

Formatives Assessment im Leseunterricht der Grundschule – Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts

Karin Hebbecke  · Elmar Souvignier

Online publiziert: 18. Juni 2018
© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

Zusammenfassung Die Komponenten Diagnostik und Feedback sowie die Ableitung von Fördermaßnahmen gelten als drei zentrale Bestandteile formativen Assessments. Es wurde untersucht, welche Effekte Materialien bewirken, die Lehrkräfte beim Einsatz von Lernverlaufsdagnostik (LVD) bei der Durchführung von Feedbackgesprächen und bei differenzierter Leseförderung unterstützen und wie gut sich solche Konzepte in dritten Klassen implementieren lassen. In einem Drei-Gruppen Design nutzten alle Lehrkräfte ($N=44$, $N=945$ Schüler) eine computerbasierte LVD. In zwei Bedingungen erhielten die Lehrkräfte zusätzliche Feedbackmaterialien (FB) sowie Feedback- und Fördermaterialien (FB+FM), um die Implementation der einzelnen Komponenten formativen Assessments zu unterstützen. Über ein Schuljahr wurden Implementationsfacetten sowie die Lesekompetenz und weitere Schülermerkmale erfasst. Während die Akzeptanz für die Materialien hoch ausfällt, wird deren Machbarkeit deutlich niedriger eingeschätzt. Im Vergleich mit der reinen LVD Bedingung erweisen sich die zusätzlichen Materialien nicht als wirksam. Diskutiert wird die Frage nach einer bestmöglichen Unterstützung von Lehrkräften beim Einsatz formativen Assessments.

Schlüsselwörter Formatives Assessment · Lernverlaufsdagnostik · Feedback · Individuelle Leseförderung · Lehrerfortbildung

K. Hebbecke (✉) · Prof. Dr. E. Souvignier (✉)
Institut für Psychologie in Bildung und Erziehung, Westfälische Wilhelms-Universität Münster,
Fliegenerstraße 21, 48149 Münster, Deutschland
E-Mail: karin.hebbecke@uni-muenster.de; elmar.souvignier@uni-muenster.de

Formative Assessment in third-grade reading instruction—Implementation and effectiveness of a modular concept including prepared teaching-material

Abstract Definitions of formative assessment include assessment, feedback and differentiated instruction as key components. We investigated the effects of prepared teaching materials designed to support teachers using learning progress assessment (LPA) to give feedback and adapt differentiated instruction. We also examined to what extent this modular approach can be implemented in regular reading lessons in third grade. In a three-group design all teachers ($N=44$, $N=945$ students) employed a computer-based LPA tool, while teachers in two conditions additionally received prepared feedback material (FB) or feedback and reading instruction material (FB+FM), to support the implementation of the different components of formative assessment. Over the course of one schoolyear, we assessed the implementation outcomes using questionnaires as well as students' reading achievement and further student outcomes. While acceptability is high, teacher ratings of feasibility are low. In comparison to the LPA group, the additional support in form of prepared materials had no effects on student outcomes. Results are discussed regarding the question of how teachers can optimally be supported in using formative assessment.

Keywords Feedback · Formative assessment · Individual reading instruction · Learning progress measurement · Teacher training

1 Einleitung

Die Grundidee des formativen Assessments – Erhebung und Nutzung diagnostischer Information zur Optimierung von Lehr-Lernsituationen – beschreibt zunächst allgemein das Konzept, Diagnostik explizit in den Dienst von Förderung zu stellen (Black und Wiliam 1998; Hattie und Timperley 2007). Die Erhebung von Schülerleistungen, die Nutzung dieser diagnostischen Informationen für individuelles Feedback und die Ableitung von Fördermaßnahmen sind entsprechend die drei zentralen inhaltlichen Komponenten, die sich in Definitionen formativen Assessments finden (Black und Wiliam 1998; Wiliam und Thompson 2008; Cizek 2010; Maier 2010). In welchem Maße diese drei Komponenten formalisiert sein sollten, wird allerdings nicht einheitlich gehandhabt. So bezeichnen beispielsweise Heritage (2007) sowie Shavelson et al. (2008) neben einem formalisierten Vorgehen, bei dem Feedback auf der Basis von Testverfahren gegeben wird („curriculum-embedded assessment“) auch Varianten adaptiven Unterrichts als formatives Assessment, bei denen spontan („on the fly“) oder nach gezielt eingesetzten Fragen („planned for interaction“) Feedback gegeben wird, ohne Schülerleistungen explizit zu messen. Im Unterschied zu einer so breiten Anwendung des Begriffs fordern Dunn und Mulvenon (2009), nur dann von formativem Assessment zu sprechen, wenn zunächst eine Leistungsbeurteilung (Assessment) vorgenommen wird, die dann (formativ) mit dem Ziel einer Verbesserung des Unterrichts oder von Fördermaßnahmen eingesetzt wird. Auch Kingston und Nash (2011) kritisieren, dass von formativem Assessment (oder *Assessment for*

learning) häufig bereits dann die Rede sei, wenn nur einzelne Komponenten dieses Konzepts verwendet werden.

Wenngleich der Ansatz formativen Assessments im angelsächsischen Sprachraum seit den 1990er Jahren und aktuell in der deutschsprachigen Literatur zunehmende Beachtung findet, liegen nur wenige ausgearbeitete Konzepte zur schulpraktischen Umsetzung, Studien zur Implementation und Befunde zur Wirksamkeit des Ansatzes formativen Assessments vor (Maier 2010; Bennett 2011). Eine gelingende Implementation wurde dabei als bedeutender Faktor hinsichtlich der Wirksamkeit von innovativen Konzepten identifiziert (Schneider und Randel 2010; Spörer und Glaser 2010; Bennett 2011). Vor diesem Hintergrund soll in der vorliegenden Studie ein konkretes Konzept für die Umsetzung von formativem Assessment im Sinne des formalisierten Ansatzes des curriculum-embedded assessment im Bereich Lesen in der Grundschule erprobt und hinsichtlich der Implementation und Wirksamkeit evaluiert werden. In einem modularen Ansatz werden Lehrkräften dabei – in je unterschiedlichem Umfang – unterstützende Materialien zu Diagnostik, Feedback und Förderung zur Verfügung gestellt: Während alle Lehrkräfte einen computerbasierten Ansatz zur Lernverlaufsdiagnostik einsetzen können, der sich in vorangehenden Studien bereits als lernwirksam erwiesen hat (Förster und Souvignier 2014, 2015), bekommen einige Lehrkräfte zusätzlich unterstützende Materialien zur Durchführung von Feedbackgesprächen. In einer dritten Bedingung wird ein „Materialpaket“ aus Lernverlaufsdiagnostik, Feedbackkonzept und Fördermaterialien angeboten. Ziel der Studie ist es zu untersuchen, wie gut sich diese Konzepte in den Unterricht implementieren lassen und welche Effekte eine Unterstützung von Lehrkräften beim Einsatz von Lernverlaufsdiagnostik (in Form von Feedback- und Unterrichtsmaterialien sowie Fortbildungen) auf die Schülerleistung und auf motivationale Schülermerkmale hat.

Im Folgenden wird zunächst die Bedeutung einer gelungenen Implementation für die Wirksamkeit von Fördermaßnahmen herausgestellt. Im Anschluss werden Befunde zur Wirksamkeit formativen Assessments zusammengefasst und abschließend evidenzbasierte Ansätze zur Gestaltung formativen Assessments (Diagnostik, Feedback und Förderung) in der Domäne des Lesens beschrieben.

1.1 Implementation formativen Assessments

In Deutschland gibt es bislang sehr wenige konkrete, evaluierte Ausarbeitungen zur Umsetzung von formativem Assessment im Schulalltag (Maier 2010, 2014). Mit Blick auf die Aus- und Fortbildung von Lehrkräften lässt sich zudem feststellen, dass Ausbildungsprogramme nur selten auf den Einsatz von formativem Assessment vorbereiten und nur wenige Kompetenzen für den Umgang mit Assessment-Informationen vermitteln (Popham 2009; Cizek 2010; Stiggins 2010; Zeuch et al. 2017). Vor diesem Hintergrund besteht ein hoher Bedarf an Fortbildungen und Maßnahmen der Unterrichtsentwicklung zur Implementation von Konzepten formativen Assessments (Bürgermeister et al. 2014).

Als zentraler Aspekt für eine gelingende Implementation wird meist die Wiedergabetreue benannt (O'Donnell 2008). Der Begriff Implementation ist jedoch weiter zu fassen und meint die Umsetzung wissenschaftlicher Erkenntnisse in die Praxis (Euler und Sloane 1998; Hasselhorn et al. 2014). Dabei sind insbesondere die

Facetten Akzeptanz, Umsetzbarkeit, Nutzung, Wiedergabetreue und Nachhaltigkeit bedeutsam (Petermann 2004; Proctor et al. 2011). Akzeptanz bezieht sich darauf, wie zufriedenstellend eine Maßnahme im Hinblick auf unterschiedliche Aspekte (z. B. Inhalt, Komplexität) in der Wahrnehmung der Beteiligten ist. Mit Umsetzbarkeit ist gemeint, inwiefern eine Maßnahme erfolgreich umgesetzt werden kann und sie sich für den jeweiligen Kontext und die entsprechenden Rahmenbedingungen eignet. Nutzung meint die tatsächliche Umsetzung einer Maßnahme. Die Wiedergabetreue beinhaltet, ob eine Maßnahme so umgesetzt wird, wie intendiert. Die Integrität und Qualität der Umsetzung stehen im Fokus. Der Aspekt der Nachhaltigkeit bezieht sich auf den dauerhaften Einsatz einer Maßnahme und die Frage nach dem Ausmaß einer Institutionalisierung oder der Entwicklung von Handlungsrou-tinen (Petermann 2004). Für eine nachhaltige Übernahme neuer Konzepte in eigene unterrichtliche Handlungsrou-tinen ist es förderlich, wenn Lehrkräfte positive praktische Erfahrungen im Umgang mit neuen Unterrichtskonzepten sammeln (Clarke und Hollingsworth 2002).

Um die Implementation eines ausgearbeiteten Konzepts formativen Assessments zu unterstützen, erscheint es hilfreich, Lehrkräften konkrete Materialien und Fortbildungsangebote zur Verfügung zu stellen (Stecker et al. 2005; Bennett 2011). Fortbildungsangebote adressieren dabei vorrangig die Facette der Akzeptanz, während die Bereitstellung ausgearbeiteter Unterrichtsmaterialien die Umsetzbarkeit sowie die Wiedergabetreue erhöhen (Souvignier und Philipp 2016). In einer Studie von Hon-drich et al. (2016) setzten Lehrkräfte häufiger Elemente formativen Assessments im Unterricht um, wenn sie eine Fortbildung zu diesem Themenbereich besucht hatten. Zudem erwies sich die zusätzliche Bereitstellung ausgearbeiteter Materialien als wirksam in Bezug auf die Häufigkeit und die Qualität der Umsetzung von Elementen formativen Assessments. Auch die Evaluation der Materialien durch die Lehrkräfte (Akzeptanz) hing positiv mit der Qualität und der Umsetzung zusammen. Diese Ergebnisse heben die Bedeutung von Fortbildungen, ausgearbeiteten Materialien sowie der Akzeptanz der Lehrkräfte für die Implementation formativen Assessments hervor (Bennett 2011).

1.2 Wirksamkeit formativen Assessments

Generell besteht ein breiter Konsens im Hinblick auf das hohe Potenzial des Ansatzes formativen Assessments (Black und Wiliam 2009). Vor dem Hintergrund der unterschiedlichen Umsetzungen der Grundidee formativen Assessments – es gibt unterschiedliche Akzentuierungen beim Standardisierungsgrad der diagnostischen Komponente, bei der Umsetzung von Feedbackkonzepten sowie bei der Verbindlichkeit von Förderempfehlungen – erscheint die Frage nach *der* Wirksamkeit formativen Assessments nur bedingt sinnvoll. Kingston und Nash (2011), die eine Metaanalyse von Studien vornahmen, in denen die Autoren explizit die Bezeichnungen der *formativen* Nutzung von Assessment oder des *Assessment for learning* nutzen, berichten beispielsweise eine starke Domänenspezifität der Befunde mit höheren Effekten bei sprachlichen und niedrigeren Effekten bei naturwissenschaftlichen Fächern. Eine eher ernüchternde Feststellung im Rahmen dieser Metaanalyse lag darin, dass aus 300 Untersuchungen lediglich 13 Studien methodische Standards

angewendet hatten (Kontrollgruppe; Bericht von Effektstärken), die eine vergleichende Untersuchung der Wirksamkeit erlauben. Im Mittel berichten Kingston und Nash (2011) eine Effektstärke von $d=0,20$. Da viele Studien die Wirksamkeit von formativem Assessment bei leistungsschwachen Schülerinnen und Schülern in den Fokus nahmen, ist eine Generalisierbarkeit der Befunde allerdings fraglich.

Jenseits der Empfehlungen, in künftigen Studien zu formativem Assessment immer eine explizite Beschreibung der konkreten Vorgehensweise vorzunehmen und Studien mit Regelschülerinnen und Regelschülern durchzuführen, weisen Kingston und Nash (2011) auf die Notwendigkeit hin, vor allem Faktoren zu untersuchen, die die Wirksamkeit von formativem Assessment beeinflussen können. In diesem Kontext ist auch von großem Interesse, wie Lehrkräfte bei der Umsetzung von formativem Unterricht im Schulalltag bestmöglich unterstützt werden können (Topping und Fisher 2003; Wiliam 2010). Forschungsbedarf wird also sowohl hinsichtlich der Implementation formativen Assessments in die Schulpraxis als auch hinsichtlich der konkreten Ausgestaltung von formativem Assessment in einzelnen Domänen gesehen (Bennett 2011; Kingston und Nash 2011). Bennett (2011) betont, dass formatives Assessment durch Lehrkräfte insbesondere dann effektiv in einzelnen Domänen umgesetzt werden kann, wenn über generelle Prinzipien und Strategien formativen Assessments hinaus auch domänenspezifisches Wissen und Hilfsmittel (z. B. Wie kann ich welche Fertigkeiten bei meinen Schülern und Schülerinnen erheben? Welche konkreten Feedback- und Fördermaßnahmen sind aus welchen Informationen abzuleiten?) an Lehrkräfte vermittelt werden.

1.3 Formatives Assessment im Bereich Lesen

Vor dem Hintergrund der Empfehlung, Studien zu formativem Assessment innerhalb einzelner Domänen zu betrachten (Bennett 2011; Kingston und Nash 2011), wurde im Rahmen der vorliegenden Studie theoriebasiert ein Konzept zur diagnosebasierten individuellen Leseförderung erstellt. Dabei stellt Lesen eine grundlegende Kompetenz mit fächerübergreifender Bedeutung dar (Hußmann et al. 2017). In der Folge soll für den vergleichsweise formalisierten Ansatz des curriculum-embedded assessment konkretisiert werden, wie die drei Komponenten Diagnostik, Feedback und Förderung in der Domäne des Lesens in der Grundschule realisiert werden können. Zu jedem der drei „Bausteine“ formativen Assessments werden nun evidenzbasierte Ansätze skizziert, die eine Grundlage für ein formalisiertes Angebot darstellen können.

Diagnostik Das Konzept der Lernverlaufdiagnostik – in kurzem zeitlichem Abstand werden äquivalente Tests durchgeführt, anhand derer sich Effekte von Förderangeboten abbilden lassen – weist eine unmittelbare Entsprechung zu dem Ansatz formativen Assessments auf. Im sonderpädagogischen Bereich hat sich insbesondere der Ansatz curriculumbasierten Messens (CBM) (Deno 1985) als eine praktikable und wirksame Möglichkeit erwiesen, diagnostische Information für instruktionale Entscheidungen einzusetzen (Stecker et al. 2005). Beim CBM-Ansatz werden beim Lesen vor allem Maße für die Leseflüssigkeit (eine Minute lautes Lesen; Lückentexte) als „robuste“ Indikatoren für Lesekompetenz eingesetzt. Mit Blick auf die

Lehrziele in der dritten und vierten Jahrgangsstufe erscheint es sinnvoll, neben der Leseflüssigkeit auch das Leseverständnis zu erfassen (vgl. Lenhard 2013). Für die Anwendung formativer Diagnostik im Schulalltag ist ein geringer zeitlicher Aufwand für die Durchführung, Auswertung und Dokumentation ein wesentliches Kriterium. Aus diesem Grund liegt der Rückgriff auf computergestützte Verfahren nahe, da diese Prozesse somit automatisiert ablaufen können (Fuchs 2004). Ein in der schulischen Praxis erprobtes und evaluiertes internetbasiertes Instrument zur Lernverlaufsdiagnostik ist die Lernverlaufsdiagnostik mit dem System quop (Souvignier et al. 2014), das sich in mehreren Evaluationsstudien als wirksam und praktikabel erwies (Förster und Souvignier 2014, 2015). Über das gesamte Schuljahr hinweg werden mehr als 90 % der bereitgestellten Computertests absolviert. Lehrerbefragungen zur Praktikabilität weisen eine Zustimmung > 80 % aus und auch in Schülerbefragungen zeigt sich ein hohes Maß an Zufriedenheit (Souvignier et al. 2014; Förster und Souvignier 2015). In Klassen, in denen quop eingesetzt wurde, zeigten sich höhere Lernzuwächse, als in Klassen, in denen die Lehrkräfte quop nicht nutzten. Wengleich diagnostische Information zu Lernverläufen ein wirksamer Ansatz zur Unterstützung von Lehrkräften ist, so zeigte sich in einer Übersichtsarbeit von Stecker et al. (2005), dass es hilfreich sein könnte, diagnostische Informationen unmittelbar mit Entscheidungshilfen zu verbinden. Auch Espin et al. (2017) weisen darauf hin, dass die Bereitstellung diagnostischer Informationen zu Lernverläufen nur dann zu höheren Lernzuwächsen führt, wenn diese für die Gestaltung entsprechender Lernangebote genutzt wird. Im Sinne des Ansatzes formativen Assessments sollte diagnostische Information als Feedback für Schülerinnen und Schüler und zur Entscheidung für passende Fördermaßnahmen genutzt werden.

Feedback Obwohl Feedback ein zentraler Bestandteil formativen Assessments ist (Black und Wiliam 1998; Wiliam und Thompson 2008), liefert die Assessmentforschung bisher nur bedingt Hinweise darauf, wie die konkrete Ausgestaltung lernförderlichen Feedbacks aussehen sollte, sodass hier häufig auf die Befunde der Feedbackforschung zurückgegriffen wird. Inwiefern diese Erkenntnisse, die häufig aus experimentellen Settings stammen und jeweils einzelne Rückmeldemerkmale untersuchen, auf Feedback als Komponente formativen Assessments übertragen werden können, ist bislang allerdings nur bedingt geklärt (Harks 2013). In der Feedbackforschung finden sich Hinweise darauf, dass Feedback neben positiven Effekten auf Leistung (z. B. Bangert-Drowns et al. 1991; Kluger und DeNisi 1996; Narciss und Huth 2004; Hattie und Timperley 2007; Shute 2008) auch förderlich für die Motivation (Krampen 1987; Dresel und Haugwitz 2008; Rakoczy et al. 2008; Harks et al. 2014), die Metakognition sowie die Fähigkeit zur Selbstregulation (Black und Wiliam 2006; Clark 2012) und die Wahrnehmung des Unterrichts durch die Schülerinnen und Schüler (Rakoczy et al. 2013; Hondrich et al. 2016) sein kann. Inzwischen liegt eine Reihe an Überblicksarbeiten und Metaanalysen vor, die jedoch verdeutlichen, dass Feedback nicht unter allen Bedingungen eine lernförderliche Wirkung hat und Hinweise auf eine förderliche Ausgestaltung von Feedbackprozessen geben (z. B. Kluger und DeNisi 1996; Hattie und Timperley 2007; Shute 2008). Feedback sollte demnach ausreichend elaboriert gestaltet sein (Narciss und Huth 2004) und so dabei helfen, Diskrepanzen zwischen dem Lernziel und dem aktuellen Lernstand

zu reduzieren (Hattie und Timperley 2007). Es sollte sich auf aktuelle Stärken und Schwächen beziehen („How am I going?“), mögliche Ziele ansprechen („Where am I going?“) und auch Möglichkeiten anbieten, dieses Ziel zu erreichen („Where to next?“) (Hattie und Timperley 2007). Darüber hinaus erscheint es hilfreich, das Feedback auf die Aufgabenleistung, Verstehens- oder Regulationsprozesse zu beziehen und Feedback auf der Ebene des Selbst (zur Person als Ganzes und ohne Bezug zu diesen Prozessen, z. B. „Du bist ein guter Schüler“) zu vermeiden (Kluger und DeNisi 1996; Hattie und Timperley 2007).

Für die Gestaltung von Feedback im Kontext formativen Assessments lässt sich aus dieser Befundlage ableiten, dass diagnostische Informationen so aufbereitet werden sollten, dass eine sachliche Feststellung von Lernstand und Entwicklung vorgenommen wird und daraus abzuleitende Ziele sowie konkrete Fördermaßnahmen besprochen werden. Eine solche strukturierte Vorgehensweise macht deutlich, dass die diagnostische Information sich zunächst an die Lehrenden richtet, deren Aufgabe darin besteht, das Feedback an die Schülerinnen und Schüler vorzubereiten und ihrerseits Entscheidungen im Hinblick auf geeignete Ziele und Fördermaßnahmen zu treffen (vgl. Espin et al. 2017).

Förderung Die Wirksamkeit von Fördermaßnahmen ist dann am höchsten, wenn sie zu den Lernvoraussetzungen der Schülerinnen und Schüler passt (Connor et al. 2004). Fördermaßnahmen sollten folglich an den individuellen Leistungsstand der Schülerinnen und Schüler angepasst werden. Die Entwicklung von Lesekompetenzen im Grundschulalter umfasst ein breites Spektrum von basalen Kompetenzen (z. B. Buchstaben- und Worterkennung sowie die Automatisierung dieser Prozesse) bis hin zu hierarchiehöheren Prozessen (satzübergreifender Aufbau einer kohärenten Textstruktur) (Richter und Christmann 2002). Die Lesegenauigkeit und die Leseflüssigkeit können als Vorläuferfertigkeiten für das Leseverstehen gesehen werden (Lenhard 2013). Durch die steigende Automatisierung dieser Prozesse werden zunehmend kognitive Ressourcen frei, die für das verstehende Lesen genutzt werden können. Durch eine Diagnostik des Lernstandes kann abgeleitet werden, auf welcher Ebene Fördermaßnahmen ansetzen sollten. Übersichtsarbeiten und Metaanalysen kommen zu dem Schluss, dass Lautleseverfahren zur Förderung der Leseflüssigkeit und lesestrategieorientierte Verfahren zur Förderung des Leseverständnisses eine hohe Wirksamkeit aufweisen (NICHD 2000; Slavin et al. 2009). Zudem erwiesen sich ein hohes Maß an Schüleraktivität und kooperative Lernformen als effektiv (Slavin et al. 2009). Für die Förderung der Lesegenauigkeit hat sich silbenbasiertes Lesen als effektiv erwiesen (z. B. Müller und Richter 2017; Müller et al. 2017). Neben konsistenten Befunden zur Leistungssteigerung gibt es auch Hinweise darauf, dass der Einsatz der Methoden Repeated Reading (Samuels 1979) und Reziprokes Lehren (Palincsar und Brown 1984) positive Effekte auf das Selbstkonzept der Schülerinnen und Schüler haben kann (Rosebrock et al. 2010; Koch und Spörer 2016).

1.4 Vorliegende Studie

Die Lernzuwächse von Schülerinnen und Schülern, deren Lehrkräfte diagnostische Informationen über deren Lernverlauf zur Verfügung haben, fallen höher aus als die Lernzuwächse von Schülerinnen und Schülern, deren Lehrkräfte diese Informationen nicht haben (Stecker et al. 2005; Förster und Souvignier 2014, 2015). Darüberhinausgehend (und unter Berücksichtigung aller Komponenten Diagnostik, Feedback, Förderung) gibt es eher wenige Befunde zur Implementation und Wirksamkeit formativen Assessments, nicht zuletzt vor dem Hintergrund uneinheitlicher Definitionen und der bisher nicht konkreten Ausgestaltung von formativem Assessment in einzelnen Domänen (Dunn und Mulvenon 2009; Maier 2010; Bennett 2011; Kingston und Nash 2011). Zwar gibt es aus der Feedback- und Interventionsforschung konsistente Hinweise auf wirksame evidenzbasierte Prinzipien in diesen Bereichen, es besteht aber ein Mangel an der Evaluation von Konzepten zur expliziten Verknüpfung der Komponenten Diagnostik, Feedback und Förderung in der Assessmentforschung. Mit Blick auf den Anspruch der Implementation eines so umfassenden theoretischen Ansatzes stellt sich dabei die Frage, wie Lehrkräfte beim Einsatz formativen Assessments unterstützt werden können (William 2010; Kingston und Nash 2011; Souvignier und Philipp 2016). Zudem wird Forschungsbedarf insbesondere im Bereich von computerbasierten Systemen in Bezug auf die pädagogischen Konsequenzen für die Nutzung solcher Programme formuliert (Topping und Fisher 2003; Maier 2014).

Im Rahmen der vorliegenden Studie wurde theoriegeleitet und orientiert an empirischen Befunden ein strukturiertes Konzept formativen Assessments für den Bereich Lesen in der Grundschule entwickelt. Dabei wurde ein modulares Prinzip für die drei Komponenten Diagnostik, Feedback und Förderung zu Grunde gelegt und eine enge Orientierung an dem formalisierten Konzept des curriculum-embedded assessments vorgenommen. Es wurden Materialien vorbereitet und Lehrerfortbildungen angeboten, um eine gelungene Implementation des erarbeiteten Konzepts zu unterstützen.

1.5 Fragestellung

Das erste Ziel der Studie war es zu prüfen, in welchem Maße die Implementation des domänenspezifischen Konzepts formativen Assessments gelingt. Darauf aufbauend sollte untersucht werden, welche zusätzlichen Effekte die Unterstützung der Lehrkräfte bei der Nutzung der diagnostischen Informationen durch ausgearbeitete Feedback- und Fördermaterialien – über die Bereitstellung eines Systems zur Lernverlaufsdagnostik hinaus – hat. Da Forschungsbefunde nicht nur auf positive Effekte auf Leistungsvariablen sondern auch auf motivationale Variablen, Selbstregulationsprozesse und die Wahrnehmung des Unterrichts durch die Schülerinnen und Schüler hinweisen, sollten auch diese in der vorliegenden Studie untersucht werden. Die Fragestellungen lauteten:

- **Implementation:** (In welchem Maße) Gelingt die materialgestützte Implementation der Lernverlaufsdagnostik mit dem internetbasierten System quop, eines Feedbackkonzepts sowie einer differenzierten Leseförderung in den Unterricht im

Hinblick auf die Kriterien Nutzung, Akzeptanz, Umsetzbarkeit, Wiedergabetreue und Nachhaltigkeit?

- **Wirksamkeit:** Welche Effekte hat die Nutzung (a) der Feedbackmaterialien und (b) der Feedback- und Fördermaterialien durch Lehrkräfte sowie eine Unterstützung durch Lehrerfortbildungen zusätzlich zum Einsatz der Lernverlaufsdagnostik mit quop auf die Lesekompetenz sowie die Lesemotivation, das Leseselbstkonzept, das Leseverhalten, die wahrgenommene Unterstützung und den Einsatz von metakognitiven und kognitiven Lesestrategien der Schülerinnen und Schüler?

Erwartet wurden positive Effekte der zusätzlichen Unterstützung der Lehrkräfte auf die abhängigen Variablen in der Bedingung FB im Vergleich zur Bedingung LVD (LVD<FB) und ebenfalls positive Effekte in der Bedingung FB+FM im Vergleich zur Bedingung LVD (LVD<FB+FM).

2 Methode

2.1 Design und Stichprobe

In einem quasi-experimentellen Drei-Gruppen-Prätest-Posttest-Design wurden die Effekte der Unterstützung von Lehrkräften durch Unterrichtsmaterialien und Lehrerfortbildungen zusätzlich zum Einsatz einer computerbasierten Lernverlaufsdagnostik über ein Schuljahr hinweg untersucht (s. Abb. 1). Die Lehrkräfte aller Bedingungen setzten die computerbasierte Lernverlaufsdagnostik mit quop (Souvignier et al. 2014) ein. Lehrkräfte der Bedingung „Feedback“ (FB) erhielten zusätzlich ausgearbeitete Materialien, die die Interpretation der Lernverläufe und die Nutzung der diagnostischen Informationen für Feedbackgespräche unterstützen. Auch die Lehrkräfte der Bedingung „Feedback + Fördermaterial“ (FB+FM) erhielten diese Feedbackmaterialien sowie zusätzliche Unterrichtsmaterialien zur individuellen Leseförderung. Die zur Verfügung gestellten Materialien sowie theoretische Hintergründe mit Bezug zur praktischen Anwendung wurden den Lehrkräften in einer Materialeinführung zu Beginn des Schuljahres vorgestellt. Die Lehrkräfte der Bedingung „LVD“ erhielten über die Lernverlaufsdagnostik hinaus keine unterstützenden Materialien und keine Einführungsveranstaltung. In allen Klassen wurden nach den Herbstferien (Oktober/

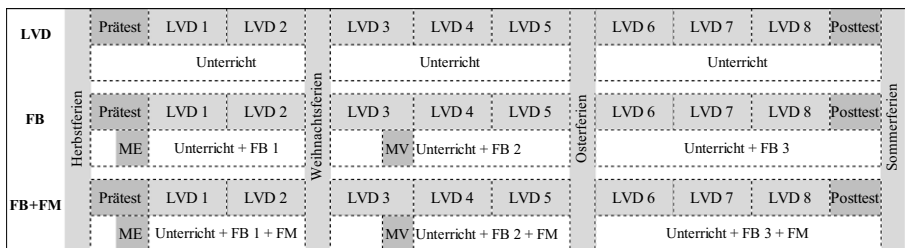


Abb. 1 Design und Ablauf der Studie für die drei Untersuchungsbedingungen „LVD“, „FB“ und „FB+FM“. ME Materialeinführung, MV Materialvertiefung, FB Feedback, FM Fördermaterial, LVD Lernverlaufsdagnostik

November 2015) und kurz vor den Sommerferien (Juni 2016) die Lesekompetenz, die Lesemotivation, das Leseselbstkonzept, das Leseverhalten, die wahrgenommene Unterstützung sowie der Einsatz metakognitiver und kognitiver Strategien beim Lesen von Texten erfasst.

Insgesamt nahmen $N=945$ Schülerinnen und Schüler der dritten Klassenstufe aus 21 Grundschulen in Nordrhein-Westfalen an der Studie teil. Das durchschnittliche Alter betrug zum Prätestzeitpunkt $M=8,67$ ($SD=0,50$). 48,0 % waren männlich, 68,8 % sprachen nach Angaben der Schülerinnen und Schüler selbst zu Hause deutsch, 22,5 % deutsch und eine andere Sprache und 8,7 % nur eine andere Sprache. Hinsichtlich der demografischen Variablen lagen keine signifikanten Unterschiede zwischen den Untersuchungsbedingungen vor. Die Teilnahme an den Schülertestungen war freiwillig und die Eltern wurden zuvor schriftlich um ihr Einverständnis gebeten. Die Daten von $N=12$ Schülerinnen und Schülern, deren Eltern der Verwendung der Daten für die Auswertung nicht zustimmten, wurden in den Analysen nicht berücksichtigt. Das durchschnittliche Alter der Lehrkräfte ($N=44$) betrug zum Prätestzeitpunkt $M=49,2$ ($SD=9,5$) Jahre. 13,6 % der Lehrkräfte waren männlich und die durchschnittliche Lehrerfahrung lag bei $M=21,0$ ($SD=10,3$) Jahren. Hinsichtlich der demografischen Variablen lagen keine signifikanten Unterschiede zwischen den Untersuchungsbedingungen vor.

Da ein erster Implementation-Check ergab, dass insgesamt $N=13$ Lehrkräfte der Untersuchungsbedingungen FB und FB+FM die erhaltenen Materialien nicht im Unterricht nutzten, musste für die folgenden Analysen eine Anpassung der Stichprobe vorgenommen werden. In die endgültigen Analysen wurden somit die Daten von $N=31$ Lehrkräften und $N=644$ Schülerinnen und Schülern einbezogen (eine genaue Beschreibung dieser Anpassung folgt in Abschn. 2.3).

2.2 Ablauf

Die Schulleitungen wurden per E-Mail und telefonisch über die Möglichkeit zur Teilnahme an der Studie informiert. Bei Interesse an einer Teilnahme wurden die Klassen (auf Schulebene) zufällig einer der drei Bedingungen zugewiesen und dann je nach Bedingung ausführlicher informiert. Da Schulen sich vereinzelt trotz einer ersten Zusage im Verlauf des Rekrutierungsprozesses gegen eine Teilnahme entschieden, ergab sich die folgende Verteilung über die Bedingungen: LVD: $N=336$ Kinder aus 16 Klassen und 8 Schulen, FB: $N=296$ Kinder aus 13 Klassen und 6 Schulen; FB+FM: $N=313$ Kinder aus 15 Klassen und 7 Schulen. Die Erhebungen wurden als Papier-Bleistift Tests von geschulten Testleitern in den Schulklassen durchgeführt.

Lernverlaufsdiagnostik Die Lehrkräfte aller drei Bedingungen setzten in ihrem Unterricht die computerbasierte Lernverlaufsdiagnostik quop ein (Souvignier et al. 2014). Dabei bearbeiteten die Schülerinnen und Schüler während der Unterrichtszeit (z. B. in Stillarbeitsphasen) zwischen den Herbst- und Sommerferien acht kurze, parallele Lesetests im Umfang von ca. 10 Minuten am Computer. Der Abstand zwischen den Tests betrug jeweils ca. drei Wochen. In jedem Test wurden die Lesegenauigkeit, Lesegeschwindigkeit und das Leseverständnis gemessen. Durch die automatisierte

Auswertung standen die Ergebnisse sowohl den Lehrkräften als auch den Schülerinnen und Schülern unmittelbar nach Bearbeitung der Tests zur Verfügung. Somit erhielten die Lehrkräfte aller Bedingungen diagnostische Informationen über die Lernverläufe ihrer Schülerinnen und Schüler. (Für eine ausführlichere Beschreibung der Lernverlaufsdagnostik mit quop s. Souvignier et al. 2014). Die Lehrkräfte in der Bedingung LVD erhielten über den Zugang zur Lernverlaufsdagnostik mit quop keine weitere Unterstützung oder Hinweise auf die Nutzung der diagnostischen Informationen für Feedback und individuelle Förderung und auch keine Fortbildung.

Feedback Die Lehrkräfte der beiden Bedingungen FB und FB+FM hatten die Möglichkeit, im quop-Lehrermenü einen individuellen Dokumentationsbogen für jede Schülerin und jeden Schüler auszudrucken. Dieser diente einerseits als Interpretationshilfe und andererseits als Vorbereitung für Feedbackgespräche mit den Schülerinnen und Schülern. Basierend auf den quop-Ergebnissen wurden auf diesem Bogen individuelle Stärken und Schwächen, mögliche Ziele im Lesen und Übungsmöglichkeiten beschrieben. Zudem waren die erreichten Punktzahlen in Tabellenform und als Grafiken dargestellt. Die Lehrkräfte erhielten ferner Hinweise für die Zusammenstellung homogener Leistungsgruppen. Als Orientierung für die Feedbackgespräche wurde ein Gesprächsleitfaden angeboten, in dem die drei Aspekte 1. Rückmeldung von Stärken und Schwächen, 2. Besprechen von individuellen Zielen im Lesen und 3. Übungsmöglichkeiten, um diese Ziele zu erreichen, adressiert wurden. Des Weiteren wurden in den Leitfäden Hinweise auf wesentliche Feedback-Prinzipien gegeben (z.B. Verwendung einer individuellen Bezugsnorm, Unterstützung angemessener Attributionsmuster). Jede Schülerin und jeder Schüler erhielt ein quop-Heft, in dem in vorstrukturierten Protokollbögen die wichtigsten Ergebnisse der Feedbackgespräche dokumentiert werden sollten.

Fördermaterial Lehrkräfte in der Bedingung FB+FM erhielten zusätzlich zu den Feedbackmaterialien ausgearbeitete Unterrichtsmaterialien zur individuellen Leseförderung. Dabei handelte es sich um eine überarbeitete und erweiterte Version des „Lese-Sportler Materials“ (Kawohl 2015), das drei Übungsmethoden auf jeweils drei Schwierigkeitsstufen enthält: Je eine Methode bezieht sich auf die Förderung der Lesegenauigkeit, der Leseflüssigkeit und des Leseverstehens. Da der Lehrkraft durch die Lernverlaufsdagnostik diagnostische Informationen über genau diese Fertigkeiten vorlagen, konnte so eine passgenaue Fördermaßnahme für jedes Kind abgeleitet und entsprechend des Leistungsstandes entschieden werden, wann ein Wechsel zwischen den Fördermethoden angemessen erscheint. Bei jeder der drei Übungsmethoden ist eine Partnerarbeit in leistungshomogenen Zweiertteams vorgesehen. Eingebettet in die Rahmengeschichte des Lese-Sportlers übernimmt bei jeder Übungsmethode abwechselnd jeweils ein Kind die Rolle des „Trainers“ (Anleitung) und ein Kind die Rolle des „Sportlers“ (Üben).

Die Methode des Lese-Slaloms dient der Förderung der Lesegenauigkeit durch das Einzeichnen von Silbenbögen sowie silbenbasiertes, wiederholendes lautes Lesen (Rosebrock et al. 2011; Müller und Richter 2017; Müller et al. 2017). Der Sportler hat die Aufgabe, bei den in einem Kasten vorgegebenen Wörtern bzw. Sätzen Silbenbögen einzuzeichnen und sie im Anschluss wiederholt vorzulesen, bis dies

fehlerfrei gelingt. Der Trainer kontrolliert das korrekte Einzeichnen der Silbenbögen und das Vorlesen. Nach jedem Kasten werden die Rollen getauscht. Von Kasten zu Kasten steigert sich die Anzahl der Wörter und Sätze schrittweise.

Die Methode des Lese-Sprinters setzt das Repeated Reading (Samuels 1979; Rosebrock et al. 2011) um und hat die Förderung der Leseflüssigkeit zum Ziel. Der Sportler liest einen vorgegebenen Text eine Minute lang laut vor. Der Trainer unterstreicht während des Lesens die Lesefehler und notiert die Zahl korrekt gelesener Wörter. In einem zweiten Durchgang liest der Sportler denselben Text erneut eine Minute lang. Auch hier werden die Rollen getauscht, sodass jedes Kind die Rolle des Sportlers und die Rolle des Trainers übernimmt.

Beim Lesetandem wird auf die Methode des Reziproken Lehrens (Palincsar und Brown 1984) zurückgegriffen, um das Leseverständnis durch die Anleitung zur Nutzung von Lesestrategien zu fördern. Zunächst führt die Lehrkraft nacheinander vier Lesestrategien (Überschrift beachten, schwierige Wörter klären, Wichtiges unterstreichen und Zusammenfassen) ein und stellt sicher, dass die Schülerinnen und Schüler den Ablauf und den Nutzen der Strategien verstanden haben. Im Anschluss sollen die Schülerinnen und Schüler die Anwendung dieser Lesestrategien auf einen Text anleiten und überwachen (Trainer) oder ausführen (Sportler). Dazu liest der Sportler einzelne Abschnitte eines Textes vor und wendet – vom Trainer angeleitet – Strategien auf die Textabschnitte an. Sowohl Trainer als auch Sportler werden dabei durch rollenspezifische Ablaufkärtchen unterstützt.

Auf der Basis der Rückmeldungen zu den Papier-Bleistift-Tests, der quop-Ergebnisse sowie der Empfehlung im Dokumentationsbogen entschieden die Lehrkräfte, welche Kinder mit welchen Materialien und welcher Schwierigkeitsstufe arbeiten sollten, um die Förderung so an die individuellen Bedürfnisse der Schülerinnen und Schüler anzupassen. Im Verlauf des Schuljahres konnten die Paare und die Methoden für die einzelnen Schülerinnen und Schüler je nach Lernverlauf immer wieder neu zusammengestellt werden. Auch die Feedbackgespräche sollten dazu genutzt werden, die Übungsmethode und den Übungspartner für die einzelnen Schülerinnen und Schüler zu besprechen.

Lehrerfortbildungen Die Lehrkräfte der Bedingungen FB und FB+FM nahmen an jeweils einer Materialeinführung zu Beginn des Schuljahres teil, in denen theoretische Grundlagen zu den Bereichen Lernverlaufsdiagnostik, Feedback und individuelle Förderung mit unmittelbarem Anwendungsbezug vermittelt wurden. Daran anknüpfend wurden die jeweiligen Materialien vorgestellt sowie Beispiele für die Instruktion der Schülerinnen und Schüler und die konkrete Umsetzung im Unterricht besprochen. Jede Lehrkraft erhielt – je nach Bedingung – die vorgestellten Feedback- und Fördermaterialien sowie ein Lehrermanual, in dem die entsprechenden Inhalte (theoretische Hintergründe, Beschreibung der Materialien, konkrete Instruktionen usw.) detailliert gesammelt waren. Mit Rückgriff auf die Ergebnisse einer vorangehenden Pilotierungsphase wurde empfohlen, mit jedem Kind bis zum Ende des Schuljahres drei Feedbackgespräche zu führen. Dabei wurden Gespräche in homogenen Gruppen mit bis zu vier Schülerinnen und Schülern vorgeschlagen. Für den Einsatz der Fördermaterialien wurde den Lehrkräften ein regelmäßiger, wöchentlicher Einsatz nahegelegt. Zwei Lehrkräfte konnten nicht an der Materialeinführung

teilnehmen und erhielten deshalb eine gesonderte Materialeinführung mit denselben Inhalten in ihrer Schule.

2.3 Instrumente

Implementation Um die Frage nach dem Gelingen der Implementation zu untersuchen, wurden die Lehrkräfte bezogen auf die Akzeptanz, Umsetzbarkeit, Nutzung, Wiedergabetreue und Nachhaltigkeit hinsichtlich der Lernverlaufsdagnostik mit quop und der Feedback- und Fördermaterialien befragt. Eine erste kurze (prospektive) Lehrerbefragung bezüglich der Akzeptanz und Umsetzbarkeit fand im Anschluss an die Materialeinführung statt. Zum Posttestzeitpunkt am Ende des Schuljahrs wurden die Lehrkräfte erneut und ausführlicher aus einer retrospektiven Perspektive befragt. Zudem wurden auch die Schülerinnen und Schüler zum Posttest bezogen auf die Akzeptanz und die subjektive Wirksamkeitserwartung befragt. Eine Übersicht über die eingesetzten Instrumente mit Beispielitems findet sich in Tab. 1. Die Skalen zur Akzeptanz und Umsetzbarkeit wurden für das Projekt in Anlehnung an Briesch et al. (2013) sowie Chafouleas et al. (2009) erstellt. Die Befragungen zur Nutzung von quop wurden von Kawohl (2015) übernommen und an die Rahmenbedingungen des Projektes angepasst. Dabei beziehen sich die Skalen „Förderung Klasse/Schüler“ und „Feedback Klasse/Schüler“ auf den allgemeinen Umgang mit den diagnostischen Informationen aus quop für die Gestaltung von Feedback und Förderung (ohne Bezug auf die in den Bedingungen FB und FB+FM bereitgestellten Feedback- und Fördermaterialien). Die Befragung zur Nutzung, Wiedergabetreue und Nachhaltigkeit in Bezug auf die Förder- und Feedbackmaterialien wurden in Anlehnung an Kawohl (2015) neu erstellt (Feedback) oder übernommen und angepasst (Förderung) und beziehen sich auf die ausgearbeiteten Materialien. Die internen Konsistenzen der Skalen sind in Tab. 1 zu finden. Wenn in Tab. 1 nicht anders angegeben, wurde eine vierstufige Likert-Skala mit den Antwortmöglichkeiten 1 = „trifft nicht zu“ bis 4 = „trifft genau zu“ oder 1 = „nie“ bis 4 = „häufig“ eingesetzt. Für die Schülerinnen und Schüler standen die Antwortmöglichkeiten 1 = „stimmt gar nicht“ bis 4 = „stimmt genau“ zur Verfügung.

Bezogen auf die Nutzung der Lernverlaufsdagnostik mit quop wurde über die Lehrer- und Schülerbefragungen hinaus als objektives Maß erhoben, wieviel Prozent der Schülerinnen und Schüler pro Klasse die acht Tests bearbeitet haben.

Bei der Abschlussbefragung der Lehrkräfte wurde zunächst in Form einer Filterfrage geprüft, ob die Ergebnisse der Lernverlaufsdagnostik im Lehrermenü angeschaut und ob die Feedback- bzw. Fördermaterialien eingesetzt wurden. In allen folgenden Analyseschritten wurden die Daten der Lehrkräfte (und deren Schülerinnen und Schüler) nur berücksichtigt, wenn angegeben wurde, dass die Ergebnisse im Lehrermenü angesehen wurden (LVD, FB, FB+FM), die Feedbackmaterialien (FB, FB+FM) und Fördermaterialien (FB+FM) im Unterricht eingesetzt wurden (vgl. Tab. 2). In der Bedingung LVD wurden die Angaben aller Lehrkräfte und die jeweiligen Schülerdaten weiterhin analysiert. In der Bedingung „Feedback“ geben nur vier (30,8 %) der Lehrkräfte an, sowohl die Lernverlaufsdagnostik als auch die Feedbackmaterialien eingesetzt zu haben. Somit wurden die restlichen acht Klassen im Folgenden aus allen Analysen ausgeschlossen. In der Bedingung FB+FM geben

Tab. 1 Übersicht über die eingesetzten Instrumente zur Erfassung der Implementationsfacetten

Konstrukt	It	α	Beispielitem
Materialeinführung			
<i>Akzeptanz</i>	3	0,71	„Ich finde es sinnvoll, Feedbackgespräche mit den Schülern zu führen.“
<i>Machbarkeit</i>	3	0,81	„Ich bin mir sicher, die Feedbackgespräche gut in meinen Unterricht integrieren zu können.“
Abschlussbefragung			
<i>Quop</i>			
<i>Nutzung</i>			
– Häufigkeit	1	–	„Wie oft pro Monat haben Sie sich durchschnittlich die Ergebnisse angesehen?“ (o. A.)
– Förderung Klasse	5	0,89	„Wie oft haben Sie aufgrund der Ergebnisse Unterrichtsinhalte mit der ganzen Klasse wiederholt?“
– Förderung Schüler	5	0,88	„Wie oft haben Sie aufgrund der Ergebnisse Unterrichtsinhalte mit einzelnen Kindern wiederholt?“
– Feedback Klasse	8	0,77	„Wie oft haben Sie aufgrund der Ergebnisse mit der ganzen Klasse über Stärken und Schwächen im Lesen gesprochen?“
– Feedback Schüler	8	0,91	„Wie oft haben Sie aufgrund der Ergebnisse mit einzelnen Kindern über Stärken und Schwächen im Lesen gesprochen?“
<i>Feedbackmaterial</i>			
<i>Akzeptanz</i>			
– Lehrer	8	0,88	„Die Feedbackgespräche waren für die Schüler hilfreich.“
– Schüler	3	0,78	„Durch die Gespräche über das Lesen habe ich Lust, das Lesen weiter zu üben.“
– Wirksamkeitserwartung L	1	–	„Die Feedbackgespräche haben dazu beigetragen, dass die Schüler sich im Lesen verbessert haben.“
– Wirksamkeitserwartung S	5	0,71	„Durch die Gespräche über das Lesen weiß ich, was mein Ziel beim Lesen ist.“
<i>Machbarkeit</i>			
– Zeitaufwand angemessen	1	–	„Der Zeitaufwand zur Durchführung der Feedbackgespräche war angemessen.“
– In Unterricht integrierbar	1	–	„Die Feedbackgespräche ließen sich gut in den Unterricht integrieren.“
<i>Nutzung</i>			
–	–	–	„Haben Sie die Dokumentationsbögen/ Gesprächsleitfäden/quop-Hefte genutzt?“ (dreistufige Likert-Skala: 1 = „nie“ bis 3 = „immer“)
<i>Wiedergabetreue</i>			
– Anzahl, Dauer	1	–	„Wie lange haben die Gespräche im Durchschnitt gedauert?“ (o. A.)
– Gruppen	1	–	„Haben Sie Feedbackgespräche in leistungsheterogenen Gruppen durchgeführt?“ (dreistufige Likert-Skala: 1 = „nie“ bis 3 = „immer“)
– Gruppengröße	1	–	„Wie groß waren die Gruppen im Durchschnitt?“ (o. A.)
– Veränderungen	1	–	„Ich habe Veränderungen gegenüber dem Gesprächsleitfaden bei der Durchführung der Feedbackgespräche vorgenommen.“

Tab. 1 (Fortsetzung)

Konstrukt	It	α	Beispielitem
Nachhaltigkeit			
– Auch im nächsten Schuljahr	1	–	„Ich werde auch im nächsten Schuljahr wieder Feedbackgespräche durchführen.“
<i>Förderung: Lese-Sportler (L-S)</i>			
Akzeptanz			
– Schüler	1	–	„Ich weiß, wie ich mit dem L-S trainieren soll.“
– Wirksamkeitserwartung S	1	–	„Durch das Training mit dem L-S habe ich mich im Lesen verbessert.“
– Wirksamkeitserwartung L	1	–	„Die Schüler haben sich durch das Training mit dem L-S verbessert.“
Machbarkeit			
– Zeitaufwand angemessen	1	–	„Der Zeitaufwand zur Durchführung des L-S war angemessen.“
Nutzung			
– Bis zum Ende? (ja/nein)	1	–	„Trainieren die Schüler bis zum Ende des Schuljahres mit dem L-S?“
Wiedergabetreue			
– Veränderungen	1	–	„Ich habe Veränderungen bei der Durchführung des L-S vorgenommen.“
Nachhaltigkeit			
– Auch im nächsten Schuljahr	1	–	„Ich werde auch im nächsten Schuljahr wieder mit dem L-S arbeiten.“

It Anzahl der Items, α Cronbachs Alpha, L Lehrer, S Schüler, L-S Lese-Sportler, o. A. offenes Antwortformat

zehn der Lehrkräfte an, quop und sowohl die Feedback- als auch die Fördermaterialien eingesetzt zu haben. Eine Lehrkraft gibt an, nur quop und die Feedbackmaterialien eingesetzt zu haben. In allen folgenden Analysen werden die Daten dieser Lehrkraft und ihrer Schülerinnen und Schüler in der Bedingung FB berücksichtigt. In der Bedingung FB+FM verbleiben somit zehn Klassen und 194 Schülerinnen und Schüler, in der Bedingung FB fünf Klassen und 108 Schülerinnen und Schüler und in der Bedingung LVD unverändert 16 Klassen und 330 Schülerinnen und Schüler. Alle im Folgenden berichteten Ergebnisse basieren auf dieser Stichprobe.

Wirksamkeit Das Leseverständnis und die basalen Lesefertigkeiten der Schülerinnen und Schüler wurden mit standardisierten Leistungstests erfasst. Die Lesemotivation, das Leseselbstkonzept, das Leseverhalten, die wahrgenommene Unterstützung sowie der selbstberichtete Einsatz von metakognitiven und kognitiven Lesestrategien der Schülerinnen und Schüler wurden mit Befragungsskalen erhoben.

Basale Lesefertigkeiten Die basalen Lesefertigkeiten der Schülerinnen und Schüler wurden mit dem Salzburger Lesescreening 2–9 (SLS 2–9; Wimmer und Mayringer 2014) gemessen. Dabei müssen die Schülerinnen und Schüler innerhalb von 3 min für so viele Sätze (z. B. „Bäume können sprechen.“) wie möglich beurteilen, ob die Aussage des Satzes richtig oder falsch ist. Verwendet wurden die beiden Pseu-

Tab. 2 Materialnutzung in den Untersuchungsbedingungen und Datengrundlage für die folgenden Analysen

	Untersuchungsbedingungen		
	LVD	FB	FB+FM
<i>Materialnutzung (%)</i>			
Nutzung Lehrermenü	100	100	100
Nutzung FB	–	30,8	73,7
Nutzung FM	–	–	66,7
<i>In die Analyse einbezogene Daten</i>			
Anzahl Klassen/Lehrkräfte	16	5	10
Anzahl Schüler	330	108	194

do-Parallelformen A1 und A2. Die Paralleltest-Reliabilitäten werden mit $r=0,95$ (2. Schulstufe) und $r=0,87$ (8. Schulstufe) angegeben. Im Hinblick auf die Validität liegen hohe Korrelationen mit lautem Lesen von Wortlisten ($r=0,89$ in der 2. Schulstufe) vor.

Leseverständnis Das Leseverständnis wurde mit einer bereits erprobten, adaptierten Version des Hamburger Lesetests (Förster und Souvignier 2014) erfasst. Dabei wurden aus der Originalversion des HAMLET 3–4 (Lehmann et al. 2006) zwei kontinuierliche Texte und ein diskontinuierlicher Text mit jeweils vier multiple choice Fragen ausgewählt und zwei weitere Fragen für den diskontinuierlichen und jeweils vier für jeden kontinuierlichen Text ergänzt. Die adaptierte Version umfasste somit 22 Items. Durch das Vertauschen der Antwortmöglichkeiten wurden zwei Pseudo-Paralleltestformen erstellt. Die interne Konsistenz der Kurzform ist hoch und vergleichbar mit der des Originaltests (Prätest: $\alpha=0,83$, Posttest: $\alpha=0,84$). Die Korrelation der Ergebnisse dieser Kurzform mit den Ergebnissen im SLS 2–9 als Indikator für die Validität liegt bei $r=0,67$ (Prätest) und $r=0,69$ (Posttest).

Schülerbefragung Die Schülerinnen und Schüler wurden gebeten, für verschiedene Aussagen zu beurteilen, inwiefern diese auf sie zutreffen. Als Antwortformat wurde, wenn im Folgenden nicht anders angegeben, jeweils eine vierstufige Likert-Skala von 1=„stimmt gar nicht“ bis 4=„stimmt genau“ verwendet. Die jeweiligen Aussagen wurden von den Testleitern vorgelesen, um sicherzustellen, dass die Schülerinnen und Schüler sie verstehen.

Um die Lesemotivation der Schülerinnen und Schüler zu erfragen, wurde der Reading Motivation Questionnaire for Elementary Students (RMQ-E; Stutz 2014) verwendet. Vier Items beziehen sich auf die intrinsische Lesemotivation (z. B. „Ich lese, weil ich mir dabei so viel vorstellen kann.“, $\alpha=0,65$ – $0,74$) und drei Items auf die extrinsische Lesemotivation („Ich lese, weil es mir wichtig ist, im Lesen immer der/die Beste zu sein“, $\alpha=0,85$ – $0,87$).

Das Leseselbstkonzept wurde mit sieben Items (z. B. „Ich kann das, was ich lese, leicht verstehen.“, $\alpha=0,79$ – $0,83$) erfasst. Dabei wurde auf die Skala aus der Berliner Längsschnittstudie 3–6 (McElvany 2008) zurückgegriffen.

Das Leseverhalten wurde in Anlehnung an McElvany (2008) mit vier Items (z. B. „Wie oft liest du zu deinem Vergnügen?“) mit einem Antwortformat von 1 = „nie“ bis 5 = „immer“ erfasst ($\alpha=0,72$).

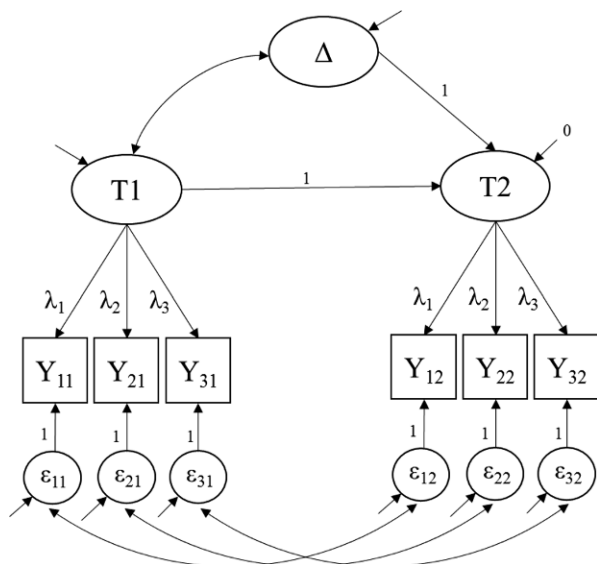
Mit einer neu konstruierten Skala wurden in Anlehnung an das Projekt „IGEL“ (Individuelle Förderung und adaptive Lerngelegenheiten in der Grundschule; Rieser et al. 2016) der selbstberichtete Einsatz kognitiver und metakognitiver Strategien der Schülerinnen und Schüler beim Lesen erfasst. Fünf Items adressierten metakognitive Strategien (z. B. „Während ich einen Text lese, überprüfe ich, ob ich alle wichtigen Dinge verstehe.“) und vier Items kognitive Strategien (z. B. „Wenn ich einen Text lese, unterstreiche ich die Dinge, die besonders wichtig sind.“; $\alpha=0,81-0,83$).

Die wahrgenommene Unterstützung der Schülerinnen und Schüler wurde mit der Skala „supportive climate“ (Fauth et al. 2014) erfasst. Die Skala bestand aus sieben Items und bezog sich auf unterstützende Tätigkeiten der Lehrkraft im Leseunterricht (z. B. „Im Leseunterricht wird mir erklärt, was ich besser machen kann, wenn ich beim Lesen Fehler mache.“; $\alpha=0,80-0,85$).

2.4 Statistische Analysen

Für alle Angaben bezüglich der Implementation aus den Schüler- und Lehrerbefragungen werden Mittelwerte und Standardabweichungen beziehungsweise relative Häufigkeiten berichtet. Für die Wirksamkeitsprüfung wurden Latent Change Modelle (LCM) in MPlus gerechnet (Version 6.1; Muthén und Muthén 2010; Abb. 2). Die Intraklassenkorrelationen lagen zwischen $ICC=0,01$ und $ICC=0,13$. Um die hierarchische Struktur der Daten zu berücksichtigen, nutzten wir das Varianz-Schätzverfahren „type=complex“ (s. Muthén und Muthén 2010). In LCM wird der Zuwachs über latente Differenzvariablen auf messfehlerfreier Ebene in die Modelle aufgenommen. Für die Ergebnisse des Leseverständnisses wurden nach dem von Little

Abb. 2 Latent-Change-Modell für ein Konstrukt mit drei beobachtbaren Variablen (Y_1 , Y_2 , Y_3), gemessen zu zwei Messzeitpunkten. Die Anzahl der beobachtbaren Variablen liegt je nach Konstrukt zwischen 3 und 9. Die Faktorladungen und Intercepts wurden über die beiden Messzeitpunkte konstant gehalten. E = Messfehlervariable, λ = Faktorladungsparameter



et al. (2002) beschriebenen Verfahren „high-to-low loadings“ aus den Items drei parcels gebildet, welche als Indikatoren für die latenten Faktoren eingesetzt wurden. Für alle anderen Variablen dienten die einzelnen Items der jeweiligen Skala als Indikatoren für die latenten Faktoren. Die Faktorladungen und Intercepts wurden über die beiden Messzeitpunkte konstant gehalten (starke faktorielle Messinvarianz), damit die gemessenen latenten Variablen über die Zeit vergleichbar waren. Für den SLS konnten keine latenten Faktoren gebildet werden, da nur der Gesamtestwert und somit nicht ausreichend Indikatoren vorlagen. Aus diesem Grund wurde eine manifeste Differenzwertvariable gebildet ($T2-T1$) und in das Modell aufgenommen. Die Zugehörigkeit zu einer Bedingung wurde mit zwei Variablen dummy-kodiert. Dabei dient die Bedingung LVD als Referenzkategorie. Zum Prä- und Posttestzeitpunkt fehlten im Mittel ca. 10% (*min*: 6%, *max*: 21%) der Daten. Wir nutzten das full information maximum likelihood Verfahren (FIML) als konventionelles Verfahren im Umgang mit fehlenden Werten (Enders 2001).

3 Ergebnisse

Das erste Ziel dieser Studie war es zu prüfen, in welchem Maße die Implementation der drei Komponenten Diagnostik, Feedback und individuelle Förderung gelingt. Im zweiten Schritt sollte dann die Wirksamkeit des Einsatzes der Materialien geprüft werden.

3.1 Implementation

Tab. 3 (oben) zeigt die Ergebnisse der Evaluation der Materialeinführung durch die Lehrkräfte der Bedingungen FB und FB+FM. Die Akzeptanz ist in beiden Bedingungen sehr hoch (FB: $M=3,60$, $SD=0,43$; FB+FM: $M=3,54$, $SD=0,35$). Die Einschätzung der Umsetzbarkeit liegt über dem Skalenmittelwert, ist jedoch geringer als die Akzeptanz (FB: $M=2,60$, $SD=0,60$; FB+FM: $M=2,83$, $SD=0,36$). Zwischen den Bedingungen bestehen keine signifikanten Unterschiede. Auch die Lehrkräfte, die die Materialien später nicht einsetzten und deren Angaben deshalb in den Analysen nicht berücksichtigt wurden, gaben für die Akzeptanz ($M=3,43$, $SD=0,35$) und die Umsetzbarkeit ($M=2,45$, $SD=0,50$) vergleichbar hohe Werte an.

Lernverlaufsdiagnostik In Tab. 4 sind die deskriptiven Statistiken für die Nutzung der Lernverlaufsdiagnostik mit quop dargestellt. Über alle Tests hinweg bearbeiteten im Mittel mehr als 95% der Schülerinnen und Schüler die quop-Tests. Die Lehrkräfte sahen sich die Ergebnisse im Lehrer Menü 1.5 bis 2.1 mal im Monat an. Die Lehrkräfte der Bedingung FB+FM geben tendenziell höhere Werte an, jedoch ist in dieser Bedingung auch eine große Streuung ($SD=1,29$) zu beobachten. Auch im Hinblick auf die Angaben zur allgemeinen Nutzung der Lernverlaufsdaten für Feedback und Förderung (unabhängig von den Feedback- und Fördermaterialien) zeigen sich – auf einem für alle drei Gruppen gleichermaßen mittleren Niveau – keine signifikanten Unterschiede zwischen den Bedingungen. Insgesamt gibt es dabei eine minimal höhere Zustimmung dafür, Maßnahmen auf der Ebene der individuellen

Tab. 3 Implementationsmaße für die Feedback- und Fördermaterialien

	Untersuchungsbedingungen			
	FB		FB+FM	
	<i>M (SD)</i>	%	<i>M (SD)</i>	%
Evaluation der Materialien (Prä)				
<i>Akzeptanz (nach Materialeinführung)</i>	3,60 (0,43)	–	3,54 (0,35)	–
<i>Machbarkeit (nach Materialeinführung)</i>	2,60 (0,60)	–	2,83 (0,36)	–
Feedbackmaterialien (Post)				
<i>Akzeptanz (FB-Post)</i>				
Lehrer	3,42 (0,36)	–	2,86 (0,42)	–
Schüler	3,30 (0,67)	–	3,22 (0,76)	–
Wirksamkeitserwartung Lehrer	2,00 (0,71)	–	2,60 (0,53)	–
Wirksamkeitserwartung Schüler	3,71 (0,45)	–	3,58 (0,48)	–
<i>Machbarkeit (FB-Post)</i>				
Zeitaufwand angemessen?	2,20 (0,11)	–	1,90 (0,57)	–
In Unterricht integrierbar?	1,40 (0,55)	–	1,89 (0,60)	–
<i>Nutzung (FB-Post)</i>				
Dokumentationsbogen				
– Nie	0	0,0	1	11,1
– Immer	5	100,0	6	66,7
Leitfaden				
– Nie	0	0,0	2	20,0
– Immer	4	80,0	3	30,0
Quop-Hefte				
– Nie	0	0,0	0	0,0
– Immer	4	80,0	6	66,7
<i>Wiedergabetreue (FB-Post)</i>				
Anzahl pro Schuljahr	1,50 (0,50)	–	2,35 (1,16)	–
Dauer pro Gespräch in Minuten	12,00 (3,56)	–	9,95 (4,38)	–
Gespräche in Gruppen?				
– Nie	0	0,0	1	10,0
– Immer	3	60,0	3	30,0
Davon leistungsheterogen				
– Nie	5	100,0	7	77,8
– Immer	0	0,0	1	11,1
Gruppengröße (Anzahl Schüler)	3,5 (0,61)	–	4,6 (1,74)	–
Veränderungen vorgenommen?	1,40 (0,55)	–	2,38 (0,74)	–
<i>Nachhaltigkeit (FB-Post)</i>				
Auch im nächsten Schuljahr?	2,00 (0,81)	–	2,60 (1,17)	–
Fördermaterialien (Post)				
<i>Akzeptanz (FM-Post)</i>				
Akzeptanz Schüler	–	–	3,45 (0,76)	–
Wirksamkeitserwartung Schüler	–	–	3,32 (0,87)	–
Wirksamkeitserwartung Lehrer	–	–	2,75 (0,89)	–

Tab. 3 (Fortsetzung)

	Untersuchungsbedingungen			
	FB		FB+FM	
	<i>M</i> (<i>SD</i>)	%	<i>M</i> (<i>SD</i>)	%
<i>Machbarkeit (FM-Post)</i>	–	–	–	–
Zeitaufwand angemessen?	–	–	2,56 (0,88)	–
<i>Nutzung (FM-Post)</i>	–	–	–	–
Bis zum Ende trainiert?	–	–	5	62,5
<i>Wiedergabetreue (FM-Post)</i>	–	–	–	–
Veränderungen vorgenommen	–	–	2,63 (0,74)	–
<i>Nachhaltigkeit (FM-Post)</i>	–	–	–	–
Auch im nächsten Schuljahr?	–	–	2,56 (1,13)	–

In den Spalten *M* (*SD*) ist in den Fällen, in denen auch Prozentzahlen angegeben sind, jeweils die Zahl absoluter Nennungen angegeben

Tab. 4 Implementationsmaße für die Lernverlaufsdagnostik (Post)

	Untersuchungsbedingungen		
	LVD	FB	FB+FM
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
<i>Nutzung</i>			
Bearbeitete quop-Tests (%)	96,25 (3,57)	95,88 (5,10)	95,41 (5,63)
Lehrermenü (pro Monat)	1,54 (0,54)	1,50 (0,87)	2,10 (1,29)
Skala Feedback Klasse	2,58 (0,58)	2,68 (0,64)	2,74 (0,60)
Skala Feedback Schüler	2,99 (0,75)	2,80 (0,45)	3,34 (0,45)
Skala Förderung Klasse	2,37 (0,51)	2,50 (0,10)	2,45 (0,67)
Skala Förderung Schüler	2,67 (0,64)	2,95 (0,50)	2,80 (0,49)

Schülerinnen und Schüler vorgenommen zu haben, als instruktionale Entscheidungen auf der Ebene der Klasse vorgenommen zu haben.

Feedbackmaterialien Die Ergebnisse zur Implementation der Feedbackmaterialien (Akzeptanz, Umsetzbarkeit, Nutzung, Wiedergabetreue und Nachhaltigkeit) sind Tab. 3 zu entnehmen.

Die *Akzeptanz* bei den Lehrkräften ist hoch und in der Bedingung FB ($M=3,42$, $SD=0,36$) tendenziell höher als in der Bedingung FB+FM ($M=2,86$, $SD=0,42$). Die Schülerinnen und Schüler geben in beiden Bedingungen vergleichbar hohe Werte an. Die Wirksamkeitserwartung von Seiten der Schülerinnen und Schüler ist hoch, während die Wirksamkeitserwartung der Lehrkräfte etwas geringer ausfällt. Die Angaben zur *Umsetzbarkeit* (Zeitaufwand; Integrierbarkeit) liegen im unterdurchschnittlichen Skalenbereich. Im Hinblick auf die *Nutzung* geben die Lehrkräfte der Bedingung FB einen intensiven Einsatz von Dokumentationsbogen, Leitfaden und quop-Heften an, während diese Materialien in der Bedingung FB+FM von rund zwei Drittel der Lehrkräfte genutzt wurden. Bei der *Wiedergabetreue* zeigt sich ein gemischtes Bild: Während die Zahl durchgeführter Feedbackgespräche in der Bedingung FB+FM etwas höher lag, orientierten sich die Lehrkräfte in der Bedingung FB

Tab. 5 Mittelwerte und Standardabweichungen im Prä- und Posttest für die drei Untersuchungsbedingungen

	Untersuchungsbedingungen					
	LVD		FB		FB+FM	
	<i>T1</i>	<i>T2</i>	<i>T1</i>	<i>T2</i>	<i>T1</i>	<i>T2</i>
	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)	(<i>SD</i>)
Basale Lesefertigkeiten	28,75 (8,04)	35,97 (8,18)	28,30 (8,60)	36,51 (9,46)	28,41 (8,56)	36,15 (9,37)
Leseverständnis	7,41 (4,47)	11,46 (4,58)	7,14 (4,84)	11,30 (5,44)	7,49 (4,68)	11,18 (4,75)
Intrinsische Motivation	3,50 (0,57)	3,42 (0,60)	3,57 (0,48)	3,53 (0,53)	3,48 (0,54)	3,38 (0,56)
Extrinsische Motivation	2,51 (0,99)	2,04 (1,00)	2,24 (1,01)	1,91 (0,87)	2,60 (1,04)	2,10 (1,05)
Leseselbstkonzept	3,32 (0,47)	3,20 (0,53)	3,48 (0,46)	3,32 (0,49)	3,31 (0,54)	3,21 (0,56)
Leseverhalten	3,29 (0,91)	3,23 (0,87)	3,48 (0,88)	3,60 (0,82)	3,54 (0,80)	3,43 (0,91)
Strategieinsatz	2,61 (0,66)	2,46 (0,69)	2,53 (0,73)	2,68 (0,72)	2,58 (0,64)	2,48 (0,68)
Wahrgenommene Unterstützung	3,22 (0,56)	3,20 (0,65)	3,15 (0,65)	3,32 (0,62)	3,25 (0,60)	3,28 (0,57)

T1 Prätest, *T2* Posttest.

enger an den Empfehlungen (Gespräche in homogenen Kleingruppen). Die Lehrkräfte der Bedingung FB+FM geben an, signifikant mehr Veränderungen an dem Material vorgenommen zu haben ($t(11) = -2,52, p < 0,05$). Die Lehrkräfte der Bedingung FB+FM stimmen im Sinne einer *Nachhaltigkeit* der Aussage, dass sie auch im nächsten Schuljahr Feedbackgespräche durchführen werden, tendenziell mehr zu als die Lehrkräfte in der Bedingung FB.

Fördermaterialien Die Ergebnisse zur Implementation der Fördermaterialien (Akzeptanz, Umsetzbarkeit, Nutzung, Wiedergabetreue und Nachhaltigkeit) sind Tab. 3 zu entnehmen. 62,5 % der Lehrkräfte ließen ihre Schülerinnen und Schüler bis zum Ende des Schuljahres mit dem Lese-Sportler trainieren. Die Akzeptanz und die Wirksamkeitserwartung der Schülerinnen und Schüler ist hoch, während die Angaben der Lehrkräfte zur Wirksamkeitserwartung, der Umsetzbarkeit und der Nachhaltigkeit nur wenig über dem Skalenmittelwert liegen. Die Lehrkräfte stimmen der Aussage, Veränderungen am Material vorgenommen zu haben, tendenziell eher zu.

3.2 Wirksamkeit

Tab. 5 zeigt die Mittelwerte und Standardabweichungen der Schulervariablen zum Prä- und Posttest getrennt für die drei Bedingungen. In Tab. 6 sind die Gütemaße für die Latent Change Modelle dargestellt. Die Ergebnisse der Latent Change Modelle sind in Tab. 7 zu finden. Das Ausgangsniveau unterscheidet sich für alle Variablen

Tab. 6 Gütemaße für die Latent Change Modelle

	χ^2 (df)	CFI/TLI	RMSEA (90% CI)	SRMR
HAMLET	22.604 (9)	0,994/0,990	0,049 (0,024, 0,074)	0,084
SLS	–	–	–	–
Intrinsische Motivation	23.615 (21)	0,996/0,995	0,014 (0,000, 0,037)	0,029
Extrinsische Motivation	22.604 (9)	0,994/0,990	0,049 (0,024, 0,074)	0,045
Selbstkonzept	112.322 (81)	0,985/0,983	0,025 (0,012, 0,035)	0,050
Leseverhalten	33.897 (21)	0,990/0,986	0,031 (0,008, 0,050)	0,027
Strategieinsatz	216.547 (141)	0,968/0,965	0,029 (0,021, 0,036)	0,046
Wahrgenommene Unterstützung	142.658 (81)	0,962/0,957	0,035 (0,025, 0,044)	0,053

χ^2 Chi-Quadrat, *CFI* Comparative fit index, *TLI* Tucker Lewis Index, *RMSEA* Root means square error of approximation, *SRMR* standardized root mean square residual, für den SLS liegen keine Angaben vor, da keine latente Modellierung vorgenommen werden konnte.

Tab. 7 Ergebnisse der Latent Change Modelle

Referenzgruppe	LVD				FB+FM				
	β	SE	z	p	β	SE	z	p	
<i>Ausgangsniveau</i>									
HAMLET	0,29	0,32	0,89	0,37	0,02	0,09	0,25	0,81	
SLS	-0,40	1,36	-0,29	0,77	-0,46	1,39	-0,33	0,74	
Intrinsische Motivation	0,09	0,09	1,01	0,31	-0,03	0,08	-0,33	0,74	
Extrinsische Motivation	-0,30	0,20	-1,49	0,14	0,09	0,24	0,38	0,70	
Selbstkonzept	0,13	0,07	1,90	0,06	-0,01	0,06	-0,08	0,94	
Leseverhalten	0,18	0,17	1,05	0,29	0,27	0,13	2,17	0,03	
Strategieinsatz	-0,08	0,14	-0,61	0,54	-0,03	0,10	-0,27	0,79	
Wahrgenommene Unterstützung	-0,06	0,09	-0,74	0,46	0,03	0,08	0,33	0,74	
<i>Veränderung über die Zeit</i>									
HAMLET	0,10	0,34	0,28	0,78	-0,09	0,17	-0,53	0,60	
SLS	0,61	0,59	1,04	0,30	0,58	0,78	0,74	0,46	
Intrinsische Motivation	0,03	0,05	0,53	0,60	0,01	0,08	0,12	0,90	
Extrinsische Motivation	0,18	0,17	1,07	0,29	-0,06	0,14	-0,39	0,69	
Selbstkonzept	-0,03	0,05	-0,62	0,54	0,00	0,05	-0,01	0,99	
Leseverhalten	0,12	0,07	1,79	0,07	-0,06	0,09	-0,69	0,49	
Strategieinsatz	0,28	0,18	1,57	0,12	0,01	0,07	0,17	0,87	
Wahrgenommene Unterstützung	0,16	0,11	1,44	0,15	0,03	0,10	0,31	0,76	

nicht signifikant von der Bedingung LVD. Lediglich beim Leseverhalten geben die Schülerinnen und Schüler in der Bedingung FB+FM im Prätest höhere Werte an, als in der Bedingung LVD ($z=2,17$, $p<0,05$). Nach einer Bonferroni-Korrektur für multiples Testen bleibt dieser Effekt jedoch nicht bestehen. Insgesamt erweisen sich die Veränderungen im Hinblick auf die Leseleistung und die motivationalen

Variablen zwischen den Bedingungen als vergleichbar und es zeigen sich keine signifikanten Effekte.¹

4 Diskussion

In dieser Studie sollte in einem ersten Schritt untersucht werden, in welchem Maße die Implementation eines ausgearbeiteten modularen Konzeptes formativen Assessments für den Bereich Lesen – unterstützt durch Unterrichtsmaterialien und Lehrerfortbildungen – in den Schulalltag gelingt. In einem zweiten Schritt sollte geprüft werden, wie wirksam die Unterstützung von Lehrkräften durch konkrete Materialien und Lehrerfortbildungen ist, wenn sie einen Ansatz zur Lernverlaufsdiagnostik nutzen. Dabei wurde der Bedarf an konkreten Konzeptualisierungen von formativem Assessment innerhalb einzelner Domänen (Bennett 2011) adressiert und im Sinne des stärker formalisierten Ansatzes des curriculum-embedded assessments ein konkretes Konzept für die drei Komponenten Diagnostik, Feedback und Förderung im Bereich Lesen erarbeitet.

4.1 Zusammenfassung der Ergebnisse

Insgesamt weisen die Befunde darauf hin, dass die Implementation der Lernverlaufsdiagnostik mit quop gelingt, während die Unterrichtsmaterialien für die beiden Komponenten Feedback und Förderung in deutlich geringerem Maße genutzt wurden. Entsprechend führen die zusätzlichen Unterstützungsangebote für die Lehrkräfte bei der Ausgestaltung von Feedback und individueller Förderung beim Einsatz von Lernverlaufsdiagnostik nicht zu einer weiteren Verbesserung der Schülerleistung, der Lesemotivation, des Leseselbstkonzepts, des Leseverhaltens, der wahrgenommenen Unterstützung und des Einsatzes von metakognitiven und kognitiven Lesestrategien. Im Folgenden sollen die Befunde im Einzelnen interpretiert und in den theoretischen Kontext eingeordnet werden. Im Anschluss werden praktische Konsequenzen sowie Limitationen der vorliegenden Studie angesprochen.

4.1.1 Implementation

Während die Akzeptanz bezogen auf das Feedback- und Fördermaterial sowohl bei den Schülerinnen und Schülern als auch bei den Lehrkräften hoch ist, wird die Umsetzbarkeit und hier insbesondere die Integrierbarkeit in den Unterricht durch die Lehrkräfte nur als mittelmäßig eingeschätzt. Ein ähnliches Bild ergibt sich für die Nutzung und Wiedergabetreue beim Einsatz der Materialien: Insgesamt werden weniger als drei Feedbackgespräche im Verlauf des Schuljahres geführt. In

¹ Eine Auswertung mittels gemischter linearer Regressionsmodelle mit random effects für die Variablen „Schülercode“, „Klassen-ID“ und „Schul-ID“ und fixed effects für die Zeit und die Zugehörigkeit zur Bedingung führte zu vergleichbaren Ergebnissen. Nach einer Bonferroni-Korrektur für multiples Testen zeigte sich lediglich ein signifikanter disordinaler Interaktionseffekt für die Bedingung FB und die Variable Strategieeinsatz ($t(590) = 3,39, p < 0,001; d = 0,44$).

der Bedingung FB+FM werden die Materialien seltener genutzt. Außerdem werden Veränderungen am Material vorgenommen und somit Abweichungen von der intendierten Maßnahme (z. B. heterogene Zusammensetzung von Gruppen bei Feedbackgesprächen) hervorgerufen. Die Materialien des Lese-Sportlers werden nur in 62 % der Fälle bis zum Ende des Schuljahres eingesetzt. Nicht zuletzt ist auch die Nachhaltigkeit in beiden Bedingungen (FB und FB+FM) nur in geringem Maße gegeben: Die Lehrkräfte stimmen der Aussage, die Materialien auch im nächsten Schuljahr zu nutzen, nur eingeschränkt zu. Im direkten Vergleich zwischen allen drei Bedingungen zeigt sich, dass die Lehrkräfte in allen Gruppen in gleichem Maße angaben, die diagnostischen Informationen zu den Lernverläufen für Feedback- und Fördermaßnahmen genutzt zu haben. Dies ließe sich als Hinweis darauf interpretieren, dass der Baustein der Lernverlaufdiagnostik bereits einen genügend hohen Aufforderungscharakter mit sich bringt, um das Konzept des formativen Assessment zu „transportieren“.

Eine mögliche Ursache für die mangelnde Implementation könnte darin gelegen haben, dass die konkrete Umsetzung der Strategie, Unterrichtsmaterialien in der Kombination mit Lehrerfortbildungen anzubieten, durch zu umfangreiche Materialien für die praktische Umsetzung eine Überforderung für viele Lehrkräfte darstellte. Da sich generell ähnliche Herangehensweisen zuvor als effektiv erwiesen hatten (Stecker et al. 2005; Bennett 2011; Hondrich et al. 2016), ist davon auszugehen, dass die konkrete Umsetzung mit Feedback- und Fördermaterialien nicht für die in den Schulen gegebenen Rahmenbedingungen passend war.

Zudem stellt sich die Frage, welcher Grad an Formalisierung für die einzelnen Komponenten Diagnostik, Feedback und Förderung förderlich ist. Für die Lernverlaufdiagnostik mit quop scheint der stark formalisierte Ansatz im Sinne des curriculum-embedded assessment (Heritage 2007; Shavelson et al. 2008) hilfreich zu sein: Durch die automatisierte Auswertung und Dokumentation sowie die ermöglichte Eigenständigkeit der Schülerinnen und Schüler bei der regelmäßigen Testbearbeitung ergibt sich für die Lehrkräfte eine Entlastung. Möglicherweise stellt allerdings die Formalisierung der beiden Komponenten Feedback und Förderung durch die ausgearbeiteten Materialien einen Eingriff in Unterrichtsroutine dar, der mit (zu) großen Einschränkungen im Hinblick auf eine unmittelbare Umsetzbarkeit einhergeht. Dafür spricht beispielsweise, dass 13 Lehrkräfte aus der anfänglichen Stichprobe die Materialien nicht nutzten und deshalb aus den Analysen ausgeschlossen werden mussten. Wie die Ergebnisse der Implementationsstudie von Hondrich et al. (2016) zeigen, ist die lehrerseitige Akzeptanz sicherlich ein förderlicher Aspekt für das Gelingen der Implementation, jedoch weisen die Ergebnisse der vorliegenden Studie darauf hin, dass dies nicht als hinreichende Implementationsbedingung gesehen werden kann. Für eine formalisierte Umsetzung des theoretisch komplexen Konzepts formativen Assessments, die mit deutlichen Veränderungen unterrichtlicher Abläufe einhergeht, scheint es notwendig zu sein, noch intensivere Unterstützungsmaßnahmen anzubieten. Diese müssten den in der Literatur diskutierten Implementationsbarrieren wie Unsicherheiten beim Einsatz neuer und ungewohnter Konzepte in den Unterricht, eingeschränkte Ressourcen zur Nutzung von Fortbildungsangeboten und fehlende Anreize für Innovationen im System Schule Rechnung tragen (van Keer und Verhaeghe 2005; Souvignier und Philipp 2016).

Nicht zuletzt vor dem Hintergrund, dass eine vorherige Version der Lese-Sportler Materialien (geringere Betonung des Feedback Bausteins und weniger differenzierte Leseförderung) sich als gut implementierbar erwies (Kawohl 2015), stellt sich die Frage nach einem optimalen Zusammenspiel von unterstützenden Rahmenbedingungen (auch administrativ) und angemessener Komplexität von Materialien.

4.1.2 *Wirksamkeit*

In der hier realisierten Form hat die zusätzliche Unterstützung der Lehrkräfte beim Einsatz der Lernverlaufsdiagnostik mit quop durch Feedback- und Fördermaterialien sowie der Einsatz dieser Materialien durch die Lehrkräfte keine signifikanten Effekte auf die abhängigen Variablen.

Vor dem Hintergrund, dass gelingende Implementationsprozesse eine maßgebliche Voraussetzung für die Wirksamkeit sind (Schneider und Randel 2010; Spörer und Glaser 2010; Bennett 2011), liegt es nahe, die wenig erfolgreiche Implementation von Feedback und Förderung als Erklärung für die ausbleibenden Effekte heranzuziehen. In der Bedingung FB+FM könnte hier insbesondere die Wiedergabetreue eine Rolle gespielt haben. Die Anpassung der Materialien an die jeweiligen Rahmenbedingungen der Schule im Sinne eines „sich zu eigen Machens“ ist sicherlich wünschenswert. Dennoch können zu starke Veränderungen dazu führen, dass zentrale Wirkfaktoren aus dem intendierten Konzept herausgenommen werden. Zudem deutet sich an, dass das Material nicht in der notwendigen Intensität und nicht ausreichend langfristig eingesetzt wurde.

Ein weiterer Erklärungsansatz für die ausbleibenden Effekte ließe sich daraus ableiten, dass die Lehrkräfte aller Bedingungen in gleichem Maße Angaben, die diagnostischen Informationen zu den Lernverläufen für Feedback- und Fördermaßnahmen genutzt zu haben. Diese Selbstauskunft sollte allerdings vorsichtig interpretiert werden, denn die Angaben zur Umsetzung der Feedback- und Förderkonzepte weisen darauf hin, dass die konkret durchgeführten unterrichtlichen Maßnahmen nur bedingt den Prinzipien evidenzbasierten Feedbacks (Kluger und DeNisi 1996; Hattie und Timperley 2007) oder evidenzbasierter Leseförderung (NICHHD 2000; Slavin et al. 2009) entsprachen. Zudem deuten Unterrichtsbeobachtungen anderer Studien darauf hin, dass Lehrkräfte im alltäglichen Unterricht ohne Unterstützung (durch Fortbildungen oder Materialien) seltener Elemente formativen Assessments einsetzen (z. B. Hondrich et al. 2016).

4.2 **Praktische Implikationen**

Die nicht zufriedenstellend gelungene Implementation von Materialien zur Umsetzung des Konzepts des formativen Assessment unterstreicht, wie wichtig es ist, Implementationsbarrieren bei der Ausarbeitung von Konzepten und deren Umsetzung im Regelschulbereich zu antizipieren und ihnen zu begegnen. In der vorliegenden Studie besteht grundsätzlich eine hohe Akzeptanz für das Konzept. Fraglich ist jedoch, wie Umsetzbarkeit, Wiedergabetreue und Nachhaltigkeit besser unterstützt werden können. Hier könnte eine verstärkte Zusammenarbeit von Lehrkräften und Forschern im Sinne des „design based research“ Ansatzes (Brown 1992) vielver-

sprechend sein, bei dem eine systematische Optimierung von Unterrichtskonzepten in der unterrichtlichen Praxis vorgenommen wird. Die Befunde der vorliegenden Studie geben erste Hinweise darauf, dass ein langfristiger gedachtes Implementationskonzept, bei dem die einzelnen Komponenten schrittweise eingeführt, erprobt und in Unterrichtsrouninen eingebunden werden, lohnenswert erscheint. Insbesondere die geringere Wiedergabetreue in der Bedingung FB+FM, in der die Lehrkräfte die Materialien für Feedbackgespräche und Fördermaterialien gleichzeitig erhielten, deuten darauf hin, dass ein zu hohes Maß an Veränderung mit einer geringen Implementationsgüte einhergehen kann. Bei einer schrittweisen Einführung könnten Lehrkräfte zunächst positive Erfahrungen mit einer Komponente sammeln, was sich als förderlich für die langfristige Übernahme von Konzepten erwiesen hat (Clarke und Hollingsworth 2002), bevor die nächste Komponente eingeführt wird. Dabei könnten die gewohnten Unterrichtsrouninen der Lehrkräfte näher analysiert und einbezogen werden. Es könnten Anknüpfungspunkte gesucht, vorhandene Ressourcen genutzt werden und eine Brücke zwischen Unterrichtsrouninen und innovativen Konzepten geschlagen werden, um mögliche Unsicherheiten der Lehrkräfte zu adressieren (van Keer und Verhaeghe 2005).

Forschungsbefunde weisen darauf hin, dass Lehrkräfte in Ihrer Ausbildung nicht ausreichend auf den Einsatz formativen Assessments vorbereitet werden (Popham 2009; Cizek 2010; Stiggins 2010; Zeuch et al. 2017). Zweifellos wäre es wünschenswert, dem Konzept des formativen Assessment zu Grunde liegende evidenzbasierte Prinzipien des Feedbacks und der Förderung intensiver in der Lehrerausbildung zu adressieren, Fortbildungsangebote auszubauen sowie Freistellungen zur Nutzung dieser Angebote zu gewähren (Bürgermeister et al. 2014; Souvignier und Philipp 2016).

4.3 Limitationen und weiterer Forschungsbedarf

Bei der Interpretation der beschriebenen Befunde sind einige einschränkende Faktoren zu beachten: Erste Einschränkungen beziehen sich auf die Erfassung der Implementationsfacetten in Form von Selbstauskünften. Faktoren wie soziale Erwünschtheit können bei der Beantwortung der Fragen eine Rolle gespielt haben. Daten aus Unterrichtsbeobachtungen könnten hier eine aussagekräftige Ergänzung darstellen (Proctor et al. 2011). Zudem beruht die Erfassung mancher Implementationsfacetten nur auf einem Item (z. B. Wirksamkeitserwartung).

Zweitens ist die geringe Stichprobengröße und insbesondere die Verringerung der in die Analyse einbezogenen Klassen auf 5 Klassen in der Bedingung FB zu nennen. Vor diesem Hintergrund ist eine Generalisierbarkeit der Ergebnisse möglicherweise eingeschränkt.

Drittens wäre eine präzise Erfassung wünschenswert gewesen, in welcher Art und in welchem Umfang Veränderungen an den Materialien vorgenommen wurden. Auf diese Weise könnte geprüft werden, wie Abweichungen vom intendierten Konzept sich auf mögliche Effekte auswirken. Auch hier bieten sich Unterrichtsbeobachtungen an, um zu prüfen, inwiefern grundsätzliche Prinzipien der Konzepte beibehalten wurden. Im Hinblick auf eine Optimierung der Materialien wäre es zudem sinnvoll gewesen, die Beweggründe der Lehrkräfte zu erfragen, die die Feedback- und Förder-

materialien nicht einsetzen, um daraus Rückschlüsse auf Implementationsbarrieren ableiten zu können. In diesem Zusammenhang wäre es wünschenswert, mehr über die Unterrichtsroutrinen der Lehrkräfte im Umgang mit formativen Assessment und schon vorhandene Konzepte und Ressourcen zu erfahren, um Anknüpfungspunkte für die Implementation herzustellen.

Viertens wurden in der vorliegenden Studie nur verhältnismäßig kurzfristige Effekte über ein Schuljahr untersucht. Da die Implementation innovativer Konzepte auf langfristige und nachhaltige Veränderung des Lehrerhandelns abzielt, liegt es nahe, auch langfristige Effekte über mehr als ein Schuljahr zu untersuchen (Euler und Sloane 1998; Hasselhorn et al. 2014). Dabei ist die Erfassung von Begleitmerkmalen (z. B. Kooperation im Kollegium, Unterstützung durch Schulleitung) oder auch möglicher Implementationsbarrieren sicherlich hilfreich, um Hinweise für eine optimale Unterstützung von Lehrkräften zu erhalten. Insbesondere der Befund, dass in den Bedingungen FB und FB+FM unterschiedlich viele Lehrkräfte die Materialien genutzt haben (31 % vs. 70 %), wirft die Frage auf, welche Faktoren die Umsetzung materialgestützter evidenzbasierter Konzepte in der Praxis fördern oder hemmen.

Auf der Basis der vorliegenden Ergebnisse können keine Aussagen darüber getroffen werden, inwiefern Implementationsmaße miteinander interagieren und inwiefern sie als Kovariaten für die Wirksamkeit fungieren. Dies ist eine Fragestellung, die sicherlich hilfreiche Hinweise für die Gestaltung der Konzepte formativen Assessments und die Implementation mit dem Ziel einer nachhaltigen Veränderung liefern könnte. Nicht erfasst wurde zudem, ob die Lehrkräfte auch die Klassenlehrer der jeweiligen Klassen waren und ob die Anzahl der Stunden, die sie in den Klassen unterrichtet haben, über die Unterrichtsbedingungen vergleichbar waren.

4.4 Fazit

Das ausgearbeitete Konzept für eine Ergänzung der Lernverlaufsdagnostik durch die Komponenten Feedback und Förderung erweist sich in der hier beschriebenen Form als zu komplex, um es in die Schulpraxis zu implementieren. Der Effekt, den die Bereitstellung diagnostischer Informationen über den Lernverlauf von Schülerinnen und Schülern hat (Stecker et al. 2005; Förster und Souvignier 2014, 2015), lässt sich durch die beschriebenen Materialien zur Umsetzung des Ansatzes formativen Assessments nicht verstärken.

Auch wenn das Konzept des formativen Assessments auf der Basis der bisherigen Forschungslage vielversprechend erscheint, stellen eine praxistaugliche und wirksame Integration der drei Komponenten Diagnostik, Feedback und Förderung sowie deren nachhaltige Implementation in den Regelschulbereich Herausforderungen dar, die noch viel Entwicklungsarbeit erfordern. Shavelson (2008) fasst eine ähnliche Einschätzung folgendermaßen zusammen: „After five years of work, our euphoria devolved into a reality that formative assessment, like so many other education reforms, has a long way to go before it can be wielded masterfully by a majority of teachers to positive ends.“ (S. 294). Dabei spielen insbesondere der Umgang mit Implementationsbarrieren und das Finden geeigneter Implementationsstrategien eine übergeordnete Rolle.

Literatur

- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213–238.
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, *18*, 5–25.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*, 7–74.
- Black, P., & Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Hrsg.), *Assessment and learning* (S. 9–25). London: SAGE.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*, 5–31.
- Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: revision of the usage rating profile-intervention. *Journal of School Psychology*, *51*, 81–96.
- Brown, A. L. (1992). Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, *2*, 142–178.
- Bürgermeister, A., Klieme, E., Rakoczy, K., Harks, B., & Blum, W. (2014). Formative Leistungsbeurteilung im Unterricht: Konzepte, Praxisberichte und ein neues Diagnoseinstrument für das Fach Mathematik. In M. Hasselhorn, W. Schneider, & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (Tests und Trends, N.F. Bd. 12, S. 19–40). Göttingen: Hogrefe.
- Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., & McCoach, D. B. (2009). Moving beyond assessment of treatment acceptability: an examination of the factor structure of the Usage Rating Profile—Intervention (URP-I). *School Psychology Quarterly*, *24*, 36–47.
- Cizek, G. J. (2010). An introduction to formative assessment: history, characteristics and challenges. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 3–17). New York: Routledge.
- Clark, I. (2012). Formative assessment: assessment is for self-regulated learning. *Educational Psychology Review*, *24*, 205–249.
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*, 947–967.
- Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: examining child x instruction interactions. *Journal of Educational Psychology*, *96*, 682–698.
- Deno, S. L. (1985). Curriculum-based measurement: the emerging alternative. *Exceptional Children*, *52*, 219–232.
- Dresel, M., & Haugwitz, M. (2008). A computer-base approach to fostering motivation and self-regulated learning. *Journal of Experimental Education*, *77*(1), 3–18.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: the limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research and Evaluation*, *14*(7), 1–11.
- Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, *8*, 128–141.
- Espin, C. A., Wayman, M. M., Deno, S. L., McMaster, K. L., & de Rooij, M. (2017). Data-based decision-making: developing a method for capturing teachers' understanding of CBM graphs. *Learning Disabilities Research and Practice*, *32*, 8–21.
- Euler, D., & Sloane, P. F. E. (1998). Implementation als Problem der Modellversuchsforschung. *Unterrichtswissenschaft*, *26*, 312–326.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9.
- Förster, N., & Souvignier, E. (2014). Learning progress assessment and goal setting: effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction*, *32*, 91–100.
- Förster, N., & Souvignier, E. (2015). Effects of providing teachers with information about their students reading progress. *School Psychology Review*, *44*, 60–75.
- Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*, 188–192.
- Harks, B., Rakoczy, K., Klieme, E., Hattie, J., & Besser, M. (2014). Indirekte und moderierte Effekte von Rückmeldung auf Leistung und Motivation. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmel-*

- dungen: Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 163–194). Münster: Waxmann.
- Harks, B. (2013). *Kompetenzdiagnostik und Rückmeldung – zwei Komponenten formativen Assessments*. Dissertation, Goethe-Universität.
- Hasselhorn, M., Köller, O., Maaz, K., & Zimmer, K. (2014). Implementation wirksamer Handlungskonzepte im Bildungsbereich als Forschungsaufgabe. *Psychologische Rundschau*, 65, 140–149.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Heritage, M. (2007). Formative assessment: what do teachers need to know and do? *Phi Delta Kappan*, 89, 140–145.
- Hondrich, A.L., Hertel, S., Adl-Amini, K., & Klieme, E. (2016). Implementing curriculum-embedded formative assessment in primary school science classrooms. *Assessment in Education: Principles, Policy & Practice*, 23(2), 353–376.
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kaper, D., Lankes, E.-M., McElvany, N., Stubbe, T.C., & Valtin, R. (Hrsg.) (2017). *IGLU 2016. Lesekompetenzen von Grundschulern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Kawohl, E. (2015). *Diagnosebasierte individuelle Leseförderung in der Grundschule*. Hamburg: Kovač.
- Kingston, N., & Nash, B. (2011). Formative assessment: a meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30, 28–37.
- Kluger, A.N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254–284.
- Koch, H., & Spörer, N. (2016). Effekte des Reziproken Lehrens im Vergleich mit einer von Lehrkräften konzipierten Unterrichtseinheit zur Förderung der Lesekompetenz. In M. Philipp & E. Souvignier (Hrsg.), *Implementation von Lesefördermaßnahmen. Perspektiven auf Gelingensbedingungen und Hindernisse* (S. 99–122). Münster: Waxmann.
- Krampen, G. (1987). Differential effects of teacher comments. *Journal of Educational Psychology*, 79, 137–146.
- Lehmann, R.H., Peek, R., & Poerschke, J. (2006). *HAMLET 3-4 – Hamburger Lesetest für 3. und 4. Klassen*. Göttingen: Hogrefe.
- Lenhard, W. (2013). *Leseverständnis und Lesekompetenz. Grundlagen – Diagnostik – Förderung*. Stuttgart: Kohlhammer.
- Little, T.D., Cunningham, W.A., Shahar, G., & Widaman, K.F. (2002). To parcel or not to parcel: exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 151–173.
- Maier, U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, 13(2), 293–308.
- Maier, U. (2014). Computergestützte, formative Leistungsdiagnostik in Primar- und Sekundarschulen. Ein Forschungsüberblick zu Entwicklung, Implementation und Effekten. *Unterrichtswissenschaft*, 42, 69–86.
- McElvany, N. (2008). *Förderung von Lesekompetenz im Kontext der Familie*. Münster: Waxmann.
- Müller, B., & Richter, T. (2017). Förderung hierarchieniedriger Leseprozesse. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 236–251). Weinheim: Juventa.
- Müller, B., Richter, T., Karageorgos, P., Krawietz, S., & Ennemoser, M. (2017). Effects of a syllable-based reading intervention in poor-reading fourth-graders. *Frontiers in Psychology*, 8, 1–9.
- Muthén, L.K., & Muthén, B.O. (2010). *Mplus – Statistical analysis with latent variables. User's guide* (6. Aufl.). Los Angeles: Muthén & Muthén.
- Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H.M. Niegemann, D. Leutner, & R. Brünken (Hrsg.), *Instructional design for multimedia learning* (S. 181–195). Münster: Waxmann.
- NICHD (2000) = National Institute of Child Health and Human Development. (2000). *Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: U.S. Government Printing Office.
- O'Donnell, C.L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78, 33–84.
- Palincsar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117–175.

- Petermann, F. (2004). Implementationsforschung: Grundbegriffe und Konzepte. *Psychologische Rundschau*, 65, 122–128.
- Popham, W.J. (2009). Assessment literacy for teachers: faddish or fundamental? *Theory into Practice*, 48, 4–11.
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., & Bunger, A. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health*, 38, 65–76.
- Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction: grading and feedback in mathematics classrooms. *Zeitschrift für Psychologie*, 216, 111–124.
- Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, 27, 63–73.
- Richter, T., & Christmann, U. (2002). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz. Bedingungen, Dimensionen, Funktionen* (S. 25–58). Weinheim: Juventa.
- Rieser, S., Naumann, A., Decristian, J., Fauth, B., Klieme, E., & Büttner, G. (2016). The connection between teaching and learning: linking teaching quality and metacognitive strategy use in primary school. *British Journal of Educational Psychology*, 68, 526–545.
- Rosebrock, C., Rieckmann, C., Nix, D., & Gold, A. (2010). Förderung der Leseflüssigkeit bei leseschwachen Zwölfjährigen. *Didaktik Deutsch*, 28, 33–58.
- Rosebrock, C., Nix, D., Rieckmann, C., & Gold, A. (2011). *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. Seelze: Klett.
- Samuels, S.J. (1979). The method of repeated readings. *The Reading Teacher*, 32, 403–408.
- Schneider, M.C., & Randel, B. (2010). Research on characteristics of effective professional development programs for enhancing educators' skills in formative assessment. In H.L. Andrade & G.J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 251–276). Abingdon: Routledge.
- Shavelson, R.J. (2008). Guest editor's introduction. *Applied Measurement in Education*, 21, 293–294.
- Shavelson, R.J., Young, D.B., Ayala, C.C., Brandon, P.R., Furtak, E.M., & Ruiz-Primo, M.A. (2008). On the impact of curriculum-embedded formative assessment on learning: a collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21, 295–314.
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189.
- Slavin, R.E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: a best-evidence synthesis. *Review of Educational Research*, 79, 1391–1466.
- Souvignier, E., & Philipp, M. (2016). Implementation – Begrifflichkeiten, Befunde und Herausforderungen. In M. Philipp & E. Souvignier (Hrsg.), *Implementation von Lesefördermaßnahmen. Perspektiven auf Gelingensbedingungen und Hindernisse*. Münster: Waxmann.
- Souvignier, E., Förster, N., & Salaschek, E. (2014). Quop: ein Ansatz internetbasierter Lernverlaufsdagnostik und Testkonzepte für Mathematik und Lesen. In M. Hasselhorn, W. Schneider, & U. Trautwein (Hrsg.), *Lernverlaufsdagnostik* (Test und Trends, N.F. Bd. 12, S. 239–256). Göttingen: Hogrefe.
- Spörer, N., & Glaser, C. (2010). Förderung selbstregulierten Lernens im schulischen Kontext. *Zeitschrift für Pädagogische Psychologie*, 24, 171–175.
- Stecker, P.M., Fuchs, L.S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: review of research. *Psychology in the Schools*, 42, 795–819.
- Stiggins, R. (2010). Essential formative assessment competencies for teachers and school leaders. In H.L. Andrade & G.J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 233–250). New York: Routledge.
- Stutz, F. (2014). Reading motivation, reading amount, and reading comprehension in the early elementary years. Dissertation, Universität Potsdam.
- Topping, K.J., & Fisher, A.M. (2003). Computerised formative assessment of reading comprehension: field trials in the UK. *Journal of Research in Reading*, 26, 267–279.
- Van Keer, H., & Verhaeghe, J.P. (2005). Comparing two teacher development programs for innovating reading comprehension instruction with regard to teachers' experiences and student outcomes. *Teaching and Teacher Education*, 21, 543–562.
- William, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H.L. Andrade & G.J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 18–40). New York: Routledge.

- William, D., & Thompson, M. (2008). Integrating assessment with learning: what will it take to make it work? In C. A. Dwyer (Hrsg.), *The future of assessment: shaping teaching and learning* (S. 53–82). New York: Lawrence Erlbaum.
- Wimmer, H., & Mayringer, H. (2014). *Salzburger Lese-Screening für die Schulstufen 2-9 (SLS 2-9)*. Bern: Huber.
- Zeuch, N., Förster, N., & Souvignier, E. (2017). Assessing teachers' competencies to read and interpret graphs from learning progress assessment. Results from tests and interviews. *Learning Disabilities Research and Practice*, 32, 61–70.