



Formative Diagnostik durch die Onlineplattform LEVUMI

Andreas Mühling · Markus Gebhardt
Kirsten Diehl

Einleitung

Basierend auf den Beschlüssen der UN-Behindertenrechtskonvention (2006) wandelt sich die Bildungslandschaft in Deutschland im Rahmen der Inklusion gerade fundamental – wie etwa im Bildungsbericht (verfügbar unter www.bildungsbericht.de) festgehalten. Kinder mit verschiedensten Formen der Behinderung sollen am Unterricht aller Schulformen teilhaben können. Der Unterricht in einem solchen inklusiven Setting stellt Lehrkräfte und Bildungsorganisationen vor große Herausforderungen. Eine zentrale Aufgabe dabei ist die, durch die notwendigerweise zunehmende Heterogenität, viel stärkere Individualisierung von Unterricht, als es bisher nötig war [10]. Der Unterricht muss nicht nur auf verschiedenste Arten „zugänglich“ – im Sinne von erfahrbar – gehalten werden, er muss auch inhaltlich und methodisch flexibel strukturiert sein, um den unterschiedlichen Ansprüchen und Fähigkeiten der Schüler gerecht werden zu können. Der Einsatz digitaler Medien bietet eine bisher nicht dagewesene Chance für die Unterstützung verschiedenster Formen der körperlichen und intellektuellen Einschränkung [17] und erlaubt damit die Partizipation am Unterricht für ein Spektrum an Personen, das – bei vergleichbar geringem Ressourcenaufwand – mit keiner herkömmlichen Form der Unterrichtsmaterialien erreicht werden kann. Die Individualisierung des Unterrichts basiert auf zwei miteinander verzahnten Aspekten: einerseits der Feststellung des individuellen Leistungsstandes und Förderbedarfs, beispielsweise durch eine formative Bewertung – eine Bewertung

also, die keine Vergleichbarkeit zwischen Subjekten herstellen möchte, sondern eine subjektive, spezifische Erhebung des aktuellen Stands leisten soll (vgl. [16]). Andererseits der tatsächlichen Individualisierung des Unterrichts, bzw. der Umsetzung von Fördermaßnahmen.

Diese beiden Phasen der Individualisierung sollen sich im Sinne eines Spiralmodells beständig wiederholen. Scriven [21], der dafür den Begriff der *formativen Evaluation* prägte, nennt als ein charakteristisches Merkmal formativer Evaluation, dass während eines Entwicklungsprozesses eine laufende Überprüfung dieses Prozesses stattfindet, welche wiederum Einfluss auf die weitere Gestaltung des Prozesses habe.

Maier [15] weist darauf hin, dass eine solche individuelle Anpassung des Unterrichts mit erheblichem organisatorischem und fachlichem Aufwand für die Lehrkräfte verbunden ist. Es muss daher Ziel der fachspezifischen wie allgemeinen Lehr-Lern-Forschung sein, den Lehrkräften praktikable Werkzeuge, Informationen und Hilfestellungen anzubieten. An dieser Stelle kann die Informatik einen wichtigen Beitrag leisten, da beide Aspekte

DOI 10.1007/s00287-017-1069-7
© Springer-Verlag Berlin Heidelberg 2017

Andreas Mühling
Universität zu Kiel, Kiel
E-Mail: am@informatik.uni-kiel.de

Markus Gebhardt
Technische Universität Dortmund, Dortmund
E-Mail: markus.gebhardt@tu-dortmund.de

Kirsten Diehl
Europa Universität Flensburg, Flensburg
E-Mail: kirsten.diehl@uni-flensburg.de

Zusammenfassung

Das rechtzeitig Erkennen von Lernschwierigkeiten stellt eine große Herausforderung im Schulunterricht dar. Insbesondere im inklusiven Unterricht mit Menschen mit Behinderung ist dies der Fall. Hier muss der Unterricht individuell angepasst werden. Durch eine Lernverlaufsmessung können individuelle Förderungen geplant und ihr Erfolg überwacht werden. Dieser Artikel stellt eine Onlineplattform vor, die ein solches Konzept im Unterricht für Lehrkräfte umsetzbar macht und dabei auch einen Austausch zwischen der empirischen Bildungsforschung und dem Alltag der Schulpraxis ermöglicht.

wirksam durch IT-Systeme unterstützt werden kann. Dieser Artikel stellt ein solches Werkzeug für den Einsatz von Lernverlaufsmessung im inklusiven Unterricht vor, das in einer Kooperation aus empirischer Bildungsforschung, Informatik und fachdidaktischer Forschung entsteht. Die Zugänglichkeit wird dabei – weit gefasst – auf drei Ebenen unterstützt:

1. Zugänglichkeit von computerbasierten Leistungstests für einen möglichst großen Kreis an Lernern.
2. Zugänglichkeit von empirisch fundierten Diagnoseinstrumenten und Förderempfehlungen für die Lehrkräfte.
3. Zugänglichkeit von empirischen Messdaten für die Bildungsforschung.

Lernverlaufsmessung

Eine spezielle Form der formativen Bewertung ist die Lernverlaufsdagnostik. Sie soll Fähigkeiten und Fertigkeiten von Schülern in Lernverläufen, also über einen längeren Zeitraum hinweg, erfassen und darstellen und somit den individuellen Lernfortschritt der Schüler an Lehrkräfte rückmelden. Anhand dieser Rückmeldungen über die Leistungsstände und -entwicklungen einzelner Schüler, spezieller Schülergruppen oder auch einer ganzen Klasse erhalten die Lehrkräfte eine wichtige Informationsbasis für pädagogische Entscheidungen [5]. Sie erkennen frühzeitig Lernlücken bzw. erwartungswidrige Leistungen, die den weiteren Lernerfolg

erschweren können und können vor diesem Hintergrund Unterricht und Förderung zeitnah an die Bedürfnisse der Schüler anpassen und modifizieren. Zusätzlich erhält die Lehrkraft Rückmeldung über den Erfolg und die Wirksamkeit ihres Unterrichts und ihrer Förderung [7].

Damit Lernverlaufsdagnostik bzw. allgemein formative Bewertung sinnvoll möglich ist, benötigen die Lehrkräfte standardisierte und empirisch erprobte Messinstrumente für den Lernfortschritt bzw. den aktuellen Lernstand. Derartige Messinstrumente, die das kleinschrittige Messen von Lernfortschritt entlang des Lehrplanes ermöglichen, wurde bereits in den 70-er Jahren unter dem Namen „Curriculum Based Measurement“ (CBM) von Stanley Deno und Kollegen vorgeschlagen [12]. Insbesondere im Bereich des Lesens kann international auf fundierte Werte in Hinblick auf Gütekriterien der CBM-Verfahren verwiesen werden [25]. Inzwischen kann auch auf wissenschaftliche Befunde für die Annahme der Vorzüge im deutschsprachigen Raum verwiesen werden [23].

Bisher werden entsprechende Instrumente oftmals in Form von Testheften von Verlagen eingekauft, beispielweise die „Hamburger Schreibprobe“ des VPM-Verlags. Diese erlauben aufgrund ihrer Konstruktion ein wiederholtes Messen in längeren Abständen, aber kein häufiges Testen oder die Darstellung des Lernverlaufs. Ein optimaler Effekt für Lernverlaufsmessung scheint im Bereich von wöchentlichem bis zweiwöchentlichem Testen zu liegen (vgl. [3]).

Während formative Bewertungen natürlich auch ohne Unterstützung durch Computersysteme möglich und sinnvoll sind, bieten diese doch Möglichkeiten, die über die reine Abnahme von organisatorischem Aufwand hinausgehen. So sind komplexere Formen der Test-Administration wie Paralleltests, zufällige Ziehung von Items und ganz besonders auch adaptives Testen nur computergestützt sinnvoll möglich [15]. Auch ein sofortiges Feedback nach dem Test – das sogar Bezug zu letzten Testergebnissen nimmt – ist nur durch den Einsatz von Computersystemen erreichbar. Der Auswertung von Daten aus Lehr-/Lernsituationen haben sich in den letzten Jahren die Forschungsrichtungen des „Educational Data Mining“ [1] bzw. der „Learning Analytics“ [14] angenommen, in denen die Informatik eine große Rolle spielt.

Die Onlineplattform LEVUMI

Wir verfolgen mit der Plattform drei Ziele:

1. Der Schulpraxis ein praktikables und frei verfügbares Onlineinstrument zur Lernverlaufsmessung – zunächst in den Bereichen Deutsch und Mathematik – anzubieten.
2. Die Forschung zur Lernverlaufsmessung und speziell auch zu ihrer Akzeptanz bei den Lehrkräften voranbringen.
3. Die gesammelten Daten dazu nutzen, die diagnostischen Maßnahmen zu verbessern und gezielt Förderungen zu entwickeln und somit den Lehrkräften Maßnahmen zur Förderung an die Hand zu geben.

Im Sinne eines „whole in one-Pakets“, wie von Hartke, Diehl und Urban [11] vorgeschlagen, soll die Plattform die Lehrkraft unterstützen, Kinder mit sehr unterschiedlichen Lernvoraussetzungen in ihrer Entwicklung angemessen zu fordern und zu fördern. Daher können die Lehrkräfte aus verschiedenen Tests auswählen. LEVUMI ist in der Pilotphase, wird aber aktuell bereits von mehr als 100 Lehrkräften benutzt. Die verfügbaren Tests sind insbesondere für Schüler mit sonderpädagogischem Förderbedarf (SPF) konzipiert. Die Lehrkräfte können zurzeit die Lernverläufe bei der Leseflüssigkeit mit quantitativen und qualitativen Rückmeldun-

gen auswerten und so mehr über die individuellen Lernfortschritte ihrer Schüler erfahren.

Basierend auf den Anforderungen, ist die Umsetzung als Onlineplattform mit Vorteilen für alle beteiligten Akteure verbunden: Die Schule gewinnt durch geringe Anforderungen an die lokale IT-Infrastruktur. Abgesehen von einem funktionierenden Netzzugang ist praktisch keine spezielle Anpassung oder Installation erforderlich. Dies vereinfacht insbesondere auch Unterrichtssituationen, in denen die Lerner mit eigenen Endgeräten am Unterricht teilnehmen („bring your own device“). Die Bedienung und das Aussehen orientiert sich an anderen Webseiten und kommt daher sowohl Lehrkräften wie auch Lernern entgegen, die auf entsprechendes Vorwissen in der Bedienung zurückgreifen können, insbesondere auch bei der Verwendung von assistiven Technologien. Die empirische Lehr-Lernforschung gewinnt durch die Benutzung der vorhandenen Tests einen „ständigen“ Feldzugang, der benutzt werden kann, um die Qualität der Tests zu überprüfen und diese ggf. zu verbessern. Durch diese ökonomischere Verfügbarkeit von Messdaten können einfacher als bisher neue Tests entwickelt und validiert werden. Diese neuen Tests können dann unmittelbar auf der Plattform verfügbar gemacht werden, wodurch die Lehrkräfte wiederum von den empirisch ermittelten Ergebnissen profitieren können. Die Lehrkräfte profitieren auch durch die ständige Verfügbarkeit der Daten z. B.

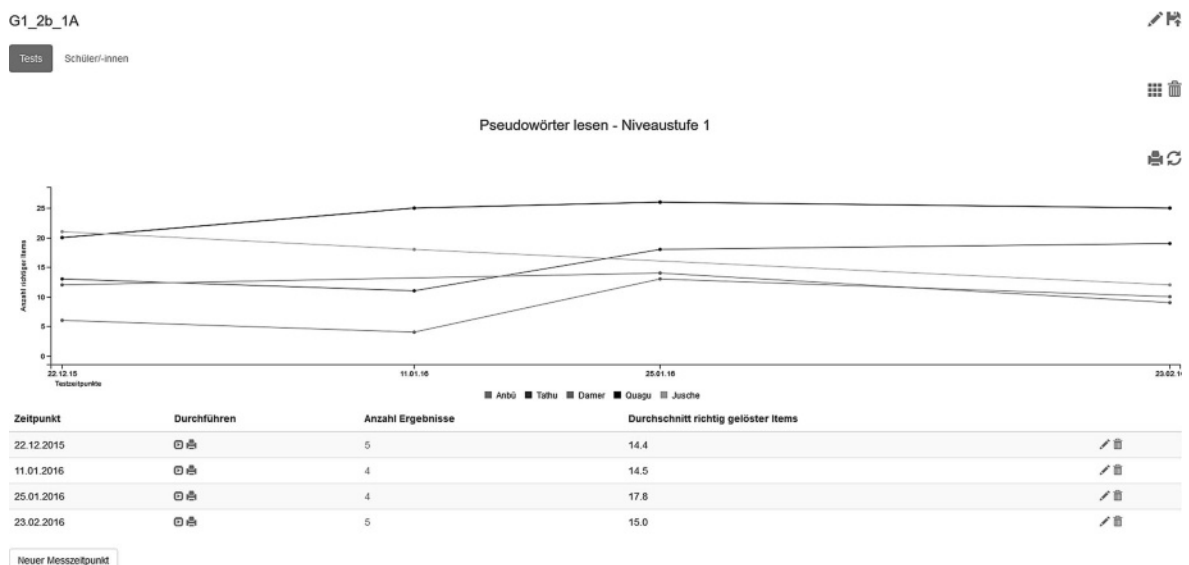


Abb. 1 Klassenansicht von LEVUMI

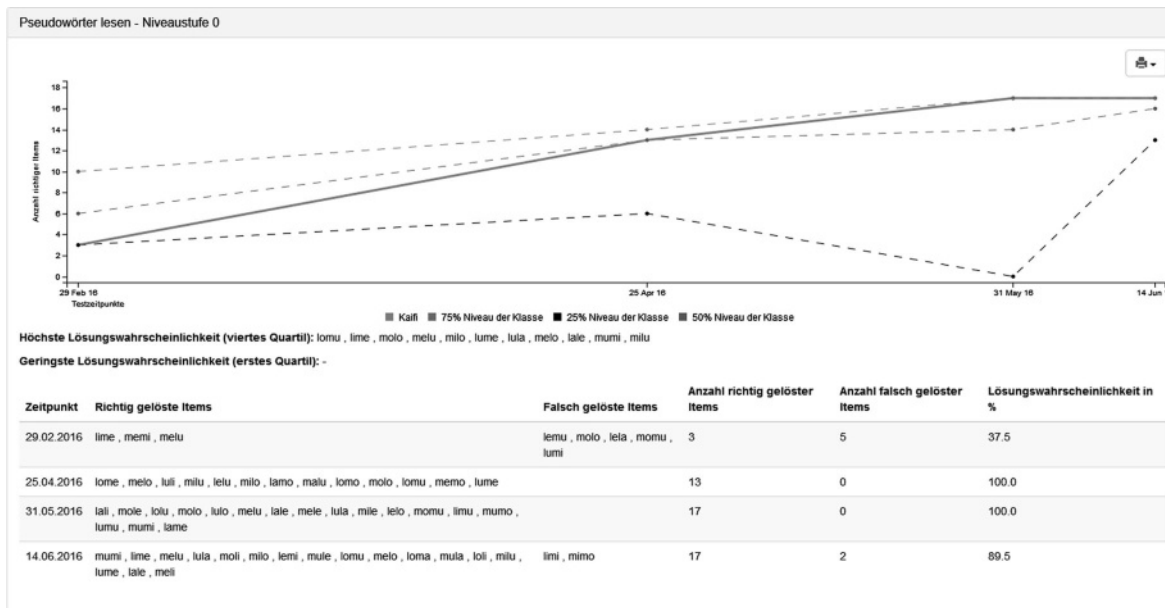


Abb. 2 Schüleransicht von LEVUMI

während der Unterrichtsvorbereitung zu Hause oder während Gesprächen mit Eltern. Es ist auch prinzipiell einfach möglich, einen direkten Zugang für Eltern oder die Lerner selbst zu schaffen, in denen diese auf die Ergebnisse Zugriff haben.

LEVUMI ist ein „Ruby on Rails“-Projekt basierend auf HTML 5 und Javascript. Die Darstellung erfolgt mithilfe des Bootstrap Frameworks und ist auf eine möglichst gute Anbindung von assistiven Technologien ausgelegt, beispielsweise durch die konsequente Verwendung von entsprechenden Metainformationen in den HTML-Dokumenten. Insbesondere wurden auch die Bereiche der Plattform, mit denen Schülern in Kontakt kommen, so gestaltet, dass sie sich im Sinne von „Universal Design“ [20] möglichst gut personalisieren lassen – etwa durch eine Einstellung der Schriftgröße bei den Lesetests.

Abbildung 1 zeigt die Klassenübersicht für einen Test, wie sie der Lehrkraft angezeigt wird. Zu erkennen sind die Lernverläufe aller Schüler für die verschiedenen Messzeitpunkte. Wichtiger für die formative Bewertung und die darauf aufbauende Förderung ist aber die in Abb. 2 dargestellte Schüleransicht. Hier wird der Lernverlauf von nur einem Lerner zusammen mit Quartilen der Klasse als Orientierungspunkt und einer automatischen Auswertung des spezifischen Antwortverhaltens gezeigt.

Die iterative Entwicklung von Tests für LEVUMI

Es gibt zwei typische Herangehensweisen für die Konstruktion von Tests für eine Lernverlaufsdiagnostik: robuste Indikatoren und Curriculum Sampling [8]. Bei Letzterem werden die am Schuljahresende geforderten Kompetenzen in Teilmengen unterteilt und aus vorab definierten Aufgabentypen dann eine Anzahl an ähnlichen „parallelen“ Tests erzeugt [24], die zu geeigneten Zeitpunkten im Unterricht eingesetzt werden können. Die so entstehenden Tests weisen, bei sorgfältiger Auswahl der Aufgaben, eine hohe Inhaltsvalidität im Hinblick auf den Lehrplan als normative Basis auf. Besonders für die Diagnose von basalen Fähigkeiten im Rahmen eines CBM-Ansatzes (Lesen, Rechnen und Schreiben) weist diese Herangehensweise aber einen Nachteil auf: So liegen zum Beispiel die schulischen Leistungen von Schülern mit sonderpädagogischem Förderbedarf „Lernen“ meist ein bis zwei Schuljahre hinter jenen ihrer Klassenkameraden zurück [13]. Nichtsdestotrotz lässt sich auch bei diesen Kindern ein schulischer Lernfortschritt feststellen [4], allerdings werden die Leistungsdifferenzen zu den Schülern ohne Förderbedarf immer größer. Beim Ansatz der robusten Indikatoren werden hingegen Typen von Aufgaben gesucht, die die geforderte Kompetenz möglichst gut repräsentieren und hoch mit den relevanten Leistungen korrelieren. Daher weisen

die so entstehenden Aufgaben eine hohe Kriteriumsvalidität hinsichtlich des externen Kriteriums des Indikators auf. In LEVUMI werden aufgrund der flexibleren Einsetzbarkeit Tests mit robusten Indikatoren bevorzugt, da diese unabhängig von einem konkreten Lehrplan sind. Für die Lesekompetenz hat sich beispielsweise das Messen der Leseflüssigkeit durch ein einminütiges lautes Lesen als ein robuster Indikator gezeigt [18]. Die Aufgaben bestehen in diesem Fall aus einzelnen Wörtern, die laut vorgelesen werden müssen. Die Anzahl der innerhalb einer Minute korrekt gelesenen Wörtern wird als Testwert verwendet. Für den Bereich Mathematik gilt die Einordnung von Zahlen auf dem Zahlenstrahl als robuster Indikator für das Stellenwertsystem.

Nach aktuellem Stand der Wissenschaft sollten diagnostische Tests aufgrund der Überprüfung von Dimensionalität, Messinvarianz und Änderungssensibilität mit Hilfe von Item-Response-Theory (IRT) entwickelt werden [12], bisher existierende Tests im CBM-Kontext werden allerdings überwiegend noch nach der klassischen Testtheorie (KTT) konstruiert [6]. Beim IRT-Ansatz [19] wird ein stochastisches Modell verwendet, das die Lösungswahrscheinlichkeit für eine Aufgabe (Item) in Abhängigkeit einer (oder mehrerer) latenten Fähigkeitsvariable(n) angibt. Diese Fähigkeit kann für einen Lerner dann aus dem Antwortverhalten geschätzt werden. Man muss für einen Test allerdings erst statistisch nachweisen, dass die enthaltenen Items tatsächlich den Annahmen des IRT-Modells genügen. Dazu gehört beispielsweise, dass die Schwierigkeit von Testitems invariant gegenüber Merkmalen abseits der Personenfähigkeit ist. Das heißt beispielweise, dass die Tatsache, ob ein Lerner einen sonderpädagogischen Förderbedarf hat oder nicht, sich nicht derart im Antwortverhalten niederschlagen darf, dass man es nicht mit der Fähigkeit des Lerners alleine erklären kann. Für diese Nachweise benötigt man – im Gegensatz zur klassischen Testtheorie – viele Messungen, diese hängen exponentiell von der Anzahl an Items ab [2].

Da derartig umfangreiche Pilotstudien an Schülern mit hohem ökonomischen Aufwand verbunden sind, gibt es bisher nur wenige solche Tests. Hier ist der Educational-Data-Mining-Ansatz von LEVUMI gut geeignet, um diese Daten nach und nach zu sammeln und stetig weiter, mit Blick auf die Testgüte, auszuwerten. Die ersten Ergebnisse für die bisher vorhandenen Lesetests sind positiv [9]. Hier zeigt sich auch der größte Unterschied zu be-

reits existierenden Plattformen wie quop [22] oder meistercody.de, bei der es sich entweder um reine Serviceangebote für Lehrkräfte – oft von Seiten eines Verlags – oder um interaktive Lernumgebungen für die Kinder handelt.

Weitere geplante Arbeiten

Die weiteren Arbeiten an LEVUMI befassen sich mit drei Aspekten: Erstens, der Erweiterung auf andere Fächer und Tests. Aktuell bietet LEVUMI hauptsächlich Tests zum Lesenlernen an – eine Erweiterung auf andere Fächer, insbesondere Mathematik ist in Vorbereitung. Zweitens, der Erweiterung auf Testformate die Schüler selbstständig am PC durchführen können. Da es bei Lesetests aktuell so ist, dass die Lehrkraft, oder eine Person, während des Tests anwesend ist und die Richtigkeit der vorgelesenen Items überprüft – diese werden während des Tests direkt von der Lehrkraft in LEVUMI eingegeben – ist der Zeitaufwand für einen Test immer noch vergleichsweise hoch. Bei Testformaten, in denen die Schüler auch selbstständig am PC die Tests durchführen können und diese auch gleich automatisch bewertet werden können – etwa bei Diktaten und Rechenaufgaben – verbessert LEVUMI die Testökonomie nochmals beträchtlich. Drittens, der Erweiterung hinsichtlich der diagnostischen Unterstützung, die das System den Lehrkräften bietet. Hier ist zum einen ein adaptives Testen geplant, also die Auswahl der jeweiligen Aufgaben basierend auf den bisherigen Leistungen eines Kindes. Zum anderen sollen durch die Implementierung entsprechender diagnostischer Modell auch detaillierte, automatisch generierte Auswertungen angeboten werden. Dies wird zunächst im Rahmen des oben erwähnten Diktattests umgesetzt werden. In der weiteren Zukunft sind auch intelligente Adaptionen, wie das automatische Auswerten von Lesetests mittels Spracherkennung spannende Erweiterungen aus informatischer Sicht.

Literatur

1. Baker RSJD, Yacef K (2009) The state of educational data mining in 2009: A review and future visions. *J Educ Data Mining* 1(1):3–17
2. Bartholomew DJ, Steele F, Moustaki I, Galbraith JI (2008) *Analysis of Multivariate Social Science Data*, 2nd edn. Chapman & Hall/CRC; CRC Press, Boca Raton
3. Black P, William D (1998) *Assessment and Classroom Learning*. *Assessment Educ Princ Pol Pract* 5(1):7–74
4. Blackorby J, Chorost M, Garza N, Guzman AM (2003) The Academic Performance of Secondary School Students with Disabilities. In: US Department of Education (ed) *The Achievement of Youth with Disabilities during Secondary School. A Report from the National Longitudinal Transition Study*, Bd. 2. SRI International, Menlo Park, CA

5. Blumenthal Y, Kuhlmann K, Hartke B (2014) Diagnostik und Prävention von Lernschwierigkeiten im Aptitude Treatment Interaction- (ATI) und Response to Intervention- (RTI-)Ansatz. In: Hasselhorn M, Schneider W, Trautwein U (Hrsg) Formative Leistungsdiagnostik. Tests & Trends, NF, vol 12. Hogrefe, Göttingen, S 61–82
6. Deno SL (2003) Developments in curriculum-based measurement. *J Spec Educ* 37(3):184–192
7. Förster N, Souvignier E (2014) Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learn Instr* 32(0):91–100
8. Fuchs LS (2004) The past, present, and future of curriculum-based measurement research. *School Psychol Rev* 33(2):188–192
9. Gebhardt M, Diehl K, Mühling A (2016) Online-Lernverlaufsmessung für alle Schülerinnen und Schüler in inklusiven Klassen. *Z Heilpädagog* 66:444–453
10. Gebhardt M, Schwab S, Krammer M, Gasteiger-Klicpera B, Sälzer C (2014) Erfassung von individualisiertem Unterricht in der Sekundarstufe I. Eine Quantitative Überprüfung der Skala „Individualisierter Unterricht“ in zwei Schuluntersuchungen in der Steiermark. *Z Bildungsforsch* 4(3):303–316
11. Hartke B, Diehl K, Vrbán R (2008) Planungshilfen zur schulischen Prävention – Früherkennung und Intervention bei Lern- und Verhaltensproblemen. In: Borchert J, Hartke B, Jogschies P (Hrsg) Frühe Förderungentwicklungsauffälliger Kinder und Jugendlicher. Kohlhammer, Stuttgart, S 218–234
12. Klauer KJ (2014) Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufdiagnostik. In: Hasselhorn M, Trautwein U, Schneider W (Hrsg) Lernverlaufdiagnostik. N.F., Bd 12. Hogrefe, Göttingen
13. Kocaj A, Kuhl P, Kroth AJ, Pant HA, Stanat P (2014) Wo lernen Kinder mit sonderpädagogischem Förderbedarf besser? Ein Vergleich schulischer Kompetenzen zwischen Regel- und Förderschulen in der Primarstufe. *Kölner Z Soziol Soz* 66:165–191
14. Larusson JA, White B (Hrsg) (2014) Computer-supported Collaborative Learning Series. Learning Analytics: From Research to Practice. Springer New York, New York, NY
15. Maier U (2014) Computergestützte, formative Leistungsdiagnostik in Primar- und Sekundarschulen – Ein Forschungsüberblick zu Entwicklung, Implementation und Effekten. *Unterrichtswissenschaft* 42(1):69–86
16. OECD (2005) Formative Assessment: Improving Learning in Secondary Classrooms. OECD Publishing and Centre for Educational Research and Innovation, Paris
17. Office of Educational Technology (2016) Future Ready Learning: Reimagining the Role of Technology in Education. US Office of Education, Washington, DC
18. Reschly AL, Busch TW, Betts J, Deno SL, Long JD (2009) Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *J School Psychol* 47:427–469
19. Rost J (2004) Psychometrische Modelle zur Überprüfung von Bildungsstandards anhand von Kompetenzmodellen. *Z Pädagogik* 50:662–678
20. Russell M, Hoffman T, Higgins J (2009) Meeting the needs of all students: A universal design approach to computer-based testing. *Innov J Online Educ* 5(4):1–8
21. Scriven M (1972) The Methodology of Evaluation. In: Tyler RW, Gagné RM, Scriven M (eds) Perspectives of Curriculum Evaluation, 5th edn. Rand McNally & Company, USA, pp 39–83
22. Souvignier E, Förster N, Salaschek M (2014) quop: ein Ansatz internet-basierter Lernverlaufdiagnostik und Testkonzepte für Mathematik und. In: Hasselhorn M, Schneider W, Trautwein U (Hrsg) Lernverlaufdiagnostik. Test und Trends N.F., Bd. 12. Hogrefe, Göttingen, S 239–256
23. Voß S, Blumenthal Y, Mahlau K, Marten K, Diehl K, Sikora S, Hartke B (2016) Der Response-to-Intervention-Ansatz in der Praxis. Evaluationsergebnisse zum Rügener Inklusionsmodell. Waxmann, Münster
24. Voß S, Hartke B (2014) Curriculumbasierte Messverfahren (CBM) als Methode der formativen Leistungsdiagnostik im RTI-Ansatz. In: Hasselhorn M, Schneider W, Trautwein UU (Hrsg) Formative Leistungsdiagnostik. Tests & Trends, NF, Bd. 12. Hogrefe, Göttingen
25. Wayman MM, Wallace T, Wiley HI, Tichá R, Espin CA (2007) Literature synthesis on curriculum-based measurement in reading. *J Spec Educ* 41:85–120