



Ann Cathrice George, Stefan Götz, Marcel Illetschko,
Evelyn Süß-Stepancik (Hrsg.)

Empirische Befunde zu Kompetenzen im Mathematikunterricht der Sekundarstufe 1 und Folgerungen für die Praxis

Ergänzende Analysen zu den Bildungsstandardüberprüfungen

In der gezeigten Konstruktion gilt:

$\alpha = 60^\circ$ $x = 18 \text{ mm}$ $\overline{AB} = 65 \text{ mm}$
 $\gamma = 120^\circ$ $y = 18 \text{ mm}$ $\overline{CD} = 65 \text{ mm}$

Die Strecken $\overline{AB}$ und $\overline{CD}$ sind parallel.
Welche Begründung dafür ist richtig, welche falsch?

Kreuze für jede Zeile an. ☒

	richtig	falsch
$\overline{AB}$ und $\overline{CD}$ sind parallel, weil α und γ gleich groß sind.	☐	☐
$\overline{AB}$ und $\overline{CD}$ sind parallel, weil $\overline{AB}$ und $\overline{CD}$ gleich lang sind.	☐	☐
$\overline{AB}$ und $\overline{CD}$ sind parallel, weil sowohl $\overline{AB}$ als auch $\overline{CD}$ normal auf g sind.	☐	☐
$\overline{AB}$ und $\overline{CD}$ sind parallel, weil x gleich lang wie y ist.	☐	☐

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <http://dnb.dnb.de> abrufbar.

**Kompetenzmessungen im österreichischen Schulsystem:
Analysen, Methoden & Perspektiven, Bd. 3**

ISSN 2628-9598

Print-ISBN 978-3-8309-4558-1

E-Book-ISBN 978-3-8309-9558-6

DOI 10.31244/9783830995586

© Waxmann Verlag GmbH, 2022

www.waxmann.com

info@waxmann.com

Umschlaggestaltung: Hannes Kaschnig-Löbel, IQS &

Anne Breitenbach, Münster

Satz: Andreas Kamenik, IQS

Dieses Werk ist unter der Lizenz CC BY-NC-SA 4.0 veröffentlicht:

Namensnennung – Nicht-kommerziell –

Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-NC-SA 4.0)



Inhalt

Vorwort des IQS-Direktors.....	7
<i>Kristina Reiss</i>	
Bildungsstandards und die Qualitätssicherung in Schule und Unterricht – ein Vorwort.....	9
<i>Andreas Schulz, Alexander Aichinger, Martina Hartl</i>	
Bildungsstandards Mathematik – von der Theorie zur Nutzung.....	15
<i>Martina Müller, Monika Musilek, Christian Wimmer</i>	
Bildungsstandardüberprüfungen in Mathematik auf der Sekundarstufe 1 über den Zeitverlauf 2009–2012–2017.....	55
<i>Stefan Götz, Ann Cathrice George</i>	
Vom Argumentieren in der Statistik oder: Konnexe zwischen Inhalts- und Handlungsbereichen.....	79
<i>Evelyn Süss-Stepancik, Michael Ober, Ann Cathrice George, Andrea Varelja-Gerber</i>	
Veränderungen im mathematischen Leistungsspektrum einer Kohorte von der Primar- zur Sekundarstufe.....	113
<i>Maria Neubacher, Eva Sattlberger, Stefan Götz</i>	
Overperformer-Schulen in den Bundesländern	133
<i>Isabella Benischek, Konrad Oberwimmer, Iris Höller</i>	
Zu Einflussfaktoren auf Bildungserwartungen von Schülerinnen und Schülern nach der Sekundarstufe 1.....	159
<i>Elisabeth Rothe, Christoph Weber, Marcel Illetschko</i>	
Ethnische und soziale Bildungsungleichheiten in Mathematik – Anregungen für den Mathematikunterricht.....	187
<i>Christina Drüke-Noe, Burkhard Gniewosz, Daniel Paasch</i>	
Mathematisches Selbstkonzept und Schülerleistungen – Zusammenhänge für den Unterricht nutzbar machen.....	211

Boris Girnat, Ramona Zintl, Regina Bruder

Strategien der Testbearbeitung.....233

Simon Plangg, Florian Stampfer, Elisabeth Fuchs

**Eine Aufgabe, viele Fehler – Ergebnisse einer qualitativen Analyse
zum Mathematisieren auf der Sekundarstufe 1 und Implikationen
für die Unterrichtspraxis.....259**

Autorinnen und Autoren.....293

Bildungsstandards Mathematik – von der Theorie zur Nutzung

1 Einleitung

Die Einführung der österreichischen Bildungsstandards und ihre regelmäßige Überprüfung hatten zum Ziel, „(1) eine nachhaltige Ergebnisorientierung in der Planung und Durchführung von Unterricht zu bewirken, (2) durch konkrete Vergleichsmaßstäbe die bestmögliche Diagnostik für individuelle Förderung sicherzustellen und (3) wesentlich zur Qualitätsentwicklung in der Schule beizutragen“ (Bildungsstandardverordnungen von 2009 und 2020, BGBl. II Nr. 1/2009 und BGBl. II Nr. 548/2020). Diese Ziele verdeutlichen, dass schulische und unterrichtliche Qualitätsentwicklung als gemeinsame Aufgabe von (1, 2) Lehrkräften mit Schulleitungen in ihrer konkreten Planung und Gestaltung von Unterricht und (3) der Bildungsadministration, bestehend aus Schulaufsicht und Politik, betrachtet wird. Aufgabe der Bildungsadministration ist es nicht nur, Bildungsstandards in Form von Beschreibungen konkreter Lernergebnisse, die von Schülerinnen und Schülern bis zu einer definierten Klassenstufe erwartet werden, zu formulieren. Verbunden damit besteht auch der Anspruch an die Bildungsadministration, das Erreichen dieser Bildungsstandards durch die Schüler/innen in Bildungsstandardüberprüfungen zu verfolgen sowie aussagekräftige und verständliche Rückmeldungen für die unterschiedlichen Zielgruppen in Auftrag zu stellen. Lehrpersonen sollen zudem, u. a. in Form von geeigneten schulischen Rahmenbedingungen und Fortbildungsangeboten, bei ihrer Nutzung der Bildungsstandardüberprüfungen und Weiterentwicklung des Unterrichts unterstützt werden.

Das vorliegende Buch/Kapitel möchte hierzu einen Beitrag leisten, indem in Abschnitt 1 zunächst die grundlegenden, intendierten Wirkungsweisen und Funktionen von Bildungsstandards im Schulsystem erläutert werden. In Abschnitt 2 werden die österreichischen Bildungsstandards Mathematik für die 8. Schulstufe vorgestellt. Abschnitt 3 diskutiert Ansätze und Modelle zur Nutzung von Bildungsstandardüberprüfungen. Abschnitt 4 fasst die zentralen Aussagen der ersten Abschnitte zusammen und begründet darüber hinaus die aktuelle Weiterentwicklung der beiden bestehenden Instrumente der Bildungsstandardüberprüfungen (kurz BIST-Ü) sowie der Informellen Kompetenzmessung (kurz: IKM) hin zur individuellen Kompetenzmessung PLUS (kurz: iKM^{PLUS}), deren Einführung einem stufenweisen Zeitplan folgt.

2 Bildungsstandards – ein Instrument zur schulischen Qualitätsentwicklung

Bildungsstandards dienen als Instrument zur Qualitätssicherung im Schulsystem und sollen gleichzeitig einen Beitrag zur überwiegend fachspezifischen Unterrichtsentwicklung leisten. Die Begriffe „Ergebnisorientierung“ und „Kompetenzorientierung“ weisen auf mit diesen Anliegen verbundene zentrale Ideen hin, welche einer Orientierung von Lehrkräften an Bildungsstandards und Ergebnissen aus Bildungsstandardüberprüfungen erst einen Sinn geben.

Ein grundlegendes Verständnis dieser beiden Begriffe erscheint daher notwendig, damit aus der Einführung von Bildungsstandards eine Nutzung der Ergebnissrückmeldungen und Weiterentwicklung des Unterrichts erwachsen kann. Die Idee der Ergebnisorientierung lässt sich auch ohne spezifische Charakterisierung für das Fach Mathematik beschreiben. Die Idee der Kompetenzorientierung mit ihrem ausgeprägten Bezug zur Unterrichtsgestaltung beinhaltet dahingegen im Fach Mathematik besondere, fachspezifische Bedeutungen, die in Abschnitt 2.2 dargelegt werden.

2.1 Ergebnisorientierung im Bildungssystem

Staatliche Schulsysteme sind eine gesellschaftliche Institution. Ihr grundlegendes Anliegen ist es, Qualität und Gleichheit von Bildungschancen zu gewährleisten. Standards im Allgemeinen kommt traditionell in allen Bereichen eines Schulsystems eine wichtige steuernde Funktion zu (Oelkers & Reusser, 2008). Standardisierte Vorgaben beziehen sich u. a. auf Lehrbücher, Versetzungskriterien, Stundenzuteilungen für Unterrichtsfächer, Raumausstattungen und die Ausbildung der Lehrkräfte. Damit wird ein hohes Maß an Sicherheit in der Gestaltung des schulischen Arbeitsalltags verbunden.

Bildungsstandards konkretisieren den Bildungsauftrag und legen fest, welche Kompetenzen die Schüler/innen bis zu einer bestimmten Jahrgangsstufe erreicht haben sollen. Darin werden die Kompetenzen so konkret beschrieben, dass sie in Aufgabenstellungen umgesetzt und prinzipiell mithilfe von Tests erfasst werden können. Dies entspricht einer ergebnisorientierten Perspektive auf die Steuerung eines Bildungssystems: Standards sind so lange unnütz und bedeutungslos, wie es keine Möglichkeiten gibt, ihr Erreichen zu messen (Klieme et al., 2003; Ravitch, 1995).

Bildungsstandards dienen somit als Referenzsystem für die Generierung von Feedback über Lernergebnisse. Eine Feedbacknutzung kann prinzipiell auf verschiedenen Ebenen des Schulsystems stattfinden (Oelkers & Reusser, 2008; Schulz, 2010):

- auf nationaler Ebene als Bildungsmonitoring
- auf Schulebene zur Schulentwicklung sowie als Schulevaluation
- auf Klassenebene für die Unterrichtsplanung und Unterrichtsevaluation
- auf Ebene der Schüler/innen als Individualdiagnostik für gezielte Fördermaßnahmen

Inwieweit ein einzelnes Instrument zur Überprüfung von Bildungsstandards geeignet ist, belastbares und aussagekräftiges Feedback auf einer oder mehrerer dieser Ebenen in einem Schulfach bereitzustellen, hängt u. a. von der jeweiligen Zielsetzung des Instruments, der fachdidaktischen Fundierung und Fokussierung der Aufgaben sowie der erfassten Kompetenzbereiche, der Form der statistischen Datenverarbeitung und der Ausgestaltung der Feedbackformate ab. Generell zeichnen sich Instrumente zum Bildungsmonitoring eher durch inhaltlich breit erfasste Kompetenzbereiche, Instrumente zur Individualdiagnostik dagegen durch inhaltlich eng fokussierte Kompetenzbereiche aus (Leuders, Schulz & Kowalk, 2019; Mislevy, Steinberg & Almond, 2003; Pellegrino, Chudowsky & Glaser, 2001; Schulz, Leuders & Rangel, 2020).

Der Begriff „Ergebnisorientierung“ als Hintergrund für Setzungen und Überprüfungen von Bildungsstandards beinhaltet zusammenfassend, dass Lernergebnisse von Schülerinnen und Schülern „zum entscheidenden Bezugspunkt für die Beurteilung des Schulsystems und für Maßnahmen zur Verbesserung und Weiterentwicklung“ werden (Klieme et al., 2003, S. 7).

2.2 Kompetenzorientierung im Fach Mathematik

Das Setzen und Überprüfen von Bildungsstandards im Sinne der Ergebnisorientierung ist eng verbunden mit der empirischen Messbarkeit von Unterrichtszielen und Lernleistungen der Schüler/innen. Dementsprechend formulieren Bildungsstandards die zu erreichenden Bildungsziele in Form von Aussagen wie: „Schülerinnen und Schüler können ...“. Hierbei kommt ein weiteres Merkmal von vielfach mit Kompetenzen beschriebenen Lernzielen zum Ausdruck, nämlich die Fähigkeit und Bereitschaft, das erworbene Wissen auch in neuen Situationen zur Lösung von Problemen anzuwenden. Dementsprechend beinhalten Kompetenzbeschreibungen teils auch motivationale Facetten, die theoretisch und empirisch fassbar sind.

Damit grenzen sich Kompetenzen von allgemeinen und weniger beeinflussbaren Persönlichkeitsmerkmalen wie Intelligenz und von kontextfreiem, nicht anwendungsbezogenem Wissen ab. In Bildungsstandards wird zur Beschreibung des Kompetenzbegriffs überwiegend auf die Definition von Weinert (zitiert nach Köller, Baumert & Bos, 2001, S. 271) verwiesen: „Dabei versteht man unter Kompetenzen die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“

Im Fach Mathematik lassen sich in Überlegungen zu kompetenzorientiertem Unterricht weitere Bezüge feststellen, die u. a. in ergänzenden Praxishinweisen zu Bildungsstandards und Weiterbildungsangeboten Verwendung finden. Diese beziehen sich auf die Bereiche a) Lernen und Unterrichten und b) Leisten und Testen (Schulz, 2010):

a) *Kompetenzorientierung beim Lernen und Unterrichten*

- *Selbstständigkeit als Ziel und Arbeitsform*: Selbstständigkeit bezieht sich auf (kognitive) Aktivitäten von Lernenden sowie auf deren Organisation. Als Konsequenz für eine kompetenzorientierte Unterrichtsgestaltung wird daraus die Notwendigkeit einer Abkehr von vorwiegend lehrerzentriertem Unterricht gefordert (z. B. Bruder, Leuders & Büchter, 2016).
- *Nachhaltigkeit und Verständnis bauen auf Grundvorstellungen und Vernetzungen auf*: Kompetenzorientierter Unterricht geht über die reine Reproduktion von Verfahren zur Lösung typischer Aufgabenstellungen hinaus. Verstehensbasierte mathematische Aktivitäten bauen auf mathematischen Grundvorstellungen (vom Hofe, 2003) auf, die in vielseitigen Lernarrangements mit Binnendifferenzierung, Aufgaben mit vielfältigen Lösungswegen sowie regelmäßigen Reflexionsphasen über Lösungswege und individuelle Lernfortschritte entwickelt werden (z. B. Blum, Drüke-Noe, Hartung & Köller, 2006).
- *Kumulatives Lernen*: Mathematische Inhalte werden über die Jahre hinweg im Sinne eines Spiralcurriculums angeordnet. Dies ermöglicht die Behandlung fundamentaler mathematischer Ideen aus verschiedenen Perspektiven und ihre Wiederholung auf verschiedenen Schwierigkeitsniveaus (z. B. Bruder et al., 2016).

b) *Kompetenzorientierung beim Leisten und Testen*

- *Fokussierte Leistungsaufgaben unterstützen Unterrichtsentwicklung*: Testaufgaben müssen auf die Inhalte und Gestaltung des vorangegangenen Unterrichts abgestimmt sein (Bruder et al., 2016). Zudem dienen Testaufgaben Schülerinnen und Schülern und ebenso Lehrkräften als Orientierung, indem sie Unterrichtsziele veranschaulichen. Eine ausschließliche Verwendung von verfahrensorientierten Testaufgaben würde einer Unterrichtsentwicklung in Richtung verstehensorientierter und problemlösender Lerngelegenheiten entgegenwirken.
- *Von der Performanz auf die Kompetenz folgern*: Nicht die Lösung einer speziellen Aufgabe wird als Lernziel angesehen, sondern die zugrunde liegende Kompetenz, die zur Lösung einer Aufgabe notwendig ist. Dies erfordert klar fokussierte Testaufgaben mit Blick auf präzise beschriebene kognitive Prozesse, die Aufgabenlösungen und Fehlern zugrunde liegen (Mislevy et al., 2003; Pellegrino et al., 2001).
- *Offenheit von Aufgaben*: Kompetenzorientierte Leistungsüberprüfung erfordert neben klar fokussierten Aufgaben auch offene Aufgaben, die eigene Lösungswege und individuelle Vorgehensweisen ermöglichen und für diagnostische Zwecke Hinweise auf die Ausprägung von Verständnis zu Beginn oder am Ende einer Lerneinheit ergeben. Bewertungen von offenen Aufgaben sind oftmals mit umfangreichen Rückmeldungen an die Lernenden verbunden. Hierbei können mehrdimensionale Bewertungsraster eine transparente und dokumentierbare Rahmung der Bewertung für Lernende und Lehrende bieten (z. B. Greefrath & Maaß, 2020).
- *Diagnose und Förderung*: Ein kompetenzorientierter Blick auf Schülerprodukte relativiert eine defizitorientierte Sichtweise. Sichtbar wird, worauf man aufbauen und was man zielgerichtet fördern kann (Sundermann & Selzer, 2013).

- *Individuelle Lernwege ermöglichen:* Lernprozesse sind individuell und verlaufen autonom. Zudem sind Könnenserfahrungen für erfolgreiches Lernen unverzichtbar (Sundermann & Selter, 2013).

Gerade die Idee der Kompetenzorientierung verdeutlicht vielfältige Bezüge aktueller bildungspolitischer Ansätze zur Qualitätssicherung im Mathematikunterricht, die über eine alleinige Umsetzung von Ergebnisorientierung beim Setzen und Überprüfen von Bildungsstandards hinausgeht. Dabei können einzelne Maßnahmen recht unterschiedliche Schwerpunktsetzungen für spezifische Entwicklungsziele beinhalten.

2.3 Funktionen von Bildungsstandards

Im Schuljahr 2008/2009 zielten in Österreich die Einführung von Bildungsstandards und ihre Überprüfung auf die nachhaltige Etablierung des kompetenzorientierten Unterrichtens sowie auf die an Daten orientierte Unterrichtsentwicklung (Schreiner & Wiesner, 2019). Bildungsstandards und ihre Überprüfung sind nicht als Einzelmaßnahmen konzipiert, sondern entsprechen der Idee evidenzbasierter Steuerung (Altrichter & Kanape-Willingshofer, 2012; Klieme et al., 2003): Von dem Wechselspiel aus stärkerer Zielorientierung, evidenzbasierten Entscheidungen, engerer Wirkungskontrolle und zeitnahen Umsteuerungsmöglichkeiten werden ein gezielterer Einsatz der knappen Mittel und eine Qualitätssteigerung des Bildungssystems erwartet.

Mit den österreichischen Bildungsstandards wurden ähnlich wie in Deutschland Regelstandards formuliert, d. h. mittlere Niveaustufen, die im Durchschnitt zum Ende der 4. Schulstufe (Deutsch, Mathematik) sowie zum Ende der 8. Schulstufe (Deutsch, Mathematik, Englisch) erreicht werden sollen (Altrichter & Kanape-Willingshofer, 2012; Schreiner & Wiesner, 2019): Die flächendeckenden Überprüfungen fanden in Österreich periodisch am Ende von Bildungsgängen statt. Ergebnisse von Bildungsstandardüberprüfungen dürfen nicht in die schulische Leistungsbeurteilung der Schüler/innen und damit auch nicht in die Vergabe von Berechtigungen über nachfolgende Bildungsgänge eingehen.

Die Ergebnisse werden Schülerinnen und Schülern, Lehrpersonen, Schulleitungen und der Schulaufsicht rückgemeldet und enthalten neben Vergleichswerten die persönlich betreffenden Resultate. Lehrpersonen erhalten aggregierte Klassenergebnisse, nicht jedoch individuelle Schülerwerte. „Faire Vergleiche“ werden auf Schul- und Unterrichtsgruppenebene angestrebt, indem verschiedene, von der Schulleitung nicht beeinflussbare Hintergrundmerkmale wie Geschlecht, Schultyp, Leistungsgruppe und Migrationshintergrund in der Rückmeldung berücksichtigt werden (Nachtigall & Kröhne, 2006).

Die Überprüfung von Bildungsstandards beinhaltet dementsprechend die Information aller beteiligten Personengruppen und Systemebenen. Flankiert wird dies durch Fortbildungen für Multiplikatorinnen und Multiplikatoren, Rückmeldemoderatorinnen und -moderatoren,

Schulqualitätsmanager/innen, Schulleitungen und Lehrende an Pädagogischen Hochschulen sowie durch die Entwicklung und Verbreitung von Unterrichtsmaterialien (z. B. Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens [BIFIE], 2011, 2012).

Neben den Bildungsstandardüberprüfungen konnten Lehrpersonen in Österreich zudem „[...] ein Instrument zur Evaluierung des eigenen Unterrichts in Volksschule, Sekundarstufe 1 und Sekundarstufe 2 [...]“ nutzen, welches „[...] einfach, zuverlässig und kostenlos sowohl über den Lernstand der ganzen Klasse bzw. Gruppe als auch über den Kompetenzstand jeder einzelnen Schülerin/jedes einzelnen Schülers informiert“ (Institut des Bundes für Qualitätssicherung im österreichischen Schulwesen [IQS], 2021). Diese informelle Kompetenzmessung (kurz: IKM) sollte zu einer erfolgreichen Umsetzung eines kompetenzorientierten Unterrichts beitragen und Lehrpersonen dabei unterstützen, individuelle Förderbedarfe ihrer Schüler/innen zu erkennen.

Die mit dem Zusammenspiel von Bildungsstandardsetzung und -überprüfung angestrebte verbesserte Qualität schulischer Arbeit soll sich in wenigstens zwei zentralen Zielgrößen niederschlagen (Altrichter & Kanape-Willingshofer, 2012; Eder, 2009; Stojanov, 2011):

1. Verbesserte Schülerkompetenzen könnten sich aus einer konsequenten Fokussierung des Unterrichts auf die Kompetenzentwicklung der Schüler/innen in der gesamten Leistungsbreite ergeben. Darüber hinaus könnten differenzierte Maßnahmen für spezielle Zielgruppen, insbesondere eine verstärkte und fokussierte Förderung von leistungsschwächeren Lernenden zur nachhaltigen Leistungssteigerung des Schulsystems beitragen.
2. Bessere Chancengleichheit und Bildungsgerechtigkeit im Schulwesen wird angesichts des durch internationale Schulleistungsvergleiche dokumentierten engen Zusammenhangs zwischen sozialer Herkunft und Bildungskarrieren angestrebt. Auch hierzu könnten Bildungsstandardüberprüfungen mit Sekundäranalysen im Sinne einer Evaluation entsprechender schulpolitischer Maßnahmen beitragen. Ziel solcher Evaluationen könnte es sein, zu überprüfen, ob beispielsweise familiäre oder soziale Hintergründe, die Zugehörigkeit zu Bildungsregionen oder das Geschlecht bei der Notengebung und der Allokation zu Bildungsgängen eine abnehmende bis hin zu keiner Rolle spielen.

Um für die Überprüfung dieser beiden Ziele nicht nur Monitoring-Daten im Sinne reiner Leistungsstudien zu generieren, wurde in Österreich von Beginn an die Strategie verfolgt, neben Leistungsdaten auch wesentliche Kontextdaten (wie z. B. Geschlecht, Erstsprache, Schulklima, Klassenklima, Selbstkonzept, Freude am Fach ...) einzubeziehen und so weitergehende Analysen und Forschung zur Generierung von Erklärungen zu den Befunden zu ermöglichen (Schreiner & Wiesner, 2019).

Die Überprüfung von Bildungsstandards in Form von Kompetenzmessungen konkretisieren nun zunächst bereichsspezifische Leistungserwartungen (Altrichter & Kanape-Willingshofer, 2012). Mit dem erweiterten Ziel der nachhaltigen Einführung, Etablierung und Unterstützung von Kompetenzorientierung soll zudem ein Wandel in der Planung, Gestaltung und Durchführung von Unterricht angeregt und unterstützt werden (Schreiner & Wiesner, 2019). Inwiefern sich die Einführung und Überprüfung von Bildungsstandards beispielsweise in einer Änderung von Lehrbüchern abbildet, wird in Kapitel 2 dieses Sammelbands untersucht. Die komplexen Ziele und Wirkungserwartungen spiegeln sich in der Orientierungsfunktion, Förderfunktion und Evaluationsfunktion der Setzung und Überprüfung von Bildungsstandards wider (Schreiner & Wiesner, 2019; Wiesner, Schreiner, Breit & Pacher, 2017):

- *Orientierungsfunktion:* Bildungsstandardüberprüfungen und IKM konkretisieren die Zielsetzungen des Lehrplans aus der Perspektive der Lehrkräfte und Schüler/innen. Dadurch soll es zu einer nachhaltigen Ergebnisorientierung bei der Planung und Durchführung von Unterricht kommen.
- *Evaluierungsfunktion:* Bildungsstandardüberprüfungen messen, inwieweit Schulen und Unterricht die Kernaufgabe der Vermittlung und Aneignung von Kompetenzen standortbezogen, regional und bundesweit erfüllen.
- *Förderfunktion:* Bildungsstandardüberprüfungen und IKM ermöglichen Lehrkräften einen Abgleich zwischen den Könnensbeschreibungen.

Diese drei Funktionen der Setzung und Überprüfung von Bildungsstandards sowie damit verbundene Ansprüche bzw. Herausforderungen werden im Folgenden weiter ausgeführt.

2.3.1 Konkretisierung von Bildungsstandards mittels Testaufgaben – Orientierungsfunktion

Bildungsstandards werden in mittlerer Abstraktionsebene formuliert, die in Aufgaben und Testskalen umsetzbar sein müssen (Klieme et al., 2003). Mit den wissenschaftlichen Diskursen, die hinter den notwendigerweise knappen und meist allgemeinen Kompetenzbeschreibungen stehen, sind Lehrkräfte oftmals nicht vertraut. Dies ist beispielsweise der Fall, wenn wissenschaftliche Diskurse, z. B. zum „Darstellungswechsel“ sowie „Modellierungskreislauf“ als Hintergründe zum „Interpretieren“, nur begrenzt in der Ausbildung thematisiert werden und/oder wenn die eigene Ausbildung schon lange zurückliegt.

Veröffentlichte Testaufgaben helfen dann, Kompetenzbeschreibungen der Bildungsstandards zu konkretisieren. Die folgende Testaufgabe (Abbildung 1) stellt ein Beispiel für den Handlungsbereich „Darstellen, Modellbilden (H1)“ und den Inhaltsbereich „Statistische Darstellung und Kenngrößen (I4)“ dar (vgl. Abschnitt 2.1 „Theoretisches Kompetenzmodell“). In der Bildungsstandardüberprüfung 2017 hatte sie in Schulstufe 8 eine Lösungsquote von ca. 70 %:

Die Lehrerin trägt die Ergebnisse der Schularbeit in eine Tabelle ein:

Schularbeitenergebnisse					
Note	Sehr gut	Gut	Befriedigend	Genügend	Nicht genügend
Anzahl	5	6	8	2	4

Mit den Daten der Tabelle wurde ein Kreisdiagramm gezeichnet. Die Unterteilung zwischen den Noten „Genügend“ und „Nicht genügend“ fehlt noch.



Durch welchen Punkt ist die Linie zu ziehen?

Kreuze an. ☒

- ☐ A
☐ B
☐ C
☐ D

Abbildung 1: Testaufgabe aus der Bildungsstandardüberprüfung Mathematik, 8. Schulstufe, 2017 (BIFIE, 2017a)

Ein möglicher Lösungsweg könnte hier über den Abgleich des Verhältnisses 2:4 bzw. 1:2 für „Genügend“ zu „Nicht genügend“ auf dem verbleibenden Kreisbogen erfolgen. Dies verdeutlicht, dass für die Beantwortung weitere Inhaltsbereiche (z. B. Zahlen und Maße [I1]) und Handlungsbereiche (z. B. Interpretieren [H3]) beim Deuten der Grafik im Kontext ebenfalls zum Tragen kommen könnten. Dementsprechend bedürfen Konkretisierungen von Bildungsstandards mittels einer Zuordnung von Testaufgaben zu Inhalts- und Handlungsbereichen überwiegend ergänzender Erläuterungen. Auf diesem Wege können Zuordnungen begründet werden, um Missverständnissen und Unklarheiten im Verständnis der Inhalts- und Handlungsbereiche vorzubeugen. Dies erfolgt beispielsweise in begleitenden Handreichungen oder Praxishandbüchern (BIFIE, 2011, 2012).

Weitergehend lässt sich anhand entsprechender Beispiele von Testaufgaben erkennen, dass es zwar das Ziel von Bildungsstandardüberprüfungen ist, die einzelnen Inhalts- oder Handlungsbereiche möglichst trennscharf zu erfassen, dass dies jedoch nicht immer realisierbar ist. Eine komplett trennscharfe Operationalisierung würde inhaltlich nochmals deutlich enger fokussierte und eindeutiger zuordenbare Testaufgaben erfordern, welche dann jedoch viele der in Bildungsstandards adressierten komplexeren Kompetenzen nicht mehr valide abbilden könnten (vgl. dazu Götz & George in diesem Band).

Aus dem Faktum beispielsweise, dass die obige Aufgabe nicht richtig gelöst wurde, lässt sich nicht schlussfolgern, ob dies auf ein mangelndes Verständnis des Kreisdiagramms oder mangelnde Vorstellungen zu Verhältnissen zwischen Zahlen zurückzuführen ist. Auch ein Punktwert einer Schülerin oder eines Schülers auf einer entsprechenden Skala einer Bildungsstandardüberprüfung zu statistischen Darstellungen und Kenngrößen kann und soll zu derart detaillierten Fragestellungen keine Antwort liefern. Hierfür wären andere Diagnoseformate wie Interviews oder enger fokussierte Diagnoseaufgaben notwendig, die für den Unterrichtsalltag angemessen sind.

Die für Bildungsstandardüberprüfungen verwendeten Aufgaben und Rückmeldungen sind dementsprechend nicht für individualdiagnostische Zwecke konzipiert, in denen aus den Antworten möglichst eindeutig z. B. auf zugrundeliegende enger umrissene Fähigkeiten gefolgert werden soll, damit nachfolgende Fördermaßnahmen geplant werden können. Im Rahmen eines Bildungsmonitorings, mit dem generelle Leistungen des Bildungssystems erfasst und beschrieben werden sollen, ist das jedoch auch nicht notwendig.

Bedarf für die Konkretisierung von Bildungsstandards mittels Testaufgaben kann sich auch daraus ergeben, wenn Kompetenzmodelle, die hinter der Entwicklung von Kompetenzmessungen stehen, anfangs oftmals nicht umfassender als in den knappen Kompetenzbeschreibungen der Bildungsstandards selbst dargestellt wurden. Hier bestehen deutliche Unterschiede zwischen den Domänen. Während z. B. zum Lesen oder zum Fremdspracherwerb (Schneider, 2005) seit Langem etablierte und empirisch bewährte Kompetenzmodelle zur Verfügung stehen, ist die theoretische Modellierung und empirische Validierung insbesondere zu Handlungsbereichen im Fach Mathematik (vgl. process standards der NCTM, National Council of Teachers of Mathematics, 2000) noch deutlich weniger weit fortgeschritten. Dementsprechend steht nicht nur die Testentwicklung bei der Kompetenzmodellierung im Fach Mathematik vor dem Problem, die allgemeinen Kompetenzbeschreibungen in konkrete Aufgaben und empirisch tragfähige Kompetenzmodelle zu übersetzen. Auch für Lehrpersonen ist es herausfordernd, die Bedeutung der Ergebnismeldungen hinsichtlich der Kompetenzbeschreibungen und Skalenwerte zu verstehen und in konkrete, unterrichtsrelevante Aussagen zu übersetzen. Dafür müssen diese allgemeinen Testwerte und Skalenbeschreibungen (gedanklich) beispielsweise in konkrete Aufgaben und Schülerlösungen, d. h. in konkret beobachtbare Leistungen rückübersetzt werden (vgl. „Rückverflüssigungsproblem“, Oelkers & Reusser, 2008, S. 505). Kapitel 10 dieses Bands diskutiert anhand einer Fehleranalyse eine solche beispielhafte „Rückverflüssigung“ von Testwerten.

Generell lässt sich zusammenfassen, dass für die Interpretation von Bildungsstandards und von darauf bezogenen Kompetenzmessungen auf Ebene der Lehrpersonen ein fachdidaktisches Hintergrundwissen zu den verwendeten Begrifflichkeiten notwendig ist, welches nicht durchgängig vorausgesetzt werden kann und daher in Begleitliteratur ergänzt und in Weiterbildungsmaßnahmen vertieft werden muss.

2.3.2 Generierung von Feedback auf Systemebene – Evaluationsfunktion

Im Bildungsmonitoring ist man nicht an Ergebnissen einzelner Klassen interessiert, sondern vorrangig an Ergebnissen von Systemen, z. B. kompletten Schulen, Bezirken, Schularten oder Ländern. Ebenso ist man nicht interessiert an Leistungsdaten zu sehr engen Lernzielen, wie sie für die Unterrichtsgestaltung (z. B. „xyz behandeln wir die kommenden beiden Wochen, siehe Schulbuch S. x–y“) relevant sind, sondern an breiteren Kompetenzbereichen, wie sie in Bildungsstandards Verwendung finden: „Sind die Leistungen von Schülerinnen und Schülern im Kompetenzbereich ‚Statistische Darstellungen und Kenngrößen‘ in vergleichbaren Schulen oder in den Bundesländern ähnlich?“

Da die möglichen Handlungsableitungen, z. B. Konzeption von Weiterbildungen, Konzeption von Schularten, Klassengrößen, Fördermaßnahmen für Sprache im Mathematikunterricht ..., allgemeinerer Art sind, können auch die benötigten Daten und Beschreibungen von Zusammenhängen hierzu allgemeinerer Art sein. Dies zeigt sich u. a. im Anliegen, zu klären, ob erfasste Unterschiede signifikant, d. h. nicht plausibel durch Zufallsschwankungen erklärbar sind oder nicht. Hierin kommt der Schluss von einer Stichprobe (meist beantworten ein paar Tausend Schüler/innen von in Österreich insgesamt ca. 80.000 eine bestimmte Aufgabe) oder Einmalmessung auf eine verallgemeinerte Aussage zum Ausdruck.

Dazu werden oftmals Vertrauensintervalle angegeben: Als Faustregel kann dabei angenommen werden, dass ein signifikanter Unterschied vorliegt, wenn sich die Vertrauensintervalle zu zwei Messwerten nicht überschneiden. Hier kommt die Möglichkeit der statistischen Analyse größerer Datensätze zur Anwendung, auch seltene Phänomene oder schwache Zusammenhänge zwischen Phänomenen, z. B. dem Zusammenhang zwischen Bildungshintergrund und Leistung, oder mögliche Leistungsunterschiede zwischen Stadt und Land oder Bundesländern aufzeigen zu können. Beispiele hierfür werden z. B. in Kapitel 5 dieses Buchs erläutert.

2.3.3 Generierung von Feedback auf Unterrichtsebene – Förderfunktion

An Lehrkräfte rückgemeldete Ergebnisse aus Bildungsstandardüberprüfungen (Abbildung 2) können sich u. a. beziehen auf:

- den Mittelwert der eigenen Klasse
 - absolut als (abstrakter) Skalenwert oder
 - im Sinne der Verteilung der Schülerleistungen auf die Kompetenzstufen (erreicht/ nicht erreicht ...) oder
 - als sozialer Vergleich mit dem Österreichmittelwert inklusive Konfidenzintervall oder

- als „fairer Vergleich“, der über statistische Verfahren bekannte Einflussgrößen auf schulische Leistungen wie soziale Herkunft der Schüler/innen oder die Erstsprache im Elternhaus berücksichtigt;
- die Leistungsstreuung
 - einzelne Schülerleistungen werden dabei als Punkte auf der Punkteskala angeordnet. Weiter oben bedeutet leistungstärker, weiter unten leistungsschwächer.

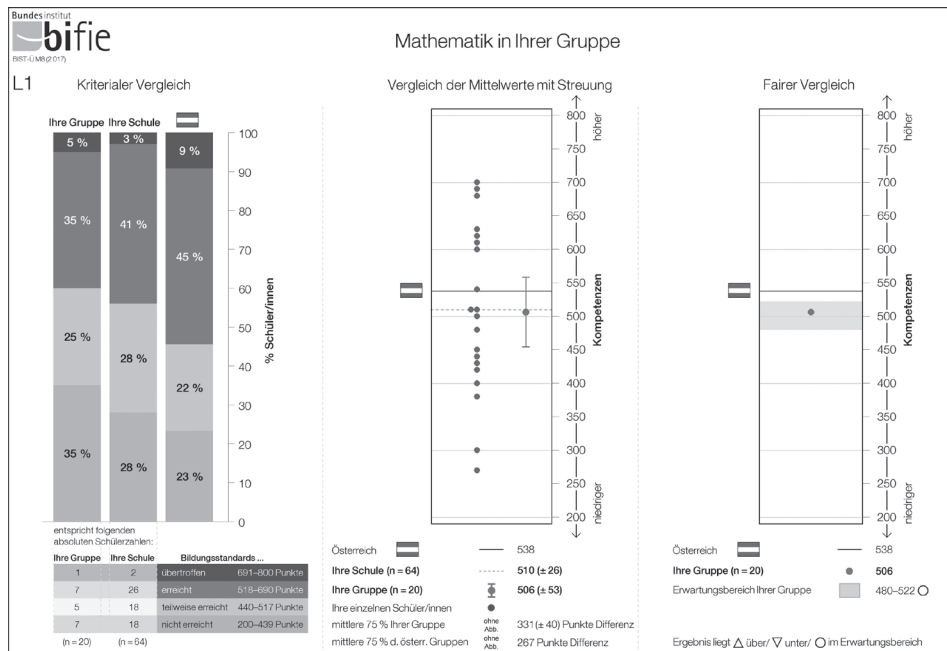


Abbildung 2: Musterrückmeldung zu einer Klasse – Gesamtleistung im Fach Mathematik inklusive fairem Vergleich (BIFIE, 2017b)

Entsprechende Rückmeldungen zum Leistungsstand der eigenen Klasse können von Lehrkräften in zwei Perspektiven interpretiert werden:

a) Rückschauerspektive/summativ:

Bei sozialen Vergleichen kann eine „faire“ Rückmeldung (Nachtigall & Kröhne, 2006) hilfreich sein, wenn leistungsrelevante Hintergrundmerkmale berücksichtigt werden (Groot-Wilken, Isaac & Schräpler, 2016; Schreiner & Wiesner, 2019). Ziel ist, die tatsächliche Qualität schulischer Arbeit vom Einfluss bekannter Hintergrundmerkmale wie beispielsweise der sozialen Herkunft oder der Erstsprache zu trennen. Allerdings geben die Leistungsrückmeldungen keinen Hinweis darüber, worin die vorhandene oder nicht vorhandene Qualität bestehen könnte, die zum Leistungsunterschied geführt hat. Die Rückmeldungen müssen demnach mit Wissen über die Unterrichtsqualität verknüpft werden, optimalerweise auch mit Wissen über alternative Realisierungen von Unterrichtsqualität als in der eigenen Klasse umgesetzt. Falls beispielsweise der vorangegangene Unterricht sehr von individualisierten

Arbeitsformen ohne kooperative Vergleiche von Lösungswegen oder ohne von der Lehrkraft sprachlich angereicherten Austausch im Plenum geprägt war, könnte das Wissen hierüber als hypothetischer Ausgangspunkt für die Erklärung einer unzufriedenstellenden Leistungsrückmeldung herangezogen werden.

b) Vorschauperspektive/formativ:

Zukünftiger Unterricht muss den bestehenden Herausforderungen gerecht werden, unabhängig davon, worauf diese statistisch zurückgeführt werden können. Dementsprechend könnten transparente und nachvollziehbare Nivellierungen von Unterschieden beim fairen Vergleich, z. B. nach sprachlichem Hintergrund der Schüler/innen, andeuten, worauf zukünftig oder weiterhin besonderer Wert gelegt werden könnte, z. B. auf sprachsensiblen Mathematikunterricht. Dies wäre eine mögliche Handlungsableitung, die generell für die eigene Gestaltung von Mathematikunterricht relevant wäre, wie er mit Bildungsstandardmessungen erfasst wird.

Davon zu unterscheiden wäre eine formative Nutzung in Bezug auf konkrete Lernziele, z. B. „statistische Darstellung und Kenngrößen“. Wie bereits erwähnt und nachfolgend in Abschnitt 4 beschrieben, sind die Testdesigns von Bildungsstandards hierfür jedoch nicht konzipiert: In Summenwerten oder Skalenwerten zu mehreren Aufgaben und teils auch in einzelnen Testaufgaben vermischen sich zu viele Teil-Kompetenzen, die später nicht statistisch voneinander getrennt und separiert berichtet werden können (vgl. Götz & George in diesem Band).

3 Bildungsstandards Mathematik für die Sekundarstufe 1

3.1 Theoretisches Kompetenzmodell

Im Zuge der Bildungsstandardverordnung von 2009 (§ 7, 2. Teil, 3. Abschnitt, gesamte Rechtsvorschrift für Bildungsstandards im Schulwesen i. d. F. BGBl. II Nr. 548/2020) wurden in Mathematik Bildungsstandards für das Ende der 8. Schulstufe definiert. Diese basieren auf dem österreichischen Lehrplan der Sekundarstufe 1 für (Neue) Mittelschulen (kurz NMS bzw. MS) (§ 2 Artikel 2, Anlage 1, gesamte Rechtsvorschrift für Lehrpläne der Mittelschulen i. d. F. BGBl. II Nr. 379/2020) und allgemeinbildende höhere Schulen (kurz AHS). Um die Kompetenzen zu gliedern, wurde ein Kompetenzstrukturmodell in drei verschiedenen Dimensionen entwickelt. Mithilfe eines Kompetenzstrukturmodells kann ein Kompetenzbereich in verschiedenen Dimensionen differenziert werden (Leutner, Klieme, Fleischer & Kuper, 2013). Das Kompetenzmodell für Mathematik auf der 8. Schulstufe (BIFIE, 2011) definiert dabei eine mathematische Inhalts-, eine mathematische Handlungs- und eine Komplexitätsdimension. Es werden jeweils vier verschiedene Inhalts- und Handlungsbereiche sowie drei Komplexitätsbereiche unterschieden (vgl. Tabelle 1).

Tabelle 1: Kompetenzbereiche im Kompetenzmodell Mathematik Sekundarstufe 1 (BIFIE, 2011)

Mathematischer Inhalt	Mathematisches Handeln	Komplexität
I1: Zahlen und Maße	H1: Darstellen, Modellbilden	K1: Einsetzen von Grundkenntnissen und -fertigkeiten
I2: Variable, funktionale Abhängigkeiten	H2: Rechnen, Operieren	K2: Herstellen von Verbindungen
I3: Geometrische Figuren und Körper	H3: Interpretieren	K3: Einsetzen von Reflexionswissen, Reflektieren
I4: Statistische Darstellung und Kenngrößen	H4: Argumentieren, Begründen	

Jede mathematische Aufgabe kann zumindest einem Bereich jeder Dimension zugeordnet werden. Wählt man jeweils einen Inhalts-, Handlungs- und Komplexitätsbereich, ergibt sich ein Knotenpunkt im Kompetenzmodell (vgl. Abbildung 3). In Abbildung 3 ist der Knotenpunkt (H1, I4, K1) im Kompetenzmodell visualisiert. Insgesamt gibt es 48 verschiedene Kombinationen für solche Knotenpunkte. In den Bildungsstandards werden Kompetenzbeschreibungen zu jedem Knotenpunkt angeführt. Zum Beispiel:

(H1, I4, K1)

Die Schülerinnen und Schüler können gegebene statistische Sachverhalte (Daten) in eine (andere) mathematische Darstellung übertragen, wobei dafür das unmittelbare Einsetzen von Grundkenntnissen erforderlich ist.

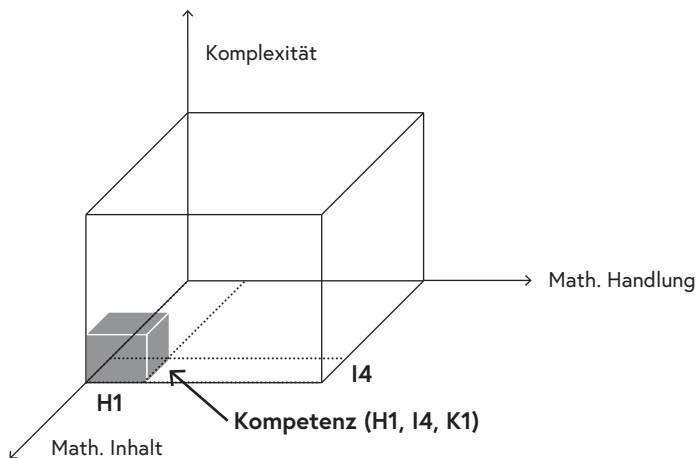


Abbildung 3: Kompetenzmodell Sekundarstufe 1 (BIFIE, 2011, S. 9)

Im Rahmen der Bildungsstandardüberprüfungen für Mathematik auf der 8. Schulstufe wurden Aufgaben verwendet, welche in diesem Modell genau einem Knotenpunkt zu-

geordnet werden können. Eine objektive Zuordnung von Aufgaben zu einem Komplexitätsbereich gestaltet sich oftmals als schwierig. Zudem erlaubt es der Umfang des Tests nicht, alle 48 Knotenpunkte mit genügend Testaufgaben abzubilden, um weiterhin eine genügend aussagekräftige Rückmeldung für jeden Knotenpunkt zu gewährleisten. Daher wurde für die BIST-Ü Mathematik auf der 8. Schulstufe entschieden, auf die Komplexitätsdimension zu verzichten und auch in der Rückmeldung nicht weiter auf diese Dimension hinzuweisen. Das Modell für das Testkonstrukt wurde um die Komplexitätsdimension reduziert. Es ergibt sich ein vereinfachtes Modell (siehe Abbildung 4). Jede Testaufgabe (vgl. 2.3 Entwicklung von Lernaufgaben aus Testaufgaben) wird genau einem mathematischen Inhalts- und Handlungsbereich zugeordnet. Im Test kommen weiterhin Aufgaben zu verschiedenen Komplexitätsbereichen zum Einsatz. Diese werden allerdings bei der Testheft-erstellung und Rückmeldung nicht gesondert berücksichtigt.

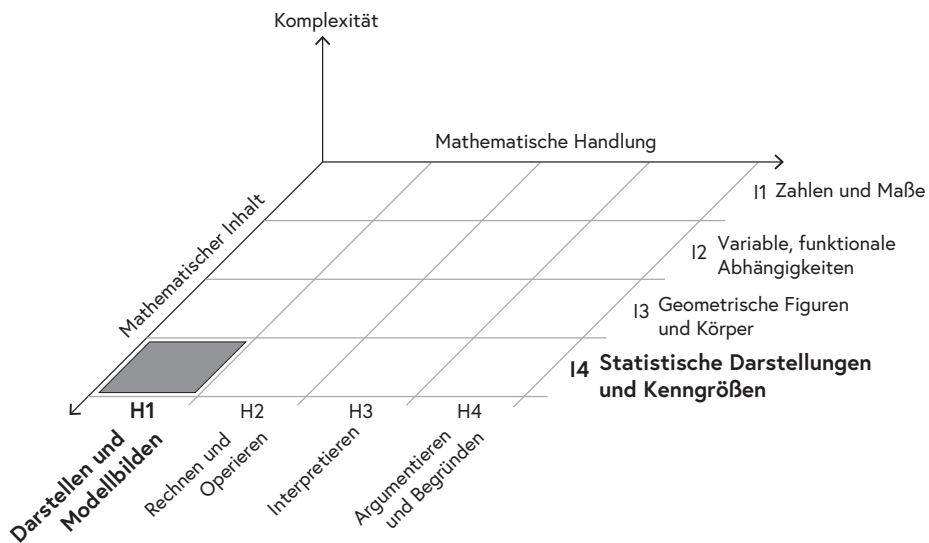


Abbildung 4: Vereinfachtes Kompetenzmodell Mathematik Sekundarstufe 1; Abbildung nach George, Robitzsch & Schreiner, 2019

Testitem M83716 (vgl. Abbildung 1) wurde im Kompetenzmodell dem Knoten (H1, I4, K1) zugeordnet. In der Aufgabe soll ein gegebener statistischer Sachverhalt (Tabelle mit Schularbeitsnoten) in eine andere mathematische Darstellungsform (Kreisdiagramm) übertragen werden. Für die Rückmeldung ist diese Aufgabe für die Kompetenzbereiche H1 und I4 relevant.

3.2 Empirische Modellierung des Kompetenzmodells

3.2.1 Aufgabenentwicklung, Testheftgenerierung, Pilotierung

Um den getesteten Kompetenzbereich auf Schul- und Systemebene möglichst in seiner Gesamtheit abzudecken, wurden im Zuge der BIST-Ü mehrere Testheftformen eingesetzt (Schreiner et al., 2018). Die Testheftformen weisen vergleichbare Schwierigkeiten auf, be-

inhalten Aufgaben zu jedem Knotenpunkt des Kompetenzmodells sowie eine möglichst gleichmäßige Verteilung der Aufgaben zu allen Inhalts- und Handlungsbereichen. Um diese Kriterien erfüllen zu können, ist für die Erstellung der Testhefte ein entsprechend breiter und ausgewogener Pool an Testaufgaben notwendig.

Für die Entwicklung und Aufrechterhaltung eines solchen Pools werden beinahe jährlich Itementwicklungsprozesse in Expertengruppen durchgeführt. Diese Prozesse zielen darauf ab, „einen möglichst großen Pool qualitativ hochwertiger Items zu erstellen, der den jeweiligen Fachbereich bzw. die gesuchte Kompetenz valide abbildet [...] und reliable und objektive Messergebnisse produziert“ (Itzlinger-Bruneforth & Kuhn, 2015, S. 3). Die Expertengruppen setzen sich dabei aus Lehrpersonen der AHS und NMS sowie Expertinnen und Experten aus den Bereichen Fachdidaktik und Psychometrie zusammen. Die Aufgaben werden dabei in erster Linie von Lehrpersonen entwickelt und durchlaufen mehrere Begutachtungsschleifen in der Expertengruppe. Dabei wird jede Aufgabe nach verschiedensten Qualitätskriterien beurteilt und weiterentwickelt.

Alle Aufgaben, die in der Bildungsstandardüberprüfung zum Einsatz kommen, werden zuvor pilotiert. In der Pilotierung werden die Aufgaben an Schülerinnen und Schülern in einer standardisierten Testsituation getestet. Jede Aufgabe wird in einer für die österreichische Schülerschaft repräsentativen Stichprobe eingesetzt. Dabei wird unter anderem auf eine repräsentative Verteilung bezüglich Geschlechts, besuchter Schulform und weiterer Merkmale geachtet.

Ziel der Pilotierung ist das Ermitteln von Itemschwierigkeiten in Form von Lösungshäufigkeiten und ein Feststellen, ob Aufgaben für den Einsatz in der eigentlichen Überprüfung geeignet sind. Ausschlussgründe sind beispielsweise eine zu hohe oder zu niedrige Schwierigkeit der Aufgabe, Antwortoptionen, welche zu häufig oder zu selten gewählt werden, Aufgaben, welche häufiger durch leistungsschwache als leistungsstarke Schüler/innen korrekt gelöst werden (Trennschärfe), eine Benachteiligung gewisser Gruppen (beispielsweise Mädchen oder Jungen) (Itzlinger-Bruneforth & Kuhn, 2015). Die ermittelte Schwierigkeit wird später auch bei der Erstellung des Testdesigns, also der methodischen und inhaltlichen Zusammenstellung der verschiedenen Testheftformen, verwendet.

3.2.2 Kompetenzstufenbeschreibungen und Standard-Setting

Wie bereits beschrieben wurde, ist ein wesentliches Ziel der Standardüberprüfungen, festzustellen, in welchem Grad Schüler/innen Bildungsstandards erreichen. Um diesen Grad zu beschreiben, werden Kompetenzstufen definiert. Im Fall der Bildungsstandards wurde ein vierstufiges Kompetenzstufenmodell entwickelt. Solche Stufenmodelle beschreiben die Ausprägung einer bestimmten Kompetenz und sind somit ein wesentlicher Bestandteil im Setzen von Standards und Kriterien (Freunberger & Yanagida, 2012; Pellegrino & Chudowsky, 2003).

Die vier Stufen unterteilen sich in „Bildungsstandards nicht erreicht“, „teilweise erreicht“, „erreicht“ und „übertrifft“. Die inhaltliche Beschreibung dieser Stufen wird in Form von schülerbezogenen „Can-do-Statements“ formuliert (Abbildung 5).

Mathematik: Inhaltliche Beschreibung der einzelnen Kompetenzstufen		
3	Bildungsstandards übertroffen	691–800 Punkte
Die Schüler/innen verfügen über grundlegende Kenntnisse und Fertigkeiten in allen Teilbereichen des Lehrplans Mathematik und über erweiterte Wissensstrukturen, welche über die Anforderungen der Stufe 2 hinausgehen, insbesondere über stärker ausgeprägtes Abstraktionsvermögen und höhere Kombinationsfähigkeit. Sie können diese eigenständig in neuartigen Situationen flexibel einsetzen.		
2	Bildungsstandards erreicht	518–690 Punkte
Die Schüler/innen verfügen über grundlegende Kenntnisse und Fertigkeiten in allen Teilbereichen des Lehrplans Mathematik und können diese flexibel nutzen. Sie können geeignete Lösungsstrategien finden und umsetzen, gewählte Lösungswege beschreiben und begründen. Sie können mit verbalen, grafischen und formalen Darstellungen mathematischer Sachverhalte flexibel umgehen und diese angemessen verwenden. Sie können relevante Informationen aus unterschiedlich dargestellten Sachverhalten (z. B. Texte, Datenmaterial, grafische Darstellungen) entnehmen und sie im jeweiligen Kontext deuten. Sie können ihre mathematischen Kenntnisse miteinander in Verbindung setzen sowie mathematische Aussagen kritisch prüfen, bewerten und/oder begründen.		
1	Bildungsstandards teilweise erreicht	440–517 Punkte
Die Schüler/innen verfügen über grundlegende Kenntnisse und Fertigkeiten in allen Teilbereichen des Lehrplans Mathematik und können damit reproduktive Anforderungen bewältigen und Routineverfahren durchführen.		
unter 1	Bildungsstandards nicht erreicht	bis 439 Punkte

Abbildung 5: Kompetenzstufenbeschreibung Bildungsstandardüberprüfung Mathematik 8 (Schreiner et al., 2018, S. 37)

In Abbildung 5 finden sich auch sogenannte Cut-Scores auf einer kontinuierlichen Punkteskala, beispielsweise 517 bzw. 518 zwischen Stufe 1 und 2. Die Cut-Scores ermöglichen es, die Testpersonen den jeweiligen Kompetenzstufen zuzuordnen (Freunberger & Yanagida, 2012). Die Cut-Scores wurden im Zuge eines Standard-Settings erarbeitet. Dabei entscheidet eine Expertengruppe über die Cut-Scores zwischen den einzelnen Kompetenzstufen. Die Expertengruppe setzte sich für Mathematik auf der 8. Schulstufe aus Vertreterinnen und Vertretern der Fachdidaktik, der Eltern, des Schulrats, der Berufsschule, des Ministeriums, der Lehrpersonen, der Testpsychologie und der Testentwicklung zusammen.

Beim Standard-Setting für Mathematik auf der 8. Schulstufe wurde auf die Item-Descriptor-Matching-Methode zurückgegriffen. Dabei erhalten die Teilnehmer/innen aus der Expertengruppe ein Booklet mit Items. Die Items sind aufsteigend nach der Schwierigkeit, die in der Pilotierung für die BIST-Ü 2012 erhoben wurde, geordnet. Die Expertengruppe ordnet jedes Item jener Kompetenzstufenbeschreibung zu, welche das zur Beantwortung des Items benötigte Wissen, die verlangten Fähigkeiten und kognitiven Prozesse am besten ausdrückt (Freunberger & Yanagida, 2012).

Im gegebenen Fall wurde die Zuordnung der Items zu den Stufenbeschreibungen für jedes Mitglied der Expertengruppe grafisch dargestellt (Abbildung 6). Anschließend wurden die Kurven geglättet, indem Ausreißer aus den Kurven entfernt wurden. An jener Stelle, an der Items „kontinuierlich“ einer höheren Stufe zugeordnet wurden, ergibt sich ein individueller Cut-Score. Aus den individuellen Cut-Scores der einzelnen Expertinnen und Experten wurde mithilfe eines Mittelwerts ein Gesamt-Cut-Score errechnet. Der Prozess wurde dabei in drei Runden durchgeführt. Zwischen den Runden wurde der Experten-

gruppe Feedback zu den Ergebnissen vorgelegt. Damit sollte die Gruppe über die drei Runden hinweg zu konsistenteren Entscheidungen kommen. Detailliertere Informationen zum Ablauf und den Ergebnissen des Standard-Setting-Prozesses finden sich in der technischen Dokumentation zum Standard-Setting (Freunberger, 2013).

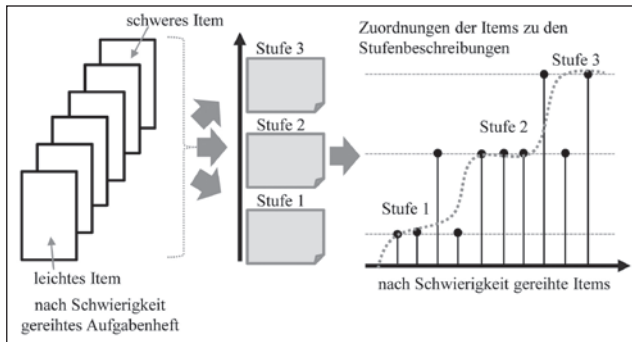


Abbildung 6: Item-Descriptor-Matching-Methode (Freunberger & Yanagida, 2012)

4 Ableitung von Handlungsmöglichkeiten

In die Implementierung und Nutzung von Bildungsstandardrückmeldungen sind im Schulsystem mehrere interagierende Teilsysteme (Klassen, Schulen, Schulverwaltung, Politik) involviert. Die individuellen Akteure in diesen Teilsystemen sind unterschiedlichen Zielen verpflichtet, von der Durchführung des täglichen Unterrichts bis zur Bildungssystemsteuerung.

Dementsprechend bestehen in der Forschungsliteratur verschiedene Modelle zu Wirkungsweisen von Bildungsstandardüberprüfungen (Kowalk, in Vorbereitung), in denen unterschiedliche Perspektiven zum Ausdruck kommen, von der Feedbackforschung (Hattie & Timperley, 2007; Koch, 2011) über die Schulevaluationsforschung (Maag Merki, 2010; Visscher & Coe, 2003) hin zur Assessmentforschung (Bennett, 2010) bis zur Schulentwicklungsforschung (Helmke & Hosenfeld, 2005; Wiesner, Schreiner, Breit & Kemethofer, 2017).

Im Kern sollen Bildungsstandardrückmeldungen Schulen und Lehrkräften helfen, die Ergebnisse und Prozesse ihrer Arbeit einzuschätzen, blinde Flecken zu erkennen und die Schul- und Unterrichtsentwicklung anzuregen (George, Süss-Stepancik, Illetschko & Wiesner, 2016). Diese Perspektive bekräftigt das Angebots-Nutzungs-Modell zu Lernstandsdiagnosen (Leuders et al., 2019) (Abbildung 7), da dieses insbesondere bestehende Befunde zur Wirksamkeit von Lehrerprofessionalisierungsmaßnahmen (vgl. Lipowsky & Rzejak, 2017) bei der Beschreibung der intendierten Wirkzusammenhänge berücksichtigt: Bildungsstandardrückmeldungen als Maßnahmen zur Schul- und Unterrichtsentwicklung sollen zu einer Förderung der professionellen Kompetenz von Lehrenden beitragen, was weitergehend zu einer Steigerung von Lernerfolgen bei Schülerinnen und Schülern führen soll.

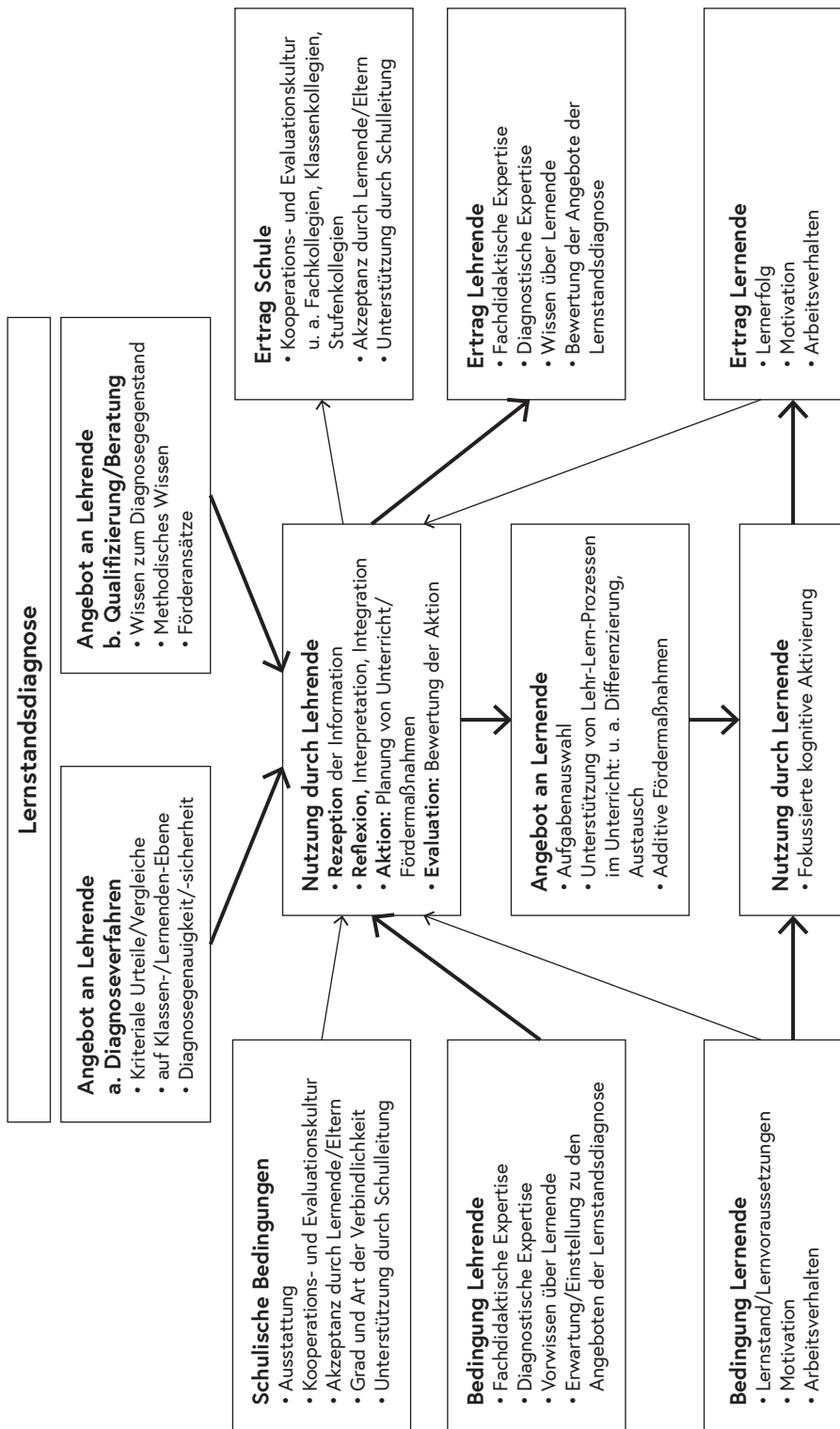


Abbildung 7: Angebots-Nutzungs-Modell zu Lernstandsdiagnosen (vgl. Leuders, Schulz & Kowalk, 2019)

Bildungsstandardüberprüfungen machen im Sinne dieser intendierten Wirkungskette ein zweifaches Angebot an Lehrkräfte: a. in Form von Ergebnissrückmeldungen zu den Schülerinnen und Schülern der eigenen Klasse oder Schule und b. in Form begleitender Qualifizierungsmaßnahmen, z. B. zum Verständnis der Ergebnisdarstellungen, zu den Kompetenzmodellen der Bildungsstandards und zu unterrichtlichen Förderansätzen.

Im Angebots-Nutzungs-Modell zu Lernstandsdiagnosen wird fachdidaktischen Kompetenzen bei den Lehrenden eine zentrale Rolle zugeschrieben, da diese die Grundlage bilden für die *Rezeption und Reflexion* der Ergebnissrückmeldungen, sodann für die *Aktion und Evaluation*. Eine solche Nutzung (Rezeption und Reflexion, Aktion und Evaluation) von Ergebnissrückmeldungen und Qualifizierungsangeboten soll sich über ein verändertes Angebot an die Lernenden (Unterrichtsmaßnahmen) auf deren Lernerfolg auswirken.

Auf Grundlage der hierbei gewonnenen neuen Erfahrungen werden zudem erweiterte fachdidaktische Kompetenzen der Lehrpersonen selbst sowie weiterführende Maßnahmen der Schulentwicklung erwartet (Leuders et al., 2019). Letztere könnten beispielsweise die Einrichtung entsprechender zusätzlicher Förderangebote für Schüler/innen mit sprachlichen Schwierigkeiten, Projektgruppen für Hochbegabte, vermehrte Fachkonferenzen oder auch eine langfristige Bildung professioneller Lerngemeinschaften im Schulkollegium bzw. in Kooperation mit benachbarten Schulen beinhalten (vgl. Wiesner & Schreiner, 2019).

Zur Nutzung von Ergebnissrückmeldungen aus Bildungsstandardüberprüfungen bzw. allgemeiner aus externen großflächigen Lernstandsdiagnosen bestehen vielfältige Befunde (Kowalk, in Vorbereitung). Nachfolgend sind solche zusammengefasst, die einen spezifischen Bezug zur fachdidaktischen Unterrichtsentwicklung aufweisen. Zu beachten ist, dass die meisten Befunde aufgrund ihres überwiegend speziellen Kontextes und zum Teil geringer Rücklaufquoten nur eingeschränkt verallgemeinerbar und möglicherweise positiv verfälscht sind (Müller, 2014). Generell lassen sich Ergebnissrückmeldungen dahingehend unterscheiden, ob sie rohe Vergleichsdaten (Punkte- oder Skalenwerte), soziale Vergleiche (Bezug zu anderen Gruppen oder Durchschnitten) oder kriteriale Vergleiche (Bezug auf einen bestimmten Kenntnis- bzw. Fähigkeitsstand) liefern (Heckhausen, 1974).

Zur *Rezeption und Reflexion* wird aus mehreren Studien berichtet (Kowalk, in Vorbereitung), dass Lehrkräfte ein höheres Interesse an kriterialen Rückmeldungen bzw. kompetenzorientierten Rückmeldungen als an sozialen Vergleichen haben und sich mit der Bedeutung von Fähigkeitsniveaus befassen (Groß Ophoff, 2013; Kühle & Peek, 2007). Es bestehen jedoch auch Schwierigkeiten beim Verständnis von kriterialen Rückmeldungen: In diesen enthaltene fachdidaktische Informationen werden teils vernachlässigt und anstattdessen Noten als Bewertungsmaßstab herangezogen, oder Lehrkräfte fokussieren zu sehr auf eingesetzte Testaufgaben, statt sich mit Fähigkeitsbeschreibungen auf Niveaus auseinanderzusetzen (Koch, 2011; Schneewind, 2006).

Auf ähnliche Befunde verweisen Maier, Ramsteck und Frühwacht (2013): Lehrkräfte greifen bei der Interpretation von Ergebnissrückmeldungen vorwiegend auf Praxiserfahrungen

und einfache Lehr-Lern-Theorien zurück. Dabei konzentrieren sie sich häufig auf den konkreten Inhalt von einzelnen Aufgaben. Kaum Bezug wird dabei zu den gemessenen Kompetenzbereichen oder Kompetenzstufen hergestellt. Zudem geschieht eine Interpretation meist ohne Verwendung fachdidaktischer Theorien. Auch eine Interpretation von rückgemeldeten Fähigkeitsniveaus mithilfe der bestehenden Kompetenzmodelle aus Lernstandserhebungen scheint Lehrkräften schwer zu fallen (Groß Ophoff, 2013), wobei Kompetenzmodellierungen mit geringem Auflösungsgrad als Hürde für die Nutzung der Ergebnisse gelten (Neumann, 2013).

Allgemein setzen sich Lehrkräfte intensiver mit Ergebnismeldungen auseinander, wenn sie diese als verständlich und nützlich einschätzen (Koch, 2011; Kühle, 2010). Hierbei scheinen intensive Auseinandersetzungen eher bei auffälligen und schlechten Ergebnissen stattzufinden, wobei positive Ergebnisse tendenziell als Bestätigung der eigenen Arbeit aufgefasst, schlechte Ergebnisse eher externen Ursachen zugeschrieben werden (Demski, 2017; Groß Ophoff, Hosenfeld & Koch, 2007; Maier, 2008; Schneewind, 2006).

Ergebnisse aus der österreichischen Bildungsstandardüberprüfung verdeutlichen, dass sich bei Schulleitungen die Reflexion von Daten stark auf die Ableitung von Maßnahmen auswirkt, wohingegen die Rezeption (lediglich das Wahrnehmen der Daten) keinen Einfluss auf die Maßnahmenableitung hat. Österreichische Lehrpersonen reflektieren deutlich weniger über Ergebnisse als Schulleitungen. Wenn sie dies jedoch tun, dann wirkt sich dies stark auf die Maßnahmenableitung aus (Wiesner & Schreiner, 2019).

Zur *Aktion und Evaluation* illustrieren verschiedene Studien, dass im Rahmen der Nutzung von Rückmeldungen durch Lehrkräfte kaum konkrete Maßnahmen mit bestenfalls niedrigschwelligen Veränderungen im Unterricht umgesetzt werden (Altrichter, 2010; Groß Ophoff, 2013; Maier, 2008). Wenn Maßnahmen berichtet werden, dann beinhalten diese die Übernahme von Aufgaben aus Tests in den Unterricht, eine Wiederholung bestimmter Unterrichtsinhalte oder eine Zunahme schulinterner Kooperation (Groß Ophoff, Koch, Hosenfeld & Helmke, 2006; Leutner, Fleischer, Spoden & Wirth, 2008; Maier, 2008). Diemer (2013) führt aus, dass selten oder nie Maßnahmen erfolgen, welche eine Veränderung von Lehr-Lern-Prozessen beinhalten. Zudem scheinen Lehrkräfte eher bereit, Ergebnismeldungen im Rahmen einer Leistungsbeurteilung und Lern diagnose zu verwenden, als didaktische Maßnahmen zur Weiterentwicklung des Unterrichts abzuleiten (Maier, 2009).

Als mögliche Ursache für solche Befunde zur relativ geringen und teils sogar rückläufigen Nutzung von Rückmeldungen aus großflächig eingesetzten, standardisierten Tests (z. B. Bildungsstandardüberprüfungen, Vergleichsarbeiten) wird angenommen, dass die Rückmeldungen keine spezifischen und für den Unterricht verwendbaren Impulse bereitstellen (Altrichter, 2010; van Ackeren, 2003). Dies wird auch so formuliert, dass Lernstandserhebungen lediglich unzusammenhängende fachdidaktische Botschaften zurückmelden, welche die Informationsnutzung nicht angemessen unterstützen (Netzwerk zur empiriegestützten Schulentwicklung, 2008).

Derartige empirische Befunde gehen in die Richtung dessen, was theoretisch-analytisch von Oelkers und Reusser (2008) als „Rückverflüssigungsproblem“ sowie von Fend (2008) als „Übersetzungsleistung“ bezeichnet wurde: Kompetenzbeschreibungen in Bildungsstandards sind notwendigerweise abstrakt und allgemein, somit zu wenig konkret und zu wenig detailliert, als dass sich aus ihnen alleine oder aus darauf bezogenen Rückmeldungen direkt konkrete Handlungen für die Unterrichtsgestaltung ableiten ließen.

Empirisch erfasst und analysiert wurde diese notwendige Rekontextualisierung von Bildungsstandards durch Lehrkräfte mittlerweile mehrfach, beispielsweise im Kontext der Implementierung der amerikanischen NCTM-Standards (Hill, 2001) oder der deutschen Bildungsstandards (Zeitler, Heller & Asbrand, 2013). Auch im Rahmen der Implementierung luxemburgischer Bildungsstandards konnte dokumentiert werden, dass für Lehrkräfte ein Verständnis dessen, was als Ergebnisorientierung oder Kompetenzorientierung bezeichnet wird, auf einem aufwändigen Rekonstruktionsprozess basiert. Hierbei werden Kompetenzbeschreibungen aus Bildungsstandards aktiv sinnstiftend und mit Rückgriff auf persönliche Erfahrungen, Wissen und Sichtweisen interpretiert (Schulz, 2010).

Speziell bei der Nutzung von Lernstandserhebungen für nachfolgende unterrichtliche Fördermaßnahmen sind zudem Diskussionsbeiträge aus der Förderdiagnostik beachtenswert (Moser Opitz, 2022; Schlee, 2002; Wember, 1998): Die Ableitung von Fördermaßnahmen stellt einen eigenständigen Konstruktionsprozess dar, der bereits vor der Diagnostik bestehendes Handlungs- und Planungswissen bedarf. Dass sich angeblich Fördermaßnahmen direkt aus Diagnoseergebnissen ableiten ließen, wird als logischer Fehlschluss angesehen.

Zusammengefasst stützen all diese Befunde und Analysen den Kern des Angebots-Nutzungs-Modells zu Lernstandsdiagnosen (Leuders et al., 2019): Fachdidaktische Kompetenzen, die u. a. ein Verständnis verwendeter Kompetenzbeschreibungen oder Wissen über entsprechend fokussierte Fördermaßnahmen beinhalten, stellen eine wesentliche Bedingung für eine effektive Nutzung von Rückmeldungen aus Lernstandserhebungen durch Lehrkräfte dar (Abbildung 7).

Erfreulicherweise gibt es Entwicklungen und Befunde, die illustrieren, wie sich durch eine gezielte fachdidaktische Unterstützung von Lehrpersonen und Schulen die Nutzungsqualität von Bildungsstandardrückmeldungen für die Unterrichtsentwicklung steigern lässt. Insbesondere zwei, optimalerweise gemeinsam realisierte Aspekte werden hierbei vorrangig diskutiert und beschrieben (Altrichter & Kanape-Willingshofer, 2012; Asbrand, Heller & Zeitler, 2012; Kowalk, in Vorbereitung; Kowalk, Leuders & Schulz, 2018; Krainer et al., 2012; Leuders et al., 2019; Schulz, 2010):

- a) *Begleitende zielgerichtete Qualifizierungsangebote für Lehrkräfte:* Dies beinhaltet die begleitende Distribution von Informations- oder Unterrichtsmaterialien, die Etablierung von institutionalisierten Multiplikatorinnen und Multiplikatoren zur Schul- und Unterrichtsentwicklung bis hin zu längerfristigen Fortbildungsangeboten. Fortbildungsinhalte können u. a. fachdidaktische Hintergründe von Kompetenzmodellen

und Ergebnismeldungen sowie darauf abgestimmte Fördermöglichkeiten sein. Beispiele hierfür finden sich auf der Website des IMST-Netzwerkes (imst.ac.at) oder in fachdidaktischen Erläuterungen und Konkretisierungen zum österreichischen Standards-Konzept M8 (Kröpfel & Schneider, 2012).

- b) Entwicklung und Implementierung von *Lernstandserhebungen*, die sich weniger an den Bedürfnissen des Bildungsmonitorings, sondern stärker an *unterrichtsrelevanter Diagnose* orientieren und dabei kriteriale Rückmeldungen mit *engem fachlichem Fokus* anstelle von Rückmeldungen zu breiten Kompetenzdimensionen generieren (vgl. Mislevy et al., 2003; Pellegrino et al., 2001). Aktuelle Beispiele hierfür sind die informellen Kompetenzüberprüfungen in Österreich „IKM“ (iqs.gv.at/ikm), die baden-württembergische Eingangsdiagnose „Lernstand 5“ (lernstand5-bw.de) (Schulz, Leuders & Rangel, 2017, 2020) oder die individuellen Lernstandsanalysen „ILeA plus“ in Berlin und Brandenburg (isq-bb.de/wordpress/werkzeuge/ileaplus/).

Inwiefern fachdidaktische Hintergründe zu Kompetenzmodellen und Ergebnismeldungen sowie darauf abgestimmte Unterrichtskonzepte einen Beitrag zur Nutzung von Ergebnismeldungen leisten können, wird in den folgenden Abschnitten an zwei Beispielen veranschaulicht:

1. Ableitung von Maßnahmen zur Differenzierung (siehe 4.1),
2. Entwicklung von Lernaufgaben aus Testaufgaben (siehe 4.2).

Hierbei werden entsprechend dem Angebots-Nutzungs-Modell zu Lernstandsdiagnosen (Abbildung 7; Leuders et al., 2019) die beiden Schritte diskutiert

- a) Rückmeldungen verstehen (Rezeption und Reflexion) sowie
- b) Handlungsableitungen entwickeln (Aktion und Evaluation).

Weitere Konkretisierungen von Handlungsableitungen aus Bildungsstandardüberprüfungen finden sich in den nachfolgenden Kapiteln dieses Sammelbandes.

4.1 Ableitung von Maßnahmen zur Differenzierung

Guter Unterricht adressiert nicht an (fiktive) durchschnittliche Schüler/innen, sondern sollte die Leistungsstreuung und Vielfalt der gesamten Lerngruppe / Klasse berücksichtigen (Leuders & Prediger, 2017). Das Anliegen, Maßnahmen zur Differenzierung anzuregen, findet sich entsprechend auch in Begleitmaterialien zur Nutzung der österreichischen Bildungsstandards (Mürwald-Scheifinger, 2012; Neureiter, 2012). Abbildung 2 (siehe oben: Musterrückmeldung zu einer Klasse – Gesamtleistung im Fach Mathematik) beinhaltet Informationen, wie eine möglicherweise typische Klasse der eigenen Schule hinsichtlich Leistungsniveaus im Fach Mathematik zusammengesetzt ist bzw. zum Erhebungszeitpunkt war. Auch wenn die entsprechende Erhebung schon etwas zurückliegt und sich auf eine Klasse eines vorangehenden Jahrgangs bezieht, kann man doch davon

ausgehen, dass die beschriebene Leistungsheterogenität einen verlässlichen Anhaltspunkt auch für aktuelle Klassenzusammensetzungen im beispielsweise gleichbleibenden Schuleinzugsgebiet liefert.

4.1.1 Rezeption und Reflexion – Streumaße verstehen und im Unterrichtskontext interpretieren

In Abbildung 2 (siehe oben: Musterrückmeldung zu einer Klasse – Gesamtleistung im Fach Mathematik) müssen die dargebotenen Informationen zunächst von einer Lehrkraft verstanden werden, bevor im summativen Sinne über mögliche Ursachen oder im formativen Sinne über mögliche nachfolgende Handlungsentscheidungen nachgedacht werden kann. Dies beinhaltet zunächst ein technisches Verständnis der Balkendiagramme, Punktwolken inkl. Streuungen und Tabellen oder des Konfidenzintervalls für den Klassenmittelwert im Vergleich zum Landesdurchschnitt. Hiermit haben Mathematiklehrkräfte der Sekundarstufe vermutlich kaum Schwierigkeiten.

Die einzelnen Punkte im Streudiagramm (rechts / Mitte in Abbildung 2) repräsentieren die Leistungsstände einzelner Schüler/innen, die sich mittels der in der Tabelle in Abbildung 2 unten links angegebenen Punktspannen den Kompetenzstufen von „nicht erreicht“ bis „übertrifft“ zuordnen lassen.

Beschreibungen für die Kompetenzstufen sind allgemeiner Art. Daher ist für ihr Verständnis beispielsweise eine Orientierung an konkreten Aufgaben hilfreich, die für die Kompetenzstufen typische Leistungsniveaus veranschaulichen (Beispiele: IQS, 2017). Eine ergänzende oder alternative Möglichkeit zur Interpretation, die weniger präzise, jedoch möglicherweise gerade für berufserfahrene Lehrkräfte praktikabel ist, kann darin bestehen, die in den Kompetenzstufen beschriebenen Leistungsniveaus als Lehrkraft mit eigenem Erfahrungswissen zum Erreichen von Mindestanforderungen zu durchschnittlichen oder zu (weit) überdurchschnittlichen Leistungen zu verbinden. Beispielsweise mit der Formulierung zu Kompetenzstufe 1 (Bildungsstandards teilweise erreicht), „Die Schüler/innen verfügen über grundlegende Kenntnisse und Fertigkeiten in allen Teilbereichen des Lehrplans Mathematik und können damit reproduktive Anforderungen bewältigen und Routineverfahren durchführen“ verbinden berufserfahrene Lehrkräfte typische Aufgaben, Lösungsbeispiele, Fehler oder auch individuelle Schüler/innen.

Eine solche Zuordnung wird sicherlich nicht exakt den psychometrischen Skalenwerten und Niveaus der Bildungsstandardüberprüfung entsprechen, kann im Kontext von Überlegungen zu Differenzierungsmaßnahmen für die eigene Klasse aber durchaus ausreichend genau sein: Falls die Rückmeldegrafik beispielsweise aussagt, dass sich ca. 25 % der Schüler/innen auf dem Niveau „teilweise erreicht“ bewegen, ca. 20 % darunter liegen („nicht erreicht“) und andererseits ca. 10 % der Schüler/innen die Bildungsstandards (d. h. ein bis drei Schüler/innen pro Klasse) sogar „übertrifft“ und damit u. a. „über stärker ausgeprägtes Abstraktionsvermögen und höhere Kombinationsfähigkeit“ verfügen sowie „diese eigenständig in neuartigen Situationen flexibel einsetzen“ können, dann verdeutlicht dies eindrücklich den Differenzierungsbedarf für die Unterrichtsgestaltung.

Die Differenz zwischen dargestellten Lernständen, z. B. zwischen der leistungsstärksten und der leistungsschwächsten Gruppe der Lernenden, deutet die Leistungsspanne einer solchermaßen womöglich typischen und auch für die Zukunft erwartbaren Klasse für die eigene Schule an. Eine solche Vorstellung ermöglicht zudem eine inhaltliche Interpretation auch der abgebildeten Differenz zwischen dem eigenen Klassenmittelwert zum Österreichmittelwert:

- Wie groß ist diese Mittelwertdifferenz im Vergleich mit angedeuteten Leistungsunterschieden zwischen entsprechenden Schülergruppen?
- Lässt sich dies zu persönlichen Erfahrungen aus dem Unterricht über möglicherweise typische Leistungsunterschiede zwischen spezifischen Leistungsgruppen von Lernenden in Bezug setzen?

Differenzen oder Leistungsspannen auf der verwendeten Kompetenzskala von 200 bis 800 Punkten können dementsprechend inhaltlich angereichert und besser hinsichtlich ihrer unterrichtlichen Bedeutung interpretiert werden. Dies ergänzt die Information, ob das Konfidenzintervall des Klassenmittelwerts den Österreichmittelwert umschließt oder nicht.

Aus derartigen Zustandsbeschreibungen, basierend auf einer Einmalerhebung, lassen sich jedoch noch keine Hinweise auf Ursachen und Rahmenbedingungen entnehmen, die einen Beitrag zum rückgemeldeten Lernstand geleistet haben könnten. Um entsprechende Vermutungen für Ursachen aufstellen zu können, müssen die Leistungsrückmeldungen mit weiteren Informationen z. B. zur Schule, zu Lehrkräften, zur zurückliegenden Unterrichtsorganisation und -gestaltung sowie zu Merkmalen der Schülerinnen und Schüler ergänzt werden. Schulorganisation und Unterrichtsgestaltung sind naturgemäß hochkomplex. Dies sollte entsprechend bei der Suche nach erklärenden Ursachen für die rückgemeldeten Leistungsdaten mitbedacht werden.

Gleichzeitig konzentrieren sich Maßnahmen immer auf ausgewählte Bedingungen und Wirkzusammenhänge und reduzieren demnach die tatsächlich vorhandene Komplexität auf ein Maß, das Handlungen ermöglicht. Diese Balance aus noch überschaubarer Komplexität und notwendiger Fokussierung auf wenige Kernmerkmale beinhaltet immer eine gewisse Subjektivität dahingehend, was als im Rahmen der eigenen Gestaltungsmöglichkeiten liegend angesehen wird, und was persönlich, z. B. seitens einer Lehrkraft, eines Fachkollegiums oder einer Schulleitung, als bedeutsam und handlungsleitend bewertet wird.

4.1.2 Aktion und Evaluation – Differenzierungsansätze auswählen, planen, durchführen und überarbeiten

Nimmt man als Lehrperson oder Fachkollegium die in einer Bildungsstandardrückmeldung dargestellte Streuung der Schülerleistungen einer Klasse der eigenen Schule als Ausgangspunkt, um über Bedarf und möglicherweise zu erweiternde Differenzierungskonzepte im eigenen Mathematikunterricht auf derselben oder einer benachbarten Klassenstufe nachzudenken, dann kann dies auch auf ein tragfähiges Zusammenspiel von Diagnose und Förderung abzielen.

Förderung ist im Unterricht immer fokussiert auf meist sehr spezifische und eng umgrenzte Lernziele, z. B. das „Darstellen statistischer Daten mit Diagrammen“, oder auf das „Argumentieren mit Verhältnissen“ oder auf eine Vernetzung mehrerer solcher Lernziele in einem Anwendungskontext (vgl. Abbildung 1). Differenzierungsmaßnahmen hingegen können eng fokussiert auf spezifische Lernziele sein, z. B. ein Vergleich multipler Lösungswege mittels Prompts oder im Klassengespräch mit hierauf abgestimmten gestuften und für eine vorangehende Arbeitsphase bereitgestellten Hilfestellungen. Differenzierungsmaßnahmen können jedoch auch breitere und wiederholte Verwendungsmöglichkeiten mit sich bringen, z. B. die Nutzung von Materialien wie Placemats sowie von Sozialformen wie Gruppenpuzzles, um verschiedene Lösungsideen oder Lösungswege der Schüler/innen in wiederholten Gelegenheiten im Schuljahr zu sammeln und systematisch in Kleingruppen vergleichen zu lassen (Barzel, Büchter & Leuders, 2018).

Hierbei wird deutlich, dass für die konkrete Ausgestaltung solcher Handlungsableitungen zur Differenzierung weiteres Wissen notwendig ist, sowohl zum Lernen spezifischer Inhalte, und ebenso zur Unterrichtsmethodik. Sofern dieses Wissen bei Lehrkräften nicht vorausgesetzt werden kann, bedarf es systematischer Begleitangebote, damit Bildungsstandardrückmeldungen wirksam werden können.

Die konkrete Umsetzung der ausgewählten Maßnahmen im Unterricht führt dann zu einer Vernetzung dieses neuen oder bestehenden Wissens mit weiteren Erfahrungen aus dem Unterricht in Bezug auf ganz konkrete Lerngruppen und zu ganz bestimmten Lernständen, z. B. zu Beginn oder gegen Ende eines Schuljahrs oder bei der Erkundung, Systematisierung, beim Üben oder einer Wiederholung. Solche Erfahrungen beinhalten oftmals bereits eine Überarbeitung der Maßnahmen, die ergänzend auch systematischer unterstützt werden kann, z. B. im Austausch eines Fachkollegiums oder in der Austauschphase längerfristiger Weiterbildungen und mit inhaltlicher Anreicherung durch Kolleginnen und Kollegen, Coaches oder Fachberater/innen.

4.2 Entwicklung von Lernaufgaben aus Testaufgaben

Die Bildungsstandardüberprüfungen melden Lernstände nicht nur global für das Fach Mathematik zurück, sondern auch in den inhaltlichen Teilkompetenzen und Handlungsbereichen. Damit geht der Anspruch einher, dass sich jede Testaufgabe möglichst eindeutig genau einem Inhalts- und einem Handlungsbereich zuordnen lässt.

Hinzu kommt der Anspruch, dass die Aufgabenstellung klar genug formuliert ist, um ohne Rückfragen beantwortet werden zu können. Aufgabenschwierigkeiten müssen angemessen sein und Aufgaben dürfen keine Teilpopulationen systematisch bevorteilen oder benachteiligen (Geschlechter, Bildungsregionen, sprachlicher Hintergrund etc.).

Zudem muss die Lösung einer Testaufgabe eindeutig sein: Ob eine Antwort korrekt ist, soll unabhängig von der beurteilenden Person entschieden werden können. Realisiert wird dies

u. a. mit einer Beschränkung auf spezifische Antwortformate, die als Antwort z. B. lediglich die Ergebniszahl oder eine Auswahl aus vorgegebenen Antwortmöglichkeiten (Single- bzw. Multiple-Choice-Format) vorsehen. Die Beurteilung bzw. Codierung von Rechenwegen oder Bewertung von Teillösungen wäre sehr aufwändig und benötigt neben einem elaborierten Codiermanual auch eine umfangreiche Schulung der beurteilenden Personen, weswegen entsprechende Antwortformate nur sehr selten in großflächigen Lernstandserhebungen Verwendung finden. Für eine vollständigere Berücksichtigung weiterer Kompetenzaspekte bei der Leistungsbeurteilung und zur Generierung diagnostischer Informationen für die nachfolgende Förderung werden daher für den Mathematikunterricht vor allem diagnostische Interviews, Fehleranalysen und Standortbestimmungen empfohlen. Standardisierten Instrumenten zur Leistungsfassung kommt für solche komplexen pädagogischen Anliegen im Unterrichtsalltag lediglich eine nachgeordnete Bedeutung zu (Moser Opitiz & Nührenbörger, 2015).

Lernaufgaben hingegen sollen Lernhandlungen mit Blick auf die erforderlichen kognitiven Aktivitäten zum Kompetenzerwerb initiieren und steuern (Bruder, 2016; George et al., 2016; Leuders, 2006): sie stellen das wichtigste Werkzeug dar, das Mathematiklehrkräften für ihre Unterrichtsplanung zur Verfügung steht. Mögliche Lernhandlungen beim Bearbeiten von Lernaufgaben sollen in ihrer Gesamtheit das gesamte Spektrum der in den Bildungsstandards beschriebenen Kompetenzen im Mathematikunterricht abdecken. Verbunden damit sind immer Überlegungen, wie das Arbeiten mit Aufgaben im Unterricht inszeniert und begleitet wird (vgl. Abbildung 7).

Lernaufgaben im Unterricht verlangen im Gegensatz zu Testaufgaben in einer Bildungsstandardüberprüfung keine eindeutige Kompetenzzuordnung, auch wenn das Bewusstsein einer Lehrkraft über das Potenzial einer Aufgabe zur Förderung spezifischer oder verschiedener Lernziele selbstverständlich wünschenswert ist. Nur so ist zu erwarten, dass das bestehende Potenzial von Aufgaben in der Unterrichtsgestaltung auch ausgeschöpft werden kann (Leuders, 2006). Mit Blick auf den in Abschnitt 3.1 beschriebenen Differenzierungsbedarf im Mathematikunterricht wird auch von Aufgaben oftmals gewünscht, dass sie verschiedenartige Zugänge, Entdeckungen und Bearbeitungen auf unterschiedlichen Leistungsniveaus ermöglichen (Neubrand, 2006). Dies ist z. B. bei vielen offenen oder selbstdifferenzierenden Aufgaben, Wahlaufgaben oder Blütenaufgaben der Fall (Bruder, 2016; Schupp, 2006).

Test- und Lernaufgaben müssen demnach teils sehr unterschiedliche Kriterien erfüllen, die in Tabelle 2 im Überblick zusammengefasst werden (vgl. George et al., 2016).

Auch wenn die Zuordnung der in Abbildung 1 (Abschnitt 1.3.1) dargestellten Testaufgabe zu den Kompetenzbereichen angesichts verschiedener denkbarer Lösungswege nicht ganz eindeutig ist (vgl. dazu Kapitel 3 in diesem Sammelband), erfüllt sie die erforderlichen Eigenschaften einer Testaufgabe. Mit etwas Kreativität bei der Unterrichtsgestaltung könnte sie jedoch auch als Lernaufgabe Verwendung finden, gerade weil unterschiedliche Lösungswege denkbar und für Vergleiche wünschenswert sind: Lösungswege können stärker prozedural

Tabelle 2: Kriterien von Testitems und Lernaufgaben (vgl. George et al., 2016)

	Testitems	Lernaufgaben
Einsatz	kompetenzbezogenes Testen	kompetenzbezogenes Unterrichten
Merkmale	in eine Atmosphäre des Prüfens eingebettet	in eine Atmosphäre des Lernens eingebettet
	dienen dem Nachweis von Kompetenzen	dienen dem langfristigen Erwerb von Kompetenzen
	messen abgrenzbare Kompetenzen einer definierten Schwierigkeit	bestehen meist aus Kompetenzbündeln und gestuften Schwierigkeitsgraden
	Fehler sind nachteilig bzw. unerwünscht	nutzen Fehler zum Lernen
	Klarheit in Aufgabenstellung über eine gewünschte Lösung	können variable Lösungswege und vielfältige Lernprozesse anbieten
	die gleiche Lösung ist unabhängig von Bewertungsinstanzen immer richtig	können variable Unterrichtssituationen hervorrufen
	erfüllen psychometrische und inhaltliche Gütekriterien	sind an den Lernstand einer Lerngruppe, Klasse oder an die Lernende/ den Lernenden angepasst
Mögliche Ziele	Rückmeldung für die Schul- und Unterrichtsentwicklung	Feedback für den Lernprozess
	Rechenschaftslegung	Förderung bestimmter Kompetenzen
		Erweiterung bzw. Ausbau von Kompetenzen
Prüfung der Gütekriterien	Lehrer/innen, Fachdidaktiker/innen, Pädagoginnen/Pädagogen, Psychologinnen/Psychologen, Erziehungswissenschaftler/innen	Lehrer/innen, Fachdidaktiker/innen, Pädagoginnen/Pädagogen ...

geprägt sein, z. B., indem die Häufigkeitsangaben für die Antwortkategorien rechnerisch in Öffnungswinkel von Kreissegmenten übersetzt werden. Ebenso können die Verhältnisse zwischen den Häufigkeitsangaben auch rein in der Vorstellung mit den sich aus den Antwortmöglichkeiten ergebenden Verhältnissen zwischen den Kreissegmenten bzw. Kreisbögen verglichen werden. Beiden Lösungswegen liegen unterschiedliche Konzepte und Vorstellungen zu Anteilen zugrunde, die vermutlich für unterschiedliche Schüler/innen mehr oder weniger anschlussfähig sind oder die im Unterricht gemeinsam als Lernziel anvisiert oder gar für eine Vernetzung von Konzepten und Vorstellungen genutzt werden können.

Derartige Überlegungen zur Inszenierung von Aufgaben im Unterricht verdeutlichen, wie bedeutsam bei Lehrkräften ein vertieftes Verständnis von Lernzielen, Kompetenzbereichen, ihrem Aufbau und ihrer gegenseitigen Vernetzung sowie von damit verbundenen Unterrichtsmethoden auch für eine Unterstützung der Idee der Kompetenzorientierung

durch beispielhafte Testaufgaben ist (vgl. Leuders, 2006). Dies spricht wiederum für Qualifizierungsangebote, welche die Nutzung von Bildungsstandartergebnissen begleiten und unterstützen, sofern entsprechende Kompetenzen bei Lehrkräften nicht bereits großflächig vorausgesetzt werden können (Leuders et al., 2019).

4.2.1 Rezeption und Reflexion – Testaufgaben finden und analysieren

Aufgaben aus Bildungsstandardüberprüfungen werden nur auszugsweise veröffentlicht. Grund ist, dass einige Aufgaben in wiederholten Durchführungen verwendet werden sollen, um derart längsschnittliche Leistungsentwicklungen über die Jahre hinweg im Schulsystem genauer erfassen und analysieren zu können. Dennoch finden sich zu allen Kompetenzbereichen Testaufgaben, welche dabei helfen, die in den Bildungsstandards beschriebenen Kompetenzen und teils auch die Kompetenzstufen zu veranschaulichen.

Entsprechende Aufgaben können als Impulsgeber für die Unterrichtsentwicklung genutzt werden. Dazu ist zunächst eine Analyse der in ihnen angesprochenen (Teil-)Kompetenzen notwendig (George et al., 2016; Leuders, 2006). In der Testaufgabe in Abbildung 1 sind dies der Handlungsbereich „Darstellen, Modellbilden (H1)“ und der Inhaltsbereich „Statistische Darstellung und Kenngrößen (I4)“. Mit Blick auf alternative Lösungswege können weitere Inhaltsbereiche bedeutsam werden, z. B. „Zahlen und Maße (I1)“ und „Interpretieren (H3)“. Derartige Zuweisungen von Lernzielen in Form von Kompetenzbereichen zu Aufgaben können u. a. dazu dienen, mit den Formulierungen der Bildungsstandards besser vertraut zu werden und nachfolgend als Lehrkraft zu überlegen, wo und wie entsprechende Kompetenzbereiche z. B. in den verwendeten Schulbüchern berücksichtigt sind.

Bei der Diskussion von möglicherweise eindeutigen oder auch unterschiedlichen Lösungswegen bieten sich Analysen der Teil-Kompetenzen an, welche Schüler/innen befähigen, eine solche Aufgabe zu lösen. In unserem Aufgabenbeispiel sind das u. a. ein Verständnis von Anteilen, sicherlich von Tabellen und Kreisdiagrammen, eventuell von Verhältnissen und möglicherweise auch von Rechenprozeduren zur Bestimmung von Kreissegmenten, die einen bestimmten Anteil repräsentieren. In solchen Analysen lassen sich weitergehend erste Vermutungen aufstellen, worauf falsche Antworten zurückzuführen sein könnten. Spätestens bei diesem Schritt liegen Überlegungen nahe, wie man ergänzend fokussierte empirische Hinweise gewinnen könnte, worauf fehlerhafte Antworten zurückzuführen sein könnten (als „vertiefte Diagnose“) und welche Lerngelegenheiten Schülerinnen und Schülern helfen könnten, um entsprechende Teil-Kompetenzen zu entwickeln, zu wiederholen oder in komplexeren Aufgabenstellungen zu vernetzen (als „fokussierte Förderung“, vgl. Prediger, 2014).

Bereits die theoretische Analyse von Aufgaben ermöglicht somit eine vernetzte Erörterung von Aufgabenmerkmalen, Lernzielen und Unterrichtsgestaltung, sei es, um bestehende Erfahrungen und Kenntnisse zu vertiefen oder um Bedarf für nachfolgende Qualifizierungsmaßnahmen für sich allein, im kollegialen Austausch oder in Weiterbildungsveranstaltungen zu identifizieren.

4.2.2 Aktion und Evaluation – Lernaufgaben finden oder entwickeln, im Unterricht einsetzen und überarbeiten

Ergibt sich aus der Aufgabenanalyse Bedarf für Lernaufgaben, die zu den in einer Testaufgabe identifizierten Kompetenzen passen, so finden sich mittlerweile vielfältige Publikationen zu beinahe allen Kompetenzbereichen, auch ergänzend zum verwendeten Schulbuch. Es kann dennoch spannend und reizvoll sein, allein oder im Austausch mit Kolleginnen und Kollegen selbstständig entsprechende Aufgaben zu entwickeln. Als Lehrkraft ermöglichen es entsprechende Erfahrungen gelegentlich, auch spontan im Unterricht, passend zur Unterrichtssituation Aufgaben abzuändern oder zu erweitern. Ein solcher kreativer Akt kann begeisterten Mathematiklehrkräften nicht selten per se Freude bereiten, insbesondere wenn diese die eigenen Handlungsspielräume im Unterricht erweitern oder gar in kollegialen Austausch integriert werden.

Anregende Ideen zur Aufgabenvariation im Fach Mathematik finden sich u. a. bei Schupp (2002, 2006), Leuders (2006, 2010) oder Behrens (2018). Speziell zur Entwicklung von Lernaufgaben aus Testaufgaben besprechen und illustrieren George, Süss-Stepancik, Illetschko und Wiesner (2016) drei Techniken:

- a) Mittels *Situationsorientierung* oder *Kontextualisierung* von Aufgaben lassen sich u. a. Alltagserfahrungen und Vorstellungen zum Kompetenzaufbau nutzen, wenn in Anlehnung an den Modellierungskreislauf von Lernenden eigene Situationsmodelle aufgestellt werden, diese mathematisiert werden, Lösungen erarbeitet und die Ergebnisse wiederum auf die Situation bezogen, interpretiert und validiert werden (Blum & Leiß, 2005; Leuders & Leiß, 2006). Bezogen auf die Testaufgabe in Abbildung 1 kann eine erweiterte Kontextualisierung darin bestehen, dass die Schüler/innen einer Klasse eigenständig Daten zu selbstgewählten Fragestellungen erheben, aufbereiten sowie ihre Interpretationen präsentieren und vergleichen.
- b) Im Sinne einer *Explorationsorientierung* werden Aufgaben geöffnet, wodurch sich Möglichkeiten zum Entdecken und intuitiven Denken bieten. Dies beinhaltet oftmals eine bewusste Vernetzung von prozessbezogenen mit inhaltsbezogenen Kompetenzen und unterstützt damit den langfristigen Kompetenzaufbau (Bruder, 2016). Die Testaufgabe in Abbildung 1 erweiternd, ließe sich fragen, woran man in einem solchen Kreisdiagramm eine leistungsstärkere oder eine leistungsschwächere Klasse erkennen könne. Dies bietet einen Anlass, Mittelwerte und Streuungen zu beschreiben, voneinander abzugrenzen oder Effekte auf diese beiden statistischen Maße bei verschiedenen Veränderungen der gegebenen Verteilung zu thematisieren.
- c) Die *Reflexionsorientierung* zielt auf eine Erweiterung von Aufgaben durch explizite Aufforderungen zur Reflexion des eigenen Lernens ab. Dem Vergleich von multiplen Lösungswegen (Neubrand, 2006) oder der Rückschau auf zurückliegende Problemlöseprozesse (Polya, 1988; Schoenfeld, 1992) wird nicht nur eine verstärkte Vernetzung von prozessbezogenen mit inhaltsbezogenen Kompetenzen zugesprochen, sondern auch eine zielgerichtete Förderung metakognitiver Fähigkeiten von Schülerinnen

und Schülern. Beim Vergleich von Lösungswegen zur Testaufgabe in Abbildung 1 kann rückblickend und zusammenfassend herausgearbeitet werden, auf welche mathematischen Konzepte zurückgegriffen wurde. Diese lassen sich derart nochmals erklären und vernetzen. Neben Mittelwerten und Streuungen können z. B. Verhältnisse zwischen Zahlen oder zwischen Anteilen von Winkeln thematisiert werden und mit weiteren bereits behandelten Unterrichtsinhalten zur Statistik, Bruchrechnung oder zu funktionalen Zusammenhängen in Verbindung gebracht werden. Auch welcher Schüler oder welche Schülerin welche Arten von Darstellungen, z. B. grafisch, tabellarisch oder mittels Rechnung, verständlicher findet, um die herausgearbeiteten Zusammenhänge zu verdeutlichen, kann thematisiert werden.

Zusammenfassend lässt sich feststellen, dass eigene Aufgabenvariationen und Aufgabenentwicklungen durch Lehrkräfte einen vielfältigen Beitrag zur eigenen Professionalisierung, zum Austausch in Fachkollegien und für eine flexible und adaptive Unterrichtsgestaltung leisten können.

Vermutlich sind allgemeinere Techniken, wie im Abschnitt zuvor beschrieben, oder z. B. zum intelligenten Üben (Leuders, 2010) schneller erlernbar und universell einsetzbar, indem sie generelle Merkmale von Aufgaben, Lernzielen oder Lerngelegenheiten wie Öffnung, Anwendungsorientierung oder Vernetzung zwischen inhaltsbezogenen und prozessbezogenen Kompetenzen in den Blick nehmen. Ein dezidiertes, sehr lerngegenstandspezifisches Wissen zu Lernverläufen, Lernvoraussetzungen, Lernhürden, Hilfen und Veranschaulichungen ist hingegen notwendig, um zielgerichtete Diagnose- und Förderaufgaben zu enger fokussierten Teil-Kompetenzen zu entwickeln. Daher wäre hierfür eher der Rückgriff auf bestehende und teils auch frei verfügbare Diagnose- und Fördermaterialien zu empfehlen (z. B. <https://mathe-sicher-koennen.dzlm.de>).

5 Ausblick

Die Bildungsstandardüberprüfungen fanden in einem zyklischen Rhythmus, zuletzt im Fach Englisch auf der 8. Schulstufe im Jahr 2019, statt. Die Kompetenzen von Schülerinnen und Schülern auf der 8. Schulstufe in Mathematik wurden zuletzt im Jahr 2017 überprüft. Die Ergebnisrückmeldungen und -berichte der unterschiedlichen Zielgruppen wurden, mitunter aufgrund der immensen Datenmengen und damit einhergehenden Verarbeitungsprozesse, jeweils zeitlich um ein Jahr verzögert zur Verfügung gestellt.

Die Bildungsstandards ermöglichten durch ihre breite Konstruktdeckung innerhalb der verschiedenen überprüften Fächer einen Fokus auf die Evaluation der Schülerleistungen insgesamt – ein wichtiger Baustein für die Schul- und Unterrichtsentwicklung auf übergeordneter Ebene sowie für das Qualitäts- und Bildungsmonitoring (vgl. Abschnitt 1.3.2 Evaluationsfunktion).

Betrachtet man allerdings die Ergebnisnutzung auf Individualebene, müssen verschiedene potenzielle Einschränkungen berücksichtigt werden: Wie bereits erwähnt wurde, sollte die Setzung von Bildungsstandards sowie deren Überprüfung zu verbesserten Schülerkompetenzen sowie zu einer nachhaltigen Leistungssteigerung des Schulsystems führen. Dies könnte einerseits durch eine konsequente Fokussierung des Unterrichts auf die Kompetenzentwicklung der Schülerinnen und Schüler realisiert werden, wie auch andererseits durch die Ableitung differenzierter Maßnahmen für spezielle Zielgruppen, z. B. für leistungsschwächere Schüler/innen (vgl. Altrichter & Kanape-Willingshofer, 2012; Eder, 2009; Stojanov, 2007).

Da die Bildungsstandardüberprüfungen jedoch erst am Ende der 4. und 8. Schulstufe stattfanden und die Ergebnisrückmeldungen nur mit einer zeitlichen Verzögerung übermittelt werden konnten, befanden sich bereits viele der getesteten Schüler/innen nicht mehr an der Schule, an der die Überprüfung stattgefunden hatte. Zudem erhielten die Lehrpersonen im Rahmen der Bildungsstandardüberprüfungen keinen Einblick in einzelne Schülerergebnisse, was wiederum die Ableitung differenzierter Maßnahmen nur auf Klassen-, nicht jedoch auf Individualebene ermöglichte.

Im Rahmen der neuen individuellen Kompetenzmessung PLUS (kurz iKM^{PLUS}) sollen die verschiedenen Bedürfnisse der unterschiedlichen Ziel- und Nutzungsebenen vereint werden (Bundesministerium für Bildung, Wissenschaft und Forschung, 2020):

- Bildungsmonitoring auf nationaler Ebene
- Schulentwicklung sowie Schulevaluation auf Schulebene
- Unterrichtsplanung und Unterrichtsevaluation auf Klassenebene
- Gezielte Fördermaßnahmen als Individualdiagnostik auf Schülerebene

Die iKM^{PLUS} wird im Rahmen eines stufenweisen Zeitplans mit einer Übergangszeit im Kalenderjahr 2021 österreichweit eingeführt. Sie vereint und entwickelt dabei wichtige Funktionen der bereits erwähnten Instrumente der IKM und BIST-Ü weiter.

So sollen jährliche Ergebnisrückmeldungen zeitnah zur Testdurchführung zur Verfügung gestellt werden und dadurch noch im selben Schuljahr unmittelbar förder- und unterrichtswirksam für Schüler/innen, Lehrpersonen und Schulleitungen werden. Im Rahmen dreijährlicher Berichterstattungen soll zudem auch die Zielebene des System- und Bildungsmonitorings weiterhin mit belastbaren Daten bedient werden.

Die Testungen finden dabei einerseits im Rahmen jährlicher und verpflichtender Basismodule statt; auf der 3. und 4. Schulstufe sind das *Deutsch (Lesen)* und *Mathematik* sowie auf der 7. und 8. Schulstufe *Deutsch (Lesen)*, *Mathematik* und *English (Receptive Skills)*. Andererseits können Lehrpersonen zusätzlich ein breites Angebot an freiwillig durchzuführenden Modulen nutzen, welche die Sichtweise auf einen Kompetenzbereich vervollständigen können.

Begleitmaterialien und umfassende Fortbildungsangebote für Lehrpersonen, Schulleitungen und Schulqualitätsmanager/innen sollen dabei auch weiterhin den unterschiedlichen Zielgruppen in den verschiedenen Phasen der Vorbereitung, der Durchführung, der Rückmeldung und Reflexion sowie der Förderung Unterstützung anbieten. Auch für die Nutzung der iKM^{PLUS} ist zu erwarten, dass dem fachdidaktischen und methodischen Wissen der Lehrpersonen zum Diagnosegegenstand eine zentrale Rolle zukommt (Leuders et al., 2019) und daher weiterhin kontinuierliche Professionalisierungsangebote notwendig sind, um langfristig das Lernen der Schülerinnen und Schüler optimal zu unterstützen.

Literatur

- Altrichter, H. (2010). Schul- und Unterrichtsentwicklung durch Datenrückmeldung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (Educational governance, Band 7, S. 219–254). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-92245-4
- Altrichter, H. & Kanape-Willingshofer, A. (2012). Bildungsstandards und externe Überprüfung von Schülerkompetenzen: Mögliche Beiträge externer Messungen zur Erreichung der Qualitätsziele der Schule. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012, Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 355–394). Graz: Leykam. doi:10.17888/nbb2012-2
- Asbrand, B., Heller, N. & Zeitler, S. (2012). Die Arbeit mit Bildungsstandards in Fachkonferenzen: Ergebnisse aus der Evaluation des KMK-Projektes for. mat. *DDS-Die Deutsche Schule*, 104 (1), 31–43. Verfügbar unter <https://www.waxmann.com/ausgabeAUG100164>
- Barzel, B., Büchter, A. & Leuders, T. (2018). *Mathematik Methodik. Handbuch für die Sekundarstufe I und II* (9. Auflage). Berlin: Cornelsen.
- Behrens, R. (2018). *Formulieren und Variieren mathematischer Fragestellungen mittels digitaler Werkzeuge* (Studien zur theoretischen und empirischen Forschung in der Mathematikdidaktik). Wiesbaden: Springer Spektrum. doi:10.1007/978-3-658-24489-7
- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A Preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research & Perspective*, 8 (2–3), 70–91. doi:10.1080/15366367.2010.508686
- Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (Hrsg.). (2006). *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen*. Berlin: Cornelsen Scriptor. Verfügbar unter: <https://edoc.hu-berlin.de/bitstream/handle/18452/3776/4.pdf?sequence=1>

- Blum, W. & Leiß, D. (2005). Modellieren im Unterricht mit der „Tanken“-Aufgabe. *Mathematik lehren*, 128/2005, 18–21. Verfügbar unter: <https://www.friedrich-verlag.de/shop/pisa-neue-ergebnisse-und-anregungen-58128>
- Bruder, R. (2016). Vielseitig mit Aufgaben arbeiten. Mathematische Kompetenzen nachhaltig entwickeln und sichern. In R. Bruder, T. Leuders & A. Büchter (Hrsg.), *Mathematikunterricht entwickeln. Bausteine für kompetenzorientiertes Unterrichten* (4. Auflage, S. 18–52). Berlin: Cornelsen.
- Bruder, R., Leuders, T. & Büchter, A. (Hrsg.). (2016). *Mathematikunterricht entwickeln. Bausteine für kompetenzorientiertes Unterrichten* (4. Auflage). Berlin: Cornelsen.
- Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (Hrsg.). (2011). *Praxishandbuch für „Mathematik“ 8. Schulstufe* (2., überarbeitete Auflage). Graz: Leykam. Verfügbar unter: https://www.iqs.gv.at/_Resources/Persistent/ff034f1cd7d23f0bb271db765aea60c7f3542e78/bist_m_sek1_praxishandbuch_mathematik_8_2012-04-16.pdf
- Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (Hrsg.). (2012). *Praxishandbuch für „Mathematik“ 8. Schulstufe. Band 2*. Graz: Leykam.
- Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens. (2017a). *Freigegebene Items. Standardüberprüfung M8 – 2017*. Verfügbar unter: <https://www.iqs.gv.at/downloads/archiv-des-bifie/bildungsstandardueberpruefungen/freigegebene-items>
- Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens. (2017b). *Rückmeldung an die Lehrer/innen. Standardüberprüfung M8 – 2017. Rückmeldung der Ergebnisse Ihrer Unterrichtsgruppe*. Verfügbar unter: <https://www.iqs.gv.at/downloads/archiv-des-bifie/bildungsstandardueberpruefungen/musterrueckmeldungen>
- Bundesministerium für Bildung, Wissenschaft und Forschung. (2020). *Das Pädagogik-Paket. Zeitgemäß. Transparent. Fair* (2., aktualisierte Auflage). Wien. Verfügbar unter: <https://www.bmbwf.gv.at/dam/jcr:326dfad6-a8b9-4e56-9d67-b4bdcc343bb1/pb.pdf>
- Demski, D. (2017). *Evidenzbasierte Schulentwicklung: Empirische Analyse eines Steuerungsparadigmas*. Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-18078-2
- Diemer, T. (2013). *Innerschulische Wirklichkeiten neuer Steuerung. Zur Nutzung zentraler Lernstandserhebungen*. Wiesbaden: Springer. doi:10.1007/978-3-658-01433-9
- Eder, F. (2009). Folgerungen für Lehrerbildung und Schulentwicklung. In C. Schreiner & U. Schwantner (Hrsg.), *PISA 2006: Österreichischer Expertenbericht zum Naturwissenschafts-Schwerpunkt* (S. 433–438.). Graz: Leykam. Verfügbar unter: https://www.iqs.gv.at/_Resources/Persistent/be0327f57e68be66fad67e077a79bdc82fd4dd79/PISA2006_NEB_web.pdf
- Fend, H. (2008). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen* (2., durchgesehene Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-91788-7

- Freunberger, R. (2013). *Standard-Setting Mathematik 8. Schulstufe. Technischer Bericht*. Salzburg: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung. Verfügbar unter: <https://www.iqs.gv.at/downloads/archiv-des-bifie/bildungsstandardueberpruefungen/technische-dokumentation>
- Freunberger, R. & Yanagida, T. (2012). Competency Diagnostics in Austria: The Standard-Setting Process. *Psychologie in Österreich*, 5, 396–403.
- George, A. C., Robitzsch, A. & Schreiner, C. (2019). Eine Diskussionsgrundlage zur Weiterentwicklung von Rückmeldungen aus standardisierten Kompetenzmessungen am Beispiel Mathematik. In A. C. George, C. Schreiner, C. Wiesner, M. Pointinger & K. Pacher (Hrsg.), *Fünf Jahre flächendeckende Bildungsstandardüberprüfungen in Österreich – Vertiefende Analysen zu Zyklus 2012 bis 2016* (S. 225–238). Münster: Waxman. Verfügbar unter: <https://www.waxmann.com/?eID=texte&pdf=3936Volltext.pdf&typ=zusatztext#page=225>
- George, A. C., Süß-Stepancik, E., Illetschko, M. & Wiesner, C. (2016). Entwicklung wirkungsvoller Lernaufgaben für den Unterricht aus Testitems der Bildungsstandardüberprüfung. *Transfer: Forschung Schule*, 2 (2), 67–87.
- Greefrath, G. & Maaß, K. (2020). *Modellierungskompetenzen – Diagnose und Bewertung* (Realitätsbezüge im Mathematikunterricht, 1st edition). Berlin, Heidelberg: Springer Spektrum. doi:10.1007/978-3-662-60815-9
- Groot-Wilken, B., Isaac, K. & Schräpler, J.-P. (Hrsg.). (2016). *Sozialindices für Schulen. Hintergründe, Methoden und Anwendung* (Beiträge zur Schulentwicklung). Münster: Waxmann. Verfügbar unter: http://www.content-select.com/index.php?id=bib_view&ean=9783830984863
- Groß Ophoff, J. (Hrsg.). (2013). *Lernstandserhebungen: Reflexion und Nutzung* (Pädagogische Psychologie und Entwicklungspsychologie, Band 85). Münster: Waxmann.
- Groß Ophoff, J., Hosenfeld, I. & Koch, U. (2007). Formen der Ergebnisrezeption und damit verbundene Schul- und Unterrichtsentwicklung. *Empirische Pädagogik*, 21 (4), 411–427.
- Groß Ophoff, J., Koch, U., Hosenfeld, I. & Helmke, A. (2006). Ergebnisrückmeldungen und ihre Rezeption im Projekt VERA. In H. Kuper & J. Schneewind (Hrsg.), *Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem* (S. 19–40). Münster: Waxmann.
- Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77 (1), 81–112. doi:10.3102/003465430298487
- Heckhausen, H. (1974). *Leistung und Chancengleichheit* (Motivationsforschung, Band 2). Göttingen: Hogrefe.
- Helmke, A. & Hosenfeld, I. (2005). Standardbezogene Unterrichtsevaluation. In G. Brägger, B. Bucher & N. Landwehr (Hrsg.), *Schlüsselfragen zur externen Schulevaluation* (Pädagogik, S. 127–151). Bern: hep.
- Hill, H. C. (2001). Policy is not enough: Language and the interpretation of state standards. *American Educational Research Journal*, 38 (2), 289–318.
- Institut des Bundes für Qualitätssicherung im österreichischen Schulwesen (Hrsg.). (2017). *Freigegebene Items der Standardüberprüfungen. Mathematik 8. Schulstufe*. Verfügbar unter: <https://iqs.gv.at/downloads/archiv-des-bifie/bildungsstandardueberpruefungen/freigegebene-items>

- Institut des Bundes für Qualitätssicherung im österreichischen Schulwesen. (2021). *Allgemeine Informationen zur Informellen Kompetenzmessung (IKM)*. Verfügbar unter: <https://www.iqs.gv.at/themen/nationales-monitoring/informelle-kompetenzmessung-ikm/allgemeine-informationen-zur-ikm>
- Itzlinger-Bruneorth, U. & Kuhn, J.-T. (2015). *Die Itemerstellung für Bildungsstandardüberprüfungen. Technische Dokumentation – BIST-Ü*. Salzburg: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung. Verfügbar unter: https://www.iqs.gv.at/_Resources/Persistent/1fdd944a0a654a413b917a3528612362712e4b5e/TD_Itemerstellung_BiSt-UE.pdf
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (Bundesministerium für Bildung und Forschung [BMBF], Hrsg.). (2003). *Expertise: Zur Entwicklung nationaler Bildungsstandards* (Bildungsreform, Band 1). Berlin: BMBF. Verfügbar unter: https://www.pedocs.de/volltexte/2020/20901/pdf/Klieme_et_al_2003_Zur_Entwicklung_Nationaler_Bildungsstandards_BMBF_A.pdf
- Koch, U. (2011). *Verstehen Lehrkräfte Rückmeldungen aus Vergleichsarbeiten? Datenkompetenz von Lehrkräften und die Nutzung von Ergebnissrückmeldungen aus Vergleichsarbeiten* (Empirische Erziehungswissenschaft, Band 31). Münster: Waxmann.
- Köller, O., Baumert, J. & Bos, W. (2001). TIMSS – Third International Mathematics and Science Study. Dritte internationale Mathematik- und Naturwissenschaftsstudie. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (2., unveränderte Auflage, S. 269–284). Weinheim: Beltz.
- Kowalk, S. (in Vorbereitung). *Datengestützte und theoriegeleitete Professionalisierung – Entwicklung und Evaluation einer Lehrer- und Lehrerinnenfortbildung im Zusammenhang mit einer zentralen Eingangsdiagnose in Klasse 5 (Lernstand 5)*. Dissertation in Vorbereitung, Institut für Mathematische Bildung, Freiburg.
- Kowalk, S., Leuders, T. & Schulz, A. (2018). *Förderung diagnostischer Kompetenzen im Zusammenhang mit zentralen Diagnostetests (Lernstand 5)*. doi:10.17877/DE290R-19476
- Krainer, K., Hanfstingl, B., Hellmuth, T., Hopf, M., Lembens, A., Neuweg, G. H. et al. (2012). Die Fachdidaktiken und ihr Beitrag zur Qualitätsentwicklung des Unterrichts. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012, Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 143–187). doi:10.17888/nbb2012-2
- Kröpfl, B. & Schneider, E. (2012). *Standards Mathematik unter der Lupe*. München: Profil-Verlag.
- Kühle, B. (2010). *Zentrale Lernstandserhebungen – ergebnisorientierte Unterrichtsentwicklung? Schulische Strategien beim Umgang mit Ergebnissen aus den Schulrückmeldungen im Kontext der ersten Lernstandserhebungen 2004/2005 in Nordrhein-Westfalen*. Berlin: Köster.
- Kühle, B. & Peek, R. (2007). Lernstandserhebungen in Nordrhein-Westfalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnissrückmeldungen in Schulen. *Empirische Pädagogik*, 21 (4), 428–447.
- Leuders, T. (2006). Kompetenzorientierte Aufgaben im Unterricht. In W. Blum, C. Drücke-Noe, R. Hartung & O. Köller (Hrsg.), *Bildungsstandards Mathematik: konkret. Sekundar-*

- stufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen. (S. 81–95). Berlin: Cornelsen Scriptor.
- Leuders, T. (2010). Intelligent üben und Mathematik erleben. In T. Leuders, L. Hefendehl-Hebeker & H.-G. Weigand (Hrsg.), *Mathemagische Momente* (S. 132–145). Berlin: Cornelsen.
- Leuders, T. & Leiß, D. (2006). Realitätsbezüge. In W. Blum, C. Drücke-Noe, R. Hartung & O. Köller (Hrsg.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen* (S. 194–206). Berlin: Cornelsen Scriptor.
- Leuders, T. & Prediger, S. (2017). Flexibel differenzieren erfordert fachdidaktische Kategorien. In J. Leuders, T. Leuders, S. Prediger, S. Ruwisch (Hrsg.), *Mit Heterogenität im Mathematikunterricht umgehen lernen* (S. 3–16). Springer Spektrum, Wiesbaden. doi:10.1007/978-3-658-16903-9_1
- Leuders, T., Schulz, A. & Kowalk, S. (2019). Lernstandsdiagnosen – Wann ist externe Diagnoseunterstützung nützlich? In C. G. Buhren, G. Klein & S. Müller (Hrsg.), *Handbuch Evaluation in Schule und Unterricht* (Pädagogik, S. 166–183). Weinheim: Beltz.
- Leutner, D., Fleischer, J., Spoden, C. & Wirth, J. (2008). Landesweite Lernstandserhebungen zwischen Bildungsmonitoring und Individualdiagnostik. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik. Zeitschrift für Erziehungswissenschaft* (Sonderheft 8/2007, S. 149–167). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-90865-6_9
- Leutner, D., Klieme, E., Fleischer, J. & Kuper, H. (2013). Editorial: Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Erziehungswissenschaft*, 16 (Supplement 1), 1–4. doi:10.1007/s11618-013-0378-0
- Lipowsky, F. & Rzejak, D. (2017). Fortbildungen für Lehrkräfte wirksam gestalten – erfolgversprechende Wege und Konzepte aus Sicht der empirischen Bildungsforschung. *Bildung und Erziehung*, 70 (4), 379–400.
- Maag Merki, K. (2010). Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 145–169). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-92245-4_6
- Maier, U. (2008). Vergleichsarbeiten im Vergleich – Akzeptanz und wahrgenommener Nutzen standardbasierter Leistungsmessungen in Baden-Württemberg und Thüringen. *Zeitschrift für Erziehungswissenschaft*, 11 (3), 453–474. doi:10.1007/s11618-008-0036-0
- Maier, U. (2009). Testen und dann? Ergebnisse einer qualitativen Lehrerbefragung zur individualdiagnostischen Funktion von Vergleichsarbeiten. *Empirische Pädagogik*, 23 (2), 191–207.
- Maier, U., Ramsteck, C. & Frühwacht, A. (2013). Lehr-lerntheoretische Argumentationsmuster bei der Interpretation und Nutzung von Vergleichsarbeitsrückmeldungen durch Gymnasiallehrkräfte. *Evidenzbasierte Steuerung im Bildungssystem*, 12, 74–96.
- Mislevy, R., Steinberg, L. S. & Almond, R. G. (2003). Focus article. On the structure of educational assessments. *Measurement: Interdisciplinary Research & Perspective*, 1 (1), 3–62. doi:10.1207/S15366359MEA0101_02

- Moser Opitz, E. (2022). Diagnostisches und didaktisches Handeln verbinden: Entwicklung eines Prozessmodells auf der Grundlage von Erkenntnissen aus der pädagogischen Diagnostik und der Förderdiagnostik. *Journal für Mathematik-Didaktik*, 43 (1), 1–26.
- Moser Opitz, E. & Nührenbörger, M. (2015). Diagnostik und Leistungsbeurteilung. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 491–512). Heidelberg: Springer.
- Müller, S. (2014). *Data Rich, Information Poor? Chancen und Grenzen vergleichender Leistungsmessung an der Förderschule mit dem Förderschwerpunkt Lernen* (Empirische Erziehungswissenschaft). Münster: Waxmann.
- Mürwald-Scheifinger, E. (2012). Mathematikunterricht in heterogenen Lerngruppen. In Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (Hrsg.), *Praxishandbuch für „Mathematik“ 8. Schulstufe. Band 2* (S. 21–34). Graz: Leykam.
- Nachtigall, C. & Kröhne, U. (2006). Methodische Anforderungen an schulische Leistungsmessung – auf dem Weg zu fairen Vergleichen. In H. Kuper & J. Schneewind (Hrsg.), *Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem* (S. 59–74). Münster: Waxmann.
- National Council of Teachers of Mathematics [NCTM]. (2000). *Principles and standards for school mathematics*. Reston: NCTM.
- Netzwerk zur empiriegestützten Schulentwicklung (2008). Nutzung und Nutzen von Schülerrückmeldungen im Rahmen standardisierter Lernstandserhebungen/Vergleichsarbeiten. Zweites Positionspapier des EMSE-Netzwerkes – verabschiedet auf der 9. EMSE-Fachtagung am 16.–17. Dezember 2008 in Nürnberg. Verfügbar unter https://uni-bielefeld.de/fakultaeten/erziehungswissenschaft/weos/hps/emse-netzwerk/tagungsmaterial/EMSE_Positionsp2_Rueckmeldungen.pdf
- Neubrand, M. (2006). Multiple Lösungswege für Aufgaben: Bedeutung für Fach, Lernen, Unterricht und Leistungserfassung. In W. Blum, C. Drüke-Noe, R. Hartung & O. Köller (Hrsg.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen* (S. 162–174). Berlin: Cornelsen Scriptor.
- Neumann, K. (2013). Mit welchem Auflösungsgrad können Kompetenzen modelliert werden? In welcher Beziehung stehen Modelle zueinander, die Kompetenz in einer Domäne mit unterschiedlichem Auflösungsgrad beschreiben? *Zeitschrift für Erziehungswissenschaft*, 16 (Sonderheft 1), 35–39. doi:10.1007/s11618-013-0382-4
- Neureiter, H. C. (2012). Selbstdifferenzierende Aufgaben. In Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (Hrsg.), *Praxishandbuch für „Mathematik“ 8. Schulstufe. Band 2* (S. 47–62). Graz: Leykam.
- Oelkers, J. & Reusser, K. (Bundesministerium für Bildung und Forschung [BMBF], Hrsg.). (2008). *Expertise: Qualität entwickeln – Standards sichern – mit Differenz umgehen* (Bildungsforschung, Band 27). Berlin: BMBF. Verfügbar unter: http://edudoc.ch/record/86369/files/expertise_oelkers_reusser_d.pdf
- Pellegrino, J. W. & Chudowsky, N. (2003). FOCUS ARTICLE: The foundations of assessment. *Measurement: Interdisciplinary Research & Perspective*, 1 (2), 103–148. doi:10.1207/S15366359MEA0102_01

- Pellegrino, J. W., Chudowsky, N., Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: National Academy Press. Verfügbar unter: <https://www.nap.edu/read/10019/chapter/1>
- Polya, G. (1988). *How to solve it. A new aspect of mathematical method* (2. Auflage). Princeton, NJ: Princeton Univ. Press.
- Prediger, S. (2014). Nicht nur individuelle, sondern auch fokussierte Förderung – Fachdidaktische Ansprüche und Forschungs- und Entwicklungsnotwendigkeiten an ein Konzept. In *Beiträge zum Mathematikunterricht 2014: Band 1*. (S. 931–934). Münster: WTM-Verlag. Verfügbar unter: https://www.dms.uni-landau.de/roth/veroeffentlichungen/2014/roth_ames_beitraege_zum_mathematikunterricht_2014_band_1.pdf
- Ravitch, D. (1995). *National standards in American education. A citizen's guide*. Washington: The Brookings Institution. Retrieved from <https://eric.ed.gov/?id=ed400617>
- Schlee, J. (2002). Was kann Diagnostik für die pädagogische Praxis leisten? Zu den Ansprüchen der sogenannten Förderdiagnostik. In W. Mutzeck & H. Bach (Hrsg.), *Förderdiagnostik. Konzepte und Methoden* (3., überarbeitete Auflage, Dr. nach Typoskript, S. 181–193). Weinheim: Beltz.
- Schneewind, J. (2006). Rückmeldungen als Motivator für die Teilnahme an Schulleistungstudien? Die Rezeptionsstudie von BeLesen. In H. Kuper & J. Schneewind (Hrsg.), *Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem* (S. 107–126). Münster: Waxmann.
- Schneider, G. (2005). Der «Gemeinsame europäische Referenzrahmen für Sprachen» als Grundlage von Bildungsstandards für die Fremdsprachen. Methodologische Probleme der Entwicklung und Adaptierung von Kompetenzbeschreibungen. *Schweizerische Zeitschrift für Bildungswissenschaften*, 27 (1), 13–36. doi:10.25656/01:4114
- Schoenfeld, A. H. (1992). Learning to think mathematically. Problem solving, metacognition and sense-making in mathematics. In D. A. Grouws (Hrsg.), *Handbook of research on mathematics teaching and learning. A project of the National Council of Teachers of Mathematics* (S. 334–370). New York NY: Macmillan.
- Schreiner, C., Breit, S., Pointinger, M., Pacher, K., Neubacher, M. & Wiesner, C. (Hrsg.). (2018). *Standardüberprüfung 2017. Mathematik, 8. Schulstufe. Bundesergebnisbericht*. Salzburg: Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens. Verfügbar unter: https://iqs.gv.at/_Resources/Persistent/4a28609fd6414dbc257274688ffa37e44e4a3cf7/BiSt_UE_M8_2017_Bundesergebnisbericht.pdf
- Schreiner, C. & Wiesner, C. (2019). Die Überprüfung der Bildungsstandards in Österreich: der erste Zyklus als Meilenstein für die Schul- und Unterrichtsentwicklung – eine gelungene Innovation im österreichischen Schulsystem. In A. C. George, C. Schreiner, C. Wiesner, M. Pointinger & K. Pacher (Hrsg.), *Fünf Jahre flächendeckende Bildungsstandardüberprüfungen in Österreich. Vertiefende Analysen zum Zyklus 2012 bis 2016* (Kompetenzmessungen im österreichischen Schulsystem, S. 13–45). Münster: Waxmann.
- Schulz, A. (2010). *Ergebnisorientierung als Chance für den Mathematikunterricht? Innovationsprozesse qualitativ und quantitativ erfassen* (Münchner Beiträge zur Bildungsforschung, Band 17). München: Utz.

- Schulz, A., Leuders, T. & Rangel, U. (2017). Arithmetische Basiskompetenzen am Übergang zu Klasse 5 – eine empirie- und modellgestützte Diagnostik als Grundlage für spezifische Förderentscheidungen. In A. Fritz, S. Schmidt & G. Ricken (Hrsg.), *Handbuch Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie* (Pädagogik, 3. Auflage, S. 396–417). Weinheim: Beltz.
- Schulz, A., Leuders, T. & Rangel, U. (2020). The use of a diagnostic competence model about children's operation sense for criterion-referenced individual feedback in a large-scale formative assessment. *Journal of Psychoeducational Assessment*, 38 (4), 426–444. doi:10.1177/0734282918823590
- Schupp, H. (2002). *Thema mit Variationen oder Aufgabenvariation im Mathematikunterricht*. Hildesheim: Franzbecker.
- Schupp, H. (2006). Variation von Aufgaben. In W. Blum, C. Drücke-Noe, R. Hartung & O. Köller (Hrsg.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen* (S. 152–161). Berlin: Cornelsen Scriptor.
- Stojanov, K. (2007). Bildungsgerechtigkeit im Spannungsfeld zwischen Verteilungs-, Teilhabe- und Anerkennungsgerechtigkeit. In M. Wimmer, R. Reichenbach & L. Pongratz (Hrsg.), *Gerechtigkeit und Bildung* (S. 29–48). Paderborn: Schöningh.
- Stojanov, K. (2011). *Bildungsgerechtigkeit. Rekonstruktionen eines umkämpften Begriffs*. Wiesbaden: VS Verl. für Sozialwissenschaften. doi:10.1007/978-3-531-94011-3
- Sundermann, B. & Selzer, C. (2013). *Beurteilen und Fördern im Mathematikunterricht* (Lehrerbücherei Grundschule, 4., überarbeitete Neuauflage). Berlin: Cornelsen Scriptor.
- Van Ackeren, I. (Bundesministerium für Bildung und Forschung [BMBF], Hrsg.). (2003). *Nutzung großflächiger Tests für die Schulentwicklung: Exemplarische Analyse der Erfahrungen aus England, Frankreich und den Niederlanden* (Bildungsreform, Band 3). Berlin: BMBF. Verfügbar unter: <https://d-nb.info/971373566/34>
- Visscher, A. J. & Coe, R. (2003). School Performance Feedback Systems: Conceptualisation, Analysis, and Reflection. *School Effectiveness and School Improvement*, 14 (3), 321–349. doi:10.1076/sesi.14.3.321.15842
- Vom Hofe, R. (2003). Grundbildung durch Grundvorstellungen. *Mathematik lehren*, 118/2003, 4–8.
- Wember, F. (1998). Zweimal Dialektik: Diagnose und Intervention, Wissen und Intuition. *Sonderpädagogik*, 28 (2), 106–120.
- Wiesner, C. & Schreiner, C. (2019). Implementation, Transfer, Progression und Transformation: Vom Wandel von Routinen zur Entwicklung von Identität. Von Interventionen zu Innovationen, die bewegen. Bausteine für ein Modell zur Schulentwicklung durch Evidenz(en). In C. Schreiner, C. Wiesner, S. Breit, P. Dobbelstein, M. Heinrich & U. Steffens (Hrsg.), *Praxistransfer Schul- und Unterrichtsentwicklung* (S. 79–140). Münster: Waxmann. Verfügbar unter: <https://www.waxmann.com/?eID=texte&pdf=3936Volltext.pdf&typ=zusatztext#page=79>
- Wiesner, C., Schreiner, C., Breit, S. & Kemethofer, D. (2017). Evidenzorientierte Schul- und Unterrichtsentwicklung. *BIFIE Online-Journal*, 1/2017. Verfügbar unter: https://www.iqs.gv.at/_Resources/Persistent/2a2407d5a60f6aec9d2ba64900acca5eafc29032/bifie_journal_1-2017-06.pdf <https://doi.org/10.17888/bifiejournal-1.2017-1-6>

- Wiesner, C., Schreiner, C., Breit, S. & Pacher, K. (2017). Bildungsstandards und kompetenzorientierter Unterricht. *BIFIE Online-Journal*, 1/2017. https://www.iqs.gv.at/_Resources/Persistent/88778ae8e56c0559e89759153e84363131b257a1/bifie_journal_1-2017-01.pdf
- Zeitler, S., Heller, N. & Asbrand, B. (2013). Bildungspolitische Vorgaben und schulische Praxis. Eine Rekonstruktion der Orientierungen von Lehrerinnen und Lehrern bei der Einführung der Bildungsstandards. *ZISU – Zeitschrift für interpretative Schul- und Unterrichtsforschung*, 2 (1), 110–127. Verfügbar unter: <https://elibrary.utb.de/doi/epdf/10.3224/zisu.v2i1.17412>