

Stichwort – Formatives Assessment

Birgit Schütze · Elmar Souvignier · Marcus Hasselhorn

Online publiziert: 29. Juni 2018

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

Zusammenfassung Formatives Assessment gilt als eines der wirksamsten Rahmenkonzepte zur Förderung schulischen Lernens. Es bezeichnet die lernbegleitende Beurteilung von Schülerleistung mit dem Ziel, diagnostische Informationen zu nutzen, um Unterricht und Lernen zu verbessern. Grundlegende Merkmale von formativem Assessment sind die Klärung von Lernzielen, die Diagnose der individuellen Leistung sowie eine darauf basierende Rückmeldung und Förderung. Die Gestaltung von formativem Assessment reicht von spontanem *on-the-fly*-Assessment bis hin zu im Voraus geplantem, formalisiertem und curricular eingebettetem Assessment. Studien untermauern die lernförderliche Wirkung von formativem Assessment, wobei diese von der konkreten Gestaltung abhängt. Obwohl politische, wissenschaftliche und schulische Entwicklungen zur Implementation von formativem Assessment beitragen, ist die Umsetzung nach wie vor herausfordernd. Im vorliegenden Beitrag wird der aktuelle Forschungsstand dargelegt, indem eine begriffliche Bestimmung und Charakterisierung des Konstrukts vorgenommen wird, empirische Befunde zur Wirksamkeit präsentiert sowie implementationsrelevante Entwicklungen und Herausforderungen beschrieben werden.

Schlüsselwörter Formatives Assessment · Leistungsdiagnostik · Rückmeldung · Individuelle Förderung

Dr. B. Schütze (✉) · Prof. Dr. E. Souvignier
Institut für Psychologie in Bildung und Erziehung, Westfälische Wilhelms-Universität Münster,
48149 Fließnerstraße 21, Deutschland
E-Mail: birgit.schuetze@uni-muenster.de

Prof. Dr. E. Souvignier
E-Mail: elmar.souvignier@uni-muenster.de

Prof. Dr. M. Hasselhorn
Abteilung Bildung und Entwicklung, Deutsches Institut für Internationale Pädagogische Forschung,
Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: hasselhorn@dipf.de

Keyword—Formative assessment

Abstract Formative assessment is considered one of the most effective frameworks to foster learning. It is defined as assessment of student performance aimed at providing diagnostic information to improve teaching and learning. Basic characteristics of formative assessment are the clarification of learning goals, the diagnosis of individual performance, as well as the provision of feedback and individual differentiation based on the diagnostic information. The design of formative assessment varies from spontaneous on-the-fly assessment to planned and formalized curricular-embedded assessment. Studies support the positive effect of formative assessment on learning with its impact varying depending on the concrete implementation. Although there are developments in politics, science, and schools contributing to the implementation of formative assessment, transfer into practice remains challenging. The present paper describes the current state of research by defining and characterizing the construct, presenting empirical results on its efficacy, and by addressing developments and challenges associated with its implementation.

Keywords Feedback · Formative assessment · Individual fostering · Performance diagnostics

1 Einleitung

Unter „formativem Assessment“ versteht man die lernprozessbegleitende Beurteilung von Leistungen mit dem Ziel, diese diagnostischen Informationen zu nutzen, um Unterricht und letztlich das individuelle Lernen zu verbessern (z. B. Black und Wiliam 1998a). Eine möglichst gute Kenntnis des individuellen Leistungsstandes und Lernzuwachses – so die Annahme – soll dazu beitragen, Lerngelegenheiten auf die Bedarfe der jeweiligen Lernenden optimal anzupassen. Formativem Assessment wird also ein hohes lernförderliches Potenzial zugeschrieben. Black und Wiliam (1998b) ordnen die Effektgröße von formativem Assessment relativ zu allen bisher untersuchten pädagogischen Interventionen als sehr hoch ein und auch andere Autoren betonen, dass formatives Assessment ein großes Potenzial zur Stimulierung des Lernzuwachses von Schülerinnen und Schülern hat (z. B. Hattie 2009; Cizek 2010; Wylie et al. 2012). Die zunehmende Zahl einschlägiger wissenschaftlicher Arbeiten sowie die Entwicklung von Instrumenten für formatives Assessment (insbesondere im angloamerikanischen Raum) verdeutlichen die großen Hoffnungen, die mit diesem Ansatz verbunden sind (z. B. Andrade und Cizek 2010; McMillan 2010; Bennett 2011).

Auch wenn Konsens über die lernförderliche Wirksamkeit von formativem Assessment besteht, wird weiterer Forschungsbedarf zur Nutzung und erfolgreichen Implementation dieses Ansatzes artikuliert. Zusätzlich werden Fragen nach den vielfältigen „Spielarten“ formativen Assessments laut und auch die Forderung, konkrete Materialien für den praktischen Einsatz zu erarbeiten (vgl. Shavelson 2008; Andrade 2010; Bennett 2011; Kingston und Nash 2011). Der vorliegende Beitrag gibt einen Überblick über den aktuellen Forschungsstand zu formativem Assessment. Hierzu

wird zunächst eine begriffliche Bestimmung und Charakterisierung des Konstrukts vorgenommen. Anschließend werden empirische Befunde zur Wirksamkeit von formativem Assessment allgemein und für verschiedene Spielarten formativen Assessments präsentiert. Schließlich werden implementationsrelevante Entwicklungen in Politik, Forschung und Unterrichtspraxis thematisiert und wird ein zusammenfassendes Fazit gezogen.

2 Was versteht man unter formativem Assessment?

Im Bildungskontext bezeichnet *Assessment* den Prozess, mit dem Hinweise zum Lernstand von Schülerinnen und Schülern erfasst und genutzt werden (Harlen 2008). Diese Erfassung und Nutzung kann auf verschiedenen Ebenen (z. B. System-, Klassen- oder Individualebene) und zu unterschiedlichen Zwecken erfolgen (formativ oder summativ; z. B. Pellegrino et al. 2001). *Formatives Assessment* erfolgt in der Regel auf Klassen- oder Individualebene und hat das Ziel, Lernen und Unterricht zu verbessern mit Hilfe von lernbegleitend erfassten diagnostischen Informationen zur Leistung der Schülerinnen und Schüler (z. B. OECD 2005; Black und Wiliam 1998a, 2009; Heritage 2007; Cizek 2010). Black und Wiliam (2009) betonen die entscheidungsunterstützende Funktion von formativem Assessment: Leistungsinformationen werden von Lehrkräften, Schülerinnen und Schülern oder Peers erhoben, interpretiert und genutzt, um Entscheidungen über die nächsten Schritte im Lehr-Lern-Prozess zu treffen. Diese Entscheidungen sind besser fundiert als solche, die ohne entsprechende Informationen getroffen werden (vgl. den Ansatz des *data-based decision-making*; Ikemoto und Marsh 2007; Mandinach 2012). In der deutschsprachigen Literatur wird *formatives Assessment* auch als *formative Leistungsmessung* (z. B. Maier 2010a) oder *formative Leistungsbeurteilung* (z. B. Klieme et al. 2010) bezeichnet. In der angloamerikanischen Literatur findet man auch die Bezeichnungen *classroom assessment* (z. B. Angelo und Cross 1993; Shepard 2000), *classroom evaluation* (z. B. Crooks 1988), *assessment for learning* (z. B. Broadfoot et al. 1999, 2002; Stiggins 2002), *assessment to assist learning* (z. B. Pellegrino et al. 2001) und *formative evaluation* (Bloom 1969).

Formatives Assessment ist vom sogenannten summativen Assessment (oder *assessment of learning*, Broadfoot et al. 1999; Stiggins 2002) abzugrenzen: Während formatives Assessment (in der Regel unterrichtsbegleitend) durchgeführt wird, um mit Hilfe der erworbenen Informationen Lehr- und Lernprozesse zu verbessern (z. B. Brookhart 2010; Cizek 2010; Maier 2010a; Gikandi et al. 2011), erfolgt summatives Assessment (häufig am Ende einer Unterrichtseinheit) mit dem Ziel, Leistung zusammenfassend zu beurteilen und, darauf aufbauend, beispielsweise Noten zu vergeben und/oder Selektionsentscheidungen zu treffen (z. B. Sadler 1989; OECD 2008; Brookhart 2010; Cizek 2010; Maier 2010a; Gikandi et al. 2011). Der Nutzungszweck der Assessmentinformationen unterscheidet also formatives von summativem Assessment. Hinzu kommt, dass Leistungsbeurteilungen bei formativem Assessment meist während einer Maßnahme erfolgen, während sie bei summativem Assessment häufig erst nach deren Abschluss stattfinden. In diesem Sinne ist also

nicht das Instrument, mit dem ein Assessment durchgeführt wird, formativ oder summativ, sondern die Nutzung des Instruments.

Historisch geht die Differenzierung von formativem und summativem Assessment auf Scrivens (1967) Konzept der formativen und summativen Evaluation von pädagogischen Maßnahmen beziehungsweise Programmen zurück, welche wiederum an Cronbachs Ausführungen zu kriteriumsorientierter (formativer) Evaluation von Lehrveranstaltungen anknüpft (Cronbach 1964, Cronbach selbst verwendete aber noch nicht den Ausdruck *formativ*). Bloom hat Scrivens Konzept der formativen und summativen Evaluation von Programmen auf die Evaluation individueller Schülerleistung übertragen (Bloom 1969) und damit den Grundstein für die heutige Unterscheidung von formativem und summativem Assessment gelegt (z. B. Cizek 2010; Bennett 2011).

2.1 Merkmale von formativem Assessment

Wiliam und Thompson (2008; vgl. auch Bennett 2011) beschreiben fünf Schlüsselmerkmale der formativen Leistungsbeurteilung. Die Ableitung dieser Merkmale erfolgte unter Berücksichtigung der von Ramaprasad (1983) postulierten drei zentralen Prozessdimensionen *Lernstand* („Where the learner is right now“), *Lernziel* („Where the learner is going“) und *Schritte zum Lernziel ermitteln* („How to get there“) sowie der Handelnden im Assessmentprozess, also Lehrkräfte, Peers und Lernende (vgl. Abb. 1):

- a) *Lernziele und Erfolgskriterien klären, teilen und verstehen*: Lernziele und Erfolgskriterien („Where the learner is going“) sollen durch die Lehrkraft individuell, spezifisch und herausfordernd formuliert sowie klar und verständlich kommuniziert werden. Zudem können sich Schülerinnen und Schüler untereinander über die Lernziele und Erfolgskriterien austauschen, um ein tieferes Verständnis dieser zu erlangen.
- b) *Lernstand durch Diskussionen, Fragen und Aufgaben erfassen*: Die Lehrkraft sollte den individuellen Lernstand im Vergleich zum Lernziel erfassen („Where the learner is right now“). Die erhobenen diagnostischen Informationen bilden die Grundlage für alle weiteren Schritte im formativen Assessmentprozess.
- c) *Lernförderliche Rückmeldung geben*: Diagnostische Informationen können durch die Lehrkraft auf zweierlei Art als Rückmeldung genutzt werden, um die Schülerinnen und Schüler dabei zu unterstützen, das Lernziel zu erreichen („How to get there“). Zum einen kann die Lehrkraft die diagnostischen Informationen als Rückmeldung nutzen, um hierauf aufbauend den *Unterricht* an den Lernstand der Schülerinnen und Schüler *anzupassen*. Eine solche Adaption kann beispielsweise in der fähigkeitsbezogenen Anpassung von Instruktions- und Aufgabenmaterial bestehen – also in der Verwendung unterschiedlicher Arbeitsblätter oder Aufgabenstellungen für unterschiedliche Schülerinnen und Schüler (Klieme und Warwas 2011). Zum anderen kann die Lehrkraft den Schülerinnen und Schülern die diagnostischen Informationen individuell *rückmelden*, so dass diese ihren Lernprozess optimieren können (zur Wirkung von Rückmeldung und zur Gestaltung von lern-

	Lernziel	Lernstand	Schritte zum Lernziel ermitteln
Lehrer	Lernziele und Erfolgskriterien klären, teilen und verstehen	Lernstand durch Diskussionen, Fragen und Aufgaben erfassen	Lernförderliche Rückmeldung geben
Peer		Schüler als instruktionale Ressourcen füreinander aktivieren	
Lernender		Schüler als Verantwortliche des eignen Lernens aktivieren	

Abb. 1 Merkmale formativen Assessments nach Wiliam und Thompson (2008)

förderlicher Rückmeldung s. z. B. Kluger und DeNisi 1996; Hattie und Timperley 2007; Shute 2008).

- d) *Schülerinnen und Schüler als Verantwortliche des eignen Lernens aktivieren*: Schülerinnen und Schüler sollten befähigt werden, die eigene Arbeit selbst zu beurteilen („*Where the learner is right now*“) und den eigenen Lernprozess zu steuern, um das Lernziel zu erreichen („*How to get there*“). Im Sinne eines Selbst-Assessments können Schülerinnen und Schüler also zur eigenen Rückmeldequelle werden (Sadler 1989; s. hierzu auch Butler und Winne 1995). Effekte von Selbst-Assessment werden bei Panadero et al. (2016) beschrieben; Beispiele zur konkreten Gestaltung finden sich unter anderem bei Maier et al. (2012).
- e) *Schülerinnen und Schüler als instruktionale Ressourcen füreinander aktivieren*: Peer-Assessment ermöglicht die Aktivierung von Schülerinnen und Schülern als wechselseitige instruktionale Ressource: Schülerinnen und Schüler beurteilen die Leistung der anderen („*Where the learner is right now*“), geben sich Rückmeldung, unterstützen sich gegenseitig („*How to get there*“), arbeiten und diskutieren miteinander (z. B. Strijbos und Sluijsmans 2010). Hinweise zur Gestaltung von Peer-Assessment gibt zum Beispiel Topping (2010), positive Leistungseffekte von Peer-Assessment werden unter anderem bei Sanchez et al. (2017) beschrieben.

Formatives Assessment sollte nicht auf eines dieser Merkmale (wie zum Beispiel Leistungsdiagnostik) reduziert werden; insbesondere die Komponente der Rückmeldung muss immer mitgedacht werden (für einen Überblick s. z. B. Harks 2013). Auch wenn der Begriff formatives Assessment oftmals bereits zur Bezeichnung eines Diagnostikinstrumentes verwendet wird, ist darauf hinzuweisen, dass Diagnostik und Feedback einerseits sowie andererseits Lehrkräfte und Schülerinnen und Schüler als Agierende notwendig sind, damit formatives Assessment realisiert wird. Diese Einordnung mag trivial erscheinen, ist jedoch zum Verständnis des Konzepts fundamental, zumal in manchen Beiträgen zum formativen Assessment spezifische Merkmale (z. B. nur die Diagnostikkomponente) oder Adressaten (z. B. nur die Schüler und Schülerinnen) isoliert betrachtet werden (Kingston und Nash 2011), so dass fälschlich der Eindruck entstehen kann, das Konzept ließe sich auf eine seiner Facetten reduzieren.

2.2 Förderung im Rahmen formativen Assessments

Während die Merkmale der Leistungsbeurteilung und der Rückmeldung in der einschlägigen Literatur vergleichsweise umfassend diskutiert werden, wird der Aspekt von Maßnahmen zur Förderung eher vernachlässigt. Möglicherweise vor dem Hintergrund jeweils sehr spezifischer Lerngegenstände und Unterrichtsszenarien wird die Ausgestaltung von Förderung als notwendige Facette formativen Assessments häufig eher knapp behandelt. Allgemeine Förderprinzipien wie die (individuelle) Anpassung instruktionaler Materialien werden zwar beschrieben, die konkrete Ausgestaltung der Förderung bleibt aber inhalts- und kompetenzspezifisch und lässt sich kaum zusammenfassend charakterisieren. Inhaltsspezifische Förderhinweise finden sich vor allem in der fachdidaktischen bzw. -bezogenen Literatur, oftmals ohne expliziten Bezug zu formativem Assessment. An dieser Stelle soll eine allgemeine Einordnung zur Frage, in welcher lerntheoretischen Tradition Fördermaßnahmen im Kontext von formativem Assessment stehen, vorgenommen werden. Laut Shepard (2000) sollten Interventionen im Rahmen formativen Assessments idealerweise auf sozial-konstruktivistischen Prinzipien basieren. Der sozial-konstruktivistische Ansatz verknüpft die konstruktivistische Perspektive (Lernen ist ein aktiver konstruktivistischer Prozess) mit kognitionspsychologischen Annahmen (u. a. Lernen baut auf bestehenden Wissensstrukturen auf) und soziokulturellen Prinzipien (Lernen resultiert aus sozialen Interaktionen). Tatsächlich werden bei formativem Assessment häufig kognitivistische und konstruktivistische Grundsätze umgesetzt (James 2006). Die Bedeutung der sozialen Dimension wird dabei oftmals betont. Explizit wird in verschiedenen aktuellen Arbeiten zum formativen Assessment auf Vygotskys sozial-konstruktivistisches Konzept der Zone der nächsten Entwicklung Bezug genommen (z. B. Heritage 2007; Black und Wiliam 2009; Clark 2012). Die Zone der nächsten Entwicklung bezeichnet den Entwicklungs- beziehungsweise Lernbereich eines Individuums, in dem bedeutende nächste Lernfortschritte möglich sind, vorausgesetzt ein Tutor (Erwachsener oder erfahrener Peer) unterstützt den Lernenden. Durch formatives Assessment können Entwicklungs- beziehungsweise Lernspielräume identifiziert und kann der Lernende in seinen Entwicklungs- und Lernprozessen gefördert werden.

3 Was ist zur Wirksamkeit formativen Assessments bekannt?

Formatives Assessment wurde als eines der wirksamsten Instrumente zur Optimierung schulischen Lernens identifiziert (Black und Wiliam 1998a). Bezugnehmend auf Übersichtsartikel von Natriello (1987) und Crooks (1988), eine Metaanalyse von Fuchs und Fuchs (1986) sowie aufbauend auf Ergebnissen zahlreicher Einzelstudien (insgesamt wurden 250 Publikationen berücksichtigt, die zum großen Teil ab dem Jahr 1988 veröffentlicht wurden) kommen Black und Wiliam (1998a) zu dem Schluss, dass die mit formativem Assessment zusammenhängenden Effekte auf die Leistung von Schülerinnen und Schülern zwischen $d=0,40$ und $d=0,70$ liegen. Neuere Arbeiten haben jedoch darauf hingewiesen, dass die empirische Grundlage der bei Black und Wiliam berichteten Effektgrößen unklar bleibt (z. B. Bennett

2011; Kingston und Nash 2011) und verweisen auf methodische Mängel in berichteten Studien (Dunn und Mulvenon 2009; Kingston und Nash 2011). Weiterhin wird kritisiert, dass Black und Wiliam relativ allgemeine Schlussfolgerungen über formatives Assessment auf der Grundlage von Studien ableiten, die sich mit unterschiedlichen Aspekten formativen Assessments beschäftigen (z. B. Assessment durch Lehrkräfte, z. B. Bergan et al. 1991; Selbstmonitoring, z. B. Fontana und Fernandes 1994; Peer-Assessment, z. B. Higgins et al. 1994) beziehungsweise sich nur auf bestimmte Schülergruppen (z. B. leistungsschwache Schülerinnen und Schüler; betrifft die Metaanalyse von Fuchs und Fuchs 1986) konzentrieren (z. B. Bennett 2011).

Dunn und Mulvenon (2009) haben auf der Grundlage von neun Studien aus den Jahren 1999 bis 2007 ein weiteres Review verfasst. Auch die hier zusammengefassten Studien beschäftigen sich mit unterschiedlichen Varianten von formativem Assessment (von Diskussionen im Klassenraum, Ruiz-Primo und Furtak 2006, bis hin zu webbasiertem Selbst-Assessment, Velan et al. 2002). Der Überblicksartikel bestätigt die insgesamt positive Einschätzung der Wirkung von formativem Assessment auf Schülerleistung, identifiziert allerdings erheblichen weiteren Forschungsbedarf.

Kingston und Nash (2011, 2015) führten eine Metaanalyse basierend auf 13 Studien aus den Jahren 1990 bis 2010 durch, in der ebenfalls unterschiedliche Varianten formativen Assessments (von Diskussionen im Klassenraum, Ruiz-Primo und Furtak 2006, bis hin zu computerbasierten Testverfahren, z. B. Poggio et al. 2007) berücksichtigt wurden. Die Autoren berichten einen kleinen bis mittleren positiven Effekt von formativem Assessment auf Leistung ($d=0,20/0,25$). Die Ergebnisse einer ebenfalls durchgeführten Moderatoranalyse sprechen dafür, dass die Wirksamkeit von formativem Assessment abhängig ist vom Schulfach sowie von der konkreten Implementation des formativen Assessments.

Eine aktuelle Literaturübersicht von McLaughlin und Yan (2017) konzentriert sich auf die Wirksamkeit von speziell *onlinebasiertem formativem Assessment* und kommt unter Berücksichtigung von 75 Studien aus den Jahren 1998 bis 2016 zu dem Schluss, dass eine positive Wirkung auf Schülerleistung und Selbstregulation nachgewiesen ist. Unter onlinebasiertem Assessment werden hier Verfahren wie *Multiple-Choice-Tests* (z. B. Klecker, 2003), *One-Minute-Papers* (kurze Reflexionen, z. B. Vonderwell 1998), *E-portfolios* (digitale Portfolios, z. B. Lin 2008), *Student-Response-Systeme* (elektronische Abstimmungssysteme, z. B. Dervan 2014) oder *Web2.0-Instrumente* (wie Blogs, Wikis etc., z. B. Chen und Chen 2012) zusammengefasst, die in diversen pädagogischen Kontexten (von Kindergarten über Primar- und Sekundarstufe bis zur Hochschule) formativ genutzt wurden.

Zusammengefasst lässt sich festhalten, dass formatives Assessment nachweislich die Leistungen von Schülerinnen und Schülern positiv beeinflussen kann. Allerdings scheint die Wirksamkeit von der Art und Weise der konkreten Umsetzung der Leistungsbeurteilung abhängig zu sein. Den Spielarten zur Umsetzung formativen Assessment kommt also eine entscheidende Rolle zu.

4 Spielarten der Umsetzung formativen Assessments

Es gibt verschiedene Kriterien, um Arten formativen Assessments zu klassifizieren. Ein Kriterium ist die *Spezifität der Assessmentinformation* (Wiliam 2010): Formatives Assessment kann Schülerinnen und Schüler identifizieren, die Probleme haben (*monitoring assessment*), zusätzlich konkrete Probleme lokalisieren (*diagnostic assessment*) oder darüber hinaus Informationen (z. B. über Fehlvorstellungen) liefern, die konkret zur Überwindung der Probleme beitragen (*instructional assessment*).

Black und Wiliam (2009) unterscheiden Arten formativen Assessments nach dem *Zeitraum*, der zwischen dem Assessment und dessen instruktionaler Nutzung liegt. Sie differenzieren *asynchronous moments of contingency* (z. B. Planung einer nachfolgenden Stunde auf Grundlage der bearbeiteten Haus- oder Schulaufgaben) und *synchronous moments of contingency* (z. B. Adaptionen im Verlauf einer Diskussion; für vergleichbare Klassifikationen s. auch Wiliam 2010; Clark 2012).

Ein weiteres (verwandtes) Klassifikationsmerkmal ist der *Planungs- und Formalitätsgrad* des Assessments (s. z. B. Cowie und Bell 1999; Heritage 2007; Shavelson et al. 2008; Baldwin und Yun 2012). Eine in diesem Kontext relativ verbreitete Unterscheidung wird von Heritage (2007) beziehungsweise Shavelson et al. (2008) beschrieben. Diese differenzieren *on-the-fly*, *planned-for-interaction* und *curriculum-embedded assessment*. Bei *on-the-fly assessment* handelt es sich um Unterrichtssequenzen (z. B. im Rahmen von Beobachtungen, offenen Fragen, Klassen-, Gruppen- oder Einzeldiskussionen), in denen Lehrkräfte spontan Fehlkonzeptionen und Verständnislücken identifizieren, Rückmeldung geben und Unterrichtsinhalte entsprechend adaptieren (für eine Studie zur positiven Wirkung von *on-the-fly assessment* s. Ruiz-Primo und Furtak 2007). *Planned-for-interaction assessment* beinhaltet geplante Unterrichtssequenzen (z. B. Frage-Antwort-Sequenzen), die diagnostische Informationen über den Lernstand der Schülerinnen und Schüler liefern (z. B. die Methode des *strukturierten Assessmentdialogs* im ASSIST-ME-Projekt, s. Grob et al. 2014). *Curriculum-embedded assessment* ist vergleichsweise stark formalisiert. Es kann dabei sowohl Bestandteil der herkömmlichen Unterrichtsaktivität sein, kann aber auch durch zusätzliche Assessmentinstrumente an Schlüsselstellen in das Curriculum eingebettet werden. Insbesondere für den Bereich des curricular eingebetteten Assessments gibt es viele mögliche Arten der Umsetzung. Diese werden im nachfolgenden Abschnitt genauer betrachtet.

4.1 Beispiele für curricular eingebettetes Assessment

Beispiele für curricular eingebettete Instrumente für das Assessment, die zugleich integraler Bestandteil des herkömmlichen Unterrichts sind, sind Hausaufgaben (z. B. Strandberg 2013) oder Präsentationen (z. B. Preiser 2000). Bei zusätzlichen Instrumenten für das Assessment kann es sich um einzelne diagnostische Aufgaben oder ganze Testverfahren bzw. -batterien handeln. Im Unterschied zu dem ökonomischeren Einsatz einzelner Diagnoseaufgaben ermöglicht die Nutzung ganzer Testverfahren eine datenbasierte Entscheidungsfindung (z. B. Mandinach 2012) auf der Basis einer psychometrisch fundierten Messung. Aufgaben und Tests können papierbasiert oder computerbasiert sein. Während papierbasierte Verfahren vielen Lehrkräften

vertrauter sein dürften und es zunächst einfacher erscheint diese in den gewohnten Unterrichtsablauf zu integrieren (s. hierzu Leahy et al. 2005), ermöglichen computerbasierte Verfahren eine effiziente und psychometrisch fundierte Diagnostik sowie eine zeitnahe automatisierte Auswertung und Erstellung individualisierter Rückmeldungen (z. B. Koeppen et al. 2008; Russel 2010).

Diagnoseaufgaben werden beispielsweise *papierbasiert* in Form von sogenannten Diagnosebögen an Schlüsselstellen im Curriculum eingesetzt und mit einer anschließenden Leistungsrückmeldung sowie ggf. mit einer instruktionalen Anpassung kombiniert. Lernförderliche Effekte einer solchen formativen Nutzung einzelner diagnostischer Aufgaben zeigten sich für den Mathematikunterricht der Sekundarstufe (s. Rakoczy et al. 2017) und den Sachkundeunterricht der Primarstufe (z. B. Decristan et al. 2015; s. a. Shavelson et al. 2008). Eine *computerbasierte* Darbietung einzelner diagnostischer Aufgaben kann mit Hilfe sogenannter *Student-Response-Systeme* realisiert werden. Hierbei können die Schülerinnen und Schüler per Eingabegerät auf eine Frage der Lehrkraft antworten, indem sie zwischen verschiedenen Antwortalternativen wählen. Die Verteilung der Antworten wird unmittelbar auf dem Computer bzw. Smartphone der Lehrkraft graphisch dargestellt, so dass diese direkt reagieren kann (z. B. Beatty und Gerace 2009). Die Nutzung solcher Systeme kann sich positiv auf die Motivation der Lernenden auswirken (McLaughlin und Yan 2017).

Die formative Nutzung ganzer *Testverfahren* kann unter anderem im Rahmen von *Lernverlaufsdagnostik* erfolgen. Diese kann *papierbasiert* (z. B. Walter 2010, 2013; Strathmann und Klauer 2012) oder *computerbasiert* sein (z. B. quop; für einen Überblick s. Souvignier et al. 2014a, b). Aufbauend auf dem Ansatz des *curriculum-based measurements* (CBM; vgl. Deno 1985; Fuchs 2017), bezeichnet Lernverlaufsdagnostik eine in kurzen Zeitintervallen wiederholte standardisierte, psychometrisch fundierte Leistungsdiagnostik mit Paralleltests zur Begleitung von Lernprozessen, die Lehrkräften eine Rückmeldung über die Leistungsentwicklung ihrer Schülerinnen und Schüler und somit über den Erfolg ihrer Fördermaßnahmen liefert. Ein positiver Effekt von Lernverlaufsdagnostik auf die Leistungsentwicklung ist belegt (Stecker et al. 2005), wobei die Unterstützung der Lehrkräfte bei der Nutzung der diagnostischen Informationen (z. B. durch Hinweise zur Unterrichtsplanung) wesentlich für die Wirksamkeit zu sein scheint. Beispiele für weitere computerbasierte Testverfahren, die Lehrkräften ebenfalls eine zusammenfassende Rückmeldung über die Leistungen ihrer Schülerinnen und Schüler anbieten, nicht aber dem Ansatz des CBM folgen, sind *Accelerated Math* (z. B. Ysseldyke und Bolt 2007) oder *e-asTTle* (s. Visible Learning Lab 2010).

Neben den bislang beschriebenen Testverfahren gibt es computerbasierte Assessments, die einen stärker interaktiven, tutoriellen Charakter aufweisen (vgl. Bennett 2002; VanLehn 2011). So erhalten die Lernenden bei der sogenannten *computer aided-instruction* (CAI) nicht erst nach Bearbeitung des gesamten Tests eine zusammenfassende Rückmeldung, sondern bekommen nach der Lösung jeder einzelnen Testaufgabe Rückmeldung und ggf. Hinweise zu ihrer Antwort (z. B. Van Lehn 2011; Maier et al. 2016; Faber et al. 2017).

Einen noch höheren Interaktionsgrad als CAI weisen sogenannte *Intelligente Tutorielle Systeme* (ITS) auf. Diese ermöglichen die Auswertung einzelner Schritte innerhalb des Lösungsprozesses, Rückmeldung zu einzelnen Lösungsschritten, kon-

texspezifische Hinweise für den nächsten Schritt (üblicherweise auf Anfrage des Schülers bzw. der Schülerin) sowie eine individualisierte Aufgabenauswahl (z. B. Koedinger und Alevan 2007; Alevan et al. 2010; VanLehn 2011). Metaanalysen und Reviews kommen zu dem Ergebnis, dass ITS ähnlich leistungsförderlich sind wie menschliches Tutoring, aber leistungsförderlicher als CAI (Van Lehn 2011; Ma et al. 2014; Steenbergen-Hu und Cooper 2014).

Der Vollständigkeit halber soll erwähnt werden, dass neben den hier beschriebenen Aufgaben- und Testvarianten auch *alternative Assessmentinstrumente* wie Lerntagebucheinträge, Portfolios oder sogenannte *One-minute-papers* (papier- oder computerbasiert) im Sinne eines *curriculum-embedded assessments* formativ genutzt werden können (für einen Überblick zur Wirksamkeit s. z. B. McLaughlin und Yan 2017).

Die dargestellten Verfahren sind nicht fachspezifisch und können jeweils in unterschiedlichen Lernbereichen angewandt werden. Jedoch eignen sich zur Erfassung bestimmter Kompetenz- oder Wissensaspekte manche Assessmentverfahren besser als andere. Während Vokabelwissen beispielsweise mit einem papierbasierten Test geprüft werden kann, lassen sich sportmotorische Kompetenzen eher durch andere Verfahren (wie z. B. Beobachtungen) erfassen (z. B. Maier et al. 2012).¹

4.2 Nutzung des Konzepts formativen Assessments auf Klassenebene und jenseits des Leistungskontexts

Formatives Assessment kann auf unterschiedlichen *Ebenen* ansetzen. Die oben beschriebenen Arten formativen Assessments erfolgen oftmals auf der Individualebene bzw. der Individual- und Klassenebene. Primär auf Klassenebene liefern beispielsweise die auf Bildungsstandards basierenden Vergleichsarbeiten (VERA, z. B. Hosenfeld und Zimmer-Müller 2009) Informationen über den Leistungsstand von Klassen, die die jeweiligen Lehrkräfte formativ nutzen können, um ihren Unterricht zu optimieren (für eine kritische Diskussion der formativen Nutzbarkeit der VERA-Ergebnisse vgl. Maier 2010b). Dass eine produktive Nutzung der Daten aus Vergleichsarbeiten nicht selbstverständlich ist, zeigte sich auch in einer Studie von Staman et al. (2017), bei der trotz ausführlicher Lehrerfortbildungsmaßnahmen keine positiven Effekte der halbjährlichen Bereitstellung von Informationen zur Kompetenzentwicklung festgestellt werden konnten. Generell gilt, dass ein Testverfahren, das mit dem Ziel einer Optimierung des Unterrichts auf der Ebene einer Klasse eingesetzt wird, ein Mittel formativen Assessments ist. Im Unterschied dazu bezieht sich formatives Assessment auf Individualebene speziell auf die Leistung einzelner Schülerinnen und Schüler. Auch wenn die beiden Ebenen theoretisch trennbar sind, kann ein und dieselbe Vorgehensweise formativen Assessments beide Ebenen einbeziehen.

¹ Es wird deutlich, dass die Klassifikation formativen Assessments eng orientiert ist an der Unterscheidung von Diagnosemöglichkeiten (z. B. Beobachtung vs. Testung). Sie geht darüber hinaus aber auch, zumindest in Teilen, mit verschiedenen Ausprägung weiterer Merkmale formativen Assessments einher (z. B. spontane mündliche vs. schriftliche Rückmeldung).

Unabhängig von der Assessmentebene können im Rahmen von formativem Assessment unterschiedliche *Objekte* beurteilt werden. Die oben beschriebenen Arten zielen gemäß der Definition von formativem Assessment (s. Abschn. 2) durchweg auf die Erfassung von Schülerleistung ab. Grundsätzlich könnten darüber hinaus aber diverse Lern- und Unterrichtsaspekte beurteilt und diese diagnostische Information dann formativ genutzt werden. Schülerinnen und Schüler können beispielsweise die Unterrichtsqualität ihrer Lehrkräfte einschätzen und ihnen hierüber Rückmeldung geben, die dann wiederum genutzt werden kann, um den Unterricht zu optimieren. Hierzu gibt es informelle Assessmentmethoden (wie das Blitzlicht, s. Bastian et al. 2003), aber auch standardisierte Assessmentinstrumente (wie SefU oder EMU, z. B. Helmke und Lenske 2013).

5 Entwicklungen zu formativem Assessment in Politik, Forschung und Unterrichtspraxis

Die Grundideen formativen Assessments sind bereits in der Reformpädagogik (insbesondere in den Arbeiten von Freinet, Kerschensteiner, Montessori und Steiner) verankert. Alternative pädagogische Konzepte erlebten in Deutschland in den 1960er-Jahren einen neuen Aufschwung und gingen unter anderem mit der Kritik an Schulnoten, Forderungen nach individualisierter Rückmeldung, prozessorientierter (statt produktorientierter) Diagnostik und motivierendem Unterricht einher. In der Folge wurden alternative Formen von Assessment wie zum Beispiel Portfolios, Wochenarbeitspläne, Lerntagebücher, Lernberichte oder Diagnosebögen entwickelt (Köller 2005). Diagnosebögen zielen beispielsweise darauf ab, Lernergebnisse zu erfassen, differenzierte Rückmeldeinformationen für Schülerinnen und Schüler und Eltern bereitzustellen, den individuellen Lernfortschritt abzubilden und den Wissenserwerb zu unterstützen (Winter 1991). In den 1980er-Jahren verschwanden Diagnosebögen wieder aus den deutschen Klassenräumen (Ingenkamp 1985). Obwohl alternative Formen von Assessment dem formativen Grundgedanken nahe sind, können sie grundsätzlich nicht mit formativem Assessment gleichgesetzt werden. Wie traditionelle Testverfahren auch, können sie sowohl summativ als auch formativ genutzt werden (Maier 2010a).

In jüngerer Zeit erfährt das Konzept des formativen Assessments in Deutschland einen Aufschwung (Maier 2010a). Sowohl auf politischer Ebene als auch in der Forschung zeigen sich vermehrt Ansätze, die langfristig zu einer Umsetzung von formativem Assessment im Unterricht beitragen können. Hierzu zählen beispielsweise Bemühungen, Unterricht kompetenzorientierter zu gestalten. Auch ist eine wachsende Auseinandersetzung mit theoretischen und methodischen Herausforderungen von Kompetenzmessung zu verzeichnen (vgl. Koeppen et al. 2008). In der schulischen Praxis ist es die zunehmende Beachtung des Anspruchs auf individuelle Förderung, die die Relevanz formativen Assessments unterstreicht. Mit den zentralen Elementen von Assessment, Feedback und Förderung stellt formatives Assessment letztlich eine Variante individueller Förderung dar (vgl. Hasselhorn et al. 2018). Auch die Verankerung von „Diagnose und Förderung individueller Lernprozesse; Leistungsmessungen und Leistungsbeurteilungen“ (KMK 2004, S. 5) als curricula-

rer Schwerpunkt in den Standards für die Lehrerbildung weist darauf hin, dass die Bedeutung diagnostischer Information als Grundlage für unterrichtliche Entscheidungen wahrgenommen wird. Mittlerweile werden im Rahmen der universitären Lehrerausbildung und im Referendariat zunehmend Ansätze zur Implementation von Diagnostik und Förderung vermittelt (vgl. Bauch 2010; Bürgermeister et al. 2011).

Die Entwicklung von Instrumenten für formatives Assessment (wie quop, z. B. Souvignier et al. 2014a, b oder *Accelerated Math*, z. B. Lehmann und Seeber 2005) sowie die Durchführung von Forschungsprojekten zur Gestaltung und Wirkung von formativem Assessment tragen maßgeblich zur Umsetzung entsprechender Maßnahmen im Unterricht bei. Beispiele für Forschungsprojekte in Deutschland sind die Projekte *Conditions and Consequences of Classroom Assessment* (Co²CA, z. B. Rakoczy et al. 2017), *Individuelle Förderung und adaptive Lern-Gelegenheiten in der Grundschule* (IGEL, z. B. Decristan et al. 2015), *Assess Inquiry in Science, Technology and Mathematics Education* (ASSIST-ME, z. B. Bernholt et al. 2013) oder das quop-Projekt (z. B. Souvignier et al. 2014a, b). Darüber hinaus bestehen in den Fachdidaktiken vielfältige Ansätze zur Gestaltung von Elementen formativen Assessments (z. B. zur Gestaltung fachspezifischer diagnostischer Instrumente oder Fördermaßnahmen), welche oftmals aber nicht explizit als formatives Assessment ausgewiesen werden.

Obwohl es also durchaus Ansätze zur Entwicklung und Evaluation formativ nutzbarer Assessmentinstrumente gibt, werden standardisierte Verfahren im regulären Unterricht nach wie vor relativ selten eingesetzt. Maier (2011) befragte Gymnasiallehrkräfte (der Sekundarstufe I) in Bayern, Thüringen und Baden-Württemberg zu ihrer Nutzung von formativer Leistungsdiagnostik im Deutsch-, Mathematik- und Fremdsprachenunterricht. Die Befragung ergab, dass schriftliche (z. B. unbenotete Kurztests) und standardisierte Diagnoseverfahren deutlich seltener formativ genutzt werden als informelle und im Unterricht einfach zu realisierende Methoden der Leistungsbeurteilung wie zum Beispiel gezielte Beobachtungen.

6 Fazit

Formatives Assessment ist ein vielversprechendes Konzept, dem ein großes lernförderliches Potenzial zugeschrieben wird. Es basiert auf der lernbegleitenden Erfassung und Nutzung diagnostischer Informationen zur Optimierung von Lehr- und Lernprozessen. Während formatives Assessment in dem Übersichtsbeitrag von Black und Wiliam (1998a, b) als einer der bedeutendsten Ansätze zur Verbesserung schulischen Lernens hervorgehoben wird, zeichnen aktuelle Überblicksarbeiten und Metaanalysen ein differenzierteres Bild der Wirksamkeit (Dunn und Mulvenon 2009; Kingston und Nash 2011). Zur Umsetzung von formativem Assessment gehören Festlegungen unterschiedlicher Assessmentmerkmale (Zielklärung, Diagnostik, Rückmeldung, Selbst-Assessment, Peer-Assessment, s. Wiliam und Thompson 2008). Dafür stehen unterschiedliche Spielarten zur Verfügung, die von *on-the-fly assessment* bis hin zu formalisiertem *curriculum-embedded assessment* (z. B. mit Hilfe von computerbasierten Testbatterien) reichen.

In Deutschland zeichnen sich zunehmend Tendenzen ab, die langfristig zu einer Umsetzung von formativem Assessment im Unterricht beitragen können. Hierzu zählen unter anderem Bemühungen, Unterricht kompetenzorientierter zu gestalten (vgl. Koeppen et al. 2008), die Definition von Diagnose und Förderung als curricularer Schwerpunkt in der Lehrerbildung (KMK 2004), die Veröffentlichung einschlägiger, praxisorientierter Lehrerhandbücher (Maier et al. 2012), die Entwicklung von Assessmentinstrumenten (wie quop, z. B. Souvignier et al. 2014a) sowie die Durchführung entsprechender Forschungsprojekte (z. B. Decristan et al. 2015; Rakoczy et al. 2017).

Trotz aller bestehenden Anstrengungen bleibt die Implementation von formativem Assessment jedoch herausfordernd und anspruchsvoll. Herausforderungen beziehen sich sowohl auf die Ausgestaltung konkreter Materialien zur Umsetzung formativen Assessments als auch auf die Unterstützung von Schülerinnen und Schülern sowie Lehrkräften, um die prinzipiellen Vorteile formativen Assessments nutzen zu können.

Mit Blick auf die konkrete Gestaltung von Assessmentmerkmalen oder -materialien zeigte sich beispielsweise, dass theoretisch wirksame Elemente wie elaboriertes Feedback (s. z. B. die Studie von Maier et al. 2016) oder individuelle Zielsetzungen (s. z. B. die Studie von Förster und Souvignier 2014) nicht immer mit den erhofften positiven Effekten einhergehen. Dies legt die Empfehlung nahe, konkrete Materialien zunächst empirisch zu prüfen und so zu erproben, ob die theoretisch lernförderliche Intervention in der Praxis den gewünschten Effekt erzielt.

Die Zielsetzung, Schülerinnen und Schülern Feedback zu geben, das diese wiederum zur Verbesserung ihres Lernens nutzen können sowie die Absicht Schülerinnen und Schüler als Verantwortliche des eigenen Lernens zu aktivieren, führt zur Einsicht, auch die Seite der Lernenden verstärkt in den Blick zu nehmen. Schülerinnen und Schüler sind diejenigen, die in letzter Konsequenz ein Lernverhalten zeigen sollen, das ihnen zu einem bestmöglichen Wissenserwerb verhilft.

Schließlich zeigen einige Studien, dass Unterstützung für Lehrkräfte notwendig ist, um das Potenzial formativen Assessments nutzbar zu machen (vgl. Stecker 2017). Diagnostische Informationen als Grundlage für Planungen zu nutzen, scheint nicht zu den etablierten Routinen von Lehrkräften zu gehören (Zeuch et al. 2017). So zeigte sich im Rahmen von Interviews, dass Lehrkräfte bei der Beschreibung von Graphen zu individuellen Lernverläufen ihren Fokus auf die Bewertung von Leistungen (auch in Form von Schulnoten) richten. Ähnlich berichten Staman et al. (2017) darüber, wie herausfordernd es ist Lehrkräfte beim Wechsel von einer summativen zu einer formativen Perspektive zu begleiten und dass die „Übersetzung“ von Informationen zu individuellen Lernverläufen in adaptive Unterrichtsangebote konkreter Unterstützungsangebote bedarf. Die Umsetzung von formativem Assessment erfordert neben spezifischem Wissen (u. a. Wissen über Diagnosemöglichkeiten und Rückmeldegestaltung, fachliches und fachdidaktisches Wissen, vgl. z. B. Heritage 2007; Hill et al. 2010) auch assessmentspezifische Kompetenzen (wie z. B. die Kompetenz diagnostische Informationen lesen und interpretieren zu können, um datenbasierte Entscheidungen zu treffen; Espin et al. 2017; Staman et al. 2017). Darüber hinaus ist sie mit Veränderungen des Lehr-Lern-Prozesses verbunden und erfordert somit das Einüben neuer Unterrichtspraktiken und letztlich den Erwerb neuer Haltungen.

Schütze et al. (2017) zeigten beispielsweise, dass Lehrkräfte nach einem Training zu formativem Assessment mehr über formatives Assessment (inkl. Rückmeldung) wissen als Lehrkräfte einer Vergleichsgruppe, dieses Wissen in künstlichen Rückmeldesituationen auch umsetzen können, ihre Rückmeldepraxis im Unterricht aus Schülerperspektive aber nicht verändern. Eine der größten Herausforderungen bei der Umsetzung formativen Assessments stellt letztlich der, mit der Etablierung eines komplexen Ansatzes verbundene, Zeitaufwand dar (Black und Wiliam 1998a; James und Pedder 2006; Bennett 2011). Die Implementation von formativem Assessment ist also ein langwieriger Prozess und sollte dementsprechend phasenweise (Brookhart et al. 2008) und in kleinen Schritten vorgenommen werden (Angelo und Cross 1993; Harlen 2008).²

Theoretisch fundierte Konzepte erfordern bei ihrer Umsetzung in die Praxis zu meist Feinabstimmungen. So ist derzeit die Frage, durch welche Fortbildungsmaßnahmen und Materialien eine adäquate Unterstützung von Lehrkräften erfolgen kann, weitgehend ungeklärt. Zu den Materialangeboten gehören idealerweise aufeinander abgestimmte Assessment-, Feedback- und Fördermaterialien. Große Hoffnungen werden aktuell in computerunterstützte Angebote gesetzt, da sie ein hohes Potenzial zur Umsetzung ökonomischen Assessments (z. B. Souvignier et al. 2014a, b), unmittelbaren Feedbacks (z. B. Maier et al. 2016) und adaptiver Förderung (z. B. Faber et al. 2017) aufweisen.

Literatur

- Aleven, V., Roll, I., McLaren, B., & Koedinger, K. (2010). Automated, unobtrusive, action-by-action assessment of self-regulation during learning with an intelligent tutoring system. *Educational Psychologist*, 45(4), 224–233. <https://doi.org/10.1080/00461520.2010.517740>.
- Andrade, H. L. (2010). Summing up and moving forward: Key challenges and future directions for research and development in formative assessment. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 344–351). New York: Routledge.
- Andrade, H. L., & Cizek, G. J. (Hrsg.). (2010). *Handbook of formative assessment*. New York: Routledge.
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2. Aufl.). San Francisco: Jossey-Bass.
- Baldwin, E. E., & Yun, J. T. (2012). *Mathematics curricula and formative assessments: Toward an error-based approach to formative data use in mathematics*. Santa Barbara: University of California Educational Evaluation Center.
- Bastian, J., Combe, A., & Langer, R. (2003). *Feedbackmethoden. Erprobte Konzepte, evaluierte Erfahrungen*. Weinheim: Beltz.
- Bauch, W. (2010). Kompetenzorientierter Unterricht – Akzente setzen, die Chancen nutzen. Das Marburger Pilotprojekt „Kompetenzorientiert unterrichten“ 2008–2010. *Schulpädagogik heute*, 1(1), 1–23.
- Beatty, I. D., & Gerace, J. W. (2009). Technology-enhanced formative assessment: A research-based pedagogy for teaching science with classroom response technology. *Journal of Scientific Educational Technology*, 18, 146–162. <https://doi.org/10.1007/s10956-008-9140-4>.
- Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment*, 1(1), 2–23.

² Angelo und Cross (1993) postulieren beispielsweise folgende Schritte bei der Implementation formativen Assessments: drei Schritte der Planung ([1] Auswahl einer Klasse in der das Assessment durchgeführt wird, [2] Fokussierung eines spezifischen Lerngegenstandes bzw. Lernziels, [3] Auswahl einer Assessmenttechnik); drei Schritte der Durchführung ([4] Durchführung der Unterrichtsstunde, die sich auf den Lerngegenstand bezieht, [5] Durchführung des Assessments, [6] Analyse der Daten) und drei Phasen der Reaktion ([7] Interpretation der Ergebnisse und Ableitung von Fördermaßnahmen, [8] Rückmeldung der Ergebnisse an die Schülerinnen und Schüler, [9] Evaluation des Assessmentprojekts).

- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>.
- Bergan, J. R., Sladeczek, I. E., Schwarz, R. D., & Smith, A. N. (1991). Effects of a measurement and planning system on kindergartners' cognitive development and educational programming. *American Educational Research Journal*, 26, 683–714.
- Bernholt, S., Rönnebeck, S., Ropohl, M., Köller, O., & Parchmann, I. (2013). *Report on current state of the art in formative and summative assessment in IBE in STM: Report from the FP7 project ASSIST-ME (Deliverable 2.4)*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–147.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5–31. <https://doi.org/10.1007/s11092-008-9068-5>.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Hrsg.), *Educational evaluation: new roles, new means: the 63rd yearbook of the National Society for the Study of Education (part II)* (Bd. 69(2), S. 26–50). Chicago: University of Chicago Press.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Gipps, C. V., Harlen, W., & James, M. (1999). *Assessment for learning: beyond the black box*. Cambridge: University of Cambridge.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge: University of Cambridge.
- Brookhart, S. M. (2010). Mixing it up: combining sources of classroom achievement information for formative and summative purpose. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 279–296). New York: Routledge.
- Brookhart, S. M., Moss, C. M., & Long, B. A. (2008). Formative assessment that empowers. *Educational Leadership*, 66(3), 52–57.
- Bürgermeister, A., Klimczak, M., Klieme, E., Rakoczy, K., Blum, W., Leiß, D., et al. (2011). Leistungsbeurteilung im Mathematikunterricht – Eine Darstellung des Projekts „Nutzung und Auswirkungen der Kompetenzmessung in mathematischen Lehr-Lernprozessen“. *Schulpädagogik – heute*, 2(3), 1–18.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65(3), 245–281. <https://doi.org/10.3102/00346543065003245>.
- Chen, L., & Chen, T.-L. (2012). Use of Twitter for formative evaluation: Reflections on trainer and trainees' experiences. *British Journal of Educational Technology*, 43, 49–52.
- 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. <https://doi.org/10.1007/s10648-011-9191-6>.
- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101–116. <https://doi.org/10.1080/09695949993026>.
- Cronbach, L. J. (1964). Evaluation for course improvement. In R. W. Heath (Hrsg.), *New Curricula* (S. 231–248). New York: Harper & Row.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438–481. <https://doi.org/10.3102/00346543058004438>.
- Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., & Hardy, I. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *Journal of Educational Research*, 108, 358–370. <https://doi.org/10.1080/002206712014.899957>.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.
- Dervan, P. (2014). Increasing in-class student engagement using Socratic (an online student response system). *The All Ireland Journal of Teaching & Learning in Higher Education*, 6, 1801–1803.
- 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.
- 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*, 32(1), 8–21.

- Faber, J.M., Luyten, H., & Visscher, A.J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education*, *76*, 83–96.
- Fontana, D., & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology*, *64*, 407–417.
- 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.
- Fuchs, L.S. (2017). Curriculum-based measurement as the emerging alternative: three decades later. *Learning Disabilities Research & Practice*, *32*, 5–7. <https://doi.org/10.1111/ldrp.12127>.
- Fuchs, L.S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, *53*(3), 199–208.
- Gikandi, J., Morrow, D., & Davis, N. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, *57*(4), 2333–2351. <https://doi.org/10.1016/j.compedu.2011.06.004>.
- Grob, R., Beerenwinkel, A., Haselhofer, M., Holmeier, M., Stübi, C., Tsvitanidou, O., & Labudde, P. (2014). *Description of the ASSIST-ME assessment methods and competences (Deliverable 4.7)*. Basel: University of Applied Sciences and Arts Northwestern Switzerland.
- Harks, B. (2013). *Kompetenzdiagnostik und Rückmeldung – zwei Komponenten formativen Assessments*. Frankfurt a. M.: Goethe-Universität. Unveröffentlichte Dissertation.
- Harlen, W. (2008). Editor's introduction. In W. Harlen (Hrsg.), *Student Assessment and Testing* (S. xix–xlvi). London: SAGE.
- Hasselhorn, M., Decristan, J., & Klieme, E. (2018). Individuelle Förderung. In O. Köller, M. Hasselhorn, F.W. Hesse, K. Maaz, J. Schrader, H. Solga, C.K. Spieß, & K. Zimmer (Hrsg.), *Das Bildungswesen in Deutschland: Bestand und Potenziale*. Bad Heilbrunn: UTB/Klinkhardt.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Helmke, A., & Lenske, G. (2013). Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung. *Beiträge zur Lehrerbildung*, *31*(2), 214–233.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, *89*, 140–145.
- Higgins, K.M., Harris, N.A., & Kuehn, L.L. (1994). Placing assessment into the hands of young children: a study of student-generated criteria and self-assessment. *Educational Assessment*, *2*, 309–324.
- Hill, M., Cowie, B., Gilmore, A., & Smith, L.F. (2010). Preparing assessment-capable teachers: What should preservice teachers know and be able to do? *Assessment Matters*, *2*, 43–64.
- Hosenfeld, I., & Zimmer-Müller, M. (2009). Was VERA Lehrern sagen kann. Über Sinn und Interpretation von Vergleichsarbeiten in der Grundschule. *Schule im Blickpunkt*, *5*, 8–9.
- Ikemoto, G.S., & Marsh, J.A. (2007). Cutting through the data decision mantra: Different conception of data-driven decision making. In P.A. Moss (Hrsg.), *Evidence and decision making* (S. 105–131). Malden: Blackwell.
- Ingenkamp, K. (1985). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (Hrsg.), *Assessment and Learning* (S. 47–60). London: SAGE.
- James, M., & Pedder, D. (2006). Professional learning as a condition for assessment for learning. In J. Gardner (Hrsg.), *Assessment and Learning* (S. 27–25). London: SAGE.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, *30*, 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>.
- Kingston, N., & Nash, B. (2015). Erratum. *Educational Measurement: Issues and Practice*, *34*(2), 55. <https://doi.org/10.1111/emip.12075>.
- Klecker, B.M. (2003). The impact of formative feedback on student learning in an online classroom. *Journal of Instructional Psychology*, *34*, 161–165.
- Klieme, E., & Warwas, J. (2011). Konzepte der individuellen Förderung. *Zeitschrift für Pädagogik*, *57*(6), 805–818.
- Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D., & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht. *Zeitschrift für Pädagogik*, *56*, 64–74.

- 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. <https://doi.org/10.1037//0033-2909.119.2.254>.
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*, 239–264.
- Koepen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie*, *216*, 61–73. <https://doi.org/10.1027/0044-3409.216.2.61>.
- Köller, O. (2005). Formative assessment in classrooms: a review of the empirical German literature. In OECD (Hrsg.), *Formative assessment: Improving learning in secondary classrooms* (S. 265–279). Paris: OECD.
- Kultusministerkonferenz (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004*. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf. Zugegriffen: 11. Juni 2018.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute-by-minute and day-by-day. *Educational Leadership*, *63*(3), 19–24.
- Lehmann, R. H., & Seeber, S. (2005). *Accelerated Mathematics in grades 4–6. Summary of a quasi-experimental study in Northrhine-Westphalia, Germany*. <http://research.renlearn.com/research/pdfs/192.pdf>. Zugegriffen: 11. Juni 2018.
- Lin, Q. (2008). Preservice teachers' learning experiences of constructing e-portfolios online. *The Internet and Higher Education*, *11*, 194–200.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918. <https://doi.org/10.1037/a0037123>.
- Maier, U. (2010a). Formative Assessment – Ein erfolversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, *13*(2), 293–308. <https://doi.org/10.1007/s11618-010-0124-9>.
- Maier, U. (2010b). Vergleichsarbeiten im Spannungsfeld zwischen formativer und summativer Leistungsmessung. *Die Deutsche Schule*, *102*(1), 60–69.
- Maier, U. (2011). Formative Leistungsdiagnostik in der Sekundarstufe I – Befunde einer quantitativen Lehrerbefragung zu Nutzung und Korrelaten verschiedener Typen formativer Diagnosemethoden in Gymnasien. *Empirische Pädagogik*, *25*(1), 25–46.
- Maier, U., Hofmann, F., & Zeitler, S. (2012). *Formative Leistungsdiagnostik. Grundlagen und Praxisbeispiele*. München: Oldenburg.
- Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education*, *75*, 85–98.
- Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, *47*, 71–85. <https://doi.org/10.1080/00461520.2012.667064>.
- McLaughlin, T., & Yan, Z. (2017). Diverse delivery methods and strong psychological benefits: A review of online formative assessment. *Journal of Computer Assisted Learning*, *33*(6), 562–574. <https://doi.org/10.1111/jcal.12200>.
- McMillan, J. (2010). The practical implications of educational aims and contexts for formative assessment. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 41–58). New York: Routledge.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, *22*(2), 155–175. https://doi.org/10.1207/s15326985Sep2202_4.
- OECD (2005). *Formative assessment: Improving learning in secondary classrooms*. <http://www.oecd.org/publications/Policybriefs>. Zugegriffen: 11. Juni 2018.
- OECD (2008). *Assessment for learning: The case for formative assessment learning in the 21st century: Research, innovation and policy*. <http://www.oecd.org/publications/Policybriefs>. Zugegriffen: 11. Juni 2018.
- Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknown and potential directions. *Educational Psychology Review*, *28*, 803–830.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington: National Academic Press.
- Poggio, A., Poggio, J., & Glasnapp, D. (2007). *The utility and impact of online computerized formative and early warning assessments on student performance*. Vortrag gehalten auf der Tagung des National Council on Measurement in Education, Chicago.

- Preiser, S. (2000). Feedback nach Referaten als hochschuldidaktisches Instrument. In G. Krampen (Hrsg.), *Psychologiedidaktik und Evaluation II* (S. 187–202). Bonn: Deutscher Psychologen Verlag.
- Rakoczy, K., Klieme, E., Leiß, D., & Blum, W. (2017). Formative assessment in mathematics instruction: Theoretical considerations and empirical results of the Co²CA project. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Hrsg.), *Competence assessment in education* (S. 447–468). Berlin: Springer.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioural Science*, 28(1), 4–13. <https://doi.org/10.1002/bs.3830280103>.
- Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3/4), 205–235. <https://doi.org/10.1080/10627197.2006.9652991>.
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding of scientific inquiry. *Journal of Research in Science and Teaching*, 44(1), 57–84. <https://doi.org/10.1002/tea.20163>.
- Russel, M. (2010). Technology-aided formative assessment and learning: New developments and applications. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 125–138). New York: Routledge.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144. <https://doi.org/10.1007/BF00117714>.
- Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H., & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology*, 109(8), 1049–1066.
- Schütze, B., Rakoczy, K., Klieme, E., Besser, M., & Leiss, D. (2017). Training effects on teachers' feedback practice: the mediating function of feedback knowledge and the moderating role of self-efficacy. *ZDM Mathematics Education*, 49(3), 475–489.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Hrsg.), *Perspectives of curriculum evaluation* (S. 38–83). Chicago: Rand McNally.
- Shavelson, R. J. (2008). Guest editor's introduction. *Applied Measurement in Education*, 21(4), 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(4), 295–314. <https://doi.org/10.1080/08957340802347647>.
- Shepard, L. A. (2000). The role of assessment in learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.3102/0013189X029007004>.
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>.
- Souvignier, E., Förster, N., & Salaschek, M. (2014a). Quop: Ein Ansatz internetbasierter Lernverlaufsdiagnostik mit Testkonzepten für Lesen und Mathematik. In M. Hasselhorn, W. Schneider, & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 239–256). Göttingen: Hogrefe.
- Souvignier, E., Förster, N., & Schulte, E. (2014b). Wirksamkeit formativen Assessments – Evaluation des Ansatzes der Lernverlaufsdiagnostik. In M. Hasselhorn, W. Schneider, & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 221–238). Göttingen: Hogrefe.
- Staman, L., Timmermanns, A. C., & Visscher, A. J. (2017). Effects of a data-based decision making intervention on student achievement. *Studies in Educational Evaluation*, 55, 58–67.
- Stecker, P. M. (2017). Reflections on teachers' data-based decision making. *Learning Disabilities Research and Practice*, 32, 71–72.
- 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.
- Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, 106(2), 331–347. <https://doi.org/10.1037/a0034752>.
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758–765.
- Strandberg, M. (2013). Homework – is there a connection with classroom assessment? A review from Sweden. *Educational Research*, 55(4), 325–346. <https://doi.org/10.1080/00131881.2013.844936>.
- Strathmann, A. M., & Klauer, K. J. (2012). *LVD-M 2–4. Lernverlaufsdiagnostik-Mathematik für zweite bis vierte Klassen*. Göttingen: Hogrefe.
- Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, 20(4), 265–269. <https://doi.org/10.1016/j.learninstruc.2009.08.002>.

- Topping, K. J. (2010). Peers as a source of formative assessment. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 61–74). New York: Routledge.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>.
- Velan, G. M., Rakesh, K. K., Mark, D., & Wakefield, D. (2002). Web-based self-assessments in pathology with questionmark perception. *Pathology*, 34, 282–284. <https://doi.org/10.1080/00313020220131372>.
- Visible Learning Lab (2010). *Educator manual. E-asTTle fitness for national standards*. <http://e-asttle.tki.org.nz/User-manuals>. Zugegriffen: 11. Juni 2018.
- Vonderwell, S. (1998). Assessing online learning and teaching: Adapting the minute paper. *TechTrends*, 48, 29–32.
- Walter, J. (2010). *LDL. Lernfortschrittsdiagnostik Lesen (LDL). Ein curriculumbasiertes Verfahren*. Göttingen: Hogrefe.
- Walter, J. (2013). *VSL. Verlaufsdiagnostik sinnerfassenden Lesens*. Göttingen: Hogrefe.
- Wiliam, 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.
- Wiliam, 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.
- Winter, F. (1991). *Schüler lernen Selbstbewertung. Ein Weg zur Veränderung der Leistungsbeurteilung und des Lernens*. Frankfurt a. M.: Peter Lang.
- Wylie, E. C., Gullickson, A. R., Cummings, K. E., Egelson, P. E., Noakes, L. A., & Norman, K. M. (2012). *Improving formative assessment practice to empower student learning*. Thousand Oakes: Corwin.
- Ysseldyke, J. E., & Bolt, D. M. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review*, 36, 453–467.
- 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.