




Anwendungsbezug in Mathematik-Large-Scale-Assessments im Bildungsmonitoring für den Sekundarstufe I – Bildungsgang des sonderpädagogischen Schwerpunkts Lernen

Sven Basendowski · Gilbert Greefrath 

Eingegangen: 13. November 2022 / Angenommen: 1. März 2024
© The Author(s) 2024

Zusammenfassung Inklusive Schulentwicklung erfordert, dass Large-Scale-Assessments zum Bildungsmonitoring die Leistungen auch auf den unteren Kompetenzstufen konzeptionell fundiert und methodisch angemessen erheben. Im Gegensatz zu bisherigen Aufgabensets in Bildungsmonitoringstudien, die für diesen Kompetenzbereich nur verzerrte Ergebnisse liefern, wird in einer ersten explorativen Neukonzeption speziell für Jugendliche aus dem sonderpädagogischen Schwerpunkt Lernen gezeigt, dass Aufgabenpools für die unteren Kompetenzbereiche möglich sind. Ziel dieses Beitrags ist es, alle 88 anwendungsbezogenen empirisch leichten bzw. schweren Aufgaben im neukonzipierten Aufgabenpool auf den unteren Kompetenzstufen hinsichtlich Modellierungsphasen, Authentizität und Relevanz zu charakterisieren. Im Vergleich zum bisherigen Aufgabenpool erwiesen sich im neu konzipierten mehr Items mit Relevanz in eher authentischen und in privaten Kontexten als geeignet. Dieser explorative Befund ist bei der Weiterentwicklung eines Kompetenzmodells für Large-Scale-Assessments in inklusiven Settings zu berücksichtigen.

Schlüsselwörter Sonderpädagogik · Anwendungen · Bildungsmonitoring · Authentizität · Relevanz

✉ Gilbert Greefrath

Institut für Didaktik der Mathematik und der Informatik, Universität Münster, Münster, Deutschland
E-Mail: greefrath@uni-muenster.de

Application Reference in Mathematics Large-scale Assessments in the Educational Monitoring of the Secondary Special Needs Education Programme

Abstract Inclusive school development requires that large-scale assessments for educational monitoring also measure performance at the lower competence levels in a conceptually sound and methodologically appropriate manner. In contrast to previous task pools in educational monitoring studies, which only provide biased results for this competence area, this first explorative new conceptualisation specifically for adolescents with a special educational focus on learning shows that it is possible to construct task pools for the lower competence areas. The aim of this article is to characterise all 88 application-related empirically easy or difficult tasks in the newly designed task pool on the lower competence levels regarding modelling phases, authenticity, and relevance. In the newly designed task pool, more items with relevance in more authentic and private contexts proved to be suitable. This exploratory finding should be taken into account in the further development of a competency model for large-scale assessment in inclusive settings.

Keywords Special education · Applications · Educational monitoring · Authenticity · Relevance

1 Einleitung

Large-Scale Assessments (LSA) im Fach Mathematik, deren Ergebnisse für die Weiterentwicklung des Bildungssystems in Deutschland bedeutsam sind, umfassen neben den internationalen Schulleistungsstudien TIMSS oder PISA auch nationale Untersuchungen wie den deutschen Bildungstrend (Mullis et al. 2021; OECD 2018; Stanat et al. 2019). Mit Hilfe der Ergebnisse können Aussagen zur Leistungsfähigkeit des Bildungssystems getroffen werden, auch um politische Entscheidungen auf der Ebene des Schulsystems vorzubereiten (Leutner et al. 2008). Mit der im Zuge des Übereinkommens über die Rechte von Menschen mit Behinderungen (UN-BRK 2008) erforderlichen Transformation von einem allgemeinbildenden hin zu einem inklusiven Bildungssystem gilt es auch LSA dahingehend weiterzuentwickeln, die unterschiedlichen Bildungs-, Beratungs- und Unterstützungsbedarfe spezifisch in den Blick zu nehmen, um die Leistungsfähigkeit des Bildungssystems für alle Lernenden sicherzustellen. Dieser Beitrag fokussiert auf die spezifischen Bedarfe der Schülerinnen und Schüler, die in einem Sekundarstufen-I-Bildungsprogramm unterhalb des Niveaus eines Ersten Allgemeinbildenden Schulabschlusses (ESA) beschult werden. Diese Gruppe findet sich gemäß KMK (2020) überwiegend an Regelschulen und an Förderschulen Lernen.

Angesichts der Relevanz von LSA gibt es eine jahrzehntelange Diskussion zu ihrer Optimierung. Dabei geht es um eine passende Aufgabenschwierigkeit und allgemeine mathematischer Kompetenzen sowie um die Erfassung des konzeptionellen Verständnisses (Burkhardt und Swan 2017). Seit einigen Jahren wird darüber hinaus vereinzelt forschungsbasiert problematisiert, ob und wie mit LSA mathe-

matische Kompetenzen der Schülerschaft erhoben werden können, die auf oder unterhalb der niedrigsten Kompetenzstufe performen. Konsens war, dass die Ergebnisse für Jugendliche mit attestiertem sonderpädagogischen Bildungs-, Beratungs- und Unterstützungsbedarf im Schwerpunkt Lernen (sBBU-L) nicht exakt gemessen werden können (Müller et al. 2017). Hinzu kommt, dass bei der Entwicklung der Testinstrumente Schülerinnen und Schüler mit sBBU-L nicht repräsentativer Teil der Probandengruppe waren. Bestehende LSA eignen sich in einem inklusiven Bildungssystem folglich nicht vollständig für Jugendliche mit Performanzen auf oder unter der niedrigsten Kompetenzstufe.

Im Diskurs zu Anpassungen der Items in LSA wird betont, dass anwendungsbezogene Aufgaben als empirisch schwerer gelten als entsprechende innermathematische Aufgaben mit gleichem mathematischem Inhalt aber ohne außermathematischen Kontext (Knoche und Lind 2004). Gründe werden in den zusätzlichen Anforderungen durch den Kontext, u. a. auf Grund der sprachlichen Herausforderungen (Prediger et al. 2015), gesehen. Mit dem Bildungstrend 2018 liegt eine explorative Weiterentwicklung eines LSA in Mathematik in der Sekundarstufe I vor. Diese legt nahe, die pauschale Einschätzung, Anwendungsaufgaben seien per se empirisch schwieriger, differenzierter zu betrachten. Der Lebensweltnähe und der Authentizität des Kontexts kommt hierbei eine entscheidende Rolle zu (Mahler et al. 2020). Allerdings wurde bisher nicht überprüft, wie jene konzeptionellen Anforderungen im Aufgabenpool tatsächlich umgesetzt wurden. In der vorliegenden Studie wird daher zum einen untersucht, ob die Aufgaben in den unteren Leistungsbereichen die Anforderungen bezüglich Anwendungsbezug im Sinne des PISA-Frameworks erfüllen können. Da das mathematische Modellieren diesbezüglich eine besondere Herausforderung darstellt, fokussiert die Analyse auf diese allgemeine mathematische Kompetenz. Zum anderen wird die Relevanz und die Authentizität des Anwendungsbezugs betrachtet. Das Ziel ist eine Analyse der Merkmale der Kontexteinbettung in einem Aufgabenpool, der Performanzen bei Anwendungsaufgaben auf den untersten Kompetenzstufen messen kann.

2 Anwendungsbezug von Items in Large-Scale-Assessments zu Mathematik

Die Funktion von summativen LSA wie den PISA- oder IQB-Bildungstrend-Studien unterscheidet sich von der formativen bzw. Prozess-Diagnostik, mit der individualdiagnostische Aussagen als Grundlage für die Gestaltung und Reflexion konkreter Bildungsangebote einzelner Lernender getroffen werden (Gold 2011; Leutner et al. 2008). Es werden Schlussfolgerungen für die konkrete Gestaltung geeigneter Bedingungen in der entsprechenden Unterrichtseinheit bzw. zugehöriger Einzelförderung als Teil einer pädagogischen Entscheidung zur Unterrichtseinheit gezogen bzw. diese Entscheidung evaluiert (Moser Opitz 2022). Konzeptionell stehen daher bei einer formativen bzw. Prozess-Diagnostik rasch änderbare spezifische Kompetenzen einzelner Lernender im Fokus. Summative LSA erheben Kompetenzen in einer zugleich fachlichen Tiefe und Breite entlang eines Kompetenzstrukturmodells wie dem der mathematischen Kompetenzen der Bildungsstandards (Leutner et al.

2008). Summative LSA unterscheiden sich weiter von den Lernstandserhebungen, die flächendeckend in verschiedenen Bundesländern eingesetzt werden. Diese Verfahren haben zum Ziel, pädagogische, didaktische bzw. curriculare Entscheidungen auf Schul- und Unterrichtsebene zu treffen (Leutner et al. 2008, S. 151). Vor dem Hintergrund dieser Unterschiede konzentrieren wir uns in diesem Beitrag auf summative LSA. Genauer betrachten wir Items in summativen LSA auf den unteren Kompetenzstufen, die einen Anwendungsbezug aufweisen.

2.1 Relevanz des Anwendungsbezugs in LSA

Anwendungsbezug ist in vielen LSA von besonderer Bedeutung. Mit der Bezeichnung *Anwendungsbezug* schließen wir hier alle Aspekte von Beziehungen zwischen Mathematik und Realität bzw. „Rest der Welt“ (Pollak 1977) ein. Häufig wird jedoch mit *Anwendungen* eher die Richtung von der Mathematik zur Realität angesprochen, während *Modellieren* eher die umgekehrte Richtung von der Realität zur Mathematik sowie die entsprechenden Prozesse kennzeichnet. Mit mathematischem Modellieren wird ein bestimmter Aspekt der angewandten Mathematik bezeichnet, der als mehrmaliges Durchlaufen eines Modellierungskreislaufs (Blum und Leiß 2005) beschrieben werden kann. Insgesamt ist Anwendungsbezug die allgemeinere Bezeichnung (Kaiser et al. 2023; Niss et al. 2007).

Insbesondere internationale LSA fokussieren bei der Erhebung mathematischer Kompetenzen im Jugend- und Erwachsenenalter Bezüge zur Realität. So wird in der PIAAC-Studie etwa *alltagsmathematische Kompetenz* erhoben, die die Fähigkeit und Bereitschaft bezeichnet, „alltägliche mathematische Informationen abzurufen, zu verwenden und zu interpretieren und somit den unterschiedlichen mathematischen Anforderungen im Alltag erfolgreich zu begegnen“ (Rammstedt 2013, S. 12). Ebenso werden in der PISA-Studie mathematische Kompetenzen „in *authentischen* Anwendungssituationen“ (Baumert et al. 2001, S. 19) erhoben. Dies gilt gleichermaßen für den Entwurf des PISA-Frameworks für 2022, das die *Relevanz* der Mathematik für Schülerinnen und Schüler verdeutlichen soll und Aufgaben weiterhin in authentischen Kontexten stellt. So bleibt ein Modellierungskreislauf ein zentraler Baustein des PISA-Frameworks (OECD 2018). Entsprechend sollen auch die Aufgaben in den Bildungstrendstudien für die Sekundarstufe I in Deutschland zu einem großen Teil „in einen authentischen Kontext aus der Alltagswelt eingebettet“ (Roppelt et al. 2019, S. 34) sein. Auch in den TIMS-Studien gibt es drei kognitive Bereiche von denen einer das Anwenden von Mathematik (applying) ist. Darin sollen Schülerinnen und Schüler Wissen und konzeptionelles Verständnis in einer Reihe von Situationen anwenden. Dieser Bereich kommt in 40 % aller Testaufgaben in der Sekundarstufe vor (Mullis et al. 2021).

Die PISA-Studien verwenden zudem ein Grundbildungskonzept mit einer Akzentuierung auf diverse Anwendungsbezüge, das zum Ziel hat, Mathematik in vielfältigen Kontexten zu formulieren, anzuwenden und zu interpretieren (Reinhold et al. 2019). Ein Anwendungsbezug ist somit ein grundlegendes Merkmal von Aufgaben in den PISA-Studien und weiteren international bekannten LSA. Dies gilt in wesentlichen Teilen auch für deutschsprachige LSA. Die Bildungstrendstudien in Deutschland (Stanat et al. 2019) beruhen auf den Bildungsstandards im Fach Mathematik

(KMK 2004, 2022), die die Kompetenzen in inhaltsbezogenen Leitideen und prozessbezogenen allgemeinen Kompetenzen beschreiben. In den Bildungsstandards ist die Kompetenz des mathematischen Modellierens eine wichtige allgemeine Kompetenz, die den Umgang mit realitätsbezogenen Aufgaben erfordert. Im Schweizer Bildungsmonitoring-Projekt zur Überprüfung von Grundkompetenzen spielen reale Probleme und der Handlungsaspekt „Mathematisieren und Modellieren“ ebenfalls eine wichtige Rolle (Konsortium ÜGK 2019). In Bildungsstandardüberprüfungen in Mathematik in Österreich wird der Handlungsbereich „Darstellen und Modellbilden“ hervorgehoben und der Aufbau von Grundvorstellungen betont (Müller et al. 2022).

Ohne Berücksichtigung der konzeptionellen Merkmale Relevanz und Authentizität von anwendungsbezogenen Aufgaben finden sich Studien, die aufzeigen, dass diese Aufgaben empirisch schwieriger sind als vergleichbare innermathematische Aufgaben (Knoche und Lind 2004). Darüber hinaus gibt es solche Unterscheide zwischen mathematisch vergleichbaren Anwendungsaufgaben mit unterschiedlichen Kontexten (Stern 2003). Weiter gibt es Studien, die Testaufgaben bezüglich der erforderlichen Modellierungsprozesse sowie weiterer allgemeiner Kompetenzen untersuchen (Neubrand et al. 2013; Scheja et al. 2023; Turner et al. 2013). So zeigte sich beispielsweise, dass Modellieren in zentralen Prüfungen nach Klasse 10 nur in geringem Maße in niedrigen Anforderungsbereichen vorkommt (Scheja et al. 2023). Dies steht im Einklang mit der Untersuchung von im Unterricht und in Klassenarbeiten eingesetzten Aufgaben im Rahmen der COACTIV-Studie (Neubrand et al. 2013). Die Ergebnisse aus einer Untersuchung zu Prüfungen zum Erwerb des Mittleren Schulabschlusses zeigen darüber hinaus, dass die Kompetenzen aus den Bildungsstandards sehr unausgewogen abgebildet werden (Kühn und Drüke-Noe 2013). Tendenziell gilt: Je niedriger das Abschlussniveau eines Bildungsgangs, desto seltener werden Modellierungsaufgaben gestellt. Der Fokus auf die Analyse von LSA im unteren Kompetenzbereich auf Items mit Anwendungsbezug hat für weitere Bildungsprogramme eine Relevanz.

2.2 Operationalisierung des Anwendungsbezugs in LSA

Die für die Charakterisierung der Bearbeitung von LSA-Aufgaben relevanten Modellierungsprozesse können aus dem Gesamtprozess des Modellierens bestehen, der durch einen Modellierungskreislauf idealisiert beschrieben werden kann, oder auch aus einzelnen Teilprozessen im Modellierungskreislauf (KMK 2022). Die Fähigkeit, einen solchen Teilprozess wie Verstehen, Vereinfachen etc. im Modellierungskreislauf durchzuführen, kann als eine spezifische Teilkompetenz des mathematischen Modellierens angesehen werden, während Modellierungskompetenz sich auf Fähigkeiten bezieht, den gesamten Modellierungsprozess durchzuführen und über ihn zu reflektieren (Kaiser 2007; Maaß 2004, 2006). Teilkompetenzen des Modellierens können durch entsprechende Beschreibungen genauer charakterisiert werden (Greefrath et al. 2013, S. 19; Maaß 2006, S. 116f., s. Anhang).

Inwiefern sich die eingesetzten Modellierungsaufgaben für unterschiedliche Zielgruppen in LSA unterscheiden, ist seither nicht analysiert worden. Dies trifft u. a. für PISA-Aufgaben zu, die in für Jugendliche *relevante Kontexte* nach unterschiedlichen

Lebensbereichen unterschieden werden können (OECD 2018). Dabei werden Kontexte differenziert, die einen *persönlichen*, einen *beruflichen*, einen *gesellschaftlichen* und einen *wissenschaftlichen* Bereich ansprechen (Reinhold et al. 2019, S. 191). Auch die in der PIAAC-Studie erhobene alltagsmathematische Kompetenz wird vergleichbar in die vier Lebensbereiche privater Alltag, Beruf, gesellschaftlicher Kontext und Kontext von Aus- und Weiterbildung unterteilt (Rammstedt 2013).

Neben der Bandbreite der zu berücksichtigenden relevanten Kontexte wird in den LSA-Frameworks die Authentizität der Testitems als Anforderung hervorgehoben. Bezüglich des außermathematischen Kontextes beschreibt z. B. die OECD *authentische Aufgaben* als „basierend auf Situationen, die ... die Arten von Problemen repräsentieren, die im wirklichen Leben vorkommen“ (OECD 2001, S. 23). Authentizität bezieht sich auf einen außermathematischen Kontext, der in der jeweiligen Situation mit mathematischen Mitteln thematisiert werden muss. Der außermathematische Kontext sollte authentisch sein und nicht nur für diese spezielle mathematische Aufgabe konstruiert werden. So sollte der Einsatz von Mathematik in dieser Situation nicht auf den Mathematikunterricht beschränkt sein.

Während die Operationalisierung von Authentizität in den PISA-Frameworks nicht konkretisiert wird, gibt es hierzu eine intensive fachdidaktische Diskussion. Authentizität von Aufgaben wird in verschiedene Klassifikationsschemata und Beschreibungen für Modellierungsaufgaben aufgenommen (Greefrath und Frenken 2021; Maaß 2010; Ross und Adleff 2022), aber auch in allgemeinen Klassifikationen von Aufgaben als Kriterium genannt (Maier et al. 2010). Authentizität von Aufgaben wird als Einbeziehung von Problemen der realen Welt und als Verankerung in der realen Welt außerhalb des Klassenzimmers charakterisiert (Garrett et al. 2016). Authentische Modellierungsaufgaben bezeichnen solche Anwendungsaufgaben, die Bezüge zum wirklichen Leben nur leicht vereinfachen (Kaiser und Schwarz 2010). Authentizität hilft den Schülerinnen und Schülern, die Aufgabe ernst zu nehmen und damit die kontextabhängige Ausweichstrategie, nur die vorgegebenen Zahlen der Aufgabe zu verwenden, zu ignorieren (Greefrath et al. 2013).

Jedoch wird der Begriff der Authentizität einer Aufgabe teilweise unterschiedlich verwendet (Turner et al. 2022). Ein zentraler Aspekt von Authentizität ist, wie sehr eine mathematische Aufgabe einer realen Situation ähnelt (Garrett et al. 2016). Dabei kann die Authentizität durch verschiedene Merkmale wie Realitätsnähe der Situation, der Fragestellung sowie der gegebenen Informationen und Hilfsmittel (Palm 2007) genauer aufgeschlüsselt werden. Wenn mehrere Dimensionen einer Aufgabe authentisch sind, ist es wahrscheinlicher, dass die Schülerinnen und Schüler die für die Lösung notwendigen Überlegungen zur realen Welt anstellen (Palm 2008). Palm (2008) betrachtet daher eine Situation als authentisch, wenn sie eine reale Aufgabensituation darstellt und wenn wichtige Aspekte dieser Situation in einem angemessenen Maße simuliert werden. Eine authentische Aufgabe ist demzufolge für die Lernenden glaubwürdig und gleichzeitig realistisch in Bezug auf eine reale Situation (Palm 2007). Maaß (2010) verwendet die Bezeichnung „authentisch“ in Abgrenzung zu (nur) realistischen Aufgaben. Sie sind dann in ein Gebiet außerhalb der Mathematik eingebettet und befassen sich mit Problemen, die auch von Experten in diesem Bereich als relevant anerkannt werden (Niss 1992). Andererseits wird die Bezeichnung authentisch verwendet, um die Art der Beziehung zwischen dem

Kontext und der Realität zu charakterisieren. Eine Fokussierung auf die Authentizität zentraler Dimensionen von Aufgaben wie Situation, Frage und Methoden ist sinnvoll (Turner et al. 2022; Vos 2018) und wird auch in diesem Beitrag verwendet.

3 Anwendungsaufgaben in Large-Scale-Assessments des sonderpädagogischen Schwerpunkts Lernen

Es gibt nur wenige LSA zu mathematischen Kompetenzen im deutschsprachigen Raum für Jugendliche mit attestiertem sBBU-L. In der Forschung dominieren standardisierte Schulleistungstests (u. a. DEMAT oder ERT), die nicht direkt für diese Gruppe von Lernenden entwickelt wurden (Moser Opitz 2022). Für Jugendliche mit attestiertem sBBU-L gibt es einige regionale Studien (KESS-7-F, LAU-F, BELLA) und Zusatzerhebungen zu (inter-)nationalen Erhebungen (z. B. zu PISA). Im Hinblick auf LSA auf den unteren Kompetenzstufen hat sich gezeigt, dass der Einsatz von Testaufgaben der allgemeinen Schulen – auch mit einigen Anpassungen – nicht ausreicht, um Schülerinnen und Schüler mit sBBU-L gerecht zu werden (K. Müller et al. 2017). So kamen bei einer Zusatzauswertung zum Nationalen Bildungspanel (Südkamp et al. 2015), zum IQB-Ländervergleich 2011 (Kocaj et al. 2016) sowie zu PISA 2012 (Müller et al. 2017) teils Aufgaben aus den bisher verwendeten Pools für die 4. bis 6. Klassenstufe mit teils veränderter Bearbeitungszeit zum Einsatz. Jene Testaufgaben erwiesen sich jedoch als nicht geeignet, auch die Kompetenzen der Jugendlichen mit attestiertem sBBU-L adäquat, d. h. ohne statistisch relevante Verzerrung, zu erfassen. Obwohl die Kompetenzen der Jugendlichen mit attestiertem sBBU-L folglich nur ungenau gemessen werden können (Müller et al. 2017), geben diese Zusatzerhebungen Hinweise darauf, dass der Gesamtmittelwert dieser Fünfzehnjährigen unter Kompetenzstufe I liegen könnte, und damit niedriger als bei den Gleichaltrigen an den übrigen allgemeinbildenden Sekundarschulbildungsprogrammen. Gleichzeitig zeigt sich eine hohe Varianz in den erreichten Kompetenzstufen unter diesen Jugendlichen mit sBBU-L, denn neben dem gültigen Anteil von 61,5 % unter Kompetenzstufe I erreichten 27,9 % die Kompetenzstufe I und 10,6 % höhere Kompetenzstufen (Müller et al. 2017, S. 186). Ein Teil der Lernenden mit sBBU-L war somit bei PISA 2012 in der Lage, Anwendungsaufgaben mathematisch zu lösen, die mehrschrittige Prozeduren verlangen.

Die Befundlage spricht konsequenterweise für die notwendige Entwicklung spezifischer Items, die geeignet sind, Performanzen in den unteren Kompetenzstufen valide und reliabel zu differenzieren. Dies gilt insbesondere für Items mit Anwendungsbezug. Einen Ansatz zeigt der Testteil Ökonomie in der Studie „BELLA“, in dem sprachlich und layouttechnisch einfacher gehaltene Anwendungsaufgaben zum Einsatz kamen (Lehmann und Hoffmann 2009), in denen arbeitsmarktrelevante „Situationen zu analysieren und [...] mit Hilfe einfacher mathematischer Modelle und Routinen zu bearbeiten“ sind (Seeber 2009, S. 91). Diese Anwendungsaufgaben sind zudem durch eine „Situierung der Aufgaben in der Lebenswelt der Schülerinnen und Schüler“ gekennzeichnet (Seeber 2009, S. 89). Die erhobenen Performanzen der teilnehmenden Jugendlichen mit sBBU-L aus 9. Klassen zeigen, dass darunter 25 % „in der Lage [sind], grundlegende ökonomische Entscheidungen auf Basis einfacher

mathematischer Prozeduren zu treffen, wobei im Aufgabenkontext beschriebene Bedingungen berücksichtigt werden“ (Seeber 2009, S. 97). In Anbetracht der geringen Item-Zahl des Tests und der nicht zufriedenstellenden Testgüte liefert das konstruierte Kompetenzstrukturmodell jedoch keine generalisierbaren Aussagen. Es lässt sich aber vermuten, dass bei geeignetem Kontext nicht nur eine Minderheit unter den Jugendlichen mit sBBU-L Anwendungsaufgaben im unteren Kompetenzbereich lösen kann (Seeber 2009). Diese Annahme knüpft an die in der sonderpädagogischen Fachrichtung Lernen aufgezeigten Beeinträchtigungen des Lernens entlang des curricularen Prinzips des „praktischen Problembezugs“ (Wember 2015) an. Dabei wird der Werkzeugcharakter mathematischer Begriffe und Konzepte für Teilhabe in verschiedenen Lebensbereichen betont (Schroeder 2015; Werner 2019). Es ist daher wichtig, kritisch zu hinterfragen, ob die Tendenz in LSA, weniger Anwendungsaufgaben für niedrigere Abschlussniveaus zu stellen, tatsächlich zutrifft und geeignet ist. Demgegenüber ist den Merkmalen des Kontextes von Anwendungsaufgaben und deren Effekten auf Lösungswahrscheinlichkeiten mehr Aufmerksamkeit zu schenken. Insofern Modellierungskompetenzen zentral für das Fach Mathematik sind (KMK 2004), so ist dieser Kompetenzbereich auch in LSA für alle Schülerinnen und Schüler in einem inklusiven Bildungssystem sicherzustellen (vgl. UN-BRK 2008).

3.1 Der Bildungstrend als Large-Scale-Assessment mit spezifischen Anpassungen für den sonderpädagogischen Schwerpunkt Lernen

Im Rahmen des Bildungstrends 2018 wurde ein zielgruppenspezifischer Aufgabenpool – zusätzlich zum bisher verwendeten Aufgabenpool früherer Ländervergleiche – entwickelt (Mahler et al. 2020), welchem die Bildungsstandards für das Fach Mathematik (KMK 2004) zu Grunde liegen. Der Bildungstrend untersucht in regelmäßigen Abständen eine repräsentative Stichprobe aller Schülerinnen und Schüler der Klassenstufen 4 und 9 ohne einen sonderpädagogischen Bildungs-, Beratungs- und Unterstützungsbedarf. In 2018 umfasst die repräsentative Stichprobe auch Schülerinnen und Schülern mit attestiertem sBBU-L und soll Aufschluss darüber geben, inwieweit die Lernenden in der Klassenstufe 9 die in den nationalen Bildungsstandards formulierten Kompetenzerwartungen erreicht haben (Stanat et al. 2019). Die Kompetenzstufe Ib wird hierbei mit den Mindestanforderungen für einen ESA assoziiert (Kölm und Mahler 2019, S. 158). Das Kompetenzstrukturmodell ist konstituiert durch drei Dimensionen: allgemeine mathematische Kompetenzen, Leitideen und Anforderungsbereiche (Frenken et al. 2022; Rupp und Vock 2007). Die Leitideen Zahl, Messen, Raum und Form, funktionaler Zusammenhang sowie Daten und Zufall entstammen dem Modell der Bildungsstandards (KMK 2004, 2022). Im Vergleich zu den bisher für die Jugendlichen mit sBBU-L spezifisch eingesetzten Items in Bildungstrendstudien weist die entsprechende Studie zur Neuentwicklung im Rahmen des Bildungstrends 2018 darauf hin, dass ein zufriedenstellend valides Instrument für Vergleiche zwischen Lernenden des Bildungsprogramms sBBU-L und weiteren – nicht-gymnasialen – Sek-I-Bildungsprogrammen gefunden werden konnte (Mahler et al. 2020, S. 135). Die Neuentwicklung fokussierte konzeptionell die unteren Kompetenzbereiche und legte Wert auf veränderte Bearbeitungszeiten,

auf die Erstellung eines möglichst leicht verständlichen Testmaterials zur Minimierung sprachlicher Hürden sowie auf die Einhaltung eines hohen Authentizitätsgrades der mehrere Aufgaben rahmenden lebensweltnahen Situationen (Mahler et al. 2020, S. 115). Jedoch wurden keine Analysen in der Auswertung vorgenommen, wie sich diese Anpassungen tatsächlich im finalen Aufgabenpool niederschlugen oder die Performanzen beeinflussten.

4 Forschungsfragen

Die Befundlage unterstreicht die konzeptionelle, didaktische und teilhabesensible Relevanz von Items mit Anwendungsbezug in LSA in Mathematik. Ferner gibt es in der Forschung keinen Konsens darüber, dass anwendungsbezogene Aufgaben in LSA in den unteren Kompetenzstufen in Mathematik für die Sekundarstufe I eliminiert werden sollten, weil sie per se zu komplex sind.

Erste Ansätze geben Hinweise darauf, dass es unter Berücksichtigung spezifischer Anpassungen der Authentizität und der Relevanz des Anwendungsbezugs gelingen kann, einen validen und reliablen Aufgabenpool zu entwickeln, der den Leistungen von Jugendlichen aus mehreren Bildungsgängen, darunter auch dem sonderpädagogischen Bildungsgang sBBU-L, gerecht wird. Angesichts der dokumentierten Testgüte des neu konzipierten Aufgabenpools für Jugendliche aus anderen Bildungsgängen ist von einer Übertragbarkeit der Anpassungen des neu konzipierten Aufgabenpools auf Jugendliche auszugehen, die auf den unteren Kompetenzstufen in den Bildungstrend-Studien performen und in anderen Bildungsgängen unterrichtet werden.

Vor einer Übertragung der vielversprechenden Ergebnisse des Bildungstrends 2018 auf weitere LSA muss jedoch noch analysiert werden (Q1), ob der neu konzipierte Aufgabenpool mit Anwendungsbezug auf den unteren Kompetenzstufen den Anforderungen an die Items von LSA bezüglich der Leitideen und allgemeinen Kompetenzen speziell hinsichtlich des Anwendungsbezugs gerecht wird. Dazu werden in der vorliegenden Studie die Items bezüglich der Kompetenz des mathematischen Modellierens genauer analysiert. In einem zweiten Schritt (Q2) soll qualitativ analysiert werden, ob sich die Items des neu konzipierten Aufgabenpools hinsichtlich der Authentizität und Relevanz des Anwendungsbezugs nach fachdidaktischen Kriterien des PISA-Frameworks je Kompetenzstufe unterscheiden. In der abschließenden Analyse (Q3) sollen die gefundenen Unterschiede in den charakterisierenden Merkmalen des neu konzipierten Aufgabenpools mit denen des bisherigen Bildungstrend-Aufgabenpools verglichen werden. Dies ist von Interesse, um Hinweise geben zu können, welche Merkmale des Kontextes sich für die Weiterentwicklung zukünftiger Aufgabenpools für LSA in einem inklusiven Bildungssystem eignen. Wir stellen also die folgenden drei Forschungsfragen für den im Bildungstrend 2018 neu konzipierten Aufgabenpool für sBBU-L:

Q1: *Entspricht* der neu konzipierte Aufgabenpool für sBBU-L mit Anwendungsbezug auf den unteren Kompetenzstufen *den konzeptionellen Anforderungen der Leitideen und prozessbezogenen Kompetenzen*, insbesondere hinsichtlich des Anwendungsbezugs?

Q2: Inwieweit *unterscheiden* sich empirisch geeignete LSA-Items mit Anwendungsbezug auf den unteren Kompetenzstufen hinsichtlich der *Authentizität und Relevanz* des Anwendungsbezugs *je Kompetenzstufe*?

Q3: Inwieweit *unterscheiden* sich empirisch geeignete LSA-Items mit Anwendungsbezug auf den unteren Kompetenzstufen *von bisher verwendeten LSA-Items* hinsichtlich Leitideen, prozessbezogenen Kompetenzen, Modellierungsphase sowie Authentizität und Relevanz des Anwendungsbezugs *je Kompetenzstufe*?

5 Methode

Die Items der IQB-Bildungstrendstudie 2018 wurden von Lehrkräften unter der Leitung des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) entwickelt. Dabei wurden teils Items aus dem Ländervergleich 2012 wiederverwendet und teils durch neu konzipierte Items spezifisch für Jugendliche mit sBBU-L ergänzt. Hintergrund der Neukonzeption war, wie bereits dargelegt, die bei bisherigen Aufgabenpools festzustellende ungenügende Messgenauigkeit für Jugendliche mit sBBU-L (Roppelt et al. 2019, S. 33). Die Rahmenkonzeption für die neuen Aufgaben ist von methodischen Prinzipien der sonderpädagogischen Disziplin Lernen geprägt, wie etwa Lebensweltnähe und angepasste Sprache (Mahler et al. 2020). Die von Sonderpädagogen und -pädagoginnen – nicht den Autoren dieses Beitrags – neu entwickelten Aufgaben wurden durch die Autoren bezüglich der kategorialen Einordnung in das Kompetenzmodell der Bildungsstandards kommentiert. Darüber hinaus wirkten die Autoren dieses Beitrags nicht an der Entstehung oder Auswahl der Aufgaben der Bildungstrendstudie 2018 mit. Vor dem Einsatz im Bildungstrend 2018 wurden die Items mit mehreren hundert Schülerinnen und Schülern sowohl an allgemeinen Schulen als auch an Förderschulen mit attestiertem sBBU-L erprobt und anschließend selektiert. Im Bildungstrend 2018 selbst fanden die aktualisierten Aufgaben Anwendung als Teil des Aufgabenpools für alle Probandinnen und Probanden ungeachtet des besuchten Bildungsprogramms (Mahler et al. 2019a). Diese Bildungsmonitoring-Studie bietet sich daher einerseits für die Bearbeitung der Forschungsfragen an, da die Personengruppe sBBU-L konzeptionell und organisatorisch repräsentativer Teil der Probandengruppe in der Pilotierungsstudie war (Mahler et al. 2019a; Mahler et al. 2019b). Andererseits ist die absolute Anzahl der Probanden aus der Personengruppe sBBU-L deutlich höher als in Studien zu dieser Zielgruppe sonst üblich (Mahler et al. 2019a). Die Aufgaben einschließlich der Itemschwierigkeitsparameter aus der Bildungstrend 2018-Studie bilden folglich eine aussagekräftige, repräsentative Grundlage für die formulierten Forschungsfragen. Zu diesem Zweck wurde daher der gesamte Aufgabenpool einschließlich Informationen zu Itemschwierigkeiten und zugeordneten Kompetenzstufen beim IQB angefragt. Diese Daten dürfen jedoch nur mit der Auflage genutzt werden, keine Items zu veröffentlichen.

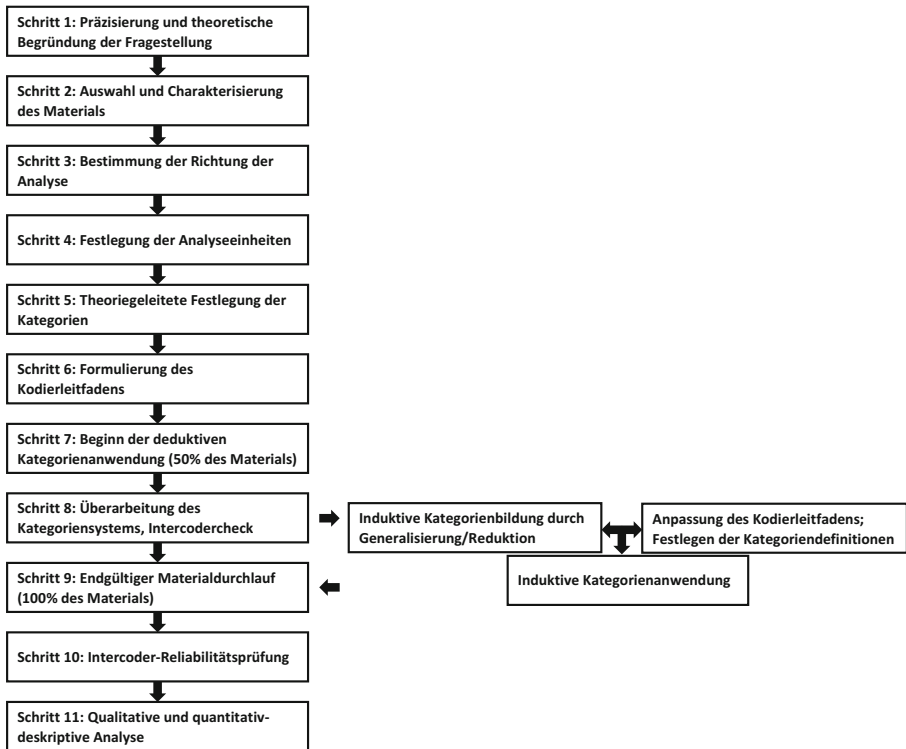


Abb. 1 Auswertungsplan

5.1 Wahl der Untersuchungsmethode

Für die angestrebte qualitative Analyse der ausgewählten fachdidaktischen Kriterien empfiehlt sich eine strukturierende qualitative Inhaltsanalyse als theorie- und empiriegeleitetes Analyseverfahren (Gläser-Zikuda et al. 2020). Dieses bewährte Verfahren eignet sich für eine systematische Analyse größerer Materialmengen in Dokumenten ohne Tiefenanalyse (Gläser-Zikuda et al. 2020; Mayring 2010, S. 611). Mit diesem Verfahren können einheitlich diskutierte fachdidaktische Merkmale zu Anwendungsaufgaben wie etwa Modellierungsphasen bei der Analyse deduktiv beachtet werden und zugleich weniger intensiv diskutierte bzw. einheitlich operationalisierte Merkmale wie etwa Authentizität und Relevanz während des Analyseprozesses durch das Prinzip der Offenheit methodisch kontrolliert berücksichtigt werden. Für die vorliegende Studie wurde das in Abb. 1 dargestellte Regelverfahren in Anlehnung an Mayring (2010, S. 605) adaptiert. Das deduktive Kategoriensystem wurde nach dem 1. Materialdurchgang hinterfragt und angepasst (Schritt 8 in Abb. 1). Das finale Kategoriensystem ist nicht auf die vordefinierten Merkmale beschränkt. Diese konnten während des Analyseprozesses bei Bedarf angepasst, subsumiert, gestrichen oder erweitert werden. Mit der begründeten Entscheidung für die strukturierende qualitative Inhaltsanalyse geht einher, dass hierbei die Gütekriterien des

Tab. 1 Übernommene Kodierungen

Konstrukt	Kategorie(n) [veröffentlichte Gesamtanteile]	Quelle
Kompetenzstufe je Item	Ia	Kodierung: IQB, Konzept: Mahler et al. 2019b, Blum et al. 2019b
	Ib	
	II	
Leitideen	Leitidee Zahl	Kodierung: IQB, Konzept: KMK (2004), vgl. auch https://www.iqb.hu-berlin.de/bt/BT2018/
	Leitidee Messen	
	Leitidee Raum und Form	
	Leitidee Funktionaler Zusammenhang	
	Leitidee Daten und Zufall	
Anforderungsbereiche	Anforderungsbereich 1: Reproduzieren	Kodierung: IQB, Konzept: KMK (2004), vgl. auch https://www.iqb.hu-berlin.de/bt/BT2018/
	Anforderungsbereich 2: Zusammenhänge herstellen	
	Anforderungsbereich 3: Verallgemeinern und Reflektieren	

Tab. 2 Verteilung der analysierten Items mit Anwendungsbezug nach Kompetenzstufen (Deskription der Kompetenzstufen s. Blum et al. 2019)

Kompetenzstufen	n _{Regel}	n _{sBBU-L}
Ia	1	36
Ib	11	39
II	28	13
Gesamt	40	88

qualitativen Paradigmas zu beachten sind (Gläser-Zikuda et al. 2020; Helsper et al. 2019).

5.2 Festlegung der Analyseeinheiten

Gegenstand der strukturierten qualitativen Inhaltsanalyse sind die im IQB-Bildungstrend 2018 (Stanat et al. 2019) verwendeten Testaufgaben, bestehend aus den Items selbst und den Auswertungsanleitungen. Einige Kategorisierungen der Aufgaben wurden direkt vom IQB übernommen: Dies gilt für die Zuordnungen aller Items zu Kompetenzstufen, Leitideen und Anforderungsbereichen (siehe Tab. 1). Die Festlegung der Kompetenzstufen am IQB beruht dabei auf dem Verfahren des Standard-Settings, bei dem die empirische Itemschwierigkeit je Testitem Berücksichtigung findet. Kompetenzstufen beschreiben Aufgabengruppen, die abgrenzbare Anforderungen beinhalten. Dabei beschreibt etwa Kompetenzstufe II den Regelstandard für den Hauptschulabschluss (Blum et al. 2019).

Unter den 521 Testitems für den Bildungstrend 2018 waren 188 für den Bildungsgang sBBU-L und 333 für die allgemeinen Schulen vorgesehen. Gemäß dem Forschungsinteresse wurden darunter die Items ausgewählt, die einen Anwendungsbezug aufweisen (etwa als eingekleidete Aufgabe, Textaufgabe oder Sachaufgabe; vgl. Greefrath et al. 2013). Die verbleibenden Items aus dem Aufgabenpool für sBBU-L stammten ausschließlich aus den Kompetenzstufen Ia, Ib und II. Daher wurden mit Blick auf das Forschungsinteresse nur Items mit diesen Kompetenzstufen als Analysegegenstand gewählt. Mit Abschluss von *Schritt 4* (vgl. Abb. 1) wurde

somit ein Datenkorpus von $n_{\text{Regel}} = 40$ Items aus dem IQB-Bildungstrend-Aufgabenpool für Regelschulen und von $n_{\text{sBBU-L}} = 88$ Items aus den neu konzipierten Items für den Bildungsgang sBBU-L fixiert (vgl. Tab. 2). Für den ersten Materialdurchgang (Schritt 7) wurden zufällig etwa die Hälfte der Items ausgewählt und für die Kodierung mit dem initialen deduktiven Kategoriensystem in MAXQDA aufbereitet. Die Auswertung erfolgte jeweils unabhängig durch zwei Coder.

5.3 Theoriegeleitete Festlegung der Kategorien und Entwicklung eines Kodierleitfadens

Das ursprüngliche deduktive Kategoriensystem basierte auf den theoretisch beschriebenen Kriterien zur Charakterisierung authentischer und relevanter Modellierungsaufgaben. Hierzu wurden die Kriterien „Modellierungsphasen“, „Authentizität der Frage“, „Authentizität der Situation“, „Authentizität der Methoden“ und „Relevanz“ herangezogen (s. Tab. 3), welche sich in Frameworks von LSA als sinnvoll erwiesen haben. Das damit anfänglich deduktive Kategoriensystem wurde vor dem 1. Materialdurchlauf um einen Kodierleitfaden ergänzt, in dem erste Regeln und Verfahrensschritte für die beiden Coder festgelegt wurden (Schritt 6). Die Anzahl der Kodierungen entlang der Authentizitäts- und Relevanzkriterien wurde auf 1 limitiert, um diesen weniger einheitlich diskutierten fachdidaktischen Merkmalen kontrolliert begegnen zu können. Dieser Bedarf wurde für die Kategorie „Modellierungsphase“ nicht gesehen und es wurden auch aus inhaltlichen Gründen Mehrfachkodierungen zugelassen. Weiter wurde vereinbart, wie während des 1. Materialdurchlaufs Diskriminierungskriterien und Ankerbeispiele bis zur qualitativen Sättigung festzuhalten sind: Memos wurden unabhängig voneinander je Kategorie notiert, bis sich diese Beschreibung als erschöpfend erwies. Das Kategoriensystem und der Kodierleitfaden wurden nach dem Prinzip der Offenheit bezüglich der Ausdifferenzierung festgelegter Kategorien in Subkategorien optimiert (Helsper et al. 2019). Es wurde vereinbart, während des 1. Materialdurchgangs zu den theoriegeleiteten Subkategorien Ankerbeispiele und Diskriminierungskriterien unabhängig voneinander zu sammeln und bei Bedarf neue Subkategorien zu entwickeln. Der anschließen-

Tab. 3 Deduktiv-induktives Kategoriensystem (Deskriptoren, Diskriminierungskriterien und Ankerbeispiele siehe Anhang, Tab. 8)

Kategorie	Subkategorien
Allgemeine mathematische Kompetenzen	Mathematisch argumentieren, Probleme mathematisch lösen, Mathematisch modellieren, Mathematische Darstellungen verwenden, Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen, Kommunizieren
Modellierungsphase	Verstehen, Vereinfachen, Mathematisieren, Mathematisch arbeiten, Interpretieren, Validieren, Vermitteln
Authentizität der Frage	Mindestens simuliert authentisch, simuliert, nicht bewertbar
Authentizität der Situation	Mindestens simuliert authentisch, simuliert, nicht bewertbar
Authentizität der Methoden	Mindestens simuliert authentisch, simuliert, nicht bewertbar
Relevanz	Persönlicher Bereich, Beruflicher Bereich, Gesellschaftlicher Bereich, Wissenschaftlicher Bereich

de 2. Materialdurchlauf (Schritt 9) wurde mit dem kompletten Datensatz und dem finalen Kategoriensystem durchgeführt.

5.4 Überarbeitung des Kategoriensystems und endgültiger Materialdurchgang einschl. Intercoder-Reliabilitätsprüfung

Vor dem Hintergrund der breiten Diskussion von Gütekriterien qualitativer Forschung (Eisewicht und Grenz 2018; Helsper et al. 2019) orientierten wir uns, ausgehend von Steinke (1999, 2019), insbesondere für die *Schritte 7–10* an Gläser-Zikuda et al. (2020). Dazu zählt die konsensuelle Validierung und die transparente Darlegung zur Nachvollziehbarkeit, wie „the researchers analyzing the qualitative data together, and discussing the development or application of categories“ (Gläser-Zikuda et al. 2020, S. 7). Demgemäß soll transparent gemacht werden, dass die regelgeleiteten Abstimmungen zwischen beiden Codern nach dem 1. Materialdurchlauf entscheidend zur Weiterentwicklung eines deduktiv-induktiven Kategoriensystems führten. Konkret wurden die Kodierregeln angepasst. Es wurden nicht mehr alle relevanten, sondern die bis zu drei für die Aufgabenstellung relevantesten Modellierungsprozesse kodiert; an den Einfachkodierungen für die weiteren Kategorien wurde unverändert festgehalten. Die im deduktiven Kategoriensystem noch auf die Arbeiten von Palm (2007) theoriebasiert übernommenen fünf Authentizitätskriterien wurden auf drei Authentizitätskategorien (Frage, Situation und Methode) reduziert. Diese

Aufgabe: „Verdienstabrechnung“

Daniel arbeitet nach seiner Ausbildung.

Jeden Monat erhält er seine Verdienstabrechnung.

In der Verdienstabrechnung steht Folgendes:

- Lohn (was Daniel im Monat insgesamt verdient)
- Abzüge und Sozialversicherung (was Daniel abgezogen wird)
- Nettoverdienst (was Daniel auf sein Konto überwiesen wird)

Verdienstabrechnung

Name: Daniel Müller

Steuerklasse: 1

Konfession: katholisch

Kinderfreibeträge: 0

	Euro
<u>Lohn:</u>	1.533,88 €
<u>Abzüge:</u>	
Lohnsteuer:	-167,87 €
Kirchensteuer:	-15,11 €
Solidaritätszuschlag:	-9,23 €
<u>Sozialversicherungsbeiträge:</u>	
Krankenversicherung:	-105,33 €
Pflegeversicherung:	-13,26 €
Rentenversicherung:	-150,59 €
Arbeitslosenversicherung:	-50,62 €
<u>Nettoverdienst:</u>	1.021,87 €

Teilaufgabe 1:

Wie viel € zahlt Daniel für die Krankenversicherung? Kreuze an.

- 9,23 €
- 15,11 €
- 105,33 €
- 167,87 €

Abb. 2 Veröffentlichtes Beispielitem aus der Neukonzeption sBBU-L; Bildungstrend 2018. (<https://www.iqb.hu-berlin.de/bt/BT2018/Beispielaufgaben>)

Tab. 4 Inter-coder-Reliabilität

Fachdidaktische Merkmale	Kappa1.Durchgang	Kappa2.Durchgang
Modellierungsphase	0,23	0,77
Authentizität Frage	0,47	0,69
Authentizität Situation	0,20	0,83
Authentizität Methode	0,18	0,72
Relevanz	0,25	0,91

Reduktion deckt sich mit anderen Arbeiten aus dem Diskussionsstand um Authentizität (Turner et al. 2022; Vos 2018). Auch eine binäre Kodierung von authentisch vs. nicht-authentisch zu allen Authentizitätssubkategorien erwies sich als nicht dem Datenmaterial entsprechend. Daher wurde eine Differenzierung von authentisch (vgl. Anhang, Tab. 8: mindestens simuliert authentisch=Frage, Situation oder Methode mindestens in Ansätzen glaubwürdig und realistisch; simuliert=Frage, Situation oder Methode unglaubwürdig; nicht bewertbar) verwendet. Das finale deduktiv-induktiv entwickelte Kategoriensystem kann dem Anhang entnommen werden. Zur Erhöhung der Nachvollziehbarkeit sind die Ankerbeispiele aus dem veröffentlichten Beispielitem aus der Neukonzeption für sBBU-L aus dem Bildungstrend (siehe Abb. 2) verwendet worden. Exemplarische Kodierungen zu weiteren Beispielen aus bisherigen Aufgabenpools finden sich im Anhang.

Die dargelegten Entscheidungen zu Kodierregeln und zu Subsummierungen von Subkategorien zwischen dem ersten halben und dem zweiten vollständigen Materialdurchgang wurden durch die Überprüfung der Inter-coder-Übereinstimmung abgesichert (siehe nachfolgend). Kommunikative Validierung als weiteres gängiges Gütekriterien der qualitativen Inhaltsanalyse gemäß Gläser-Zikuda et al. (2020, S. 7) bedeutet für die vorliegende Analyse von Dokumenten eine kollegiale Validierung (Mayring 2020) durch weitere Expertinnen und Experten. Diese konnte jedoch aufgrund der Auflagen des IQB zur Verschwiegenheit zu den Testitems nicht vorgenommen werden. Daher wurde besonderer Wert auf die konsequente Einhaltung der Auswertungsschritte der strukturierten qualitativen Inhaltsanalyse (s. Abb. 1) gelegt.

Als drittes gängiges Gütekriterium für qualitative Inhaltsanalysen nennen Gläser-Zikuda et al. (2020) die Prüfung mittels einer Inter-coder-Übereinstimmung, d. h. die Prüfung „of the stability of applying the coding frame, the replicability of coding by groups of two or more coders, and the accuracy of the coding procedure“ (S. 7) und geben hierfür einen anzustrebenden Wert von $Kappa = 0,70$ als „sufficient indicator“ an. Die mittels MAXQDA ermittelten Kappa-Werte je Oberkategorie im 1. Durchgang mit der Hälfte der Items und unabhängig davon im 2. Durchgang mit den verbleibenden Items zwischen Coder 1 und 2 sind Tab. 4 zu entnehmen. Es zeigt sich, dass die Stärke der Übereinstimmung bei allen fachdidaktischen Merkmalen durch die oben benannten Kodierregeln auf das angestrebte Konsistenzniveau gesteigert werden konnte. Es liegt nach dem 2. Durchgang folglich ein Ergebnis vor, das keine weiteren Überarbeitungen des deduktiv-induktiven Kategoriensystems notwendig macht.

5.5 Einschätzung entlang der Gütekriterien qualitativer Forschung

Aus der Diskussion zur Einhaltung der Gütekriterien für die qualitative Inhaltsanalyse lassen sich abschließend Einschätzungen zum Grad der Verallgemeinerung (Helsper et al. 2019) vornehmen. Diese beziehen sich auf die mit den Forschungsfragen angestrebte Beschreibung von Unterschieden zwischen Aufgabengruppen. Diese Beschreibung von Mustern in den Aufgaben basiert zwar nicht auf einer vollständigen kommunikativen kollegialen Validierung, kann aber auf Basis der erreichten Konsistenzwerte im 2. Materialdurchgang und aller Maßnahmen zur kommunikativen Validierung als gesichert betrachtet werden.

Andererseits basiert die Beschreibung von Unterschieden in den Aufgabengruppen zu inhaltlichen Merkmalen auf den empirisch in der Bildungstrend-2018-Studie gesetzten Kompetenzstufen Ia, Ib und II (Stanat et al. 2019). Dennoch wurde von Alternativen zu Spektren von Itemschwierigkeiten, z. B. Quartilen, Abstand genommen, da in der Diskussion Kompetenzstufen dominieren. Auch wenn die Schwierigkeitsparameter des Bildungstrend-2018-Aufgabenpools auf kontrollierten, und insbesondere für die Personengruppe sBBU-L geeigneten Modellen beruht, bleibt darüber hinaus die vorgenommene Beschreibung von Unterschieden der Aufgabengruppen an den spezifischen Aufgabenpool gebunden. Die fachdidaktischen Merkmale können daher nicht auf andere LSA übertragen werden. Gleiches gilt für Bildungsmonitoring-Studien, die organisatorisch die Zielgruppe sBBU-L nicht umfassen. Die beschriebenen Unterschiede im Kontext des Bildungstrends 2018 können für die Weiterentwicklung von LSA in einem inklusiven Bildungssystem Auskunft über zu berücksichtigende fachdidaktische Merkmale geben, die die Wahrscheinlichkeit erhöhen, dass Anwendungsaufgaben in einem ausreichend validen Modell auch auf unteren Kompetenzstufen Teil der erhobenen Kompetenzen sind.

6 Ergebnisse – qualitative und quantitative Analyse der Kategorien

6.1 Konzeptionelle Anforderungen bezüglich der Leitideen und prozessbezogenen Kompetenzen insbesondere hinsichtlich des Anwendungsbezugs (Forschungsfrage 1)

Die im Bildungstrend verwendeten Aufgaben mit Anwendungsbezug für Lernen mit sBBU-L decken, wie für dieses LSA erforderlich, alle Leitideen und alle prozessbezogenen Kompetenzen der seinerzeit gültigen Bildungsstandards (KMK 2004) je Kompetenzstufe ab (s. Tab. 5). Die Anforderungsbereiche I und II sind etwa gleich häufig vertreten, jedoch konnte der Anforderungsbereich III erwartungsgemäß nicht gefunden werden. Weiter lässt sich feststellen, dass das mathematische Arbeiten unter allen Teilschritten des Modellierens am häufigsten vertreten ist. Nahezu gleichermaßen häufig ist das Mathematisieren, während das Vereinfachen nur eine untergeordnete Rolle spielt (s. Tab. 6). Bezogen auf alle Teilschritte des Modellierens ist festzustellen, dass auf der höchsten hier betrachteten Kompetenzstufe II die Bandbreite geringer ist als bei den unteren beiden Kompetenzstufen. Auf Kompetenzstufe II wurden nur Mathematisieren, mathematisch Arbeiten und Interpretieren

Tab. 5 Anzahl der Items für sBBU-L in den Kategorien der Bildungsstandards

Kompetenzstufe	Ia	Ib	II
Mathematisch argumentieren	2	4	1
Probleme mathematisch lösen	2	2	2
Mathematisch modellieren	11	4	2
Mathematische Darstellungen verwenden	12	7	4
Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen	54	24	19
Kommunizieren	32	54	12
Leitidee Zahl	36	24	10
Leitidee Messen	24	11	5
Leitidee Raum und Form	2	12	8
Leitidee Funktionaler Zusammenhang	4	1	5
Leitidee Daten und Zufall	17	25	5
Anforderungsbereich 1	36	52	8
Anforderungsbereich 2	40	17	19
Anforderungsbereich 3	0	0	0

Tab. 6 Anzahl der Items für sBBU-L in den Kategorien zum Anwendungsbezug

Kompetenzstufe	Ia	Ib	II
Vermitteln	0	0	0
Validieren	5	3	0
Interpretieren	21	14	6
Math. Bearbeiten	63	38	21
Mathematisieren	43	35	23
Vereinfachen	1	2	0
Verstehen	0	0	0
<i>Authentizität der Frage</i>			
Nicht bewertbar	36	33	15
Simuliert	32	23	6
Mindestens simuliert authentisch	8	15	6
<i>Authentizität der Situation</i>			
Nicht bewertbar	0	0	0
Simuliert	41	26	22
Mindestens simuliert authentisch	35	35	5
<i>Authentizität der Methoden</i>			
Nicht bewertbar	44	31	14
Simuliert	26	22	8
Mindestens simuliert authentisch	4	18	5
<i>Relevanz</i>			
Privat	18	21	1
Gesellschaftlich	2	4	1
Beruflich	2	2	0
Wissenschaftlich	54	47	25

gefunden, während auf den anderen beiden Stufen auch weitere Teilschritte identifiziert wurden (s. Tab. 6).

Es lässt sich somit konstatieren, dass der neu konzipierte Aufgabenpool von Anwendungsaufgaben insbesondere auf den untersten Kompetenzstufen Ia und Ib den Anforderungen von LSA an die Leitideen und allgemeinen Kompetenzen einschließlich den Teilprozessen des Modellierens entspricht.

6.2 Unterschiede hinsichtlich der Authentizität und Relevanz des Anwendungsbezugs je Kompetenzstufe (Forschungsfrage 2)

Wir vergleichen hier die Items aus der Neukonzeption der Kompetenzstufe Ia mit den Items der beiden höheren Kompetenzstufen Ib und II (s. Tab. 6) nach Authentizität und Relevanz.

In den Authentizitätskategorien gibt es wenige deutliche Unterschiede zwischen den Aufgaben der verschiedenen Kompetenzstufen. Tendenziell gibt es in der Kompetenzstufe Ia weniger authentische Items bezogen auf die Frage und die Methoden. Diese Unterschiede sind aber gegenüber Stufe II lediglich bei der Authentizität der Situation deutlich. Übergeordnet zeigt sich das Muster, dass die meisten Items auf allen Kompetenzstufen keine authentische, sondern eine simulierte Frage stellen und dabei kaum authentische Hilfsmittel eine Rolle spielen. Während auf den Kompetenzstufen Ia und Ib zumindest in nahezu jeder zweiten Aufgabe eine authentische Situation gegeben ist (s. Tab. 6), werden nur bei Ib für die Situation einige authentische Fragen gestellt bzw. authentische Hilfsmittel zur Verfügung gestellt.

Bezogen auf das Kriterium der Relevanz zeigt sich, dass die meisten Testaufgaben für Jugendliche in wissenschaftlichen oder privaten Kontexten relevant sind, aber nicht in Bezug auf berufliche oder gesellschaftlich-öffentliche Lebensbereiche (s. Tab. 6). Diese Verteilung gilt gleichermaßen für die Aufgaben auf der Kompetenzstufe Ia zu Ib, während die Aufgaben auf der Kompetenzstufe II nahezu ausschließlich in wissenschaftlichen Kontexten relevant sind.

Es kann geschlussfolgert werden, dass geeignete LSA-Anwendungsaufgaben auf den untersten Kompetenzstufen für private Kontexte für die Jugendlichen tatsächlich relevant sowie Frage bzw. Situation mindestens simuliert authentisch sind. Darunter fallen folglich nicht eingekleidete Textaufgaben, bei denen Frage und Situation letztlich beliebig austauschbar sind. Die Authentizität der Methoden ist demgegenüber eine Kategorie, die sich vergleichsweise als schwieriger in der LSA-Itementwicklung herausgestellt hat und daher diese Items zu einem geringen Maße den Kern geeigneter LSA-Aufgaben mit Anwendungsbezug ausmachen.

6.3 Unterschiede zwischen LSA-Items mit Anwendungsbezug auf den unteren Kompetenzstufen von bisher verwendeten LSA-Items je Kompetenzstufe (Forschungsfrage 3)

Im Vergleich zwischen der bisherigen Konzeption und der Neukonzeption für sBBU-L in den Kompetenzstufen I und II zeigen sich erwartungsgemäß die auffälligsten Unterschiede bezüglich der Anforderungsbereiche (s. Tab. 7). Weitere Unterschiede liegen im Anteil der Allgemeinen mathematischen Kompetenz *mit*

Tab. 7 Anzahl der Items in der bisherigen Konzeption und sBBU-L in den ausgewählten Kategorien

	Kompetenzstufe Ia		Kompetenzstufe Ib		Kompetenzstufe II	
	Bisherige Konzeption	Neukonzeption sBBU-L	Bisherige Konzeption	Neukonzeption sBBU-L	Bisherige Konzeption	Neukonzeption sBBU-L
N Items	22	76	2	71	56	22
<i>Allgemeine mathematische Kompetenzen</i>						
Mathematisch argumentieren	2	2	0	4	4	1
Probleme mathematisch lösen	5	2	0	2	17	2
Mathematisch modellieren	5	11	1	4	14	2
Mathematische Darstellungen verwenden	6	12	1	7	14	4
Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen	5	54	0	24	26	19
Kommunizieren	17	32	2	54	32	12
<i>Leitideen</i>						
Leitidee Zahl	4	36	0	24	19	10
Leitidee Messen	0	24	0	11	2	5
Leitidee Raum und Form	2	2	0	12	8	8
Leitidee Funktionaler Zusammenhang	6	4	0	1	13	5
Leitidee Daten und Zufall	11	17	2	25	21	5
<i>Anforderungsbereiche</i>						
Anforderungsbereich 1	10	36	1	52	20	8
Anforderungsbereich 2	12	40	1	17	34	19
Anforderungsbereich 3	0	0	0	0	2	0
<i>Modellierungskompetenzen</i>						
Vermitteln	0	0	0	0	0	0
Validieren	1	5	0	3	8	0
Interpretieren	9	21	2	14	23	6
Math. arbeiten	10	63	0	38	34	21
Mathematisieren	12	43	0	35	31	23
Vereinfachen	0	1	0	2	4	0
Verstehen	0	0	0	0	2	0
<i>Authentizität der Frage</i>						
Nicht bewertbar	11	36	0	33	25	15
Simuliert	11	32	2	23	21	6
Mindestens simuliert authentisch	1	8	0	15	9	6
<i>Authentizität der Situation</i>						
Nicht bewertbar	0	0	0	0	0	0
Simuliert	14	41	0	26	31	22
Mindestens simuliert authentisch	7	35	2	35	22	5

Tab. 7 (Fortsetzung)

	Kompetenzstufe Ia		Kompetenzstufe Ib		Kompetenzstufe II	
	Bisherige Konzeption	Neukonzeption sBBU-L	Bisherige Konzeption	Neukonzeption sBBU-L	Bisherige Konzeption	Neukonzeption sBBU-L
<i>Authentizität der Methoden</i>						
Nicht bewertbar	13	44	0	31	31	14
Simuliert	8	26	2	22	19	8
Mindestens simuliert authentisch	1	4	0	18	6	5
<i>Relevanz</i>						
Privat	0	18	0	21	0	1
Gesellschaftlich	0	2	0	4	0	1
Beruflich	0	2	0	2	0	0
Wissenschaftlich	22	54	2	47	50	25

symbolischen, formalen und technischen Elementen der Mathematik umgehen, entlang einiger Teilprozesse des mathematischen Modellierens und den Kriterien Authentizität der Situation bzw. Relevanz:

- In beiden Aufgabenpools entsprechen Items mit Anwendungsbezug auf den Kompetenzstufen I und II allesamt dem Anforderungsbereich I oder II.
- In Bezug auf die prozessbezogenen Kompetenzen ist in beiden Aufgabenpools das *Umgehen mit symbolischen, formalen und technischen Elementen der Mathematik* sowie das *Kommunizieren* besonders häufig. In der bisherigen Konzeption ist das *Problemlösen* aber im Vergleich häufiger vertreten.
- In beiden Pools ist das *mathematische Arbeiten* der am häufigsten benötigte Teilschritt des Modellierens. Bei der Neukonzeption ist der Abstand zu den anderen Teilschritten hier allerdings größer und das *Interpretieren* wird tendenziell etwas seltener benötigt. Die Bedeutung des Mathematisierens ist in beiden Pools etwa gleich.
- Die Authentizität der Aufgaben in den beiden Pools ist generell gering. Es scheint aber einen Trend zu geben, dass die neu konzipierten Aufgaben in der Kategorie *Situation* zu einem größeren Anteil als die bisher verwendeten Aufgaben mindestens simuliert authentisch sind. In den anderen Authentizitätskategorien sind keine Unterschiede festzustellen.
- Bezogen auf die *Relevanz* der Aufgaben sind größere Unterschiede sichtbar. Zwar ist in beiden Gruppen die Relevanz im Wesentlichen auf wissenschaftliche Kontexte beschränkt, d. h. es werden exemplarische Fragen beschrieben, die privaten, gesellschaftlichen und zukünftigen beruflichen Leben der Schülerinnen und Schüler nicht von Bedeutung sind. Es zeigt sich aber, dass es unter den neu entwickelten Aufgaben deutlich mehr Items gibt, die im privaten Kontext als relevant eingestuft werden können.

7 Diskussion

7.1 LSA-Aufgabenpool mit Anwendungsbezug auf den unteren Kompetenzstufen

Die Analyse hat ergeben, dass es durchaus möglich ist, geeignete und valide Anwendungsaufgaben auf den untersten Kompetenzstufen zu entwickeln. Die im Rahmen des Bildungstrend 2018 neu konzipierten Aufgaben erfüllen insbesondere auf den unteren Kompetenzstufen Ia und Ib die konzeptionellen Anforderungen an LSA-Aufgaben an die Breite und Tiefe der Items, indem alle Leitideen, prozessbezogene Kompetenzen und Teilprozesse des Modellierens abgebildet werden. Wichtige internationale Studien wie TIMSS oder PISA (Mullis et al. 2021; OECD 2018) basieren auf diesen Anforderungen an die Konzeption von LSA-Items mit Anwendungsbezug. Im deutschen Bildungstrend werden aufgrund des zugrunde gelegten Kompetenzmodells jedoch auch innermathematische Aufgaben verwendet (Stanat et al. 2019). Selbst wenn man die innermathematischen Aufgaben außer Acht lässt, erfüllt das verbleibende Testinstrument, das aus anwendungsbezogenen Testaufgaben besteht, wichtige Anforderungen an Breite, Tiefe und Vollständigkeit der Inhalte sowie der prozessbezogenen Kompetenzen (Roppelt et al. 2019).

Die betrachteten Aufgaben stellen also eine mögliche Entwicklungsrichtung im Sinne des Anwendungsbezugs aus dem PISA-Framework (OECD 2018) für Lernen- de mit sBBU-L dar. Es ist daher möglich, allein für die unteren Kompetenzstufen Ia und Ib Testaufgaben mit Anwendungsbezug zu entwickeln, die sowohl den empirischen als auch inhaltlichen Voraussetzungen aktueller LSA-Frameworks genügen. Solche Ergebnisse sind für mögliche Entscheidungen auf der Ebene des Schulsystems (Leutner et al. 2008) von großer Bedeutung. Insbesondere vor dem Hintergrund, dass auch in der Schweiz und in Österreich Studien mit stärkerem Anwendungsbezug durchgeführt werden, sind sie relevant (Konsortium ÜGK 2019; Müller et al. 2022).

Da die dokumentierten Analysen ergaben, dass mit dem neu konzipierten Aufgabenpool die Performanzen von Jugendlichen in unterschiedlichen Bildungsgängen unterhalb des Gymnasiums, einschließlich des Bildungsgangs sBBU-L, valide erhoben werden können, kann diese zentrale Schlussfolgerung auf die große Mehrheit der Jugendlichen ungeachtet des Bildungsgangs übertragen werden, die auf den untersten Kompetenzstufen performen. Im Rahmen der mathematischen Anforderungen in LSA-Items bedarf es jedoch tendenziell einiger Anpassungen hinsichtlich der Kriterien Authentizität und Relevanz.

Erstaunlich ist, dass auf der höchsten hier betrachteten Kompetenzstufe II die Bandbreite der erforderlichen Teilschritte des Modellierens sogar geringer als bei den unteren Kompetenzstufen ist. Dies verdeutlicht die Möglichkeit vielfältiger Modellierungsaktivitäten auf allen Kompetenzstufen. Die mögliche Dominanz des technischen Arbeitens auf der Kompetenzstufe Ia ist daher vermutlich nicht an die Kompetenzstufe gekoppelt, sondern ein Phänomen, das bereits bei Prüfungsaufgaben mehrfach beobachtet wurde (Kühn und Druke-Noe 2013; Scheja et al. 2023).

Die fehlenden authentischen Kontexte auf allen Ebenen (Fragestellung, Situation und Hilfsmittel) in den Aufgaben sollten im Hinblick auf das PISA-Framework

(OECD 2018) diskutiert werden. Insbesondere für die Zielgruppe mit sBBU-L sind authentische Kontexte von großer Bedeutung (Mahler et al. 2020). Möglicherweise besteht hier noch Entwicklungsbedarf.

Es besteht die Möglichkeit, diese Kontexte an dem für die Zielgruppe betonten Prinzip des lebensweltsensiblen Problembezugs anzupassen (Schroeder 2015; Seeber 2009; Wember 2015). Daher ist die Frage interessant, ob es überhaupt möglich ist, entsprechend authentische und relevante LSA-Items für diese Zielgruppe zu entwickeln. Wenn durch authentische und relevante Kontexte der Anspruch der Aufgaben steigen würde, könnte man durch den Vergleich der unteren Kompetenzstufe mit den darüberliegenden ggf. Unterschiede herausarbeiten. In dieser Untersuchung konnten jedoch keine deutlichen Unterschiede gefunden werden. Dies könnte bedeuten, dass authentische und relevante Kontexte in allen Kompetenzstufen konstruiert werden können. Tatsächlich gibt es aber Indizien, dass die Authentizität in der untersten Kompetenzstufe Ia etwas seltener vorkommt. Daher können diese Bedenken nicht vollständig ausgeräumt werden.

7.2 LSA-Aufgabenpool mit Anwendungsbezug für Lernende mit sBBU-L vor dem Hintergrund bisher verwendeter Testaufgaben

Der Vergleich zwischen der bisherigen Konzeption und der neuen Konzeption von Testaufgaben im Bildungstrend berücksichtigte ausschließlich anwendungsbezogene Aufgaben in den unteren Kompetenzstufen.

Es wurde festgestellt, dass es auffällige Unterschiede zwischen den Konzeptionen gibt. Dies lässt vermuten, dass die neue Konzeption (Mahler et al. 2020) tatsächlich neue Impulse für die Aufgabenentwicklung von LSA-Items mit Anwendungsbezug auf den unteren Kompetenzstufen liefern kann. Insgesamt konnte im Rahmen dieser Studie genauer der Blick auf bisherige und neu entwickelte Testitems in den unteren Kompetenzstufen geworfen werden. Auffällig ist, dass in der bisherigen Konzeption die empirisch leichten Aufgaben mit Anwendungsbezug besonders häufig aus der Leitidee Daten zu Zufall stammten. Hier konnte im Rahmen der Neukonzeption eine inhaltliche Neuorientierung durchgeführt werden, die mit der Leitidee Zahl für die Zielgruppe passend ist. Interessant ist, dass im Rahmen der Neukonzeption das Problemlösen weniger berücksichtigt wurde als bei den bisherigen Aufgaben auf gleicher Kompetenzstufe. Dies ist eine auffällige Veränderung, die für weitere Entwicklungen überdacht werden könnte, da es in der bisherigen Konzeption möglich ist auf diesen Kompetenzstufen Aufgaben zum Problemlösen zu erstellen.

Es ist zu beachten, dass die Verteilung der Leitideen in der Explorationsstudie zum Bildungstrend 2018 möglicherweise unbewusst von der Orientierung an Anwendungsaufgaben in Leistungstests zu Basiskompetenzen für den sonderpädagogischen Schwerpunkt Lernen beeinflusst wurde (Gebhardt et al. 2013; Moser Opitz et al. 2021). Diese Vermutung kann dadurch bestätigt werden, dass im Bereich der prozessbezogenen Kompetenzen bei den neu konzipierten Aufgaben – insbesondere auf Kompetenzstufe II – das technische Arbeiten sehr im Vordergrund steht. Dieses Phänomen wird durch die erhobenen Authentizitätsmerkmale unterstrichen und wurde auch bei der Untersuchung von Prüfungsaufgaben gefunden, bei denen das Modellieren und andere prozessbezogene Kompetenzen nur in geringem Maße ver-

treten waren (Kühn und Drüke-Noe 2013; Scheja et al. 2023). Dies führt dazu, dass das mathematische Arbeiten besonders häufig vorkommt, wenn man die Teilschritte des Modellierens betrachtet.

Natürlich müssen die untersuchten neuen Aufgaben nicht alle Aspekte der Bildungsstandards abdecken (KMK 2004; Roppelt et al. 2019), da wir nur Aufgaben mit Anwendungsbezug auf den unteren Kompetenzstufen betrachten und die vollständigen Testinstrumente (durch innermathematische Aufgaben) umfangreicher sind. Dennoch lässt sich feststellen, dass insbesondere die neu konzipierten Aufgaben auf Kompetenzstufe Ia und Ib aufzeigen, wie eine große Bandbreite an Inhalten (Leitideen) und allgemeinen Kompetenzen abgedeckt werden kann. Im Gegensatz dazu trifft dies bei den bisherigen Aufgabenpools für allgemeine Schulen eher auf Kompetenzstufe II zu. Überraschend ist der Befund, dass in den bisherigen Aufgaben häufiger als im neukonzipierten Aufgabenpool der Kompetenzstufe II authentische Situationen aufgedeckt wurden. Es lässt sich vermuten, dass insbesondere auf Kompetenzstufe Ia und Ib Authentizitätskriterien bei der Gestaltung von für LSA geeigneten anwendungsbezogenen Testitems von Bedeutung sind.

Es gibt verschiedene Testverfahren für den sonderpädagogischen Schwerpunkt Lernen. Einige dieser Verfahren sind jedoch weniger geeignet für Large-Scale-Analysen (Moser Opitz 2022) oder haben weniger Anwendungsbezug (Moser Opitz et al. 2021). Da insbesondere von sehr heterogenen schulmathematischen Leistungen von Lernenden mit attestiertem sBBU-L auszugehen ist (Gebhardt et al. 2013), ist es wichtig, für die Erhebung der Performanzen in größerem Rahmen auch geeignete Large-Scale-Instrumente zur Verfügung zu haben. Diese Instrumente müssen auch entsprechend authentische und relevante Kontexte nutzen, um an internationale Schulleistungsstudien wie PISA anschlussfähig zu sein (OECD 2018).

7.3 Untersuchungsmethode

Als wichtiges Werkzeug und Ergebnis des methodischen Prozesses ist das entwickelte Kategoriensystem herauszustellen, welches intensiv diskutiert und innerhalb von Konsensratings erweitert wurde. Auch wenn auf Grund der qualitativen Auswertungsmethode kein Anspruch auf Generalisierung der Ergebnisse auf andere LSA bestehen kann, können die dabei entstandenen Kategorisierungen als umfassend angesehen werden und eine Grundlage für weitere Studien sein. Auf dieser Basis können also auch Aufgaben anderer LSA hinsichtlich ihres Anwendungsbezuges beurteilt werden. Trotz der guten Übereinstimmung der Bewertung der Coder ist zu bedenken, dass die theoretische Herleitung zu einer Reihe von Kategorien geführt hat, die aber bei einer noch breiteren Sichtweise, z. B. auch in Bezug auf innermathematische Aufgaben, zu weiteren Ergebnissen führen könnte. Es ist jedoch nicht nur die Aufgabe, sondern die Aufgabe einschließlich der Auswertungsanleitung als eine Einheit zu betrachten, so dass auch eine Veränderung der Lösungshinweise zu einer anderen Kodierung führen könnte. Die Kodierung bezieht sich jeweils auf die einzelnen Items, so dass ggf. die Rolle des Stimulus nicht immer eindeutig zu den jeweiligen Items erfasst werden kann.

Eine weitere Limitation ist, dass die Items aufgrund der geplanten Folgenutzungen nicht publiziert werden können, ähnlich wie bei vergleichbar angelegten Studien

(z.B. OECD 2001). Daher konnten die Kategorien zwar auf der Basis von viel Material entwickelt werden, jedoch können sie im Rahmen dieses Beitrags nicht mit mehreren Beispielen belegt werden.

Es ist zu beachten, dass die untersuchten Aufgaben zwar von Kolleginnen und Kollegen mit dem Hintergrund für sBBU-L entwickelt wurden, jedoch anschließend das Ergebnis eines Auswahlprozesses sind. Dieser berücksichtigte nicht nur fachdidaktische Kriterien im Hinblick auf die Zielgruppen, sondern auch Itemschwierigkeit sowie Passung zum Rasch-Modell nach Pilotierung. Dies könnte auch zu Verschiebungen bezüglich der untersuchten Kriterien führen. Auf der anderen Seite sichert die Aufgabenauswahl auf empirischer Basis, dass die Items von den Schülerinnen und Schüler mit sBBU-L im Mittel auch tatsächlich gelöst werden können. Daher können wir auch darlegen, dass Items mit den gewünschten Eigenschaften auch empirisch lösbar sind. Die stärkere fachdidaktische Diskussion der Items hingegen sollte mit dem hier entwickelten Kategoriensystem einerseits angeregt werden. Es lässt sich andererseits das Ergebnis aus weiteren LSA manifestieren, dass keine zu vernachlässigende Minderheit dieser Jugendlichen Kompetenzen aufweisen, anwendungsbezogene Aufgaben auf den Kompetenzstufen Ia bis II zu lösen. Da die summativen LSA keine individualdiagnostischen Aussagen erlauben, kann ein Impuls für die Forschung dieses Typs der Diagnostik gegeben werden.

Es ist zu diskutieren, dass für die bisher verwendeten Testaufgaben lediglich Aufgaben aus den unteren Kompetenzstufen ausgewählt wurden. Diese Entscheidung wurde auf Grund der Vergleichbarkeit der Aufgaben für sBBU-L getroffen, die nur diese Kompetenzstufen umfassen. Für weitere Studien wäre auch ein Vergleich der gesamten Tests miteinander denkbar. Eine gewisse Beschränkung ergibt sich aus der Festlegung des Schwerpunktes auf das Lernen, der jedoch bereits durch die Aufgabenentwicklung festgelegt wurde.

8 Fazit

Die Analyse der Testaufgaben für den Bildungstrend mit dem Schwerpunkt sBBU-L hat gezeigt, dass die fachdidaktische Beurteilung der Items mit Anwendungsbezug nicht nur die aus den Bildungsstandards bekannten prozessbezogenen Kompetenzen in geeigneter Weise einbeziehen kann, sondern weitere Aspekte wie Authentizität und Relevanz für die Konzeption solcher Testinstrumente einschließen sollte. Diese Aspekte können auch Anregungen für andere Testinstrumente liefern, wie im Vergleich mit bisherigen Items herausgearbeitet wurde. Die neue Konzeption der Aufgaben für Lernende mit attestiertem sBBU-L für den Bildungstrend (Mahler et al. 2020) zeigt im Vergleich auch neue Eigenschaften anwendungsbezogener Aufgaben auf niedrigen Kompetenzstufen, die in der bisherigen Konzeption des Bildungstrends so deutlich nicht vorhanden waren. Die tatsächlichen Testitems weichen zwar in Bezug auf ihre theoretische Konzeption (Mahler et al. 2020) von der gewünschten Authentizität und Relevanz ab. Es ist jedoch tendenziell bedeutsam, den Problemcharakter der Situation authentischer zu gestalten und insbesondere relevante private Kontexte stärker zu berücksichtigen, um die Kompetenzen von Jugendlichen mit sBBU-L ohne Verzerrungen zu erfassen. Diese Entwicklung ist

vor dem Hintergrund der bekannten internationalen Schulleistungstudien (OECD 2018) sehr positiv zu beurteilen. Angesichts der dokumentierten Testgüte dieses neukonzipierten Aufgabenpools lassen sich die Schlussfolgerungen auf Jugendliche in weiteren Bildungsgängen übertragen, die im Bildungstrend 2018 auf den unteren Kompetenzstufen bei Items mit Anwendungsbezug performten.

9 Anhang

9.1 Vollständiges Deduktives Kategoriensystem

Tab. 8 Kategoriensystem für anwendungsbezogene Testaufgaben

Kategorie	Subkategorie	Deskriptoren
Allgemeine mathematische Kompetenzen	Mathematisch argumentieren	Mathematische Aussagen hinterfragen und auf Korrektheit überprüfen
		Vermutungen zu mathematischen Zusammenhängen aufstellen
		Eigene Begründungen formulieren und Begründungen anderer nachvollziehen
		A1 → nicht kodiert
		A2 → nicht kodiert
		A3 → nicht kodiert
Probleme mathematisch lösen		Lösungsideen zu Aufgaben entwickeln, zu denen bislang keine Lösungsrouninen bekannt sind
		Lösungsstrategien entwickeln, heuristische Hilfsmittel auswählen und nutzen
		Vorgehensweisen überdenken und ggf. anpassen
		(D) wenn Lösung mittels einer bekannten Lösungsroutine primäre Anforderung des Items ist, dann ist Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen zu kodieren
		A1 → nicht kodiert
		A2 → nicht kodiert
		A3 → nicht kodiert
Mathematisch modellieren		Für die mathematische Bearbeitung einer Fragestellung relevante Informationen u. a. aus Texten, Darstellungen oder der Lebenswirklichkeit entnehmen
		Sachprobleme in die Sprache der Mathematik übersetzen
		Mathematische Lösungen zu Sachproblemen in Bezug auf die Ausgangssituationen prüfen und interpretieren
		Sachaufgaben zu Termen, Gleichungen und bildlichen Darstellungen formulieren
		A1 → kodiert, da Item verlangt, aus einem kontinuierlichen und einem nicht-kontinuierlichen Text die relevanten Informationen zu entnehmen
		A2 → nicht kodiert
		A3 → nicht kodiert

Tab. 8 (Fortsetzung)

Kategorie	Subkategorie	Deskriptoren
	Mathematische Darstellungen verwenden	<p>Geignete mathematische Darstellungsformen für das Bearbeiten mathematischer Fragestellungen auswählen, nutzen und entwickeln</p> <p>Eine mathematische Darstellungsform in eine andere übertragen</p> <p>Mathematische Darstellungsformen miteinander vergleichen und bewerten</p> <p>(D) sind nicht-mathematische Darstellungen, wie z. B. gegenständliche Darstellungen von rechten Winkeln oder Parallelität, zu verarbeiten, dann ist Kommunizieren zu kodieren</p> <p>A1 → nicht kodiert</p> <p>A2 → nicht kodiert</p> <p>A3 → nicht kodiert, da eine nicht-mathematische Darstellung (hier eine Temperaturskala) auf den Anwendungskontext übertragen werden muss</p>
	Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen	<p>Symbolische und formale Sprache in Alltagssprache übersetzen und umgekehrt</p> <p>Mathematische Fachbegriffe und Zeichen sachgerecht verwenden</p> <p>Mathematische Objekte bei der Bearbeitung mathematischer Aufgaben- und Problemstellungen sicher und flexibel verwenden</p> <p>(D) wenn Lösung mittels einer bekannten Lösungsroutine primäre Anforderung des Items ist, dann ist Probleme mathematisch lösen zu kodieren</p> <p>A1 → nicht kodiert</p> <p>A2 → kodiert, da Alltagssprache in symbolische Sprache der Mathematik übersetzt und für die Lösung eine bekannte Lösungsroutine angewendet werden muss</p> <p>A3 → nicht kodiert</p>
	Mathematisch kommunizieren	<p>Überlegungen zu mathematischen Sachverhalten, Lösungswegen und Ergebnissen adressatengerecht beschreiben und erklären</p> <p>Mathematische Zusammenhänge erläutern</p> <p>Lösungen und Lösungswege anderer nachvollziehen und hinterfragen sowie gemeinsam weiterentwickeln</p> <p>A1 → nicht kodiert</p> <p>A2 → nicht kodiert</p> <p>A3 → kodiert, da eine nicht-mathematische Darstellung (Temperaturskala) auf den Anwendungsbezug übertragen werden muss</p>

Tab. 8 (Fortsetzung)

Kategorie	Subkategorie	Deskriptoren
Modellierungsphase Maß (2006), Greefrath et al. (2013)	Verstehen	Gegebene Problemsituation verstehen Text lesen und/oder graphische Darstellung erfassen (D) zu kodieren, falls die Problemsituation für Jugendliche komplex ist oder eine Fülle an Angaben ggf. aus unterschiedlichen Quellen interpretiert werden muss >A1 → kodiert, da die Problemsituation an einer Verdienstabrechnung kontextualisiert wird, bei der fachsprachliche Bezeichnungen und die Beziehungen zwischen den zahlreichen Angaben verstanden werden müssen A2 → nicht kodiert A3 → >kodiert, da die Temperaturskala mit einer Körpertemperatur in Beziehung gebracht werden muss; dabei ist davon auszugehen, dass Jugendliche überwiegend elektronische Fieberthermometer kennen
	Vereinfachen	Mathematische Fragestellung aus Problemsituation entwickeln Gegebene Informationen als un-/relevant interpretieren ggf. fehlende Informationen identifizieren ggf. angegebene Größen mit Hilfsmittel ermitteln (D) Nur zu kodieren, falls für die gegebene Fragestellung irrelevante Informationen aus der Problemsituation herausgefiltert werden müssen A1 → nicht kodiert A2 → nicht kodiert A3 → nicht kodiert
	Mathematisieren	Passende Formel auswählen Term aufstellen Mathematische Darstellung erstellen (einschl. Tabellen) (D) zu kodieren, falls das Item v. a. durch die Anforderung gekennzeichnet ist, eine Formel, einen Term oder eine Darstellung aufzustellen A1 → nicht kodiert A2 → kodiert, da die Anforderung des Items u. a. darin liegt, die angegebenen Punktwerte korrekt in die bekannte Prozentformel einzusetzen A3 → nicht kodiert

Tab. 8 (Fortsetzung)

Kategorie	Subkategorie	Deskriptoren
	Mathematisch arbeiten	<p>Mathematische Operationen durchführen</p> <p>(D) zu kodieren, falls das Item v. a. durch die Anforderung gekennzeichnet ist, eine oder mehrere mathematische Operationen durchzuführen bis heuristische Strategie abwägend bei der Bearbeitung auszuwählen</p> <p>(D) auch zu kodieren, wenn keine weitere Modellierungsphase kodiert wird</p> <p>A1 → nicht kodiert</p> <p>A2 → kodiert, da die Anforderung des Items primär in der Verarbeitung einer bekannten Formel liegt</p> <p>A3 → nicht kodiert</p>
		<p>Interpretieren</p> <p>Mathematische Resultate in reale Resultate übersetzen</p> <p>ggf. Größen, math. Darstellung usw. angemessen berücksichtigen</p> <p>(D) zu kodieren, falls das Item v. a. durch die Anforderung gekennzeichnet ist, mindestens ein Resultat umzurechnen bzw. umzuwandeln [z. B., falls das Beispielim 1 folgende Zielsetzung hätte: <i>Es wären Bruttolohn und Krankenerstreckungsbeitrag in mehreren Ländern mit unterschiedlicher Landeswährung mit der Funktion angegeben, dass die Differenz in Euro ausgerechnet werden soll.</i>]</p> <p>A1 → nicht kodiert</p> <p>A2 → nicht kodiert</p> <p>A3 → nicht kodiert</p>
		<p>Validieren</p> <p>Ermittelte Resultate auf Plausibilität überprüfen</p> <p>Ermittelte Resultate ggf. mit Vergleichsobjekten kontrastieren</p> <p>(D) zu kodieren, falls das Item v. a. durch die Anforderung gekennzeichnet ist, mindestens ein Resultat auf Plausibilität zu überprüfen [z. B. falls das Beispielim 1 folgende Zielsetzung hätte: <i>Der Anteil der Krankenerstreckungsbeiträge vom Bruttolohn sind an einem Schaubild zu überprüfen, bei dem die Beitragsätze einiger GKV gegenübergestellt sind.</i>]</p> <p>A1 → nicht kodiert</p> <p>A2 → nicht kodiert</p> <p>A3 → nicht kodiert</p>
	Vermitteln	<p>Ermittelte Resultate für andere Personen verständlich darlegen</p> <p>(D) zu kodieren, falls das Item v. a. durch die Anforderung gekennzeichnet ist, dass der Bearbeitungsweg zu mindestens einem Resultat schriftlich begründet werden muss [z. B. falls das Beispielim 1 folgende Zielsetzung hätte: <i>Die graphische oder arithmetische Ermittlung des Anteils der Krankenerstreckungsbeiträge vom Bruttolohn Schritt für Schritt erklären und vergleichen.</i>]</p> <p>A1 → nicht kodiert</p> <p>A2 → nicht kodiert</p> <p>A3 → nicht kodiert</p>

Tab. 8 (Fortsetzung)

Kategorie	Subkategorie	Deskriptoren
Authentizität der Frage (Palm 2007, 2008; Turner et al. 2022; Vos 2018)	Mindestens Simuliert authentisch	<p>Problemlösung, die für die beschriebene Situation tatsächlich glaubwürdig und realistisch ist ggf. entspricht lediglich der Grad der Objektivität nicht der beschriebenen Situation (z. B. stark gerundete Lösung)</p> <p>(D) inwiefern die Angaben selbst vereinfacht ggü. der Problemsituation sind, wird unter Authentizität der Situation kodiert</p> <p>(D) es ist „nicht bewertbar“ zu kodieren, falls aus der konkret gegebenen Problemstellung erschlossen werden kann, in der eine mathematische Problemlösung glaubwürdig ist</p> <p>(D) es ist „simuliert“ zu kodieren, falls aus der konkret gegebenen Problemstellung zwar eine Relevanz der gewünschten mathematischen Anforderungen erschlossen werden kann, die jedoch für die konkret gegebene Situation unglaubwürdig ist</p> <p>A1 → kodiert, da die Problemsituation, dass der Auszubildende einen neuen monatlichen Lohnauszug erhält und seinen Krankenversicherungsbeitrag prüft, zwar vage und damit simuliert, aber nicht unglaubwürdig ist</p> <p>A2 → nicht kodiert</p> <p>A3 → kodiert, da die Problemstellung der Messung einer Körpertemperatur durch Ablesen eines Fieberthermometers nicht unglaubwürdig ist</p>
	Simuliert	<p>Problemlösung, die für die beschriebene Situation unglaubwürdig ist ggf. vereinfacht der Grad der Objektivität völlig unangemessen</p> <p>(D) inwiefern die Angaben selbst vereinfacht ggü. der Problemsituation sind, wird unter Authentizität der Situation kodiert</p> <p>A1 → nicht kodiert</p> <p>A2 → kodiert, da die Kommunikation zwischen Schülerinnen und Schülern zu den erreichten Punktzahlen nach Rückgabe eines Tests glaubwürdig ist, die sie direkt aus dem rückgemeldeten Testresultat entnehmen können. Unglaubwürdig ist für den Kommunikationsanlass jedoch, dass der prozentuale Anteil berechnet werden soll</p> <p>A3 → nicht kodiert</p>
	Nicht bewertbar	<p>Bei dem Item liegt keine Problemsituation vor, die in die Sprache der Mathematik glaubwürdig übersetzt werden soll</p> <p>(D) inwiefern die Angaben selbst vereinfacht ggü. der Problemsituation sind, wird unter Authentizität der Situation kodiert</p> <p>A1 → nicht kodiert</p> <p>A2 → nicht kodiert</p> <p>A3 → nicht kodiert</p>

Tab. 8 (Fortsetzung)

Kategorie	Subkategorie	Deskriptoren
Authentizität der Situation (Palm 2007, 2008; Turner et al. 2022; Vos 2018)	Mindestens simuliert authentisch	<p>Mathematische Angaben in der Aufgabenstellung usw. sind für die gegebene Problemsituation tatsächlich glaubwürdig und realistisch ggf. sind Angaben komplexitätsreduzierend</p> <p>(D) es ist „nicht bewertbar“ zu kodieren, falls aus der konkret gegebenen Problemstellung nicht die Glaubwürdigkeit der mathematischen Angaben erschlossen werden kann</p> <p>(D) es ist „simuliert“ zu kodieren, falls die konkret gegebenen mathematischen Angaben für die konkret gegebene Situation unglaubwürdig ist [z. B. zu kodieren, falls das <i>Beispielitem 1 folgende Angaben hätte: Der Lohnauszug würde nur aus wenigen Angaben bestehen und/oder bspw. der angegebene Bruttolohn etc. entspräche eindeutig nicht dem Bruttolohn von Auszubildenden</i>]</p> <p>A1 → kodiert, da zwar die Fülle der Angaben auf dem Lohnauszug vereinfacht ist, dies jedoch ohne die Situation zu verzerren; die Werte selbst entsprechen einem möglichen Lohnauszug von Auszubildenden</p> <p>A2 → kodiert, da die angegebenen Punktwerte im Kontext der Schüleräußerung zur erreichten Punktzahl in einer Mathematikarbeit glaubwürdig und nicht beliebig austauschbar sind</p> <p>A3 → nicht kodiert</p>
	Simuliert	<p>Mathematische Angaben widersprechen oder verzerren die gegebenen Problemsituation</p> <p>Angaben sind nicht für die Situation beliebig austauschbar</p> <p>A1 → nicht kodiert</p> <p>A2 → nicht kodiert</p> <p>A3 → kodiert, da das Ablesen eines Fieberthermometers ohne eine betroffene Person nicht glaubwürdig ist; in der Situation mag die gemessene Körpertemperatur sehr unterschiedliche Folgen haben</p>
	Nicht bewertbar	<p>Angaben sind derart, dass sie beliebig austauschbar sind</p> <p>Bei dem Item liegt keine Problemsituation vor, die in die Sprache der Mathematik glaubwürdig übersetzt werden soll</p> <p>A1 → nicht kodiert</p> <p>A2 → nicht kodiert</p> <p>A3 → nicht kodiert</p>

Tab. 8 (Fortsetzung)

Kategorie	Subkategorie	Deskriptoren
Authentizität der Methoden (Palm 2007, 2008; Turner et al. 2022; Vos 2018)	Mindestens simuliert authentisch	(illustrierte/beschriebene) Hilfsmittel sind in der beschriebenen Problemsituation tatsächlich glaubwürdig und realistisch; d. h. man würde sie in der Situation nutzen Typische Hilfsmittel sind z. B. Messgeräte, Taschenrechner usw A1 → kodiert, da im Item alle Hilfsmittel illustriert werden, die in der Problemsituation glaubwürdig vorliegen würden A2 → nicht kodiert A3 → nicht kodiert
	Simuliert	(illustrierte/beschriebene) sind für die beschriebenen Problemsituation unglaubwürdig Es fehlen Hilfsmittel, die in der Problemsituation vorliegen Typische Hilfsmittel sind z. B. Messgeräte, Taschenrechner usw A1 → nicht kodiert
	Nicht bewertbar	A2 → kodiert, da zum einen der Kommunikationsanlass durch den zurück erhaltenen Test mit den angegebenen Testergebnissen gekennzeichnet ist, der nicht illustriert ist; und zum anderen hätte der Schüler einen Taschenrechner und nicht unterschiedliche Antworten zur Verfügung A3 → kodiert, da das Fieberthermometer nur virtuell vorhanden ist Hilfsmittel sind beliebig austauschbar Keine außermathematische Problemsituation, die Einschätzung über Hilfsmittel erlaubt Typische Hilfsmittel sind z. B. Messgeräte, Taschenrechner usw A1 → nicht kodiert A2 → nicht kodiert

Tab. 8 (Fortsetzung)

Kategorie	Subkategorie	Deskriptoren
Relevanz (OECD 2001; Rammstedt 2013; Reinhold et al. 2019, S. 191)	<i>Persönlicher Bereich</i>	<p>Die konkret gegebene Problemsituation hat eine lebendige Stellung im Interessen- oder Bedürfnisbereich vieler Jugendlichen (vgl. Shell-Studien) außerhalb beruflicher bis öffentlicher (z. B. Schule) Kontexte</p> <p>Problemsituation ist vielen aus eigenen Erfahrungen heraus glaubwürdig und muss daher nicht erst in den Fragehorizont gebracht werden</p> <p>(D) „Beruflicher Bereich“ ist zu kodieren, falls konkrete Problemsituation glaubwürdig für (zukünftige) Tätigkeiten am Arbeitsplatz – aber nicht für die Berufsschule – für Jugendliche ist (z. B. bei diesem fiktiven Item: <i>Ein Item verlangt die Ermittlung des Wassers, welches bei der Zubereitung einer Suppe in der Großküche erhitzt werden soll. Neben dem Rezept wird ein Kochkessel mit integrierter Füllstandsanzeige abgebildet. Die Problemsituation kann für die meisten Jugendlichen mit den beruflichen Anforderungen an einem Arbeitsplatz. In der Großküche verknüpft wenden.</i>)</p> <p>(D) „Gesellschaftlicher Bereich“ ist zu kodieren, falls konkrete Problemsituation glaubwürdig für (zukünftige) Handlungen im öffentlichen Leben (z. B. im Rahmen von Behördengängen, bei civil engagement, usw.) für Jugendliche ist</p> <p>A1 → nicht kodiert A2 → nicht kodiert A3 → kodiert, da diese Problemsituation aus eigener Erfahrung heraus glaubwürdig ist</p>
	Beruflicher Bereich	<p>Die konkret gegebene Problemsituation hat eine lebendige Stellung in der beruflichen Welt, in die die Jugendlichen hineinwachsen</p> <p>Von den meisten Jugendlichen kann mit wenigen Sätzen begründet werden, dass die Problemstellung in der beruflichen Gegenwart eine Bedeutung hat oder in der Zukunft erhalten wird oder müsste</p> <p>Konkrete Problemstellung leistet Beitrag zum Erreichen genereller Ziele wie Selbstverantwortung, Mündigkeit etc. in beruflichen Kontexten</p> <p>A1 → nicht kodiert A2 → nicht kodiert A3 → nicht kodiert</p>
	Gesellschaftlicher Bereich	<p>Von den meisten Jugendlichen kann mit wenigen Sätzen begründet werden, dass die Problemstellung im Leben als Bürger/in (Pflichten, Rechte, Behörden etc.) der Gegenwart eine Bedeutung hat oder in der Zukunft erhalten wird oder müsste</p> <p>Die konkret gegebene Problemstellung leistet Beitrag zum Erreichen genereller Ziele wie Selbstverantwortung, Mündigkeit etc. in beruflichen Kontexten</p> <p>A1 → kodiert, da Problemsituation und Fragestellung entlang eines Dokuments des öffentlichen Lebens (= Lohnauszug) glaubwürdig für die meisten Jugendlichen eine Handlungsfähigkeit authentisch simuliert, die in der Zukunft eine Bedeutung erhalten wird</p> <p>A2 → nicht kodiert A3 → nicht kodiert</p>

Tab. 8 (Fortsetzung)

Kategorie	Subkategorie	Deskriptoren
	<i>Wissenschaftlicher</i> Bereich	Nicht die konkrete Problemstellung, sondern die Problemstellung oder math. Prozeduren verdeutlicht exemplarisch allgemeinere Zusammenhänge, Beziehungen, Gesetzmäßigkeiten, Strukturen, Widersprüche oder Handlungsmöglichkeiten (Techniken, Methoden usw.) A1 → nicht kodiert A2 → kodiert, da das Beispielitem nicht zum Ziel hat, die konkrete Problemsituation mathematisch zu bearbeiten, sondern verallgemeinert eine Technik, hier die Prozentformel, anzuwenden A3 → nicht kodiert

(D) Diskriminierungskriterien

- Ankerbeispiele A1 (kodiert = Beispielitem 1, s. Abb. 2, ist ein Ankerbeispiel)
 Ankerbeispiele A2 (kodiert = Bsp.-item 2, s. Abb. 3 Anhang, ist ein Ankerbeispiel)
 Ankerbeispiele A3 (kodiert = Bsp.-item 3, s. Abb. 4 Anhang, ist ein Ankerbeispiel)

Mathematikarbeit

Rolf sagt: „In der letzten Mathematikarbeit habe ich 48 von 60 Punkten erreicht, das sind der Gesamtpunktzahl.“

Welcher der folgenden Prozentsätze muss eingesetzt werden?

Kreuze an.

12 %

12,5 %

48 %

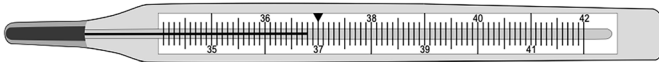
60 %

80 %

Abb. 3 Mathematikarbeit, Beispielim 2 passend zum bisherigen Aufgabenpool. (Quelle: <https://www.iqb.hu-berlin.de/vera/aufgaben/ma1/>)

Fieberthermometer

Die Abbildung zeigt ein Fieberthermometer. Die schwarze dicke Linie zeigt die gemessene Körpertemperatur in Grad Celsius an.



Grafik: © IQB

Teilaufgabe 1

Gib an, wie viel °C die gemessene Körpertemperatur in der Abbildung beträgt.

_____ °C

Abb. 4 Beispielim 3 passend zum bisherigen Aufgabenpool. (Quelle: <https://www.iqb.hu-berlin.de/vera/aufgaben/ma1/>)

Danksagung Die Autoren bedanken sich beim Institut zur Qualitätsentwicklung im Bildungswesen für die Bereitstellung der untersuchten Testaufgaben und Materialien.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., & Weiß, M. (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Leske + Budrich. <https://doi.org/10.1007/978-3-322-83412-6>.
- Blum, W., & Leiß, D. (2005). Modellieren im Unterricht mit der „Tanken“-Aufgabe. *mathematik lehren*, 128, 18–21.
- Blum, W., Roppelt, A., & Müller, M. (2019). Kompetenzstufenmodelle für das Fach Mathematik. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 60–71). Waxmann.
- Burkhardt, H., & Swan, M. (2017). Design and development for large-scale improvement. In G. Kaiser (Hrsg.), *Proceedings of the 13th International Congress on Mathematical Education* (S. 177–200). Springer. https://doi.org/10.1007/978-3-319-62597-3_12.
- Eisewicht, P., & Grenz, T. (2018). Die (Un)Möglichkeit allgemeiner Gütekriterien in der Qualitativen Forschung – Replik auf den Diskussionsanstoß zu „Gütekriterien qualitativer Forschung“ von Jörg Strübing, Stefan Hirschauer, Ruth Ayaß, Uwe Krähnke und Thomas Scheffer. *Zeitschrift für Soziologie*, 47(5), 364–373. <https://doi.org/10.1515/zfsoz-2018-0123>.
- Frenken, L., Libbrecht, P., Becker, B., & Greefrath, G. (2022). Dynamic geometry tasks in standardized assessment—analysis of solution processes and consequences for practice. *International Journal of Mathematical Education in Science and Technology*. <https://doi.org/10.1080/0020739X.2022.2036838>.
- Garrett, L., Huang, L., & Charleton, M. (2016). A framework for authenticity in the mathematics and statistics classroom. *The Mathematics Educator*, 25, 32–55.
- Gebhardt, M., Oelkrug, K., & Tretter, T. (2013). Das mathematische Leistungsspektrum bei Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf in der Sekundarstufe. Ein explorativer Querschnitt der fünften bis neunten Klassenstufe in Münchner Förderschulen. *Empirische Sonderpädagogik*, 5(2), 130–143. <https://doi.org/10.25656/01:8913>.
- Gläser-Zikuda, M., Hagenauer, G., & Stephan, M. (2020). The potential of qualitative content analysis for empirical educational research. *Forum: Qualitative Social Research*, 21(1), 17. <https://doi.org/10.17169/fqs-21.1.3443>.
- Gold, A. (2011). *Lernschwierigkeiten: Ursachen, Diagnostik, Intervention*. Kohlhammer.
- Greefrath, G., & Frenken, L. (2021). Fermi problems in standardized assessment in grade 8. *Quadrante*, 30(1), 52–73. <https://doi.org/10.48489/quadrante.23587>.
- Greefrath, G., Kaiser, G., Blum, W., & Ferri, B.R. (2013). Mathematisches Modellieren – Eine Einführung in theoretische und didaktische Hintergründe. In R. Borromeo Ferri, G. Greefrath & G. Kaiser (Hrsg.), *Mathematisches Modellieren für Schule und Hochschule* (S. 11–37). Springer. https://doi.org/10.1007/978-3-658-01580-0_1.
- Helsper, W., Kelle, H., & Koller, H.-C. (2019). Qualitätskriterien der Begutachtung qualitativer Forschungsvorhaben. *Zeitschrift für Pädagogik*, 62, 738–748.
- Kaiser, G. (2007). Modelling and modelling competencies in school. In C. R. Haines, P.L. Galbraith, W. Blum & S. Khan (Hrsg.), *Mathematical modelling (ICTMA 12): education, engineering and economics* (S. 110–119). Horwood. <https://doi.org/10.1533/9780857099419.3.110>.
- Kaiser, G., & Schwarz, B. (2010). Authentic modelling problems in mathematics education—examples and experiences. *Journal für Mathematik-Didaktik*, 31(1), 51–76. <https://doi.org/10.1007/s13138-010-0001-3>.
- Kaiser, G., Blum, W., Borromeo Ferri, R., & Greefrath, G. (2023). Mathematisches Modellieren. In R. Bruder, A. Büchter, H. Gasteiger, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 399–428). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-662-66604-3_13.
- KMK (Hrsg.). (2004). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 4.12.2003*. Luchterhand.
- KMK (Hrsg.). (2020). Sonderpädagogische Förderung in Schulen 2009 bis 2018. https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/Dok223_SoPae_2018.pdf
- KMK (Hrsg.). (2022). *Bildungsstandards für das Fach Mathematik. Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA) (Beschluss der Kultusministerkonferenz vom 15.10.2004 und vom 04.12.2003, i.d.F. vom 23.06.2022)*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-ESA-MSA-Mathe.pdf

- Knoche, N., & Lind, D. (2004). Bedingungsanalysen mathematischer Leistung: Leistungen in den anderen Domänen, Interesse, Selbstkonzept und Computernutzung. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA 2000* (S. 205–226). VS. https://doi.org/10.1007/978-3-322-80661-1_11.
- Kocaj, A., Haag, N., Weirich, S., Kuhl, P., Pant, H. A., & Stanat, P. (2016). Aspekte der Testgüte bei der Erfassung schulischer Kompetenzen von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf. In V. Moser & B. Lütje-Klose (Hrsg.), *Schulische Inklusion* (Bd. 62, S. 212–234). Beltz Juventa.
- Kölm, J., & Mahler, N. (2019). Kompetenzstufenbesetzungen im Ländervergleich. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 157–168). Waxmann.
- Konsortium ÜGK (2019). *Überprüfung der Grundkompetenzen. Nationaler Bericht der ÜGK 2016: Mathematik 11. Schuljahr*. EDK, SRED. <https://doi.org/10.18747/PHSG-coll3/id/386>.
- Kühn, S. M., & Drike-Noe, C. (2013). Qualität und Vergleichbarkeit durch Bildungsstandards und zentrale Prüfungen? Ein bundesweiter Vergleich von Prüfungsanforderungen im Fach Mathematik zum Erwerb des Mittleren Schulabschlusses. *Zeitschrift für Pädagogik*, 59(6), 912–932.
- Lehmann, R. H., & Hoffmann, E. (Hrsg.). (2009). *BELLA: Berliner Erhebung arbeitsrelevanter Basiskompetenzen von Schülerinnen und Schülern mit Förderbedarf „Lernen“*. Waxmann.
- 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* (S. 149–167). VS. https://doi.org/10.1007/978-3-531-90865-6_9.
- Maaß, K. (2004). *Mathematisches Modellieren im Unterricht. Ergebnisse einer empirischen Studie*. Franzbecker.
- Maaß, K. (2006). What are modelling competencies? *ZDM*, 38(2), 2. <https://doi.org/10.1007/BF02655885>.
- Maaß, K. (2010). Classification scheme for modelling tasks. *Journal für Mathematik-Didaktik*, 31(2), 285–311. <https://doi.org/10.1007/s13138-010-0010-2>.
- Mahler, N., Schipolowski, S., & Weirich, S. (2019a). Anlage, Durchführung und Auswertung des IQB-Bildungstrends 2018. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 99–124). Waxmann.
- Mahler, N., Weirich, S., & Becker, B. (2019b). Auswertung, Trendschätzung und Ergebnisdarstellung. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 125–130). Waxmann.
- Mahler, N., Kölm, J., & Werner, B. (2020). Entwicklung von Mathematiktestaufgaben für Schüler*innen mit einem sonderpädagogischen Förderbedarf im Lernen – Konzeption und empirische Ergebnisse. In C. Gresch, P. Kuhl, M. Grosche, C. Sälzer & P. Stanat (Hrsg.), *Schüler*innen mit sonderpädagogischem Förderbedarf in Schulleistungserhebungen* (S. 109–146). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-27608-9_5.
- Maier, U., Kleinknecht, M., Metz, K., & Bohl, T. (2010). Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben. *Beiträge zur Lehrerbildung*, 28(1), 84–96. <https://doi.org/10.25656/01:13734>.
- Mayring, P. (2010). Qualitative Inhaltsanalyse. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 601–613). VS. https://doi.org/10.1007/978-3-531-92052-8_42.
- Mayring, P. (2020). Qualitative content analysis: demarcation, varieties, developments. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3), 16. <https://doi.org/10.17169/FQS-20.3.3343>.
- 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), 205–230. <https://doi.org/10.1007/s13138-022-00201-1>.
- Moser Opitz, E., Labhart, D., Grob, U., & Prediger, S. (2021). *BASIS-MATH-G 6+ : Gruppentest zur Basisdiagnostik Mathematik für das vierte Quartal der 6. Klasse und das erste Quartal der 7. Klasse*
- Müller, K., Prenzel, M., Sälzer, C., Mang, J., Heine, J.-H., & Gebhardt, M. (2017). Wie schneiden Schülerinnen und Schüler an Sonder- und Förderschulen bei PISA ab? Analysen aus der PISA 2012-Zusatzerhebung zu Jugendlichen mit sonderpädagogischem Förderbedarf. *Unterrichtswissenschaft*, 2, 194–211. <https://doi.org/10.3262/UW1702175>.

- Müller, M., Musilek, M., & Wimmer, C. (2022). Bildungsstandardüberprüfungen in Mathematik auf der Sekundarstufe 1 über den Zeitverlauf 2009–2012–2017. In A. C. George, S. Götz, M. Illetschko & E. Süß-Stepancik (Hrsg.), *Empirische Befunde zu Kompetenzen im Mathematikunterricht der Sekundarstufe I und Folgerungen für die Praxis. Ergänzende Analysen zu den Bildungsstandardüberprüfungen* (S. 55–78). Waxmann. <https://doi.org/10.31244/9783830995586>.
- Mullis, I. V. S., Martin, M. O., & von Davier, M. (Hrsg.). (2021). *TIMSS 2023 Assessment Frameworks*. Boston College, TIMSS & PIRLS.
- Neubrand, M., Jordan, A., Krauss, S., Blum, W., & Löwen, K. (2013). Task analysis in COACTIV: examining the potential for cognitive activation in German mathematics classrooms. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (S. 125–144). Springer. https://doi.org/10.1007/978-1-4614-5149-5_7.
- Niss, M. (1992). Applications and modelling in school mathematics—directions for future development. In I. Wirsup & R. Streit (Hrsg.), *Development in school mathematics education around the world* (Bd. 3, S. 346–361). NCTM.
- Niss, M., Blum, W., & Galbraith, P. L. (2007). Introduction. In W. Blum, P. L. Galbraith, H.-W. Henn & M. Niss (Hrsg.), *Modelling and Applications in Mathematics Education. The 14th ICMI study* (S. 3–32). Springer. https://doi.org/10.1007/978-0-387-29822-1_1.
- OECD (Hrsg.). (2001). *Knowledge and skills for life. First results from the OECD programme for international student assessment*. OECD.
- OECD (2018). PISA 2022 mathematics framework (draft). <https://pisa2022-maths.oecd.org>
- Palm, T. (2007). Features and impact of the authenticity of applied mathematical school tasks. In W. Blum, P. L. Galbraith, H.-W. Henn & M. Niss (Hrsg.), *Modelling and applications in mathematics education. The 14th ICMI Study*, (Bd. 10, S. 201–208). Springer. https://doi.org/10.1007/978-0-387-29822-1_20.
- Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics*, 67(1), 37–58. <https://doi.org/10.1007/s10649-007-9083-3>.
- Pollak, H. O. (1977). The interaction between mathematics and other school subjects (including integrated courses). In H. Athen & H. Kunle (Hrsg.), *Proceedings of the Third International Congress on Mathematical Education* (S. 255–264). Zentralblatt für Didaktik der Mathematik.
- Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E., & Benholz, C. (2015). Sprachkompetenz und Mathematikleistung – Empirische Untersuchung sprachlich bedingter Hürden in den Zentralen Prüfungen 10. *Journal für Mathematik-Didaktik*, 36(1), 77–104. <https://doi.org/10.1007/s13138-015-0074-0>.
- Rammstedt, B. (Hrsg.). (2013). *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich: Ergebnisse von PIAAC 2012*. Waxmann. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-360687>
- Reinhold, F., Reiss, K., Diedrich, J., Hofer, S. I., & Heinze, A. (2019). Mathematische Kompetenz in PISA 2018 – aktueller Stand und Entwicklung. In K. Reiss, M. Weis, E. Klieme & O. Köller (Hrsg.), *PISA 2018: Grundbildung im internationalen Vergleich* (S. 187–209). Waxmann. <https://doi.org/10.31244/9783830991007>.
- Roppelt, A., Blum, W., Pöhlmann, C., Mahler, N., & Greefrath, G. (2019). Beschreibung der untersuchten mathematischen Kompetenzen. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 21–35). Waxmann.
- Ross, N., & Adleff, A.-K. (2022). Anforderungsstufen von Modellierungskomplexität: Transfermöglichkeiten des theoretischen Konstrukts aus dem Aufgabenklassifikationssystem von TEDS-Validierung in Maßnahmen der Lehrkräfteprofessionalisierung. In N. Buchholtz, B. Schwarz & K. Vorhölter (Hrsg.), *Initiationen mathematikdidaktischer Forschung* (S. 341–367). Springer. https://doi.org/10.1007/978-3-658-36766-4_18.
- Rupp, A., & Vock, M. (2007). National educational standards in Germany: methodological challenges for developing and calibrating standards-based tests. In D. Waddington, P. Nentwig & S. Schanze (Hrsg.), *Making it comparable. Standards in science education* (S. 173–198). Waxmann.
- Scheja, B., Schreck, A., & Rott, B. (2023). Prozessbezogene Kompetenzen in zentralen Abschlussarbeiten im Fach Mathematik? – Eine Analyse aller ZP10-Aufgaben Nordrhein-Westfalens von 2007 bis 2019. *Zeitschrift für Erziehungswissenschaft*. <https://doi.org/10.1007/s11618-023-01191-7>.
- Schroeder, J. (2015). *Pädagogik bei Beeinträchtigungen des Lernens*. Kohlhammer.
- Seeber, S. (2009). Ökonomisches Verständnis. In R. H. Lehmann & E. Hoffmann (Hrsg.), *BELLA: Berliner Erhebung arbeitsrelevanter Basiskompetenzen von Schülerinnen und Schülern mit Förderbedarf „Lernen“* (S. 89–118). Waxmann.

- Stanat, P., Schipolowski, S., Mahler, N., Weirich, S., & Henschel, S. (Hrsg.). (2019). *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich*. Waxmann.
- Steinke, I. (1999). *Kriterien qualitativer Forschung: Ansätze zur Bewertung qualitativ-empirischer Sozialforschung*. Juventa.
- Steinke, I. (2019). Gütekriterien Qualitativer Forschung. In U. Flick, E. von Kardorff & I. Steinke (Hrsg.), *Qualitative Forschung – Ein Handbuch* (13. Aufl. S. 319–331). Rowohlt.
- Stern, E. (2003). Früh übt sich – Neuere Ergebnisse aus der LOGIK-Studie zum Lösen mathematischer Textaufgaben. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), *Rechenschwäche: Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie ; ein Handbuch* (S. 116–130). Beltz.
- Südkamp, A., Pohl, S., Hardt, K., Jordan, A.-K., & Duchhardt, C. (2015). Kompetenzmessung in den Bereichen Lesen und Mathematik bei Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf. In P. Kuhl, P. Stanat, B. Lütje-Klose, C. Gresch, H. A. Pant & M. Prenzel (Hrsg.), *Inklusion von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf in Schulleistungserhebungen* (S. 243–272). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-06604-8_9.
- Turner, E. E., Bennett, A. B., Granillo, M., Ponnuru, N., Roth McDuffie, A., Foote, M. Q., Aguirre, J. M., & McVicar, E. (2022). Authenticity of elementary teacher designed and implemented mathematical modeling tasks. *Mathematical Thinking and Learning*. <https://doi.org/10.1080/10986065.2022.2028225>.
- Turner, R., Dossey, J., Blum, W., & Niss, M. (2013). Using mathematical competencies to predict item difficulty in PISA: a MEG study. In M. Prenzel, M. Kobarg, K. Schöps & S. Rönnebeck (Hrsg.), *Research on PISA* (S. 23–37). Springer. https://doi.org/10.1007/978-94-007-4458-5_2.
- UN-BRK (2008). *Übereinkommen der Vereinten Nationen über die Rechte von Menschen mit Behinderung*. Bonn: Bundesministerium für Arbeit und Soziales. Beauftragter der Bundesregierung für die Belange von Menschen mit Behinderungen, Hrsg.
- Vos, P. (2018). "How real people really need mathematics in the real world" - Authenticity in Mathematics Education. *Education Sciences*, 8(4), 195. <https://doi.org/10.3390/educsci8040195>.
- Wember, F. B. (2015). Didaktische Prinzipien und Qualitätssicherung im Förderunterricht. In U. Heimlich & F. B. Wember (Hrsg.), *Didaktik des Unterrichts im Förderschwerpunkt Lernen: Ein Handbuch für Studium und Praxis* (S. 79–95). Kohlhammer.
- Werner, B. (2019). *Mathematik inklusive: Grundriss einer inklusiven Fachdidaktik*. Kohlhammer. <https://doi.org/10.17433/978-3-17-033865-4>.

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.