



Überprüfung des Erreichens der Grundkompetenzen (ÜGK) im Fach Mathematik im 11. Schuljahr

Technische Dokumentation zur Testentwicklung und Skalierung

Domenico Angelone

Florian Keller

Mai 2019

Inhaltsverzeichnis

1	Einleitung	4
2	Kompetenztest Mathematik 11. Schuljahr	4
3	Testdesign	6
4	Skalierung	7
4.1	Datengrundlage.....	7
4.2	Itemanalyse.....	8
4.3	Itemkalibrierung	10
4.4	Plausible Value Imputation	11
5	Standardsetting	13
6	Anhang	15
	Literaturverzeichnis	19

1 Einleitung

Mit der interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (Harmos-Konkordat) wurde beschlossen, für die obligatorische Schule nationale Bildungsstandards zu entwickeln, einzuführen und im Rahmen des nationalen Bildungsmonitorings regelmässig zu überprüfen (EDK 2007). Im Jahr 2011 hat die Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) nationale Bildungsstandards in den Fächern Schulsprache, Fremdsprachen, Mathematik und Naturwissenschaften verabschiedet (EDK 2011a, 2011b, 2011c, 2011d). Die nationalen Bildungsstandards sind als *Mindeststandards* formuliert und beschreiben Grundkompetenzen, die von praktisch allen Schülerinnen und Schülern bis zu einem bestimmten Schuljahr 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 der obligatorischen Schule einsetzt. Die nationalen Bildungsstandards sind als Zielvorgaben in die sprachregionalen Lehrpläne – *Lehrplan 21*, *Plan d'études romand* und *Piano di studio* – eingeflossen.

Mit den Erhebungen «Überprüfung des Erreichens der Grundkompetenzen (ÜGK)» wird 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).

Die erste Erhebung zur Überprüfung des Erreichens der Grundkompetenzen fand im Frühjahr 2016 im Fach Mathematik im 11. Schuljahr am Ende der obligatorischen Schulzeit statt (Konsortium ÜGK 2019). An der ÜGK-Erhebung 2016 nahmen insgesamt 22'423 Schülerinnen und Schüler aus allen Kantonen der Schweiz teil. Der vorliegende Bericht beschreibt die technischen Grundlagen der Testentwicklung und der Skalierung der Leistungsdaten.

2 Kompetenztest Mathematik 11. Schuljahr

Die mathematischen Grundkompetenzen für das 11. Schuljahr sind in fünf *Kompetenzbereiche* – «Zahl und Variable», «Form und Raum», «Grössen und Masse», «Funktionale Zusammenhänge» und «Daten und Zufall» – und acht *Handlungsaspekte* – «Wissen, Erkennen und Beschreiben», «Operieren und Berechnen», «Instrumente und Werkzeuge verwenden», «Darstellen und Kommunizieren», «Mathematisieren und Modellieren», «Argumentieren und Begründen», «Interpretieren und Reflektieren der Resultate» und «Erforschen und Explorieren» – der Mathematik gegliedert. Für den Kompetenztest Mathematik im 11. Schuljahr wurden insgesamt 180 Testaufgaben zu den fünf *Kompetenzbereichen* – «Daten und Zufall», «Grössen und Masse», «Funktionale Zusammenhänge», «Zahl und Variable» und «Form und Raum» der Mathematik und zu fünf von acht *Handlungsaspekten* – «Argumentieren und

Begründen», «Darstellen und Kommunizieren», «Wissen, Erkennen und Beschreiben», «Mathematisieren und Modellieren» und «Operieren und Berechnen» eingesetzt (vgl. Tabelle 2-1). Auf die Überprüfung der drei Handlungsaspekte «Instrumente und Werkzeuge verwenden», «Interpretieren und Reflektieren der Resultate» und «Erforschen und Explorieren» wurde aufgrund technischer und finanzieller Einschränkungen verzichtet. Die Testaufgaben wurden von Fachdidaktikerinnen und Fachdidaktikern von pädagogischen Hochschulen und Universitäten der drei Sprachregionen der Schweiz (PH FHNW, HEP Valais, HEP BEJUNE, HEP Vaud, Université Genève und DECS/Kanton TI) in Zusammenarbeit mit der Geschäftsstelle Aufgabendatenbank EDK (ADB) auf der Testplattform TAO (v2.5.9) entwickelt.

Tabelle 2-1: Anzahl Items des Kompetenztests nach Handlungsaspekten und -bereichen der Mathematik

		Kompetenzbereiche					Total
		Daten und Zufall	Grössen und Masse	Funktionale Zusammenhänge	Zahl und Variable	Form und Raum	
Handlungsaspekte	Argumentieren und Begründen	5	9	8	7	8	37
	Darstellen und Kommunizieren	6	4	9	6	5	30
	Wissen, Erkennen und Beschreiben	6	3	11	7	7	34
	Mathematisieren und Modellieren	11	7	10	3	5	36
	Operieren und Berechnen	6	5	8	14	10	43
Total		34	28	46	37	35	180

Tabelle 2-2: Anzahl Items des Kompetenztests nach Itemformat und Handlungsaspekten und -bereichen der Mathematik

		Geschlossen	Halboffen	Offen	Total
Handlungsaspekte	Argumentieren und Begründen	7	0	30	37
	Darstellen und Kommunizieren	20	8	2	30
	Wissen, Erkennen und Beschreiben	26	8	0	34
	Mathematisieren und Modellieren	7	27	2	36
	Operieren und Berechnen	16	27	0	43
	Total	76	70	34	180
Kompetenzbereiche	Daten und Zufall	13	14	7	34
	Grössen und Masse	9	11	8	28
	Funktionale Zusammenhänge	20	18	8	46
	Zahl und Variable	21	10	6	37
	Form und Raum	13	17	5	35
	Total	76	70	34	180

Die Testaufgaben liegen in den drei Sprachversionen Deutsch, Französisch und Italienisch vor und bestehen jeweils aus einem kurzen Aufgabenstamm (Stimulus) mit einer oder mehreren Fragen bzw. Aufgabenstellungen. Die Fragen bzw. Aufgabenstellungen werden als Items bezeichnet. Als Antwortformat

kamen geschlossene (76 Items), halboffene (70 Items) und offene Formate (34 Items) vor (vgl. Tabelle 2-2). Die Items mit geschlossenem Format umfassen Multiple Choice Items mit einer korrekten Antwortoption und mindestens drei Distraktoren sowie Complex-Multiple-Choice Items mit mehreren Richtig-Falsch-Items. Items mit halboffenem Antwortformat erfordern die Eingabe einzelner Informationen wie eine Zahl oder ein Wort in ein vorgegebenes Textfeld. Items mit offenem Format hingegen verlangen eine eigenständig formulierte Antwort (mehrere Worte oder Sätze).

Die Testaufgaben wurden im Frühjahr 2015 an einer national repräsentativen Stichprobe ($N = 1'249$) erprobt. Für die Haupterhebung wurden zusätzliche Testaufgaben im unteren Fähigkeitsbereich entwickelt.

3 Testdesign

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 ein Youden-Square-Design (YSD) verwendet (Frey et al. 2009).

Zur Erstellung des YSD wurden in einem ersten Schritt alle Testaufgaben zu 13 ähnlich schwierigen Aufgabenblöcken («M1» bis «M13») zusammengefasst, wobei die Aufgabenmenge pro Block (durchschnittlich 14 Aufgaben) so bemessen wurde, dass ein Block in 25 Minuten vollständig bearbeitet werden kann (vgl. Tabelle 3-1). Jeder Aufgabenblock enthält Testaufgaben zu mindestens drei Handlungsaspekten und zwei Kompetenzbereichen der Mathematik. Innerhalb eines Aufgabenblocks wechseln sich einfachere und schwierigere Aufgaben ab, wobei jeder Block mit einer eher einfachen Aufgabe beginnt.

In einem zweiten Schritt wurden die 13 Aufgabenblöcke systematisch auf 13 unterschiedliche Testhefte verteilt (vgl. Tabelle 3-2). Jedes Testheft setzt sich aus vier Aufgabenblöcken zusammen. Jeder Aufgabenblock (und damit jede Testaufgabe) erscheint in vier der 13 Testhefte, einmal an allen vier Positionen, und jede Kombination von Aufgabenblöcken erscheint nur in einem einzigen Testheft. Die 13 Testhefte wurden zufällig auf die Schülerstichprobe verteilt. Jede Schülerin und jeder Schüler bearbeitete somit nur einen Teil (4/13) des gesamten Aufgabenmaterials.

Tabelle 3-1: Anzahl Items pro Aufgabenblock nach Handlungsaspekten und -bereichen der Mathematik

		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
Handlungsaspekte	Argumentieren und Begründen	0	5	6	3	3	2	5	4	5	0	1	1	2
	Darstellen und Kommunizieren	2	3	0	3	3	2	3	0	0	6	3	1	4
	Wissen, Erkennen und Beschreiben	2	5	4	2	1	1	2	0	2	0	8	6	1
	Mathematisieren und Modellieren	4	2	2	7	3	5	1	1	1	4	0	0	6
	Operieren & Berechnen	5	0	1	0	3	3	3	6	2	5	6	7	2
	Total	13	15	13	15	13	13	14	11	10	15	18	15	15
Kompetenzbereiche	Daten und Zufall	4	1	4	3	0	7	6	0	6	0	2	0	1
	Grössen und Masse	2	4	0	0	1	2	2	2	1	5	3	0	6
	Funktionale Zusammenhänge	2	0	1	6	8	2	3	0	2	5	9	6	2
	Zahl und Variable	0	6	4	5	0	2	0	3	1	2	3	9	2
	Form und Raum	5	4	4	1	4	0	3	6	0	3	1	0	4
	Total	13	15	13	15	13	13	14	11	10	15	18	15	15

Tabelle 3-2: Testdesign Mathematik

Position	Testheft-Nr.												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
2	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M1
3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M1	M2	M3
4	M10	M11	M12	M13	M1	M2	M3	M4	M5	M6	M7	M8	M9

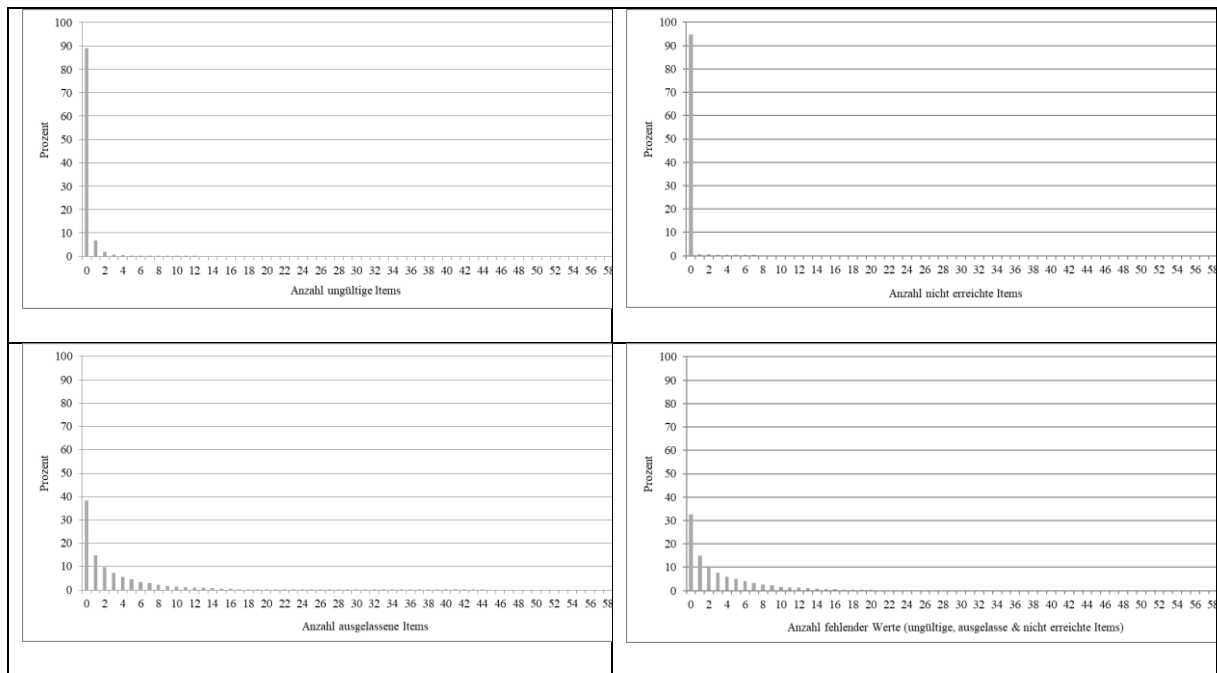
4 Skalierung

4.1 Datengrundlage

In die Skalierung sind Daten von 22'423 Schülerinnen und Schülern mit mindestens einer gültigen Antwort im Kompetenztest oder Schülerfragebogen eingegangen. 282 Schülerinnen und Schüler (1.3%) haben kein Item des Kompetenztests beantwortet. Von den restlichen Schülerinnen und Schüler haben 89.1 Prozent keine ungültigen Werte, 94.9 Prozent das Testende erreicht und 38.3 Prozent kein Item ausgelassen (vgl. Abbildung 4.1). Alle Item-Antworten wurden dichotom (0/1) kodiert.¹ Fehlende Werte aufgrund ungültiger, nicht erreichter und ausgelassener Items wurden als falsch gewertet (mit Ausnahme der Schülerinnen und Schüler, die nur fehlende Werte aufweisen).

¹ Die vier polytomen Items wurden zu dichotomen Items rekodiert: Die Codes 1 («partial credit») und 2 («full credit») wurden dabei zu Code 1 («correct») zusammengefasst.

Abbildung 4.1: Fehlende Werte pro Person



4.2 Itemanalyse

Vor der Schätzung der Itemschwierigkeiten wurden die Items auf der Grundlage eindimensionaler Raschmodelle sowohl auf nationaler wie auch auf sprachregionaler Ebene auf ihre psychometrische Güte geprüft. Dazu wurden die Kriterien Item- und Distraktorentrennschärfe und Modellpassung herangezogen. Jedes Item wurde von mindestens 6'766 Schülerinnen und Schülern bearbeitet. Fehlerhafte Items sowie Items mit einer Trennschärfe $r_{pb} < .20$, schlechtem Itemfit (Infit > 1.2 oder $< .8$) oder nicht modellkonformer IC-Kurve wurden aus den weiteren Analysen ausgeschlossen (vgl. Tabelle 4-1).

Die Items wurden zudem hinsichtlich DIF nach Sprachregion, Geschlecht und zu Hause gesprochener Sprache (Testsprache: Ja/Nein) der Schülerinnen und Schüler überprüft. Die DIF-Analysen wurden mithilfe logistischer Regression (Zumbo 1999) mit dem R Package «sirt» (Robitzsch 2017) unter Berücksichtigung der latenten Personenfähigkeit (WLE) durchgeführt. Items mit signifikantem DIF grösser 0.64 oder kleiner -0.64 Logits (entspricht starkem DIF gemäss der ETS-Klassifikation, vgl. Longford et al. 1993) wurden von einer Expertengruppe aus Fachdidaktikern und Lehrpersonen dahingehend überprüft, ob die vermutete Ursache des DIF konstruktrelevant oder -irrelevant ist (Camilli 2006; Trendtel et al. 2016). Items mit vermutetem konstruktirrelevantem DIF wurden aus den weiteren Analysen ausgeschlossen (vgl. Tabelle 4-1).

Tabelle 4-1: Anzahl ausgeschlossene Items

Kompetenztest	180
Fehlerhaft	9
Schlechter FIT (Trennschärfe, MNSQ, ICC)	35
DIF	4
Skalierung	132

Für die Skalierung verbleiben 132 von insgesamt 180 Items im Itempool (vgl. Tabelle 4-2 und Tabelle 4-3).

Tabelle 4-2: Anzahl Items in der Skalierung nach Handlungsaspekten und -bereichen der Mathematik

		Kompetenzbereiche					Total
		Daten & Zufall	Grössen & Masse	Funktionale Zusammenhänge	Zahl & Variable	Form & Raum	
Handlungsaspekte	Argumentieren und Begründen	4	6	5	6	6	27
	Darstellen und Kommunizieren	2	3	8	4	4	21
	Wissen, Erkennen und Beschreiben	6	3	3	5	4	21
	Mathematisieren und Modellieren	7	6	10	1	5	29
	Operieren und Berechnen	2	2	6	14	10	34
	Total	21	20	32	30	29	132

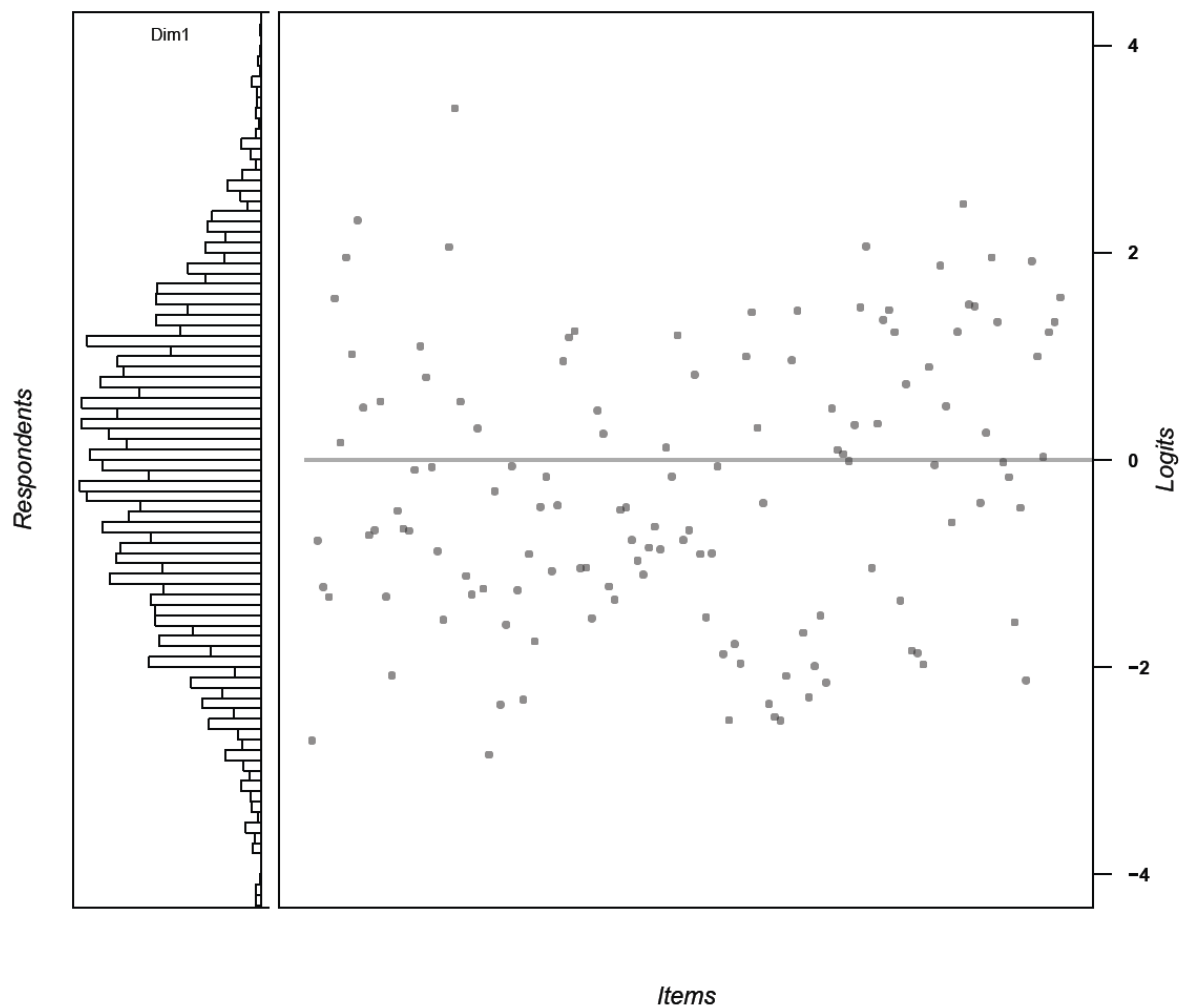
Tabelle 4-3: Anzahl Items in der Skalierung nach Itemformat und Handlungsaspekten und -bereichen der Mathematik

		Geschlossen	Halboffen	Offen	Total
Handlungsaspekte	Argumentieren und Begründen	2	0	25	27
	Darstellen und Kommunizieren	12	7	2	21
	Wissen, Erkennen und Beschreiben	15	6	0	21
	Mathematisieren und Modellieren	4	23	2	29
	Operieren und Berechnen	8	26	0	34
	Total	41	62	29	132
Kompetenzbereiche	Daten und Zufall	5	10	6	21
	Grössen und Masse	3	11	6	20
	Funktionale Zusammenhänge	10	16	6	32
	Zahl und Variable	15	9	6	30
	Form und Raum	8	16	5	29
	Total	41	62	29	132

4.3 Itemkalibrierung

Die Itemschwierigkeiten wurden anhand eines eindimensionalen Raschmodells geschätzt. Die Parameterschätzung erfolgte mit dem Marginal Maximum Likelihood Verfahren (MML) unter Berücksichtigung der Schülergewichte mit dem R-Paket «TAM» (Robitzsch et al. 2017), wobei der Mittelwert der Personenparameter auf null fixiert wurde. Die Verteilung der Item- und Personenparameter ist in Abbildung 4.2 dargestellt. Die Itemparameter sind im Anhang in Tabelle 6-1 ausgewiesen.

Abbildung 4.2: Verteilung der Item- und Personenparameter (RP = .62)



4.4 Plausible Value Imputation

Mit der ÜGK 2016 wurden Mathematikkompetenzen auf einer Globalskala und in fünf Handlungsaspekten und fünf Kompetenzbereichen der Mathematik erfasst. Für diese Dimensionen wurden Plausible Values als Schätzer für Kompetenzverteilungen gezogen. In das Plausible Values Ziehungsmodell gingen ausserdem acht Kovariaten² (Geschlecht, Schultyp, Spezielle Lernbedürfnisse, Höchste Ausbildung der Eltern, HISEI, Migrationsstatus, zu Hause gesprochene Sprache und Anzahl Bücher) sowie das Samplingstratum kodiert mittels Dummyvariablen ein. Da die Kovariaten teilweise auch fehlende Werte aufwiesen, wurde ein gemeinsames Imputationsmodell zur Behandlung fehlender Werte in den Kovariaten sowie der Ziehung der Plausible Values eingesetzt (vgl. Robitzsch et al. 2016).

Als Skalierungsmodell kam das eindimensionale Raschmodell zum Einsatz, mit dem für jedes Item eine Itemschwierigkeit bestimmt wurde (vgl. Kapitel 4.3). Diese Itemschwierigkeiten wurden in der Plausible Values Ziehung fixiert. Jedes Item wurde dabei genau einem Kompetenzbereich und einem Kompetenzaspekt zugeordnet. In das Imputationsmodell gingen die Stratumvariable S , Kovariaten X_1, \dots, X_p , die globale Mathematikkompetenz θ_G , die Kompetenzwerte für die fünf Kompetenzbereiche $\theta_{B1}, \dots, \theta_{B5}$ und die Kompetenzwerte für die fünf Handlungsaspekte $\theta_{A1}, \dots, \theta_{A5}$ ein. Das Imputationsmodell folgte dem Ansatz der chained equations, in dem alle Variablen mit fehlenden Werten iterativ aus bedingt eindimensionalen und mehrdimensionalen Modellen, gegeben alle anderen Variablen, imputiert werden (van Buuren 2012). Die Kovariaten wurden nacheinander imputiert. Anschliessend wurden für die 11 Kompetenzwerte (eine Globalskala Mathematik, fünf Kompetenzbereiche und fünf Handlungsaspekte) Plausible Values generiert. Die Imputationen wurden zyklisch in einer grossen Anzahl von Iterationen wiederholt.

Spezifikation der bedingt eindimensionalen Imputationsmodelle: Für die Imputation einer Kovariaten X_k wurden alle übrigen Prädiktoren sowie alle 11 Kompetenzwerte als Prädiktoren in einem Regressionsmodell verwendet. Zusätzlich zu den linearen Effekten wurden Interaktionseffekte aller Kovariaten mit den Kovariaten (sofern sie nicht imputiert werden) Geschlecht, Schultyp, spezielle Lernbedürfnisse, HISEI, Migrationsstatus, zu Hause gesprochene Sprache und globale Mathematikkompetenz einbezogen. Für die Imputation einer Kovariaten wurden für die linearen Effekte Prädiktoren mit einer minimalen absoluten Korrelation von .07 und für die Interaktionseffekte mit einer minimalen absoluten Korrelation von .12 ausgewählt. Die Mehrebenenstruktur der Clusterung von Schülerinnen und Schülern in Schulen wurde im Imputationsmodell repräsentiert, indem von allen Kovariaten und der globalen Mathematikkompetenz jeweils ein Schulmittelwert linear in die Regression aufgenommen wurde. Für die abhängige Variable (d.h. die zu imputierende Variable) wurde die Mehrebenenstruktur näherungsweise

² Kodierung der Kovariaten: Geschlecht (1 = Mädchen / 2 = Knaben), Schultyp (0 = Erweiterte- und Grundanforderungen / 1 = Hohe Anforderungen), Spezielle Lernbedürfnisse (0 = Keine spezielle Lernbedürfnisse / 1 = Spezielle Lernbedürfnisse), Höchste Ausbildung der Eltern (0 = Sek I / 1 = Sek II / 2 = Tertiär), HISEI, Migrationsstatus (0 = Einheimisch / 1 = 1./2. Generation), Zu Hause gesprochene Sprache (0 = andere Sprache / 1 = Testsprache) und Anzahl Bücher (1 = keine / 2 = 1 - 10 / 3 = 11 - 50 / 4 = 51 - 100 / 5 = 101 - 250 / 6 = 251 - 500 / 7 = mehr als 500). Die Angaben zu den speziellen Lernbedürfnissen, Geschlecht und Schultyp stammen aus den Sampling-Schülerlisten und die Angaben zu den Hintergrundmerkmalen aus dem Schülerfragebogen.

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 Regressionsmodell dadurch sehr viele zu schätzende Regressionskoeffizienten enthielt, wurde die Dimensionalität der Prädiktoren mit der Methode der Partial Least Squares Regression reduziert (vgl. Robitzsch et al. 2016). Dabei wurden 15 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 wurde aber auf wenige Faktoren reduziert. Die eigentliche Imputation einer Kovariaten erfolgte 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 der sechs Kovariaten (Schultyp, Höchste Ausbildung der Eltern, HISEI, Migrationsstatus, zu Hause gesprochene Sprache und Anzahl Bücher) mit fehlenden Werten eingesetzt.

Für die Plausible Value Imputation wurde für die globale Mathematikkompetenz ein eindimensionales Regressionsmodell (Hintergrundmodell) sowie für die fünf Kompetenzbereiche und fünf Handlungsaspekte jeweils ein mehrdimensionales Modell spezifiziert (latente Korrelationen: vgl. Tabelle 4-4). Die Konstruktion der Prädiktoren folgte dabei demselben Prinzip wie bei der Imputation von Kovariaten mit dem Unterschied, dass 25 anstelle von 15 Faktoren mit der Partial Least Squares Regression ermittelt wurden. Das eindimensionale Modell sowie die beiden mehrdimensionalen Modelle für die Plausible Values Ziehung wurden mit der Markov Chain Monte Carlo (MCMC) Schätzmethode des Gibbs Sampling mit der Funktion TAM::tam.pv.mcmc() im R-Paket «TAM» (Robitzsch et al. 2017) geschätzt. Dabei wurden die Itemparameter auf die Schätzungen aus einem Raschmodell fixiert. In dem MCMC-Schätzansatz erfolgte ein simultanes Sampling von Regressionsparametern (Regressionskoeffizienten und Kovarianz- und Varianzparameter). Dadurch ist die Unsicherheit in der Schätzung von Regressionskoeffizienten und Varianzen in den gezogenen Plausible Values abgebildet. Für die Ziehung eines Plausible Values der globalen Mathematikkompetenz gingen alle anderen Plausible Values der Kompetenzbereiche und Handlungsaspekte nicht als Prädiktoren ein, da diese Variablen Gruppen voneinander abhängen. Analog wurden bei der Ziehung der Plausible Values für die fünf Kompetenzbereiche die globale Mathematikkompetenz als auch die fünf Handlungsaspekte nicht als Prädiktoren verwendet. Für die Repräsentation der Mehrebenenstruktur in den Kompetenzskalen wurde jeweils der adjustierte Schulmittelwert der globalen Mathematikskala als Prädiktor aufgenommen.

Die Imputation von Kovariaten und Plausible Values wechselten sich im Imputationsmodell in 150 Iterationen ab. Die Iterationen 105, 110, ..., 150 stellten 20 Imputationen dar, die in den Analysen als Imputationsdatensätze verwendet wurden. Um die einzelnen Skalen der Kompetenzbereiche und Handlungsaspekte auf derselben Metrik wie der globalen Mathematikskala interpretieren zu können, wurden

die Plausible Values dieser Subdimensionen linear transformiert, so dass die jeweiligen Mittelwerte und Standardabweichungen mit dem Mittelwert und der Standardabweichung der globalen Mathematikskala übereinstimmen.

Tabelle 4-4: Latente Korrelationen zwischen den Handlungsaspekten und -bereichen der Mathematik

Handlungsaspekte	Argumentieren und Begründen	Darstellen und Kommunizieren	Wissen, Erkennen und Beschreiben	Mathematisieren und Modellieren
Argumentieren und Begründen				
Darstellen und Kommunizieren	.908			
Wissen, Erkennen und Beschreiben	.910	.860		
Mathematisieren und Modellieren	.939	.907	.862	
Operieren und Berechnen	.927	.885	.881	.933
Kompetenzbereiche	Daten und Zufall	Grössen und Masse	Funktionale Zusammenhänge	Zahl und Variable
Daten und Zufall				
Grössen und Masse	.849			
Funktionale Zusammenhänge	.901	.907		
Zahl und Variable	.826	.863	.884	
Form und Raum	.840	.917	.920	.900

5 Standardsetting

Im Anschluss an die Skalierung der Testdaten wurde unter der Leitung der Geschäftsstelle Aufgabendatenbank EDK (ADB) ein Standardsetting durchgeführt. Ziel des Standardsetting war es, auf der globalen Mathematikskala einen Schwellenwert festzulegen, der es erlaubt, die Schülerinnen und Schülern dahingehend einzustufen, ob sie die mathematischen Grundkompetenzen erreicht haben oder nicht. Der Schwellenwert wurde in einem Workshop von einer Expertengruppe anhand einer modifizierten Bookmark-Methode (Karantonis & Sireci 2006; Mitzel et al. 1999) bestimmt.

Die Expertengruppe bestand aus 13 Personen und setzte sich aus Fachdidaktikerinnen und Fachdidaktikern der Aufgabenentwicklung (5 Personen), externen Fachdidaktikerinnen und Fachdidaktikern, die nicht an der Aufgabenbeteiligung beteiligt waren (2 Personen), praktizierenden Mathematiklehrpersonen der Sekundarstufe I (5 Personen) und einer praktizierenden Berufsschullehrperson der Sekundarstufe II zusammen. 5 Personen der Expertengruppe kamen aus der Deutschschweiz und je 4 Personen aus der französisch- und italienischsprachigen Schweiz. Die Auswahl der externen Fachdidaktikerinnen und Fachdidaktikern und der Lehrpersonen erfolgte durch die Geschäftsstellen der EDK Regionalkonferenzen.

Zu Beginn des Standardsetting wurden die Expertinnen und Experten einen Vormittag lang geschult: Nach einer Einführung in die Überprüfung des Erreichens der Grundkompetenzen (ÜGK) in der Schweiz und der Standardsetting-Methode wurden die nationalen Bildungsstandards mit ihren «Cando-Deskriptoren» erläutert (vgl. EDK 2011b). Unklarheiten in den Begrifflichkeiten wurden in Kleingruppen und im Anschluss daran im Plenum diskutiert.

Anschliessend mussten die Experten in sprachregional zusammengesetzten Gruppen auf der Grundlage eines nach Itemschwierigkeiten gereihten Itembuchs, an der Stelle ein Lesezeichen (Bookmark) setzen, an der sie die Schwelle zwischen mathematischen Grundkompetenzen und höheren Kompetenzen vermuteten. Das Itembuch umfasste den gesamten Itempool. Pro Seite wurde ein Item dargestellt. Jede Seite enthielt den Itemtext mit dazugehörigen Abbildungen, die Itembezeichnung, die Lösung bzw. die Kodieranweisung, die Itemschwierigkeiten und die Seitenzahl.

Als Feedback wurden den Expertinnen und Experten eine Tabelle mit den gruppenspezifischen Schwellenwerten und den daraus resultierenden Konsequenzdaten zurückgemeldet. Die Bewertungen wurden im Plenum diskutiert und der Schwellenwert in einem Konsensentscheid bei Item «M_9_103_006» festgelegt. Dies entspricht auf der globalen Mathematikskala einem Wert von -0.421 Logits (RP = 62%). Tabelle 5-1 zeigt die Anzahl Items des Kompetenztests oberhalb und unterhalb des Schwellenwerts «Grundkompetenz» nach Handlungsaspekten und Kompetenzbereichen der Mathematik.

Tabelle 5-1: Anzahl Items des Kompetenztests unterhalb und oberhalb des Schwellenwert «Grundkompetenz» nach Handlungsaspekten und Kompetenzbereichen der Mathematik

Anzahl Items oberhalb Schwellenwert «Grundkompetenz»							
		Handlungsaspekte					Total
		Wissen, Erkennen und Beschreiben	Operieren und Berechnen	Darstellen und Kommunizieren	Mathematisieren und Modellieren	Argumentieren und Begründen	
Kompetenzbereiche	Zahl und Variable	3	8	4	0	6	21
	Form und Raum	1	5	3	5	6	20
	Grössen und Masse	3	1	0	6	6	16
	Funktionale Zusammenhänge	2	2	6	6	3	19
	Daten und Zufall	1	0	2	5	3	11
Total		10	16	15	22	24	87
Anzahl Items unterhalb Schwellenwert «Grundkompetenz»							
		Handlungsaspekte					Total
		Wissen, Erkennen und Beschreiben	Operieren und Berechnen	Darstellen und Kommunizieren	Mathematisieren und Modellieren	Argumentieren und Begründen	
Kompetenzbereiche	Zahl und Variable	2	6	0	1	0	9
	Form und Raum	3	5	1	0	0	9
	Grössen und Masse	0	1	3	0	0	4
	Funktionale Zusammenhänge	1	4	2	4	2	13
	Daten und Zufall	5	2	0	2	1	10
Total		11	18	6	7	3	45

6 Anhang

Tabelle 6-1: Itemparameter

Item Code	Handlungsaspekt	Kompetenzbereich	Format	Aufgaben- block	N ungewichtet	% korrekt ungewichtet	Xsi (RP = .62)
M_9_001_002	Wissen, Erkennen und Beschreiben	Funktionale Zusammenhänge	Halboffen	4	6811	89%	-2.221
M_9_004_003	Operieren und Berechnen	Zahl und Variable	Halboffen	8	6799	62%	-0.290
M_9_006_001_1	Wissen, Erkennen und Beschreiben	Daten und Zufall	Halboffen	9	6826	71%	-0.738
M_9_006_001_2	Wissen, Erkennen und Beschreiben	Daten und Zufall	Halboffen	9	6826	73%	-0.834
M_9_015_004	Argumentieren und Begründen	Zahl und Variable	Offen	3	6794	22%	2.048
M_9_018_001	Mathematisieren und Modellieren	Daten und Zufall	Halboffen	7	6839	46%	0.655
M_9_018_002	Argumentieren und Begründen	Daten und Zufall	Offen	7	6839	17%	2.444
M_9_018_005	Argumentieren und Begründen	Daten und Zufall	Offen	2	6829	31%	1.510
M_9_019_009	Argumentieren und Begründen	Zahl und Variable	Offen	4	6811	13%	2.802
M_9_027_001	Argumentieren und Begründen	Funktionale Zusammenhänge	Offen	5	6807	39%	0.994
M_9_030_001	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Halboffen	13	6766	62%	-0.238
M_9_030_006_1	Mathematisieren und Modellieren	Grössen und Masse	Halboffen	13	6766	61%	-0.190
M_9_030_006_2	Mathematisieren und Modellieren	Grössen und Masse	Halboffen	13	6766	39%	1.054
M_9_032_008	Operieren und Berechnen	Form und Raum	Geschlossen	8	6799	72%	-0.832
M_9_034_008	Mathematisieren und Modellieren	Daten und Zufall	Offen	6	6843	82%	-1.592
M_9_034_009	Mathematisieren und Modellieren	Daten und Zufall	Offen	6	6843	56%	-0.002
M_9_056_004	Argumentieren und Begründen	Form und Raum	Geschlossen	3	6794	62%	-0.176
M_9_057_004	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Geschlossen	13	6766	61%	-0.196
M_9_057_008	Argumentieren und Begründen	Form und Raum	Geschlossen	7	6839	51%	0.393
M_9_062_001	Darstellen und Kommunizieren	Funktionale Zusammenhänge	Geschlossen	5	6807	29%	1.586
M_9_062_004	Operieren und Berechnen	Form und Raum	Halboffen	7	6839	36%	1.287
M_9_071_001	Wissen, Erkennen und Beschreiben	Grössen und Masse	Geschlossen	13	6766	49%	0.420
M_9_073_004	Darstellen und Kommunizieren	Daten und Zufall	Geschlossen	6	6843	64%	-0.392
M_9_073_005	Darstellen und Kommunizieren	Form und Raum	Geschlossen	13	6766	76%	-1.055
M_9_074_004	Darstellen und Kommunizieren	Form und Raum	Offen	10	6827	17%	2.545
M_9_075_001	Operieren und Berechnen	Zahl und Variable	Halboffen	12	6811	6%	3.884
M_9_077_004	Operieren und Berechnen	Zahl und Variable	Halboffen	8	6799	39%	1.051
M_9_101_003	Operieren und Berechnen	Form und Raum	Halboffen	1	6810	70%	-0.631
M_9_101_005	Operieren und Berechnen	Zahl und Variable	Halboffen	6	6843	72%	-0.811
M_9_101_006	Operieren und Berechnen	Form und Raum	Halboffen	13	6766	43%	0.794
M_9_101_007	Operieren und Berechnen	Grössen und Masse	Halboffen	13	6766	72%	-0.756
M_9_101_008	Operieren und Berechnen	Zahl und Variable	Halboffen	10	6827	90%	-2.357
M_9_101_010	Mathematisieren und Modellieren	Form und Raum	Halboffen	1	6810	55%	0.187
M_9_102_002	Operieren und Berechnen	Funktionale Zusammenhänge	Halboffen	10	6827	85%	-1.874

Item Code	Handlungsaspekt	Kompetenzbereich	Format	Aufgaben- block	N ungewichtet	% korrekt ungewichtet	Xsi (RP = .62)
M_9_102_005	Operieren und Berechnen	Zahl und Variable	Halboffen	8	6799	76%	-1.102
M_9_102_007	Operieren und Berechnen	Funktionale Zusammenhänge	Halboffen	10	6827	50%	0.428
M_9_102_008	Operieren und Berechnen	Funktionale Zusammenhänge	Halboffen	12	6811	70%	-0.769
M_9_103_001	Operieren und Berechnen	Funktionale Zusammenhänge	Halboffen	10	6827	85%	-1.826
M_9_103_006	Operieren und Berechnen	Daten und Zufall	Halboffen	1	6810	66%	-0.421
M_9_103_010	Darstellen und Kommunizieren	Funktionale Zusammenhänge	Geschlossen	11	6802	78%	-1.264
M_9_104_001	Darstellen und Kommunizieren	Funktionale Zusammenhänge	Geschlossen	11	6802	58%	0.035
M_9_104_002	Operieren und Berechnen	Form und Raum	Geschlossen	3	6794	53%	0.326
M_9_104_004	Operieren und Berechnen	Funktionale Zusammenhänge	Halboffen	7	6839	68%	-0.585
M_9_104_005	Operieren und Berechnen	Funktionale Zusammenhänge	Halboffen	7	6839	57%	0.050
M_9_104_009	Wissen, Erkennen und Beschreiben	Zahl und Variable	Geschlossen	12	6811	32%	1.442
M_9_111_009_3	Wissen, Erkennen und Beschreiben	Funktionale Zusammenhänge	Geschlossen	11	6802	28%	1.673
M_9_112_001	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Halboffen	6	6843	27%	1.733
M_9_112_002	Operieren und Berechnen	Zahl und Variable	Halboffen	12	6811	67%	-0.557
M_9_112_003	Operieren und Berechnen	Form und Raum	Halboffen	8	6799	67%	-0.552
M_9_113_002	Mathematisieren und Modellieren	Zahl und Variable	Halboffen	13	6766	75%	-1.040
M_9_113_005	Operieren und Berechnen	Zahl und Variable	Halboffen	12	6811	40%	0.966
M_9_114_002	Darstellen und Kommunizieren	Funktionale Zusammenhänge	Geschlossen	12	6811	44%	0.741
M_9_121_001	Wissen, Erkennen und Beschreiben	Zahl und Variable	Geschlossen	2	6829	70%	-0.733
M_9_121_002	Wissen, Erkennen und Beschreiben	Zahl und Variable	Geschlossen	2	6829	72%	-0.861
M_9_121_003	Wissen, Erkennen und Beschreiben	Zahl und Variable	Geschlossen	3	6794	58%	0.007
M_9_121_004	Wissen, Erkennen und Beschreiben	Zahl und Variable	Geschlossen	3	6794	58%	0.030
M_9_121_009	Operieren und Berechnen	Zahl und Variable	Geschlossen	12	6811	62%	-0.285
M_9_122_002	Operieren und Berechnen	Zahl und Variable	Geschlossen	11	6802	66%	-0.483
M_9_122_004	Operieren und Berechnen	Zahl und Variable	Geschlossen	11	6802	67%	-0.619
M_9_122_006	Operieren und Berechnen	Zahl und Variable	Geschlossen	11	6802	64%	-0.358
M_9_124_002	Darstellen und Kommunizieren	Zahl und Variable	Geschlossen	2	6829	60%	-0.156
M_9_124_003	Darstellen und Kommunizieren	Zahl und Variable	Geschlossen	2	6829	63%	-0.375
M_9_124_004	Darstellen und Kommunizieren	Zahl und Variable	Geschlossen	4	6811	45%	0.608
M_9_124_007	Darstellen und Kommunizieren	Zahl und Variable	Geschlossen	4	6811	50%	0.328
M_9_140_002	Darstellen und Kommunizieren	Funktionale Zusammenhänge	Geschlossen	11	6802	27%	1.693
M_9_140_005	Operieren und Berechnen	Zahl und Variable	Geschlossen	12	6811	62%	-0.285
M_9_140_006	Operieren und Berechnen	Zahl und Variable	Geschlossen	12	6811	60%	-0.189
M_9_141_001	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Halboffen	5	6807	34%	1.311
M_9_141_002	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Geschlossen	5	6807	65%	-0.420
M_9_141_003	Darstellen und Kommunizieren	Funktionale Zusammenhänge	Geschlossen	5	6807	75%	-1.030
M_9_141_004	Darstellen und Kommunizieren	Funktionale Zusammenhänge	Halboffen	10	6827	65%	-0.413
M_9_141_007	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Halboffen	1	6810	50%	0.426
M_9_141_008	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Halboffen	1	6810	80%	-1.386
M_9_141_010	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Halboffen	4	6811	86%	-2.024

Item Code	Handlungsaspekt	Kompetenzbereich	Format	Aufgaben- block	N ungewichtet	% korrekt ungewichtet	Xsi (RP = .62)
M_9_141_011	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Halboffen	4	6811	78%	-1.289
M_9_141_012	Mathematisieren und Modellieren	Funktionale Zusammenhänge	Geschlossen	4	6811	81%	-1.477
M_9_150_003	Argumentieren und Begründen	Form und Raum	Offen	3	6794	32%	1.487
M_9_150_004	Argumentieren und Begründen	Form und Raum	Offen	8	6799	11%	1.915
M_9_150_005	Argumentieren und Begründen	Form und Raum	Offen	2	6829	30%	0.799
M_9_150_014	Argumentieren und Begründen	Form und Raum	Offen	5	6807	56%	0.072
M_9_151_001_2	Wissen, Erkennen und Beschreiben	Form und Raum	Geschlossen	2	6829	85%	-1.868
M_9_151_001_3	Wissen, Erkennen und Beschreiben	Form und Raum	Geschlossen	2	6829	87%	-1.993
M_9_151_003	Wissen, Erkennen und Beschreiben	Form und Raum	Geschlossen	11	6802	86%	-2.028
M_9_152_001_1	Operieren und Berechnen	Form und Raum	Halboffen	5	6807	82%	-1.597
M_9_152_001_2	Operieren und Berechnen	Form und Raum	Halboffen	5	6807	32%	1.451
M_9_152_002	Operieren und Berechnen	Form und Raum	Halboffen	10	6827	25%	1.929
M_9_160_001	Wissen, Erkennen und Beschreiben	Daten und Zufall	Halboffen	11	6802	78%	-1.180
M_9_160_002	Mathematisieren und Modellieren	Daten und Zufall	Halboffen	6	6843	84%	-1.802
M_9_160_006	Argumentieren und Begründen	Daten und Zufall	Offen	7	6839	81%	-1.502
M_9_161_003	Wissen, Erkennen und Beschreiben	Daten und Zufall	Geschlossen	1	6810	75%	-1.015
M_9_161_004	Wissen, Erkennen und Beschreiben	Daten und Zufall	Geschlossen	3	6794	83%	-1.661
M_9_161_005	Argumentieren und Begründen	Daten und Zufall	Offen	9	6826	38%	0.985
M_9_161_006	Darstellen und Kommunizieren	Daten und Zufall	Halboffen	13	6766	47%	0.585
M_9_162_001_1	Mathematisieren und Modellieren	Daten und Zufall	Halboffen	4	6811	47%	0.546
M_9_162_002	Mathematisieren und Modellieren	Daten und Zufall	Halboffen	4	6811	49%	0.477
M_9_170_001	Argumentieren und Begründen	Zahl und Variable	Offen	4	6811	41%	0.827
M_9_170_002	Argumentieren und Begründen	Zahl und Variable	Offen	12	6811	23%	1.962
M_9_170_003	Argumentieren und Begründen	Funktionale Zusammenhänge	Offen	11	6802	7%	2.550
M_9_170_004	Argumentieren und Begründen	Funktionale Zusammenhänge	Offen	5	6807	66%	-0.554
M_9_170_006	Argumentieren und Begründen	Funktionale Zusammenhänge	Offen	7	6839	9%	0.838
M_9_170_008	Argumentieren und Begründen	Zahl und Variable	Offen	9	6826	25%	1.841
M_9_170_010	Argumentieren und Begründen	Zahl und Variable	Offen	3	6794	23%	1.936
M_9_170_012	Argumentieren und Begründen	Grössen und Masse	Offen	6	6843	26%	1.724
M_9_170_014	Argumentieren und Begründen	Funktionale Zusammenhänge	Offen	9	6826	72%	-0.869
M_9_181_001	Wissen, Erkennen und Beschreiben	Funktionale Zusammenhänge	Halboffen	11	6802	36%	1.221
M_9_182_003_1	Darstellen und Kommunizieren	Grössen und Masse	Halboffen	10	6827	79%	-1.353
M_9_182_003_2	Darstellen und Kommunizieren	Grössen und Masse	Halboffen	10	6827	80%	-1.375
M_9_182_003_3	Darstellen und Kommunizieren	Grössen und Masse	Halboffen	10	6827	81%	-1.486
M_9_186_002	Argumentieren und Begründen	Grössen und Masse	Offen	13	6766	32%	1.385
M_9_186_008	Darstellen und Kommunizieren	Funktionale Zusammenhänge	Offen	5	6807	50%	0.441
M_9_190_001	Argumentieren und Begründen	Grössen und Masse	Offen	7	6839	18%	2.365
M_9_192_001	Wissen, Erkennen und Beschreiben	Form und Raum	Halboffen	3	6794	38%	1.008
M_9_200_001	Argumentieren und Begründen	Grössen und Masse	Offen	2	6829	59%	-0.114
M_9_200_002	Argumentieren und Begründen	Grössen und Masse	Offen	13	6766	27%	1.725

Item Code	Handlungsaspekt	Kompetenzbereich	Format	Aufgaben- block	N ungewichtet	% korrekt ungewichtet	Xsi (RP = .62)
M_9_200_003	Argumentieren und Begründen	Grössen und Masse	Offen	6	6843	11%	2.960
M_9_200_007	Wissen, Erkennen und Beschreiben	Grössen und Masse	Geschlossen	7	6839	24%	1.989
M_9_200_008	Wissen, Erkennen und Beschreiben	Grössen und Masse	Geschlossen	11	6802	24%	1.973
M_9_200_009	Mathematisieren und Modellieren	Grössen und Masse	Halboffen	2	6829	56%	0.074
M_9_200_010	Mathematisieren und Modellieren	Grössen und Masse	Halboffen	2	6829	43%	0.751
M_9_200_011	Mathematisieren und Modellieren	Grössen und Masse	Halboffen	10	6827	19%	2.444
M_9_200_013	Operieren und Berechnen	Grössen und Masse	Halboffen	1	6810	27%	1.821
M_9_200_014	Mathematisieren und Modellieren	Grössen und Masse	Halboffen	10	6827	52%	0.466
M_9_201_001	Wissen, Erkennen und Beschreiben	Daten und Zufall	Geschlossen	1	6810	52%	0.323
M_9_201_003	Operieren und Berechnen	Daten und Zufall	Halboffen	9	6826	75%	-1.079
M_9_201_004	Mathematisieren und Modellieren	Daten und Zufall	Geschlossen	9	6826	57%	0.026
M_9_201_005	Operieren und Berechnen	Form und Raum	Halboffen	1	6810	83%	-1.639
M_9_201_006	Darstellen und Kommunizieren	Form und Raum	Halboffen	13	6766	17%	2.408
M_9_201_007	Darstellen und Kommunizieren	Form und Raum	Halboffen	1	6810	32%	1.488
M_9_201_009	Mathematisieren und Modellieren	Form und Raum	Halboffen	8	6799	49%	0.517
M_9_201_010	Mathematisieren und Modellieren	Form und Raum	Halboffen	1	6810	29%	1.722
M_9_201_011	Mathematisieren und Modellieren	Form und Raum	Halboffen	10	6827	26%	1.821
M_9_201_012	Mathematisieren und Modellieren	Form und Raum	Halboffen	4	6811	23%	2.059

Literaturverzeichnis

- Camilli, G. (2006). Test fairness. In R. L. Brennan (Hrsg.), *Educational measurement* (S. 221–256). Westport CT: American Council on Education; Praeger Publishers.
- EDK. (2007). Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) vom 14. Juni 2007. https://edudoc.ch/record/24711/files/HarmoS_d.pdf. Zugegriffen: 22. Februar 2019.
- EDK. (2011a). Grundkompetenzen für die Fremdsprachen. Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16. Juni 2011. https://edudoc.ch/record/96780/files/grundkomp_fremdsprachen_d.pdf. Zugegriffen: 22. Februar 2019.
- EDK. (2011b). Grundkompetenzen für die Mathematik. Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16. Juni 2011. https://edudoc.ch/record/96784/files/grundkomp_math_d.pdf. Zugegriffen: 22. Februar 2019.
- EDK. (2011c). Grundkompetenzen für die Naturwissenschaften. Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16. Juni 2011. https://edudoc.ch/record/96787/files/grundkomp_nawi_d.pdf. Zugegriffen: 22. Februar 2019.
- EDK. (2011d). Grundkompetenzen für die Schulsprache. Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16. Juni 2011. https://edudoc.ch/record/96791/files/grundkomp_schulsprache_d.pdf. Zugegriffen: 22. Februar 2019.
- EDK. (2014). Organisationsreglement über die Durchführung der Überprüfung des Erreichens der Grundkompetenzen vom 8. Mai 2014, EDK. https://edudoc.ch/record/113128/files/Organisationsregl_UeGK_d.pdf. Zugegriffen: 22. Februar 2019.
- 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.
- Konsortium ÜGK. (2019). *Überprüfung der Grundkompetenzen. Nationaler Bericht der ÜGK 2016: Mathematik 11. Schuljahr*. Bern und Genf: EDK und SRED.
- 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). Hillsdale, NJ: Erlbaum.
- Robitzsch, A. (2017) sirt: Supplementary item response theory models [Computer software].

- Robitzsch, A., Kiefer, T. & Wu, M. (2017) TAM: Test analysis modules [Computer 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). Wien: facultas.
- Trendtel, M., Schwabe, F. & Fellingner, 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). Wien: facultas.
- van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton: CRC Press.
- 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*. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense.