ist die Konzepttreue (Förderqualität) gewährleistet?
  - a) Werden die Kinder, die in der Diagnostik auffällig waren, in die Fördergruppen aufgenommen?
  - b) Findet die Förderung regelmäßig statt?
  - c) Nehmen die Kinder regelmäßig an der Förderung teil?
  - d) Inwieweit wird die Förderung konzepttreu umgesetzt?
2. Inwieweit wird die Fortbildung als Teil der Implementation von den Lehrkräften akzeptiert (Evaluation der Fortbildung)?
3. Inwieweit wird das Training von den Lehrkräften akzeptiert?

### **Teil A: Kompetenzbereich Lesen**

Teil A der Studie untersucht die Entwicklung der Lese- und Schreibkompetenzen der Schülerinnen und Schüler, die das Lms-Training absolviert haben. Da ein Vergleich der gesamten Experimentalgruppe mit der Kontrollgruppe nicht geeignet ist, um die Wirksamkeit zu überprüfen (siehe Anhang Abschnitt I und IV), wurden Differentialgruppen mit den Schülerinnen und Schülern gebildet, die an Lms-Trainingsbausteinen teilgenommen haben (siehe auch Abbildung 53).

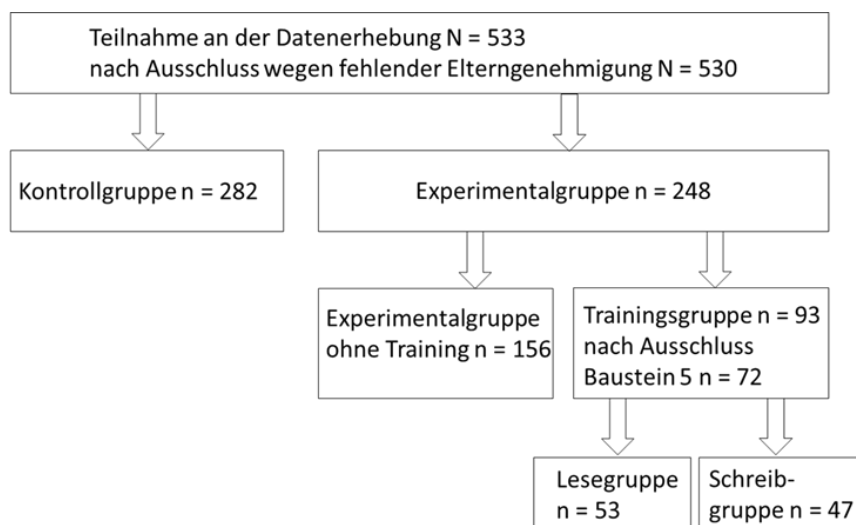


Abb. 53: Auswahl der Differentialgruppen

Für die Überprüfung der Wirksamkeit des Lesetrainings, welches in den Bausteinen 1, 2 und 5 durchgeführt wurde, wurden Schülerinnen und Schüler, die an mindestens einem Baustein Lesen teilnahmen, mit den Schülerinnen und Schülern der Experimentalgruppe verglichen, die kein Lms-Training erhielten, sowie mit den Schülerinnen und Schülern der Kontrollgruppe. Zusammenfassend lässt sich sagen, dass Schülerinnen und Schüler, die an mindestens einem Lms-Trainingsbaustein zum Lesen teilgenommen haben, einen stärkeren Zuwachs der Wortverständnisfähigkeit von Messzeitpunkt 1 zu Messzeitpunkt 2 im Vergleich zur Kontrollgruppe ( $F(1, 144) = 5,061$ ;  $p = ,026$ ; partielles  $\text{Eta}^2 = ,034$ ) aufweisen. D.h., die Schülerinnen und Schüler, die am Lesetraining teilnahmen, zeigen innerhalb eines Schulhalbjahres einen größeren Leistungszuwachs als die Kontrollgruppe. Vergleicht man die Leistungszuwächse der Kinder, die an mindestens einem Baustein zur Leseförderung teilgenommen haben, mit denen der Kinder des gleichen Klassenverbundes – also Kinder der Experimentalschulen, die jedoch nicht in der Fördergruppe waren, erhält man sogar noch deutlichere Effekte: Hier zeigt sich ein signifikant größerer Leistungszuwachs mit einer höheren Effektstärke ( $F(1, 94) = 11,149$ ;  $p < ,001$ ); partielles  $\text{Eta}^2 = ,106$ ). Hier ist zu vermuten, dass die Kinder der Kontrollgruppe – also Kinder anderer Schulen, die nicht am Programm „Lesen macht stark“ teilnahmen – eine andere Förderung erhielten, die sich positiv auf die Leistungsentwicklung auswirkte, während Kinder der Experimentalklassen keine zusätzliche Förderung erhielten. Um einen Einblick in die Förderung an den Kontrollschulen zu erlangen, wurde auch hier ein Fragebogen eingesetzt. Auf die Frage hin, in welcher Form die Förderung der entsprechenden Schülerinnen und Schüler umgesetzt wird, gibt die Mehrheit an, dass die Förderung in den regulären Deutschunterricht integriert wird (81,8 %). Es wird also deutlich, dass die Kontrollschulen keine zusätzliche intensive (trainingsförmige) Lese- und Schreibförderung für die Schülerinnen und Schüler anbieten.

Hingegen war der Zuwachs der Satzverständnisfähigkeit von Messzeitpunkt 1 zu Messzeitpunkt 2 für die drei Gruppen ähnlich stark. Ein Grund für den ausbleibenden Effekt bei der Satzverständnisfähigkeit gegenüber der Wortverständnisfähigkeit könnte darin begründet sein, dass Kinder der ersten Jahrgangsstufe sich im Schriftspracherwerb noch auf der Stufe der Laut-Buchstaben-Zuordnung befinden bzw. erste Wörter erlesen können. Das Lesen und Verstehen ganzer Sätze könnte für Kinder der Experimental- sowie der Kontrollgruppe im ersten Schuljahr noch schwieriger sein als das Wortverständnis.

### **Teil A: Kompetenzbereich Schreiben**

Da die Studie pandemiebedingt abgebrochen werden musste, liegt zum Kompetenzbereich „Schreiben“ lediglich ein Querschnitt vor. Daher ist keine Berechnung des Leistungszuwachses möglich. Ein Vergleich der Schülerinnen und Schüler, die an mindestens einem Baustein des Schreibtrainings (Baustein 3 und/oder 4) teilgenommen haben, mit der Experimentalgruppe ohne Training sowie der Kontrollgruppe zeigt, dass alle drei Gruppen ähnlich viele orthographische Fehler machen. Hingegen zeigt sich ein sehr bedeutsamer Unterschied bei den lautgetreuen Fehlern (N-Fehler): Die Schreibgruppe machte demnach signifikant mehr N-Fehler (lautierend korrekt schreiben) als die Kontrollgruppe ( $t(402) = -10,481$ ;  $p < ,001$ ; Cohens  $d = 1,70$ ) sowie mehr Fehler als die Experimentalgruppe ohne Training ( $t(402) = 9,586$ ;  $p < ,001$ ; Cohens  $d = 1,66$ ). Die Effektstärke gibt einen sehr starken Effekt an. Bezüglich der Fehler in der Groß- und Kleinschreibung (G-Fehler) zeigt sich, dass die Schreibgruppe ebenfalls signifikant mehr Fehler macht als die Kontrollgruppe ( $t(402) = -2,482$ ;  $p = ,013$ ) sowie die Experimentalgruppe ohne Training ( $t(402) = 2,15$ ;  $p = ,037$ ). Der Effekt ist als mittlerer Effekt einzuordnen (Cohens  $d = 0,40$  bzw. Cohens  $d = 0,43$ ). Die Experimentalgruppe ohne Training erzielt ähnlich viele N- bzw. G-Fehler wie die Kontrollgruppe. Eine Förderung der Rechtschreibkompetenz im Sinne orthographischer Schreibungen ist nicht explizit Teil des Trainings. Im Schreibtraining wurde zu diesem frühen Zeitpunkt der Schreibentwicklung auf die Sicherung und Automatisierung der Phonem-Graphem-Korrespondenzen auf Buchstaben- und Silbenebene fokussiert, trotzdem weist die Schreibgruppe eine hohe Anzahl von N-Fehlern, also den lautorientierten Schreibungen, auf. Es sollte dabei jedoch beachtet werden, dass von den vorgesehenen Fördereinheiten zum Bereich Schreiben nur 78 % (Baustein 3) bzw. 58 % (Baustein 4) der Fördereinheiten angeboten wurden. Außerdem ist hier nur eine Darstellung des Querschnitts möglich, so dass nicht ersichtlich wird, mit welcher Ausgangsleistung die Kinder gestartet sind und welche Entwicklung sie genommen haben.

Teil B der Studie geht der Frage nach, wie das Lms-Training in den Schulen umgesetzt und wie insbesondere die Förderung realisiert wurde.

## **Teil B: Werden die Kinder, die in der Diagnostik auffällig waren, in die Fördergruppen aufgenommen?**

Der Grundgedanke des Programms beinhaltet, dass die Zuteilung in die Fördergruppen anhand des „Lesen macht stark“-Diagnosematerials erfolgen sollte. Insgesamt lässt sich jedoch feststellen, dass viele Daten zur Dokumentation der Lms-Diagnostik fehlen, weil die Lehrkräfte ihr Vorgehen nicht durchgängig dokumentiert haben. Die Auswertung der vorhandenen Daten zeigt jedoch, dass die Lms-Diagnostik nicht konsequent von den Lehrkräften als Basis für die Zuteilung der Kinder in die Fördergruppen genutzt wurde. Hinsichtlich der Zuteilung der Kinder in die Fördergruppen zeigt sich über alle Meilensteine (Messzeitpunkte mit dem Lms-Diagnosematerial) hinweg, dass in 81 % der Fälle unauffällige Leistungen in den Aufgaben der Meilensteine beobachtet wurden. Trotzdem wurden in 6 % der Fälle Kinder, deren Leistungen nicht unter dem Cut-off-Wert lagen, einer Trainingsgruppe zugeordnet. Auf der anderen Seite wurde in 19 % der Fälle ein Förderbedarf festgestellt, der jedoch nur in 10 % der Fälle zu einer zusätzlichen Förderung der Kinder führte. Insgesamt wurden also in 48 % der Fälle, in denen eine unterdurchschnittliche Leistung beobachtet wurde, darauf verzichtet, diese Kinder den Fördergruppen zuzuteilen. Viele Kinder mit festgestelltem Förderbedarf erhielten demnach keine zusätzliche Förderung. Dies bestätigt auch der Lehrkräftefragebogen, denn hier geben knapp 40 % an, dass sie die Ergebnisse der Lms-Diagnostik nicht für ihre Förderung bzw. für die Zusammensetzung der Fördergruppe genutzt haben. Trotzdem halten die Lehrkräfte die Bestimmung des Förderbedarfs aufgrund der Lms-Diagnostik über alle Bausteine hinweg für sehr zuverlässig (4,13 – 4,5). Auf die Frage hin, inwieweit die Lehrkräfte es für sinnvoll erachten, nach jedem Baustein erneut eine Lms-Diagnostik durchzuführen (s. Abb. 41), antworteten diese ebenso mit im Mittel hoher Zustimmung (4,75 – 5,0). Hier liegt also eine Diskrepanz zwischen der Bewertung der Lms-Diagnostik und der tatsächlichen Nutzung dieser zur Einteilung in die Fördergruppen vor.

Um einen vertieften Einblick in die Zuteilung der Kinder zu den Fördergruppen zu erhalten, wurde im Nachgang an die Studie ein leitfadengestütztes Interview durchgeführt. Die Nachbefragung (s. Anlage) der Lehrkräfte ergab, dass die schwächsten Kinder häufig nicht in die Fördergruppen aufgenommen wurden. Folgende Frage wurde seitens einer Mitarbeiterin des IQSH an die Lehrkräfte (N = 5) gestellt:

„Wie haben Sie als Lehrkraft entschieden, ob ein Kind in die Trainingsgruppe aufgenommen wurde oder nicht?“

Die Lehrkräfte antworten auf diese Frage, dass sie zum einen die Fördergruppen aufgrund der Lms-Diagnostik gebildet haben, aber zum anderen auch andere Aspekte bei der Einteilung berücksichtigt haben. So nennen sie mehrfach, dass Kinder, die nach Einschätzung der Lehrkraft die Eingangsphase in drei Jahren (statt zwei Jahren) durchlaufen werden, nicht mit

in die Trainingsgruppe aufgenommen wurden zugunsten der Kinder, die weniger schwache Leistungen zeigten.

Zum Teil wurden pragmatische Entscheidungen getroffen, wenn Kinder aufgrund von Verhaltensauffälligkeiten nicht tragbar für die Gruppe waren oder wenn die Chance auf eine regelmäßige Teilnahme nicht gesehen wurde. Die Lehrkräfte geben weiterhin an, dass sie aufgrund ihrer Einschätzung und Erfahrung die Kinder in ihren Leistungen beurteilt und die Fördergruppen gebildet haben.

### **Teil B: Findet die Förderung regelmäßig statt?**

Vorgesehen ist pro Förderbaustein die Förderung an vier Tagen in der Woche für jeweils etwa 20 Minuten. Es wurde deutlich, dass in den ersten drei Förderbausteinen zumindest 70-80 % der vorgesehenen Fördereinheiten angeboten wurden. Ab dem vierten Baustein sinkt jedoch das Förderangebot. In den letzten beiden Förderbausteinen wurden im Durchschnitt nur noch 60 % der vorgesehenen Fördereinheiten durchgeführt. An einer Schule liegt der Anteil der durchgeführten Förderung sogar nur noch bei rund 20 %. Im Baustein 5 konnte keine regelmäßige Teilnahme der Kinder mehr stattfinden, da die Pandemie zu diesem Zeitpunkt ihren Anfang nahm.

Es ist also festzustellen, dass die Anzahl der durchgeführten Fördereinheiten je Baustein im Laufe der Zeit abnimmt. Dies ist ein Indiz dafür, dass das Trainingsprogramm anspruchsvoll in der Umsetzung ist, da es eine regelmäßige Vorbereitung und Durchführung zusätzlich zum weiterhin stattfindenden Regelunterricht erforderlich macht. Hier ist zu überlegen, inwieweit bei der Implementation systematischer Trainingsprogramme in den Schulalltag berücksichtigt werden sollte, dass diese Form der Förderung nicht im regulären Deutschunterricht durchgeführt, sondern in Form eines Förderbandes realisiert werden sollte.

Zudem zeigen sich hinsichtlich des Angebots der Förderung deutliche Unterschiede zwischen den Schulen. Offenbar begrenzen schulinterne bzw. schulorganisatorische Faktoren die Möglichkeiten, das Förderangebot regelmäßig und im vorgesehenen Maße anzubieten (z.B. fehlende personelle Ressourcen). Unterstützen könnte auch hier die Einrichtung eines Förderbandes in die Stundentafel, was bislang nur von 11 % der Schulen umgesetzt wird.

### **Teil B: Nehmen die Kinder regelmäßig an der Förderung teil?**

Die Schülerinnen und Schüler, die einmal einer Fördergruppe zugeteilt wurden, nehmen auch regelmäßig an der angebotenen Förderung teil. Allein im Baustein 5 zeigt sich eine deutlich verminderte Teilnahme an nur 58 % der angebotenen Förderung. Da die Pandemie zu diesem Zeitpunkt ihren Anfang nahm, konnten die Kinder nicht mehr regelmäßig an der Förderung teilnehmen. Über alle Bausteine gemittelt besuchen die Kinder 82 % der angebotenen



Förderung. Insofern kann davon ausgegangen werden, dass die Teilnahme an der zusätzlichen Förderung für die Kinder größtenteils in den Schulalltag integriert werden konnte.

### **Teil B: Inwieweit wird die Förderung konzepttreu umgesetzt?**

Um einen Einblick in die tatsächlich durchgeführte Förderung zu erhalten, wurden Beobachtungen in den Trainingsgruppen durchgeführt. Diese ergeben, dass die Umsetzung der Fördereinheiten in einigen Aspekten konzepttreu erfolgte. Beispielsweise wurden die Aspekte Levelteilung (zur Binnendifferenzierung) und Arbeit in Tandems sinnvoll eingesetzt. Darüber hinaus wirkten sowohl die Schülerinnen und Schüler als auch die Lehrkräfte sehr routiniert und motiviert in der Nutzung des Förderkonzepts. Allerdings ergibt die Beobachtung auch, dass die Umsetzung nicht durchgängig konzepttreu erfolgte. Bei einzelnen Aspekten waren immer wieder Abweichungen vom Konzept zu beobachten. So griffen in etwa der Hälfte der Einheiten die verschiedenen Phasen des Trainingsprogramms nicht ineinander und in ca. jeder fünften Einheit gingen die Lehrkräfte nicht auf den spezifischen Unterstützungsbedarf der einzelnen Schülerinnen und Schüler ein. Neben diesen allgemeinen Auffälligkeiten ergeben sich außerdem Abweichungen, die ausschließlich bei Baustein 4 auftraten. Denn hier hielten sich die Lehrkräfte durchweg nicht genau an die im Manual beschriebenen Vorgehensweisen, sondern brachten eigene Ideen ein. Des Weiteren wirkten die Kinder in diesem Baustein häufig durch die Arbeit mit dem Trainingsprogramm überfordert und die Trainingsintervalle von 15-20 Min. erschienen bei der nicht-teilnehmenden Beobachtung in fast allen Fällen vom Umfang her nicht angemessen.

Diese Beobachtungen lassen darauf schließen, dass es den Lehrkräften bei einigen Merkmalen des Förderkonzepts leicht fiel, diese in der Förderpraxis im Schulalltag umzusetzen, bei anderen Merkmalen hingegen nicht. Ein Beispiel für Letzteres ist das Eingehen auf spezifische Unterstützungsbedarfe der einzelnen Schülerinnen und Schüler, was vermutlich zeitlichen Mehraufwand in der Umsetzung bedeutet gegenüber anderen Konzeptprinzipien. Die darüber hinaus gehenden Auffälligkeiten, die ausschließlich bei Baustein 4 auftraten, wie bspw. der Überforderung der Lernenden, könnten mit der oben konstatierten abnehmenden Regelmäßigkeit des Förderangebots erklärt werden. Es ist aber auch möglich, dass die Lehrkräfte sich durch ihre wachsende Vertrautheit mit dem Förderkonzept sowie ihre bereits gemachten Erfahrungen sicher genug in der Umsetzung fühlten, um Adaptionen vorzunehmen. Diese Annahme liegt auch deshalb nahe, da Baustein 3 und 4 didaktisch identisch gestaltet sind. Die Variationen des Konzepts durch die Lehrkräfte führten jedoch nicht zu einer Verbesserung, sondern bei den Schülerinnen und Schülern zu Schwierigkeiten in der Annahme der Förderung.

---

## **Teil B: Inwieweit wird die Fortbildung als Teil der Implementation von den Lehrkräften akzeptiert (Evaluation der Fortbildung)?**

Die teilnehmenden Lehrkräfte wurden aufgefordert, zwei Fragebögen zu bearbeiten. Das Ziel war zum einen, eine Einschätzung der Lehrkräfte zu den Materialien und zur Umsetzung in der Schulpraxis abzufragen, und zum anderen, eine Bewertung der flankierenden Fortbildungsmaßnahmen des IQSH zu erhalten.

Insgesamt lässt sich feststellen, dass die Fortbildungen, die durch das IQSH durchgeführt wurden, von den teilnehmenden Lehrkräften sowohl bezogen auf die allgemeinen Aspekte der Durchführung, wie z.B. die Expertise der Fortbildnerinnen, als auch bezogen auf die spezifischen fachlichen Inhalte sehr positiv bewertet wurden. Da die Durchführung der Förderung jedoch tendenziell abgenommen hat, die flankierenden Fortbildungen der Lehrkräfte jedoch vor jedem Meilenstein angeboten wurden, ist nicht davon auszugehen, dass hier ein Zusammenhang zwischen der Fortbildung und der Kontinuität der Durchführung von Förderung vorliegt. Vielmehr ist zu vermuten, dass andere schulinterne Aspekte dazu führten, dass einzelne Schulen auffällig wenig Förderung zum Ende der Intervention durchgeführt haben.

## **Teil B: Inwieweit wird das Training von den Lehrkräften akzeptiert?**

Die Lehrkräfte bewerten die Fokussierung des Trainingsprogramms auf die Automatisierung der basalen Lese- und Schreibfähigkeiten als positiv. Sie geben weiterhin an, dass sie sich durch das Training in ihrer Förderarbeit unterstützt fühlen und eine bedarfsgerechte Förderung durch die Nutzung des Trainingsprogramms erfolgte. Auch die Passung zwischen dem Lms-Diagnosematerial und dem ergänzenden Trainingsprogramm bewerten die Lehrkräfte als gut, obwohl feststellbar war, dass die Zuteilung der Kinder in die Fördergruppen oftmals nicht auf den Ergebnissen der Lms-Diagnostik basierte. Weniger positiv wurde der Aufwand bei der Vorbereitung auf die Arbeit mit dem Trainingsprogramm bewertet. Diesen schätzen die Lehrkräfte als eher hoch ein. Weiterhin schätzen einige Lehrkräfte die Fördergruppen als zu groß ein, obwohl die Fördergruppen auf eine Größe von sechs bis acht Kindern beschränkt wurden, was darauf schließen lässt, dass die Lehrkräfte die Förderung stark an- und begleiten mussten. Die beteiligten Lehrkräfte bewerteten das Material insgesamt jedoch grundsätzlich eher positiv bis sehr positiv.

## **Limitationen der Studie**

Studiendesign:

Die Ergebnisse bezüglich der Wirksamkeit des Lms-Trainings müssen vor dem Hintergrund betrachtet werden, dass sich die Kontrollschulen deutlich leistungsstärker zeigten als die

Experimentalschulen. Die Gründe hierfür sind unklar, da die Auswahl der Schulen auf Basis der VERA-Ergebnisse erfolgte. Die Schulen beider Gruppen ähneln sich sowohl in Bezug auf sozioökonomische Merkmale als auch hinsichtlich der weiteren Kontrollvariablen. Zudem lassen sich die Leistungsunterschiede zwischen den Kontroll- und Experimentalschulen nicht auf ein zusätzliches Training der Schülerinnen und Schüler im Bereich des Lesens bzw. Schreibens an den Kontrollschulen zurückführen.

Eine weitere Einschränkung der Generalisierbarkeit der Ergebnisse liegt im Studiendesign begründet. Vorgesehen war der Vergleich der Klassenleistungen der Schulen in der Experimentalgruppe mit denen der Kontrollgruppe. Das Lms-Training wurde jedoch nicht mit den gesamten Klassen, sondern nur mit einem ausgewählten Teil durchgeführt. Ein möglicher Effekt der zusätzlichen Förderung dieser kleinen Gruppe von Schülerinnen und Schülern wird somit durch die Betrachtung und den Vergleich der Gesamtleistung der Klassen verwässert. Um also den direkten Effekt des Lms-Trainings auf die Leistungsentwicklung der leistungsschwächeren Kinder feststellen zu können, wäre es sinnvoll gewesen, Lms-Diagnostik nicht nur in der Experimentalgruppe, sondern auch in der Kontrollgruppe einzusetzen. Die Entwicklung und der Leistungsstand der hierdurch identifizierten Gruppen hätten dann miteinander verglichen werden können. Da die regelmäßige Lms-Diagnostik in den Klassen der Kontrollschulen nicht durchgeführt wurde und somit eine entsprechende Identifizierung potenziell förderbedürftiger Kinder nicht möglich war, konnte die Leistung bzw. Leistungsentwicklung zwischen den Subgruppen „Trainingsgruppe der Experimentalgruppe“ und „potenziell förderbedürftige Kinder der Kontrollgruppe“ nicht analysiert werden.

**Pandemie:**

Aufgrund der Pandemie wurde die Durchführung der Förderung zunehmend schwieriger und nahm im weiteren Verlauf der Studie immer stärker ab. Die Studie musste schließlich vorzeitig beendet werden. Die Entwicklung der Schreibkompetenz konnte deshalb nicht betrachtet werden.

**Passung zwischen Diagnostik und Fördermaterial:**

Obwohl die Lehrkräfte die Passung zwischen der Diagnostik und dem Fördermaterial als gut einschätzen, ist eine derartige Passung nicht gegeben. Die Diagnostik erfasst differenziert den Leistungsstand der Kinder in unterschiedlichen Teilbereichen des Lesens und des Schreibens. So wurde z.B. das Schreiben von Pseudowörtern und realen Wörtern abgefragt, um orthographische Kompetenzen zu erfassen. Die Trainingsbausteine bilden jedoch diese Differenzierung nicht ab, da hier der Fokus auf das Trainieren automatisierter Lese- und Schreibfähigkeiten gefördert wurde. Kinder, die aufgrund schwächerer Leistungen im Schreiben der Trainingsgruppe zugeordnet wurden, erhielten so möglicherweise dem vorgesehenen Baustein entsprechend ein Lms-Training der Lesefähigkeit oder Kinder fielen im

Lesen auf und erhielten dem Baustein entsprechend ein Schreibtraining. Es ist aber davon auszugehen, dass die Lehrkräfte ihrer Erfahrung und subjektiven Einschätzung nach die Kinder den Trainingsgruppen zugeordnet haben und nicht nach der individuellen Leistung in der Lms-Diagnostik. Diese Annahme wird durch die Interviews im Nachgang zu der Studie bestätigt (s. oben).

Stichprobe:

Die anvisierte Stichprobengröße konnte aufgrund der hohen Anzahl fehlender Werte nicht erreicht werden. Die fehlenden Werte ergeben sich aus der relativ geringen Rücklaufquote des Elternfragebogens sowie aus der zum Teil lückenhaften Dokumentation der Lehrkräfte.

Varianz schulinterner Faktoren:

Die Daten geben Hinweise auf eine hohe Varianz schulinterner Faktoren, die einen erheblichen Einfluss auf die Durchführung und somit auch auf die Wirksamkeit des Lms-Trainings haben können. Dazu zählen insbesondere personelle Ressourcen (z.B. Vertretung im Krankheitsfall), die Einbindung des Lms-Trainings in den schulischen Alltag sowie die Nutzung der Lms-Diagnostik für die Zusammensetzung der Fördergruppen.

## 6. Fazit

Zusammenfassend lässt sich festhalten, dass das Lms-Training eine sinnvolle Ergänzung zur Lms-Diagnostik darstellt. In der ersten „Lesen macht stark – Grundschule“ Studie zeigten die trainierten Kinder keine Verbesserung ihrer Schreib- und Lesefähigkeiten. Folglich wurde für die zweite „Lesen macht stark – Training – Studie“ ein neues Trainingsmaterial entwickelt. Das strukturierte Trainingsmaterial sowie das systematische Vorgehen bei der Durchführung der Trainingsinhalte wurden sehr positiv von den Lehrkräften aufgenommen. Zwar ergab die begleitende Überprüfung der konzepttreuen Umsetzung, dass sich die Lehrkräfte sowohl bei der Auswahl der Kinder für die Trainingsgruppen sowie in einigen Aspekten der Trainingsdurchführung nicht vollständig an die Vorgaben hielten, dennoch konnte ein positiver Effekt der Wirksamkeit des Lesetrainings auf die Entwicklung des Wortschatzverständnisses der Schülerinnen und Schüler gefunden werden.

## 7. Literatur

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles: Policy and Practice*, 5(1), 7-74.

- 
- Hasselhorn, M., & Gold, A. (2017). *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren* (4. Aufl.). Stuttgart: Kohlhammer.
- Hattie, J., & Beywl, W., (2013). *Lernen sichtbar machen* (1. Aufl.). Baltmannsweiler: Schneider-Verlag.
- Heppt, B., Olczyk, M., & Volodina, A. (2022). Number of books at home as an indicator of socioeconomic status: Examining its extensions and their incremental validity for academic achievement. In *Social Psychology of Education: an international journal*, 25, 903-928.
- Lenhard, W., & Schneider, W. (2006). *ELFE 1-6: Ein Leseverständnistest für Erst-bis Sechstklässler* (1. Aufl.). Göttingen: Hogrefe.
- Mayer, A. (2011). *Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit (TEPHOBE)*. München/ Basel: Ernst Reinhardt Verlag.
- Moll, K., & Landerl, K. (2010). *SLRT-II: Lese- und Rechtschreibtest; Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)* (2. Aufl.). Bern: Hans Huber Verlag.
- Philipp, M. (2012). *Selbstreguliertes Lesen: Ein Überblick über wirksame Leseförderansätze* (1. Aufl.). Seelze: Kallmeyer.
- Simons, P. R.J. (1992). *Lernen selbständig zu lernen – ein Rahmenmodell* (1. Aufl.). In H. Mandl & H. F. Friedrich (Hrsg.), *Lern- und Denkstrategien – Analyse und Intervention* (1. Aufl.) (S. 251-264). Göttingen: Hogrefe.
- Weiß, R. H., & Osterland, J. (2012). *Grundintelligenztest Skala 1 – Revision (CFT 1 – R)* (1. Aufl.). Göttingen: Hogrefe.
- Rüßmann, L., Steinhoff, T., Marx, N. Wenk, A. (2016) Schreibförderung durch Sprachförderung? Zur Wirksamkeit sprachlich profilierter Schreibarrangements in der mehrsprachigen Sekundarstufe | unterschiedlicher Schulformen. In: *Didaktik Deutsch Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur* 21 (40), S. 41-59.

---

## 8. Anhang

### 8.1 Darstellung der Ergebnisse des Vergleichs der Experimentalgruppe mit der Kontrollgruppe

#### **Vergleich der Kontrollgruppe mit der gesamten Experimentalgruppe hinsichtlich der Kontrollvariablen**

##### Demographische Variablen

Um zu überprüfen, ob die beiden Gruppen hinsichtlich spezifischer demographischer und sozioökonomischer Merkmale vergleichbar sind, wurden verschiedene Kontrollvariablen mittels Fragebogen erhoben. Der deskriptive Vergleich zwischen den Gruppen zeigt, dass eine Vielzahl von Merkmalen in beiden Gruppen annähernd gleich verteilt ist. So beträgt das mittlere Alter der Schülerinnen und Schüler zum Zeitpunkt der ersten Datenerhebung im September 2018 in beiden Gruppen 6,8 Jahre (EG:  $SD = 0,48$ ;  $Min = 6,0$ ;  $Max = 8,4$ ; KG:  $SD = 0,46$ ;  $Min = 5,8$ ;  $Max = 8,3$ ). Auch das Geschlecht ist in beiden Gruppen annähernd gleich verteilt (EG: Mädchen = 51 %; KG: Mädchen = 50 %). Ein partieller oder vollständiger Migrationshintergrund liegt bei 21 % der Schülerinnen und Schüler in der Experimentalgruppe vor (59 % kein Migrationshintergrund, 20 % fehlende Angaben) und bei 29 % Kindern in der Kontrollgruppe (55 % kein Migrationshintergrund, 16 % fehlende Angaben). Zusätzlich wurde der sprachliche Hintergrund mit drei Variablen erfasst. Die Ergebnisse zeigen, dass in 56 % der Familien in der Experimentalgruppe nur Deutsch gesprochen wird, in 21 % der Familien wird Deutsch und eine andere Sprache gesprochen und von 23 % der Familien fehlen die Angaben. In der Kontrollgruppe sprechen 54 % der Familien ausschließlich Deutsch Zuhause, in 27 % der Familien wird Deutsch und eine andere Sprache gesprochen und von 19 % der Familien fehlen die Angaben. Die Familien, in denen auch eine andere Sprache als Deutsch gesprochen wird, geben zusätzlich an, welche Sprache das Kind Zuhause spricht und welche Sprache hauptsächlich in der Familie gesprochen wird. Dabei zeigt sich, dass diese Variablen nicht ganz gleich in beiden Gruppen verteilt sind. In der Experimentalgruppe sprechen 15 % der Kinder nur Deutsch Zuhause und 85 % der Kinder Deutsch und eine andere Sprache. In der Kontrollgruppe sprechen hingegen 26 % der Kinder nur Deutsch und 74 % der Kinder sprechen Deutsch und eine andere Sprache. Auch in Bezug auf die Familiensprache zeigt sich, dass in der Experimentalgruppe etwas weniger Familien nur Deutsch sprechen (44 %) als in der Kontrollgruppe (50 %) und etwas mehr Familien in der Experimentalgruppe (56 %) sprechen zusätzlich noch eine andere Sprache Zuhause als in Familien der Kontrollgruppe (50 %). Befragt nach den Deutschkenntnissen der Kinder geben mehr Eltern in der Kontrollgruppe an, dass ihre Kinder sehr gute bis gute Kenntnisse verfügen (96 %) als in der Experimentalgruppe

(82 %). Der sozioökonomische Status der Eltern wurde mittels des Bücherbesitzes erfasst. Die beiden Gruppen unterscheiden sich nicht signifikant hinsichtlich der Anzahl des Besitzes von Bücher Zuhause ( $t(422) = -1,224$ ;  $p = ,22$ ; Cohens  $d = 0,12$ ).

## TEPHOBE

Zur ersten Kontrolle der Voraussetzungen der Schülerinnen und Schüler wurden die Untertests „Anlaute“ und „Reimen“ des standardisierten Tests TEPHOBE zu Beginn der 1. Klassenstufe eingesetzt. Es liegen valide Daten von 411 Schülerinnen und Schülern im Untertest „Reimen“ und  $N = 365$  für den Untertest „Anlaute“ vor. Die Ergebnisse zeigen keine signifikanten Unterschiede zwischen den Gruppen (Reimen:  $t(409) = 0,236$ ;  $p = ,81$ ; Cohens  $d = 0,02$ ; Anlaute:  $t(363) = -0,163$ ;  $p = ,87$ ; Cohens  $d = 0,02$ ). Abbildung A1 zeigt die Mittelwerte der Untertests beider Gruppen. Folglich unterscheiden sich die beiden Gruppen nicht hinsichtlich ihrer Vorläuferfähigkeiten des Schriftspracherwerbs.

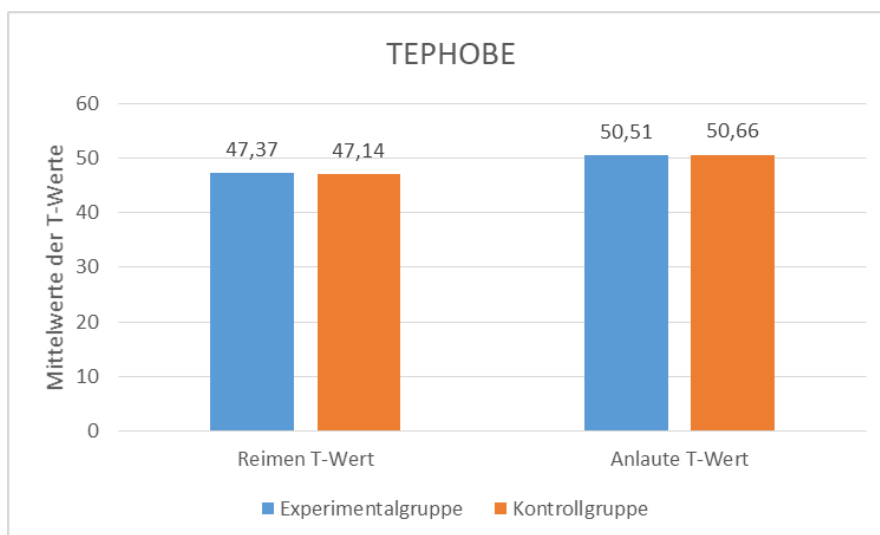


Abb. A1: Mittelwerte der T-Werte in den beiden TEPHOBE Untertests „Anlaute“ und „Reime“ für die beiden Gruppen (Experimentalgruppe vs. Kontrollgruppe)

## CFT 1-R

Als Kontrollvariable und zur Beschreibung der Stichprobe wurde der CFT 1-R zum letzten Erhebungsmesszeitpunkt eingesetzt. Es liegen Daten von 427 Schülerinnen und Schülern vor. Die Analyse zeigt, dass sich die beiden Gruppen nicht signifikant voneinander unterscheiden ( $t(425) = 1,512$ ;  $p = ,13$ ; Cohens  $d = 0,15$ ). Die Mittelwerte der beiden Gruppen sind in Abbildung A2 dargestellt.

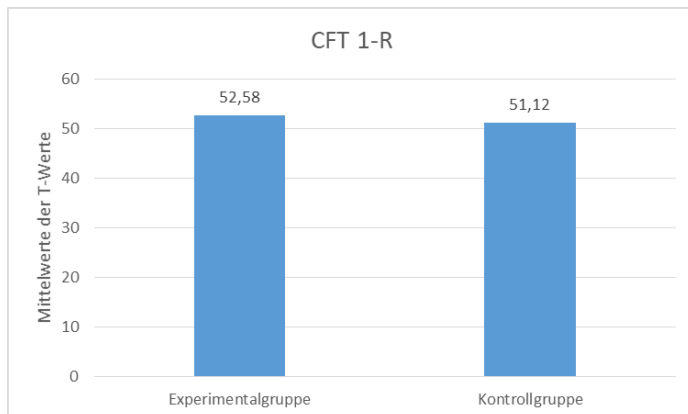


Abb. A2: Mittelwerte der T-Werte im CFT 1-R für die beiden Gruppen (Experimentalgruppe vs. Kontrollgruppe)

## Vergleich der Kontrollgruppe mit der gesamten Experimentalgruppe hinsichtlich der Fördermaßnahmen

### Entwicklung der Lesekompetenz

Der ELFE-Lesekompetenztest besteht aus den Untertests Wort-, Satz- und Textverständnis. Da nicht zu beiden Zeitpunkten alle Untertests durchgeführt wurden, konnte keine Veränderung des Gesamttestwerts errechnet werden. Die Untertests wurden daher getrennt betrachtet. Durchgeführt wurde eine Varianzanalyse (ANOVA) mit dem Messwiederholungsfaktor *Zeit* und der unabhängigen Variable *Gruppe* (EG, KG). Abhängige Variablen waren die T-Werte in den beiden entsprechenden Untertests (Wort- und Satzverständnis).

Für den Untertest „Wortverständnis“ zeigen die Ergebnisse einen signifikanten Haupteffekt des Faktors *Gruppe* ( $F(1, 368) = 7,709; p = ,006$ ; partielles  $\eta^2 = ,021$ ). Das bedeutet, dass sich die beiden Gruppen über die beiden Messzeitpunkte hinweg hinsichtlich der Ergebnisse im Wortverständnis unterscheiden. Dabei zeigt die Kontrollgruppe einen höheren Wert im Wortverständnistest als die Experimentalgruppe. Der signifikante Haupteffekt des Faktors *Zeit* ( $F(1, 369) = 15,155; p < ,001$ ; partielles  $\eta^2 = ,04$ ) gibt an, dass sich das Wortverständnis über die Zeit verbessert hat (über beide Gruppen hinweg). In der Abbildung A3 lässt sich erkennen, dass sich der Zuwachs der Wortverständnissfähigkeit von Messzeitpunkt 1 zu Messzeitpunkt 2 zwischen den Gruppen signifikant unterscheidet ( $F(1, 369) = 4,439; p = ,036$ ; partielles  $\eta^2 = ,012$ ). Dabei ist der Zuwachs der Wortverständnissfähigkeit von Messzeitpunkt 1 zu Messzeitpunkt 2 höher für die Kontrollgruppe als für die Experimentalgruppe (siehe Abbildung A3). Allerdings ist der Effekt als klein zu bezeichnen.



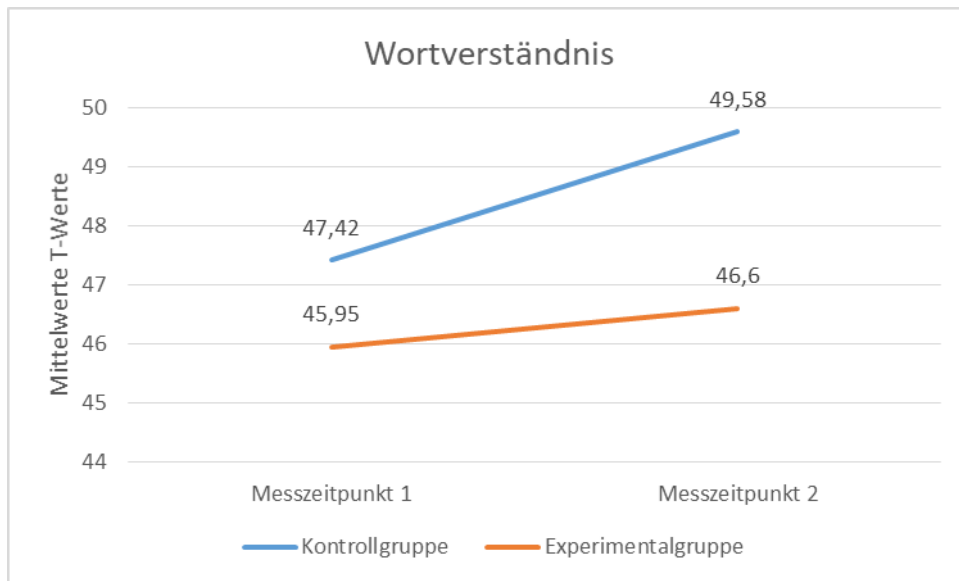


Abb. A3: Mittelwerte im Wortverstehen für beide Gruppen (Experimentalgruppe und Kontrollgruppe) zum ersten und zweiten Messzeitpunkt

Der Vergleich der gesamten Experimentalgruppe mit der Kontrollgruppe bezüglich der Satzverständnissfähigkeit zeigt einen signifikanten Haupteffekt des Faktors Gruppe ( $F(1, 369) = 17,834$ ;  $p < ,001$ ; partielles  $\eta^2 = ,046$ ) sowie einen signifikanten Haupteffekt des Faktors Zeit ( $F(1, 369) = 97,760$ ;  $p < ,001$ ; partielles  $\eta^2 = ,209$ ). Das bedeutet, dass sich zum einen die beiden Gruppen über die beiden Messzeitpunkte hinweg hinsichtlich der Ergebnisse im Wortverständnis unterscheiden und zum anderen, dass sich über beide Gruppen hinweg das Wortverständnis über die Zeit verbessert hat. Allerdings kann kein unterschiedlich starker Anstieg der Satzverständnissfähigkeit vom ersten zum zweiten Messzeitpunkt in Abhängigkeit der Gruppe gefunden werden ( $F(1, 369) = 3,361$ ;  $p = ,068$ ; partielles  $\eta^2 = ,009$ ). In Abbildung A4 sind die Werte graphisch verdeutlicht. Beide Gruppen zeigen somit einen vergleichbaren Leistungsanstieg.

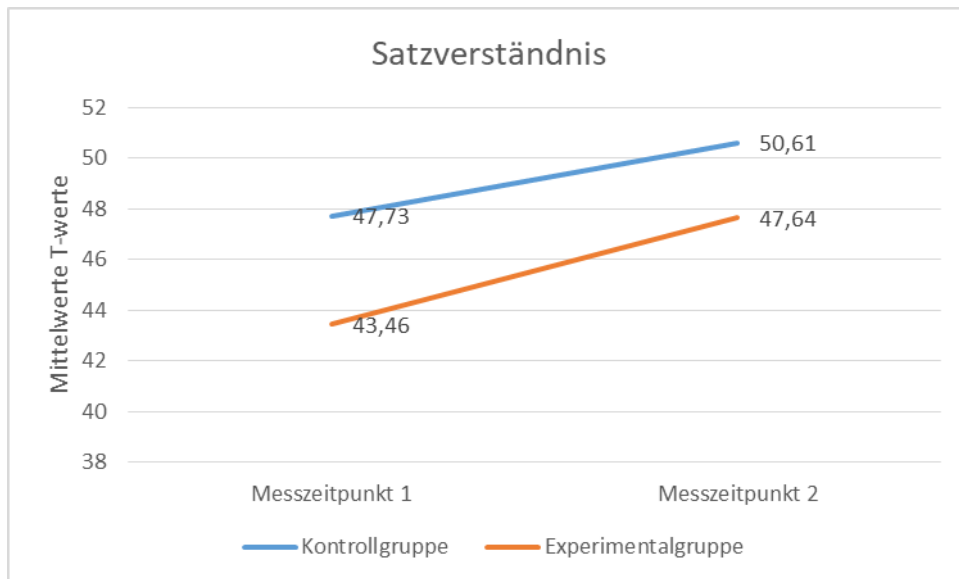


Abb. A4: Mittelwerte im Satzverstehen für beide Gruppen (Experimentalgruppe und Kontrollgruppe) zum ersten und zweiten Messzeitpunkt

Da der Untertest „Textverständnis“ des ELFE nur zu einem Messzeitpunkt erhoben wurde, liegen nur querschnittliche Daten vor. Die Ergebnisse zeigen, dass sich der Mittelwert des Untertests Textverständnis ( $M = 46,96$ ) der Experimentalgruppe signifikant vom Mittelwert des Untertests Textverständnis der Kontrollgruppe ( $M = 49,82$ ) unterscheidet ( $t(418) = 3,535$ ;  $p < ,001$ ; Cohens  $d = 0,35$ ).

### Vergleich der Kontrollgruppe mit der gesamten Experimentalgruppe bezüglich der Schreibkompetenz

Da der SLRT nur zu einem Messzeitpunkt erhoben wurde, liegen nur querschnittliche Daten vor und ein Vergleich der Leistungsentwicklung ist nicht möglich. Der SLRT liefert drei verschiedene abhängige Variablen. Hinsichtlich der O-Fehler gibt es keinen signifikanten Unterschied zwischen der Experimentalgruppe und der Kontrollgruppe ( $t(436) = ,487$ ;  $p = ,627$ ; Cohens  $d = 0,05$ ). Die Schülerinnen und Schüler beider Gruppen machen somit ähnlich viele orthographische Fehler (siehe Abbildung A5). Ein Unterschied zeigt sich bei den lautgetreuen Fehlern (N-Fehler). Die Experimentalgruppe macht signifikant mehr N-Fehler als die Kontrollgruppe ( $t(436) = 5,542$ ;  $p < ,001$ ; Cohens  $d = 0,53$ ). Bezüglich der Fehler in der Groß- und Kleinschreibung (G-Fehler) zeigt sich, dass die Experimentalgruppe ebenfalls signifikant mehr Fehler macht als die Kontrollgruppe ( $t(436) = 2,101$ ;  $p = ,36$ ; Cohens  $d = 0,20$ ).

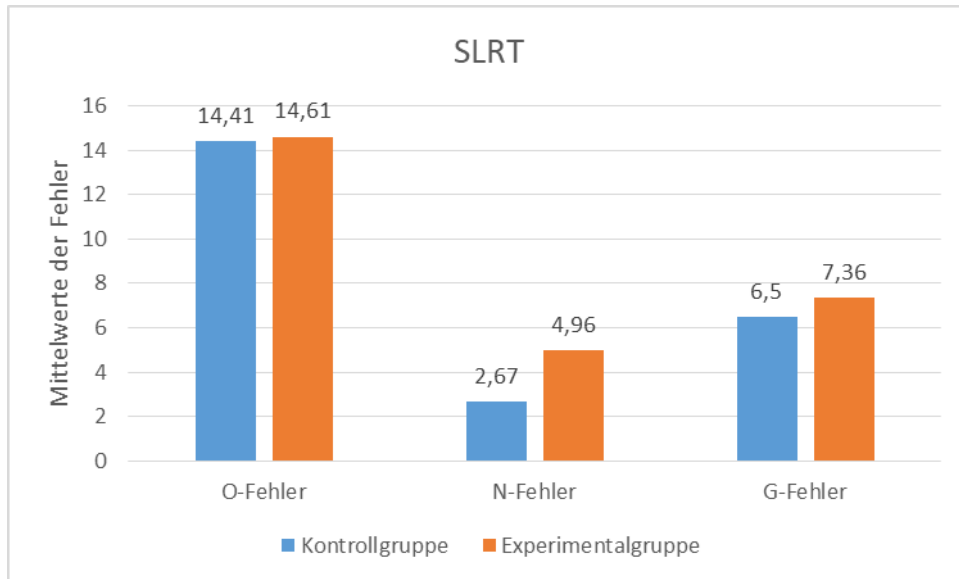


Abb. A5: Mittlere Fehleranzahl für drei Fehlerarten: orthographische Fehler (O-Fehler), lautgetreue Fehler (N-Fehler) und Fehler der Groß-Kleinschreibung (G-Fehler) für beide Gruppen (Experimentalgruppe vs. Kontrollgruppe)

## 8.2 Darstellung der Ergebnisse des Vergleichs der Trainingsgruppe mit der Experimentalgruppe ohne Training und der Kontrollgruppe

Für die folgenden Analysen wurden die Kinder der Experimentalgruppe ausgewählt, die an mindestens einem Baustein des Trainings teilgenommen haben (Trainingsgruppe). Kinder, die nur am fünften Baustein teilgenommen haben, wurden nicht in die Gruppe eingeschlossen, da der fünfte Trainingsbaustein aufgrund der Corona Pandemie erst nach der Erhebung der Leistungstest stattgefunden hat. Die Trainingsgruppe wurde zum einen mit der Kontrollgruppe und zum anderen mit den Kindern der Experimentalgruppe, die an keinem Trainingsbaustein teilgenommen haben, verglichen.

Der deskriptive Vergleich zwischen den Gruppen zeigt, dass einige Merkmale in den Gruppen annähernd gleich verteilt sind. So beträgt das mittlere Alter der Schülerinnen und Schüler zum Zeitpunkt der ersten Datenerhebung im September 2018 in allen drei Gruppen 6,8 Jahre (TG:  $SD = 0,39$ ,  $Min = 6,08$ ,  $Max = 8,08$ ; EGoT:  $SD = 0,49$ ;  $Min = 6,0$ ;  $Max = 8,4$ ; KG:  $SD = 0,46$ ;  $Min = 5,8$ ;  $Max = 8,3$ ). Ein partieller oder vollständiger Migrationshintergrund liegt bei 18 % der Schülerinnen und Schüler der Trainingsgruppe (53 % kein Migrationshintergrund, 29 % fehlende Angaben) und bei 22 % der Schülerinnen und Schüler in der Experimentalgruppe ohne Training vor (61% kein Migrationshintergrund, 17 % fehlende Angaben) sowie bei 29 % Kindern in der Kontrollgruppe (55 % kein Migrationshintergrund, 16 % fehlende Angaben). Zusätzlich wurde der sprachliche Hintergrund mit drei Variablen erfasst. Die Ergebnisse

zeigen, dass in 50 % der Familien in der Trainingsgruppe nur Deutsch gesprochen wird, in 18 % der Familien wurden Deutsch und eine andere Sprache gesprochen und von 32 % der Familien fehlen die Angaben. In der Experimentalgruppe ohne Training sprechen 58 % der Familien ausschließlich Deutsch Zuhause, in 22 % der Familien wurden Deutsch und eine andere Sprache gesprochen und von 20 % der Familien fehlen die Angaben. In der Kontrollgruppe sprechen 54 % der Familien ausschließlich Deutsch Zuhause, in 27 % der Familien wurden Deutsch und eine andere Sprache gesprochen und von 19 % der Familien fehlen die Angaben. Die Familien, in denen auch eine andere Sprache als Deutsch gesprochen wird, geben zusätzlich an, welche Sprache das Kind Zuhause spricht und welche Sprache hauptsächlich in der Familie gesprochen wird. Dabei zeigt sich, dass diese Variablen nicht ganz gleich zwischen den Gruppen verteilt sind. In der Trainingsgruppe sprechen 8 % der Kinder nur Deutsch Zuhause und 92 % der Kinder Deutsch und eine andere Sprache. In der Experimentalgruppe ohne Training sprechen 16 % der Kinder nur Deutsch Zuhause und 84 % der Kinder Deutsch und eine andere Sprache. In der Kontrollgruppe sprechen hingegen 26 % der Kinder nur Deutsch und 74 % der Kinder sprechen Deutsch und eine andere Sprache. In Bezug auf die Familiensprache zeigt sich, dass in der Experimentalgruppe ohne Training etwas weniger Familien nur Deutsch sprachen (40 %) als in der Trainingsgruppe (50 %) und der Kontrollgruppe (50 %). Befragt nach den Deutschkenntnissen der Kinder, geben mehr Eltern in der Kontrollgruppe (96 %) an, dass ihre Kinder sehr gute bis gute Kenntnisse verfügen als in der Experimentalgruppe ohne Training (85 %) und in der Trainingsgruppe (83 %). Da die letzten drei erhobenen Variablen nur von Eltern beantwortet wurden, deren Familiensprache nicht ausschließlich Deutsch war, wurden diese Variablen aufgrund der somit geringen Stichprobengröße nicht als Kovariaten verwendet.

Der sozioökonomische Status der Eltern wurde mittels des Bücherbesitzes erfasst. Die Gruppen unterscheiden sich nicht signifikant hinsichtlich der Anzahl des Besitzes von Bücher Zuhause ( $F(2, 405) = 1,164$ ;  $p = ,313$ ; partielles  $\eta^2 = ,006$ ).

Bei der Betrachtung des Geschlechts zeigt sich, dass mehr Jungen in der Trainingsgruppe (TG: Jungen = 63 %) waren als in der Experimentalgruppe ohne Training bzw. der Kontrollgruppe (EGoT: Jungen = 42 %; KG: Jungen = 50 %). Diese Variable wurde somit als Kovariate mit in die Analysen aufgenommen.

#### TEPHOBE

Zur ersten Kontrolle der Voraussetzungen der Schülerinnen und Schüler wurden die Untertests „Anlaute“ und „Reimen“ des standardisierten Tests TEPHOBE zu Beginn der 1. Klassenstufe eingesetzt. Es liegen valide Daten von 397 Schülerinnen und Schülern im Untertest „Reimen“ und  $N = 351$  für den Untertest „Anlaute“ vor. Die Ergebnisse zeigen signifikante Unterschiede zwischen den Gruppen (Reimen:  $F(2, 394) = 16,521$ ;  $p < ,001$ ; partielles  $\eta^2 = ,077$ ; Anlaute:  $F(1, 348) = 10,859$ ;  $p < ,001$ ; partielles  $\eta^2 = ,059$ ). Abbildung

A6 zeigt die Mittelwerte der Untertests beider Gruppen. Folglich unterscheiden sich die Gruppen hinsichtlich ihrer Vorläuferfähigkeiten des Schriftspracherwerbs und die beiden TEPHOBE T-Werte der Untertests werden als Kovariaten mit in die Analysen aufgenommen.

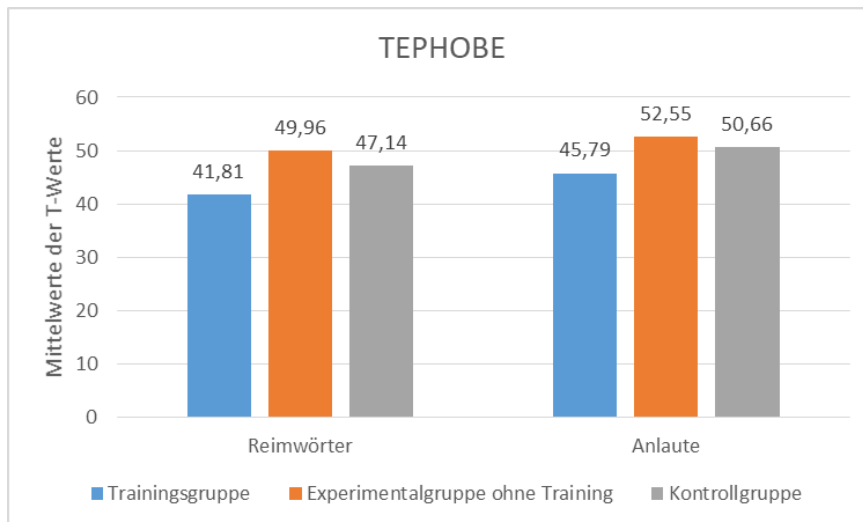


Abb. A6: Mittelwerte der T-Werte in den beiden TEPHOBE Untertests „Anlaute“ und „Reime“ für die drei Gruppen (Trainingsgruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe)

#### CFT 1-R

Als Kontrollvariable und zur Beschreibung der Stichprobe wurde der CFT 1-R zum letzten Erhebungsmesszeitpunkt eingesetzt. Es liegen Daten von 408 Schülerinnen und Schülern vor. Die Analyse zeigt, dass sich die drei Gruppen signifikant voneinander unterscheiden ( $F(2, 405) = 5,832$ ;  $p = ,003$ ; partielles  $\eta^2 = ,028$ ). Folglich wird der Intelligenzwert als Kovariate mit in die Analysen aufgenommen. Die Mittelwerte der drei Gruppen sind in Abbildung A7 dargestellt.

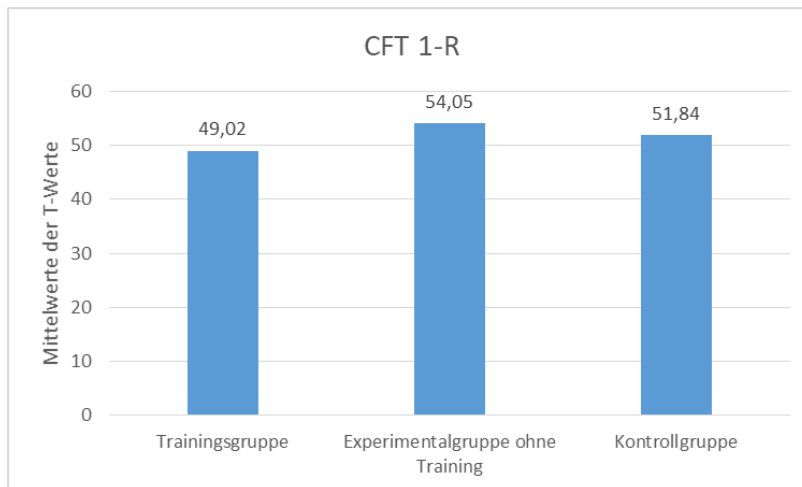


Abb. A7: Mittelwerte der T-Werte im CFT 1-R für die drei Gruppen (Trainingsgruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe)

### 8.3 Vergleich der Trainingsgruppe mit der Kontrollgruppe und der Experimentalgruppe ohne Training hinsichtlich der Fördermaßnahmen

#### Entwicklung der Lesekompetenz

Zur Überprüfung der Wirksamkeit des Lms-Trainings wurden nur die Schülerinnen und Schüler aus der Experimentalgruppe ausgewählt, die an mindestens einem Trainingsbaustein (Baustein 1, 2, 3 oder 4) teilgenommen haben. Diese Trainingsgruppe wurde zum einen mit der gesamten Kontrollgruppe verglichen und zum anderen mit den Kindern der Experimentalgruppe, die kein Training erhalten haben. Der ELFE-Lesekompetenztest besteht aus den Untertests Wort-, Satz- und Textverständnis. Da nicht zu beiden Zeitpunkten alle Untertests durchgeführt wurden, kann keine Veränderung des Gesamtwertes errechnet werden. Die Untertests werden daher getrennt betrachtet. Durchgeführt wurde eine Varianzanalyse mit Kovariaten (ANCOVA) mit dem Messwiederholungsfaktor *Zeit* und der unabhängigen Variable *Gruppe* (TG, KG, EGoT). Abhängige Variablen waren die T-Werte in den beiden entsprechenden Untertests (Wort- und Satzverständnis).

Für den Untertest „Wortverständnis“ liegen insgesamt Daten für 246 Schülerinnen und Schüler vor. Die Ergebnisse zeigen einen signifikanten Haupteffekt des Faktors *Gruppe* ( $F(2, 239) = 7,764; p < ,001$ ; partielles  $\eta^2 = ,061$ ) sowie einen signifikanten Haupteffekt des Faktors *Zeit* ( $F(1, 239) = 11,767; p < ,001$ ; partielles  $\eta^2 = ,047$ ). Das bedeutet, dass sich zum einen die beiden Gruppen über die beiden Messzeitpunkte hinweg hinsichtlich der Ergebnisse im Wortverständnis unterscheiden und zum anderen, dass sich über beide Gruppen hinweg das Wortverständnis über die Zeit verbessert hat. Weiterhin zeigt sich, dass sich der Zuwachs der

Wortverständnisfähigkeit von Messzeitpunkt 1 zu Messzeitpunkt 2 zwischen den Gruppen signifikant unterscheidet ( $F(2, 239) = 8,247; p < ,001; \text{partielles } \eta^2 = ,065$ ). Eine differenzierte Betrachtung der Gruppen zeigt, dass der Leistungszuwachs der Schülerinnen und Schüler, die an mindestens einem Trainingsbaustein teilnahmen innerhalb eines Schulhalbjahres einen signifikant größeren Leistungszuwachs als die Experimentalgruppe ohne Training erzielen ( $F(1, 106) = 5,779; p < ,001; \text{partielles } \eta^2 = ,052$ ). Hingegen ist der Leistungszuwachs der Kontrollgruppe und der Trainingsgruppe ähnlich. In der Abbildung A8 sind die Werte graphisch verdeutlicht.

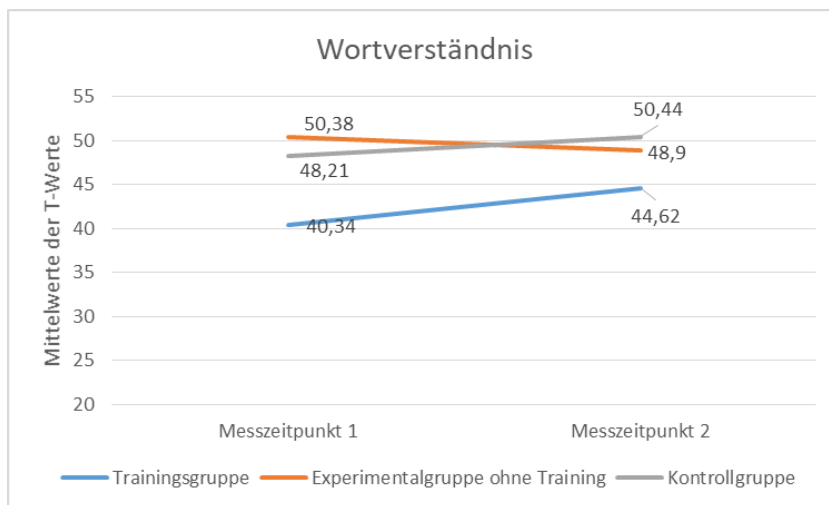


Abb. A8: Mittelwerte im Wortverstehen für die drei Gruppen (Trainingsgruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe) zum ersten und zweiten Messzeitpunkt

## Satzverstehen

Für den Untertest „Satzverständnis“ liegen Daten für 247 Schülerinnen und Schüler vor. Die Ergebnisse zeigen einen signifikanten Haupteffekt des Faktors Gruppe ( $F(2, 240) = 9,643; p < ,001; \text{partielles } \eta^2 = ,074$ ) sowie einen signifikanten Haupteffekt des Faktors Zeit ( $F(1, 240) = 9,009; p = ,003; \text{partielles } \eta^2 = ,036$ ). Das bedeutet, dass sich zum einen die beiden Gruppen über die beiden Messzeitpunkte hinweg hinsichtlich der Ergebnisse im Satzverständnis unterscheiden und zum anderen, dass sich über beide Gruppen hinweg das Satzverständnis über die Zeit verbessert hat. Allerdings kann kein unterschiedlich starker Anstieg der Satzverständnisfähigkeit vom ersten zum zweiten Messzeitpunkt in Abhängigkeit der Gruppe gefunden werden ( $F(2, 240) = 1,635; p = ,197; \text{partielles } \eta^2 = ,013$ ). Eine differenzierte Betrachtung der Gruppen zeigt, dass der Leistungszuwachs der Schülerinnen und Schüler, die an mindestens einem Trainingsbaustein teilnahmen innerhalb eines Schulhalbjahres einen signifikant größeren Leistungszuwachs als die Kontrollgruppe erzielen ( $F(1, 157) = 4,395; p = ,038; \text{partielles } \eta^2 = ,027$ ). Hingegen ist der Leistungszuwachs der Trainingsgruppe und der

Experimentalgruppe ohne Training ähnlich ( $F(1, 104) = 0,653$ ;  $p = ,421$ ; partielles  $\eta^2 = ,006$ ). In Abbildung A9 sind die Werte graphisch verdeutlicht.

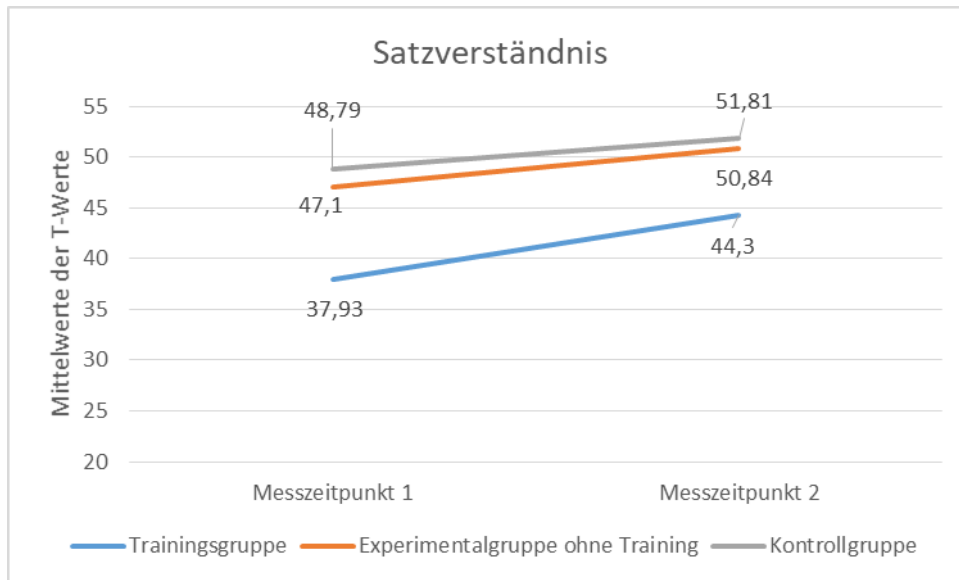


Abb. A9: Mittelwerte im Satzverstehen für die drei Gruppen (Trainingsgruppe vs. Experimentalgruppe ohne Training vs. Kontrollgruppe) zum ersten und zweiten Messzeitpunkt

### Textverständnis

Da der Untertest „Textverständnis“ des ELFE nur zu einem Messzeitpunkt erhoben wurde, liegen nur querschnittliche Daten vor. Die Ergebnisse zeigen, dass sich das Textverständnis der Trainingsgruppe ( $M = 43,47$ ) signifikant vom Textverständnis der Experimentalgruppe ohne Training ( $M = 49,10$ ) ( $t(398) = 34,173$ ;  $p < ,001$ ;  $d = 0,69$ ) und der Kontrollgruppe ( $M = 49,82$ ) unterscheidet ( $t(398) = 45,020$ ;  $p < ,001$ ;  $d = 0,78$ ).

### Schreibkompetenz

Da der SLRT nur zu einem Messzeitpunkt erhoben wurde, liegen nur querschnittliche Daten vor. Der SLR liefert drei verschiedene abhängige Variablen. Die einfaktorielle ANOVA zeigt, dass sich die Gruppen in Bezug auf alle Fehlerarten signifikant voneinander unterscheiden (O-Fehler:  $F(2,404) = 4,613$ ;  $p < ,001$ ; partielles  $\eta^2 = .022$ ; N-Fehler:  $F(2,404) = 71,233$ ;  $p < ,001$ ; partielles  $\eta^2 = ,261$ ; G-Fehler:  $F(2,404) = 8,518$ ;  $p < ,001$ ; partielles  $\eta^2 = .040$ ). Eine differenzierte Betrachtung der Gruppen mittels geplanter Kontrastanalysen zeigt, dass es keinen signifikanten Unterschied zwischen der Trainingsgruppe und der Kontrollgruppe ( $t(421) = 1,755$ ;  $p = ,080$ ; Cohens  $d = 0,25$ ) in Bezug auf O-Fehler (orthographische Fehler) gibt. Die Schülerinnen und Schüler beider Gruppen machen somit ähnlich viele orthographische



Fehler (siehe Abbildung A10). Hingegen zeigt sich, dass die Trainingsgruppe mehr orthographische Fehler macht als die Experimentalgruppe ohne Training ( $t(421) = -2,668, p = ,008$ ; Cohens  $d = 0,41$ ). Ein sehr bedeutsamer Unterschied zeigt sich bei den lautgetreuen Fehlern (N-Fehler). Die Trainingsgruppe macht signifikant mehr N-Fehler als die Kontrollgruppe ( $t(421) = -11,387, p < ,001$ ; Cohens  $d = 1,59$ ) sowie mehr Fehler als die Experimentalgruppe ohne Training ( $t(421) = 10,247; p < ,001$ ; Cohens  $d = 1,57$ ). Die Effektstärke gibt einen sehr starken Effekt an. Bezüglich der Fehler in der Groß- und Kleinschreibung (G-Fehler) zeigt sich, dass die Trainingsgruppe ebenfalls signifikant mehr Fehler macht als die Kontrollgruppe ( $t(421) = -3,969; p < ,001$ ) sowie die Experimentalgruppe ohne Training ( $t(421) = 3,44; p < ,001$ ). Der Effekt ist als mittlerer Effekt einzuordnen (Cohens  $d = 0,56$  bzw.  $d = 0,53$ ).

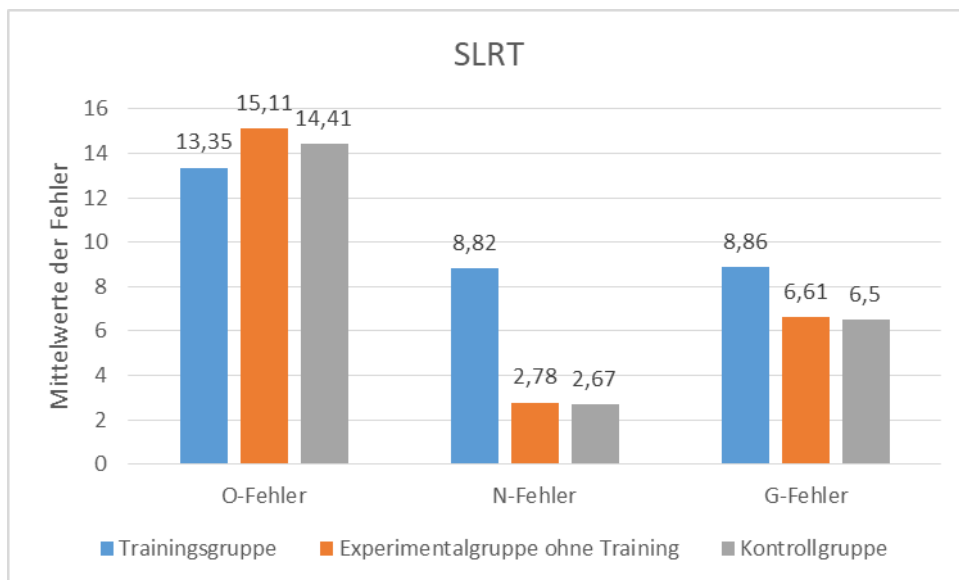


Abb. A10: Mittlere Fehleranzahl für drei Fehlerarten: orthographische Fehler (O-Fehler), lautgetreue Fehler (N-Fehler) und Fehler der Groß-Kleinschreibung (G-Fehler) für die drei Gruppen (T vs. EGoT vs. KG)

#### 8.4 Kurzdarstellung weiterer Analysen

##### Vergleich der Schreibgruppe mit der Experimentalgruppe und Kontrollgruppe bezüglich der Leseentwicklung

Der Vergleich der Schülerinnen und Schüler, die an mindestens einem Baustein des Schreibtrainings (Baustein 3 und/oder 4) teilgenommen haben mit der Experimentalgruppe ohne Training und der Kontrollgruppe, zeigt ähnliche Ergebnisse wie die anderen Analysen (Anhang II und Hauptanalysen Abschnitt 4.1.3).

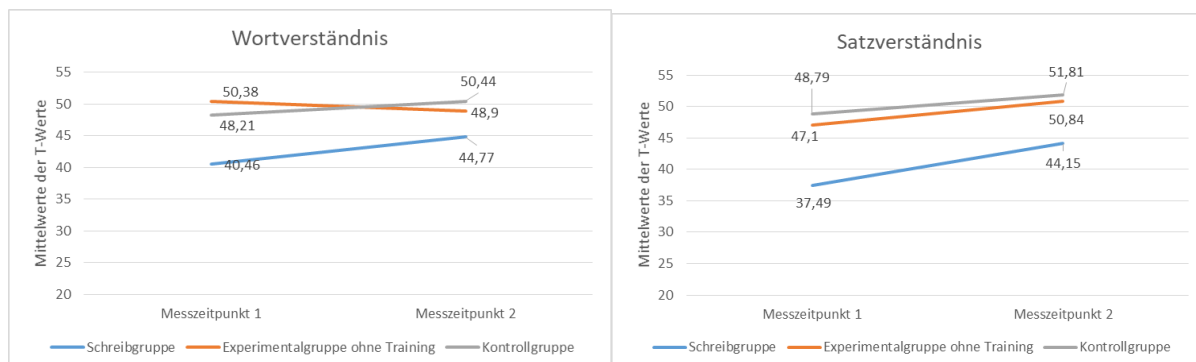


Abb. A11: Mittelwerte der T-Tests für den Untertest Wortverständnis (links) und Satzverständnis (rechts für die drei Gruppen (SG vs. EGoT vs. KG)

	Wortverständnis	Satzverständnis	Textverständnis
Haupteffekt Gruppe	$F(2,232) = 5,714; p = ,004; \text{partielles } \eta^2 = ,047$	$F(2,233) = 8,051; p < ,001; \text{partielles } \eta^2 = ,065$	$F(2,382) = 10,317; p < ,001; \text{partielles } \eta^2 = ,05$
Haupteffekt Zeit	$F(1,232) = 2,977; p = ,086; \text{partielles } \eta^2 = ,013$	$F(1,233) = 10,029; p = ,002; \text{partielles } \eta^2 = ,041$	n.a.
Interaktionseffekt Gruppe x Zeit	$F(2,232) = 5,215; p = ,006; \text{partielles } \eta^2 = ,043$	$F(2,233) = 1,457; p = ,235; \text{partielles } \eta^2 = ,012$	n.a.

Tab. A1: Darstellung der Ergebnisse der ANCOVA für die 3 Untertests des ELFE des Vergleichs der drei Gruppen (SG vs. EGoT vs. KG)

## 8.5 Zusammenfassung der zusätzlichen Analysen

Die Zusatzanalysen zeigen, dass beim Vergleich zwischen der gesamten Experimentalgruppe und der Kontrollgruppe keine Effekte der Wirksamkeit des Trainings nachgewiesen werden können. Die Lesefähigkeit sowie Schreibfähigkeit der Experimentalgruppe sind sogar teilweise als schlechter einzuschätzen als die der Kontrollgruppe, obwohl sich die beiden Gruppen hinsichtlich der Kontrollvariablen nicht unterscheiden. Auch steigt die Wortverständnisfähigkeit der Kontrollgruppe über die Zeit stärker an als in der Experimentalgruppe. Da nicht alle Schülerinnen und Schüler der Experimentalgruppe an den Trainingsbausteinen teilnahmen, kann der Effekt des Lms-Trainings maskiert sein.

Somit erscheint es sinnvoll, Differentialgruppen zu bilden und nur die Schülerinnen und Schüler für die Analysen auszuwählen, die an dem Training teilnahmen. Die Hauptanalysen (Abschnitt 4.1.3) zeigen einen positiven Effekt des Lms-Lesetrainings auf die Lesefähigkeit. Die

Lesegruppe (Schüler und Schülerinnen, die an mindestens einem Trainingsbaustein Lesen teilnahmen) verbesserte ihre Wortverständnisfähigkeit im Vergleich zu Kindern aus der Experimentalgruppe, die kein Training erhalten haben, deutlich. Eine Verbesserung der Lesegruppe im Wortverständnis zeigt sich auch, wenn auch in einem geringeren Ausmaß, im Vergleich zur Kontrollgruppe, also Schulen, an denen keine Lms-Diagnostik bzw. Lms-Training durchgeführt wurde. Bezüglich der Kompetenzen im Satzverständnis zeigt die Lesegruppe nur im Vergleich mit der Kontrollgruppe eine Verbesserung, nicht jedoch mit der Experimentalgruppe ohne Training. Bezüglich der Schreibfähigkeiten sind keine Aussagen zur Wirkung des Trainings möglich, da es nur einen Messzeitpunkt gab. Die Ergebnisse zeigen, dass die Kinder der Schreibgruppe mehr Fehler (lautgetreue Fehler und Fehler der Groß- und Kleinschreibung) als die Experimentalgruppe ohne Training und die Kontrollgruppe machen. Hinsichtlich der orthographischen Fehler zeigen sich jedoch keine deutlichen Unterschiede zwischen den Gruppen. Da es keine Ausgangsmessung der Schreibfähigkeit gibt, kann nicht geklärt werden, ob die Unterschiede in den Fehlern nicht bereits vor dem Training sogar noch stärker ausgeprägt waren.

Zusätzlich wurden die Analysen mit den Schülerinnen und Schülern wiederholt, die an mindestens einem Baustein teilgenommen haben (Trainingsgruppe). Die Ergebnisse sind vergleichbar mit den Ergebnissen der Lesegruppe. Da die Trainingsgruppe auch Schülerinnen und Schüler der Lesegruppe enthält, ist das Ergebnis nicht überraschend.

## 8.6 Interview Schulen

Schul-ID	Wie haben Sie als LK entschieden, ob ein Kind in die Trainingsgruppe aufgenommen wurde oder nicht?	Wie wurde Ihnen in der Fortbildung vermittelt, welche Kinder Sie in die Trainingsgruppen aufnehmen sollten?
201	<ul style="list-style-type: none"> <li>• Mit Hilfe der Ergebnisse aus Klassenaufgabe und Einzelaufgabe</li> <li>• Da es nur SuS aus einer Klasse waren, haben sie Kinder nicht so einen großen Pool gehabt</li> <li>• SuS von denen klar war, dass sie die Eingangsphase in 3 Jahren durchlaufen werden, wurden nicht mit in die Trainingsgruppe genommen zugunsten derer, die in der KA nicht ganz so auffällig waren, aber in EA und Unterricht</li> </ul>	<ul style="list-style-type: none"> <li>• Die SuS, die in KA und EA unter den CutOff fielen, sollen in die Trainingsgruppe</li> </ul>
202	<ul style="list-style-type: none"> <li>• Zum einen anhand des Ergebnisses des Meilensteins</li> <li>• Die eine Klasse hatte die LK selbst in Deutsch, sie kannte daher die Schwierigkeiten der Kinder</li> <li>• Im Unterricht hat sie Kinder erlebt, von denen sie dachte, dass sie unbedingt in die Trainingsgruppe müssten, die aber in den Aufgaben nicht unter den Cut off fielen</li> <li>• Mit der anderen LK hat sie sich immer besprochen, da war es ähnlich.</li> <li>• Die Anzahl von Kindern, die auffällig waren, passte immer ungefähr mit der Anzahl der Kinder, die in die Trainingsgruppe sollten. Manchmal mussten sie ein Kind dazu nehmen, damit es genug und eine grade Anzahl waren.</li> </ul>	<ul style="list-style-type: none"> <li>• Die Ergebnisse vom Meilenstein</li> <li>• Haben gezielt passende Aufgaben angesehen</li> <li>• Wenn man sich vor Augen führt, dass es um das flüssige Lesen geht, könnten auch andere Kinder in die Trainingsgruppe</li> </ul>
203	<ul style="list-style-type: none"> <li>• Mit zunehmender Erfahrung mit einzelnen Kindern floss auch das in die Entscheidung mit ein.</li> <li>• Es gab an der Schule auch noch andere Angebote; manchmal wurden SuS dann anstatt in die Lms-Trainingsgruppe in ein anderes Angebot geschickt.</li> </ul>	<ul style="list-style-type: none"> <li>• Vorgabe war: Ergebnisse der Diagnostik angucken und im Team beraten</li> </ul>

	<ul style="list-style-type: none"> <li>• Zum Teil waren es pädagogische Entscheidungen, z.B. ein Kind, das in der Gruppe war, aber aufgrund der großen Verhaltensauffälligkeiten nicht tragbar war oder eines, was nicht zu motivieren war und deshalb zu Gunsten eines motivierten Kindes nicht weitergemacht hat,</li> <li>• Grundlage für die Entscheidung waren die diagnostischen Ergebnisse (KA + EA), die Gruppengröße (6-8 Kinder) und pädagogische Gründe</li> </ul>	
204	<ul style="list-style-type: none"> <li>• Zwei unterschiedliche LK</li> <li>• In einer Klasse sehr genau nach Diagnostik</li> <li>• In der zweiten Klasse passte es nicht so ganz</li> <li>• Evtl. hat der Unterricht andere Ergebnisse gebracht als die Testung</li> <li>• Es könnte auch die Qualifikation der LK ein Punkt sein</li> <li>• Es wurde im Einzelfall entschieden. Wenn SuS auch nicht eine DaZ Förderung oder eine Soz.-emo. Gruppe hatten, haben sie die zum Teil nicht mit reingenommen, sondern sich für andere schwache SuS entschieden, die noch keine Förderung erhalten.</li> <li>• Bei SuS, bei denen schon früh klar wurde, dass sie wahrscheinlich 3 Jahre Eingangsphase haben, haben sie andere Kinder vorgezogen.</li> <li>• Sie haben die SuS genommen, wo sie eine Chance auf Erfolg gesehen haben.</li> </ul>	<ul style="list-style-type: none"> <li>• Die SuS, die <ul style="list-style-type: none"> <li>○ auffällig aus Meilenstein</li> <li>○ auffällig aus Beobachtungen im Unterricht</li> </ul> </li> <li>• sollten mit 6-8 Kindern eine homogene Gruppe bilden</li> </ul>
205	<ul style="list-style-type: none"> <li>• Hat in erster Linie organisatorische Gründe:</li> <li>• Das Training liegt im Anschluss an den Unterricht. Einigen Eltern war die Organisation zu groß, da die Kinder nicht den regulären Bus nehmen konnten, bzw. sie die Kinder dann gesondert abholen mussten. Diese Eltern wollten nicht, dass die Kinder teilnehmen, haben z.T. schriftlich widersprochen.</li> <li>• In einigen Fällen haben die Kinder ein unzuverlässiges Arbeitsverhalten gezeigt, sind z.B. nach dem Unterricht immer nach Hause entwischt anstatt zum Training zu kommen. Da</li> </ul>	<ul style="list-style-type: none"> <li>• Im Grunde: Hinweis auf die Diagnostik und individuell entscheiden, welche Möglichkeiten bei einzelnen Kindern gesehen wurden</li> <li>• Es soll geguckt werden, was passt</li> </ul>

	<p>haben wir uns für die SuS entschieden, mit denen eine verbindlichere Arbeit möglich war.</p> <ul style="list-style-type: none"> <li>• Es gab auch SuS, die durch ihr Verhalten das Training gesprengt, die Qualität für andere Kinder beeinträchtigt haben</li> <li>• Wir haben die Kinder ausgesucht: was praktikabel war. Chancen auf regelmäßige Teilnahme am größten, waren ggf. die, die laut ihren Werten nicht ganz so auffällig waren.</li> <li>• Bei einigen musste es auch abgebrochen werden</li> </ul>	
--	---	--

Tab. A2: Notizen Telefoninterview