

Testentwicklung und Skalierung

ÜGK / COFO / VECOF 2024

Schulsprache und Mathematik, 4. Schuljahr

Technischer Bericht

Domenico Angelone [Hrsg.]

Bericht im Auftrag der Konferenz der kantonalen Erziehungsdirektorinnen
und -direktoren (EDK)

u^b

^b
UNIVERSITÄT
BERN

adb
bdt

Aufgabendatenbank der EDK
Banque de données de tâches de la CDIP
Banca dati degli esercizi della CDPE
Banca da datas da pensums da la CDEP

Impressum

Herausgeber:	Domenico Angelone, Geschäftsstelle der Aufgabendatenbank EDK (ADB), c/o Schweizerische Koordinationsstelle für Bildungsforschung (SKBF), Entfelderstrasse 61, 5000 Aarau
Auftraggeberin:	Konferenz der kantonalen Erziehungsdirektorinnen und -direktoren (EDK)
Zitationsvorschlag:	Angelone, Domenico (Hrsg.) (2025). Testentwicklung und Skalierung. ÜGK / VECOF / COFO 2024, Schulsprache und Mathematik, 4. Schuljahr. Technischer Bericht. Geschäftsstelle der Aufgabendatenbank EDK (ADB). https://doi.org/10.48620/90933
Autorinnen und Autoren:	Domenico Angelone, Florian Keller (ADB) Giang Pham (PHSG) Linda Boos-Leutwiler, Miriam Dittmar, Sabine Geiger, Thomas Lindauer (PH FHNW) Nadine Nell-Tuor (PH ZUG) Vincent Capt, Anne-Claire Blanc (HEP VD) Nicodemo Cannavò, Alessandra Moretti (DECS) Stephan Schönenberger (PH TG) Alberto Piatti SUPSI Jean-Luc Dorier (Uni GE)
ÜGK-Webseite:	www.uegk-schweiz.ch
Layout:	Interfaculty Centre for Educational Research (ICER)
Weitere Sprachversionen:	Französisch, Italienisch
Copyright:	Creative Commons: Namensnennung 4.0 International (CC BY 4.0). Das Material unter der Creative-Commons-Lizenz darf unter den folgenden, von den Urheberinnen und Urhebern festgelegten Bedingungen verwendet werden: Sie dürfen das Material in jedem Format weitergeben, vervielfältigen, frei verwenden und weiterverteilen, vorausgesetzt, dass die Herausgebenden und die Nutzungsrechte in angemessener Weise genannt werden.

Inhalt

Impressum.....	3
1 Einleitung	6
2 Nationale Bildungsstandards in der Schulsprache.....	7
2.1 Im Bereich Lesen überprüfte Grundkompetenzen.....	7
Operationalisierung	8
Aufgabenbeispiele.....	11
Zusatztest Leseflüssigkeit	15
2.2 Im Bereich Hören überprüfte Grundkompetenzen.....	17
Operationalisierung	18
Aufgabenbeispiele.....	23
2.3 Aufgabenentwicklung.....	29
Machbarkeitsstudie	29
Internes Review	29
CogLabs	30
Externes Review und Präpilotierung	30
Übersetzung der Aufgaben im Bereich Lesen und Hören	30
Pilotierung der Aufgaben	30
3 Nationale Bildungsstandards in der Mathematik.....	31
3.1 Im Bereich Mathematik überprüfte Grundkompetenzen	32
Operationalisierung	34
Aufgabenbeispiele.....	36
3.2 Aufgabenentwicklung.....	41
Machbarkeitsstudie	41
Internes Review	41
CogLabs	42
Externes Review und Präpilotierung	42
Übersetzung der Aufgaben im Bereich Mathematik.....	42
Pilotierung der Aufgaben	42
4 Testdesign.....	43
5 Skalierung	46
5.1 Itemkalibrierung und Linking	46
5.2 Itemanalyse	47
5.3 Plausible Value Imputation.....	49
6 Standardsetting	51

7	Anhang	54
8	Literaturverzeichnis.....	61

1 Einleitung

Domenico Angelone, Florian Keller und Giang Pham

Mit der interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (Harmos-Konkordat) wurde beschlossen, für die obligatorische Schule nationale Bildungsziele zu entwickeln, einzuführen und regelmässig zu überprüfen (Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat), 2007). Im Jahr 2011 hat die Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) nationale Bildungsziele in den Fächern Schulsprache, Fremdsprachen, Mathematik und Naturwissenschaften verabschiedet (EDK, 2011d, 2011a, 2011b, 2011c). Die nationalen Bildungsziele sind als *Mindeststandards* formuliert und beschreiben Grundkompetenzen, die von *allen* Schülerinnen und Schülern bis zu einer bestimmten Schulstufe erreicht werden sollten. Für die Fächer Schulsprache, Mathematik und Naturwissenschaften wurden Grundkompetenzen definiert, die bis am Ende des 4., 8. und 11. Schuljahrs der obligatorischen Schule zu erreichen sind. Für die Fremdsprachen wurden Grundkompetenzen definiert, die bis am Ende des 8. und 11. Schuljahrs zu erreichen sind, da der Fremdsprachenunterricht in der Regel erst nach dem 4. Schuljahr einsetzt. Die nationalen Bildungsstandards sind als Zielvorgaben in die sprachregionalen Lehrpläne – Lehrplan 21, Plan d'études romand und Piano di studio della scuola dell'obbligo ticinese – eingeflossen.

Mit den Erhebungen «Überprüfung des Erreichens der Grundkompetenzen (ÜGK)» wird im Rahmen des nationalen Bildungsmonitorings anhand standardisierter, computerbasierter Kompetenztests schweizweit überprüft, inwieweit die nationalen Bildungsstandards auf bestimmten Schulstufen erreicht worden sind. Ziel der ÜGK ist die Evaluation von Leistungen auf der Ebene des Bildungssystems und nicht die Beurteilung von Schulen, Lehrpersonen oder Schülerinnen und Schülern (EDK, 2014).

Im Frühjahr 2024 wurde das Erreichen der Grundkompetenzen in der Schulsprache (Deutsch, Französisch und Italienisch) und in der Mathematik im 4. Schuljahr anhand tabletbasierter Kompetenztests überprüft. In der Schulsprache wurden die Kompetenzbereiche Lesen und Hören und in der Mathematik die Kompetenzbereiche Zahl und Variable sowie Form und Raum überprüft.

Tabelle 1-1: Überprüfte Fach- und Kompetenzbereiche ÜGK 2024, 4. Schuljahr

Fachbereiche	Kompetenzbereiche
Schulsprache	Lesen Hören
Mathematik	Zahl und Variable Form und Raum

Der vorliegende Bericht beschreibt die technischen Grundlagen der Testentwicklung und Skalierung der Leistungsdaten.

2 Nationale Bildungsstandards in der Schulsprache

Linda Boos-Leutwiler, Nadine Nell-Tuor, Miriam Dittmar, Sabine Geiger, Thomas Lindauer, Vincent Capt, Anne-Claire Blanc, Alessandra Moretti, Nicodemo Cannavò, Domenico Angelone und Florian Keller

Die nationalen Bildungsziele im Fachbereich Schulsprache basieren auf einem handlungsorientierten Verständnis von Sprachkompetenz: Gemeint ist demnach «die Fähigkeit, mit Sprache situationsangemessen zu handeln» (EDK, 2011d, S. 5). Sprachhandlungen umfassen dabei rezeptive wie produktive Kompetenzen und werden medial mündlich oder schriftlich realisiert. Daraus ergeben sich die vier Kompetenzbereiche Lesen, Hören, Schreiben und Sprechen. Zudem gehören auch die beiden Bereiche Orthografie und Grammatik zum schulsprachlichen Unterricht, wobei der Bereich Grammatik im 4. Schuljahr noch wenig ausgebaut und der Bereich Orthografie im 4. Schuljahr im Bereich Schreiben integriert ist. Aus diesem Grund sind Orthografie und Grammatik erst ab Ende des 2. Zyklus in den Bildungsstandards aufgeführt (EDK, 2011d). Lesen, Hören, Schreiben und Sprechen werden im schulsprachlichen Unterricht gezielt gefördert, liegen gleichzeitig aber jeglichem schulischen Lernen zugrunde. Somit sind sie für das Lernen in allen Fächern von zentraler Bedeutung.

Im Rahmen der ÜGK 2024 werden die beiden Kompetenzbereiche Lesen und Hören am Ende des 1. Zyklus getestet. In diesem Zyklus wird die Basis des schulischen Kompetenzerwerbs im Bereich Lesen und Hören gelegt: Die Kinder erwerben schulsprachliche Grundkompetenzen, welche zentral sind für die weitere Schullaufbahn, da sie – wie ausgeführt – für das Lernen in allen Fächern relevant sind.

2.1 Im Bereich Lesen überprüfte Grundkompetenzen

In den von der EDK verabschiedeten Bildungsstandards ist mit Lesen die Fähigkeit gemeint,

Texte unterschiedlicher Textsorten (literarische Texte und Sachtexte, kontinuierliche und diskontinuierliche Texte usw.) sinngemäss auf verschiedenen Ebenen verstehen zu können, Texte als Ausdruck einer Kultur wahrzunehmen und, wenn angemessen, als ästhetisches Produkt geniessen zu können sowie das Gelesene in der eigenen Vorstellung ausgestalten, über das Lesen nachdenken und sich mit anderen darüber austauschen zu können. (EDK, 2011d, S. 7)

Ein solch weites Verständnis von Lesekompetenz basiert auf den Modellen der Lesesozialisationsforschung: Diese unterstreichen die identitätsbildende Wirkung des Lesens, die motivational-emotionale Beteiligung der Leserinnen und Leser an der Bedeutungskonstruktion und schliessen die Reflexion und Anschlusskommunikation über das Gelesene ein (Giasson, 2011; Hurrelmann, 2007; Rosebrock, 2012). Die Grundkompetenzen nehmen die Prozessebene des Lesens in den Blick und beziehen sich damit auf einen engeren Kompetenzbegriff, wie er in der kognitionspsychologischen Forschung fokussiert wird: Leserinnen und Leser entnehmen aus Wörtern und Sätzen Sinn, stellen lokale und globale Kohärenz her, bringen das Gelesene mit dem eigenen

Vorwissen in Verbindung und konstruieren so ein Situationsmodell (Giasson, 1990; Richter & Christmann, 2009; van Dijk & Kintsch, 1983).

Die Grundkompetenzen am Ende des 4. Schuljahres beschreiben, welche Fähigkeiten alle Schülerinnen und Schüler im Bereich Lesen am Ende des Zyklus 1 erreichen sollen:

«Die Schülerinnen und Schüler können die Beziehungen zwischen Lauten und Buchstaben herstellen (Graphem-Phonem-Korrespondenzen) sowie den Sinn von Wörtern und Sätzen verstehen.

Die Schülerinnen und Schüler können kurze Texte verschiedener Art verstehen, wenn Thema, Textaufbau und Wortschatz vertraut sind, z.B. Bilderbücher, Märchen, einfache Anleitungen und Sachtexte.

Insbesondere können sie, wenn sie mithilfe gezielter Fragen dazu angeleitet werden, einen Text als Ganzes verstehen, Informationen erfassen sowie naheliegende Informationen aus dem Kontext ableiten sowie Text und Bild zueinander in Beziehung setzen.» (EDK, 2011d, S. 12)

Wie gut ein Text oder Satz verstanden wird, hängt eng mit textseitigen Merkmalen sowie Aufgabeneigenschaften zusammen. Diese sind in den Bildungsstandards wie folgt festgelegt:

«Der Wortschatz ist geläufig und schulnah, einzelne wenig vertraute Wörter sind aus dem Kontext erschliessbar oder erklärt.

Der Text ist sowohl inhaltlich als auch formal (Layout) übersichtlich strukturiert und kann von Bildern und anderen grafischen Elementen begleitet sein.

Die Frage- und Antwortformate sind einfach. Die Fragen haben einen klaren Bezug zur Textstelle. Die Beantwortung offener Fragen ist mit einem einfachen Wortschatz möglich.» (EDK, 2011d, S. 12)

Operationalisierung

Im Kompetenzbereich Lesen werden im Rahmen der ÜGK 2024 die beiden Dimensionen Textverstehen und Satzverstehen überprüft.

Das Lesen einzelner Sätze stellt dabei andere Anforderungen an die Lesekompetenz als die Lektüre ganzer Texte und kann nicht lediglich als eine im Textverstehen integrierte Kompetenz oder Vorläuferfertigkeit gehandelt werden: Während die Textmenge bei Sätzen zwar gemeinhin geringer ausfällt als bei Texten, muss aber die propositionale Bedeutung des Satzes ohne Kontext aufgebaut werden. Dies bedingt u. a. sehr genaues Lesen, etwa wenn nach räumlichen Relationen zwischen Objekten gefragt wird.

Zur Testung des Lesens wurden die folgenden geschlossenen Aufgabenformate eingesetzt (weitere Präzisierungen siehe weiter unten):

- Multiple Choice text (Wahl der richtigen Textantwort aus mehreren Optionen)
- Multiple Choice image (Wahl richtigen Bildes aus mehreren Optionen)
- Drag & Drop (Zuordnung von Texten zu Bildern)

- Image Map (Antippen des richtigen Bildbereichs)

Textverstehen: Kinder sollten am Ende des 4. Schuljahres «z.B. Bilderbücher, Märchen, einfache Anleitungen und Sachtexte» (EDK, 2011d, S. 12) verstehen können, so dass in der ÜGK 2024, auch in Übereinstimmung mit dem Bereich Hören, das Verstehen narrativer, instruktiver und informativer Texte überprüft wird. Die Stimuli wurden mit Blick auf die Grundkompetenzen anhand der folgenden Kriterien ausgewählt:

- thematische Relevanz für Kinder im Alter von 8 Jahren
- schulnaher und geläufiger Wortschatz
- einzelne wenig vertraute Wörter sind aus dem Kontext erschliessbar oder erklärt
- einfache Syntax
- Vermeidung allfälliger Biases (Geschlecht, Herkunft, Sprachregion)
- inhaltlich und formal übersichtliche Strukturierung
- geeignet für eine Übersetzung in die anderen Landessprachen

Jedem Stimulus ist ein Titel vorangestellt ausser wenn in einem Item nach einem möglichen Titel gefragt wird. Die Texte umfassen max. 400 Zeichen. Zudem können die Stimuli von einem Bild begleitet werden, welches das Verstehen entlastet.

Mit Bezug zu den nationalen Bildungsstandards und analog zur ÜGK 2017 HarmoS 8. (Konsortium ÜGK, 2019) wurden zu jedem Stimulus Items entwickelt, die

- das Erfassen expliziter Informationen im Text,
- das Ableiten impliziter Informationen und
- das globale Verstehen, das heisst das Verstehen des Texts als Ganzes, prüfen.

Bei der Mehrheit der Items wurde auf das Multiple-Choice-Format zurückgegriffen: Bei 38 Items muss die richtige Textantwort und bei 6 Items das korrekte Bild ausgewählt werden. Bei einem Item wird das Drag&Drop-Format eingesetzt.

Satzverstehen: Kinder sollten am Ende des 4. Schuljahres «den Sinn von Wörtern und Sätzen verstehen» (EDK, 2011d, S. 12) können. Zur Überprüfung des Satzverstehens wurden die folgenden Aufgabenformate unter Einbezug von Bildern eingesetzt:

- Image Map: Der Stimulussatz enthält eine Aufforderung, welche das Kind dazu veranlasst, auf einen bestimmten Bildbereich zu tippen.
- Multiple Choice: Als Stimulus dient ein Bild, zu welchem das Kind den passenden Satz antippt, oder als Stimulus dient ein Satz, zu welchem das Kind das passende Bild wählt.
- Drag & Drop: Mehrere Sätze werden dem jeweils passenden Bild zugeordnet.

Die Entwicklung der Aufgaben zum Satzverstehen orientiert sich an denselben Kriterien bzgl. Wortschatz, Syntax, Übersetzbarkeit etc. wie die Aufgaben zum Textverstehen.

Zur Überprüfung der Grundkompetenzen im Bereich Lesen in der Schulsprache wurden in der Dimension Textverstehen 20 Lesetexte (Stimuli) mit insgesamt 36 Items (1-3 pro Stimuli) und in der Dimension Satzverstehen 15 Items eingesetzt (vgl. Tabelle 2-1).

Tabelle 2-1: Anzahl Lesetexte und Items des Kompetenztests zum Lesen in der Schulsprache

Dimension	Textsorte	Verstehensleistung			Total
		Explizite Informationen	Implizite Informationen	Globale Informationen	
Textverstehen	Informative Texte (N = 6)	4	6	2	12
	Instruktive Texte (N = 6)	1	4	3	8
	Narrative Texte (N = 8)	9	5	2	16
Satzverstehen					15
Total					51

In der ÜGK 2024 im 4. Schuljahr wird darauf verzichtet, die Herstellung von Graphem-Phonem-Korrespondenzen gesondert zu testen. Dies liegt einerseits an den Rahmenbedingungen der ÜGK: Um zu überprüfen, ob die Schülerinnen und Schüler die Beziehung zwischen Buchstaben und Lauten herstellen können, müsste man ihnen Pseudowörter präsentieren, die sie laut vorlesen müssten und die dann mit einer Normalverschriftung verglichen würden (Graphem → Phonem). Ein solches Setting ist in einem tabletbasierten Large-Scale-Assessment wie der ÜGK nicht möglich. Andererseits handelt es sich bei dieser sogenannten phonologischen Rekodierung um eine hierarchieniedrige Lesekompetenz, welche als Voraussetzung für das verstehende Lesen dient (Morais et al., 2003; Richter & Christmann, 2009) und somit in den Aufgaben zur Überprüfung des Verstehens von Sätzen und Texten inkludiert ist: Schülerinnen und Schüler, welche die Beziehungen zwischen Buchstaben und Lauten nicht herstellen können, können auch keine propositionale Textrepräsentation aufbauen und erreichen die Grundkompetenzen «Sätze und Texte verstehen» somit nicht.

Aufgabenbeispiele

Die folgenden Abbildungen 2-1 bis 2-4 zeigen Aufgabenbeispiele, die zur Überprüfung der Grundkompetenzen im Bereich *Lesen* eingesetzt wurden. Schülerinnen und Schüler, welche die Grundkompetenzen erreichen, können diese Aufgaben in der Regel richtig lösen.

Abbildung 2-1: Aufgabenbeispiel zum Satzverstehen

SL04_105_001

Welcher Satz passt zu diesem Bild? Tipp die richtige Antwort an.



Das Kind malt ...

- ☐ einen Tisch.
- ☐ einen Engel.
- ☐ einen Stift.



Fachdidaktischer Kommentar

Als Stimulus dient in diesem Item ein Foto, auf welchem ein Kind einen Engel malt. Die Schülerinnen und Schüler müssen den Satz richtig vervollständigen, sodass er zum Bild passt. Die Distraktoren «Tisch» und «Stift» sind zwar beide auf dem Foto zu sehen, sind im Satzkontext aber eindeutig falsch (das Kind malt auf einem Tisch und nicht einen Tisch, es malt mit einem Stift und nicht einen Stift). Neben Wortschatzwissen werden mit diesem Item also auch syntaktische Kenntnisse getestet, wodurch ein umfassendes Satzverstehen auf einfachem Niveau überprüfbar wird.

Lösung

Das Kind malt einen Engel

Abbildung 2-2: Aufgabenbeispiel zu narrativem Text mit expliziter Verstehensleistung

SL04_305_002

Tipp die richtige Antwort an.

Das Kind, das lachte

Es war einmal ein Kind, das immer lachte. Eines Tages ging es mit seinem Grossvater spazieren. Da begann es zu regnen. Sie hatten keinen Regenschirm dabei. Deshalb stellten sie sich mit anderen Leuten unter ein Vordach.

Die Leute waren wütend. Nur das Kind lachte.

Mit wem stellt sich das Kind unter das Vordach?

- ☐ Alleine.
- ☐ Mit dem Grossvater.
- ☐ Mit der Mutter.



Fachdidaktischer Kommentar

Bei diesem Stimulus handelt es sich um eine kurze Erzählung über ein Kind, das mit seinem Grossvater spazieren geht. Der Wortschatz ist einfach und nahe an der Lebenswelt der Schülerinnen und Schüler. Der Stimulustext hat zwar eine Pointe, die nicht leicht zu verstehen ist, das Verstehen dieser Pointe am Ende des Textes ist aber zur Beantwortung dieses Items nicht erforderlich.

Das Item erfordert eine explizite Verstehensleistung: Es wird danach gefragt, mit wem sich das Kind unter das Vordach stellt, was mit der Verknüpfung der beiden Textstellen «Eines Tages ging es mit seinem Grossvater spazieren.» und «Deshalb stellten sie sich mit anderen Leuten unter das Vordach.» beantwortet werden kann. Dazu müssen die Schülerinnen und Schüler in der Lage sein, das Pronomen «sie» in der zweiten Textstelle richtig auf das Kind und den Grossvater in der ersten Textstelle zu beziehen. Mit dem Item wird das Herstellen einfacher pronominaler Bezüge über Sätze hinweg überprüft.

Lösung

Mit dem Grossvater.

Abbildung 2-3: Aufgabenbeispiel zu informativem Text mit impliziter Verstehensleistung

SL04_109_001

Lies den Text. Tipp dann die richtige Antwort an.

Wie bewegt sich ein Regenwurm?

Regenwürmer haben keine Beine und keine Knochen. Sie können sich aber trotzdem bewegen. Wie geht das?

Auf der Haut hat der Regenwurm viele kleine Borsten. Er kann diese Borsten in verschiedene Richtungen bewegen. So kann er vorwärts und rückwärts kriechen.

Was braucht der Regenwurm, um sich zu bewegen?

- ☐ Beine.
- ☐ Knochen.
- ☐ Borsten.



Fachdidaktischer Kommentar

Bei diesem Stimulus handelt es sich um einen kurzen Sachtext zum Thema Regenwürmer, welcher in einfach strukturierten Sätzen die Bewegung der Regenwürmer beschreibt. Sachtexte werden schulisch in bedeutendem Masse zur Wissensvermittlung eingesetzt und das Verstehen solcher Texte hängt eng mit dem schulischen Lernerfolg von Schülerinnen und Schülern zusammen (Baumert & Schümer, 2001; Rincke, 2010).

Die Information, um das Item richtig zu beantworten, nämlich dass sich Regenwürmer mithilfe ihrer Borsten bewegen, sollte den meisten Schülerinnen und Schülern im 4. Schuljahr unbekannt sein. Somit kann dieses Item nicht mit Weltwissen beantwortet werden, sondern die korrekte Antwort muss aus dem Text erschlossen werden. Dafür sind Verstehen und Verknüpfen der letzten drei Sätze notwendig. Mit dem Item werden Verstehensleistungen durch lokale Kohärenzherstellung überprüft.

Lösung

Borsten.

Abbildung 2-4: Aufgabenbeispiel zu instruktivem Text mit impliziter Verstehensleistung

SL04_104_001

Lies die Anleitung. Tipp dann die richtige Antwort an.

Einen Stern basteln

Nimm ein Blatt Papier.

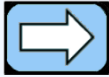
Zeichne einen Stern auf das Papier.

Schneid den Stern dann aus.

Jetzt ist dein Stern fertig!

Was brauchst du, um den Stern zu basteln ?

- ☐ Leim.
- ☐ Schnur.
- ☐ Schere.



Fachdidaktischer Kommentar

Bei diesem Stimulus handelt es sich um einen Instruktionstext in Form einer Anleitung. Aus diesem Grund enthält die Aufgabenformulierung (Zeile ganz oben) die Aufforderung „Lies die Anleitung. Tipp dann die richtige Antwort an.“ Die Schüler und Schülerinnen erhalten dadurch einen Hinweis auf die Textsorte und eine explizite Aufforderung, wie das Item zu lösen ist.

Der Stimulus besteht aus für Bastelanleitungen typischen Textbausteinen (auch Textprozeduren genannt, vgl. Feilke & Rezat, 2020), deren zugrundeliegenden Handlungsschemata den Schülern und Schülerinnen bekannt sein sollten: Das benötigte Material wird genannt („Nimm ein Blatt Papier.“), die Leserinnen und Leser werden explizit zu Handlungen aufgefordert (zeichne, schneid), die Handlungen werden in eine zeitliche Reihenfolge gebracht (dann) und der Vorgang des Bastelns wird am Schluss bewertet („Jetzt ist dein Stern fertig!“). Mit dem Item wird überprüft, inwieweit Abfolgen in Instruktionen verstanden werden.

Lösung

Schere

Zusatztest Leseflüssigkeit

In der ÜGK 2024 wird im Fachbereich Schulsprache erstmals die Leseflüssigkeit erhoben. Die Leseflüssigkeit bezeichnet die genaue, automatisierte (und somit schnelle) sowie sinn-konstruierende Fähigkeit zur Textlektüre, welche es erlaubt, einen Text verstehend zu lesen (Kuhn et al., 2010; Rosebrock & Nix, 2006). Die Leseflüssigkeit ist zentral fürs Leseverstehen, denn sie nimmt eine Brückenfunktion zwischen Dekodieren und Verstehen ein (Gold, 2018; Richter & Müller, 2017; Rosebrock et al., 2017; Schindler & Richter, 2018): Nur wer weitgehend automatisiert und ausreichend flüssig liest, ist in der Lage, einen Text zu verstehen.

In der Schulpraxis nimmt die Leseflüssigkeit einen immer grösseren Stellenwert ein: Sie hat Eingang in die neuen, sprachregionalen Lehrpläne genommen, welche den Ausbau derselben über alle drei Zyklen als Kompetenzziel festhalten (CIIP, 2010; D-EDK, 2016; Repubblica e Cantone Ticino, 2015). Auch in der neueren Lesekompetenz-Forschung rückt die Leseflüssigkeit in den Fokus, werden Förderkonzepte entwickelt und nimmt sie Eingang in Large-Scale-Assessments. So testete etwa PISA 2018 erstmals die Leseflüssigkeit bei 15-Jährigen (Becker-Mrotzek et al., 2019). Da es sich bei der Leseflüssigkeit um eine basale Lesefertigkeit handelt, wird sie ab Beginn des Schriftspracherwerbs entwickelt. Eine Überprüfung im 4. Bildungsjahr ist deshalb möglich und sinnvoll, da sich bereits da interindividuelle Unterschiede ausbilden (Schneider, 2017).

In den nationalen Bildungsstandards der EDK wird die Leseflüssigkeit nicht explizit erwähnt. Implizit liegt sie jedoch dem folgenden Kriterium zur Erreichung der Grundkompetenzen zugrunde: «Die Schülerinnen und Schüler können die Beziehung zwischen Lauten und Buchstaben herstellen (Graphem-Phonem-Korrespondenzen) sowie den Sinn von Wörtern und Sätzen verstehen.» Die Herstellung von Graphem-Phonem-Korrespondenzen (= Dekodieren) wie auch das Erkennen von Wörtern (= Automatisieren) muss ausreichend effizient sein, um kognitive Ressourcen zur lokalen und globalen Kohärenzbildung freizugeben und dadurch die Voraussetzung zum Aufbau eines mentalen Modells zu gewährleisten. Richtig zu dekodieren genügt also nicht, sondern führt erst gekoppelt mit der Automatisierung bzw. Lesegeschwindigkeit zu flüssigem, verstehendem Lesen. Insofern ist die Leseflüssigkeit als Voraussetzung des sinnkonstruierenden Lesens Teil der Grundkompetenzen.

Die Leseflüssigkeit wird in Anlehnung an PISA 2018 und das Salzburger Lesescreening SLS 1–4 mit der Sentence Verification Technique gemessen (Mayringer & Wimmer, 2003; Reiss et al., 2019): Den Schülerinnen und Schülern werden während drei Minuten nacheinander Sätze präsentiert, die dahingehend beurteilt werden müssen, ob sie stimmen oder nicht.

Die Sätze entsprechen dem Weltwissen von Schülerinnen und Schülern im 4. Schuljahr und umfassen 3 bis 6 Wörter. Gemessen wird, wie viele Sätze die Schülerinnen und Schüler pro Minute richtig beurteilen können, das heisst, neben dem genauen und automatisierten Lesen wird auch die Verstehenskomponente berücksichtigt.

Die Schülerinnen und Schüler werden dazu angehalten, während des 3-minütigen Tests so viele Sätze wie möglich zu bearbeiten. Vorbereitend lösen sie zwei Beispielaufgaben. Sobald die Kinder «Ja» oder «Nein» antippen, erscheint automatisch der nächste Satz.

Abbildung 2-5: Aufgabenbeispiel Leseflüssigkeit

Satz	Ja	Nein
Wasser kann man trinken.	Ja	Nein
Die Sonne ist dreieckig.	Ja	Nein

Da die Leseflüssigkeit in den nationalen Bildungsstandards nicht explizit genannt wird und es (noch) keine Anhaltspunkte dazu gibt, wie flüssig ein Kind am Ende des 4. Schuljahres lesen sollte, damit es die Grundkompetenzen erreicht, werden die Ergebnisse zur Leseflüssigkeit gesondert publiziert.

2.2 Im Bereich Hören überprüfte Grundkompetenzen

In den nationalen Bildungsstandards ist Hören definiert als

«Fähigkeit, Hörtexte unterschiedlicher Art (z.B. Radiosendungen, Bahnhoftsansagen, Hörbücher, Anweisungen der Lehrperson, Erzählungen, Debatten) sinngemäss auf verschiedenen Ebenen – in einfachen Verstehens- sowie in Gesprächssituationen – verstehen zu können, das Gehörte als Ausdruck einer Kultur wahrzunehmen, damit sein Wissen zu erweitern sowie das Gehörte in der eigenen Vorstellung auszugestalten, über das Gehörte nachzudenken und sich mit anderen darüber austauschen zu können.» (EDK, 2011d, S. 7)

Im Fokus steht demnach die Verarbeitung auditiver Informationen, wie sie im Modell der kognitiven Informationsverarbeitung von Imhof (2010) erfasst ist (ferner Gschwend, 2014). In den Grundkompetenzen kommt zum Ausdruck, dass Hören in zwei verschiedenen Grundkonstellationen vorkommt: in einfachen Verstehenssituationen, in welchen die Schülerinnen und Schüler ausschliesslich zuhören (monologische Hörsituationen), sowie in Gesprächssituationen, in welchen die Schülerinnen und Schüler auch als Sprechende agieren können (dialogische Hörsituationen). Darüber hinaus kommt in der obigen Beschreibung zum Ausdruck, dass Zuhörkompetenz facettenreich ist und sowohl sozio-kulturelle wie reflexive Faktoren einschliesst (vgl. Krelle, 2010). Mit Blick auf die Grundkompetenzen, wie sie für das Ende des 4. Schuljahres formuliert sind, zeigt sich eine klare Fokussierung auf das Hören in einfachen Verstehenssituationen:

«Die Schülerinnen und Schüler können kurze Hörtexte verstehen, deren Thema, Struktur und Wortschatz ihnen vertraut sind, z.B. Märchen, Spielregeln, Reportagen für Kinder, kurze Vorträge. Insbesondere sind sie fähig, explizite Informationen zu entnehmen, wenn sie durch zielgerichtete Hinweise oder Fragen dazu angeleitet werden, und sie handelnd umzusetzen und/oder sie mit ihrer Lebenswelt in Verbindung zu setzen. Sie können elementare nonverbale Signale deuten (z.B. vertraute Gesten).» (EDK, 2011d, S. 11)

Spezifiziert werden weiter die Eigenschaften von Texten und Aufgaben im Bereich Hören: So wird u. a. festgehalten, dass die Texte sowohl monologisch als auch dialogisch sein können, dass der Wortschatz alltagsnah, vertraut und sehr gebräuchlich ist, die Hörtexte kurz und einfach strukturiert sind, die Thematik vertraut und die Aussprache deutlich ist (Thomas et al., 2013, S. 11)

Operationalisierung

Die Grundkompetenzen für das Ende des 4. Schuljahres im Kompetenzbereich Hören fokussieren das Textverstehen in monologischen und dialogischen Verstehenssituationen. Bei der Testung kamen ausschliesslich Multiple Choice-Aufgaben zum Einsatz. Als Stimuli wurden gemäss Kompetenzbeschreibung und in Übereinstimmung mit der Aufgabenentwicklung im Bereich Lesen narrative (z.B. Märchen), informative (z.B. Sachvorträge, Radioreportagen) und instruktive Texte (z.B. Spielanleitungen) gewählt. Dabei wurde darauf geachtet, dass die Themen aus der Erfahrungswelt der Kinder stammen und dass unter den durchgehend fiktionalen Figuren sowohl wirklichkeitsnahe (Verkäufer, Radioreporterin) wie auch phantastische Figuren (sprechende Tiere, Märchenkönig) sind. Zugleich sollten die Texte so beschaffen sein, dass die Lernenden etwas Neues erfahren und die Items nur aufgrund des Textes, nicht aufgrund von Vorwissen beantworten können (somit kam ein Märchen wie «Dornröschen» oder ein Sachtext mit gängigen Informationen über die Planeten nicht infrage).

Die Hörtexte sind kurz. Ihre Dauer beträgt max. 3 Minuten. Längere Stimuli (3–4 Minuten) und auch solche, welche unterschiedliche Aspekte betonen, wurden in zwei Teile aufgeteilt und je mit einer Kontextualisierung bzw. einem Hinweis für die Aufmerksamkeitsfokussierung versehen. So ist bei diesen kein detailliertes Verstehen notwendig, bei welchem die Gedächtnisleistung einen grossen Einfluss auf das Testergebnis hätte.

Unterschieden werden die Texte danach, ob sie monologisch oder dialogisch sind. Bei dialogischen Texten wird zudem vermerkt, ob die Dialoge von einem Sprecher oder einer Sprecherin gesprochen sind oder von mehreren Sprechenden.

Bei den Texten handelt es sich um für die Testentwicklung verfasste Texte, also um keine in einer natürlichen Situation aufgenommenen Gespräche. Dies liegt darin begründet, dass die Texte in drei verschiedenen Sprachen vorliegen sollen und sich daher zur Übersetzung in allen Sprachen eignen bzw. in allen Sprachversionen vergleichbar sein müssen. Es entspricht auch der gängigen Praxis (VERA, HV-Tests in Schulen). Eine Ausnahme stellt hier die BIFIE-Aufgabenentwicklung dar (Thomas et al., 2013, S. 14), welche sich allerdings auf *eine* Sprache beschränkt.

In standardisierten, tabletbasierten Tests ist es nicht möglich, dass die Schülerinnen und Schüler Nachfragen stellen. Daher wurden die Stimuli so gewählt, dass sie Hörsituationen wiedergeben, die auch in einer natürlichen Situation keine Nachfragen und Wiederholungen erlauben, etwa eine Radiosendung oder ein medienvermitteltes Märchen (Behrens, 2010; Krelle & Prengel, 2014). Zu den Stimuli wurden ein bis drei Items formuliert. Ein Item umfasst eine Frage mit drei Antwortoptionen. Die Items werden danach unterschieden, ob sie explizite oder implizite Informationen abfragen. Erstere werden in den Grundkompetenzen ausdrücklich genannt. Zugleich schliesst das Verstehen der aufgeführten Textsorten immer auch das Verständnis von Implizitem ein, so dass auch dazu Items entwickelt wurden. Es liegen aber weniger implizite als explizite Items vor, da der Fokus gemäss Grundkompetenzbeschreibung auf Letzteren liegt: «Insbesondere sind sie [die Schülerinnen und Schüler am Ende des 4. Schuljahres] fähig, explizite Informationen zu entnehmen, [...]» (EDK, 2011d, S. 11)

Einschränkungen bei der Testung gibt es bezüglich zweier Aspekte: Erstens sind der handelnden Umsetzung aufgrund der tabletbasierten Testung Grenzen gesetzt. Im Sinne einer Annäherung wurden jedoch Items formuliert, die eine Anweisung durch eine fiktive Lehrperson beinhalten. Von den Schülerinnen und Schülern wird verlangt, dass sie die Handlung mental ausführen und dann aus drei Fotos oder Bildern das passende auswählen (Fragen in der Art von «Wie sieht dein Pult jetzt aus?»). Zweitens sprechen technische Gründe gegen die Überprüfung der Deutung nonverbaler Signale: Diese würde das Abspielen von Videos bedingen, das aber aufgrund der zur Verfügung stehenden Software nicht reibungslos gewährleistet werden könnte. Zudem würde die Produktion audio-visueller Stimuli, die über die drei Sprachen hinweg vergleichbar sind, einen erheblichen Aufwand mit sich bringen. Das Audio-Visuelle, das im Alltag der Schülerinnen und Schüler von hoher Relevanz ist (gerade auch in der Unterrichtskommunikation, vgl. Nell-Tuor, 2025). Wie angeführt ist es in den Grundkompetenzen für das Ende des 4. Schuljahres aber auch nur marginal erwähnt. Entsprechend ist dies nur eine geringe Einschränkung im Kontext der Überprüfung der Grundkompetenzen.

Nachfolgend wird das Vorgehen bei den Tonaufnahmen sowie der Aufbau und die Präsentation der Aufgaben beschrieben.

Tonaufnahmen

Bei der Übersetzung der schriftlichen Vorlagen für die Tonaufnahmen wurde auf eine möglichst grosse Vergleichbarkeit zwischen den drei Sprachen geachtet. Anschliessend wurden die Texte von L1-Sprechenden und damit von je unterschiedlichen Sprechenden angefertigt.

Pro Sprache gibt es drei verschiedene Sprechende. Bei diesen handelt es sich um Personen mit besonderem Bezug zum Sprechen, z.B. Leute vom Radio, Sprechwissenschaftler/-innen, Theaterleute, Sprachdidaktiker/-innen. Sie achteten auf eine deutliche Aussprache sowie ein langsames bis mittleres Sprechtempo. Die Aufnahmen fertigten sie in kleinen bis mittelgrossen Räumen ohne störende Hintergrundgeräusche und ohne Hall an.

Das Schneiden der Audios wurde so gehandhabt, dass bei grossen Unterschieden in der Dauer die Sprechgeschwindigkeit angepasst bzw. Sprechpausen angeglichen wurden. Es wurde aber darauf verzichtet, die Audios darüber hinaus möglichst einheitlich machen zu wollen, da dies der Authentizität der Sprechenden und der Spezifika der Einzelsprachen abträglich gewesen wäre.

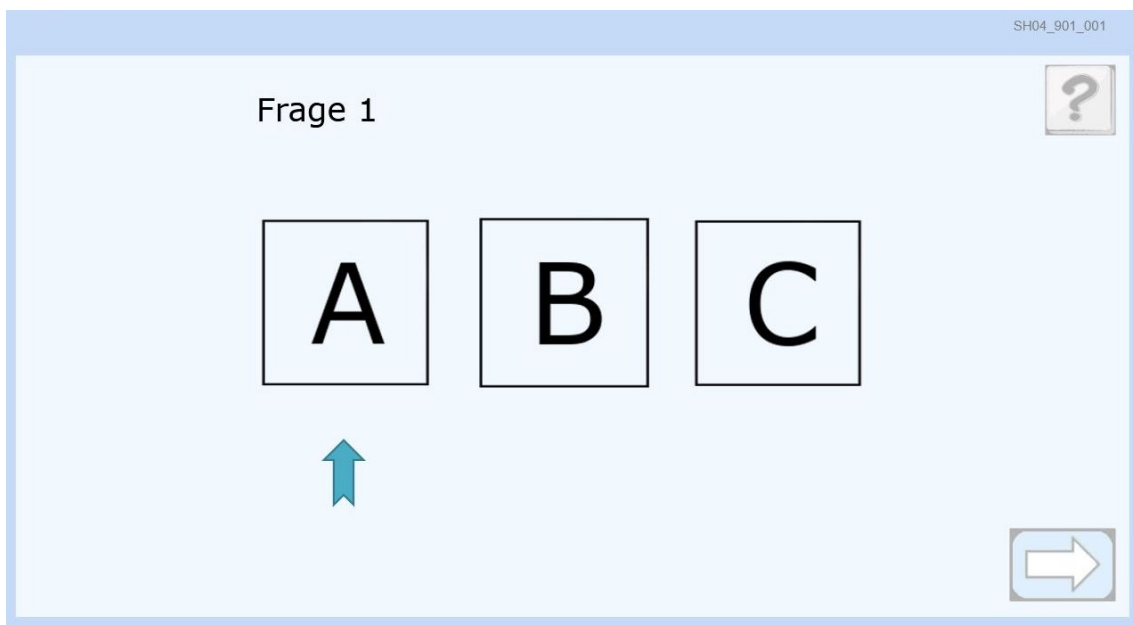
Aufbau und Präsentation der Aufgaben

Zum Anhören der Stimuli tippen die Schülerinnen und Schüler auf ein Ohr-Symbol. Zugleich ist auf dem Screen ein Bild zur Illustration des Hörtextes zu sehen (vgl. Abbildung 2-6). Der Stimulus enthält eine Kontextualisierung, häufig auch mit einem Hinweis, worauf besonders geachtet werden soll (z.B. «Du hörst ein Märchen von einem alten König. Du hörst den ersten Teil. Darin gibt der König Kindern Samen. Höre genau zu, was sie mit diesen Samen machen sollen.»). Dies erlaubt den Schülerinnen und Schülern, eine Zuhörabsicht zu bilden und die Aufmerksamkeit auf das Wesentliche zu fokussieren, zwei zentrale Bedingungen für erfolgreiches Zuhören (Imhof, 2010).

Der Stimulus wird einmalig abgespielt. Eine zwingende Wiederholung würde wertvolle Testzeit brauchen, eine Wiederholung auf freiwilliger Basis wäre der Standardisierung abträglich. Durch die einmalige Präsentation steigt der Einfluss der Aufmerksamkeits- und Konzentrationsfähigkeit. Zugleich werden – insbesondere bei längeren Texten – Informationen, die zur Beantwortung der Fragen notwendig sind, wiederholt. Auch die genannte Kontextualisierung (häufig mit einem Hinweis, worauf zu achten ist) sowie die grundsätzliche Kürze der Hörtexte kann die Aufmerksamkeitsfokussierung und die Konzentration erleichtern. Das messtheoretische Problem von Fluktuationen der Aufmerksamkeit lässt sich aber nicht gänzlich lösen (Thomas et al., 2013).

Anschliessend folgt die akustische Präsentation der Frage(n) und Antwortmöglichkeiten. Die Antwortoptionen sind durch die Buchstaben A, B und C visualisiert, in einigen Fällen auch durch Fotos oder Zeichnungen (vgl. Abbildung 2-6 und Abbildung 2-7). In letzteren Fällen besteht zwischen gehörtem Inhalt und Fotos bzw. Zeichnungen eine redundante Beziehung, d. h. es kommen visuell keine zusätzlichen Informationen dazu.

Abbildung 2-6: Akustisch präsentierte Antwortoptionen durch Buchstaben visualisiert



Zur besseren Orientierung zeigt ein Pfeil auf den Buchstaben A, B oder C, je nach Antwort, die gerade hörbar ist. Im Gegensatz zum Hörtext, welcher nur einmal angehört werden kann, gibt es bei den Items die Möglichkeit, über einen Fragezeichen-Button die Frage und Antworten nochmals abzuspielen.

Abbildung 2-7: Antwortoptionen ausschliesslich als Bilder



Wenn sich die Schülerinnen und Schüler für eine Antwort entschieden haben und diese antippen, wird die entsprechende Antwort automatisch nochmals abgespielt (während des Abspielens ist es aber möglich, auf «Weiter» zu tippen, so dass die Schülerinnen und Schüler nicht zwingend die ganze Antwort nochmals anhören müssen). So können die Schülerinnen und Schüler ihre Eingabe korrigieren, wenn sie merken, dass sie sich im Buchstaben geirrt haben.

Die Aufgaben werden also rein akustisch präsentiert, um eine trennscharfe, valide Messung zu erreichen: Durch die ausschliesslich akustische Präsentation kann vermieden werden, dass Lesekompetenzen hineinspielen. Zugleich ist zu bedenken, dass die Anforderungen an die Zuhörkompetenz der Schülerinnen und Schüler ausgeweitet werden und schwächere Zuhörende noch stärker gefordert sind, wenn auch die Frage und Antworten rein akustisch zu rezipieren sind (Behrens & Weirich, 2019; Chang & Read, 2013).

Wie oben ausgeführt, kann dem aber teilweise dadurch begegnet werden, dass die ausgewählte Antwort wiederholt wird, damit die Schülerinnen und Schüler sie nochmals hören und sich vergewissern können, dass sie sich nicht im Buchstaben geirrt haben.

Zur Überprüfung der Grundkompetenzen im Bereich Hören in der Schulsprache wurden 16 Hörtexte (Stimuli) mit insgesamt 34 Items (1-3 pro Stimuli) eingesetzt (vgl. Tabelle 2-2).

Tabelle 2-2: Anzahl Hörtexte und Items des Kompetenztests zum Hören in der Schulsprache

Textsorte	Verstehensleistung		Total
	Explizite Informationen	Implizite Informationen	
Informative Texte (N = 6)	10	5	15
Instruktive Texte (N = 5)	7	0	7
Narrative Texte (N = 5)	7	5	12
Total	24	10	34

Aufgabenbeispiele

Die folgenden Abbildungen 2-8 bis 2-10 zeigen Aufgabenbeispiele, die zur Überprüfung der Grundkompetenzen im Kompetenzbereich Hören eingesetzt wurden. Unter den Abbildungen ist schriftlich wiedergegeben, was die Schülerinnen und Schüler ausschliesslich hören. Schülerinnen und Schüler, welche die Grundkompetenzen erreichen, können die Aufgaben in der Regel richtig lösen.

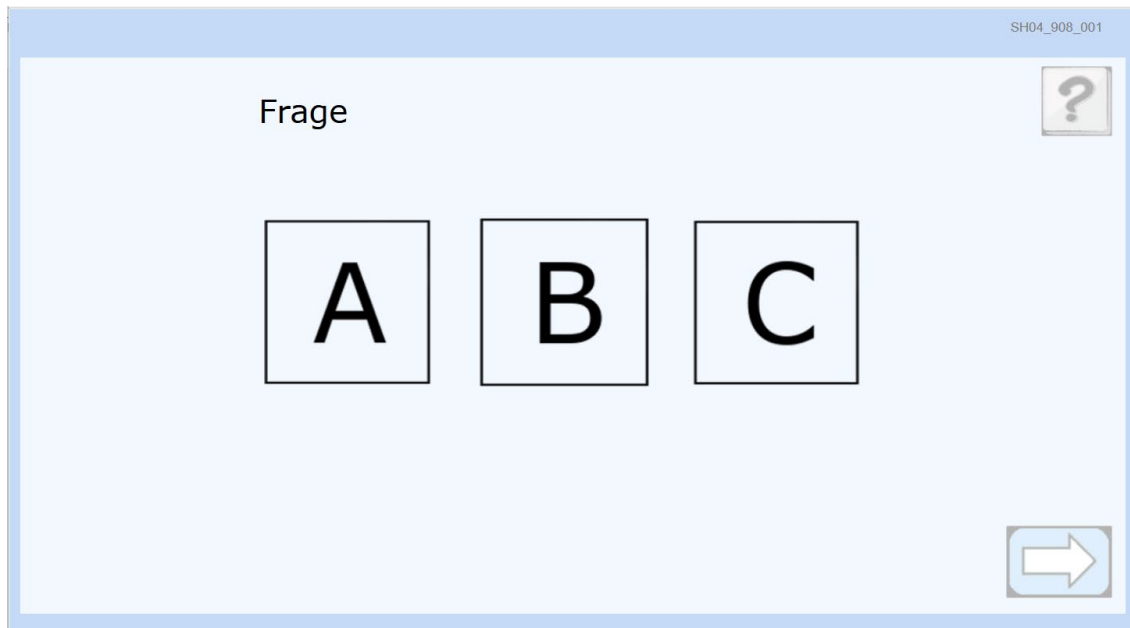
Abbildung 2-8: Aufgabenbeispiel «Märchen neuer König»



Du hörst ein Märchen von einem alten König. Du hörst den ersten Teil. Darin gibt der König Kindern Samen. Höre genau zu, was sie mit diesen Samen machen sollen.

Es war einmal ein alter König. Der hatte keine Kinder und deshalb machte er sich Sorgen, wer nach seinem Tod König oder Königin werden könnte. Es sollte jemand sein, der ehrlich war, also immer die Wahrheit sagte. Aber wie konnte er herausfinden, wer ehrlich war?

Er überlegte lange. Da kam ihm eine Idee. Er liess die Kinder in seinem Land rufen, gab ihnen Samen und sagte: «Nehmt diese Samen und pflanzt sie in einen Blumentopf mit Erde. Gebt ihnen Wasser und schaut ihnen gut zu. Wer am Schluss die schönsten Blumen hat, soll König oder Königin werden.»



Der König gibt den Kindern Samen. Was sollen sie damit machen?

- A. An einem geheimen Ort im Wald vergraben.
- B. In ihren Hosentaschen tragen, bis sie König oder Königin sind.
- C. In einen Blumentopf mit Erde pflanzen und mit Wasser giessen.

Tipp die richtige Antwort an.

Fachdidaktischer Kommentar

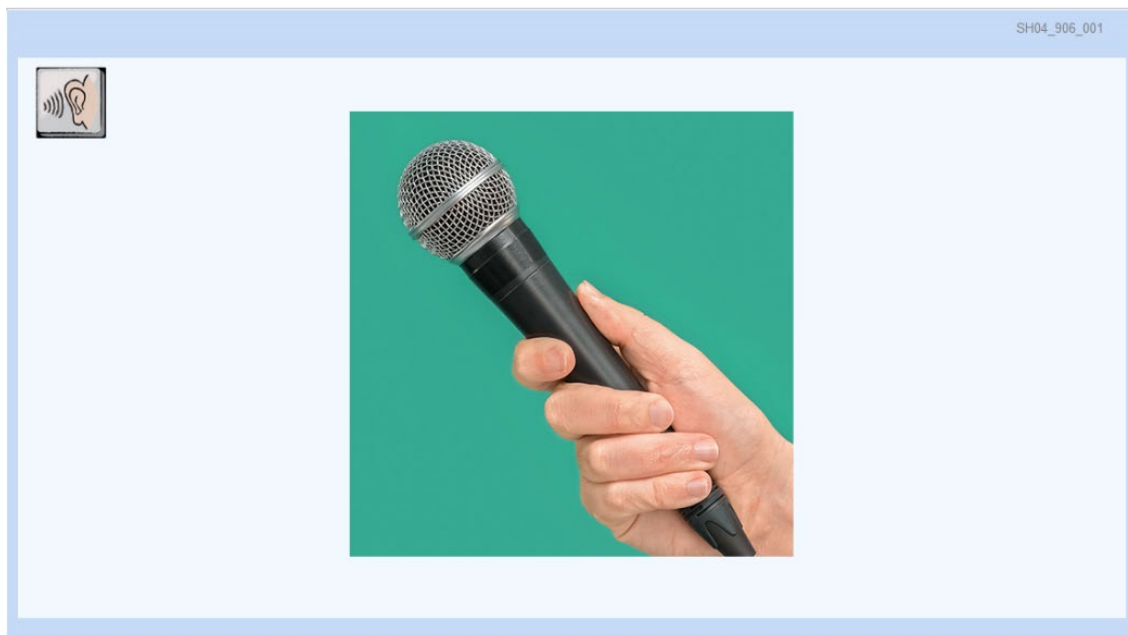
Die Schülerinnen und Schüler sind bereits vor dem Abspielen des Hörtextes darauf hingewiesen worden, dass sie darauf achten müssen, was die Kinder mit den Samen machen sollen (Aufmerksamkeitsfokussierung). Dieser inhaltliche Aspekt ist von hoher Relevanz für das Gesamtverständnis.

Die Aussage in Antwort C wird im Text explizit genannt. Um die Aufgabe richtig zu lösen, müssen die Schülerinnen und Schüler demnach ihre Aufmerksamkeit auf den in Antwort C enthaltenen thematischen Aspekt richten und die Aussage verstehen. Es ist davon auszugehen, dass der Wortschatz des Hörtextes und der Fragestellung den Schülerinnen und Schülern vertraut ist, ebenso die Textsorte Märchen.

Lösung

In einen Blumentopf mit Erde pflanzen und mit Wasser giessen.

Abbildung 2-9: Aufgabenbeispiel «Radioreportage»



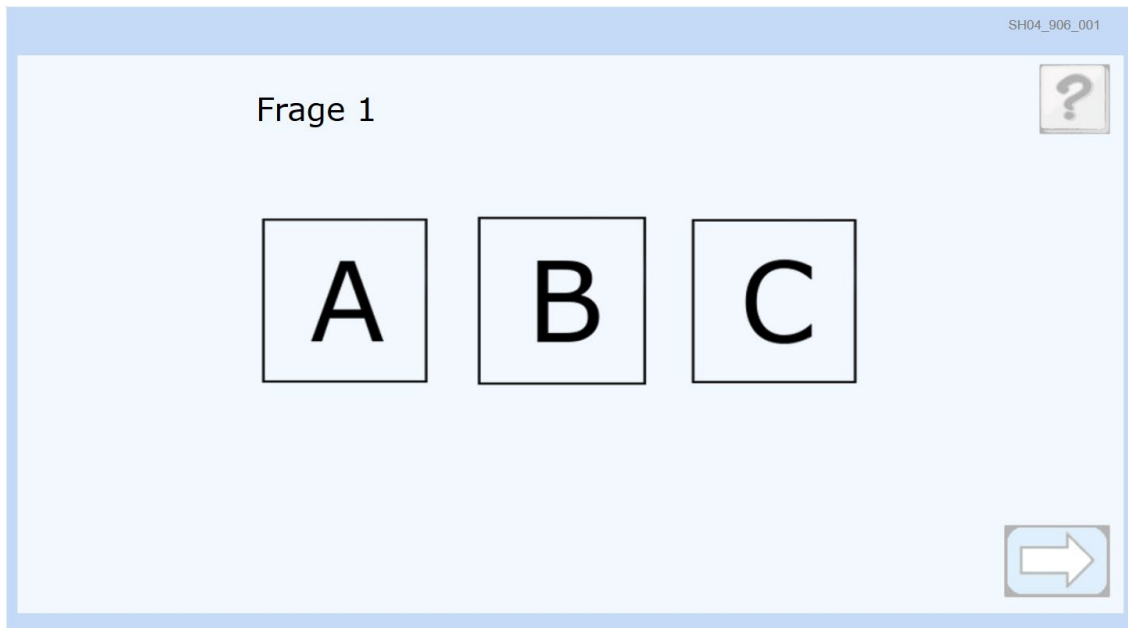
Du hörst einen Ausschnitt aus einer Radiosendung. Darin ist eine Reporterin mit ihrem Mikrofon unterwegs in einer Stadt. Sie spricht mit Leuten, die sie dort antrifft. Hör genau zu, was diese Leute machen und wofür sich die Reporterin interessiert.

Hallo, ich bin Katie. Heute bin ich fürs Radio in der Stadt unterwegs. Mich nimmt wunder, was die Leute hier in ihrer Freizeit tun. Da drüben sehe ich etwas Seltsames. Eine Gruppe junger Frauen klettert über eine Mauer, jetzt springen sie über einen Busch. Weshalb tun sie das? Das möchte ich wissen. Ich gehe mal hin und frage sie.

«Entschuldigung, ich bin vom Radio und habe Sie beobachtet. Ist das Sport, was Sie da machen? Sieht anstrengend aus.»

Frau (hörbar atmend): «Ja genau, sehr streng. Wir trainieren. Wir haben ein grosses Ziel: Wir möchten die ganze Stadt durchqueren, und zwar auf einer graden Linie. Wir gehen also immer gradeaus, wir biegen nie ab. Wenn ein Container im Weg ist, steigen wir darüber. Wenn eine Mauer im Weg ist, klettern wir darüber. Das ist natürlich nicht der einfachste Weg. Aber je mehr wir trainieren, desto leichter geht es.»

Radiomoderatorin: «Interessant. Davon habe ich ja noch nie gehört. Vielen Dank für die Auskunft!»



Was ist das Ziel der Sportlerinnen?

- A. Die Stadt auf einer geraden Linie durchqueren.
- B. Auffallen und im Radio kommen.
- C. Den einfachsten Weg durch die Stadt finden.

Tipp die richtige Antwort an.

Fachdidaktischer Kommentar

Die Schülerinnen und Schüler hören einen Ausschnitt aus einer fiktiven Radiosendung, in welcher die Reporterin in einer Stadt unterwegs ist und eine Gruppe von Frauen anspricht, welche sie dort trifft. Im Verlauf der Radiosendung sagt eine der Sportlerinnen: «Wir haben ein grosses Ziel: Wir möchten die ganze Stadt durchqueren, und zwar auf einer graden Linie.» Damit enthält der Text explizit die Information, welche zur Auswahl von Antwort A notwendig ist.

Da die Reporterin auf die Sportlerinnen zugeht, die gerade am Trainieren sind, gibt es keinerlei Hinweise darauf, dass Antwort B zutrifft. Diese Antwort kann somit eindeutig ausgeschlossen werden.

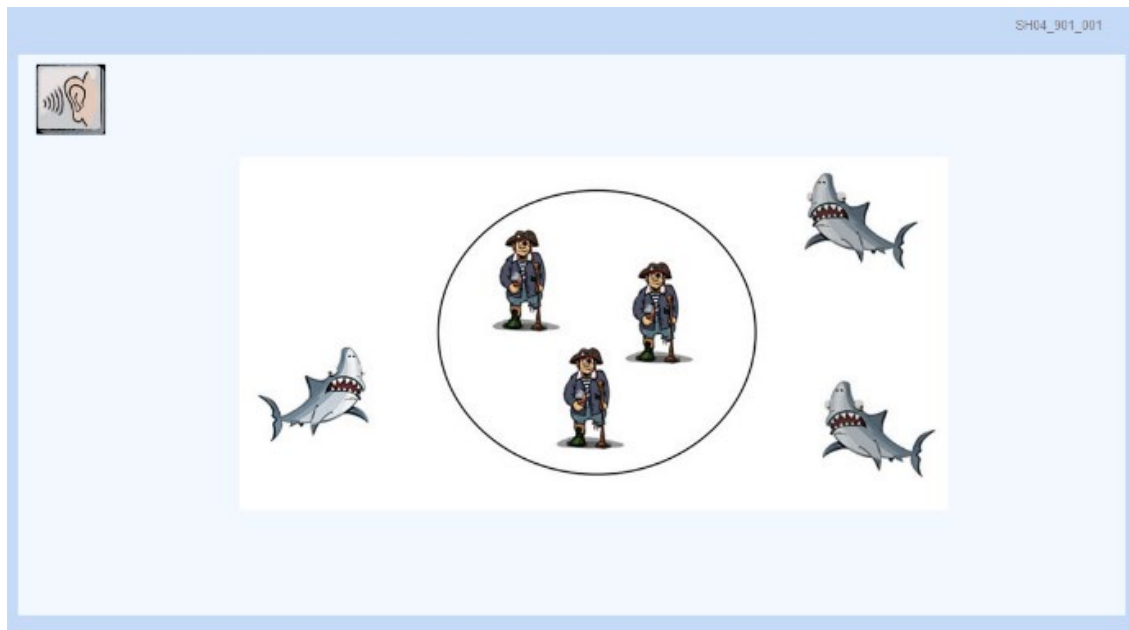
Antwort C nimmt zwar Bezug auf sprachliches Textmaterial (so sagt eine Sportlerin: «Das ist natürlich nicht der einfachste Weg.»), inhaltlich wird aber gerade das Gegenteil ausgesagt. Somit fällt auch Antwort C weg.

Da bei einer Radiosendung in der Regel keine Nachfragen und Wiederholungen möglich sind, entspricht das lebensweltliche Setting dem Testsetting.

Lösung

Die Stadt auf einer geraden Linie durchqueren.

Abbildung 2-10: Aufgabenbeispiel «Piraten»



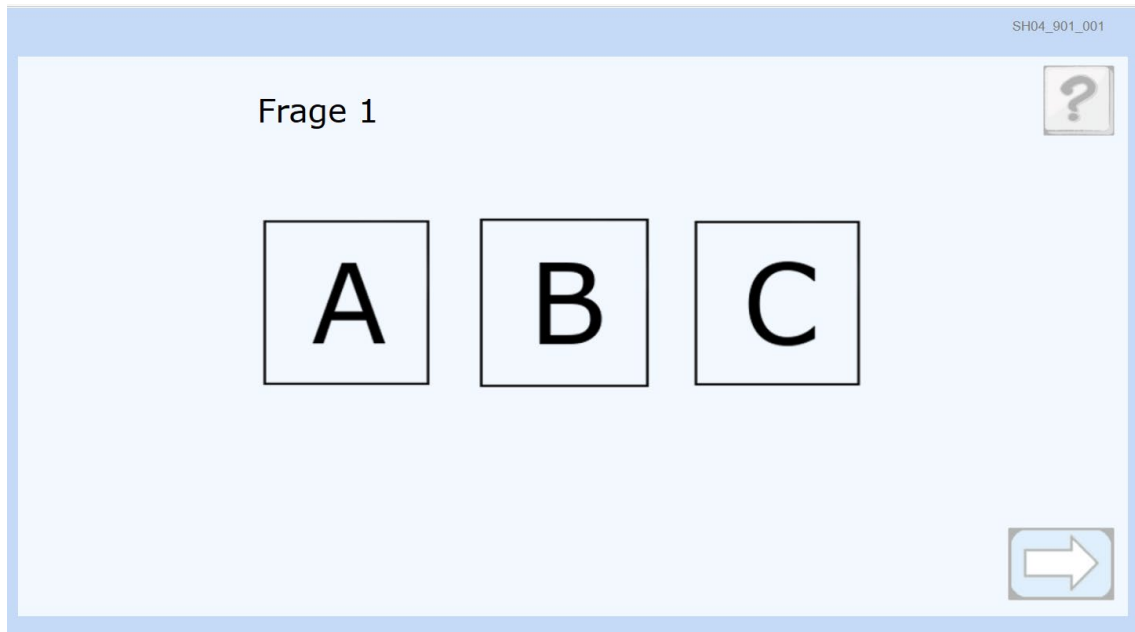
Du bist am Geburtstagsfest von Jan. Ihr macht ein Piraten-Spiel. Die Mutter von Jan erklärt euch, wie das Spiel geht. Hör genau zu.

Jetzt kommt das Piraten-Spiel! Wir machen zwei Gruppen. In einer Gruppe sind Piraten. In der anderen sind Haie, also ganz gefährliche Fische.

Am Boden hat es einen Ring. Alle Piraten stellen sich in den Ring. Die Haie sind aussen und dürfen nicht in den Ring. Sie versuchen, die Piraten aus dem Ring zu ziehen. Sie dürfen die Piraten aber nur an den Armen ziehen.

Das Spiel ist fertig, wenn alle Piraten aus dem Ring gezogen wurden.

Denkt daran, ihr dürft nur an den Armen ziehen.



Wann ist das Spiel fertig?

- A. Das Spiel ist fertig, wenn alle Haie im Ring sind.
- B. Das Spiel ist fertig, wenn ein Pirat aus dem Ring gezogen wurde.
- C. Das Spiel ist fertig, wenn alle Piraten aus dem Ring gezogen wurden.

Tipp die richtige Antwort an.

Fachdidaktischer Kommentar

Diese Aufgabe enthält als Stimulus eine Spielanleitung und damit einen instruktiven Text. In der Kontextualisierung wird eine konkrete, den Schülerinnen und Schülern aus ihrem Alltag bekannte Situation geschildert: Die Zuhörenden sind auf einer Geburtstagsfeier und die Mutter des Geburtstagskindes erklärt ein Spiel. Die Information zum Ende des Spiels wird explizit so genannt, wie in Antwort C formuliert. Die Erinnerung an sie wird dadurch erleichtert, dass der Text kurz ist und die Information am Schluss des Textes genannt wird. Zudem handelt es sich um eine den Schülerinnen und Schülern vertraute Textsorte, zu welcher als essenzieller Bestandteil ein Hinweis zum Ende des Spiels gehört.

Lösung

Das Spiel ist fertig, wenn alle Piraten aus dem Ring gezogen wurden.

2.3 Aufgabenentwicklung

Die Testaufgaben zur Überprüfung der Grundkompetenzen in der Schulsprache wurden von Fachdidaktikerinnen und Fachdidaktikern von pädagogischen Hochschulen der drei Sprachregionen der Schweiz in Zusammenarbeit mit der Geschäftsstelle Aufgabendatenbank EDK (GS ADB) entwickelt:

- Linda Boos-Leutwiler, Miriam Dittmar, Sabine Geiger, Thomas Lindauer, Pädagogische Hochschule der Fachhochschule Nordwestschweiz (PH FHNW, Wissenschaftliche Leitung)
- Nadine Nell-Tuor, Zentrum Mündlichkeit, Pädagogische Hochschule Zug (PH Zug)
- Anne-Claire Blanc, Vincent Capt, Haute école pédagogique Vaud (HEP Vaud)
- Nicodemo Cannavò, Alessandra Moretti, Dipartimento dell'educazione, della cultura e dello sport (DECS / Repubblica e Cantone Ticino)
- Petra Mazzoni, Miriam Salvisberg, Scuola universitaria professionale della Svizzera italiana (SUPSI; Umsetzung der Testaufgaben mit der Plattform CBA Itembuilder).

Die Testaufgaben im Kompetenzbereich Lesen wurden von Fachdidaktikerinnen und Fachdidaktikern der PH FHNW, der HEP Vaud und des DECS entwickelt, die Testaufgaben im Kompetenzbereich Hören von Fachdidaktikerinnen der PH Zug. Die Umsetzung der Testaufgaben in computerbasierte Testaufgaben erfolgte durch das SUPSI mit der Plattform CBA ItemBuilder (Rölke, 2012).

Qualitäts- und Validitätssicherung

Folgende Prozesse dienen der Sicherstellung von Qualität und Validität der Aufgabenentwicklung.

Machbarkeitsstudie

2018/19 wurde unter der Leitung der Geschäftsstelle der Aufgabendatenbank EDK (GS ADB) eine Machbarkeitsstudie für die Fachbereiche Mathematik, Naturwissenschaften und Schulsprache durchgeführt (Angelone et al., 2019). Die Machbarkeitsstudie sollte insbesondere aufzeigen, welche der formulierten Bildungsstandards im Rahmen einer Erhebung zur Überprüfung des Erreichens der Grundkompetenzen (ÜGK) überprüft werden können. Fachdidaktikerinnen und Fachdidaktiker erarbeiteten theoretische Assessment-Frameworks. Diese beschreiben, welche der formulierten Grundkompetenzen im Rahmen einer ÜGK im 4. Schuljahr getestet werden können und zeigen anhand von Beispielaufgaben, wie die Grundkompetenzen operationalisiert werden können. Zudem erprobte die GS ADB in vier Schulklassen ein mögliches standardisiertes Testdurchführungsszenario.

Internes Review

Die Testaufgaben für die Überprüfung des Erreichens der Grundkompetenzen am Ende des 4. Schuljahres wurden sowohl institutionsintern in den beteiligten Fachdidaktik-Teams unter Einbezug von Lehrpersonen als auch institutionsübergreifend in Gesamttreffen der Aufgabenentwicklungsgruppe wiederholt begutachtet und überarbeitet.

CogLabs

Im November 2020 fanden CogLabs statt. Dabei lösten Schülerinnen und Schüler des 4. Schuljahres die Aufgaben und gaben in retrospektiven Interviews Auskunft über ihren Lösungsprozess und ihr Verständnis der Aufgaben. So konnten allfällige inhaltliche Probleme identifiziert und die Aufgaben überarbeitet werden.

Externes Review und Präpilottierung

Im Sommer / Herbst 2021 wurden die Aufgaben durch externe Expertinnen und Experten (Lehrpersonen, externe Fachdidaktik und Heilpädagogik) begutachtet und in einer kleineren Stichprobe von 21 Schulklassen (7 pro Sprachregion) präpilottiert. Die qualitativen und quantitativen Rückmeldungen erlaubten eine weitere Überarbeitung der Testaufgaben im Hinblick auf die Piloterhebung.

Übersetzung der Aufgaben im Bereich Lesen und Hören

Die Aufgaben in der Schulsprache wurden in den drei Landessprachen Italienisch, Französisch und Deutsch erstellt. Die Übersetzung der Aufgaben gestaltete sich als rekursiver Prozess im Endstadium der Aufgabenentwicklung. Ziel war es, bei möglichst guter Vergleichbarkeit der Übersetzungen (bspw. die lexikalische Geläufigkeit einzelner Wörter oder die Komplexität syntaktischer Strukturen betreffend) doch flüssig lesbare und natürlich klingende Texte zu erarbeiten (siehe dazu auch die internationalen Guidelines in Koršňáková et al. 2020).

Pilotierung der Aufgaben

Die Aufgaben wurden im Frühjahr 2022 in einer sprachregional repräsentativen Stichprobe von rund 4'300 Schülerinnen und Schülern pilotiert. Die Ergebnisse zeigten, dass genügend valide Items vorhanden waren, um die Grundkompetenzen im Lesen und Hören zu testen. Allerdings zeigte sich, dass einzelne Aufgaben in den drei Sprachversionen unterschiedlich schwierig waren (Differential Item Functioning). In Absprache mit der Geschäftsstelle der Aufgabendatenbank der EDK (ADB) überarbeiteten die Fachdidaktik-Teams einige Aufgaben behutsam bzw. entschieden sich bei gewissen Aufgaben, diese nicht in die Haupterhebung aufzunehmen. Im Anschluss an die Pilotierung fand im Februar 2023 ein Standardsetting statt.

3 Nationale Bildungsstandards in der Mathematik

Stephan Schönenberger, Alberto Piatti, Jean-Luc Dorier, Domenico Angelone und Florian Keller

Die nationalen Bildungsstandards im Fachbereich Mathematik beschreiben ein Kompetenzmodell, dem ein Konzept von Grundbildung zu Grunde liegt, die für eine gesellschaftliche Teilhabe als wichtig erachtet wird (EDK, 2011b). Der im Jahr 2009 veröffentlichte wissenschaftliche Kurzbericht (Konsortium Mathematik, 2009) dokumentiert die theoretischen Grundlagen und das Vorgehen, um ein Kompetenzmodell für die mathematischen Grundkompetenzen zu formulieren. Basierend auf den Tests aus dem Jahr 2007 wurde ein Kompetenzmodell formuliert und mit Aufgabenbeispielen für den Anhörungsprozess vom 25.1.2010 aufgearbeitet (EDK, 2010). Schliesslich wurden die Grundkompetenzen von der Plenarversammlung der EDK am 16.6.2011 verabschiedet (EDK, 2011b). Grundkompetenzen werden dabei einem von fünf Kompetenzbereichen sowie einem von acht Handlungsaspekten der Mathematik zugeordnet (vgl. Abbildung 3-1).

Abbildung 3-1: Kompetenzmodell Mathematik (EDK, 2011b)



Für das 4. Schuljahr werden Grundkompetenzen innerhalb der beiden Kompetenzbereiche Zahl und Variable sowie Form und Raum in allen Handlungsaspekten formuliert. Die Kompetenzbereiche Grössen und Masse, Funktionale Zusammenhänge und Daten und Zufall werden nicht separat beschrieben. Inhalte aus diesen Bereichen, die für das 4. Schuljahr relevant sind, sind in den gegebenen Formulierungen enthalten. Gleichermassen sind zwar alle Handlungsaspekte berücksichtigt, sollten aber mit Blick auf die Jahrgangsstufe 4 ganzheitlich gelesen werden (EDK, 2011b; Konsortium Mathematik, 2009).

Insgesamt wurden 30 Grundkompetenzen auf mittlerem Abstraktionsniveau formuliert (EDK, 2011b). Eine spezifische Erläuterung der Grundkompetenzen, insbesondere bezüglich Anforderungsniveau, ist den Basisdokumenten zur Anhörung (EDK, 2010, S. 8–25) sowie dem wissenschaftlichen Kurzbericht (Konsortium Mathematik, 2009, S. 41–43) zu entnehmen. Sie bilden zusammen mit dem Aufgabenpool der Untersuchung im Vorfeld zur Verabschiedung der Grundkompetenzen aus dem Jahr 2007 und fachdidaktischer Literatur (Linneweber-Lammerskitten, 2013; Walther et al., 2011) die Grundlage für die Test- und Aufgabenentwicklung zur Überprüfung der Grundkompetenzen im 4. Schuljahr.

3.1 Im Bereich Mathematik überprüfte Grundkompetenzen

Wie gut die Schülerinnen und Schüler die Grundkompetenzen erreichen, wird im Rahmen der ÜGK-Erhebungen anhand von standardisierten Tests erfasst, die von den Schülerinnen und Schülern selbstständig auf Tablets bearbeitet werden. Insbesondere müssen die Testaufgaben ein geschlossenes Aufgabenformat aufweisen. Die Rahmenbedingungen für die ÜGK 2024 wurden von der EDK vorgegeben und im Rahmen einer Machbarkeitsstudie geprüft (Angelone et al., 2019). Tabelle 3-1 zeigt die für die Erhebung relevanten Grundkompetenzen.

Tabelle 3-1: ÜGK 2024: Überprüfte Grundkompetenzen in der Mathematik

	Zahl und Variable	Form und Raum
Wissen, Erkennen und Beschreiben	Die Schülerinnen und Schüler • kennen Ziffern, Stellenwertschreibweise und Zahlwörter bis 100 und können bis 100 zählen; • können Vorgänger und Nachfolger einer Zahl im Zahlenraum bis 100 bestimmen; • können die Zahlen von 1 bis 9 auf 10 ergänzen	Die Schülerinnen und Schüler • verstehen relative Angaben zur Raumlage (insbesondere zwischen, auf, unter, über, darunter, vor, hinter, links von, rechts von) bzw. zur Richtung (links, rechts, geradeaus, vorwärts, rückwärts) und können diese Ausdrücke auch selbst korrekt anwenden; • erkennen und benennen einfache geometrische Figuren (insbesondere Kreis, Rechteck, Quadrat, Dreieck).
Operieren und Berechnen	Die Schülerinnen und Schüler können • Zahlen (< 100) auf dem Zahlenstrahl darstellen; • eine ungeordnete Anzahl von Objekten abzählen (Zahlenraum bis 20); • im Zahlenraum bis 100 auf den nächsten Zehner ergänzen; • Additions- sowie Subtraktionsrechnungen ohne Zehnerüberschreitung im Zahlenraum bis 100 ausführen; • eine einfache Tabelle (z.B. 2 Spalten und 3 Zeilen) lesen und ergänzen.	Die Schülerinnen und Schüler können • komplexere Figuren in einfachere Figuren zerlegen und wieder zusammensetzen; • eine Folge mit einfachen geometrischen Figuren fortsetzen.
Instrumente und Werkzeuge verwenden	Die Schülerinnen und Schüler können • verschiedene Anordnungen von Zahlen (z.B. Hundertertafel) lesen und nutzen.	Die Schülerinnen und Schüler können • Alltagsgegenstände als Hilfsmittel gebrauchen, um Längen miteinander zu vergleichen; • Raster verwenden, um die Position eines Gegenstandes in der Ebene zu beschreiben.
Mathematisieren und Modellieren	Die Schülerinnen und Schüler können • ein Problem in eine Addition oder Subtraktion übersetzen, die im Zahlenraum bis 20 oder bei den Zehnerzahlen bis 100 gelöst werden kann.	Die Schülerinnen und Schüler können • Probleme lösen, die das Wiedererkennen von Formen oder den Gebrauch von geometrischen

		Transformationen erfordern (z.B. die Vervollständigung einfacher Muster).
Interpretieren und Reflektieren der Resultate	Die Schülerinnen und Schüler können • eine gefundene Lösung zu einer arithmetischen Aufgabe überprüfen, wenn sie explizit dazu aufgefordert werden; • entscheiden, ob ein gegebenes Resultat eine richtige Lösung darstellt.	Die Schülerinnen und Schüler können • entscheiden, ob eine Figur oder ein Muster eine Lösung eines gegebenen Problems darstellt.

Spezifisch für die Mathematik hält der Audit Bericht ÜGK 2016 (Mathematik, 11. Schuljahr) fest, dass die von der EDK freigegebenen Grundkompetenzen (Mindeststandards) für die Mathematik auf einem mittlerem Abstraktionsniveau definiert sind und hinsichtlich Anforderungs- und Schwierigkeitsniveau teilweise einen grossen Interpretationsspielraum zulassen (Fischbach & Ugen, 2018, S. 14).

Deshalb wurden die Grundkompetenzen für die Aufgabenentwicklung im Rahmen eines Projekts konkretisiert (Schönenberger et al., 2025). Analog zur Formulierung der Basisstandards (CDPE, 2010) wurden wo nötig in einem ersten Schritt konkrete Beschreibungen der Grundkompetenzen (Niveau 1) und einer erhöhten Schwierigkeit (Niveau 2) formuliert. Aufgaben wurden sowohl zu Grundkompetenzen wie auch schwieriger als Grundkompetenzen entwickelt. Dies ist erforderlich, da für die ÜGK ein Standardsetting vorgesehen, bei dem ein Schwellenwert (Cut Score) zur Festlegung der Erreichung der Grundkompetenzen bestimmt wird. Ein valides Standardsetting ist möglich, wenn der Test Aufgaben unterschiedlicher Schwierigkeit enthält. Zu einfache oder zu schwierige Aufgaben tragen nicht zu einer fundierten Bestimmung eines Schwellenwerts bei. Die für den Test verwendeten konkretisierten Beschreibungen wurden mit Vorliegen der Daten aus Präpilotierung und externen Reviews überarbeitet. Schliesslich wurden die Konkretisierungen anlässlich des Standardsettings durch die Expertengruppe geprüft.

Operationalisierung

Für die Testung der Mathematik wurden folgende geschlossene Aufgabenformate eingesetzt: Multiple Choice, Image Map, Drag and Drop und Textentry. Offene Aufgabenformate werden nicht eingesetzt, weil die Kodierung der Schülerantworten kostenintensiv ist und bei offenen Aufgabenformaten nicht ausgeschlossen werden kann, dass die Lösungswahrscheinlichkeit einer Testaufgabe durch individuelle Fertigkeiten der Schülerinnen und Schüler im Umgang mit der Tablettastatur beeinflusst wird.

Für die Aufgabenformulierungen werden, abhängig vom Aufgabenformat, einheitliche Aufforderungen verwendet. Diese sind mit dem Fachbereich Schulsprache abgesprochen und gelten somit fachübergreifend. Schülerinnen und Schüler, die über die Grundkompetenzen im Lesen verfügen, verstehen die Mathematikaufgaben. Die Anweisungen in den Aufgaben werden schriftlich gegeben. Zusätzlich wird eine Audiodatei bereitgestellt, so dass die Schülerinnen und Schüler die Möglichkeit haben, sich die Aufgaben, Anweisungen und allenfalls Lösungsvorschläge vorlesen zu lassen.

Zur Überprüfung der Grundkompetenzen in der Mathematik wurden insgesamt 90 Items eingesetzt (vgl. Tabelle 3-2).

Tabelle 3-2: Anzahl Items des Kompetenztests zur Mathematik

		Kompetenzbereiche		Total
		Zahl und Variable	Form und Raum	
Handlungsaspekte	Wissen, Erkennen und Beschreiben	19	8	27
	Operieren und Berechnen	13	7	20
	Instrumente und Werkzeuge verwenden	3	9	12
	Mathematisieren und Modellieren	7	7	14
	Interpretieren und Reflektieren der Resultate	10	7	17
Total		52	38	90

Aufgabenbeispiele

Die folgenden Abbildungen 3-2 bis 3-6 zeigen Aufgabenbeispiele, die zur Überprüfung der Grundkompetenzen in der Mathematik eingesetzt wurden. Schülerinnen und Schüler, welche die Grundkompetenzen erreichen, können diese Aufgaben in der Regel richtig lösen.

Abbildung 3-2: Aufgabe zum Bereich Zahl und Variable

MA04_815_001

In jeder Linie soll die Rechnung 10 ergeben.
Welche Zahl musst du ergänzen?
Zieh die passenden Zahlen in die Felder.



10		
8+		
5+		
9+		
3+		

1 2 3 4 5 6 7 8 9



Fachdidaktischer Kommentar

Die Kinder tippen auf die Zahlen und 'ziehen' diese in die Felder. Das Layout erinnert an ein bekanntes Aufgabenformat 'Zahlenhaus'. Indem die Kinder die Ziffern korrekt an die Stellen im Haus ziehen, zeigen sie die Fähigkeit, Zahlen auf 10 zu ergänzen. Die Aufgabe operationalisiert die Grundkompetenz:

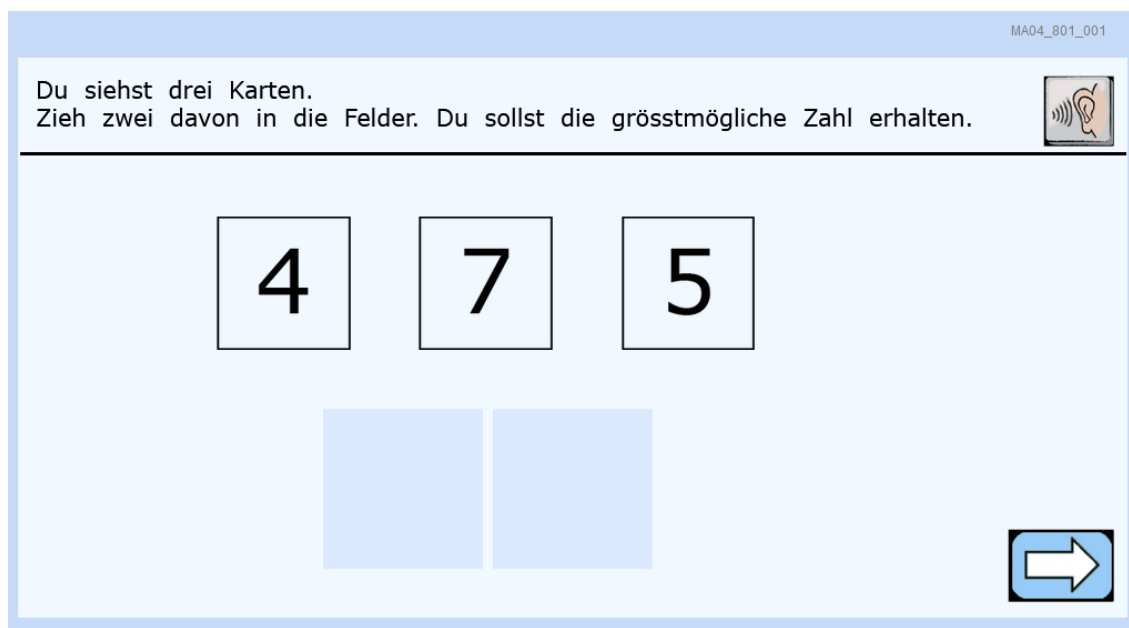
Die Schülerinnen und Schüler können die Zahlen von 1 bis 9 auf 10 ergänzen.

Diese Kompetenz bedarf keiner weiteren Konkretisierung.

Lösung

2, 5, 1, 7 (von oben nach unten)

Abbildung 3-3: Aufgabe zum Bereich Zahl und Variable



Fachdidaktischer Kommentar

Die Kinder tippen auf dem Tablet die Ziffernkarten und 'ziehen' diese in die Felder. Indem die Kinder die grösste Ziffer als Zehner und die zweitgrösste Ziffer als Einer wählen, demonstrieren sie ein Verständnis von 'Stellenwertschreibweise', das einer Grundkompetenz entspricht. Die Aufgabe operationalisiert die Grundkompetenz:

Die Schülerinnen und Schüler kennen Ziffern, Stellenwertschreibweise und Zahlwörter bis 100 und können bis 100 zählen.

In dieser Kompetenz ist es wichtig, dass die Zahlen auch gehört werden können. In Niveau 1 sind gegebene Zahlen stets im angegebenen Zahlbereich.

Anforderungen Grundkompetenz (Niveau 1)	Anforderungen höher als Grundkompetenz (Niveau 2)
Schülerinnen und Schüler können • von gegebenen Zahlen die Grösste auswählen, • aus drei gegebenen 'Ziffernkarten' eine zweistellige Zahl bilden, die möglichst gross ist, • eine geschriebene oder gesprochene Zahl eintippen, • aus 'Ziffernkarten' die zu einem Zahlwort (z.B. «Siebzehn») entsprechende Zahl bilden und umgekehrt.	Schülerinnen und Schüler können • aus drei 'Ziffernkarten' eine Zahl bilden, die zwischen zwei Zahlen aus dem Zahlenraum bis 100 liegt, • aus drei 'Ziffernkarten' eine Zahl bilden, möglichst nahe an einer gegebenen Zahl kleiner 100 ist.

Lösung

Abbildung 3-4: Aufgabe zum Bereich Zahl und Variable

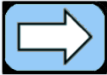
MA04_339_001

Im ersten Topf hat Lisa 7 Gänseblumen. Im zweiten Topf hat sie 5 orange Rosen und 5 gelbe Rosen.



Wie viele Blumen hat Lisa?
Tipp die richtige Rechnung an.

- ☐ $2 + 7 + 5 + 5$
- ☐ $7 + 5 + 5$
- ☐ $2 + 5 + 7$

Fachdidaktischer Kommentar

Die Kinder müssen die zur Aufgabe passende Rechnung wählen. Sie zeigen so ihre Kompetenz, ein Problem in eine Addition zu übersetzen. Die Aufgabe operationalisiert die Grundkompetenz:

Die Schülerinnen und Schüler können ein Problem in eine Addition oder Subtraktion übersetzen, die im Zahlenraum bis 20 oder bei den Zehnerzahlen bis 100 gelöst werden kann.

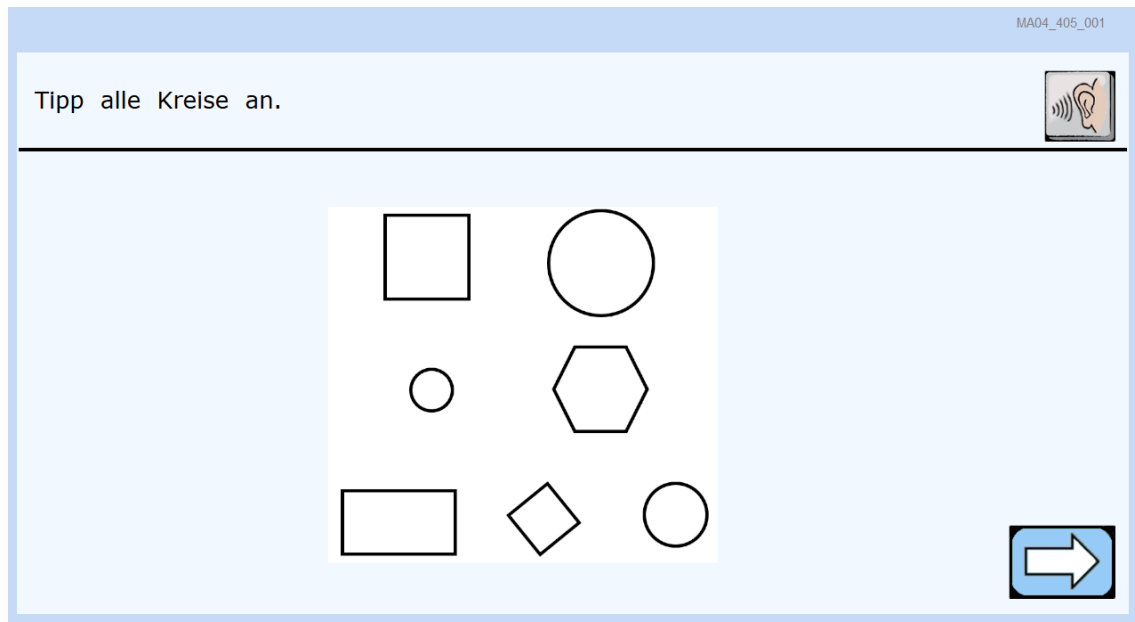
Eine Beschreibung additiver oder subtraktiver Sachaufgaben findet sich z.B. bei Hasemann und Gasteiger (2020) oder Gagnebin et al. (1998). Die Konkretisierung der 'Probleme' umfasst zwei Aspekte. Die *Anzahl der Rechenschritte* (einschrittige oder mehrschrittige Probleme) und den *Aufgabentyp* (im Sinn von Vergnaud 1986): Bei *Kombinationsaufgaben* werden zwei Zustände zu einem neuen Zustand kombiniert (Beispiel: Maria hat 3 Murmeln, Hans hat 5. Wie viele haben sie gemeinsam?). *Austauschaufgaben* enthalten eine Transformation (Beispiel: Hans hat 5 Murmeln, Maria gibt ihm 3. Wie viele Murmeln hat Hans nun?). Andere Typen (über Grundkompetenz) fragen nach einem *Vergleich* (Beispiel: Hans hat 8 Murmeln, Maria hat 5. Wie viele Murmeln hat Hans mehr als Maria?) (Consortio HarmoS Matematica, 2009, S. 42). In der französischsprachigen Schweiz wird die Typisierung additiver Aufgaben verwendet, die von Gagnebin et al. (1998) beschrieben wird und von Vergnaud (1986) inspiriert ist. Diese Typisierung wird deshalb auch für die Konkretisierung der Grundkompetenz verwendet.

Anforderungen Grundkompetenz (Niveau 1)	Anforderungen höher als Grundkompetenz (Niveau 2)
Schülerinnen und Schüler können im Zahlenraum bis 20 oder bei den Zehnerzahlen Kombinationsaufgaben mit ein bis zwei Schritten oder Austauschaufgaben lösen.	Schülerinnen und Schüler können mehrschrittige Aufgaben oder Aufgaben eines anderen Typs lösen.

Lösung

7+5+5

Abbildung 3-5: Aufgabe zum Bereich Form und Raum



Fachdidaktischer Kommentar

Die Schülerinnen und Schüler erkennen Kreise und tippen sie an. Andere Figuren haben Ecken (keine Ovale, Ellipsen oder ähnlich). Die Aufgabe operationalisiert die Grundkompetenz:

Die Schülerinnen und Schüler erkennen und benennen einfache geometrische Figuren (insbesondere Kreis, Rechteck, Quadrat, Dreieck).

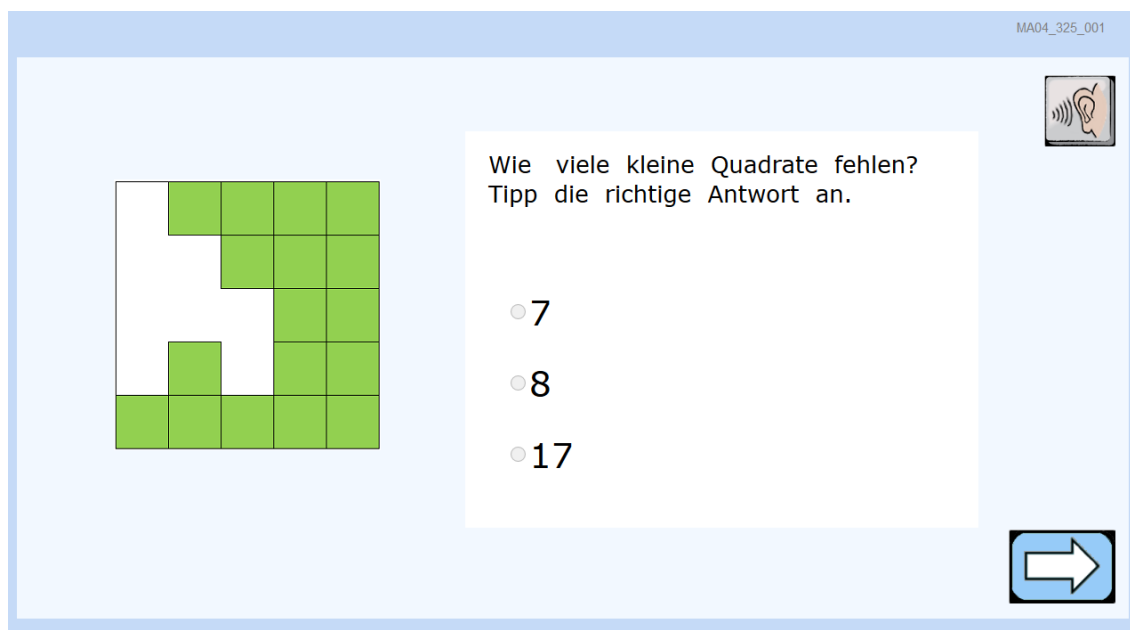
Zu dieser Kompetenz sind Aufgaben mit unterschiedlichem Anspruch möglich, die eine Konkretisierung nötig machen.

Anforderungen Grundkompetenz (Niveau 1)	Anforderungen über Grundkompetenz (Niveau 2)
Schülerinnen und Schüler können geometrische Formen in der Aufzählung erkennen und benennen (z.B. Unterscheidung Dreieck / Kreis)	Schülerinnen und Schüler können schwierigere räumliche und geometrische Fachbegriffe verstehen und richtig anwenden (z.B. Unterscheidung Rechteck / Quadrat oder Kreis / Oval).

Lösung

Alle drei vorhandenen Kreise wurden angewählt.

Abbildung 3-6: Aufgabe zum Bereich Form und Raum



Fachdidaktischer Kommentar

Die Schülerinnen und Schüler erkennen die quadratische Form und zählen, wie viele kleine Quadrate ergänzt werden müssen. Der fehlende Teil enthält kein Raster. Um die richtige Anzahl zu finden, müssen sie im fehlenden Teil die Form der kleinen Quadrate erkennen und das Muster vervollständigen können. Die Aufgabe operationalisiert die Grundkompetenz:

Die Schülerinnen und Schüler können Probleme lösen, die das Wiedererkennen von Formen oder den Gebrauch von geometrischen Transformationen erfordern (z.B. die Vervollständigung einfacher Muster).

Inhaltlich ist eine Konkretisierung nötig, die die Art der Probleme beschreibt. Eine mögliche Problemstellung besteht in der Fortsetzung bzw. Vervollständigung einer Folge von einfachen geometrischen Figuren (vgl. oben). Weiter können Flächenmuster mit einfachen Figuren und Symmetrien vorkommen (Translationen, Rotationen um 90° , Spiegelungen an vertikalen oder horizontalen Achsen).

Anforderungen Grundkompetenz (Niveau 1)	Anforderungen höher als Grundkompetenz (Niveau 2)
Schülerinnen und Schüler können • einfache Muster weiterführen (z.B. einfache Bandornamente, Muster beginnt an der Startstelle). • in einfachen Figuren fehlende Teile ergänzen.	Schülerinnen und Schüler können • komplexe Muster weiterführen (Flächenornamente, Muster beginnt an beliebiger Stelle).

Lösung

3.2 Aufgabenentwicklung

Basierend auf den überprüfbaren Grundkompetenzen sowie den vorangegangenen Arbeiten wurden von Fachdidaktikerinnen und Fachdidaktiker der drei Sprachregionen in Zusammenarbeit mit der Geschäftsstelle Aufgabendatenbank EDK (GS ADB) Testaufgaben entwickelt. Beteiligt waren

- Stephan Schönenberger (PHTG, wissenschaftliche Leitung)
- Katrin Kocher (PHLU)
- Sarah Jandl (PHZH)
- Jean-Luc Dorier (Uni GE)
- Yana Lacek (Uni GE)
- Céline Vendeira (Uni GE)
- Isaline Ruf (IRDp)
- Audrey Daina (HEPL)
- Martine Balegno (HEPL)
- Alberto Piatti (SUPSI)
- Elena Franchini (SUPSI)
- Petra Mazzoni (SUPSI; Umsetzung der Testaufgaben mit CBA Itembuilder)

Die Umsetzung der Testaufgaben in computerbasierte Testaufgaben erfolgte durch das SUPSI mit der Plattform CBA ItemBuilder (Rölke, 2012).

Qualitäts- und Validitätssicherung

Um die Passung von Testaufgaben zu den formulierten Grundkompetenzen sicherzustellen und diese zu konkretisieren, wurden verschiedene Validierungsschritte vorgenommen (Pellegrino et al., 2016).

Machbarkeitsstudie

Die Machbarkeitsstudie von 2018/2019 sollte auch für Mathematik aufzeigen, welche der formulierten Bildungsstandards im Rahmen einer Erhebung zur Überprüfung des Erreichens der Grundkompetenzen (ÜGK) überprüft werden können. Fachdidaktikerinnen und Fachdidaktiker erarbeiteten theoretische Assessment-Frameworks. Diese beschreiben, welche der formulierten Grundkompetenzen im Rahmen einer ÜGK im 4. Schuljahr getestet werden können und zeigen anhand von Beispielaufgaben, wie die Grundkompetenzen operationalisiert werden können. Zudem erprobte die GS ADB in vier Schulklassen ein mögliches standardisiertes Testdurchführungsszenario.

Internes Review

Die Aufgaben wurden in den unterschiedlichen Sprachregionen entwickelt, wiederholt begutachtet und auf Grund der Diskussionen in den einzelnen Fachteams überarbeitet. Im Rahmen der Testentwicklung

wurden einzelne Grundkompetenzen konkretisiert und erste mögliche Operationalisierungen des Anspruchsniveaus bezüglich Grundkompetenz erarbeitet.

CogLabs

Einzelne Testaufgaben wurden mittels retrospektiven Interviews (Hofmeyer et al., 2015; Leighton, 2017) bezüglich Kompetenzzuordnung überprüft. Schülerinnen und Schüler lösten dabei eine Testaufgabe und wurden anschliessend dazu befragt. So konnte einerseits festgestellt werden, ob die Denkweise bei der Aufgabenlösung passend zur Kompetenz ist, andererseits wurden inhaltliche, sprachliche wie auch technische Probleme der Testaufgaben offengelegt.

Externes Review und Präpilotierung

Die Testaufgaben wurden unter Einbezug von Lehrpersonen aus verschiedenen Kantonen aller Sprachregionen, sowie einer Fachperson für Fachdidaktik und Heilpädagogik je aus der Deutschschweiz und der Romandie zur Begutachtung gegeben, insb. auch bezüglich Anspruchsniveau. Damit wurde die Passung zum Schulalltag und die fachdidaktische Qualität der Aufgaben eingeschätzt. Eine Präpilotierung in 21 Schulklassen (7 Schulklassen pro Sprachregion) diente der testtheoretischen Aufgabenvalidierung. Auffällige Aufgaben (z.B. mit Differential Item Functioning) wurden in einem gemeinsamen Workshop diskutiert und entweder überarbeitet oder verworfen.

Übersetzung der Aufgaben im Bereich Mathematik

Die Aufgaben wurden in den drei Landessprachen Italienisch, Französisch und Deutsch erstellt. Die Übersetzung der Aufgaben gestaltete sich als rekursiver Prozess im Endstadium der Aufgabenentwicklung. Ziel war es, bei möglichst guter Vergleichbarkeit der Übersetzungen die kulturellen Eigenheiten und spezifischen Ausdrucksweisen zu berücksichtigen. Spezifische Formulierungen einzelner Lehrmittel konnten dabei nicht berücksichtigt werden. Formulierungen, die von kantonal üblichen Lehrmitteln abweichen, wurden im Rahmen der CogLabs geprüft.

Pilotierung der Aufgaben

Die Aufgaben wurden im Frühjahr 2022 in einer sprachregional repräsentativen Stichprobe von rund 4'300 Schülerinnen und Schülern pilotiert. Der Pilottest diente einerseits dazu, die Testdurchführung zu prüfen, andererseits wurden so die Aufgaben empirisch validiert. Die Auswertungen der Resultate zeigten, dass genügend valide Items vorhanden waren, um die Grundkompetenzen in Mathematik zu testen. In Absprache mit der Geschäftsstelle der Aufgabendatenbank der EDK (ADB) wurden einzelne auffällige Aufgaben überarbeitet bzw. aus der Erhebung entfernt. Im Anschluss an die Pilotierung fand im Februar 2023 ein Standardsetting statt. Im Rahmen des Standardsettings für Mathematik wurden die vorgeschlagenen Konkretisierungen der Grundkompetenzen diskutiert und finalisiert.

4 Testdesign

Giang Pham, Domenico Angelone und Florian Keller

Die an der ÜGK-Erhebung teilnehmenden Schülerinnen und Schüler bearbeiteten jeweils ein Testheft, das nur eine Teilmenge aller eingesetzten Testaufgaben enthält. Damit die Aufgabenschwierigkeiten und die Leistungen der Schülerinnen und Schüler trotzdem auf einer gemeinsamen Messskala abgebildet werden können, wurden die Testhefte so zusammengestellt, dass die verschiedenen Testhefte zum Teil gleiche Aufgabengruppen enthalten. Im Rahmen der ÜGK wurde für jeden Fach- und Kompetenzbereich jeweils ein Youden-Square-Design (Frey et al., 2009) verwendet.

Dafür wurden die Testaufgaben in einem ersten Schritt zu unterschiedlichen Aufgabenblöcken mit gleich langer Bearbeitungszeit gruppiert (Schulsprache Lesen: SL1-SL8; Schulsprache Hören; SH1-SH9; Mathematik: M1-M10). In einem zweiten Schritt wurden die Aufgabenblöcke systematisch auf verschiedene Testhefte verteilt (vgl. Tabelle 4-1 bis Tabelle 4-3). In einem zweiten Schritt wurden die Aufgabenblöcke eines Kompetenzbereichs systematisch auf unterschiedliche Testhefte verteilt (vgl. Tabelle 4-4 und 4-5). Jeder Aufgabenblock (und damit jede Testaufgabe) erscheint in gleicher Anzahl von Testheften, einmal an jeder Position im Testheft, und jede Kombination von Aufgabenblöcken erscheint nur in einem einzigen Testheft.¹ Die Testhefte wurden dann zufällig auf die Schülerstichprobe verteilt.

Tabelle 4-1: Kompetenztests zum Lesen in der Schulsprache: Anzahl Items pro Aufgabenblock

Kompetenztest zum Lesen in der Schulsprache								
Bereich	SL1	SL2	SL3	SL4	SL5	SL6	SL7	SL8
Satzverstehen (N = 15)	2	2	2	2	2	1	2	2
Textverstehen								
Explizite Informationen (N = 14)	2	1	1	2	2	2	2	2
Implizite Informationen (N = 15)	0	2	4	1	2	3	1	2
Globale Informationen (N = 7)	2	1	0	1	1	0	2	0
Total (N = 51)	6	6	7	6	7	6	7	6

¹ Eine Ausnahme stellt der Block SH9 im Kompetenztest zum Hören in der Schulsprache dar, der nur eine Testaufgabe enthält und immer am Ende eines Testhefts erscheint.

Tabelle 4-2: Kompetenztests zum Hören in der Schulsprache: Anzahl Items pro Aufgabenblock

Bereich	SH1	SH2	SH3	SH4	SH5	SH6	SH7	SH8	SH9
Explizite Informationen (N = 24)	5	3	3	4	2	2	1	3	1
Implizite Informationen (N = 10)	0	0	1	1	1	2	3	1	1
Total (N = 34)	5	3	4	5	3	4	4	4	2

Tabelle 4-3: Kompetenztest Mathematik: Anzahl Items pro Aufgabenblock

Bereich	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Zahl und Variable (N = 52)	5	4	6	3	5	6	5	6	6	6
Form und Raum (N = 38)	4	5	3	6	4	3	4	3	3	3
Total (N = 90)	9	9	9	9	9	9	9	9	9	9

Tabelle 4-4: Testheftdesign zu den Kompetenztests in der Schulsprache

Kompetenztest zum Lesen in der Schulsprache										
Position	Dauer	Testheft-Nr.								
		1	2	3	4	5	6	7	8	
1	7.5 Min	SL1	SL2	SL3	SL4	SL5	SL6	SL7	SL8	
2	7.5 Min	SL2	SL3	SL4	SL5	SL6	SL7	SL8	SL1	

Kompetenztest zum Hören in der Schulsprache										
Position	Dauer	Testheft-Nr.								
		1	2	3	4	5	6	7	8	
1	7 Min.	SH1	SH2	SH3	SH4	SH5	SH6	SH7	SH8	
2	7 Min.	SH2	SH3	SH4	SH5	SH6	SH7	SH8	SH1	
3	7 Min.	SH3	SH4	SH5	SH6	SH7	SH8	SH1	SH2	
4	4 Min.	SH9	SH9	SH9	SH9	SH9	SH9	SH9	SH9	

Tabelle 4-5: Testheftdesign zu den Kompetenztests in der Mathematik

Position	Dauer	Testheft-Nr.									
		1	2	3	4	5	6	7	8	9	10
1	7.5 Min.	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
2	7.5 Min.	M2	M3	M4	M5	M6	M7	M8	M9	M10	M1

3	7.5 Min.	M3	M4	M5	M6	M7	M8	M9	M10	M1	M2
4	7.5 Min.	M4	M5	M6	M7	M8	M9	M10	M1	M2	M3

Zusatztest zur Leseflüssigkeit in der Schulsprache: Die Leseflüssigkeit wurde mithilfe der *Sentence Verification Technique* (Mayringer & Wimmer, 2003; Reiss et al., 2019) gemessen. Die Schülerinnen und Schüler erhielten während drei Minuten nacheinander Sätze präsentiert, die sie anschliessend hinsichtlich ihrer inhaltlichen Richtigkeit beurteilen mussten (Ja-/Nein-Antworten). Für den Test wurden zwei Testhefte mit jeweils 50 Items eingesetzt. Beide Testhefte enthielten dieselben Items, jedoch in umgekehrter Reihenfolge. Die Verteilung der Testhefte auf die Schülerstichprobe erfolgte zufällig.

Testabfolge: In der Piloterhebung zeigte es sich ein signifikanter Positionseffekt in Bezug auf die Testabfolge (Angelone et al., 2025). Die Leistungen im Lesetest fielen signifikant geringer aus, wenn dieser nicht zu Beginn, sondern im Anschluss an den Mathematiktest vorgelegt wurde. Für den Mathematiktest konnte hingegen kein entsprechender Positionseffekt festgestellt werden. Aufgrund dieser Befunde wurde für die Haupterhebung ein einheitlicher Testablauf festgelegt: (1) Lesen in der Schulsprache (15 Minuten), (2) Hören in der Schulsprache (25 Minuten), (3) Mathematik (30 Minuten); (4) Zusatztest Leseflüssigkeit in der Schulsprache (3 Minuten).

5 Skalierung

Giang Pham, Domenico Angelone und Florian Keller

Im Frühjahr 2022 wurde eine Piloterhebung mit einer sprachregional repräsentativen Stichprobe durchgeführt, um (1) die Durchführbarkeit der geplanten Erhebung zu evaluieren und (2) empirische Itemparameter zur Festlegung von Schwellenwerten zum Erreichen der Grundkompetenzen zu ermitteln (vgl. Kapitel 6; Angelone et al., 2025). Die auf Basis der Piloterhebung ermittelten Schwellenwerte müssen auf die Haupterhebung übertragen werden. Dafür ist es erforderlich, die Itemparameter der Haupterhebung auf die Metrik der Piloterhebung zu transformieren.

In die Skalierung sind Daten von 19'883 Schülerinnen und Schülern mit mindestens einer gültigen Antwort in einem der Kompetenztests eingegangen. Alle Item-Antworten wurden dichotom (0/1) kodiert. Fehlende Werte aufgrund ausgelassener oder nicht erreichter Items wurden als falsch gewertet.

5.1 Itemkalibrierung und Linking

Zur Überprüfung der Invarianzannahme wurde zunächst untersucht, ob ein potenzielles differenzielles Itemfunktionieren (DIF) zwischen der Pilot- und Haupterhebung vorliegt (Trendtel, Pham, et al., 2016). Zu diesem Zweck wurde getrennt nach den Kompetenzbereichen Mathematik, Lesen und Hören jeweils eine Differential-Item-Functioning-Analyse durchgeführt, bei der die Erhebungszugehörigkeit (Piloterhebung vs. Haupterhebung) als Gruppierungsvariable diente (Item-Parameter-Drift-Analyse, IPD-Analyse). Die Analysen wurden mithilfe logistischer Regression (Zumbo, 1999) mit dem R Package «sirt» (Robitzsch, 2024) unter Berücksichtigung der latenten Personenfähigkeit (WLE) durchgeführt. Die WLE's stammen aus einer gemeinsamen Skalierung der Daten beider Erhebungen nach dem Rasch-Modell. Die Ergebnisse haben gezeigt, dass sich die Itemparameter zwischen der Pilot- und Haupterhebung nur geringfügig unterscheiden. In Absprache mit den Fachdidaktikerinnen und Fachdidaktikern der Aufgabenentwicklung wurde entschieden, dass die Übertragbarkeit der Ergebnisse der Schwellenwertsetzung aus der Piloterhebung auf die Haupterhebung gegeben ist und einzelne Items mit starkem IPD (Typ C gemäss ETS-Klassifikation, vgl. Longford et al., 1993) aus den weiteren Berechnungen auszuschliessen.

Anschliessend wurden die Itemparameter der Haupterhebung auf die Metrik der Piloterhebung transformiert. Für diese Skalentransformation standen drei methodische Vorgehensweisen zur Verfügung (Trendtel, Pham, et al., 2016): (1) Übernahme der Itemparameter aus der Piloterhebung als Fixwerte für die Haupterhebung, (2) gemeinsame Kalibrierung beider Stichproben in einem simultanen Skalierungsmodell oder (3) Transformation der Skala der Haupterhebung auf die Metrik der Piloterhebung unter der Annahme, dass die Items im Mittel invariant funktionieren (Linking). Aufgrund zweier zentraler Erwägungen wurde die dritte Vorgehensweise gewählt: erstens besteht ein geringes Ausmass an IPD zwischen den beiden Erhebungen und zweitens sind geschätzte Itemparameter in jeder Erhebung mit Messfehlern behaftet.

Zum Linking der Itemparameter zwischen der Haupterhebung und Piloterhebung wurde die Haberman-Methode (Haberman, 2009) gewählt – eine in der Bildungsforschung etablierte Linking-Strategie, die insbesondere bei komplexen Designs mit mehreren Gruppen Anwendung findet (Robitzsch, 2025). Die Methode schätzt die Itemparametern (Schwierigkeit und Trennschärfe) gemeinsam über beide Stichproben (Pilot- und Haupterhebung) hinweg und berücksichtigt zugleich Unterschiede im Itemfunktionieren zwischen den Stichproben. Die Haberman-Methode stellt somit eine flexible und robuste Methode für stichprobenübergreifendes Linking dar (Robitzsch, 2021).

Im ersten Schritt der Umsetzung wurde die Itemkalibrierung nach dem Rasch-Modell getrennt für die getesteten Kompetenzbereiche Lesen, Hören, Mathematik auf Grundlage der Daten der Haupterhebung durchgeführt. Die Schätzung der Itemschwierigkeiten wurde mit dem R-Paket *TAM* (Robitzsch et al., 2024) durchgeführt, wobei das Marginal Maximum Likelihood (MML)-Verfahren unter Berücksichtigung der korrigierten Stichprobengewichte zum Einsatz kam. Die Mittelwerte der Personenparameter wurden jeweils auf null fixiert. Im nächsten Schritt wurden die frei geschätzten Itemparameter der Haupterhebung mit den Itemparametern der Piloterhebung anhand der Haberman-Methode (Haberman, 2009) verlinkt. Dies erfolgte mithilfe der Funktion «*linking.haberman()*» aus dem R-Paket *sirt* (Robitzsch, 2024). Für weitere Analysen und Berechnungen wurden die verlinkten *joint*-Itemparameter verwendet.

5.2 Itemanalyse

Nach der Schätzung der *joint*-Itemparameter wurden die Items auf weitere psychometrische Gütekriterien geprüft. Als Kriterien wurden die Item- und Distraktorentrennschärfe und die Modellpassung (Itemfit) herangezogen. Items mit einer Trennschärfe $rpb < .20$, einem schlechten Itemfit ($Inf\hat{it} > 1.2$) oder nicht modellkonformer IC-Kurve wurden überprüft. Die Items wurden zudem hinsichtlich Differential-Item-Functioning (DIF) nach Sprachregion sowie nach Geschlecht und zu Hause gesprochener Sprache der Schülerinnen und Schüler überprüft. Die DIF-Analysen wurden mithilfe logistischer Regression mit dem R Package «*sirt*» (Robitzsch, 2024) unter Berücksichtigung der latenten Personenfähigkeit (WLE) durchgeführt. Items mit starkem DIF (Typ C gemäss ETS-Klassifikation, vgl. Longford et al. 1993) wurden von einer Expertengruppe aus Fachdidaktikerinnen und Fachdidaktiker dahingehend überprüft, ob die vermutete Ursache des DIF konstruktrelevant oder -irrelevant ist (Camilli, 2006; Trendtel, Schwabe, et al., 2016). Items mit vermutetem konstruktirrelevantem starkem DIF wurden aus den weiteren Analysen ausgeschlossen. Für die Skalierung verblieben je nach Kompetenztest zwischen 34 und 87 Items im Itempool (vgl. Tabelle 5-1 bis Tabelle 5-3). Anhand des finalen Itempools wurden die *joint*-Parameter erneut geschätzt. Die Itemparameter aus der Piloterhebung, die freigeschätzten Parameter aus der Haupterhebung sowie die verlinkten Parameter sind im Anhang in den Tabellen 7-1 bis 7-3 dargestellt.

Tabelle 5-1: Schulsprache Lesen: Anzahl Items in der Skalierung

Dimension	Textsorte	Verstehensleistung			Total
		Explizite Informationen	Implizite Informationen	Globale Informationen	
Textverstehen	Informative Texte (N = 6)	4	6	1	11
	Instruktive Texte (N = 6)	1	4	3	8
	Narrative Texte (N = 8)	8	4	2	14
Satzverstehen					15
Total					48

Tabelle 5-2: Schulsprache Hören: Anzahl Items in der Skalierung

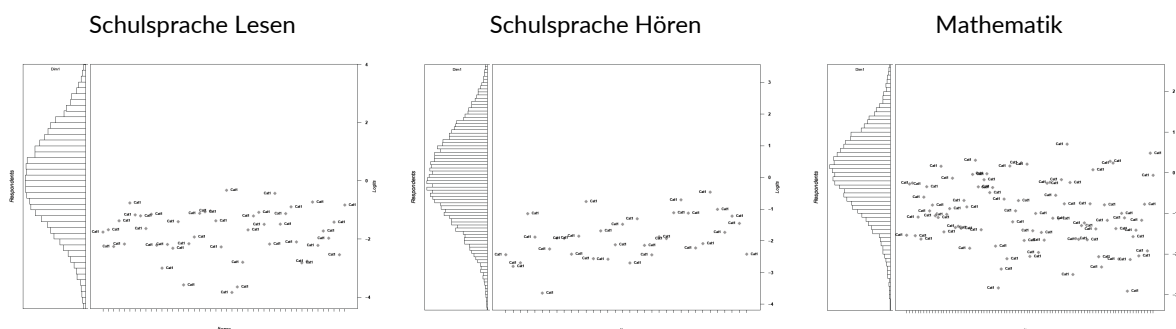
Textsorte	Verstehensleistung		Total
	Explizite Informationen	Implizite Informationen	
Informative Texte (N = 6)	10	5	15
Instruktive Texte (N = 5)	7	0	7
Narrative Texte (N = 5)	7	5	12
Total	24	10	34

Tabelle 5-3: Mathematik: Anzahl Items in der Skalierung

		Kompetenzbereiche		Total
		Zahl und Variable	Form und Raum	
Handlungsaspekte	Wissen, Erkennen und Beschreiben	17	8	25
	Operieren und Berechnen	12	7	19
	Instrumente und Werkzeuge verwenden	3	9	12
	Mathematisieren und Modellieren	7	7	14
	Interpretieren und Reflektieren der Resultate	10	7	17
Total		52	38	87

Die Verteilungen der Item- und Personenparameter sind in Abbildung 5-1 dargestellt.

Abbildung 5-1: Verteilung der Item- und Personenparameter (RP = .625)



5.3 Plausible Value Imputation

Mit der ÜGK 2024 wurden die Kompetenzen in den Kompetenzbereichen Lesen, Hören und Mathematik getestet. Für jeden dieser 3 Kompetenzbereiche wurden 20 Plausible Values (PV) als Schätzer für die Kompetenzverteilungen gezogen. Die PV wurden aus einer A-Posteriori-Verteilung möglicher Leistungswerte bei Berücksichtigung des Skalierungsmodells und der Zusammenhänge zwischen Leistungs- und Kontextdaten gezogen. Neben den Leistungswerten der anderen Kompetenzbereiche wurden alle Variablen aus dem Schülerfragebogen und Elternfragebogen sowie zentrale Informationen aus den Trackingdaten (z. B. Geschlecht, SEN-Status, Alter, Schulprogramme, vgl. Sampling-Bericht) in das Hintergrundmodell einbezogen. Dabei wurden sowohl lineare wie auch quadratische Zusammenhänge zwischen Leistungs- und Kontextdaten modelliert. Da einige Kovariaten fehlende Werte aufwiesen, wurden diese zunächst imputiert. Die Imputation der Kovariaten und die Ziehung der PV der Leistungsdaten erfolgten in einem gemeinsamen Prozess nach Robitzsch et al. (2016) unter Verwendung des MICE-Ansatzes und des R-Pakets mice (van Buuren, 2012). Die Mehrebenenstruktur der Clusterung von Schülerinnen und Schülern in Schulen wurde im Imputationsmodell repräsentiert, indem von allen Kovariaten und Kompetenzwerten jeweils ein Schulmittelwert linear in die Regression aufgenommen wurde. Für die abhängige Variable (d.h. die zu imputierende Variable) wurde die Mehrebenenstruktur näherungsweise abgebildet, indem der um den jeweiligen Schüler adjustierte Schulmittelwert (d.h. für Schüler i in Schule j wird der Schulmittelwert aller Schüler von Schule j exklusive des Schülers i ermittelt) der abhängigen Variable als Kovariate aufgenommen wurde. Weil das Hintergrundmodell dadurch sehr viele zu schätzende Regressionskoeffizienten enthielt, wurde die Dimensionalität der Prädiktoren mit der Methode der Partial Least Squares Regression reduziert (Mevik & Wehrens, 2007). Dabei wurden 50 unabhängige Faktoren auf Basis der spezifizierten Prädiktoren bestimmt, so dass die Kovarianz zwischen den Faktoren und der abhängigen Variablen maximiert wird. Dadurch bleibt die erklärte Varianz im Imputationsmodell praktisch erhalten, die grosse Anzahl der Prädiktoren wird aber auf wenige Faktoren reduziert. Die Imputation der Kovariaten erfolgte schliesslich mit der linearen Regression der zu imputierenden Variablen auf die bestimmten Faktoren nach der Methode des predictive mean matching (van Buuren, 2012), die Imputationen aus der Menge der

beobachteten Werte einer Variablen generiert. Dadurch ist diese Imputationsmethode auch für die Imputation von dichotomen oder ordinalen Variablen geeignet. Das Vorgehen wurde für die Imputation aller Kovariaten mit fehlenden Werten eingesetzt. Die Imputation erfolgte separat nach Kanton und Sprachregion. Bei dieser Vorgehensweise werden alle Zusammenhänge zwischen Leistungs- und Kontextdaten kantonspezifisch geschätzt.

Für die Ziehung der PV pro Kompetenzbereich wurde jeweils ein Skalierungsmodell mit fixierten Itemparameter aus der Itemkalibrierung (nach dem Linking) spezifiziert. Die Konstruktion der Prädiktoren folgte dabei demselben Prinzip wie bei der Imputation der Kovariaten. Die Ziehung der PV erfolgten mit Hilfe des R-Paket «TAM» (Robitzsch et al., 2024). Für die Imputation der Kovariaten und die Ziehung der PV wurden 20 separate Startmodelle mit unterschiedlichen Startwerten initiiert. Die imputierten Werte der Kontextdaten sowie die zugehörigen PV eines Modelllaufs nach 40 Iterationen wurden gespeichert.

6 Standardsetting

Domenico Angelone und Giang Pham

Das Standardsetting zur ÜGK 2024 basiert auf der Datengrundlage der sprachregional repräsentativen Piloterhebung (Angelone et al., 2025), die im Frühjahr 2022 mit rund 4'300 Schülerinnen und Schüler durchgeführt wurde. Im Anschluss an die Skalierung der Testdaten wurde unter der Leitung der Geschäftsstelle der Aufgabendatenbank der EDK getrennt nach den Fachbereichen Schulsprache und Mathematik je ein zweitägiger Expertenworkshop zur Schwellenwertsetzung durchgeführt. Ziel dabei war es, auf einer kontinuierlichen Testskala jeweils einen Schwellenwert festzulegen, der es erlaubt, Schülerinnen und Schüler dahingehend einzustufen, ob sie die Grundkompetenzen in den entsprechenden Fach- und Kompetenzbereichen erreicht haben oder nicht.

Die Schwellenwerte wurden von Expertengruppen anhand einer modifizierten Bookmark-Methode (Karantonis & Sireci, 2006; Mitzel et al., 2001) bestimmt. Die Expertengruppen setzten sich aus Fachdidaktikerinnen und Fachdidaktikern der Aufgabenentwicklung, externen Fachdidaktikerinnen und Fachdidaktikern, die nicht an der Aufgabenentwicklung beteiligt waren, sowie praktizierenden Lehrpersonen und Heilpädagoginnen und Heilpädagogen der Primarschule zusammen (vgl. Tabelle 6-1). Die Auswahl der Lehrpersonen sowie Heilpädagoginnen und Heilpädagogen erfolgte durch die Geschäftsstellen der EDK Regionalkonferenzen.

Tabelle 6-1: Standardsetting Schulsprache Lesen und Hören: Zusammensetzung der Expertengruppen

Sprachregion	Fachdidaktik	Lehrpersonen	Heilpädagogik	Total
Deutschsprachige Schweiz	2	2	1	5
Französischsprachige Schweiz	2	2	1	5
Italienischsprachige Schweiz	2	2	1	5
Total	6	6	3	15

Tabelle 6-2: Standardsetting Schulsprache Mathematik: Zusammensetzung der Expertengruppen

Sprachregion	Fachdidaktik	Lehrpersonen	Heilpädagogik	Total
Deutschsprachige Schweiz	2	2	1	5
Französischsprachige Schweiz	2	2	1	5
Italienischsprachige Schweiz	2	2	1	5
Total	6	6	3	15

Zu Beginn des Standardsettings wurden die Expertinnen und Experten jeweils geschult: Nach einer Einführung in die Überprüfung des Erreichens der Grundkompetenzen (ÜGK) in der Schweiz und die Standardsetting-Methode wurden die entsprechenden nationalen Bildungsstandards

(«Grundkompetenzen») mit ihren «Can-do»-Deskriptoren erläutert. Unklarheiten in den Begrifflichkeiten wurden in Kleingruppen und im Plenum diskutiert.

Im Anschluss daran erfolgte in mehreren Bewertungsrunden die Bestimmung der Schwellenwerte. Dazu wurde der Expertengruppe ein Itembuch vorgelegt, in dem alle Testaufgaben bzw. Items des entsprechenden Kompetenzbereiches nach der empirisch ermittelten Schwierigkeit geordnet waren. Jede Seite enthielt den Stimulus und die Fragestellung bzw. Aufforderung der Testaufgabe, die Lösung und die Seitenzahl.

In Runde 1 wurden die Expertinnen und Experten aufgefordert, jede Testaufgabe des Itembuchs individuell danach zu beurteilen, ob die definierten Grundkompetenzen ausreichen, um die Aufgabe in der Regel richtig zu lösen. Die genaue Frage lautete: «Sollte ein Schüler, der gerade über Grundkompetenzen verfügt, dieses Item in der Regel (d. h. in 2 von 3 Fällen) richtig lösen können? – Ja oder nein?». Im Anschluss daran wurde der Expertengruppe eine Grafik mit den aggregierten Bewertungsergebnissen pro Item gezeigt. Die Ergebnisse wurden mit Blick auf unterschiedlich bewertete Aufgaben im Plenum diskutiert.

In Runde 2 wurden die Expertinnen und Experten aufgefordert, im Itembuch an jener Stelle ein Lesezeichen (Bookmark) zu setzen, an der sie die Schwelle zwischen Grundkompetenzen und höheren Kompetenzen vermuteten. Dazu mussten die Expertinnen und Experten für jedes Item dieselbe Frage wie in Runde 1 beantworten. Bei dem Item, bei dem die Antwort «Nein» lautete, sollte ein Lesezeichen gesetzt werden, allerdings erst nachdem geprüft worden war, ob nicht nachfolgende (d. h. empirisch schwierigere) Items wieder als Grundkompetenz-Items eingestuft werden müssten. Das Ziel bestand darin, den definitiven Übergang zu den Items zu finden, die von Schülerinnen und Schülern mit Grundkompetenzen (in der Regel) nicht mehr korrekt beantwortet werden können. Nachdem in der Gruppe alle Lesezeichen gesetzt worden waren, wurden die entsprechenden Seitenzahlen gesammelt und grafisch aufbereitet. Die Grafik mit den individuellen Bewertungen (Seitenzahlen) sowie einem ersten Schwellenwert – dem Mittelwert aller Seitenzahlen – wurde der Expertengruppe als Rückmeldung präsentiert. Zusätzlich erhielten die Expertinnen und Experten eine Rückmeldung darüber, wie hoch der Anteil der Schülerinnen und Schüler wäre, der bei diesem Schwellenwert die Grundkompetenzen nicht erreichen würde. Diese Information – ebenso wie die Streuung der Seitenzahlen – sollte dazu anregen, die eigene Einschätzung kritisch zu hinterfragen. Die Rückmeldungen wurden im Plenum diskutiert, mit besonderem Fokus auf die Gründe für unterschiedliche Bewertungen. In der dritten Runde wurden die Expertinnen und Experten dazu aufgefordert, das Itembuch unter Berücksichtigung der vorangegangenen Diskussionen nochmals durcharbeiten und ihre Einschätzungen gegebenenfalls zu überarbeiten. Anschliessend wurden erneut die Streuung der Seitenzahlen, der Schwellenwert sowie der Anteil der Schülerinnen und Schüler, die bei diesem Schwellenwert die Grundkompetenzen nicht erreichen würden, präsentiert und im Plenum diskutiert. Nach diesen drei Bewertungsrunden konnte in allen Fach- bzw. Kompetenzbereichen ein abschliessendes Urteil getroffen werden. Die festgelegten Schwellenwerte wurden im Konsensentscheid als vertretbar erachtet (vgl. Tabelle 6-3).

Tabelle 6-3: Expertengruppen, Bewertungsrunden und Schwellenwerte

Kompetenzbereich	Anzahl Personen der Expertengruppe	Anzahl Bewertungs- runden	Schwellenwert Linking (Logits, RP = 62.5%)
Schulsprache Lesen	15	3	-1.31
Schulsprache Hören	15	3	-1.65
Mathematik	15	3	-0.66

7 Anhang

Tabelle 7-1: Itemparameter Schulsprache Lesen

Item	Xsi Joint (RP = .50)	Xsi Main Study (RP = .50)	Xsi Pilot Study (RP = .50)	Xsi Joint (RP = .625)
SL04_101_001	-3.45		-3.24	
SL04_102_001	-3.30		-3.10	
SL04_104_001	-2.27	-2.00	-2.33	-1.76
SL04_104_002	-2.19	-2.19		-1.68
SL04_105_001	-2.77	-2.58	-2.75	-2.26
SL04_106_001	-1.88	-1.93	-1.63	-1.37
SL04_106_002	-2.68	-2.78	-2.37	-2.17
SL04_107_001	-1.27	-1.15	-1.18	-0.76
SL04_107_002	-1.68	-1.69	-1.47	-1.17
SL04_107_003	-1.71	-1.75	-1.47	-1.20
SL04_109_001	-2.15	-2.25	-1.84	-1.64
SL04_109_002	-1.66	-1.82	-1.29	-1.15
SL04_110_001	-2.72	-2.30	-2.94	-2.21
SL04_113_001	-3.51	-3.49	-3.32	-3.00
SL04_113_002	-2.42		-2.22	
SL04_113_003	-1.77		-1.56	
SL04_115_001	-2.70	-2.66	-2.52	-2.18
SL04_115_002	-2.83	-2.93	-2.51	-2.32
SL04_115_003	-1.91	-2.05	-1.56	-1.40
SL04_116_001	-4.09	-3.93	-4.03	-3.58
SL04_117_001	-2.67	-2.92	-2.21	-2.16
SL04_117_002	-2.44	-2.64	-2.03	-1.93
SL04_118_001	-1.63	-1.63		-1.11
SL04_204_001	-1.57	-1.64	-1.30	-1.06
SL04_204_002	-1.54	-1.53	-1.35	-1.03
SL04_205_001	-1.88	-1.93	-1.63	-1.37
SL04_206_001	-2.78	-2.96	-2.40	-2.27
SL04_206_002	-0.83	-0.68	-0.78	-0.32
SL04_208_001	-4.35	-4.29	-4.20	-3.84
SL04_209_001	-4.16	-4.06	-4.04	-3.65

SL04_211_001	-3.31	-3.50	-2.91	-2.80
SL04_212_001	-2.20	-2.36	-1.82	-1.68
SL04_217_001	-1.72	-1.67	-1.56	-1.21
SL04_217_002	-1.59	-1.67	-1.31	-1.08
SL04_301_001	-2.01	-1.74	-2.06	-1.50
SL04_303_001	-2.68	-2.43	-2.72	-2.17
SL04_305_001	-0.94	-0.96	-0.71	-0.43
SL04_305_002	-2.00	-2.18	-1.61	-1.49
SL04_306_002	-1.64	-1.72	-1.34	-1.13
SL04_307_001	-1.48		-1.27	
SL04_307_002	-1.40	-1.44	-1.16	-0.89
SL04_308_001	-2.61	-2.44	-2.58	-2.10
SL04_312_001	-3.33	-3.54	-2.91	-2.82
SL04_313_001	-3.27	-3.08	-3.26	-2.76
SL04_314_001	-1.24	-1.33	-0.94	-0.73
SL04_315_001	-2.73	-3.09	-2.16	-2.22
SL04_315_002	-2.23	-2.53	-1.73	-1.72
SL04_317_001	-2.48	-2.44	-2.31	-1.97
SL04_318_001	-1.93	-1.63	-2.02	-1.42
SL04_319_001	-3.05	-2.81	-3.08	-2.54
SL04_320_001	-1.34	-1.20	-1.27	-0.83

Tabelle 7-2: Itemparameter Schulsprache Hören

Item	Xsi Joint (RP = .50)	Xsi Main Study (RP = .50)	Xsi Pilot Study (RP = .50)	Xsi Joint (RP = .625)
SH04_901_001	-2.95	-3.00	-2.95	-2.44
SH04_901_002	-3.32	-3.41	-3.28	-2.81
SH04_902_001	-3.21	-3.15	-3.32	-2.69
SH04_902_002	-1.65	-1.65		-1.14
SH04_902_003	-2.39	-2.32	-2.52	-1.88
SH04_903_001	-4.16	-4.16	-4.21	-3.65
SH04_903_002	-2.77	-2.68	-2.91	-2.26
SH04_904_001	-2.42	-2.67	-2.23	-1.91
SH04_905_001	-2.43	-2.31	-2.60	-1.92
SH04_905_002	-2.93	-2.84	-3.07	-2.42
SH04_906_001	-2.36	-2.28	-2.50	-1.85
SH04_906_002	-1.27	-1.27		-0.76
SH04_907_001	-3.07	-3.07	-3.13	-2.56
SH04_907_002	-2.20	-2.05	-2.39	-1.69
SH04_908_001	-3.09	-3.12	-3.12	-2.58
SH04_908_002	-2.64	-2.80	-2.52	-2.12
SH04_908_003	-1.99	-2.15	-1.88	-1.48
SH04_909_001	-3.21	-3.15	-3.32	-2.70
SH04_910_001	-1.82	-2.03	-1.65	-1.30
SH04_911_001	-2.65	-2.61	-2.75	-2.14
SH04_911_002	-2.95	-2.82	-3.14	-2.44
SH04_911_003	-2.38	-2.42	-2.40	-1.87
SH04_912_001	-2.44	-2.44		-1.93
SH04_912_002	-1.62	-1.62		-1.11
SH04_913_001	-1.22	-1.20	-1.29	-0.71
SH04_913_002	-1.64	-1.60	-1.72	-1.13
SH04_913_003	-2.74	-2.75	-2.78	-2.23
SH04_914_001	-2.60	-2.60		-2.09
SH04_914_002	-0.97	-0.97		-0.46
SH04_916_001	-1.52	-1.59	-1.50	-1.01
SH04_916_002	-2.24	-2.18	-2.36	-1.73

SH04_916_003	-1.73	-1.78	-1.73	-1.22
SH04_917_001	-1.96	-1.93	-2.05	-1.45
SH04_917_002	-2.93	-2.86	-3.06	-2.42

Tabelle 7-3: Itemparameter Mathematik

Item	Xsi Joint (RP = .50)	Xsi Main Study (RP = .50)	Xsi Pilot Study (RP = .50)	Xsi Joint (RP = .625)
MA04_301_001	-2.05	-2.23	-1.80	-1.54
MA04_302_001	-0.79	-0.88	-0.64	-0.28
MA04_303_001	-0.76	-1.04	-0.42	-0.25
MA04_304_001	-2.06	-2.17	-1.89	-1.55
MA04_306_001	-1.60	-1.75	-1.40	-1.09
MA04_307_001	-2.14	-2.11	-2.12	-1.63
MA04_308_001	-1.11	-0.98	-1.17	-0.60
MA04_311_001	-0.85	-0.91	-0.74	-0.34
MA04_312_001	-1.45	-1.37	-1.46	-0.94
MA04_313_001	-1.31	-1.25	-1.30	-0.80
MA04_314_001	-1.57	-1.41	-1.67	-1.06
MA04_315_001	-1.61	-1.54	-1.62	-1.10
MA04_316_001	-0.35	-0.34	-0.30	0.16
MA04_318_001	-1.97	-2.00	-1.87	-1.46
MA04_320_001	-1.53	-1.54	-1.47	-1.02
MA04_322_001	-1.39	-1.40	-1.32	-0.88
MA04_323_001	-0.64	-0.65	-0.57	-0.13
MA04_324_001	-1.86	-1.90	-1.76	-1.35
MA04_325_001	-1.83	-1.92	-1.68	-1.32
MA04_326_001	-1.88	-1.94	-1.75	-1.37
MA04_327_001	-1.19	-1.14	-1.19	-0.68
MA04_328_001	-1.35	-1.63	-1.02	-0.84
MA04_329_001	0.21		0.28	
MA04_332_001	-0.08		-0.02	
MA04_333_001	-2.37	-2.56	-2.11	-1.86
MA04_334_001	-0.55	-0.53	-0.51	-0.04
MA04_336_001	-0.20	-0.15	-0.19	0.31
MA04_338_001	-0.86	-0.90	-0.76	-0.35
MA04_339_001	-1.91	-2.00	-1.77	-1.40
MA04_340_001	-0.68	-0.59	-0.71	-0.17
MA04_341_001	-0.54	-0.44	-0.57	-0.02
MA04_343_001	-1.00	-1.08	-0.86	-0.49
MA04_344_001	-0.87	-0.80	-0.89	-0.36

MA04_401_001	-2.32	-2.47	-2.10	-1.81
MA04_402_001	-3.35	-3.38	-3.25	-2.83
MA04_403_001	-2.88	-3.00	-2.70	-2.37
MA04_404_001	-1.16	-1.16	-1.10	-0.65
MA04_405_001	-2.63	-2.79	-2.40	-2.12
MA04_406_001	-0.35	-0.36	-0.27	0.17
MA04_409_001	-1.72	-1.55	-1.83	-1.21
MA04_410_001	-1.45	-1.58	-1.25	-0.94
MA04_411_001	-1.20	-1.23	-1.10	-0.69
MA04_412_001	0.35		0.42	
MA04_413_001	-0.27	-0.21	-0.26	0.25
MA04_414_001	-2.18	-2.19	-2.10	-1.67
MA04_415_001	-0.30	-0.36	-0.18	0.21
MA04_417_001	-2.57	-2.79	-2.28	-2.06
MA04_418_001	-1.94	-1.80	-2.02	-1.43
MA04_419_001	-1.51	-1.63	-1.33	-1.00
MA04_421_001	-2.47	-2.60	-2.28	-1.96
MA04_422_001	-1.98	-1.71	-2.18	-1.47
MA04_423_001	-2.17	-2.24	-2.04	-1.66
MA04_424_001	-0.77	-0.71	-0.77	-0.26
MA04_425_001	-0.75	-0.73	-0.70	-0.24
MA04_427_001	-1.63	-1.72	-1.48	-1.12
MA04_428_001	-1.07	-1.13	-0.94	-0.56
MA04_429_001	-1.64	-1.46	-1.76	-1.13
MA04_430_001	-0.68	-0.59	-0.71	-0.17
MA04_801_001	-1.28	-1.33	-1.16	-0.76
MA04_801_002	0.19	0.34	0.11	0.70
MA04_801_003	-0.75	-0.65	-0.79	-0.24
MA04_802_001	-3.02	-2.85	-3.12	-2.51
MA04_802_002	-2.13	-2.02	-2.18	-1.62
MA04_803_001	-2.15	-2.05	-2.19	-1.64
MA04_808_001	-1.81	-1.99	-1.57	-1.30
MA04_808_002	-1.75	-1.92	-1.51	-1.24
MA04_810_001	-2.16	-2.06	-2.19	-1.65
MA04_810_002	-1.27	-1.16	-1.33	-0.76

MA04_811_001	-0.44	-0.44		0.07
MA04_811_002	-1.89	-1.89		-1.38
MA04_812_001	-1.67		-1.61	
MA04_813_001	-2.58	-2.62	-2.48	-2.07
MA04_814_001	-3.70		-3.63	
MA04_815_001	-2.83	-2.70	-2.90	-2.32
MA04_821_001	-1.30	-1.16	-1.38	-0.79
MA04_822_001	-1.71		-1.65	
MA04_825_001	-1.68	-1.75	-1.55	-1.17
MA04_832_001	-0.23	-0.26	-0.13	0.28
MA04_833_001	-0.27	-0.32	-0.16	0.24
MA04_838_001	-1.89	-1.85	-1.86	-1.38
MA04_848_001	-2.63	-2.38	-2.82	-2.12
MA04_850_002	-1.51	-1.51		-1.00
MA04_851_001	-1.62	-1.68	-1.51	-1.11
MA04_860_001	-3.43	-3.44	-3.35	-2.91
MA04_861_001	-2.65	-2.43	-2.80	-2.14
MA04_862_001	-2.08	-2.06	-2.05	-1.57
MA04_863_001	-1.92	-1.96	-1.83	-1.41
MA04_864_001	-2.56	-2.51	-2.54	-2.05
MA04_865_001	-1.68	-1.69	-1.61	-1.17
MA04_866_001	-1.29	-1.27	-1.24	-0.78
MA04_867_001	-2.43	-2.39	-2.41	-1.92
MA04_868_001	-0.03	0.11	-0.10	0.48
MA04_869_001	-0.57	-0.49	-0.60	-0.06

8 Literaturverzeichnis

- Angelone, D., Keller, F., Arnold, J., Jaun-Holderer, B., Leutwiler, L., Nell-Tuor, N., Schönenberger, S., Hauser, M., Eckstein, B., & Erzinger, A. (2019). *Machbarkeitsstudie zur Überprüfung der Grundkompetenzen im 4. Schuljahr*. Geschäftsstelle der Aufgabendatenbank EDK.
- Angelone, D., Keller, F., & Pham, G. (2025). *Überprüfung des Erreichens der Grundkompetenzen (ÜGK) in den Fächern Schulsprache und Mathematik im 4. Schuljahr. Technische Dokumentation zur Testentwicklung und Skalierung der Piloterhebung 2022*. Geschäftsstelle der Aufgabendatenbank EDK (ADB).
- Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiss (Hrsg.), *PISA 2000 Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–407). Leske + Budrich.
- Becker-Mrotzek, M., Lindauer, T., Pfost, M., Weis, M., Strohmaier, A., & Reiss, K. (2019). Lesekompetenz heute—Eine Schlüsselqualifikation im Wandel. In K. Reiss, M. Weis, E. Klieme, & O. Köller (Hrsg.), *PISA 2018. Grundbildung im internationalen Vergleich. 1. Auflage* (S. 21–46). Waxmann.
- Behrens, U. (2010). Aspekte eines Kompetenzmodells zum Zuhören und Möglichkeiten ihrer Testung. In M. Imhof & V. Bernius (Hrsg.), *Zuhörkompetenz in Unterricht und Schule: Beiträge aus Wissenschaft und Praxis* (S. 31–50). Vandenhoeck & Ruprecht.
- Behrens, U., & Weirich, S. (2019). Stim-mig: Assessing prosodic comprehension in primary school. *L1-Educational Studies in Language and Literature*, 1–26. <https://doi.org/10.17239/L1ESLL-2019.19.03.03>
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Hrsg.), *Educational measurement* (S. 221–256). American Council on Education; Praeger Publishers.
- CDPE. (2010). *Standard di base per la matematica: Documenti per il procedimento d'audizione*, 25.1.2010. CDPE. <https://edudoc.ch/record/36467?ln=it>
- Chang, A. C.-S., & Read, J. (2013). Investigating the effects of multiple-choice listening test items in the oral versus written mode on L2 listeners' performance and perceptions. *System*, 41(3), 575–586. <https://doi.org/10.1016/j.system.2013.06.001>
- CIIP. (2010, Januar 1). *Plan d'études romand (PER)*. Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP). www.plandetudes.ch/per
- Consorzio HarmoS Matematica. (2009). *Matematica. Rapporto scientifico di sintesi e modello di competenza. Versione provvisoria (prima dell'adozione degli standards di base)*. Stato: Dicembre 2009. <https://edudoc.ch/record/87028?ln=it>
- D-EDK. (2016, Januar 1). *Lehrplan 21*. D-EDK Geschäftsstelle. <https://www.lehrplan21.ch/>
- EDK. (2010). *Basisstandards für die Mathematik. Unterlagen für den Anhörungsprozess*. 25. 1. 2010. <https://edudoc.ch/record/36469?ln=de>
- EDK. (2011a). *Grundkompetenzen für die Fremdsprachen. Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16. Juni 2011*. Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK). https://edudoc.ch/record/96780/files/grundkomp_fremdsprachen_d.pdf?version=1
- EDK. (2011b). *Grundkompetenzen für die Mathematik. Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16. Juni 2011*. Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK). https://edudoc.ch/record/96784/files/grundkomp_math_d.pdf?ln=de
- EDK. (2011c). *Grundkompetenzen für die Naturwissenschaften. Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16. Juni 2011*. Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK). https://edudoc.ch/record/96787/files/grundkomp_nawi_d.pdf?ln=de

- EDK. (2011d). *Grundkompetenzen für die Schulsprache. Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16. Juni 2011.* Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
https://edudoc.ch/record/96791/files/grundkomp_schulsprache_d.pdf?ln=de
- EDK. (2014). *Organisationsreglement über die Durchführung der Überprüfung des Erreichens der Grundkompetenzen vom 8. Mai 2014.*
https://edudoc.ch/record/113128/files/Organisationsregl_UeGK_d.pdf?ln=de
- Feilke, H., & Rezat, S. (2020). Textprozeduren—Werkzeuge für Schreiben und Lesen. *Praxis Deutsch*, 281, 4–13.
- Fischbach, A., & Ugen, S. (2018). *ÜGK/COFO Mathematics 2016 Audit Report.* Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
https://edudoc.ch/record/204067/files/UeGK_Audit_Report_230218.pdf?ln=en
- Frey, A., Hartig, J., & Rupp, A. A. (2009). An NCME Instructional Module on Booklet Designs in Large-Scale-Assessments of Student Achievement: Theory and Practice. *Educational Measurement: Issues and Practice*, 28(3), 39–53. <https://doi.org/10.1111/j.1745-3992.2009.00154.x>
- Gagnebin, A., Guignard, N., & Jaquet, F. (1998). *Apprentissage et enseignement des mathématiques: Commentaires didactiques sur les moyens d'enseignement pour les degrés 1 à 4 de l'école primaire.* COROME, Commission romande des moyens d'enseignement.
- Giasson, J. (1990). *La compréhension en lecture.* De Boeck.
- Giasson, J. (2011). *La lecture: Apprentissage et difficultés.* G. Morin.
- Gold, A. (2018). *Lesen kann man lernen. Wie man die Lesekompetenz fördern kann.* 3., völlig überarbeitete Auflage. Vandenhoeck & Ruprecht.
- Gschwend, R. (2014). Zuhören und Hörverstehen – Aspekte, Ziele und Kompetenzen. In G. Elke & S. Carmen (Hrsg.), *Konzeptionen des Mündlichen—Wissenschaftliche Perspektiven und didaktische Konsequenzen* (S. 143–160). hep.
- Haberman, S. J. (2009). *Linking parameter estimates derived from an item response model through separate calibrations (ETS Research Report Nr. RR-09-40).* Educational Testing Service.
- Hasemann, K., & Gasteiger, H. (2020). *Anfangsunterricht Mathematik.* Springer.
- Hofmeyer, A., Sheingold, B. H., & Taylor, R. (2015). Do You Understand What I Mean? How Cognitive Interviewing Can Strengthen Valid, Reliable Study Instruments And Dissemination Products. *Journal of International Education Research*, 11(4), 261–267.
- Hurrelmann, B. (2007). Modelle und Merkmale der Lesekompetenz. In A. Bertschi-Kaufmann (Hrsg.), *Lesekompetenz – Leseleistung – Leseförderung. Grundlagen, Modelle und Materialien* (1. Aufl., S. 18–28). Klett Kallmeyer.
- Imhof, M. (2010). Zuhören lernen und lehren. Psychologische Grundlagen zur Beschreibung und Förderung von Zuhörkompetenzen in Schule und Unterricht. In M. Imhof & V. Bernius (Hrsg.), *Zuhörkompetenz in Unterricht und Schule: Beiträge aus Wissenschaft und Praxis* (S. 15–20). Vandenhoeck & Ruprecht.
- Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) (2007). <https://edudoc.ch/record/24711>
- Karantonis, A., & Sireci, S. G. (2006). The Bookmark Standard-Setting Method: A Literature Review. *Educational Measurement: Issues and Practice*, 25(1), 4–12. <https://doi.org/10.1111/j.1745-3992.2006.00047.x>
- Konsortium Mathematik. (2009). *HarmoS Mathematik: Wissenschaftlicher Kurzbericht und Kompetenzmodell: Provisorische Fassung (vor Verabschiedung der Basisstandards): Stand: 13. Dezember 2009.* EDK. <https://edudoc.ch/record/87030?ln=de>
- Konsortium ÜGK (Hrsg.). (2019). *Überprüfung der Grundkompetenzen. Nationaler Bericht der ÜGK 2017: Sprachen 8. Schuljahr.* Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK); Service de la recherche en éducation (SRED). <https://doi.org/10.18747/PHSG-coll3/id/385>
- Krelle, M. (2010). Zuhördidaktik – Anmerkungen zur Förderung rezeptiver Fähigkeiten des mündlichen Sprachgebrauchs im Deutschunterricht. In V. Bernius & M. Imhof (Hrsg.), *Zuhörkompetenz in*

- Unterricht und Schule. Beiträge aus Wissenschaft und Praxis* (S. 51–68). Vandenhoeck & Ruprecht.
- Krelle, M., & Prengel, J. (2014). Zur Konzeption von Zuhören im Rahmen der Vergleichsarbeiten für die dritte Klassen im Fach Deutsch. In E. Grundler & C. Spiegel (Hrsg.), *Konzeptionen des Mündlichen: Wissenschaftliche Perspektiven und didaktische Konsequenzen* (S. 208–226). hep.
- Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning Theory and Assessment of Reading Fluency: Automaticity, Prosody, and Definitions of Fluency. *Reading Research Quarterly*, 45(2), 230–251. <https://doi.org/10.1598/RRQ.45.2.4>
- Leighton, J. P. (2017). *Using Think-Aloud Interviews and Cognitive Labs in Educational Research*. Oxford University Press.
- Linneweber-Lammerskitten, H. (2013). *Fachdidaktik Mathematik. Grundbildung und Kompetenzaufbau im Unterricht der Sek. I und II*. Klett und Balmer Verlag.
- Longford, N. T., Holland, P. W., & Thayer, D. t. (1993). Stability of the MH D-DIF statistics across populations. In P. W. Holland & H. Wainer (Hrsg.), *Differential Item Functioning* (S. 171–196). Erlbaum.
- Mayringer, H., & Wimmer, H. (2003). *SLS 1-4. Salzburger Lese-Screening für die Klassenstufen 1-4*. Verlag Hans Huber.
- Mevik, B.-H., & Wehrens, R. (2007). The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software*, 18(2), 1–23.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark standard setting procedure: Psychological perspectives. In G. J. Cizek (Hrsg.), *Setting performance standards: Concepts, methods and perspectives*. Erlbaum Associates Publishers.
- Morais, J., Pierre, R., & Kolinsky, R. (2003). Du lecteur compétent au lecteur débutant: Implications des recherches en psycholinguistique cognitive et en neuropsychologie pour l'enseignement de la lecture. *Revue des sciences de l'éducation*, 29(1), 51–74. <https://doi.org/10.7202/009492ar>
- Nell-Tuor, N. (2025). Zuhören in der schulischen Interaktion – mündliche Auftragserteilung im Fokus. In S. Fornol & N. Nell-Tuor (Hrsg.), *Zuhören in der Schule – interdisziplinär*. Hep.
- Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A Framework for Conceptualizing and Evaluating the Validity of Instructionally Relevant Assessments. *Educational Psychologist*, 51(1), 59–81.
- Reiss, K., Weis, M., Klieme, E., & Köller, O. (Hrsg.). (2019). *PISA 2018 Grundbildung im internationalen Vergleich*. Waxmann.
- Repubblica e Cantone Ticino. (2015, Januar 1). *Piano di studio della scuola dell'obbligo ticinese*. www.pianodistudio.ch
- Richter, T. ., & Müller, B. (2017). Entwicklung hierarchieniedriger Leseprozesse. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 51–66). Beltz Juventa.
- Richter, T., & Christmann, U. (2009). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (3. Aufl., S. 25–58). Juventa.
- Rincke, K. (2010). Von der Alltagssprache zur Fachsprache. Bruch oder schrittweiser Übergang? In G. Fenkart, A. Lembens, & E. Erlacher-Zeitlinger (Hrsg.), *Sprache, Mathematik und Naturwissenschaften* (Bd. 16, S. 47–62). Studienverlag.
- Robitzsch, A. (2021). A Comparison of Linking Methods for Two Groups for the Two-Parameter Logistic Item Response Model in the Presence and Absence of Random Differential Item Functioning. *Foundations*, 1(1), Article 1. <https://doi.org/10.3390/foundations1010009>
- Robitzsch, A. (2024). *Sirt: Supplementary Item Response Theory Models. R package version 4.1-15* [Software]. <https://CRAN.R-project.org/package=sirt>
- Robitzsch, A. (2025). Linking Error Estimation in Haberman Linking. *AppliedMath*, 5, Article 1. <https://doi.org/10.3390/appliedmath5010007>

- Robitzsch, A., Kiefer, T., & Wu, M. (2024). TAM: Test Analysis Modules. R package version 4.2-21. [Software]. <https://CRAN.R-project.org/package=TAM>
- Robitzsch, A., Pham, G., & Yanagida, T. (2016). Fehlende Daten und Plausible Values. In S. Breit & C. Schreiner (Hrsg.), *Large-Scale Assessment mit R: Methodische Grundlagen der österreichischen Bildungsstandardüberprüfung* (S. 259–293). facultas.
- Rölke, H. (2012). The ItemBuilder: A graphical authoring system for complex item development. In T. Bastiaens & G. Marks (Hrsg.), *Proceedings of E-Learn 2012-World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 1* (S. 344–353). Association for the Advancement of Computing in Education (AACE).
- Rosebrock, C. (2012). Was ist Lesekompetenz, und wie kann sie gefördert werden? *Leseforum*, 3.
- Rosebrock, C., & Nix, D. (2006). Forschungsüberblick: Leseflüssigkeit (Fluency) in der amerikanischen Leseforschung und -didaktik. *Didaktik Deutsch*, 20, 90–112.
- Rosebrock, C., Nix, D., Rieckmann, C., & Gold, A. (2017). *Leseflüssigkeit fördern: Lautleseverfahren für die Primar- und Sekundarstufe* (5. Auflage). Klett Kallmeyer.
- Schindler, J., & Richter, T. (2018). Reading comprehension: Individual differences, disorders, and underlying cognitive processes. In A. Bar-On & D. Ravid (Hrsg.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistic perspectives* (S. 503–524). De Gruyter.
- Schneider, W. (2017). *Lesen und Schreiben lernen*. Springer Berlin Heidelberg.
- Schönenberger, S., Jean-Luc, D., Alberto, P., & Domenico, A. (2025). Überprüfung des Erreichens der Grundkompetenzen in der Mathematik im 4. Schuljahr. Schlussbericht zur Konkretisierung der Grundkompetenzen Mathematik für die Testentwicklung. Geschäftsstelle der Aufgabendatenbank EDK.
- Thomas, A., Latschen, D., Weishaupt, A., Weishaupt, J., Frank, E., & Unterköfler-Klatzer, D. (2013). Item-Generation für die Bildungsstandards D4 Hören. *Forschungszeitung der Pädagogischen Hochschule Kärnten - Viktor Frankl Hochschule*, 4/5, 12–15.
- Trendtel, M., Pham, G., & Yanagida, T. (2016). Skalierung und Linking. In S. Breit & C. Schreiner (Hrsg.), *Large-Scale Assessment mit R: Methodische Grundlagen der österreichischen Bildungsstandardüberprüfung* (S. 185–224). facultas.
- Trendtel, M., Schwabe, F., & Feller, R. (2016). Differenzielles Itemfunktionieren in Subgruppen. In S. Breit & C. Schreiner (Hrsg.), *Large-Scale Assessment mit R: Methodische Grundlagen der österreichischen Bildungsstandardüberprüfung* (S. 111–147). facultas.
- van Buuren, S. (2012). *Flexible Imputation of Missing Data* (1. Aufl.). CRC Press. <https://doi.org/10.1201/b11826>
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- Vergnaud, G. (1986). *Psychologie du développement cognitif et didactique des mathématiques. Un exemple: Les structures additives*. 38, 21–40.
- Walther, G., van den Heuvel-Panhuizen, M., Granzer, D., & Köller, O. (2011). *Bildungsstandards für die Grundschule: Mathematik konkret* (5. Aufl.). Cornelsen Verlag Scriptor GmbH & Co. KG.
- Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense.