

**Fundament zur Entwicklung eines Messinstruments zur grundlegenden  
Programmierkompetenz im Sachunterricht**

Inaugural-Dissertation  
zur  
Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Universität zu Köln  
vorgelegt von

Johanna Schlieben  
aus Frechen

Berichtersteller\*in:  
(Gutachter\*in)

Prof. Dr. Daniela Schmeinck

Prof. Dr. Jörg Großschedl

Tag der mündlichen Prüfung:

10.07.2024

## **Abstrakt**

Spätestens seit dem Beschluss der Kultusministerkonferenz im Jahr 2016 ist das Erlernen von Programmierinhalten von Kindern in der Grundschule ein gefordertes Ziel. In der aktuellen Forschung werden insbesondere Programmiermaterialien für den didaktischen Einsatz in der Grundschule diskutiert und ihre Wirksamkeit erprobt. Wenige Studien erforschen den Aufbau von Programmierkompetenzen im Primarbereich. Die vorliegende Dissertation nähert sich dieser Thematik durch die Entwicklung eines fundamentalen Messinstruments zur grundlegenden Programmierkompetenz im Sachunterricht. In diesem Kontext werden die Basisprogrammierbausteine Sequenz, Alternative und Iteration erforscht. Zur Entwicklung des Messinstruments wird dem methodischen Vorgehen nach Wilson (2005) gefolgt und es werden die Bausteine theoretisches Konstrukt, Itementwicklung, Ergebnisraum und Messinstrument dargelegt. In diesem Zusammenhang wird von der Autorin ein selbstentwickeltes theoretisches Konstrukt präsentiert. Darauf basierend erfolgt die Itementwicklung mit Hilfe einer Feldstudie, einer Vorstudie und einer Hauptstudie. Das Ergebnis bildet ein Itempool von 35 Items. Der Ergebnisraum definiert für die Items Kategorien und Bepunktungen. In diesem Rahmen werden zu den Themen Sequenz, Alternative und Iteration Items festgelegt. Die Überprüfung des Messmodells erfolgt mit Hilfe der Rasch-Analyse. Die Ergebnisse liefern eine gute Basis zur Entwicklung eines Messinstruments zur grundlegenden Programmierkompetenz im Sachunterricht. Aus diesem Ergebnis lassen sich weitere Schritte für die Testentwicklung ableiten. Für zukünftige Forschungen bietet sich eine Erweiterung des Itempools an, welcher das theoretische Konstrukt angemessen repräsentiert. Daran anschließend sollte der Itempool mittels quantitativer Studie überprüft werden. Zusätzlich sind weitere Validierungsstudien zu mathematischen Fähigkeiten, der Intelligenz oder dem Hörverstehen zu empfehlen. Hiermit können Überschneidungen und Diskrepanzen zu anderen Konstrukten herausgefunden werden. Nachfolgend kann eine Endversion des Tests als Basis für Interventionsstudien genutzt werden, um konkrete und individuelle Vorschläge für Programmierseinheiten abzuleiten.

**Abstract**

A required goal of the resolution of the Conference of Ministers of Education and Cultural Affairs in 2016 is to teach programming content already starting from elementary school. Current studies especially test programming tools in didactic context to prove their effectiveness. Just a few studies investigate the development of programming competencies in primary school. This dissertation wants to face the situation by developing a basic measurement model for fundamental programming competence in primary science and social science education. In this context, the basic programming building blocks of Sequence, Alternative, and Iteration are focused. To develop the measurement model, the methodological approach outlined by Wilson (2005) is followed, presenting the construct map, items design, outcome space, and measurement model as a result. In this context the author wants to present a self-developed construct map. Based on this, an item pool is carried out using a field study, a pilot study, and a main study. As a result, an item pool of 35 items can be presented. In the outcome space the items are categorized and scored. In this context the topics Sequence, Alternative, and Iteration will be classified. The measurement model is validated using Rasch analysis. The results provide a good basis for the development of a measurement model for fundamental programming competence in primary science and social science education. From this outcome further steps for the test development can be derived. For future research an extension of the item pool, which provides a good representation of the construct is recommended. Following this, the item pool should be validated through a quantitative study. Additionally, further validation studies on mathematical abilities, intelligence, or auditory comprehension should be done. These can help to identify overlaps and discrepancies to other constructs. Subsequently, a final version of the programming test can be used as a basis for intervention studies to derive concrete and individual suggestions for programming units.

## Danksagung

Zu Beginn möchte ich all denjenigen danken, die mich auf dem Weg meiner Dissertation begleitet haben. Ihr wart mir stets eine große Hilfe und Unterstützung! Mein besonderer Dank gilt Frau Prof.\*in Dr.\*in Daniela Schmeinck dafür, dass sie mir die Möglichkeit eröffnet hat, neue Wege zu gehen, und ich dadurch viele spannende, interessante und schöne Erfahrungen sammeln konnte. Weiterer Dank gilt ihrer Unterstützung, dem fachlichen Austausch und den produktiven Ideen zu meinem Forschungsprojekt. Auch möchte ich mich an dieser Stelle bei Prof. Dr. Jörg Großschedl für den hilfreichen statistischen Input und die effektiven Anregungen zu meinem Projekt bedanken. Darüber hinaus gilt allen Beteiligten der Graduiertenschule des ZfsL der Universität zu Köln unter der Leitung von Frau Dr. Gaby Schwager-Büschges und Frau Dr. Eva Becker meine Dankbarkeit für ihr Engagement, die zugeschnittenen Workshops, die individuellen Beratungsangebote und die hilfreichen Methodenberatungen. Als nächstes gilt mein Dank meinen Kolleg\*innen des Instituts für Didaktik des Sachunterrichts der Universität zu Köln. Danke für die vielen spannenden Diskussionen, den Zusammenhalt, die Unterstützung und den gegenseitigen Respekt. Es gibt viele Momente, an die ich mich sehr gerne zurückerinnern werde. Ihr wart mir fachlich und persönlich eine große Unterstützung. Zuletzt möchte ich meine große Dankbarkeit meiner Familie gegenüber zum Ausdruck bringen, welche mich auf vielfältige Weise unterstützt hat und immer für mich da war. In diesem Zusammenhang möchte ich besonders meinem geliebten Ehemann Oliver dafür danken, dass er mir so oft den Rücken freigehalten, an mich geglaubt und mir so viel Kraft und Motivation geschenkt hat. Danke für die vielen Stunden, die du mir für die Arbeit selbstverständlich ein- und freigeräumt hast. Danke, dass du mir immer eine so große Stütze bist, immer für mich da bist und ich mich immer auf dich verlassen kann! Zuletzt danke ich meiner über alles geliebten Tochter Mika. Danke, dass du so viel Geduld mit mir hattest und alles so wunderbar mitgemacht hast. Du hast mir in so vielen Situationen gezeigt, was wirklich wichtig ist.

**INHALTSVERZEICHNIS**

<u>ABKÜRZUNGSVERZEICHNIS</u>	<u>VI</u>
<u>ABBILDUNGSVERZEICHNIS</u>	<u>VII</u>
<u>TABELLENVERZEICHNIS</u>	<u>VIII</u>
<u>DIAGRAMMVERZEICHNIS</u>	<u>IX</u>
<u>1 ANLASS</u>	<u>1</u>
<u>2 FORSCHUNGSFRAGEN</u>	<u>6</u>
<u>3 GRUNDLAGEN UND METHODIK</u>	<u>9</u>
3.1 THEORETISCHES KONSTRUKT	9
3.1.1 ZWEIDIMENSIONALES KOMPETENZSTRUKTURMODELL (FACHBEITRAG 1)	9
3.1.2 DREIDIMENSIONALES KOMPETENZSTRUKTURMODELL (FACHBEITRAG 2.1)	13
3.2 ITEMENTWICKLUNG	16
3.2.1 FELDSTUDIE (FACHBEITRAG 2.2)	17
3.2.2 VORSTUDIE (FACHBEITRAG 3)	18
3.2.3 HAUPTSTUDIE	19
3.3 ITEMANALYSE	25
3.3.1 ROHWERTVERTEILUNG	25
3.3.2 ITEMSCHWIERIGKEITEN	26
3.3.3 ITEMtrennschärfe	28
3.3.4 DIMENSIONALITÄTSPRÜFUNG	29
3.4 WEITERE BERECHNUNGEN	30
3.4.1 RELIABILITÄT	30
3.4.2 HÄUFIGKEITSVERTEILUNGEN	30
3.5 ERGEBNISRAUM	31
3.6 MESSMODELL	32
3.6.1 KLASSISCHE TESTTHEORIE UND ITEM-RESPONSE-THEORIE	32
3.6.2 RASCH-ANALYSE	34
3.7 TESTVALIDIERUNG	37
3.7.1 HINFÜHRUNG	37
3.7.2 VALIDIERUNGSSTUDIE	41
3.7.3 KONSTRUKTVALIDITÄT	42
<u>4 ERGEBNISSE</u>	<u>45</u>
4.1 ITEMANALYSE	45
4.1.1 ROHWERTVERTEILUNG	45
4.1.2 ITEMSCHWIERIGKEIT	48

4.1.3	ITEMTRENNSCHÄRFE	51
4.1.4	DIMENSIONALITÄTSPRÜFUNG	53
4.2	WEITERE BERECHNUNGEN	54
4.2.1	RELIABILITÄT	54
4.2.2	HÄUFIGKEITSVERTEILUNGEN	55
4.3	ERGEBNISRAUM	63
4.4	MESSMODELL (RASCH-ANALYSE)	66
4.4.1	EINDIMENSIONALITÄT	67
4.4.2	PERSONEN- UND ITEMRELIABILITÄT SOWIE PERSONEN- UND ITEMSEPARATION	67
4.4.3	ZIELSETZUNG SOWIE BODEN- UND DECKENEFFEKTE	68
4.4.4	FUNKTIONSFÄHIGKEIT DES MESSINSTRUMENTS UND KONSTRUKTVALIDITÄT	69
4.5	TESTVALIDIERUNG	70
4.5.1	PRO SU 2 UND DEMAT 2+	71
4.5.2	PRO SU 2 UND CFT 1-R	74
4.5.3	PRO SU 2	77
<u>5</u>	<u>DISKUSSION DER ERGEBNISSE</u>	<u>79</u>
5.1	GRUNDLEGENDE PROGRAMMIERKOMPETENZ IM SACHUNTERRICHT	79
5.2	SCHWIERIGKEITSERZEUGENDE KRITERIEN	86
5.3	GEDREHTE ODER GESPIEGELTE ABBILDUNGEN	90
5.4	PROZEDURALES WISSEN	91
5.5	GESCHLECHTERUNTERSCHIEDE	92
5.6	KONSTRUKTÜBERSCHNEIDUNG MATHEMATISCHE FÄHIGKEITEN	94
5.7	KONSTRUKTÜBERSCHNEIDUNG INTELLIGENZ	95
<u>6</u>	<u>ZUSAMMENFASSUNG UND AUSBLICK</u>	<u>97</u>
	<u>LITERATUR</u>	<u>100</u>
	<u>ANHANG</u>	<u>109</u>
A	FACHBEITRÄGE	109
B	FELDSTUDIE	110
C	VORSTUDIE	113
D	HAUPTSTUDIE	118
E	BERECHNUNGEN DER RASCH-ANALYSE	140

## Abkürzungsverzeichnis

AL	Alternative
CFT 1-R	Culture Fair Intelligence Test, Grundintelligenztest: Skala 1
DEMAT 2+	Der Deutsche Mathematiktest für zweite Klassen
FF	Forschungsfrage
GDSU	Gesellschaft für Didaktik des Sachunterrichts
GI	Gesellschaft für Informatik
IRT	Item-Response Theorie
IT	Iteration
KMgPS	Kompetenzstrukturmodell zur grundlegenden Programmierkompetenz im Sachunterricht
KMK	Kultusministerkonferenz
KTT	Klassische Testtheorie
MNSQ	Mean-Square
MTMM	Multitrait-Multimethod
NRW	Nordrhein-Westfalen
PCAR	Principal Component Analysis of Residuals
PRO SU 2	Programmieren im Sachunterricht für zweite Klassen
SE	Sequenz

## Abbildungsverzeichnis

Abbildung 1: Erweitertes zweidimensionales KMgPS .....	11
Abbildung 2: Nomologisches Netzwerk .....	13
Abbildung 3: Dreidimensionales KMgPS.....	15
Abbildung 4: Kompetenzbeschreibungen des KMgPS.....	16
Abbildung 5: Übersicht über die durchgeführten Studien .....	17
Abbildung 6: Beispiel einer parallelen Testaufgabe .....	20
Abbildung 7: SEQUENZ Testinstruktion zu der Aufgabe 14.....	21
Abbildung 8: Person-Item-Map .....	35
Abbildung 9: MTMM-Matrix .....	39
Abbildung 10: Angepasste MTMM-Matrix.....	41
Abbildung 11: Ergebnisraum Sequenz.....	64
Abbildung 12: Ergebnisraum Alternative .....	65
Abbildung 13: Ergebnisraum Iteration.....	66
Abbildung 14: Person-Item-Map Sequenz.....	81
Abbildung 15: Person-Item-Map Alternative .....	83
Abbildung 16: Person-Item-Map Iteration.....	85

## Tabellenverzeichnis

Tabelle 1: Einordnung der Testitems in das KMgPS.....	21
Tabelle 2: Verteilung der Klassen auf den Sozialindex .....	23
Tabelle 3: Verteilung der Testpersonen auf die Testhefte .....	23
Tabelle 4: Aufschlüsselung der Testversionen zur Berechnung der Schwierigkeitsindexe.....	28
Tabelle 5: Itemschwierigkeiten .....	50
Tabelle 6: Trennschärfe.....	52
Tabelle 7: Itempool nach der Itemanalyse .....	54
Tabelle 8: Kennwerte des PRO SU 2 .....	57
Tabelle 9: Personen- und Itemreliabilität sowie Personen- und Itemseparation.....	68
Tabelle 10: Korrelationen zwischen dem PRO SU 2 (Version A) und dem DEMAT 2+ .....	71
Tabelle 11: Korrelationen zwischen dem PRO SU 2 (Version B) und dem DEMAT 2+.....	72
Tabelle 12: Vergleich zwischen dem PRO SU 2 (Version A) und PRO SU (Version B) mit dem DEMAT 2+ .....	73
Tabelle 13: Korrelationen zwischen dem PRO SU 2 (Version A) und dem CFT R-1 .....	75
Tabelle 14: Korrelationen zwischen dem PRO SU 2 (Version B) und dem CFT R-1.....	76
Tabelle 15: Vergleichende Korrelationen zwischen dem PRO SU 2 (Version A) und PRO SU (Version B) mit dem CFT 1-R .....	77
Tabelle 16: schwierigkeits erzeugende Merkmale Sequenz .....	87
Tabelle 17: schwierigkeits erzeugende Merkmale Alternative .....	88
Tabelle 18: schwierigkeits erzeugende Merkmale Iteration .....	89

## Diagrammverzeichnis

Diagramm 1: Rohwertverteilung der Sequenz .....	46
Diagramm 2: Rohwertverteilung der Alternative.....	47
Diagramm 3: Rohwertverteilung der Iteration .....	47
Diagramm 4: Häufigkeitsverteilung der Testwerte (Sequenz).....	58
Diagramm 5: Häufigkeitsverteilung der Testwerte (Alternative) .....	59
Diagramm 6: Häufigkeitsverteilung der Testwerte (Iteration).....	60
Diagramm 7: Häufigkeitsverteilung der Testwerte nach Geschlecht (Sequenz) .....	61
Diagramm 8: Häufigkeitsverteilung der Testwerte nach Geschlecht (Alternative).....	62
Diagramm 9: Häufigkeitsverteilung der Testwerte nach Geschlecht (Iteration) .....	63

## 1 Anlass

Aus informatischer und sachunterrichtlicher Sicht wird ein Kompetenzaufbau informatischer Inhalte im Sachunterricht befürwortet (Borowski et al., 2010; GDSU, 2021; GI, 2019; Goecke et al., 2017; Schmeinck, 2023; Straube et al., 2019). Spätestens seit dem Beschluss der Kultusministerkonferenz (KMK) im Jahr 2016 ist jedes Bundesland dazu verpflichtet Medienkompetenzen bereits in der Primarstufe aufzubauen. Das Strategiepapier der KMK (2016) definiert sechs Bereiche zum Aufbau von „Kompetenzen in der digitalen Welt“ (S. 16). Hierzu zählen neben dem Aufbau von Programmierkompetenzen unter dem Punkt „Problemlösen und Handeln“ im Unterpunkt „Algorithmen erkennen und formulieren“ fünf weitere Bereiche der Medienkompetenzförderung (KMK, 2016, S. 18). Diese sind „Suchen, Verarbeiten und Aufbewahren“, „Kommunizieren und Kooperieren“, „Produzieren und Präsentieren“, „Schützen und sicher Agieren“ und „Analysieren und Reflektieren“ (KMK, 2016, S. 16ff). Basierend auf dem Beschluss der KMK (2016) hat das Land Nordrhein-Westfalen (NRW) den Medienkompetenzrahmen NRW konzipiert (Medienberatung NRW, 2018), der ebenso aus sechs Kompetenzbereichen besteht, welche wiederum konkrete Kompetenzformulierungen enthalten. Im sechsten Bereich des Medienkompetenzrahmens NRW sind unter dem Punkt „Problemlösen und Modellieren“ informatische Kompetenzen zu finden. Die Förderung der definierten Medienkompetenzen soll in jedem Unterrichtsfach geschehen. Denn jedes Unterrichtsfach verfügt mit seinen fachspezifischen Herangehensweisen über einen Mehrwert zum bestmöglichen Aufbau von Medienkompetenzen (KMK, 2016). Das Fach Sachunterricht eignet sich besonders gut zur Förderung von Programmierkompetenzen, welches über das von Schmeinck (2018) geforderte Bedienen und Nutzen von digitalen Geräten wie beispielweise Smartphones hinaus geht. Denn charakteristisch für den Sachunterricht ist eine forschende, entdeckende und handlungsorientierte Auseinandersetzung mit dem Sachgegenstand (GDSU, 2013). Diese Herangehensweise bietet im Vergleich zu anderen Fächern den Vorteil, dass sich die Kinder mit dem Sachgegenstand „Computer“ und der Tätigkeit des „Programmierens“, also dem Aufbau und der Funktionsweise von Programmen, auseinandersetzen können (Borowski et al., 2010). Eine weitere zentrale Aufgabe des Sachunterrichts liegt darin, für Kinder zunächst unerklärliche Phänomene und Zusammenhänge aus ihrer Lebenswelt erfahrbar zu machen (GDSU, 2013). So gehören einerseits Programmierungen zu der Lebenswelt der Kinder, wie auch die Studie (2022) Kindheit, Internet, Medien (KIM) bestätigt. Denn Kinder verwenden im Alter von sechs bis dreizehn Jahren jeden bzw. fast jeden Tag den Fernseher (67 Prozent), das Handy/Smartphone (48 Prozent) oder das Internet (33 Prozent) (mpfs, 2023). Andererseits laufen die Programme im

Verborgenen ab und sind nicht direkt ersichtlich oder greifbar, sodass die Funktion der Handy- oder Fernsehsoftware für Kinder zunächst unerklärlich bleibt. Hier kann der Sachunterricht durch seine spezifische Herangehensweise ansetzen und ein sachbezogenes Verständnis bei den Schülerinnen und Schülern aufbauen. Über die Auseinandersetzung mit dem Sachgegenstand wird eine kompetente Teilhabe an der Gesellschaft eröffnet (GDSU, 2013; Schmeinck, 2023). Ein kompetentes Handeln in der Umwelt ist ebenso aus informatischer Sicht relevant, damit die Kinder in der Lebenswelt urteilsfähig sowie handlungs- und gestaltungsfähig teilhaben können. Dazu gehört vor allem zunehmend der Aufbau von informatischen Kompetenzen (GI, 2019). Damit Anknüpfungspunkte zur Förderung von informatischen Kompetenzen identifiziert werden können, ist ein Test hilfreich, welcher die Programmierkompetenzen von Grundschulkindern misst. Über die ermittelten Kompetenzen lassen sich geeignete Implikationen zur individuellen Förderung der Kinder ableiten. Dazu zählen insbesondere didaktische Entscheidungen hinsichtlich eines dienlichen Unterrichtsablaufs und eines passenden Programmiermaterials.

Studien bestätigen, dass die kognitiven Voraussetzungen bei Kindern vorliegen, um informatische Inhalte zu verstehen. So wertet Andreas Schwill (2001) verschiedene Studien zu den „fundamentalen Ideen der Informatik“ (Schwill, 2001, S. 14) hinsichtlich Rekursion, Greedy-Methode und Strukturierte Zerlegung aus. Die Rekursion, also ein sich unendlich wiederholender Vorgang, kann von Kindern in einem frühen Stadium genutzt werden. So können die Kinder rekursive Denkweisen ausbilden sowie rekursive Abbildungen verwenden. Bei der Greedy-Methode werden Schritt für Schritt Teillösungen genutzt, um zur Gesamtlösung zu gelangen. Dabei wird in jedem Schritt die günstigere Lösung gewählt, wodurch eine geringere Gedächtnisleistung benötigt wird. Bei dieser „fundamentalen Idee der Informatik“ (Schwill, 2001, S. 14) ist davon auszugehen, dass Kinder diese Methode ab einem Alter von 5 Jahren unbewusst anwenden. Weiter kommt Schwill (2001) zu dem Rückschluss, dass eine strukturierte Zerlegung, also die Problemlösestrategie durch Zerlegung des Problems in zielgerichtete Einzelschritte, von Kindern vorgenommen werden kann. In einer weiteren Studie von Gibson (2012) konnte gezeigt werden, dass Kinder, die weder schreiben noch lesen können, ab einem Alter von 5 Jahren in der Lage sind, grafisch dargestellte Algorithmen zu verstehen (Gibson, 2012). Eine weitere Studie mit 125 Kindern der dritten und vierten Klasse bestätigt, dass Kinder ein Verständnis von algorithmischen Strukturen haben. Die Studie hat gezeigt, dass Kinder verschiedene Instruktionen korrekt ausführen können (Weigend, 2009). In einem 5-wöchigen Workshop mit 4- bis 7-jährigen Kindern konnte ein Wissenszuwachs über

Programmierungen nachgewiesen werden (Bers, 2007). Insgesamt wird davon ausgegangen, dass Kinder aus der Primarstufe informatische Inhalte kognitiv verarbeiten, verstehen und umsetzen können (Straube et al., 2013). Diese Aspekte sind eine wesentliche Voraussetzung für den Aufbau und die Abfrage von Kompetenzen.

Im Bereich der Testentwicklung zum Thema Aufbau von Programmierkompetenzen in der Primarstufe, speziell im Sachunterricht, liegen bisher nur lückenhafte Ergebnisse vor. Eine wesentliche Studie von Hoffmann et al. (2017) soll folgend ausführlicher dargelegt werden. Die Studie beschreibt die Entwicklung von Aufgaben basierend auf drei Dimensionen. Die erste Dimension beinhaltet die grundlegenden Programmierbausteine: Sequenzielle Befehle, Alternativen und Iterationen. Die Inhalte der Programmierkompetenz werden wie folgt definiert (Schlieben, 2022, S. 356): „Die Sequenz wird als eine Abfolge von Befehlen verstanden. Die Alternative wird als eine Bedingung verstanden, die entweder wahr oder falsch sein kann (Boolescher Ausdruck) und dadurch bestimmt, welche von zwei Anweisungsfolgen im Anschluss ausgeführt wird. Die Iteration (Schleife) wird genutzt, um Anweisungen wiederholt auszuführen. Dabei wird zwischen Zählschleifen und Bedingten Schleifen unterschieden. Bei einer Zählschleife ist die Anzahl der vorzunehmenden Wiederholungen vorher numerisch festgelegt. Bei der Bedingten Schleife wird die Anzahl der Wiederholungen durch die Bedingung bestimmt, die zum Beenden des Schleifenprozesses erfüllt sein muss (Ficher & Hofer, 2011; GI, 2019)“. Die zweite Dimension beschreibt das Problem der Raumgröße und die dritte Dimension greift das Problem der Komplexität zur Lösung der Aufgabe auf. Die Ergebnisse zeigen, dass signifikante Leistungsunterschiede zwischen Kindern der zweiten und dritten Klasse bestehen. Besonders auffällig ist, dass bis auf ein Kind ( $n = 33$ ) der zweiten Klasse kein Kind die Aufgaben aus dem Bereich der Iteration lösen konnte. In der dritten Klasse waren es knapp ein Viertel der getesteten Kinder (16 von 66 Kindern), welche das richtige Ergebnis erzielten. Je höher die Komplexität der Aufgabe wurde, desto weniger Kinder konnten die Aufgaben richtig lösen. Hoffmann et al. (2017) kommen zu dem Schluss, dass das Konzept der Iteration erst ab der dritten Klasse angewendet und verstanden werden kann. Um mit Hilfe einer Interventionsstudie herauszufinden, inwieweit die Programmierelemente trainierbar sind, ist ein Test notwendig, welcher die Programmierelemente abfragt (Hoffmann et al., 2017). Jedoch liegen bisher nur wenige Studienergebnisse zum Aufbau von informatischen Kompetenzen bei Grundschulkindern vor (Bergner et al., 2018).

Aktuelle Studien beschäftigen sich vordergründig mit dem Einsatz von unterschiedlichen Programmiermaterialien in der Primarstufe. Forschungsergebnisse aus einer Studie mit 28 Kindern aus der ersten bis dritten Klasse zeigen, dass Kinder ein Interesse an Programmiermaterialien haben. Es zeichnet sich dabei eine geschlechterspezifische Auswahl der Programmiermaterialien ab (Straube et al., 2019). Eine weitere Studie untersucht über einen explorativen Zugang den kindlichen Umgang mit Programmiermaterialien. Hier konnten verschiedene Typen bei der Auseinandersetzung mit dem Material identifiziert werden (Goecke et al., 2017). Weiter konnte gezeigt werden, dass Spiele der Bewegungssteuerung Kindern geholfen haben, Algorithmierungen mithilfe von Befehlskärtchen darzustellen (Wagner, 2018). In einer Studie zum Programmieren mit der App Scratch Jr.® (Scratch Foundation Inc, 2014) wurde untersucht, welche Programmierblöcke Kinder (Kindergartenalter bis zweite Klasse) nach einer Lehrinheit wählen. Es hat sich gezeigt, dass Kinder unterschiedlichen Alters unterschiedliche symbolisch dargestellte Blöcke auswählen, um die Scratch Jr.® (Scratch Foundation Inc, 2014) Projekte zu bearbeiten (Portelance, 2016). Zusätzlich zu berücksichtigen sind Studien, welche die Einstellung der Grundschullehrkräfte hinsichtlich Programmiermaterialien untersuchen. Hier konnte Schmeinck (2020) zeigen, dass „die ausgewählten Tools, Softwareapplikationen und Programmiersprachen hinsichtlich ihrer Akzeptanz durch Grundschullehrkräfte [...] als grundsätzlich für den Einsatz in der Grundschule als geeignet angesehen werden“ (Schmeinck, 2020, S. 117). Jedoch zeigen die Ergebnisse weiter, dass „sowohl bei der Einstellungsakzeptanz als auch bei der Verhaltensakzeptanz gewisse Unterschiede seitens der Lehrkräfte hinsichtlich der Testmaterialien“ (Schmeinck, 2020, S. 119) vorliegen. Resultierend lässt sich konstatieren, dass das Lehren und Lernen von informatischen Unterrichtsinhalten aktuell ein unerschlossenes Forschungsfeld darstellt und großes Forschungspotential bietet (Magenheim & Romeike, 2019).

Insgesamt zeigen die Forschungsergebnisse, dass die kognitiven Voraussetzungen vorhanden sind, um Programmierkompetenzen bei Grundschulkindern aufzubauen. Jedoch werden Programmiermaterialien hinsichtlich ihrer Zweckdienlichkeit in der Primarstufe aktuell nur vordergründig untersucht. Ein Aufbau von Programmierkompetenzen basierend auf einem passenden Messinstrument steht zurzeit nicht im Fokus der Forschung. Daher kann ein entsprechender Programmierertest, welcher prozedurales Wissen abfragt, eine zusätzliche und hilfreiche Möglichkeit für Lehrkräfte sein, um informatische Kompetenzen bei Grundschulkindern zu identifizieren und um Kinder individuell zu fördern und zu fordern. An

dieser Stelle möchte die vorliegende Dissertation anknüpfen und ein Fundament zur Entwicklung eines Messinstruments zur grundlegenden Programmierkompetenz im Sachunterricht für zweite Klassen präsentieren. Dafür werden zu Beginn die Forschungsfragen und resultierenden Hypothesen beschrieben (Kapitel 2). Daran anschließend werden die Grundlagen und die Methodik der Testkonstruktion (Kapitel 3) darlegt. Dafür wird über die Zusammenfassung des Fachbeitrags 1 und Fachbeitrags 2.1<sup>1</sup> das theoretische Konstrukt vorgestellt (3.1). Daran anschließend werden die Itementwicklung (3.2), die Itemanalyse (3.3), weitere Kennwerte (3.4), der Ergebnisraum (3.5), das Messmodell (3.6) und die Testvalidierung (3.7) beschrieben. In dem Nachfolgenden Kapitel 4 werden die Ergebnisse der zuvor beschriebenen methodischen Schritte präsentiert. Darauffolgend werden in Kapitel 5 die Forschungsfragen und Hypothesen durch die Interpretation und Diskussion der Ergebnisse beantwortet. Abschließend folgt eine Zusammenfassung wesentlicher Ergebnisse und eine Darlegung des Ausblicks (Kapitel 6).

---

<sup>1</sup> Der Fachbeitrag 2 wird auf Grund von zwei unterschiedlichen inhaltlichen Schwerpunkten in dieser Arbeit in Fachbeitrag 2.1 und Fachbeitrag 2.2 unterteilt.

## 2 Forschungsfragen

Ziel des Forschungsprojektes ist es, ein Fundament für ein objektives, reliables und valides Messinstrument zu entwickeln, das die grundlegende Programmierkompetenz von Grundschulkindern der zweiten Klasse im Sachunterricht misst. Dafür wurden von der Autorin ein dreidimensionales Kompetenzstrukturmodell und darauf basierend ein Itempool entwickelt. Zum einen gilt es, herauszufinden, inwiefern Überschneidungen zu anderen Konstrukten wie mathematischen Inhalten oder der Intelligenz vorliegen. Zum anderen gilt zu überprüfen, ob Kinder aus der zweiten Klasse Kompetenzen zu den Programmierbausteinen Sequenz, Alternative und Iteration aufweisen. Darüber hinaus wird untersucht, ob geschlechterspezifische Unterschiede zwischen den jeweiligen Kompetenzen vorliegen. Aus diesen übergeordneten Zielsetzungen lassen sich folgende konkrete Forschungsfragen und Hypothesen formulieren.

**FF 1:** Lässt sich die *grundlegende Programmierkompetenz im Sachunterricht* von Kindern der zweiten Klasse objektiv, reliabel und valide messen?

**FF 2:** Lassen sich allgemeine schwierigkeiterzeugende Merkmale für Items in Bezug auf die Personenfähigkeit ableiten?  
Lassen sich gruppenspezifische Merkmale für Items wiederfinden, welche den a priori postulierten Kompetenzformulierungen des *Kompetenzstrukturmodells zur grundlegenden Programmierkompetenz im Sachunterricht (KMgPS)* entsprechen?

**FF 3:** Was bewirken gedrehte oder gespiegelte Abbildungen in den Items?

**Hypothese 1:** Aufgaben, welche der Leserichtung folgen sind von Kindern einfacher zu lösen.

**FF 4:** Verfügen Kinder der zweiten Klasse über prozedurales Wissen in Bezug auf die Basisprogrammierbausteine Sequenz, Alternative und Iteration?

**Hypothese 2:** Kinder der zweiten Klasse verfügen über prozedurales Wissen in Bezug auf die Sequenz.

**Hypothese 3:** Kinder der zweiten Klasse verfügen über prozedurales Wissen in Bezug auf die Alternative.

**Hypothese 4:** Kinder der zweiten Klasse verfügen über prozedurales Wissen in Bezug auf die Iteration.

**FF 5:** Wie unterscheiden sich die Teilgruppen hinsichtlich des Geschlechts in Bezug auf die Basisprogrammierbausteine Sequenz, Alternative und Iteration?

**Hypothese 5:** Jungen verfügen über ein höheres prozedurales Wissen als Mädchen in Bezug auf den Programmierbaustein der Sequenz.

**Hypothese 6:** Jungen verfügen über ein höheres prozedurales Wissen als Mädchen in Bezug auf den Programmierbaustein der Alternative.

**Hypothese 7:** Jungen verfügen über ein höheres prozedurales Wissen als Mädchen in Bezug auf den Programmierbaustein der Iteration.

**FF 6:** Lässt sich eine Konstruktüberschneidung zu mathematischen Fähigkeiten herstellen?

**Hypothese 8:** Es liegen diskriminante Validitäten zwischen der Sequenz und mathematischen Inhalten (DEMAT 2+<sup>2</sup>) vor.

**Hypothese 9:** Es liegen diskriminante Validitäten zwischen der Alternative und mathematischen Inhalten (DEMAT 2+) vor.

**Hypothese 10:** Es liegen konvergente Validitäten zwischen der Iteration und mathematischen Inhalten (DEMAT 2+) vor.

**FF 7:** Lässt sich eine Konstruktüberschneidung zur Intelligenz herstellen?

**Hypothese 11:** Es liegen diskriminante Validitäten zwischen der Sequenz und der fluiden Intelligenz (CFT 1-R<sup>3</sup>) vor.

**Hypothese 12:** Es liegen diskriminante Validitäten zwischen der Alternative und der fluiden Intelligenz (CFT 1-R) vor.

---

<sup>2</sup> Der Deutsche Mathematiktest für zweite Klassen (Krajewski et al., 2004)

<sup>3</sup> Grundintelligenztest: Skala 1 (Weiß & Osterland, 2013)

**Hypothese 13:** Es liegen diskriminante Validitäten zwischen der Iteration und der fluiden Intelligenz (CFT 1-R) vor.

Zur Beantwortung der FF1 werden die Ergebnisse aus der Hauptstudie, der Validierungsstudie und der Rasch-Analyse herangezogen. Dazu zählen die Testgütekriterien der Objektivität, der Reliabilität und der Validität. Die FF2 wird insbesondere durch die erstellten Person-Item-Maps und eine Analyse der Aufgaben beantwortet. Die FF3 wird unter dem Punkt der Itemschwierigkeiten (4.1.2) aufgegriffen und behandelt. Zur Beantwortung der FF4 und FF5 werden insbesondere die Tabellen der Häufigkeitsverteilungen resultierend aus der Hauptstudie betrachtet. Die FF6 und FF7 werden basierend auf den Ergebnissen der durchgeführten Validierungsstudie (4.5) beantwortet.

### **3 Grundlagen und Methodik**

In diesem Kapitel werden wesentliche theoretische Grundlagen und methodische Schritte der Testentwicklung dargelegt. Bei der Testentwicklung wird dem etablierten Entwicklungskreislauf nach Wilson (2005) gefolgt. Der Kreislauf besteht aus den vier Bausteinen theoretisches Konstrukt, Itementwicklung, Ergebnisraum und Messmodell. Die Realisierung der Testentwicklungsschritte beinhaltet zu Beginn die Darlegung des theoretischen Konstrukts mittels Zusammenfassung des Fachbeitrags 1 und des Fachbeitrags 2.1 (Kapitel 3.1). Anschließend wird die Itementwicklung (3.2) resümierend durch den Fachbeitrag 2.2 und den Fachbeitrag 3 vorgestellt. Im Zentrum dieser Betrachtung steht das methodische Vorgehen der Feldstudie, der Vorstudie und der Hauptstudie. Im nächsten Kapitel wird die Itemanalyse (3.3) durch die Schritte Rohwertverteilung, Itemschwierigkeit, Itemtrennschärfe und Dimensionalitätsprüfung dargelegt. Anknüpfend werden in Kapitel 3.4 weitere Berechnungen, wie die Reliabilität und Häufigkeitsverteilungen beschrieben. Darauf folgt die Skizzierung des Ergebnisraums (3.5). In diesem Kontext werden die Erwartungshorizonte der Testhefte und die vorgenommenen Bepunktungen beschrieben. Als Nächstes wird das methodische Vorgehen zur Bestimmung des Messmodells ausgeführt (3.6). Abschließend wird die Testvalidierung (3.7) beschrieben.

#### **3.1 Theoretisches Konstrukt**

In der Literatur existieren konkrete Ansätze zur Umsetzung des Entwicklungskreislaufes nach Wilson (2005). Die Realisierung des theoretischen Konstrukts wird zu Beginn in Kapitel 3.1.1 beschrieben, indem ein Grundgerüst präsentiert wird (Eggert & Bögeholz, 2014; Kleickmann et al., 2010; Mayer & Wellnitz, 2014). Kapitel 3.1.2 legt die Erweiterung des Kompetenzstrukturmodells dar.

##### **3.1.1 Zweidimensionales Kompetenzstrukturmodell (Fachbeitrag 1)**

Wenn keine Tests im zu beforschenden Bereich vorliegen, muss ein theoretisches Konstrukt definiert werden. Zur Darlegung theoretischer Konstrukte sind Kompetenzstrukturmodelle geeignet (Kauertz et al. 2010; Klieme, 2004; Schecker & Parchmann, 2006). Kompetenzstrukturmodelle definieren erstens für einen bestimmten Bereich verschiedene und charakteristische Dimensionen. Zweitens legen sie für jede Teildimension verschiedene Kompetenzniveaus fest. Die Niveaueprägungen beschreiben konkrete Kompetenzen, welche es zu erfüllen gilt (BMBF, 2003; Klieme, 2004; Klieme & Leutner, 2006; Mayer & Wellnitz,

2014). Zur Herleitung eines theoretischen Konstrukts ist eine umfassende Literaturrecherche notwendig (Eggert & Bögeholz, 2014; Wilson, 2005). Dafür gilt es folgende Aspekte zu berücksichtigen: den präskriptiven Rahmen, empirische Befunde, theoretische Grundlagen und das nomologische Netzwerk (Mayer & Wellnitz, 2014). Infolgedessen ergibt sich ein neu definiertes theoretisches Konstrukt, also eine nicht direkt beobachtbare (latente) Variable. Eine detaillierte Herleitung des theoretischen Konstrukts ist dem Fachbeitrag 1 und dem Fachbeitrag 2.1<sup>4</sup> zu entnehmen. Die Zusammenfassung der Beiträge findet sich nachfolgend.

### Fachbeitrag 1

Heinrichs, J. (2020). Programmieren im Sachunterricht. In K. Kaspar, M. Becker-Mrotzek, S. Hofhues, J. König & D. Schmeinck (Hrsg.), *Bildung, Schule, Digitalisierung* (S. 339–344). Münster: Waxmann. <https://doi.org/10.31244/9783830992462>

**Hintergrund:** Der erste Artikel greift den ersten Baustein der Testentwicklung auf, die Entwicklung des theoretischen Konstrukts (Wilson, 2005). In diesem Kontext werden über eine umfassende Literaturrecherche eine erste Strukturierung des theoretischen Konstrukts zur *grundlegenden Programmierkompetenz im Sachunterricht* vorgenommen und zentrale Definitionen festgelegt.

**Methode:** Die Literaturanalyse folgt dem Vorschlag von Mayer und Wellnitz (2014) und beschreibt den präskriptiven Rahmen (nationale und internationale Bildungsdokumente), empirische Befunde (Schüler:innenvorstellungen, Lernvoraussetzungen, Kompetenzausprägungen) und theoretische Grundlagen (Theorien, Modelle, Fachkonzepte). Zur Beschreibung des theoretischen Konstrukts gehört zusätzlich eine Darstellung des nomologischen Netzwerks, welches im Anschluss an die Zusammenfassung des Fachbeitrags 1 ergänzend dargelegt wird (Mayer & Wellnitz, 2016).

**Ergebnis:** Das Ergebnis bildet eine erste Strukturierung des theoretischen Konstrukts. Es wird ein Grundgerüst für das *Kompetenzstrukturmodell zur grundlegenden Programmierkompetenz im Sachunterricht* (KMgPS) (Heinrichs, 2020, S. 341) vorgestellt. Bezugnehmend auf die anhaltende wissenschaftliche Auseinandersetzung mit dem theoretischen Modell möchte die

---

<sup>4</sup> Siehe Fußnote 1

Autorin an dieser Stelle gerne ein erweitertes Kompetenzstrukturmodell (Abbildung 1) vorstellen.

### Abbildung 1

*Erweitertes zweidimensionales KMgPS*

<b>inhaltliche Dimension</b> <b>kognitive Dimension</b>	<b>Ausprägung 1</b>	<b>Ausprägung 2</b>	<b>Ausprägung 3</b>	<b>Ausprägung 4</b>
<b>Sequenz</b>				
<b>Alternative</b>				
<b>Iteration</b>				

*Anmerkung.* Angepasste Grafik aus Heinrichs 2020, S. 341.

Das erweiterte zweidimensionale KMgPS (Abbildung 1) wird durch die fachliche Dimension (Sequenz, Alternative, Iteration) und die Kognitive Dimension (vier Ausprägungsniveaus) aufgespannt. Das oben abgebildete Modell enthält zwei Änderungen zum Fachbeitrag 1: Die erste Spalte „grundlegende Programmierkompetenz“ (ursprünglich vor der Spalte inhaltliche und kognitive Dimension) wurde entfernt und nun im Titel durch die Bezeichnung *Kompetenzstrukturmodell zur grundlegende Programmierkompetenz im Sachunterricht* (KMgPS) aufgegriffen. Durch diese Modifikation soll hervorgehoben werden, dass das gesamte Kompetenzstrukturmodell die *grundlegende Programmierkompetenz im Sachunterricht* repräsentiert. Zweitens wird der Begriff *Konzept* durch die Begriffe *inhaltliche* und *kognitive Dimension* weiter ausdifferenziert. Durch diese Änderungen soll deutlich gemacht werden, welche Dimensionen im zukünftigen Kompetenzmodell (siehe Fachbeitrag 2.1) ihre Berücksichtigung finden. Zusätzlich ist an dieser Stelle anzumerken, dass der im ersten Artikel verwendete Begriff *Programmierkonzepte* durch den Begriff *Programmierbausteine* zu ersetzen ist. Im Laufe der Forschung hat sich herausgestellt, dass in diesem Zusammenhang der Begriff *Bausteine* die gängigere Bezeichnung ist.

**Ergänzung – Nomologisches Netzwerk:** Das nomologische Netzwerk zeigt Beziehungen zwischen dem zu messenden Konstrukt und anderen Konstrukten auf und ist entscheidend bei der Untersuchung der Konstruktvalidität (Hartig & Jude, 2007). Um das nomologische Netzwerk zu beschreiben, werden im Folgenden informatische Inhalte (Sequenz, Alternative und Iteration) mit mathematischen Inhalten und mit der Intelligenztheorie nach Cattell (1963) in Beziehung zueinander gesetzt. Dafür wird im Bereich der Mathematik auf die Bildungsstandards aus dem Primarbereich zurückgegriffen (KMK, 2005). Der Fokus liegt auf dem literaturbasierten Identifizieren von Überschneidungen und Abgrenzungen zu informatischen Inhalten (Sequenz, Alternative und Iteration<sup>5</sup>).

Zur Aufschlüsselung der mathematischen Kompetenzen wird der Beschluss der Kultusministerkonferenz der Länder (15.10.2004) herangezogen (KMK, 2005). Im Zentrum des Beschlusses werden fünf inhaltsbezogene mathematische Kompetenzen definiert. Die festgesetzten Themen sind: „Zahlen und Operationen“, „Raum und Form“, „Muster und Strukturen“, „Größen und Messen“ sowie „Daten, Häufigkeiten und Wahrscheinlichkeiten“ (KMK, 2005, S. 8). Die Themenfelder sind weiterhin aktuell und werden von der KMK in der Fassung vom 23.06.2022 ausdifferenziert dargelegt (KMK, 2022). Schwerpunkte aus den fünf mathematischen Bereichen, welche Überschneidungen zu den Programmierbereichen (Sequenz, Alternative, Iteration) aufweisen, werden im Folgenden kurz dargelegt. Im Bereich „Zahlen und Operationen“ ist eine Orientierung im Zahlenraum bis 100 (entsprechend der zweiten Klassenstufe) und ein Operationsverständnis zu nennen. Hier liegt eine Überschneidung zu den Inhalten der Iteration vor. Die Inhalte Alternative und Sequenz lassen sich im Bereich „Zahlen und Operation“ laut Kompetenzbeschreibungen nicht wiederfinden. In den Bereichen „Raum und Form“, „Größen“, „Muster und Strukturen“ sowie „Daten und Zufall“ sind keine überschneidenden Schwerpunkte zu den Inhalten der Sequenz, Alternative oder Iteration zu identifizieren (Abbildung 2).

Die Intelligenztheorie nach Cattell (1963) unterscheidet zwei Arten von Intelligenz: „fluide Intelligenz“ und „kristalline Intelligenz“. Die fluide Intelligenz bezeichnet biologische Voraussetzungen, welche die intellektuelle Entwicklung des Individuums beeinflussen. Es handelt sich dabei um angeborene, vererbte Eigenschaften. Die fluide Intelligenz ist die entscheidende zu messende Eigenschaft (Horn & Cattell, 1966). Die kristalline Intelligenz beschreibt allgemeine Fähigkeiten, welche erlernt werden und sich als „qualifizierte Urteilsgewohnheiten herauskristallisiert“ (Cattell, 1963, S. 2) haben. Es handelt sich um sprachliche und numerische Fähigkeiten sowie logisches Denken, welche im Laufe des Lebens

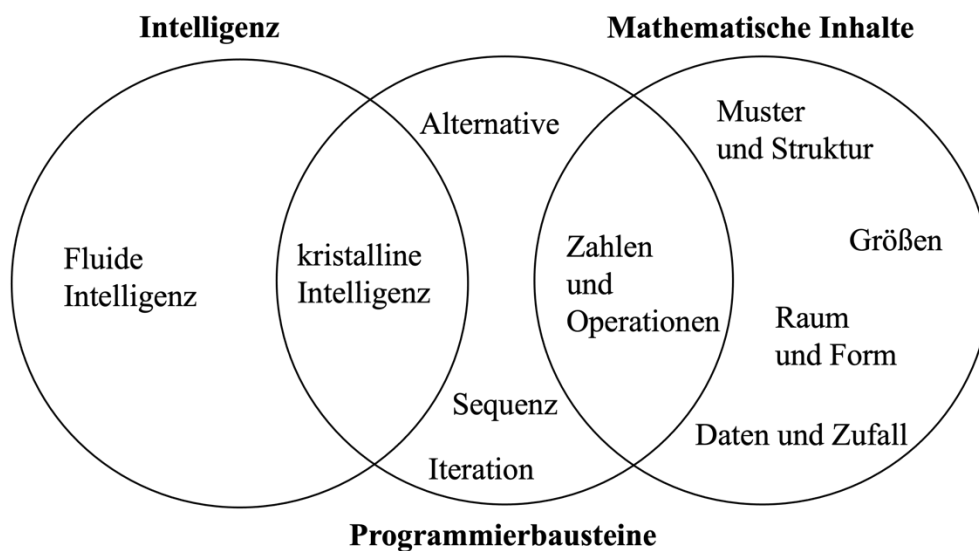
---

<sup>5</sup> Die Definitionen sind im Anhang 1 wiederzufinden.

durch erzieherische und kulturelle Einflüsse erlernt werden. In der Regel werden sie in der Schule vermittelt (Horn & Cattell, 1966; Weiß & Osterland, 2013). Es ist davon auszugehen, dass informatische Inhalte der Sequenz, Alternative und Iteration eine Überschneidung zu der kristallinen Intelligenz aufweisen (Abbildung 2). Denn im Laufe der Zeit werden informatische Fähigkeiten über die Schule, das Elternhaus oder die kulturellen gesellschaftlichen Prägungen vermittelt und erlernt. Von einem korrelativen Zusammenhang zwischen Programmierbausteinen zur fluiden, angeborenen Intelligenz ist nicht auszugehen. Es ist eine Überschneidung zwischen Programmierbausteinen und der kristallinen Intelligenz zu erwarten.

## Abbildung 2

*Nomologisches Netzwerk zwischen Programmierbausteine, Intelligenz und mathematischen Inhalten*



*Anmerkung.* Überschneidungen von Programmierbausteinen mit der Intelligenztheorie (Cattell, 1963) und den mathematischen Inhalten (KMK, 2005) befinden sich innerhalb der Schnittmengen.

### 3.1.2 Dreidimensionales Kompetenzstrukturmodell (Fachbeitrag 2.1)

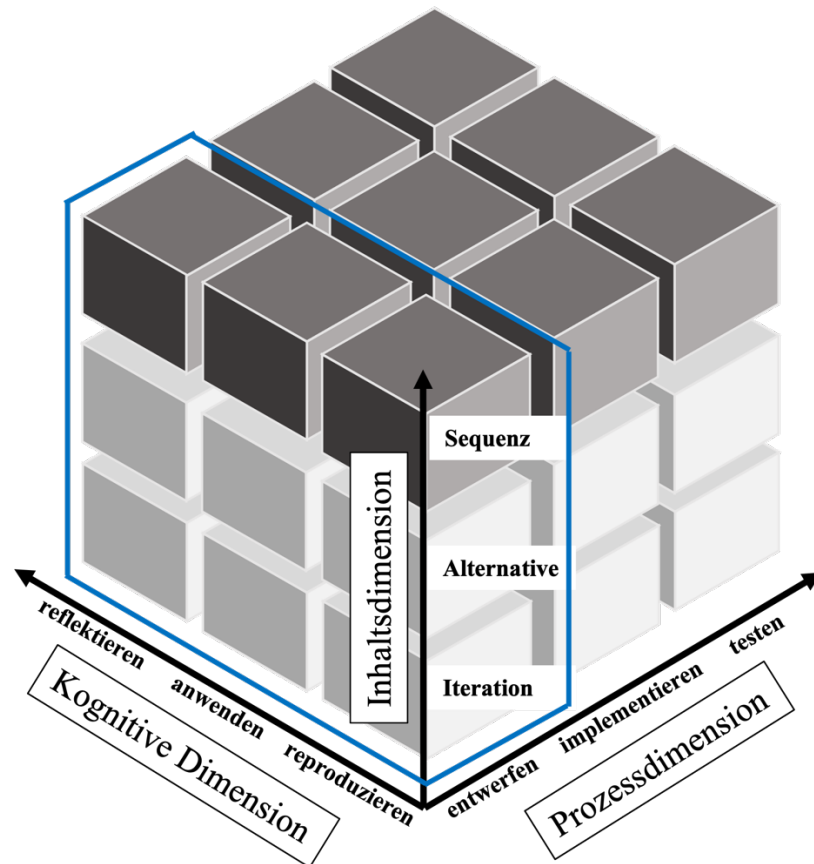
Im zweiten Artikel wird zum einen das theoretische Konstrukt (erster Baustein) finalisiert und zum anderen wird die Entwicklung eines grundlegenden Aufgabenpools (zweiter Baustein) dargelegt. Im folgenden Kapitel wird der erste Teil, die Finalisierung des theoretischen Konstrukts, zusammenfassend beschrieben (Schlieben, 2022, S. 353 – S. 358). Im Kapitel 3.2.1 wird der zweite Teil des Fachbeitrags zur Itementwicklung skizziert (Schlieben, 2022, S. 358 – S. 362).

Schlieben, J. (2022). Testinstrumentenentwicklung zur Messung grundlegender Programmierkompetenz im Sachunterricht. In B. Brandt, L. Böll, L. & H. Dausend (Hrsg.), *Digitales Lernen in der Grundschule III. Fachdidaktiken in der Diskussion* (S. 353–366). Münster: Waxmann. <https://doi.org/10.31244/9783830995913>

**Hintergrund:** Der erste Teil des Fachbeitrags 2 finalisiert das theoretische Konstrukt. Dafür wird das zweidimensionale Kompetenzstrukturmodell um die Prozessdimension erweitert sowie inhaltlich mit Kompetenzbeschreibungen gefüllt.

**Methode:** Auf Grundlage einer Literaturanalyse wird das theoretische Konstrukt finalisiert. Dafür werden unterschiedliche Kompetenzmodelle im Bereich der informatischen Bildung analysiert. Zur Festsetzung der inhaltlichen Dimension werden Definitionen literaturbasiert hergeleitet. Die Prozessdimension wird basierend auf den Programmentwicklungsprozess festgesetzt. Die Kognitive Dimension ist auf die Anforderungsbereiche der Bildungsstandards zurückzuführen. Darüber hinaus wird zu jedem Würfelement des Modells eine Kompetenzbeschreibung vorgenommen.

**Ergebnis:** Als Ergebnis wird ein dreidimensionales KMgPS (Abbildung 3) inklusive Kompetenzbeschreibungen (Abbildung 4) präsentiert. Hierbei beschreibt die Kognitive Dimension mit den Ausprägungen „reproduzieren“, „anwenden“ und „reflektieren“ das jeweilige Anspruchsniveau. Die Prozessdimension wird in die Handlungen „entwerfen“, „implementieren“ und „testen“ untergliedert und legt somit die prozessbezogenen Kompetenzerwartungen (das Handlungswissen) fest. Die Inhaltsdimension beschreibt das Fachwissen über die Sequenz, die Alternative und die Iteration. Die Definitionen der einzelnen Dimensionen sind dem Fachbeitrag 2 zu entnehmen (Schlieben geb. Heinrichs, 2022, S. 356f).

**Abbildung 3***Dreidimensionales KMgPS*

*Anmerkung.* Grafik aus Schlieben (geb. Heinrichs) 2022, S. 359.

Aus dem Zusammenschluss aller drei Dimensionen resultieren 27 Kompetenzbeschreibungen. Dabei werden die Kompetenzbeschreibungen für die Bedingte Schleife und die Zählschleife zusammen betrachtet (Iteration). Jedes Würfelement des Modells definiert eine Kompetenz. Abbildung 4 legt die Kompetenzbeschreibungen des KMgPS mit dem inhaltlichen Schwerpunkt der Sequenz dar. Es handelt sich hierbei um die in Abbildung 3 dunklen Würfelemente. Die Kompetenzbeschreibungen gelten auch substituierend für die Inhalte der Alternative und der Iteration. In dieser Dissertation wird für die Testentwicklung die Prozessdimension „entwerfen“ fokussiert, welche in Abbildung 4 blau umrandet ist.

**Abbildung 4***Kompetenzbeschreibungen des KMgPS mit dem inhaltlichen Schwerpunkt der Sequenz*

<b>Prozessdimension</b> <b>Kognitive Dimension</b>	<b>reproduzieren</b>	<b>anwenden</b>	<b>reflektieren</b>
<b>entwerfen</b>	Zu einem Sachverhalt oder einer Problemstellung wird ein Algorithmus (bestehend aus Sequenzen) entworfen, indem ein passender Algorithmus identifiziert wird.	Zu einem Sachverhalt oder einer Problemstellung wird ein Algorithmus (bestehend aus Sequenzen) entworfen, indem der Algorithmus vervollständigt wird.	Zu einem neuen Sachverhalt oder einer neuen Problemstellung wird ein Algorithmus (bestehend aus Sequenzen) entworfen, indem eigene Lösungswege gefunden werden.
<b>implementieren</b>	Ein Algorithmus (bestehend aus Sequenzen) wird implementiert, indem eine bekannte Entwurfsform in ein entsprechend bekanntes Programm übersetzt wird.	Ein Algorithmus (bestehend aus Sequenzen) wird implementiert, indem bekannte Entwurfsformen in neuen Programmen vervollständigt werden.	Ein Algorithmus (bestehend aus Sequenzen) wird implementiert, indem unbekannte Entwurfsformen flexibel in neue Programme übersetzt werden.
<b>testen</b>	Zu einem Sachverhalt oder einer Problemstellung wird ein Algorithmus (bestehend aus Sequenzen) getestet, indem Störungen und verzichtbare Anweisungen benannt werden.	Zu einem Sachverhalt oder einer Problemstellung wird ein Algorithmus (bestehend aus Sequenzen) getestet, indem Fehler behoben und Vereinfachungen vorgenommen werden.	Zu einem Sachverhalt oder einer Problemstellung wird ein Algorithmus (bestehend aus Sequenzen) getestet, indem Anpassungen beurteilt werden.

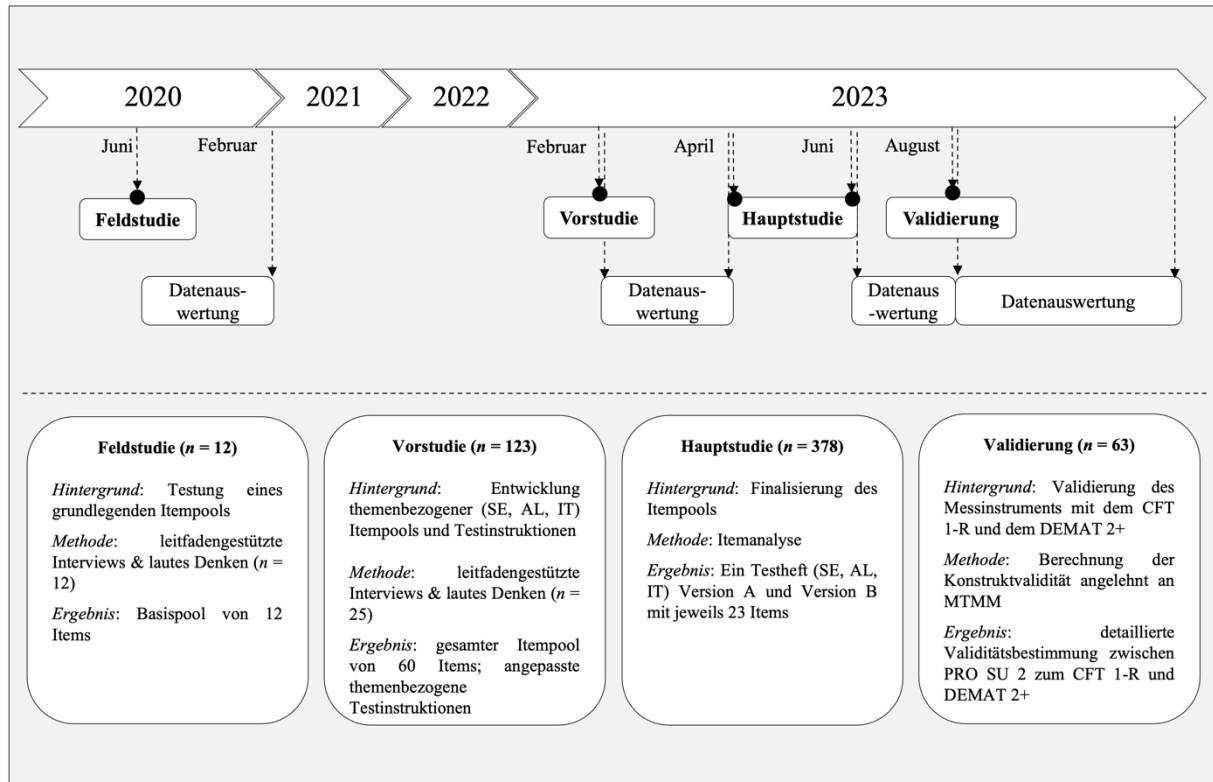
*Anmerkung.* Grafik aus Schlieben (geb. Heinrichs) 2022, S. 358.

**3.2 Itementwicklung**

Basierend auf dem vorgestellten theoretischen Konstrukt lassen sich folglich geeignete Items konzipieren (Kauertz et al., 2010; Klieme, 2004). Die Items bilden die Fähigkeiten der Probandinnen und der Probanden ab (Kauertz, 2014; Wilson, 2005). Es gilt, Situationen zu kreieren, in denen die Probandinnen und Probanden das beobachtbare (manifeste) Verhalten zeigen (Wilson, 2005). Dafür werden geeignete Testaufgaben benötigt. Wenn keine Testaufgaben vorhanden sind, welche für die Entwicklung des Tests genutzt werden können, müssen die Testaufgaben neu konzipiert werden (Eggert & Bögeholz, 2014; Kauertz, 2014). Hierfür eignet sich eine Feldstudie (Bergius, 2022), die im zweiten Teil des Fachbeitrags 2 (3.2.1) erläutert wird. Daran anschließend ist eine Vorstudie zu empfehlen, welche im Kapitel 3.2.2 als Zusammenfassung des Fachbeitrags 3 dargelegt wird. Abschließend wird das methodische Vorgehen der Hauptstudie (3.2.3) beschrieben. Eine Übersicht über den zeitlichen Ablauf der durchgeführten Studien und über die Zeiträume der Datenauswertung ist der Abbildung 5 zu entnehmen. Zusätzlich sind die Stichprobengrößen, die theoretischen Hintergründe, die Methoden und die Ergebnisse zu den einzelnen Studien in der Abbildung skizziert.

## Abbildung 5

## Übersicht über die durchgeführten Studien



Anmerkung. SE Sequenz, AL Alternative, IT Iteration.  $n$  = Größe der Stichprobe.

Die in der Dissertation ermittelten Primärdaten, ihre Analyse und die Reproduzierbarkeit der dargestellten Ergebnisse setzt die Verfügbarkeit von Datenmaterial voraus. Die im Rahmen der Dissertation gewonnenen Daten und Materialien sind entsprechend den Vorgaben des Instituts für Didaktik des Sachunterrichts der Universität zu Köln hinterlegt und über die Institutsleitung zugänglich.

### 3.2.1 Feldstudie (Fachbeitrag 2.2)

#### Fachbeitrag 2.2

Schlieben, J. (2022). Testinstrumentenentwicklung zur Messung grundlegender Programmierkompetenz im Sachunterricht. In B. Brandt, L. Böll, L. & H. Dausend (Hrsg.), *Digitales Lernen in der Grundschule III. Fachdidaktiken in der Diskussion* (S. 353–366). Münster: Waxmann. <https://doi.org/10.31244/9783830995913>

**Hintergrund:** Im zweiten Teil des Fachbeitrags 2 wird mittels einer Feldstudie<sup>6</sup> ein Basispool von Items entwickelt.

**Methode:** Zur Entwicklung von Testaufgaben wird zuerst das methodische Vorgehen zur Identifikation von schwierigkeiterzeugenden Kriterien in Aufgaben der Bundesweiten Informatikwettbewerbe (BWINF, 2023) beschrieben. Gemeinsam mit dem KMgPS werden im Kontext des Promotionsvorhabens 48 neu entwickelte Aufgaben zu den drei Inhalten (Sequenz, Alternative und Iteration) vorgestellt (Anhang: B Feldstudie). Anschließend wird die Feldstudie beschrieben, welche das Ziel hat, den Aufgabenpool zu testen und zu optimieren. Das Vorgehen beinhaltet die Durchführung von leitfadengestützten Interviews, welche mit der Methode des *Lauten Denkens* umgesetzt wird (Ericsson & Simon, 1993). Die Interviews werden videografiert und anschließend mit der Methode des *Zusammenfassenden Protokolls* (Trautmann, 2010; Gläser-Zikuda, 2015) ausgewertet.

**Ergebnis:** Basierend auf dem von der Autorin entwickelten KMgPS und den Ergebnissen der Feldstudie wird ein Basisitempool von zwölf Items zu der Prozessdimension „entwerfen“ präsentiert. Davon sind drei Items der Sequenz, drei Items der Alternative und sechs Items der Iteration zuzuordnen, wobei jeweils drei Items der Bedingten Schleife beziehungsweise der Zählschleife zuzuweisen sind.

### 3.2.2 Vorstudie (Fachbeitrag 3)

#### Fachbeitrag 3

Schlieben, J. (i. D.). Testinstrumentenentwicklung zur Messung grundlegender Programmierkompetenz im Sachunterricht – Eine Beschreibung des methodischen Vorgehens und der Ergebnisse aus der Vorstudie. In B. Brandt, L. Böll, L. & H. Dausend (Hrsg.), *Digitales Lernen in der Grundschule IV*. Waxmann.

**Hintergrund:** Der dritte Artikel behandelt weitere Schritte der Itementwicklung. Zu Beginn werden vier vorbereitende Schritte, welche vor der Vorstudie zu beachten sind, und ihre

---

<sup>6</sup> Innerhalb des Fachbeitrags 2 wurde der Begriff der Vorstudie verwendet, welcher nun durch den Begriff der Feldstudie zu ersetzen ist.

Ergebnisse dargelegt (Eggert & Bögeholz, 2014; Kauertz, 2014). Daran anschließend werden das methodische Vorgehen und die daraus resultierenden Ergebnisse der Vorstudie vorgestellt.

**Method:** Als vorbereitende, methodische Schritte für die Vorstudie werden eine Konzeption einer Anleitung für die Itementwicklung, eine Diskussion der Aufgaben mittels Expertisegremium, eine Formulierung der Testinstruktion und eine Erstellung von Leitfäden für Einzelinterviews (Anhang: C Vorstudie) präsentiert. Innerhalb der Vorstudie werden die Prozesse der Testdurchführung, wie die Bearbeitungsdauer und das eingesetzte Material, thematisiert. Zusätzlich werden weitere methodische Vorgehensweisen zur Optimierung der Testinstruktionen und der vier Testhefte (Sequenz, Alternative, Bedingte Schleife und Zählschleife) dargestellt. Dies beinhaltet eine Beschreibung des methodischen Vorgehens zur Durchführung und Auswertung von Einzelinterviews. Zusätzlich werden Analysemittel zur Auswertung der Testhefte dargelegt.

**Ergebnisse:** Es werden neue wichtige Erkenntnisse hinsichtlich der Formulierungen in den Testinstruktionen, den Testheften und für die Testdurchführung präsentiert. Die Ergebnisse werden als Grundlage zur Optimierung der Testhefte und der Testinstruktionen für die anschließende Hauptstudie<sup>7</sup> verwendet (3.2.3).

### 3.2.3 Hauptstudie

Bevor die Hauptstudie durchgeführt wird, erfolgt basierend auf den Ergebnissen der Vorstudie und im Sinne der Testökonomie (Bortz & Döring, 2016; Lienert & Raatz, 1998; Moosbrugger & Kelava, 2020) eine Optimierung der Testhefte und der Testinstruktionen. Folgende Änderungen werden im Testheft vorgenommen, um eine einfachere Handhabung zu ermöglichen (Bortz & Döring, 2016). Die Einführung des Spielfeldes und der Symbole zu Beginn der Testhefte wird entgegen der ursprünglichen Planung nicht umgesetzt. Alternativ wird vor jedem neuen Aufgabenformat eine Beispielaufgabe eingefügt (Moosbrugger & Kelava, 2020). Zusätzlich werden kurze Arbeitsanweisungen verwendet. Diese beginnen mit den Operatoren *ankreuzen*, *anmalen*, *nutzen* oder *einsetzen*. Lange Textpassagen werden von den Testheften in die Testinstruktionen überführt und es wird auf eine klare, übersichtliche Aufmachung des Testlayouts geachtet (Kauertz, 2014). Das Testheft der Bedingten Schleife und der Zählschleife wird aus inhaltlichen Gründen zu einem Testheft zusammengefügt. Ein

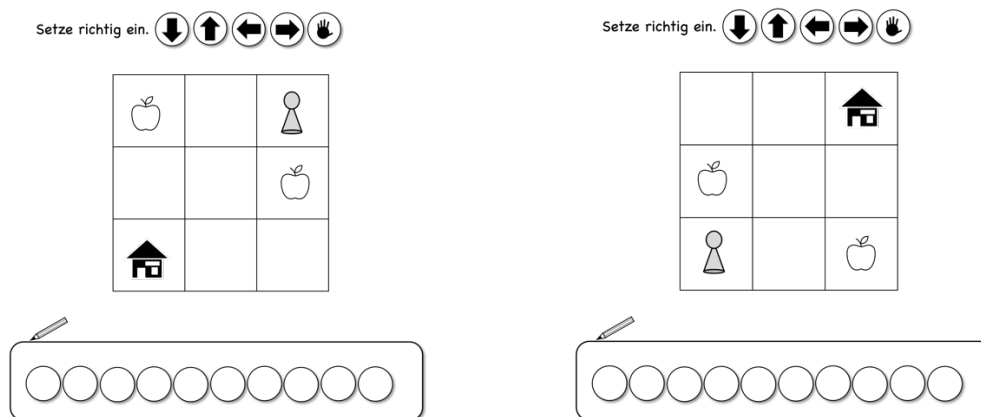
---

<sup>7</sup> Der in dem Artikel verwendete Begriff der Pilotierung wird fortan durch den gängigeren Begriff der Hauptstudie ersetzt.

weiterer Punkt der Testökonomie beinhaltet die Konzeption eines Gruppentests (Bortz & Döring, 2016). Aus diesem Grund wird eine parallele Testversion B zur Minimierung von Abschreibeffekten für alle drei Testhefte entwickelt (ebd., 2016). Dafür werden in der Testversion B der Sequenz und Iteration gedrehte oder gespiegelte Abbildungen verwendet (Abbildung 6).

## Abbildung 6

### SEQUENZ Beispiel einer parallelen Testaufgabe



*Anmerkung.* Die Abbildung zeigt die Aufgabe 14 aus dem Testheft der Sequenz Version A (links) und der Sequenz Version B (rechts). Das gesamte Spielfeld ist bei Version B um 180 Grad gedreht dargestellt.

Zusätzlich werden in einigen Aufgaben das Start- und Zielfeld oder die Symbolfelder, wie zum Beispiel das Apfel-Feld mit dem Blumen-Feld, vertauscht. Lediglich bei dem Testheft der Alternative wird ein echter Pseudo-Paralleltest genutzt. Die Abbildungen sind identisch und es werden nur andere Symbole verwendet. Bei dem Testheft der Sequenz und der Iteration befinden sich die Antwortmöglichkeiten der Multiple-Choice-Aufgaben an unterschiedlichen Positionen. Bei den halboffenen und offenen Aufgaben wird eine parallele Testaufgabe verwendet<sup>8</sup>. So unterscheiden sich die Positionen der Symbole auf dem Feld oder das Feld an sich ist in der parallelen Testform gedreht oder gespiegelt. Die Lösungsfelder sind bei den offenen Aufgabenformaten identisch. Die Kinder zeichnen beispielsweise in leere Kreise die passenden Befehle zur Steuerung der Figur vom Start zum Ziel. Bei den halboffenen Aufgaben sind die Lösungsfelder passend zu den Abbildungen. Die Beispielaufgaben vor jedem neuen Aufgabenformat sind in allen Testheften gleich. Die Testinstruktionen werden entsprechend der Ergebnisse aus der Vorstudie hinsichtlich adressatengerechter Sprache, wiederkehrender

<sup>8</sup> Dieses Vorgehen wirkt sich auf die Berechnungen der Itemanalyse und des Messmodells aus.

Aufforderungen, der Betonung von relevanten Wörtern und des Zeigens auf wichtige Aspekte angepasst. Ein passendes Beispiel aus der Testinstruktion zu der Aufgabe 14 der Sequenz (Abbildung 6) ist der Abbildung 7 zu entnehmen.

### Abbildung 7

#### *SEQUENZ Testinstruktion zu der Aufgabe 14*

Du hast nun das Blatt mit der **Aufgabe 14** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
 Die Figur geht **einen** schnellen Weg in das Haus.  
 Auf dem Weg hebt die Figur **alle Äpfel** (*auf den Befehl zeigen*) auf.  
 Achtung: Dafür gibt es einen neuen Befehl.  
 Setze die richtigen Befehle (*auf die Lösung zeigen*) ein.  
 Es können auch Kreise leer bleiben (*zeigen*).  
 Nutze: nach unten, nach oben, nach rechts, nach links, aufheben (*zeigen*).  
 Du kannst die Befehle mehrmals benutzen.  
 Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
 Blättere nun um.

*Anmerkung.* Die fett gedruckten Begriffe werden betont vorgelesen. Die kursiv geschriebenen Wörter geben der Testleitung Instruktionen.

Aus den zuvor beschriebenen vorbereitenden Schritten ergeben sich für das Testheft der Sequenz (Version A und B) jeweils 15 Items, für das Testheft der Alternative (Version A und B) und der Iteration (Version A und B) jeweils 13 Items (Tabelle 1). Der gesamte Itempool umfasst 60 Items. In diesem Kontext ist zu beachten, dass alle verschiedenen Aufgaben gezählt werden. Somit liegen für die Sequenz (A und B) 25 unterschiedliche Items, für die Alternative (A und B) 13 unterschiedliche Items und für die Iteration (A und B) 22 unterschiedliche Items vor (Anhang: D Hauptstudie).

### Tabelle 1

#### *Einordnung der Testitems in das KMgPS*

Testheft	Aufgabenniveau		
	reproduzieren	anwenden	reflektieren
Items - Sequenz	1 bis 5	6 bis 10	11 bis 15
Items - Alternative	1 bis 7	8, 9	10 bis 13
Items - Iteration	1 bis 4, 8 bis 10	5, 6, 11, 12	7, 13

Innerhalb des Testhefts der Sequenz (Version A und B) repräsentieren die Items 1 bis 5 die Kognitive Dimension „reproduzieren“, die Items 6 bis 10 die Kognitive Dimension „anwenden“ und die Items 11 bis 15 die Kognitive Dimension „reflektieren“. Im Testheft der Alternative sind die Aufgaben 1 bis 7 der Kognitiven Dimension des Reproduzierens, die Items 8 und 9 dem Kognitiven Bereich des Anwendens und die Aufgaben 10 bis 13 dem Kognitiven Bereich des Reflektierens zuzuordnen. Die Items aus dem Testheft der Iteration (Version A und B) verteilen sich folgendermaßen: Die Items 1 bis 4 und 8 bis 10 repräsentieren den Kognitiven Bereich des Reproduzierens, die Items 5, 6, 11 und 12 den Kognitiven Bereich des Anwendens und die Items 7 und 13 den Kognitiven Bereich des Reflektierens (Tabelle 1). Darüber hinaus wird darauf geachtet, dass aus testökonomischen Gründen der Test innerhalb einer Schulstunde (45 Minuten) zu bewältigen ist (Bortz & Döring, 2016; Eid & Schmidt, 2014; Lienert & Raatz, 1998). Die Testzeit der Aufgaben richtet sich danach, ob ein Powertest oder ein Speedtest vorliegt. Bei einem Speedtest werden leicht zu bearbeitende Aufgaben gestellt. Der Speedtest prüft, wie viele Aufgaben in einer vorgegebenen Zeit geschafft werden. Ein Powertest enthält leichte bis schwierige Aufgaben, welche ohne Zeitbeschränkung bearbeitet werden (Eid & Schmidt, 2014; Lienert & Raatz, 1998; Moosbrugger & Kelava, 2020; Weise, 1975). Für diese Testentwicklung werden die Fähigkeiten der Kinder mittels Powertest erhoben. Aus diesem Grund haben die Kinder ausreichend Zeit, um die Testaufgaben zu bearbeiten. Es wird zusätzlich darauf geachtet, dass jedes Kind die Aufgabe zu Ende bearbeiten kann. Zudem werden für eine schnelle und angenehme Testauswertung drei Auswertungsschablonen erstellt (Bortz & Döring, 2016).

**Eichstichprobe:** Die Hauptstudie wurde Ende April bis Mitte Juni 2023 an neun Grundschulen in Köln durchgeführt. Bei der Wahl der Eichstichprobe wurde darauf geachtet, dass die zu testende Population angemessen repräsentiert wird (Bortz & Döring, 2016; Lienert & Raatz, 1998). Die Wahl der zweiten Schuljahre erfolgte zufällig. Von über 100 angeschriebenen Schulen konnten 20 Klassen das Projekt unterstützen. Die Schulen lagen in acht unterschiedlichen Stadtteilen (Ehrenfeld, Sülz, Meschenich, Zollstock, Riehl, Lindental, Bayenthal und Buchforst). Um die soziale Verteilung der beteiligten Klassen zu verdeutlichen, wird eine Einteilung der Klassen in Sozialindexe vorgenommen (Tabelle 2). Der Sozialindex setzt sich aus vier Bereichen zusammen: Anteil der Schüler und Schülerinnen mit Deutsch als Zweitsprache, mit einem Zuzug aus dem Ausland, mit einem Förderschwerpunkt Lernen, emotionale und soziale Entwicklung und/oder Sprache sowie dem Punkt Kinder- und Jugendarmut. Der Sozialindex kann einen Wert von 1 bis 10 annehmen, wobei 1 eine geringe

soziale Herausforderung charakterisiert (Schräpler & Jeworutzki, 2021). In Kölner Grundschulen verteilen sich die Grundschulen auf die Indizes 1 bis 8 (MSB, o.J.). Die Betrachtung der Schulsozialindizes für alle Kölner Grundschulen ergibt einen Mittelwert von  $M = 3.29$ , eine Standardabweichung von  $SD = 1.74$  und eine Varianz von  $V = 3.02$ .

**Tabelle 2**

*Verteilung der Klassen auf den Sozialindex*

Sozialindex								
$x_i$	1	2	3	4	5	6	7	8
$H_{x_i}$	6	4	5	0	1	4	0	0

*Anmerkung.*  $x_i$  = Sozialindex 1 bis 8, wobei ein hoher Sozialindex eine hohe soziale Herausforderung definiert.  $H_{x_i}$  = absolute Häufigkeit;  $M = 2.9$ .

Die Berechnungen des Schulsozialindex für die teilgenommenen Schulen ergeben einen Mittelwert von  $M = 2.9$ , eine Standardabweichung von  $SD = 1.89$  und eine Varianz von  $V = 3.57$ , was für eine gute Repräsentation der beteiligten Kölner Grundschulen im Vergleich zu allen Kölner Schulen spricht. An der Studie haben insgesamt  $N = 378$  Kinder teilgenommen und es wurden  $n = 378$  Testhefte ausgewertet. Es wurde keine Unterscheidung zwischen Kindern mit oder ohne Förderschwerpunkt vorgenommen. Innerhalb einer Klasse haben die Kinder immer ein Themenheft in der Version A oder B bearbeitet. Die Verteilung der Kinder auf die Themenhefte (Sequenz, Alternative, Iteration) sowie die Unterteilung in Testversion A und B lässt sich der folgenden Tabelle 3 entnehmen.

**Tabelle 3**

*HAUPTSTUDIE Verteilung der Testpersonen auf die Testhefte A und B*

	Sequenz	Alternative	Iteration
Testversion	$n$	$n$	$n$
A	70	60	61
B	64	65	58

*Anmerkung.*  $n$  = Anzahl der Probandinnen und Probanden in Bezug auf die Testhefte (Sequenz, Alternative und Iteration) und die jeweilige Testversion (A oder B).

Das Testheft der Sequenz (Version A) haben 70 Kinder und der Version B 64 Kinder bearbeitet. Das Testheft der Alternative haben insgesamt 125 Kinder bearbeitet. An der Testung der Iteration haben 61 Kinder (Version A) und 58 Kinder (Version B) teilgenommen.

**Objektivität:** Die Objektivität eines Tests wird definiert als der „Grad, in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind“ (Lienert & Raatz, 1998, S. 7). Es wird zwischen der Durchführungs-, Auswertungs- und Interpretationsobjektivität unterschieden. Die Durchführungsobjektivität charakterisiert die Unabhängigkeit eines Testergebnisses einer Person von der Testleitung. Eine hohe Durchführungsobjektivität kann erreicht werden, wenn die Instruktionen präzise schriftlich festgehalten werden und die Umsetzung durch die Testleitung exakt nach diesen Vorgaben erfolgt (z.B. Testinstruktion oder Bearbeitungszeit). Die Auswertungsobjektivität betrifft die Auswertung des Tests, welche durch verschiedene Personen zu einem identischen Ergebnis führen sollte. Dies kann mittels Auswertungsschablonen oder Computerprogrammen erfolgen. Die Interpretationsobjektivität ist gegeben, wenn unterschiedliche Personen die gleichen Rückschlüsse beziehungsweise Interpretationen in Bezug auf den Testwert vornehmen. Dies kann mithilfe von Interpretationsbeispielen im Testmanual erfolgen (Bortz & Döring, 2016; Lienert & Raatz, 1998).

Die Durchführungsobjektivität ist gewährleistet, da eine validierte Testinstruktion vorliegt, welche von der Testleitung exakt abgelesen werden soll. Das Einhalten und genaue Vorlesen der Testinstruktion ist besonders wichtig, da innerhalb des Tests nur kurze Arbeitsanweisungen zu finden sind. An dieser Stelle ist anzumerken, dass bisher keine Zeitangabe pro Subtest existiert und so lange gewartet wird, bis davon ausgegangen werden kann, dass die Aufgabe von allen Kindern gelöst werden konnte. Hier ist zu überlegen, ob für zukünftige Tests eine Zeitangabe festgesetzt wird, in der davon ausgegangen werden kann, dass die Lösung gefunden wird. Dadurch kann die Durchführungsobjektivität erhöht und lange Wartezeiten können reduziert werden. Die Auswertungsobjektivität ist gegeben, da für den gesamten Test eine Auswertungsschablone vorliegt. Da eine Unterscheidung zwischen richtigen und falschen Antworten vorgenommen wird, sind in der Lösungsschablone für jede Aufgabe alle möglichen richtigen Ergebnisse aufgelistet. Nur Ergebnisse, die in der Auswertungsschablone zu finden sind, werden auch als richtig bepunktet. Schwierig zu beurteilen können Aufgaben sein, bei denen die Kinder die Pfeilrichtung ungenau zeichnen, Radierungen nicht exakt vorgenommen werden, die Zahlen nicht zu lesen sind oder die Lösungen nicht richtig eingetragen werden. Aus

diesem Grund werden die Kinder vorab auf diese Fehlerquelle hingewiesen. Um die Interpretationsobjektivität in diesem Zusammenhang zu erhöhen sind Beispiele von ungenauen Lösungen und Interpretationsvorgaben hinzuzufügen. Die Interpretationsobjektivität kann zu diesem Zeitpunkt der Testkonstruktion nur unzureichend gewährleistet werden, da hierfür eine Endversion des Programmiertests benötigt wird. Dafür bedarf es jedoch einer weiteren quantitative Studie zur abschließenden Testinstrumententwicklung (siehe Diskussion der Forschungsfrage 1 im Kapitel 5.1).

### **3.3 Itemanalyse**

Über eine Itemanalyse gilt es herauszufinden, welche Items das Konstrukt angemessen repräsentieren und welche Items aus dem Itempool entfernt werden sollten. Dafür wird basierend auf Bortz & Döring (2016), Eid & Schmidt (2014) oder auch Kelava & Moosbrugger (2020) das methodische Verfahren der Itemanalyse durchgeführt. Der erste Analyseschritt beinhaltet die Überprüfung der Rohwerte der Testscores hinsichtlich einer Normalverteilung (3.3.1). Dies erfolgt durch den Kolmogoroff-Smirnov-Test und eine grafische Betrachtung der Histogramme und der Q-Q-Diagramme. Daran anschließend erfolgt eine Berechnung der Itemschwierigkeiten (3.3.2). Die Itemschwierigkeitsgrade werden basierend auf der Formel nach Lienert und Raatz (1998) bestimmt. Der nächste Analyseschritt beinhaltet die Itemtrennschärfe (3.3.3). In diesem Unterpunkt werden die intervallskalierten Merkmale mittels Produkt-Moment-Korrelation berechnet. Abschließend wird das Vorgehen der Dimensionalitätsprüfung (3.3.4) beschrieben.

#### **3.3.1 Rohwertverteilung**

Bei der Überprüfung der Rohwerte des Testscores auf eine Normalverteilung wird der Kolmogoroff-Smirnov-Tests verwendet (Bortz & Döring, 2016; Bortz & Lienert, 2008; Field, 2018). Ab einem Wert von  $p > 0.05$  ist die Verteilung nicht signifikant und es liegt eine Normalverteilung der Werte vor. Ab einem Wert von  $p < 0.05$  ist die Verteilung signifikant und es liegt keine Normalverteilung vor (Field, 2018). Zusätzlich können über Histogramme und Q-Q-Diagramme die Daten grafisch hinsichtlich einer Normalverteilung geprüft werden (Bortz & Döring, 2016; Field, 2018). Eine Überprüfung der Werte auf eine Normalverteilung ist notwendig, da einige inferenzstatistische Verfahren dies voraussetzen (Bortz & Döring, 2016). Falls die Berechnungen eine Abweichung der Normalverteilung aufweisen, kann mithilfe des zentralen Grenzwerttheorems eine Normalverteilung ab einer Stichprobengröße von 30

Personen als vorausgesetzt angesehen werden (Bortz & Döring, 2016; Pólya, 1920). Denn das zentrale Grenzwerttheorem besagt, dass für ein „großes  $n$  die Stichprobenverteilung des Mittels normal“ (Bortz & Schuster, 2010, S. 86) ist.

### 3.3.2 Itemschwierigkeiten

Der nächste deskriptivstatistische Schritt der Itemanalyse ist die Berechnung der Itemschwierigkeiten. Die Berechnung erfolgt mittels Schwierigkeitsindex, welcher für jedes Item einzeln berechnet wird. Es gilt herauszufinden, mit welcher Lösungsrate die Items von den Testpersonen gelöst werden. Ein hoher numerischer Wert bedeutet, dass viele Testpersonen die Aufgaben lösen konnten und es sich folglich um eine einfache Aufgabe handelt. Ein niedriger numerischer Wert definiert eine schwierige Aufgabe und impliziert, dass sie nur von wenigen Testpersonen gelöst werden konnte. Das Skalenniveau der Items entscheidet darüber, welcher Schwierigkeitsindex gewählt wird. Bei nominalskalierten Items wird der Schwierigkeitsindex über den Prozentsatz der richtig gelösten Aufgaben angegeben (Bortz & Döring, 2016; Eid & Schmidt, 2014; Lienert & Raatz, 1998; Moosbrugger & Kelava, 2020). Niveautests werden mithilfe folgender Formel berechnet:

$$P = 100 \times \frac{N_R}{N} \text{ (Lienert \& Raatz, 1998, S. 74)}$$

$P$  ist definiert als der Quotient aus der Anzahl der Probanden und Probandinnen, welche die Aufgabe richtig gelöst haben ( $N_R$ ) dividiert durch die Gesamtanzahl der Probanden und Probandinnen ( $N$ ), multipliziert mit 100 (Lienert & Raatz, 1998). Items können einen Schwierigkeitsindex zwischen 0.0 und 1.0 annehmen (Fisseni, 1990, S.102; Weise, 1975). Fisseni besagt bereits 1990 das Schwierigkeitsindexe, welche einen Wert von  $0.50 > p < 0.50$  aufweisen berücksichtigt werden sollten. Allgemein werden Items innerhalb eines Schwierigkeitsindexes von  $0.20 < p < 0.80$  bevorzugt. (Bortz & Döring, 2016). In dieser Dissertation wird eine Spanne von  $0.19 < p < 0.84$  gewählt, damit die Homogenität des Itempools entzerrt wird und eine breitere Differenzierung von Items berücksichtigt werden kann (Fisseni, 1990). Zur Kategorisierung der Items wird von der Autorin eine Schwierigkeitsspanne in die drei Bereiche leicht, mittel und schwierig vorgeschlagen. Items mit einem Schwierigkeitsindex von  $0.19 < p < 0.40$  sind als schwierig zu betrachten. Items, welche einem Wert von  $0.40 < p < 0.60$  entsprechen, sind einem mittleren Anspruchsniveau zuzuordnen. Items, im Bereich von  $0.60 < p < 0.84$  werden als einfach zu lösende Aufgaben eingestuft.

Aufgrund des in Kapitel 3.2.3 beschriebenen Vorgehens zur Entwicklung der drei parallelen Testversionen B, werden die Itemschwierigkeiten für die Testversion A und B teilweise

zusammen bzw. getrennt berechnet. Bei Aufgaben, bei denen es sich in der Version B um eine Pseudo-Parallelaufgabe handelt, werden die Itemschwierigkeiten zusammen berechnet. Dadurch kann eine größere Stichprobe für einzelne Items berücksichtigt werden und eine zuverlässigere Aussage hinsichtlich der Berechnungen abgeleitet werden. Eine Übersicht über die Berechnung der Schwierigkeitsindexe ist Tabelle 4 zu entnehmen. Beispielsweise wird innerhalb des Testhefts der Sequenz für die Aufgabe 1 ein Schwierigkeitsindex angegeben, da es sich bei der Version B um eine Pseudo-Parallelaufgabe handelt. Bei den Aufgaben 6 bis 15 aus dem Testheft der Sequenz (Version A und B) werden die Schwierigkeitsindexe hingegen einzeln pro Testversion A oder B berechnet, da es sich hier nicht um eine Pseudo-Parallelaufgabe sondern um eine parallele Aufgabe handelt. Zusätzlich wird eine Änderung in der Testinstruktion zu dem Testheft der Sequenz vorgenommen, was dazu führt, dass mehr Aufgaben als richtig gewertet werden können. So werden innerhalb des Testhefts der Sequenz die Aufgaben 14 und 15 als richtig gewertet, wenn die Figur in das Haus navigiert wird. Es ist nicht von Relevanz einen möglichst schnellen Weg zu finden, wie ursprünglich angenommen. Diese Entscheidung hat zur Folge, dass die Items 14 und 15 beibehalten werden können (siehe Ergebniskapitel).

**Tabelle 4***Aufschlüsselung der Testversionen zur Berechnung der Schwierigkeitsindexe*

Testversion	Aufgaben														
Sequenz	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	-	-	-	-	-	x	x	x	x	x	x	x	x	x	x
B	-	-	-	-	-	x	x	x	x	x	x	x	x	x	x
A + B	x	x	x	x	x	-	-	-	-	-	-	-	-	-	-

Alternative	1	2	3	4	5	6	7	8	9	10	11	12	13
A + B	x	x	x	x	x	x	x	x	x	x	x	x	x

Iteration	1	2	3	4	5	6	7	8	9	10	11	12	13
A	-	-	x	x	-	-	-	-	-	x	-	-	x
B	-	-	x	x	-	-	-	-	-	x	-	-	x
A + B	x	x	-	-	x	x	x	x	x	-	x	x	-

*Anmerkung.* A = Testitem der Testversion A. B = Testitem der Testversion B. A + B = identische Testitems. x = gewähltes Item. - = nicht gewähltes Item.

**3.3.3 Itemtrennschärfe**

Der Trennschärfekoeffizient ist definiert „als die Korrelation der Beantwortung dieses Items mit dem Summenwert der übrigen Skala“ (Bortz & Döring, 2016, S. 478). Ein hoher Trennschärfekoeffizient einer Aufgabe unterscheidet *gute* und *schlechte* Personenleistungen voneinander. Testpersonen, welche im Gesamtergebnis einen hohen Testwert erreichen, werden ein trennscharfes Einzelitem richtig lösen. Ebenso werden Personen mit einem niedrigen Testwert ein trennscharfes Einzelitem falsch lösen. Trennscharfe Items lassen somit bereits Rückschlüsse zu, welche Person eine hohe oder niedrige Ausprägung bezüglich des Konstrukts besitzt (Bortz & Döring, 2016; Lienert & Raatz, 1998). Das Skalenniveau der Items entscheidet darüber, welche Berechnungsmethode des Korrelationskoeffizient angewendet wird. Intervallskalierte Merkmale werden mittels Produkt-Moment-Korrelation nach der Part-Whole-Korrektur berechnet (Bortz & Döring, 2016; Bortz & Lienert, 2008; Moosbrugger & Kelava, 2012). Der Trennschärfekoeffizient kann einen Wert zwischen -1 und +1 annehmen (Bortz & Döring, 2016; Moosbrugger & Kelava, 2020). Dabei definieren Werte, welche um  $r < 0.30$  einen niedrigen Zusammenhang, ein Wert zwischen  $r = 0.30$  und  $r = 0.50$  einen mittleren

Zusammenhang und ein Wert  $r > 0.50$  einen hohen Zusammenhang (Bortz & Döring, 2016; Cohen, 1988; Fisseni, 1990; Weise, 1975). Über eine inhaltliche Interpretation wird entschieden, welche Items nach den Korrelationsberechnungen beibehalten werden. Die Itemtrennschärfen sind von ihrer Itemschwierigkeit abhängig. So liefern Items mit einer mittleren Schwierigkeit die höchsten Trennschärfen (Bortz & Döring, 2016).

Zur Berechnung des Trennschärfekoeffizienten werden folgende Skalen gebildet: Je eine Skala zu der Sequenz, zu der Alternative, zu der Zählschleife und zu der Bedingten Schleife. Zusätzlich wird auf Grund der unterschiedlich verwendeten Aufgaben in der Testversion A und B zwischen Skala A und B unterschieden. Eine Ausnahme bildet weiterhin das Testheft der Alternative, welches als eine Skala gefasst wird. Für die Berechnung der Trennschärfe werden alle Items berücksichtigt, welche nach der Schwierigkeitsanalyse vorliegen. Diese Items werden mittels Trennschärfeberechnung weiter untersucht. Es werden Items beibehalten, wenn sie mittels Trennschärfeberechnung ein Korrelationskoeffizient von  $r > 0.28$  aufweisen. Das Herabsetzen des Korrelationskoeffizienten bietet den Vorteil, dass drei Items der Iteration (Aufgabe 1 B, 2 A und 6 A), welche nah an 0.30 liegen weiterhin genutzt werden können (Bortz & Döring, 2016).

### 3.3.4 Dimensionalitätsprüfung

Bei einer Dimensionalitätsprüfung eindimensionaler Tests wird getestet, ob alle Items ein Konstrukt messen. In diesem Zusammenhang wird eine hohe Homogenität angestrebt. Eine Betrachtung der Dimensionalität erfolgt in der Regel mit Hilfe einer Faktorenanalyse oder, wenn die Voraussetzungen nicht erfüllt sind durch die Berechnung der Mittelwerte der Item-Interkorrelation (Bortz & Döring, 2016). Die Durchführung einer Faktorenanalyse ist abhängig von der Itemanzahl, der Stichprobengröße, der Kommunalität und den Ladungswerten der Variablen (Bortz & Döring, 2016; Bühner, 2021). Insgesamt ist eine Anzahl von mindestens vier Items, eine Kommunalität von mindestens 0.60 und eine Stichprobengröße von 250 Personen Voraussetzung zur Durchführung einer Faktorenanalyse (Bühner, 2021). Weiterhin ist ein Stichprobenumfang von 150 Personen ausreichend, wenn pro Faktor zehn oder mehr Variablen vorliegen. Wenn hingegen nur eine geringe Variablenanzahl pro Faktor vorliegt, wird eine Stichprobe von über 300 Personen empfohlen. Darüber hinaus reicht ein Ladungswert von über 0.60 bzw. 0.40 aus, wenn pro Faktor vier bzw. zehn bis zwölf Variablen identifiziert sind, um die Stichprobengröße zu vernachlässigen (Bortz & Döring, 2016; Guadagnoli & Velicer, 1988). Eine weitere Möglichkeit der Dimensionalitätsprüfung bietet die Berechnung des

Mittelwertes der Item-Interkorrelation. Bei der Betrachtung der Dimensionalität über den Mittelwert der Item-Interkorrelation sollten die Korrelationswerte positiv ausfallen, sich signifikant von Null unterscheiden und inhaltlich homogen sein. Die Homogenität wird über den Mittelwert der Item-Interkorrelation angegeben (Bortz & Döring, 2016). Als akzeptierte Werte gelten 0.20 bis 0.40 (Briggs & Cheek, 1986). Items mit einer geringen Homogenität sollten aus dem Itempool entfernt werden, da sie inhaltlich etwas anderes messen als die restlichen Items. Für die Dimensionalitätsprüfung werden die Berechnungen der Bedingten Schleife und Zählschleife aus inhaltlichen Gründen getrennt betrachtet.

### **3.4 Weitere Berechnungen**

In diesem Kapitel werden weitere deskriptivstatistische Schritte dargelegt. Zu Beginn wird die Reliabilität (3.4.1) beschrieben. Danach wird auf die Häufigkeitsverteilungen der Testversionen A und B sowie auf die Häufigkeitsverteilung der Geschlechter eingegangen (3.4.2).

#### **3.4.1 Reliabilität**

Die Reliabilität ist definiert als der Grad der Messgenauigkeit, mit welcher der Test ein bestimmtes Persönlichkeitsmerkmal misst. Hierfür wird die interne Konsistenz mithilfe des Cronbachs Alpha-Koeffizienten ermittelt (Bortz & Döring, 2016; Lienert & Raatz, 1998; Moosbrugger & Kelava, 2020). Ein Reliabilitätskoeffizient mit einem Wert von über 0.90 charakterisiert eine hohe Korrelation zwischen den Items. Ein Wert von über 0.80 wird in der Literatur als ausreichend beschrieben. Jedoch werden häufig bereits Skalen mit Werten um 0.70 verwendet (Bortz & Döring, 2016; Moosbrugger & Kelava, 2020). Zusätzlich wird angemerkt, dass die Wertevorgaben für die Studie individuell zu prüfen sind (Bortz & Döring, 2016; Field, 2018; Schecker, 2014; Schmitt, 1996). Gerade in der frühen Phase der Forschung sind Werte unter 0.50 ausreichend, da eine gewisse inhaltliche Abdeckung eines Bereichs notwendig ist (Field, 2018; Schecker, 2014). Zusätzlich haben Leistungstests auf Grund einer begrenzten Testlänge Schwierigkeiten, einen hohen Cronbachs Alpha Wert zu erreichen, da der Itempool nur begrenzt erhöht werden kann. Je mehr Items jedoch in einem Itempool vorhanden sind, desto besser fällt der Cronbachs Alpha Wert aus. Insgesamt ist zwischen der Bandbreite des Tests und der Gleichheit von Items (hohe Reliabilität) abzuwägen (Schecker, 2014).

#### **3.4.2 Häufigkeitsverteilungen**

Zur Betrachtung der Häufigkeitsverteilungen werden Histogramme verwendet, welche als besonders geeignet gelten (Bortz & Döring, 2016). In dieser Arbeit werden die

Häufigkeitsverteilungen der Testwerte für die Themen Sequenz, Alternative und Iteration (Version A und B) dargestellt. Die Berechnungen basieren auf den Items, welche nach der Itemanalyse zur Verfügung stehen. Zusätzlich werden die Testwerte der einzelnen Themen nach Geschlechtern getrennt für jedes Testheft beschrieben. In diesem Kontext werden die Gruppenunterschiede auf Signifikanz überprüft. Zur Überprüfung der Mittelwertsunterschiede von zwei unabhängigen Stichproben (wie das Geschlecht) wird ein  $t$ -Test durchgeführt. Die Voraussetzungen zur Durchführung von  $t$ -Tests bei unabhängigen Stichproben sind zwei voneinander unabhängige Variablen, wobei die  $y$ -Variablen metrisch skaliert sind (Bortz & Döring, 2016). Diese Voraussetzung ist für alle Testhefte erfüllt. Zusätzlich sollten die  $y$ -Variablen innerhalb der Gruppe eine Normalverteilung aufweisen (Field, 2018). Zur Überprüfung, ob die Daten normalverteilt sind, wird der Kolmogoroff-Smirnov-Test verwendet (Bortz & Döring, 2016; Bortz & Lienert, 2008; Field, 2018). Eine Normalverteilung liegt bei einem Wert von  $p > 0.05$  vor (Field, 2018). Die Überprüfung auf Normalverteilung zeigt, dass für alle Summenwerte der Testhefte keine Normalverteilung vorliegt. In diesem Fall wird auf das zentrale Grenzwerttheorem zurückgegriffen, welches eine Normalverteilung ab einer Stichprobengröße von 30 Personen voraussetzt (Bortz & Döring, 2016; Bortz & Schuster, 2010; Pólya, 1920). Da in allen Gruppen eine Stichprobengröße von über 30 Kindern vorliegt, wird für alle Daten eine Normalverteilung angenommen. Anschließend erfolgt die Überprüfung auf die Varianzhomogenität, welche mit Hilfe des Levene-Tests erfolgt (Field, 2018; Levene, 1960). Ein signifikantes Ergebnis ( $p < 0.05$ ) bedeutet, dass sich die Varianzen unterscheiden und eine Varianzheterogenität vorliegt. In diesem Fall wird der  $t$ -Test für die Mittelwertgleichheit hinsichtlich ungleicher Varianzen (Welch-Test) auf Signifikanz hin überprüft. Wenn  $p > 0.05$  ist, sind die Varianzen gleich groß und der  $t$ -Test wird dahingehend betrachtet. Zur Bestimmung der Mittelwertunterschiede wird im letzten Schritt die zwei-seitige Signifikanz angeschaut. Liegt der Wert über  $p > 0.05$ , dann ist kein Gruppenunterschied zu verzeichnen. Wenn der Wert mit  $p < 0.05$  signifikant ist, liegen Gruppenunterschiede vor (Bortz & Döring, 2016; Field, 2018). In diesem Fall wird die Effektstärke nach Cohens  $d$  (1988) betrachtet. Cohen (1988) definiert ab  $d = 0.2$  einen geringen, ab  $d = 0.5$  einen mittleren und ab  $d = 0.8$  einen starken Effekt.

### 3.5 Ergebnisraum

Ein weiterer methodischer Schritt beinhaltet die Festlegung des Ergebnisraums. Der Ergebnisraum definiert Kategorien und Bepunktungen der Antworten, welche als Grundlage

für zu ziehende Schlussfolgerungen genutzt werden. Dies beinhaltet die Konzeption einer Codieranleitung. Die Codieranleitung umfasst alle Items eines Tests. Die Kategorien der Antworten sollten eindeutig formuliert und nach Schwierigkeit sortiert werden. Für die Bepunktung der Antwortkategorien werden bei geschlossenen Antwortformaten beispielsweise eine Einordnung der Antworten in die Kategorien *richtig* und *falsch* vorgenommen. Bei offenen Antwortformaten werden unterschiedliche Kategorien gebildet, welche den Ergebnisraum darstellen. Die gebildeten Kategorien werden anschließend gescort. Zusätzlich wird eine Beziehung zum theoretischen Konstrukt hergestellt. Bei der Erstellung der Codieranleitung handelt es sich um einen iterativen Prozess, welcher bereits während des Prozesses der Aufgabenentwicklung mit bedacht wurde (Eggert & Bögeholz, 2014; Kauertz, 2014; Wilson, 2005).

### **3.6 Messmodell**

In diesem Kapitel wird das methodische Vorgehen zur Bestimmung des Messmodells dargelegt. Hierfür werden zu Beginn die theoretischen Grundlagen der Klassischen Testtheorie (KTT) und die Item-Response-Theorie (IRT) vorgestellt (3.6.1). Anschließend werden die Schritte zur Durchführung einer Rasch-Analyse detailliert beschrieben (3.6.2).

#### **3.6.1 Klassische Testtheorie und Item-Response-Theorie**

Der vierte Baustein der Testentwicklung ist durch die Wahl eines passenden Messinstruments charakterisiert. Das Messmodell setzt die Itemantworten in Beziehung zu dem theoretischen Konstrukt (Eggert & Bögeholz, 2014; Wilson, 2005). Es existieren unterschiedliche statistische Messmodelle, wie das True-Score Modell der klassischen Testtheorie, das Domain-Score Modell, Modelle der Faktorenanalyse, Item-Response-Modelle oder latente Klassenmodelle (Wilson, 2005). Vorherrschend in der Testtheorie sind die Klassische Testtheorie (KTT) und die Item Response Theorie (IRT), welche im Folgenden skizziert werden.

Die KTT wird überwiegend bei kontinuierlichen Antwortformaten und zur Bestimmung der Reliabilität und Validität der Testwerte eingesetzt. Ein Schwerpunkt der Theorie bildet die Annahme, dass sich die beobachtbare Merkmalsausprägung aus dem wahren Wert (True-Score) und einem zufälligen Messfehler zusammensetzt. Es wird ein linearer Zusammenhang zwischen den beobachtbaren Itemvariablen und den latenten Variablen angenommen, wodurch eine Interpretation der Kennwerte möglich ist. Zusätzlich lassen sich die Schätzungen der

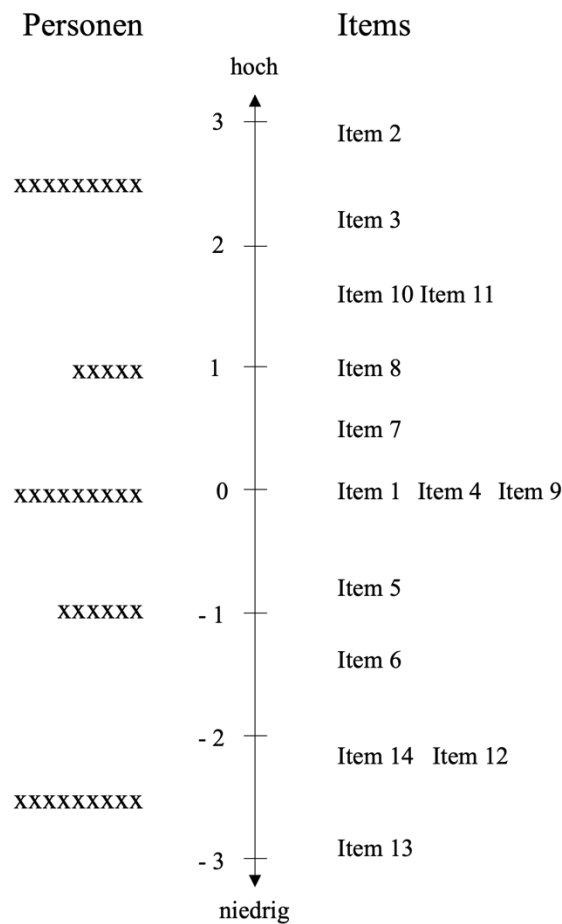
Reliabilität und der Validität über die Testsummenwerte einfach vornehmen. Ein Nachteil ist, dass die Gültigkeit der Modellannahmen in der Vergangenheit nicht geprüft werden konnten. Mittlerweile existieren auch für die KTT Modelltests, die zur Analyse der Modellannahmen dienen, weshalb eine Kritik an der KTT unbegründet ist (Moosbrugger & Kelava, 2020). Die Gültigkeit des Modells wird häufig mittels konfirmatorischer Faktorenanalyse ermittelt. Darüber wird geprüft, welches Messmodell passend zu den Daten ist und welches Reliabilitätsmaß (Kuder-Richardson-Formel, Cronbachs Alpha oder McDonalds Omega) für weitere Berechnungen verwendet werden darf (Bortz & Döring, 2016; Moosbrugger & Kelava, 2020). Die Voraussetzungen zur Durchführung der Faktorenanalyse sind, wie bereits in Kapitel 3.3.4 beschrieben, abhängig von der Itemanzahl, der Stichprobengröße, der Kommunalität und den Ladungswerten der Variablen (Bortz & Döring, 2016; Bühner, 2021; Klopp, 2010).

Die IRT wird schwerpunktmäßig bei dichotomen (kategorialen) Antwortkategorien angewendet. Der Fokus der IRT liegt darauf, dass aus dem beobachteten Antwortverhalten die latenten Variablen, die Personenfähigkeit und die Itemschwierigkeit geschätzt werden. Die Zusammenhänge werden kurvilinear über probabilistische Modelle dargestellt. Ein Vorteil liegt darin, dass die Modellannahmen mit Modelltests überprüft werden können (Moosbrugger & Kelava, 2020). Kompetenzmodellierungen werden üblicherweise mit Messmodellen der Item-Response-Theorie vorgenommen (Eggert & Bögeholz, 2014; Hartig, 2007). Hier wird zwischen Latent-Trait-Modellen, Latent-Class-Modellen und Mischverteilungs-Modellen unterschieden. Relevant an dieser Stelle sind die Latent-Trait-Modelle. Diese erfassen latente Merkmale als ein kontinuierliches Merkmal, welche diverse Abstufungen in einer minimalen bis maximalen Spanne aufweisen. Innerhalb der Latent-Trait-Modelle wird weiter zwischen dichotomen und ordinalen Modellen unterschieden. Dichotomen Modellen liegen dichotome Items zu Grunde und ordinale Modelle basieren auf mindestens ordinalskalierten Items. Weiter von Bedeutsamkeit sind in diesem Kontext dichotome Latent-Trait-Modelle. Dazu zählen das ein-Parameter-Logistische-Modell (dichotomes Rasch-Modell), das zwei-Parameter-Logistische-Modell (Birnbaum-Modell) und das drei-Parameter-Logistische-Modell. Es werden immer die Itemlösungswahrscheinlichkeiten basierend auf der Personenfähigkeit errechnet, plus ein (Itemschwierigkeit), zwei (Itemschwierigkeit, Trennschärfe) oder drei (Itemschwierigkeit, Trennschärfe, Ratewahrscheinlichkeit) weitere Parameter, je nach Modell (Bortz & Döring, 2016).

Insgesamt lässt sich konstatieren, dass die KTT und die IRT einige Gemeinsamkeiten aufweisen und dass die beiden Theorien nicht miteinander konkurrieren, sondern sich ergänzen. Jede Theorie hat ihre beschriebenen Vorteile, welche es entsprechend zu nutzen gilt (Moosbrugger & Kelava, 2020). Da es sich bei den verwendeten Testheften (Sequenz, Alternative und Iteration) der Hauptstudie um Leistungstests handelt und die Tests ein dichotomes Antwortformat (richtig/falsch) aufweisen, wird zur Kompetenzmodellierung das dichotome Rasch-Modell verwendet.

### **3.6.2 Rasch-Analyse**

Das Rasch-Modell wurde von dem dänischen Mathematiker Georg Rasch (Rasch, 1960) entwickelt und von Benjamin Wright (Wright & Stone, 1979) erweitert. Die mathematischen Feinheiten des Modells nahmen Wright und Masters (1982) vor. Die Rasch-Analyse eignet sich besonders gut, um Testdaten basierend auf Multiple-Choice Fragen, welche in richtig oder falsch gescort sind, zu analysieren. Die Analyse liefert wichtige Daten zur Vorbereitung von statistischen Auswertungen und wesentliche Aspekte zur Bewertung des Messinstruments (Boone et al., 2014). Verschiedene Aspekte unterstreichen die Anwendung einer Rasch-Analyse. Das Verfahren bietet beispielweise die Möglichkeit, die Funktionsfähigkeit des Messinstruments zu evaluieren. Weiter werden die Personenfähigkeiten auf einer linearen Skala abgebildet. Zusätzlich können die Itemschwierigkeiten berechnet werden, welche in einer Person-Item-Map dargestellt werden können (Abbildung 8) (Eggert & Bögeholz, 2014; Neumann, 2014). Diese wiederum kann zur Überprüfung der Validität des Messinstruments verwendet werden (Holmefur & Krumlinde-Sundholm, 2016). Die Person-Item-Map spannt ein Kontinuum von niedrig bis hoch auf. Auf der linken Seite der Person-Item-Map ist die Personenfähigkeit und auf der rechten Seite die Itemschwierigkeit abgebildet (Wilson, 2005; Wright et al., 1980; Wright & Masters, 1982).

**Abbildung 8***Person-Item-Map*

*Anmerkung.* Angelehnt an Wilson, 2005, S. 27 und Wright & Masters, 1982, S. 159.

Insgesamt ist die Raschanalyse unabhängig von der Stichprobe durchzuführen. Jedoch können geringe Stichproben zu ungenauen Parametern führen. Allgemein wird eine Mindestgröße von 100 Probandinnen und Probanden empfohlen (Neumann, 2014). Die Auswertung der Rasch-Analyse orientiert sich an dem methodischen Vorgehen von Davis & Boone (2021) und beinhaltet vier Schritte: Die Überprüfung der Eindimensionalität, die Bewertung der Zielsetzung sowie der Boden- und Deckeneffekte, die Evaluation der Funktionsfähigkeit des Messinstruments und die Konstruktvalidität. Die theoretischen Grundlagen dieser Schritte werden folgend dargelegt.

**Eindimensionalität:** Die Eindimensionalität ist eine Voraussetzung, um die Rasch-Analyse durchzuführen (Wright & Stone, 1979). Zur Evaluation der Eindimensionalität werden drei

Verfahren empfohlen: *fit statistics*, *Principal Component Analysis of Residuals* (PCAR) und die *Point-Measure Correlations*. Mithilfe der Methode *fit statistics* wird überprüft, ob die Daten zu dem Modell passen. Die Überprüfung erfolgt mittels Infit Mean-Square (MNSQ) und Outfit Mean-Square (MNSQ). Ein Wert von 1.0 repräsentiert eine perfekte Passung zum Raschmodell. Ein Wert, welcher innerhalb der Range von 0.5 und 1.5 liegt wird als akzeptiert bezeichnet (O'Conner et al. 2016; Wright & Linacre 1994). Wenn die Daten in das Modell passen, dann gilt dies als ein Beweis für die Eindimensionalität (Brown et al. 2016; Li et al. 2016). Mit der Methode *Principal Component Analysis of Residuals* (PCAR) wird untersucht, ob mehr als eine Varianzkomponente existiert, auf welche das Antwortverhalten hindeutet. Die Werte für die unerklärte Varianz sollten unter einem Wert von 2.0 für den ersten Kontrast liegen (Boone & Staver, 2020; Linacre, 1998). Die Berechnungen der *Point-Measure Correlations* sollten einen positiven Wert von über 0.3 aufweisen, damit die Items als akzeptiert angesehen werden können (Li et al. 2016).

**Personen- und Itemreliabilität sowie Personen- und Itemseparation:** Diese Daten geben Aufschluss darüber, wie zuverlässig der Test bei wiederholter Durchführung in Bezug auf Personen (Personenreliabilität) oder Items (Itemreliabilität) ist. Die Item- und Personenreliabilität sind gleichzusetzen mit Cronbach Alpha. Ein höherer Wert entspricht einer größeren Zuverlässigkeit (Davis & Boone, 2021). Die Personenseparation gibt an, wie gut ein Itemsatz die gemessenen Personen trennt. Die Itemseparation gibt an, wie gut eine Stichprobe von Testpersonen die verwendeten Items trennt (Wright & Stone, 1999). Werte für die Personenreliabilität sollten bei 0.80 und für die Itemreliabilität bei 0.90 liegen. Die Werte für Personenseparation sollten über 2.0 und für die Itemseparation über 4.0 sein (Malec et al. 2007).

**Zielsetzung sowie Boden- und Deckeneffekte:** Die Berücksichtigung der Zielsetzung sowie Boden- und Deckeneffekten sind wesentliche Messgrößen, welche in Raschanalysen berücksichtigt werden (Khadka et al. 2016; Wong et al. 2018). Die Zielsetzung betrachtet die Unterschiede der durchschnittlichen Personenfähigkeit und der durchschnittlichen Itemschwierigkeit. Wenn der Wert unter 1.00 Logits liegt, wird von einer guten Zielsetzung der Personen und Items gesprochen (Finger et al. 2012; Wong et al. 2018). Bei Betrachtung der Bodeneffekte sollten nicht mehr als 5 Prozent der Personen unterhalb des 5 Prozentbereichs liegen und bei der Betrachtung der Deckeneffekte sollten nicht mehr als 5 Prozent der Personen oberhalb des 95 Prozentbereichs liegen (Fisher 2007).

**Funktionsfähigkeit des Messinstruments und Konstruktvalidität:** Um zu überprüfen, ob das Messinstrument funktioniert, wird die Person-Item-Map genutzt. Darüber kann die Anordnung und der Abstand der Items genauer überprüft werden (Boone & Staver, 2020). Die Person-Item-Map trennt durch eine vertikale Linie die Itemschwierigkeiten auf der rechten Seite von den Personenfähigkeiten auf der linken Seite. Die Itemschwierigkeiten und Personenfähigkeiten sind von leicht (unten) bis schwer (oben) bzw. schwach (unten) und stark (oben) angeordnet (Wilson, 2005). Bei Betrachtung der Person-Item-Map hinsichtlich der Itemanordnung ist es wichtig, zu untersuchen, ob Items nebeneinander liegen oder große Abstände zwischen den Items vorherrschen. Im ersten Fall liegt eine Dopplung von Items vor und im zweiten Fall werden Bereiche aufgedeckt in denen Items fehlen (Boone et al. 2014).

### 3.7 Testvalidierung

In diesem Kapitel werden zuerst in der Hinführung (3.7.1) die drei verschiedenen Validitätsformen: Inhalts-, Konstrukt- und Kriteriumsvalidität vorgestellt. Insbesondere wird auf die Konstruktvalidität eingegangen, da diese in der Dissertation zentral behandelt wird. Anschließend wird in Kapitel 3.7.2 das methodische Vorgehen der Validierungsstudie beschrieben. Kapitel 3.7.3 beschreibt basierend auf dem skizzierten nomologischen Netzwerk in Kapitel 3.1.1 mögliche Zusammenhänge. Die Überprüfung der Konstruktvalidität erfolgt hinsichtlich mathematischer Inhalte mit dem Deutschen Mathematiktest für zweite Klassen (DEMAT 2+) und hinsichtlich der Intelligenz mit dem Culture Fair Intelligence Test (CFT 1-R).

#### 3.7.1 Hinführung

Die Validität (Gültigkeit) gibt an, ob der Test das misst, was er zu messen vorgibt (Bortz & Döring, 2016; Eid & Schmidt, 2014; Moosbrugger & Kelava 2020). Die internationalen Test-Standards (AERA, APA & NCME, 2014) unterscheiden zwischen der Inhalts-, der Konstrukt- und Kriteriumsvalidität. Mit der Inhaltsvalidität wird inhaltlich überprüft, inwiefern die Testitems die schulischen Inhalte im gesamten Umfang repräsentieren. Die Beurteilung erfolgt mittels Expertisegremium. Bei der Validitätsbestimmung liegt kein numerischer Wert vor (Bortz & Döring, 2016; Eid & Schmidt, 2014). Die Kriteriumsvalidität gibt den Zusammenhang zwischen dem Verhalten innerhalb des Tests und einem Außenkriterium an. Je höher der positive, korrelierte Wert ist, desto enger ist die Beziehung zwischen dem Testwert und dem Kriteriumswert. Es wird zwischen konkurrender, prognostischer und retropektiver Validität

unterschieden. Konkurrente Validität definiert eine zeitgleiche Messung von Test- und Kriteriumswert. Mit der prognostischen Validität wird mittels Testwert geprüft, ob ein später erhobenes Kriterium vorhergesagt werden kann. Die retropektive Validität nutzt Messwerte aus früheren Messverfahren (Bortz & Döring, 2016; Eid & Schmidt, 2014; Moosbrugger & Kelava, 2012). Hinzuzufügen ist, dass die Kriteriumswerte für andere Stichproben andere Validitäten ergeben und somit von der gewählten Stichprobe abhängen (differenzielle Validität). Die Bestimmung des Validitätskoeffizienten erfolgt mittels Korrelationsberechnung (Bortz & Döring, 2016). Die Kriteriums- und Inhaltsvalidität werden in dieser Dissertation nicht thematisiert. Der Fokus liegt auf der Bestimmung der Konstruktvalidität, weshalb das methodische Vorgehen folgend detailliert dargelegt wird.

Die Konstruktvalidität gibt an, was der Test misst, indem a priori theoretische Aussagen über das zu untersuchende Konstrukt aufgestellt werden. Die Gesamtheit aller aufgestellten Hypothesen wird als nomologisches Netzwerk bezeichnet. Das nomologische Netzwerk gibt Beziehungen zwischen dem zu messenden Konstrukt und anderen Konstrukten an. Es geht darum, herauszufinden, welche Überschneidungen und Abgrenzungen zu anderen Konstrukten vorliegen und wie das zu messende Konstrukt in diesem Netz verortet werden kann (Hartig & Jude, 2007). Wenn die Testwerte den formulierten Hypothesen entsprechen, liegt ein erstes Anzeichen für eine Konstruktvalidität vor. Es wird zwischen der konvergenten und diskriminanten Validität unterschieden. Die konvergente Validität überprüft, ob ein Zusammenhang mit ähnlichen Konstrukten, also eine hohe Korrelation, vorliegt. Die diskriminante Validität definiert eine niedrige Korrelation zu entfernten Konstrukten (Bortz & Döring, 2016; Moosbrugger & Kelava, 2012). Zur Bestimmung der Konstruktvalidierung werden Korrelationsberechnungen durchgeführt (Bortz & Döring, 2016; Eid & Schmidt, 2014). Einen umfassenden Ansatz zur Bestimmung der konvergenten und diskriminanten Validität bietet der Multitrait-Multimethod-Ansatz (MTMM) nach Campbell und Fiske (1959). Mit diesem Ansatz lassen sich die diskriminante und konvergente Validität mittels Korrelationsberechnungen bestimmen. Die Zusammenhänge zwischen Konstrukten und angewendeten Methoden (hier Tests) werden mithilfe von Multitrait-Multimethod-Matrixen visualisiert (MTMM-Matrix). Eine vollständige MTMM-Matrix besteht aus vier Teilmatrixen: dem Monotrait-Monomethod-Block, dem Monotrait-Heteromethod-Block, dem Heterotrait-Monomethod-Block und dem Heterotrait-Heteromethod-Block (Abbildung 9).

## Abbildung 9

## MTMM-Matrix

		Methode 1			Methode 2		
		SE	AL	IT	SE	AL	IT
Methode 1	SE	1					
	AL		1				
	IT			1			
Methode 2	SE				1		
	AL					1	
	IT						1

Anmerkung. Angelehnt an Campbell und Fiske, 1959, S. 82.

Blau = Heterotrait-Monomethod (verschiedene Konstrukte mit derselben Methode). 1 = Monotrait-Monomethod Korrelationen (ein Konstrukt mit derselben Methode). Lila = Heterotrait-Heteromethod (verschiedene Konstrukte mit verschiedenen Methoden). Gelb = Monotrait-Heteromethod (ein Konstrukt mit verschiedenen Methoden). *SE* Sequenz, *AL* Alternative, *IT* Iteration.

Bei der Betrachtung der Monotrait-Monomethod (Abbildung 9) wird ein Konstrukt mit einer Methode gemessen. Die Werte werden mit sich selbst korreliert und es ergibt sich ein Korrelationswert von  $r = 1$ . Zur Bestimmung der konvergenten Validität führen Messungen desselben Konstrukts mit verschiedenen Methoden (Monotrait-Heteromethod) zu einer hohen Korrelation innerhalb der MTMM-Matrix, welche innerhalb des gelben Kastens wiederzufinden sind. Messungen zwischen inhaltlich unterschiedlichen Konstrukten in Bezug auf eine Methode (Heterotrait-Monomethod, blaue Dreiecke) und verschiedenen Methoden (Herterotrait-Heteromethod, lilafarbene Dreiecke) ergeben geringe Korrelationen innerhalb der MTMM-Matrix. Sie charakterisieren die diskriminante Validität (Bortz & Döring, 2016; Campbell und Fiske, 1959; Moosbrugger & Kelava, 2012). Campbell und Fiske (1959) definieren vier Kriterien, um die diskriminante und konvergente Validität zu bestimmen. Eine konvergente Validität liegt vor, wenn die Monotrait-Heteromethod-Korrelationen hoch sind und sich signifikant von Null unterscheiden. Andernfalls werden mit unterschiedlichen Methoden verschiedene Konstrukte gemessen. Für den Nachweis der diskriminanten Validität sind zwei Punkte zu prüfen. Erstens sollte eine geringere Korrelation bei Betrachtung des Heterotrait-Monomethod-Blocks im Vergleich zum Monotrait-Heteromethod-Blocks vorliegen. Das bedeutet, dass verschiedene Konstrukte innerhalb einer Methode geringer

korrelieren sollten als gleiche Konstrukte, die mit verschiedenen Methoden erfasst wurden. Zweitens sollten die Heterotrait-Heteromethod-Koeffizienten niedriger sein als die Monotrait-Heteromethod-Koeffizienten und die Heterotrait-Heteromethod-Koeffizienten sollten den kleinsten Wert aufweisen. Wenn das Kriterium nicht erfüllt ist, bedeutet dies, dass die Konstrukte sich inhaltlich nicht unterscheiden und innerhalb der Konstrukte ein gemeinsames Merkmal existiert. Darüber hinaus werden zum Schluss die Teilmatrizen hinsichtlich gleicher Rangreihen von Korrelationen überprüft. Dafür wird jede Teilmatrize einzeln betrachtet. Innerhalb der vier Dreiecke werden die Korrelationen der Größe nach sortiert und entsprechend inhaltlich beschriftet. Im nächsten Schritt wird überprüft, ob die Reihenfolge für alle Inhalte gleich ist. Ist dies der Fall wird eine echte Varianz gemessen (Bortz & Döring, 2016; Campbell und Fiske, 1959; Moosbrugger & Kelava, 2012).

In dieser Dissertation wird eine angepasste Variante der oben beschriebenen Methode verwendet, da kein vergleichbarer Programmieretest vorliegt, um mit verschiedenen Methoden das gleiche Konstrukt zu messen. So werden Tests verwendet, welche potenziell Zusammenhänge zwischen der Programmierkompetenz aufweisen könnten. Es wurde hierfür ein Intelligenztest und ein Mathematiktest herangezogen. In diesem Kontext geht es darum herauszufinden, welche Korrelationswerte innerhalb des grünen Kastens (Abbildung 10), vorliegen und ob sich darüber gleiche Konstrukte beziehungsweise unterschiedliche Konstrukte identifizieren lassen. Zur Visualisierung der Korrelationsberechnungen wird auf die MTMM-Matrix nach Campell und Fiske (1959) zurückgegriffen.

**Abbildung 10**

*Angepasste MTMM-Matrix zur Darstellung von Korrelationsberechnungen*

	SE	AL	IT	SUBST	LABYR	ÄHNL	REIH	KLASS	MATR
SE	1								
AL		1							
IT			1						
SUBST				1					
LABYR					1				
ÄHNL						1			
REIH							1		
KLASS								1	
MATR									1

*Anmerkung.* Angelehnt an Campbell und Fiske, 1959, S. 82.

blau = Heterotrait-Monomethod (verschiedene Konstrukte mit derselben Methode) Korrelationen. 1 = Monotrait-Monomethod Korrelationen (ein Konstrukt mit derselben Methode). grün = Heterotrait-Heteromethod Korrelationen (verschiedene Konstrukte mit verschiedenen Methoden) und Monotrait-Heteromethod (ein Konstrukt mit verschiedenen Methoden). *SE* Sequenz, *AL* Alternative, *IT* Iteration, *SUBST* Substitutionen, *LABYR* Labyrinth, *ÄHNL* Ähnlichkeiten, *REIH* Reihenfortsetzen, *KLASS* Klassifikationen, *MATR* Matrizen.

Für die Kriteriums- und Konstruktvalidität sind Korrelationswerte wünschenswert, welche möglichst nah bei Eins liegen. Insgesamt sind Werte, welche  $r \leq 0.40$  sind als niedrig, Werte zwischen 0.40 und 0.60 als mittelmäßig und Werte über  $r \geq 0.60$  als hoch zu werten (Bortz & Döring, 2016; Bühner, 2011; Fisseni, 1990, S. 102; Weise, 1975). Bortz und Döring (2016) weisen darauf hin, dass die Korrelationseinstufungen nicht in jedem Fall übernehmen werden sollten und dass ebenso eine inhaltliche Bewertung entscheidend ist.

### 3.7.2 Validierungsstudie

Für die Validierungsstudie wird ein Testheft (Version A und B) aus den Themen Sequenz, Alternative und Iteration erstellt. Die gewählten Items basieren auf den Ergebnissen der Itemanalyse und nicht auf den Ergebnissen der Rasch-Analyse. Dies hängt damit zusammen, dass die Ergebnisse der Rasch-Analyse erst nach der Durchführung der Validierungsstudie vorlagen. Für das Thema der Sequenz ergeben sich neun Items<sup>9</sup>, für die Alternative fünf Items,

<sup>9</sup> Item 8 unterscheidet sich von der Aufgabe aus der Hauptstudie. Ein Apfel befindet sich an einer anderen Stelle. Die Aufgabe wird weiter berücksichtigt, da inhaltlich keine relevanten Unterschiede zu identifizieren sind.

für die Bedingte Schleife fünf<sup>10</sup> Items und für die Zählschleife vier Items. Jedes neue Thema wird mit einer einfachen und zu den folgenden Testitems vergleichbaren Beispielaufgabe eingeführt. Durch das Vorgehen wird das Prinzip der Aufgabe verdeutlicht (Krajewski et al., 2004; Weiß & Osterland, 2013). Die parallele Testversion, wie in der Hauptstudie, wird beibehalten. Für den Fall, dass nur für eine Testversion A oder B ein Item aus der Itemanalyse existiert, wird jenes übriggebliebene Item ebenso für die entsprechende Version A oder B als identisches Item genutzt (Sequenz A: Item 8, Iteration A: Item 2 und 10, Iteration B: Item 5). Somit ergeben sich insgesamt zwei Testhefte (Version A und B) aus jeweils 23 Items, basierend auf den Items der Itemanalyse (Tabelle 7). Dieser Zusammenschluss wird als vorläufige Testversion mit dem Titel *Programmieren im Sachunterricht für zweite Klassen (PRO SU 2)* definiert. Für die Testinstruktion werden die Formulierungen der verwendeten Items aus den drei Testinstruktionen aus der Hauptstudie zu einer Instruktion zusammengefasst. Die Validierungsstudie wurde Ende August 2023, zeitnah nach den Sommerferien, in einer Grundschule in Köln-Ehrenfeld in vier dritten Klassen durchgeführt. Da die Hauptstudie am Ende des zweiten Schuljahres (April bis Mitte Juni) stattgefunden hat wurden in diesem Kontext dritte Schuljahre gewählt, um einen vergleichbaren Wissensstand vorzufinden. Insgesamt haben  $N = 69$  Kinder an der Studie teilgenommen und  $n = 63$  Testhefte konnten ausgewertet werden. Es wurde keine Unterscheidung zwischen Kindern mit oder ohne Förderschwerpunkt vorgenommen. Zur Bestimmung der Konstruktvalidität wurden der von der Autorin entwickelte Programmieretest im Sachunterricht für zweite Klassen (PRO SU 2), ein Mathematiktest (DEMAT 2+) und ein Intelligenztest (CFT 1-R) an drei aufeinander folgenden Tagen durchgeführt. Im Sinne der Durchführungsobjektivität wurde die vorab festgesetzte Testinstruktion von der Testleitung abgelesen (Bortz & Döring, 2016; Lienert & Raatz, 1998; Moosbrugger & Kelava, 2020).

### 3.7.3 Konstruktvalidität

**DEMAT 2+ und PRO SU 2:** Der Deutsche Mathematiktest für zweite Klassen (DEMAT 2+) basiert auf den Mathematiklehrplänen aller deutscher Bundesländer. Der Test besteht aus den Subthemen Zahleneigenschaften, Längenvergleich, Addition und Subtraktion, Verdoppeln, Division, Halbieren, Rechnen mit Geld, Sachaufgaben und Geometrie. Die Subthemen lassen sich in drei von den fünf beschriebenen inhaltsbezogenen mathematischen Kompetenzen innerhalb des nomologischen Netzwerkes einordnen und zwar in die Bereiche Arithmetik,

---

<sup>10</sup> Innerhalb der Iteration wurde aufgrund eines Berechnungsfehlers das Item 1 nicht in der Validierungsstudie berücksichtigt.

Größe und Geometrie. Die Bereiche „Muster und Strukturen“ sowie „Daten, Häufigkeiten und Wahrscheinlichkeiten“ (KMK, 2005, S. 8) sind im DEMAT 2+ nicht vertreten. In den Themenbereich der Arithmetik fallen die DEMAT 2+ Aufgaben zur Addition, Subtraktion, Division, zum Halbieren und Verdoppeln, zu Zahleigenschaften (gerade/ungerade) sowie Sachaufgaben. Der Bereich der Größe deckt die Teilbereiche Längenvergleich (größer/kleiner) und Rechnen mit Geld ab. Im Bereich der Geometrie werden Aufgaben zu Würfelgebäuden untersucht (Krajewski et al., 2004). Eine Beziehung zwischen den Subtests (Sequenz, Alternative und Iteration) des PRO SU 2 und den mathematischen Inhalten Arithmetik, Größe und Geometrie lässt sich lediglich zwischen der Iteration und der Arithmetik inhaltlich begründen (siehe Kapitel 3.1.1), da die Schleife wiederholt ausgeführt wird und dafür mathematische Fähigkeiten in der Addition bzw. Multiplikation erforderlich sind. Siehe hierzu auch die formulierten Hypothesen zu Forschungsfrage 6.

**CFT 1-R und PRO SU 2:** Der Grundintelligenztest (CFT 1-R) lautet im Original *Culture Fair Intelligence Test*. Dieser basiert auf der von Cattell (1963) entwickelten Intelligenztheorie. Die Intelligenztheorie unterscheidet zwischen fluider (angeborener) und kristalliner (erlernter) Intelligenz. Der Test besteht aus sechs Subtests: Substitution, Labyrinth, Ähnlichkeiten, Reihenfortsetzen, Klassifikationen und Matrizen. Die Subtests decken in unterschiedlicher Intensität lediglich die fluide Intelligenz ab. Zum Lösen der Substitutionsaufgaben ist eine durchschnittliche Beteiligung an der fluiden Intelligenz und ein reproduzierender Anteil der optischen Wahrnehmung notwendig. Die Labyrinth haben eine durchschnittliche Beteiligung an der fluiden Intelligenz. Die optische Wahrnehmung ist in produktiven Anteilen vorhanden. Der Subtest der Ähnlichkeiten umfasst einen mittleren Anteil der fluiden Intelligenz und ein exaktes Wiedererkennen von Figuren. Die drei Subtests haben große Überschneidungen in dem Bereich der Wahrnehmung (Schnelligkeit, Voraussicht im Denken, optische Differenzierung), der visuellen Aufmerksamkeit und dem visomotorischen Entwicklungsstand. Der Subtest Reihenfortsetzen beinhaltet schlussfolgerndes Denken und einen hohen Anteil der fluiden Intelligenz. Die Klassifikationsaufgaben erfordern eine Differenzierung zwischen Details und sie weisen einen hohen Anteil der fluiden Intelligenz auf. Der Subtest der Matrizen erfordert das Erfassen von Fähigkeiten, Regeln und Zusammenhängen bei figuralen Denkaufgaben. Zusätzlich lässt sich eine hohe Beteiligung der fluiden Intelligenz identifizieren. Die drei Subtests Reihenfortsetzen, Klassifikation und Matrizen beinhalten die Merkmale der Grundintelligenz (Weiß, 1969): Beziehungstiftendes Denken sowie das Erkennen von Regelmäßigkeiten und Gesetzmäßigkeiten. In allen drei Subtests hat die fluide Intelligenz einen

hohen Anteil (Weiß & Osterland, 2013). Die zu erwartenden Beziehungen zwischen dem Programmieretest (PRO SU 2) und dem Intelligenztest (CFT 1-R) werden insgesamt als geringfügig angenommen, da die fluide Intelligenz in allen CFT 1-R Aufgaben dominiert. Bei den Programmieraufgaben handelt es sich hingegen um Aufgaben der kristallinen Intelligenz (siehe Forschungsfrage 7).

## 4 Ergebnisse

In diesem Kapitel werden die Ergebnisse des methodischen Vorgehens der Reihe nach dargelegt. Dafür werden zu Beginn innerhalb des Kapitels der Itemanalyse (4.1) die Ergebnisse aus den Berechnungen der Rohwertverteilung, der Itemschwierigkeiten, der Itemtrennschärfe und der Dimensionalitätsprüfung präsentiert. Daran anschließend erfolgt eine Betrachtung weiterer Berechnungen (Kapitel 4.2) aus der Hauptstudie. Hier zu nennen sind die Ergebnisse der Reliabilität und der Häufigkeitsverteilungen. Kapitel 4.3 thematisiert den Ergebnisraum. In diesem Kontext sind die finalen Codieranleitungen abgebildet. Die Ergebnisse aus der Rasch-Analyse sind im nachfolgenden Kapitel (4.4) dargelegt. Beendet wird das Kapitel mit der Beschreibung der Korrelationsberechnungen aus der Validierungsstudie zur Bestimmung von konvergenten und diskriminanten Validitäten (4.5).

### 4.1 Itemanalyse

Dieses Kapitel behandelt die Ergebnisse der Rohwertverteilung, der Itemschwierigkeiten, der Itemtrennschärfe und der Überprüfung der Dimensionalität.

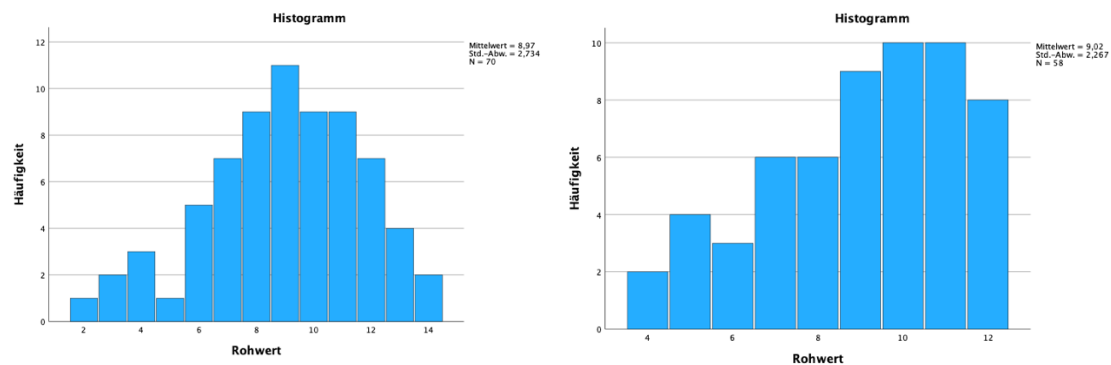
#### 4.1.1 Rohwertverteilung

Für die Berechnung des Kolmogoroff-Smirnov-Tests werden die Testhefte der Sequenz A, Sequenz B, Iteration (Bedingte Schleife und Zählschleife) A und Iteration (Bedingte Schleife und Zählschleife) B getrennt voneinander berechnet, da unterschiedliche Aufgaben in den Testheften vorhanden sind und es sich um einen Paralleltest mit unterschiedlichen Aufgaben handelt. Für das Testheft der Alternative werden die Testversionen A und B zusammen berücksichtigt, da es sich um eine echte Pseudo-Parallelform des Tests handelt. Die Lösungen der Pseudo-Parallelaufgaben sind lediglich anders sortiert. Für alle Testhefte, außer dem Testheft A der Sequenz, ergeben sich bei der Berechnung des Kolmogoroff-Smirnov-Tests signifikante Werte von  $p < 0.05$ . Somit liegt bei den Testheften der Sequenz B, der Alternative A/B und der Iteration A/B keine Normalverteilung vor. Lediglich das Testheft A der Sequenz weist mit einem Wert von  $p = 0.06$  (auf zwei Stellen hinter dem Komma gerundet) keine Signifikanz auf, weshalb eine Normalverteilung angenommen werden kann. Da eine Normalverteilung vorliegt wird für das Testheft der Sequenz A zusätzlich auf die Schiefe und Kurtosis (= Wölbung) eingegangen. Die Schiefe liegt bei  $-0.43$ , weshalb die Daten nach links geneigt sind. Hinsichtlich der Kurtosis liegt der Exzesswert bei  $-0.15$ , weshalb der Graph als flachgipflig beschrieben werden kann.

Auch die grafische Prüfung mittels Histogramme und Q-Q-Diagramme auf eine Normalverteilung hin bestätigen die zuvor beschriebenen Ergebnisse. Die Normalverteilung für die Sequenz (Version A) lässt sich mittels grafischer Prüfung zusätzlich bestätigen (linkes Diagramm 1).

## Diagramm 1

### *Rohwertverteilung der Sequenz*

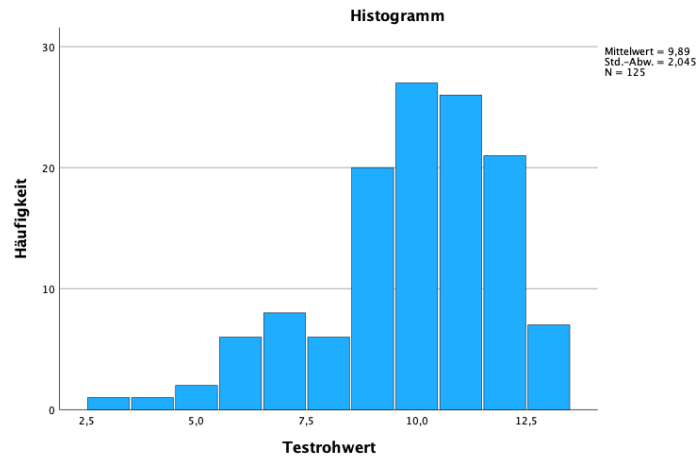


*Anmerkung.* Diagrammerstellung mittels IBM® SPSS®. Links = Sequenz A. Rechts = Sequenz B.

Bei näherer Betrachtung des Histogramms zu dem Testheft B der Sequenz (rechtes Diagramm 1) ist zu erkennen, dass im Rohwertbereich von 9 bis 12 Punkten (Spanne 4 bis 12 Punkten) über die Hälfte der Kinder zu verorten ist. Die Betrachtung des Histogramms zu dem Testheft der Alternative (Version A und B) weist eine Schiefe von -0.88 auf und beschreibt somit linksschiefe Daten, da die Verteilung auf der linken Seite flacher abfällt als auf der rechten Seite. Die entsprechende Verteilung ist folgenden Diagramm 2 zu entnehmen.

## Diagramm 2

### Rohwertverteilung der Alternative

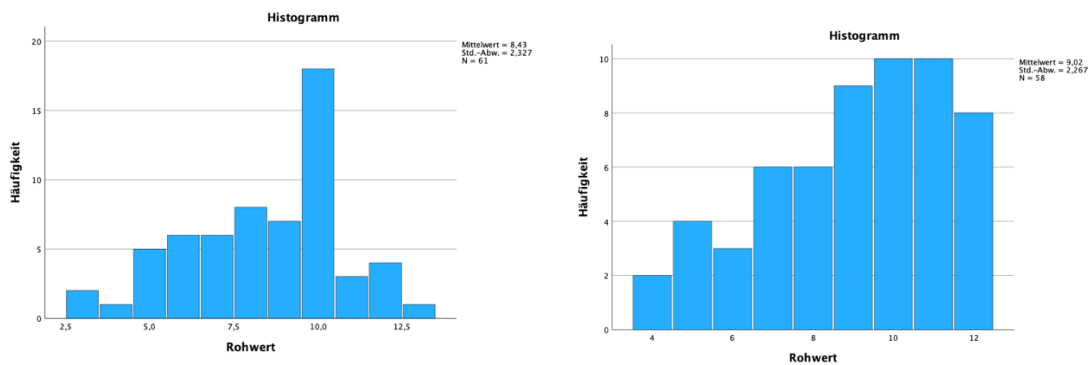


Anmerkung. Diagrammerstellung mittels IBM® SPSS®.

Bei der Betrachtung des Histogramms des Testhefts A der Iteration wird deutlich, dass einige Aufgaben leicht zu lösen waren (linkes Diagramm 3). Die Betrachtung des Histogramms zu dem Testheft der Iteration B (rechtes Diagramm 3) ergibt einen hohen Anteil von Kindern mit einem Rohwert von 9 bis 12 Punkten (Spanne 4 bis 12 Punkte).

## Diagramm 3

### Rohwertverteilung der Iteration



Anmerkung. Diagrammerstellung mittels IBM® SPSS®. Links = Iteration A. Rechts = Iteration B.

Insgesamt lässt sich festhalten, dass nur für das Testheft der Sequenz A eine Normalverteilung vorliegt. Das Anwenden inferenzstatistischer Verfahren kann dennoch für alle Testhefte erfolgen, da ab einer Stichprobengröße von über 30 Kindern der zentrale Grenzwertsatz gilt und eine Normalverteilung angenommen werden kann (Bortz & Döring, 2016; Bortz & Schuster, 2010; Pólya, 1920). Dadurch können Rückschlüsse von der Stichprobe auf die Grundgesamtheit gezogen werden.

#### 4.1.2 Itemschwierigkeit

Es werden nacheinander die Berechnungen der Schwierigkeitsindexe für das Testheft der Sequenz A, der Sequenz B, der Alternative, der Iteration A und der Iteration B vorgestellt. Dabei wird die Bedingte Schleife (Version A und B) durch die Items 1 bis 7 und die Zählschleife (Version A und B) durch die Items 8 bis 15 repräsentiert.

**Sequenz A:** Die Itemschwierigkeiten für das Testheft der Sequenz A liegen bei einem prozentualen Schwierigkeitsindex zwischen 28.57 und 61.43 (Tabelle 5). Der mittlere prozentuale Schwierigkeitsindex liegt bei 47.66 und liegt somit im mittleren Bereich. Die Itemschwierigkeiten sind von leicht bis schwierig überwiegend gleichmäßig verteilt. Lediglich im leichten Bereich sind keine Aufgaben zu verzeichnen. Insgesamt bleiben zehn Items erhalten: Aufgabe 3 aus dem Kognitiven Bereich „reproduzieren“, die Aufgaben 7 bis 10 aus dem Kognitiven Bereich „anwenden“ und die Aufgaben 11 bis 15 aus dem Kognitiven Bereich „reflektieren“.

**Sequenz B:** Die Itemschwierigkeiten für das Testheft der Sequenz B liegen bei einem prozentualen Schwierigkeitsindex zwischen 31.25 und 67.19 (Tabelle 5). Der mittlere prozentuale Schwierigkeitsindex liegt bei 47.87 und befindet sich somit im mittleren Bereich. Bei genauerer Betrachtung der Verteilung der prozentualen Schwierigkeiten in der Spanne von 19.0 und 84.0 ist eine gleichmäßige Verteilung festzustellen. Lediglich in den Randbereichen sind keine Aufgaben zu finden. Insgesamt bleiben neun Items erhalten: Aufgabe 3 aus dem Kognitiven Bereich „reproduzieren“, die Aufgaben 7, 9 und 10 aus dem Kognitiven Bereich „anwenden“ und die Aufgaben 11 bis 15 aus dem Kognitiven Bereich „reflektieren“.

**Alternative:** Die Berechnung der Itemschwierigkeiten für das Testheft der Alternative ergeben für die Items 1 bis 6 und 10, 12 und 13 einen prozentualen Schwierigkeitsindex zwischen 60.0 und 78.37 (Tabelle 5). Somit sind alle Items in einem leichten Schwierigkeitsbereich zu

verorten. Der mittlere prozentuale Schwierigkeitsindex liegt bei 69.33 und ist ebenfalls im leichten Bereich anzusiedeln. Insgesamt bleiben neun Items erhalten: Die Aufgaben 1 bis 6 aus dem Bereich des Reproduzierens und die Aufgaben 10, 12 und 13 aus dem Bereich des Reflektierens.

**Iteration A:** Für das Testheft der Iteration A ergeben die Berechnung der Itemschwierigkeiten für die Items 1 bis 6 und 10 bis 13 einen prozentualen Schwierigkeitsindex zwischen 19.33 und 78.69 (Tabelle 5). Der mittlere prozentuale Schwierigkeitsindex liegt bei 58.45, was einer mittleren Aufgabenschwierigkeit entspricht. Bei Betrachtung der Verteilung der prozentualen Schwierigkeiten in der Spanne von 19.0 bis 84.0 ist eine ungleiche Verteilung zu verzeichnen. So sind fünf Aufgaben leicht, zwei Aufgaben mittelmäßig und zwei Aufgaben schwierig zu lösen. Insgesamt bleiben zehn Items erhalten: die Aufgaben 1 bis 4 und 10 aus dem Bereich des Reproduzierens, die Aufgaben 5, 6, 11, 12 aus dem Bereich des Anwendens und die Aufgabe 13 aus dem Bereich des Reflektierens.

**Iteration B:** Für das Testheft der Iteration B ergibt sich für die Items 1 bis 6 und 11 bis 13 ein prozentualer Schwierigkeitsindex zwischen 19.33 und 84.48 (Tabelle 5). Der mittlere prozentuale Schwierigkeitsindex liegt bei 61.20, was einem leichten Schwierigkeitsbereich entspricht. So sind sechs von neun Aufgaben in einem leichten Schwierigkeitsbereich anzusiedeln. Eine Aufgabe ist im mittleren Bereich und zwei Aufgaben sind im schwierigen Bereich wiederzufinden. Insgesamt bleiben neun Items erhalten: die Aufgaben 1 bis 4 aus dem Bereich des Reproduzierens, die Aufgaben 5, 6, 11 und 12 aus dem Bereich des Anwendens und Aufgabe 13 aus dem Bereich des Reflektierens. Die Aufgabe 11 (B) wird weiter beibehalten, damit ausreichend Items im Bereich der Zählschleife (B) für weitere Berechnungen zur Verfügung stehen.

Weiter zeigen die Ergebnisse, dass die gedrehten oder gespiegelten Abbildungen der in der Testversion B der Sequenz und Iteration beibehaltenen Items zu unterschiedlichen prozentualen Schwierigkeiten führten. Dafür werden die Aufgaben hinsichtlich der Schwierigkeit und Leserichtung miteinander verglichen. Es werden nur Aufgaben gewählt, welche von der Laufrichtung der Figur abhängig sind. Somit werden die Sequenz-Aufgaben 7, 9, 10, 11, 12, 13, 14 und 15 sowie die Aufgaben 11 und 12 der Iteration analysiert. Die Aufgaben der Bedingten Schleife entfallen, da die Abbildungen anders aufgebaut sind. Der Vergleich aller Aufgaben ergibt, dass in acht von zehn Fällen Aufgaben, in denen die Figur von links nach rechts bewegt

werden muss, einfacher für die Kinder zu lösen sind als Aufgaben, in denen die Figur von rechts nach links bewegt werden muss. Zusätzlich zeigt die Auswertung der Aufgabe 13 der Zählschleife, dass die Aufgabe, in der die Figur von oben nach unten bewegt wird, einfacherer für die Kinder zu lösen ist als die Aufgabe, in der die Figur von unten nach oben bewegt werden muss.

Ein Zusammenschluss aller Testitems zur Erfassung der grundlegenden Programmierkompetenz im Sachunterricht ergibt nach der Berechnung des Schwierigkeitsindexes einen Pool von 42 Einzelaufgaben. Testitems, welche in der Testversion A und B gleich sind, werden nur einfach gezählt.

**Tabelle 5**

*Itemschwierigkeiten der Testhefte Sequenz, Alternative und Iteration*

Item	Sequenz A	Sequenz B	Alternative A und B	Iteration A	Iteration B
	P	P	P	P	P
1	x	x	78.37	19.33	
2	x	x	70.0	58.40	
3	41.79		67.6	68.03	71.55
4	x	x	67.6	55.74	76.72
5	x	x	76.0	72.27	
6	x	x	74.8	36.97	
7	44.29	51.56	x	x	x
8	81.43	x	x	x	x
9	57.14	42.19	x	x	x
10	58.57	67.19	60.0	75.41	x
11	50.0	54.69	x	78.69	84.48
12	61.43	65.63	60.8	55.74	68.97
13	57.14	40.63	68.8	63.93	62.07
14	30.0	35.94	-	-	-
15	28.57	31.25	-	-	-

*Anmerkung.*  $P$  = prozentualer Schwierigkeitsindex. Werte zwischen den Spalten berücksichtigen errechnetes  $P$  aus A und B. x= aussortierte Items über Schwierigkeitsberechnung, da  $P < 0.19$  oder  $P > 0.84$ . Bedingte Schleife: Aufgabe 1 bis 7. Zählschleife: Aufgabe 8 bis 13. Grün = Kognitive Dimension „reproduzieren“. Gelb = Kognitive Dimension „anwenden“. rot = Kognitive Dimension „reflektieren“. - = keine Items vorhanden.

### 4.1.3 Itemtrennschärfe

Die Ergebnisse der Trennschärfenberechnungen sind Tabelle 6 zu entnehmen. Die Tabelle zeigt die Trennschärfenberechnungen aufgeschlüsselt nach den sieben Skalen: Sequenz A, Sequenz B, Alternative A/B, Bedingte Schleife A (Item 1 bis 7), Bedingte Schleife B (Item 1 bis 7), Zählschleife A (Item 8 bis 15) und Zählschleife B (Item 8 bis 15). Für die Sequenz A ergibt sich ein mittlerer (Item 7, 8, 10 und 13) bis hoher Zusammenhang (Item 9, 11, 12, 14 und 15) der Items mit der Skala der Sequenz A. Die Berechnungen für das Testheft B der Sequenz ergeben überwiegend einen hohen bis sehr hohen (Item 7, 9, 11 bis 15) Zusammenhang der Items mit der Skala der Sequenz B. Lediglich das Item 10 zeigt eine mittlere Korrelation mit der Skala. Die Korrelationsberechnungen der Alternative liefern mittlere (Item 1, 2, 5 und 6) bis hohe (Item 4) Zusammenhänge. Die Betrachtung der Korrelationsberechnungen für die Bedingte Schleife A zeigen schwache (Item 2 und 6) bis mittlere (Item 1, 3 und 4) Korrelationen mit der Skala. Die Berechnungen der Zählschleife A ergeben mittlere (Item 11 bis 13) bis hohe (Item 10) Zusammenhänge zwischen den Items und der Skala. Für die Bedingte Schleife B (Item 1, 3 bis 6) ergeben sich gerundet für alle Items mittlere Zusammenhänge zwischen den Items und der Skala. Bei Betrachtung der Korrelationskoeffizienten der Zählschleife B liegen mittlere (Item 11 und 12) bis hohe (Item 13) Zusammenhänge vor.

Tabelle 6

*Itemtrennschärfe (korrigierte Item-Skala-Korrelationen)*

Aufgabe	Sequenz A	Sequenz B	Alternative A und B	Iteration A	Iteration B
	r	r	r	r	r
1	x	x	.39	.31	.29
2	x	x	.44	.28	.11
3	.15	.01	.22	.31	.33
4	x	x	.59	.34	.45
5	x	x	.41	.20	.32
6	x	x	.45	.28	.46
7	.38	.47	x	x	x
8	.40	x	x	x	x
9	.57	.56	x	x	x
10	.46	.41	-.02	.53	x
11	.52	.49	x	.46	.34
12	.56	.55	.04	.43	.45
13	.37	.55	-.03	.48	.53
14	.55	.58	-	-	-
15	.51	.61	-	-	-

*Anmerkung.*  $r$  = Itemtrennschärfen (korrigierte Item-Skala-Korrelationen). x = aussortierte Items über die Schwierigkeitsberechnungen, da  $P < 0.19$  oder  $P > 0.84$ . Grau hinterlegte Felder entsprechen den aussortierten Items mittels Trennschärfeermittlung, da  $r < 0.28$ . Bedingte Schleife (Aufgaben 1 bis 7) und Zählschleife (Aufgaben 8 bis 13) werden jeweils als eine separate Skala erfasst.

Zusammenfassend lässt sich sagen, dass von ursprünglich 42 Items sieben Items (Item 3 der Sequenz ist identisch) aufgrund mangelnder Trennschärfe entfernt werden. Somit verbleiben 35 Items im Itempool. In der Tabelle sind insgesamt 39 Items abgebildet, wobei vier Aufgaben identisch sind. Somit sind die vier Items 1, 6, 11 und 12 der Iteration von den 39 Items zu subtrahieren, um auf den Itempool von 35 Items zu gelangen.

#### 4.1.4 Dimensionalitätsprüfung

Eine Durchführung der Dimensionalitätsprüfung mittels Faktorenanalyse erscheint an dieser Stelle als wenig sinnvoll, da die Überprüfung der Voraussetzungen nur unzureichend erfüllt werden kann. Eine Itemanzahl von vier Items pro Themenheft, Sequenz A mit neun Items, Sequenz B mit acht Items, Alternative A/B mit fünf Items und Iteration A und B mit jeweils sieben Items, gilt als erfüllt. Jedoch werden die anderen Voraussetzungen nur bedingt bis gar nicht eingehalten. So liegt bei wenigen Items eine Kommunalität von über 0.60 vor: Innerhalb der Sequenz A und B finden sich jeweils drei Items, bei der Alternative ebenfalls drei Items, bei der Iteration A kein Item und bei der Iteration B drei Items. Zusätzlich ist in allen Testheften eine deutlich kleinere Stichprobengröße als 250 Personen (Tabelle 3) vorhanden. Weiter existiert insgesamt eine geringe Variablenanzahl pro Faktor (maximal 4), was eine Stichprobe von über 300 Personen und einen Ladungswert von 0.60 fordert. Nur wenige Items erreichen einen Ladungswert von über 0.60 und könnten somit berücksichtigt werden: Sequenz A mit drei Items, Sequenz B mit vier Items, Alternative A/B mit drei Items, Iteration A mit drei Items und Iteration B mit zwei Items. Nur in dem Testheft der Sequenz B werden die geforderten vier Variablen pro Faktor erreicht. Dies setzt wiederum eine Stichprobengröße von mehr als 300 Personen voraus. Insgesamt ist eine Faktorenanalyse aus den zuvor genannten Gründen ungeeignet, weshalb die Dimensionalitätsprüfung über die Berechnung des Mittelwertes der Item-Interkorrelation erfolgt.

Der Mittelwert der Item-Interkorrelation für das Testheft der Sequenz A hat einen Wert von 0.30, weshalb alle Items (7 bis 15) beibehalten werden können. Für das Testheft der Sequenz B liegt ein Korrelationswert von 0.36 vor. Auch hier können alle Items (7, 9 bis 15) weiter berücksichtigt werden. Der Mittelwert der Item-Interkorrelation für das Testheft der Alternative A/B liegt bei 0.40. Daraus ergibt sich folglich, dass alle Items (1, 2, 4 bis 6) übernommen werden. Die Korrelationsberechnungen für den Mittelwert der Bedingte Schleife A (Item 1 bis 4 und 6) weist einen Wert von 0.19 auf, welcher nah dem geforderten Wert von 0.20 liegt und somit akzeptiert wird. Alle Items der Bedingten Schleife A werden berücksichtigt. Der Korrelationswert für die Bedingte Schleife B weist einen Wert von 0.25 auf, weshalb alle Aufgaben (1, 3 bis 6) beibehalten werden können. Die Berechnungen für den Mittelwert der Item-Interkorrelation weist für die Zählschleife A einen Wert von 0.36 und für die Zählschleife B einen Wert von 0.35 auf. Alle Items der Zählschleife A (10 bis 13) und der Zählschleife B (11 bis 13) werden weiter berücksichtigt.

Nach abschließender Itemanalyse besteht nach Zusammenschluss aller Testitems zur Erfassung der *grundlegenden Programmierkompetenz im Sachunterricht* ein Itempool von 35 Einzelaufgaben (Tabelle 7), welcher einer geforderten Mindestmenge in einfachen Leistungstests entspricht (Lienert & Raatz, 1998). In diesem Zusammenhang existieren 17 unterschiedliche Items zum Inhalt der Sequenz, 5 unterschiedliche Items zum Inhalt der Alternative und 13 unterschiedliche Items zum Inhalt der Iteration.

**Tabelle 7**

*Itempool nach der Itemanalyse*

Testversion	Aufgaben														
Sequenz	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A						--	x	x	x	x	x	x	x	x	x
B						--	x	--	x	x	x	x	x	x	x
Alternative	1	2	3	4	5	6	7	8	9	10	11	12	13		
A	x	x	--	x	x	x	--	--	--	--	--	--	--		
B	x	x	--	x	x	x	--	--	--	--	--	--	--		
Iteration	1	2	3	4	5	6	7	8	9	10	11	12	13		
A	x	x	x	x	--	x	--	--	--	x	x	x	x		
B	x	--	x	x	x	x	--	--	--	--	x	x	x		

*Anmerkung.* -- = aussortierte Items. x = beibehaltende Items, graue Felder = Pseudo-Parallelaufgaben in Testversion A und B.

## 4.2 Weitere Berechnungen

In diesem Kapitel werden die Reliabilität und die Häufigkeitsverteilungen der Testwerte auch nach Geschlechtern getrennt dargelegt. Die Berechnungen basieren auf den erhobenen Daten der Hauptstudie.

### 4.2.1 Reliabilität

An dieser Stelle werden die notwendigen Daten aus der Hauptstudie zusammenfassend für die Reliabilitätsberechnungen dargelegt. Die Eichstichprobe besteht aus  $N = 378$  Grundschulkindern der zweiten Klasse. Das Testheft der Sequenz (Version A) haben  $n = 70$

Kinder, der Sequenz (Version B) haben  $n = 64$  Kinder, der Alternative (Version A und B) haben  $n = 125$ , der Iteration (Version A) haben  $n = 61$  und der Iteration (Version B) haben  $n = 58$  Kinder bearbeitet. Zur Berechnung der Zuverlässigkeit der Testteile werden die Items, welche nach der Itemanalyse im Itempool verblieben sind, mit Hilfe des Konsistenzkoeffizienten Cronbachs Alpha errechnet. Die Bezifferung der Aufgaben wird weiter beibehalten.

Die Berechnung des Reliabilitätskoeffizienten mittels Cronbachs Alpha liefern statistisch gute bis zufriedenstellende gesicherte Ergebnisse. Für die Sequenz (Version A) ergibt sich unter Berücksichtigung der Aufgaben 7 bis 15 ein Wert von  $\alpha = 0.80$ , daher ist keine Entfernung von Items notwendig. Für die Sequenz (Version B) ergibt sich ein  $\alpha = 0.82$  für die Items 7 und 9 bis 15. Alle Items können weiter benutzt werden. Für die Alternative (Version A/B) ergibt sich ein Cronbachs Alpha von  $\alpha = 0.77$  für die Items 1, 2, 4 bis 6, infolgedessen können alle Items weiter beibehalten werden. Für die Bedingte Schleife (Version A) liegt für die Items 1, 2, 3, 4 und 6 ein  $\alpha = 0.53$  vor. Durch die Entfernung der Items 1, 2 und 6 wird ein Wert von  $\alpha = 0.70$  erreicht. Da allerdings eine Bandbreite von verschiedenen Items im Testheft angestrebt wird, werden die Items 1, 2 und 6 weiter berücksichtigt. Der Reliabilitätskoeffizient für die Bedingte Schleife (Version B) liegt bei  $\alpha = 0.62$ . Alle Items (1, 3 bis 6) werden weiter berücksichtigt, da durch eine Entfernung der Items eine niedrigere Korrelation erreicht würde. Die Reliabilitätsberechnungen für die Zählschleife (Version A) ergeben ein  $\alpha = 0.68$  (Items 10 bis 13) und für die Zählschleife (Version B) ein  $\alpha = 0.62$  (Items 11 bis 13). Da insgesamt nur drei beziehungsweise vier Items innerhalb der Zählschleife vorliegen und eine Bandbreite von verschiedenen Items im Testheft angestrebt wird, werden alle Items beibehalten.

Resümierend lässt sich festhalten, dass für die Testhefte der Sequenz (Version A und Version B) und der Alternative (Version A/B) ausreichende Reliabilitäten ( $\alpha \geq 0.77$ ) vorliegen. Lediglich für die Themenhefte der Iteration liegen die Korrelationskoeffizienten bei  $\alpha = 0.53$  und  $\alpha = 0.70$ , was in der Literatur als akzeptierte Werte für frühe Forschungsstadien angesehen wird (Field, 2018; Schecker, 2014). Zusätzlich spricht für die Beibehaltung der Items, dass es sich um einen Leistungstest handelt. Diese haben das Ziel eine inhaltliche Bandbreite abzudecken, was mit einer niedrigeren Reliabilität einhergeht (Schmitt, 1996).

#### 4.2.2 Häufigkeitsverteilungen

In diesem Kapitel werden zu Beginn basierend auf den Ergebnissen der Itemanalyse erneut die Trennschärfe und die interne Konsistenz ermittelt, da in diesem Kontext die Items der

Bedingten Schleife und der Zählschleife zusammen berechnet werden. Basierend auf den voran beschriebenen Berechnungen erfolgt die Ermittlung der Häufigkeitsverteilungen für die Testwerte (Version A und B). Anschließend werden die Häufigkeitsverteilungen der Testwerte in Bezug auf das Geschlecht dargelegt. In diesem Kontext werden die Ergebnisse in Folgender Reihenfolge ausgeführt: Sequenz (A und B), Alternative (A und B) und Iteration (A und B).

**Kennwerte:** Eine Zusammenfassung der Kennwerte Aufgabenschwierigkeit, Trennschärfe und interne Konsistenz sind in Tabelle 8 dargestellt. Insgesamt liegen die Werte der Trennschärfe über einem Wert von  $r > 0.30$  außer innerhalb der Iteration A (Items 1 und 2) und Iteration B (Items 5 und 12). Hier liegen die Werte zwischen 0.20 und 0.24. Trotz der geringen Trennschärfen innerhalb der Iteration werden die Items aus inhaltlichen Aspekten beibehalten (Bortz & Döring, 2016). Lediglich das Item 6 der Iteration A wird aufgrund einer geringen Trennschärfe aus dem Itempool entfernt. Die Reliabilitätsberechnungen für die Testhefte ergeben akzeptierte Werte zwischen  $\alpha = 0.67$  und  $\alpha = 0.80$ , denn häufig werden Skalen mit Werten um 0.70 verwendet (Bortz & Döring, 2016; Moosbrugger & Kelava, 2020).

**Tabelle 8***Kennwerte des PRO SU 2*

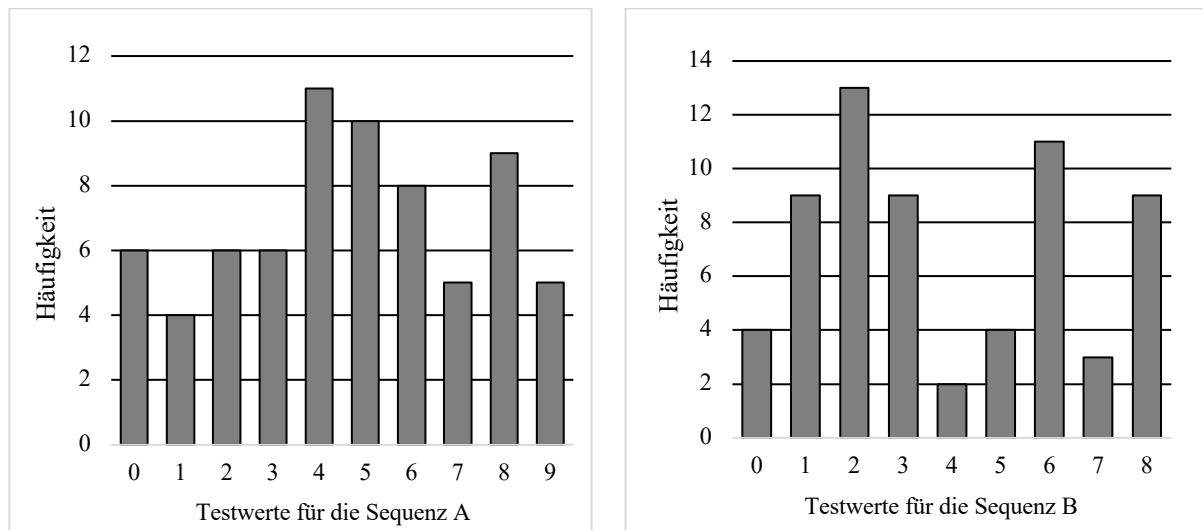
Aufgaben	Aufgabenschwierigkeit	Trennschärfe	Interne Konsistenz
Sequenz A			
7	.44	.40	
8	.81	.39	
9	.57	.55	
10	.59	.44	.80
11	.50	.53	
12	.61	.57	
13	.57	.40	
14	.30	.56	
15	.29	.52	
Sequenz B			
7	.52	.46	
9	.42	.50	
10	.67	.40	
11	.55	.52	.82
12	.66	.58	
13	.41	.58	
14	.36	.64	
15	.31	.63	
Alternative			
1	.78	.46	
2	.70	.37	
4	.68	.74	.77
5	.76	.51	
6	.75	.65	
Iteration A			
1	.19	.20	
2	.58	.23	
3	.68	.55	
4	.56	.47	.69
10	.75	.40	
11	.79	.43	
12	.56	.48	
13	.64	.37	
Iteration B			
1	.19	.33	
3	.72	.47	
4	.77	.43	
5	.72	.24	.67
6	.37	.35	
11	.84	.32	
12	.69	.24	
13	.62	.54	

*Anmerkung.* Angelehnt an Krajewski et al., 2004, S. 29. Die interne Konsistenz für die Bedingte Schleife und die Zählschleife wurde zusammen berechnet.

**Häufigkeitsverteilung Sequenz:** Die Häufigkeitsverteilungen der Testwerte für die Sequenz (Diagramm 4) weisen für die Version A einen höheren Mittelwert ( $M = 4.7$ ) im Vergleich zur Version B ( $M = 3.9$ ) auf. So wird in der Version A durchschnittlich ein höherer Testwert als in der Version B erreicht. Die Standardabweichung liegt bei beiden Versionen mit  $SD = 2.65$  (Version A) und  $SD = 2.59$  (Version B) nah beieinander. In diesem Zusammenhang sei darauf hingewiesen, dass in der Testversion A neun Punkte und in der Testversion B acht Punkte maximal erreicht werden können.

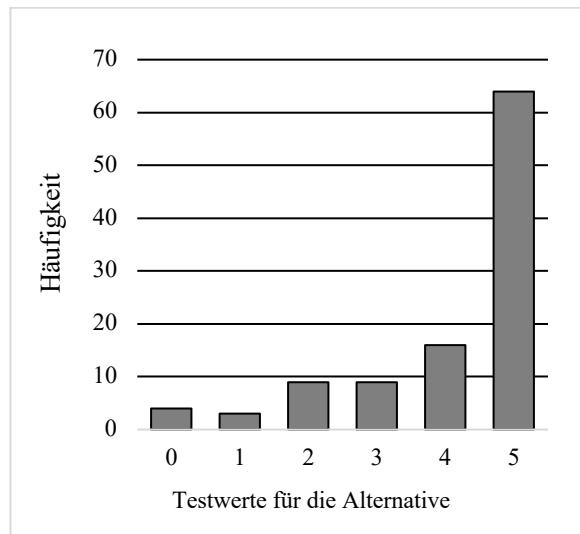
#### Diagramm 4

##### SEQUENZ Häufigkeitsverteilung der Testwerte



*Anmerkung.* Sequenz A (links):  $N = 70$ ;  $n = 70$ ;  $M = 4.7$ ;  $SD = 2.65$ . Sequenz B (rechts):  $N = 64$ ;  $n = 64$ ;  $M = 3.9$ ;  $SD = 2.59$ .

**Häufigkeitsverteilung Alternative:** Die Häufigkeitsverteilung der Testwerte für die Alternative (Diagramm 5) zeigen, dass im Durchschnitt 4 von 5 möglichen Punkten von den Kindern erreicht wurden. Die durchschnittliche Streuung liegt bei  $SD = 1.38$ .

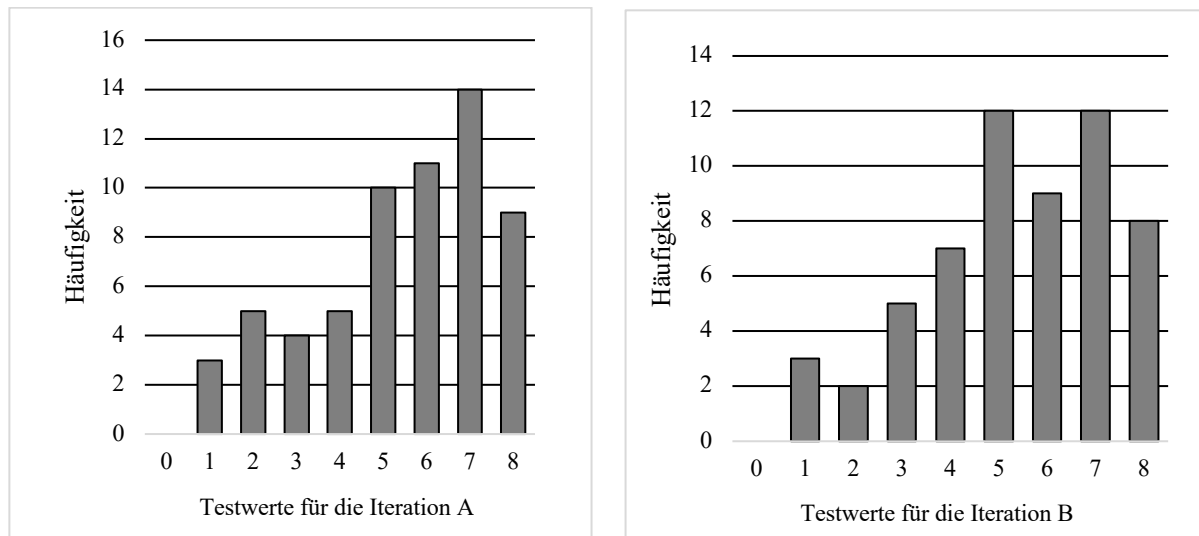
**Diagramm 5***ALTERNATIVE Häufigkeitsverteilung der Testwerte*

Anmerkung.  $N = 125$ ;  $n = 105$ ;  $M = 4.1$ ;  $SD = 1.38$ .

**Häufigkeitsverteilung Iteration:** Die Häufigkeitsverteilung der Testwerte für die Iteration (Diagramm 6) weisen in den Versionen A und B eine durchschnittliche Punktzahl von 5 Punkten auf. Die Betrachtung des Mittelwerts der Versionen A und B entspricht einem hohen Wert in Bezug auf die Gesamtpunktzahl von 8 Punkten. Die durchschnittliche Streuung rund um den Mittelwert liegt bei beiden Versionen gerundet bei  $SD = 2$ .

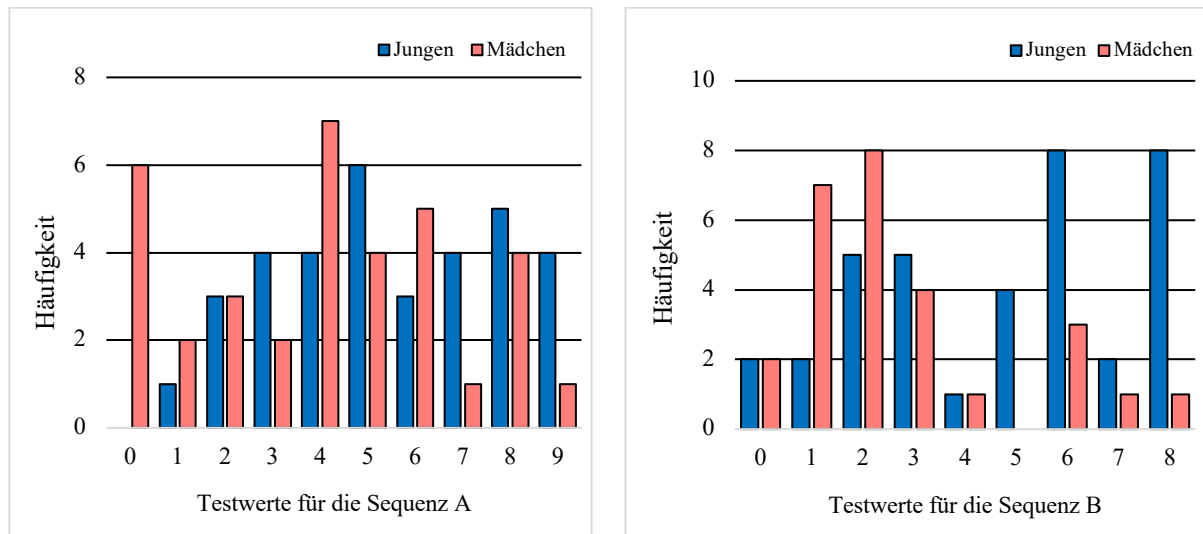
## Diagramm 6

### ITERATION Häufigkeitsverteilung der Testwerte



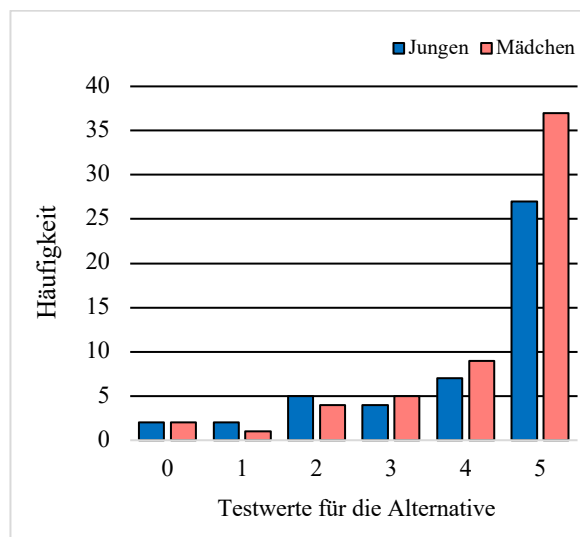
Anmerkung. Iteration A (links):  $N = 61$ ;  $n = 61$ ;  $M = 5.4$ ;  $SD = 1.95$ . Iteration B (rechts):  $N = 58$ ;  $n = 58$ ;  $M = 5.4$ ;  $SD = 2.04$ .

**Häufigkeitsverteilung nach Geschlecht – Sequenz:** Die Häufigkeitsverteilung der Testwerte nach Geschlecht für die Sequenz (Version A) zeigen (Diagramm 7), dass Jungen einen durchschnittlichen Punktwert von 5.5 Punkten erzielen. Mädchen erzielen durchschnittlich einen Wert von 4.0. Die Standardabweichung innerhalb der Jungen-Gruppe liegt mit  $SD = 2.35$  etwas niedriger als bei der Mädchen-Gruppe ( $SD = 2.71$ ). Der Levene-Test ist mit  $p = 0.661$  nicht signifikant, weshalb die Varianzen gleich sind. Der  $t$ -Test liefert mit einem Wert von  $p = 0.421$  ebenfalls keine Signifikanz. Bei der Sequenz (Version B) ist festzustellen, dass die Jungen einen Mittelwert von  $M = 4.8$  und eine Standardabweichung von  $SD = 2.55$  erzielen. Die Mädchen erreichen im Vergleich zu den Jungen (Version B) einen deutlich geringen Mittelwert von  $M = 2.7$  mit einer niedrigeren Standardabweichung von  $SD = 2.15$ . Die Überprüfung der Varianzen ergeben eine Varianzhomogenität ( $p = 0.094$ ). Die Überprüfung der Mittelwertunterschiede ergeben einen signifikanten Unterschied ( $p < 0.001$ ), weshalb im nächsten Schritt die Effektstärke betrachtet wird. Die Effektstärke nach Cohen (1988) liegt bei  $d = 0.886$ , was einen starken Effekt charakterisiert. In diesem Zusammenhang bedeutet dies, dass das Geschlecht ausschlaggebend für das Erreichen einer höheren oder niedrigeren Punktzahl ist.

**Diagramm 7***SEQUENZ Häufigkeitsverteilung der Testwerte nach Geschlecht*

*Anmerkung.* Sequenz A (links) Jungen:  $N = 35$ ;  $n = 34$ ;  $M = 5.5$ ;  $SD = 2.35$ . Sequenz A (links) Mädchen:  $N = 35$ ;  $n = 35$ ;  $M = 4.0$ ;  $SD = 2.71$ . Sequenz B (rechts) Jungen:  $N = 37$ ;  $n = 37$ ;  $M = 4.8$ ;  $SD = 2.55$ . Sequenz B (rechts) Mädchen:  $N = 27$ ;  $n = 27$ ;  $M = 2.7$ ;  $SD = 2.15$ .

**Häufigkeitsverteilung nach Geschlecht – Alternative:** Die Häufigkeitsverteilung der Testwerte nach Geschlecht für die Alternative (Diagramm 8) weisen ähnliche Werte auf. So liegt der Mittelwert bei beiden Gruppen gerundet bei 4.0 und die Standardabweichung bei 1.48 (Jungen) beziehungsweise 1.3 (Mädchen). Die Varianzüberprüfung mittels Levene-Tests zeigt, dass eine Varianzhomogenität vorliegt. Die Betrachtung der  $t$ -Werte konnte keine signifikanten Unterschiede ( $p = 0.375$ ) hinsichtlich des Geschlechts erfassen.

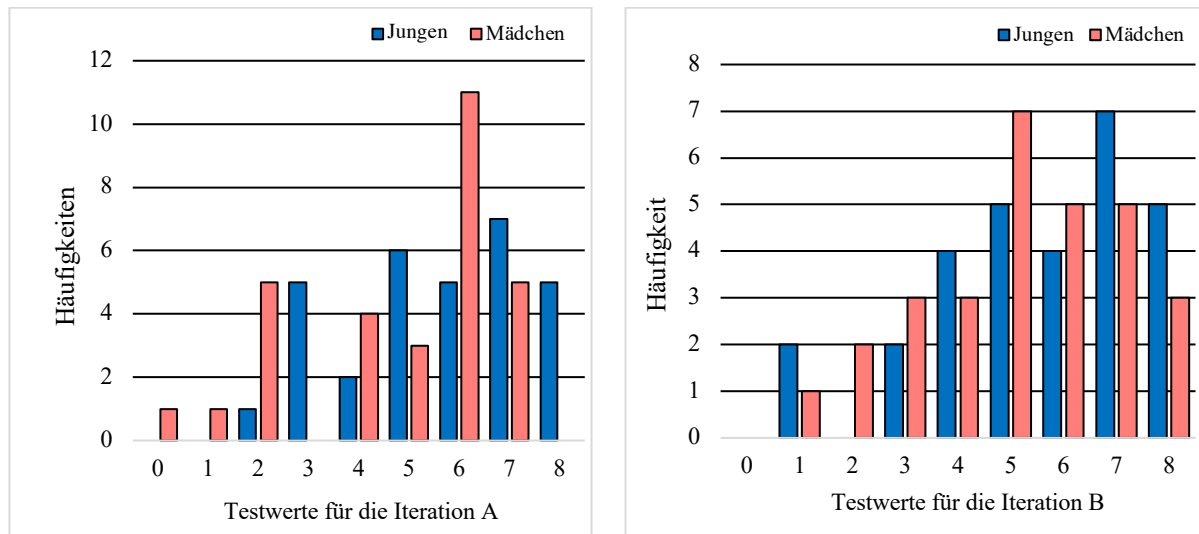
**Diagramm 8***ALTERNATIVE Häufigkeitsverteilung der Testwerte nach Geschlecht*

*Anmerkung.* Alternative Jungen:  $N = 55$ ;  $n = 47$ ;  $M = 4.0$ ;  $SD = 1.48$ . Alternative Mädchen:  $N = 67$ ;  $n = 58$ ;  $M = 4.2$ ;  $SD = 1.30$ .

**Häufigkeitsverteilung nach Geschlecht – Iteration:** Die Häufigkeitsverteilung der Testwerte in Bezug auf das Geschlecht für die Iteration (Version A) weisen ähnliche Mittelwerte auf (Diagramm 9). So liegt für die Gruppe der Mädchen ein Mittelwert von  $M = 5.1$  und für die Gruppe der Jungen ein durchschnittlicher Wert von  $M = 5.7$  vor. Die Standardabweichungen liegen nah beieinander. Innerhalb der Iteration (Version B) sind die Mittelwerte der Jungen- und der Mädchen-Gruppe ähnlich eng beisammen. Auch die Standardabweichungen unterscheiden sich lediglich um 0.1 Punkte. Die Betrachtung beider Levene-Tests zeigt, dass sowohl in Version A als auch in Version B eine Varianzhomogenität vorliegt. Die Mittelwertunterschiede sind nach Berechnungen des  $t$ -Tests in beiden Versionen (A und B) nicht signifikant.

## Diagramm 9

*ITERATION Häufigkeitsverteilung der Testwerte nach Geschlecht*



*Anmerkung.* Iteration A (links) Jungen:  $N = 31$ ;  $n = 31$ ;  $M = 5.7$ ;  $SD = 1.80$ . Iteration A (links) Mädchen:  $N = 30$ ;  $n = 30$ ;  $M = 5.1$   $SD = 2.04$ . Iteration B (rechts) Jungen:  $N = 29$ ;  $n = 29$ ;  $M = 5.6$ ;  $SD = 1.99$ . Iteration B (rechts) Mädchen:  $N = 29$ ;  $n = 29$ ;  $M = 5.2$ ;  $SD = 1.89$ .

### 4.3 Ergebnisraum

Codieranleitungen legen die Beziehung zwischen dem theoretischen Konstrukt der Sequenz, der Alternative bzw. der Iteration und dem Ergebnisraum dar. In diesem Kapitel sind die finalen Codieranleitungen basierend auf dem Itempool aus der Itemanalyse, den Abbildung 11 bis Abbildung 13 zu entnehmen. Das theoretische Konstrukt (links in den Abbildungen) beschreibt die Kognitive Dimension entwerfen mit den Niveaustufen (1) „reproduzieren“, (2) „anwenden“ und (3) „reflektieren“. Zusätzlich sind die entsprechenden Definitionen zur Veranschaulichung erneut aufgelistet (siehe auch KMgPS). Rechts in den Abbildungen wird für das jeweilige theoretische Konstrukt der Ergebnisraum definiert. Alle *richtigen* Items werden mit einem Punkt und alle *falschen* Items mit null Punkten gescort. In diesem Zusammenhang wird keine Unterscheidung zwischen den Niveaustufen der Items gemacht, sodass pro Item maximal ein Punkt erreicht werden kann. Die Items werden innerhalb der Testversion A und B betrachtet außer bei der Alternative. Hier konnten die Testversion A und B als Pseudo-Parallelform gewertet und somit im Gesamten betrachtet werden. Die Kategorisierung der Aufgaben erfolgt über die a priori vorgenommene Zuordnung der Items in die Stufen „reproduzieren“, „anwenden“ und „reflektieren“. Es werden nur die Items berücksichtigt, welche nach der

Itemsanalyse weiterhin im Itempool verblieben sind. Die Nummerierungen der Items sind den Testheften aus der Hauptstudie zu entnehmen.

Zu Beginn wird der Ergebnisraum der Sequenz betrachtet (Abbildung 11). Hier ergeben sich für die Testversionen A und B der Sequenz keine Items in der Niveaustufe 1. Vier Items (7, 8, 9 und 10) innerhalb der Testversion A und drei Items (7, 9 und 10) innerhalb der Testversion B sind der zweiten Niveaustufe zuzuordnen. Im Bereich des Reflektierens (Niveaustufe 3) befinden sich in beiden Testversionen die Items 11 bis 15.

### Abbildung 11

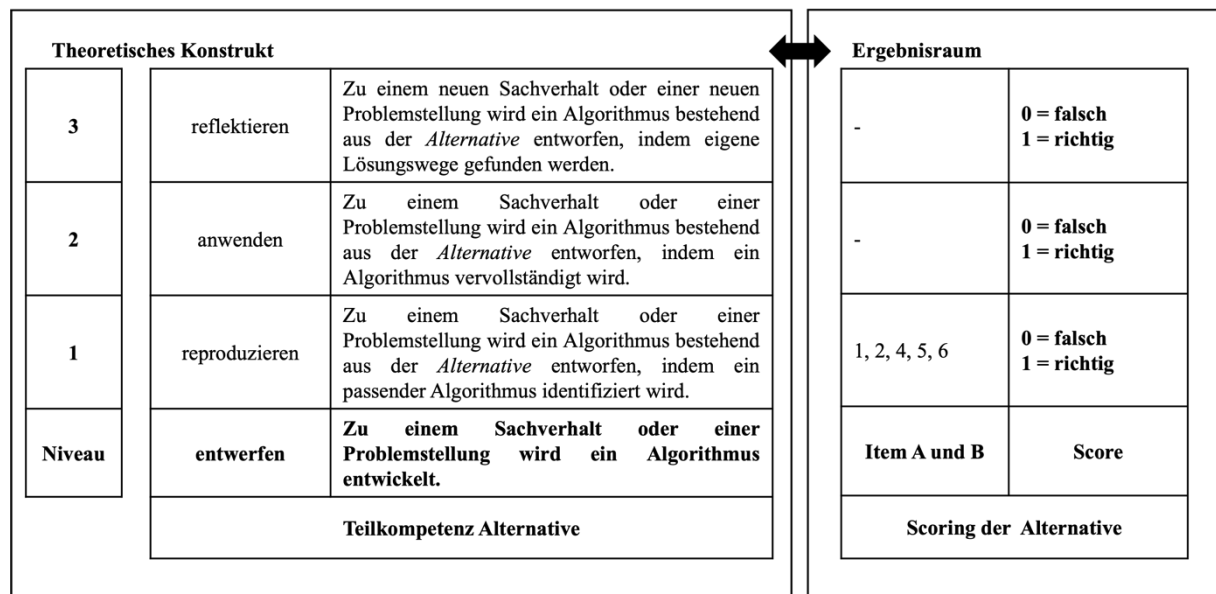
#### SEQUENZ Beziehung zwischen dem theoretischen Konstrukt und dem Ergebnisraum

Theoretisches Konstrukt			Ergebnisraum		
3	reflektieren	Zu einem neuen Sachverhalt oder einer neuen Problemstellung wird ein Algorithmus bestehend aus <i>Sequenzen</i> entworfen, indem eigene Lösungswege gefunden werden.	11, 12, 13, 14, 15	11, 12, 13, 14, 15	0 = falsch 1 = richtig
2	anwenden	Zu einem Sachverhalt oder einer Problemstellung wird ein Algorithmus bestehend aus <i>Sequenzen</i> entworfen, indem ein Algorithmus vervollständigt wird.	7, 8, 9, 10	7, 9, 10	0 = falsch 1 = richtig
1	reproduzieren	Zu einem Sachverhalt oder einer Problemstellung wird ein Algorithmus bestehend aus <i>Sequenzen</i> entworfen, indem ein passender Algorithmus identifiziert wird.	-	-	0 = falsch 1 = richtig
Niveau	entwerfen	<b>Zu einem Sachverhalt oder einer Problemstellung wird ein Algorithmus entwickelt.</b>	A	B	Score
			Item		
Teilkompetenz Sequenz			Scoring der Sequenz		

Anmerkung. Angelehnt an die Visualisierung von Eggert und Bögeholz, S. 378 (2014)

Der Ergebnisraum der Alternative (Abbildung 12) ergibt eine Zuordnung aller fünf Items (1, 2, 4, 5 und 6) in ein reproduzierendes Aufgabenniveau. Aufgaben aus dem Bereich des Anwendens oder Reflektierens sind innerhalb des Prozesses der Itemselektion weggefallen.

## Abbildung 12

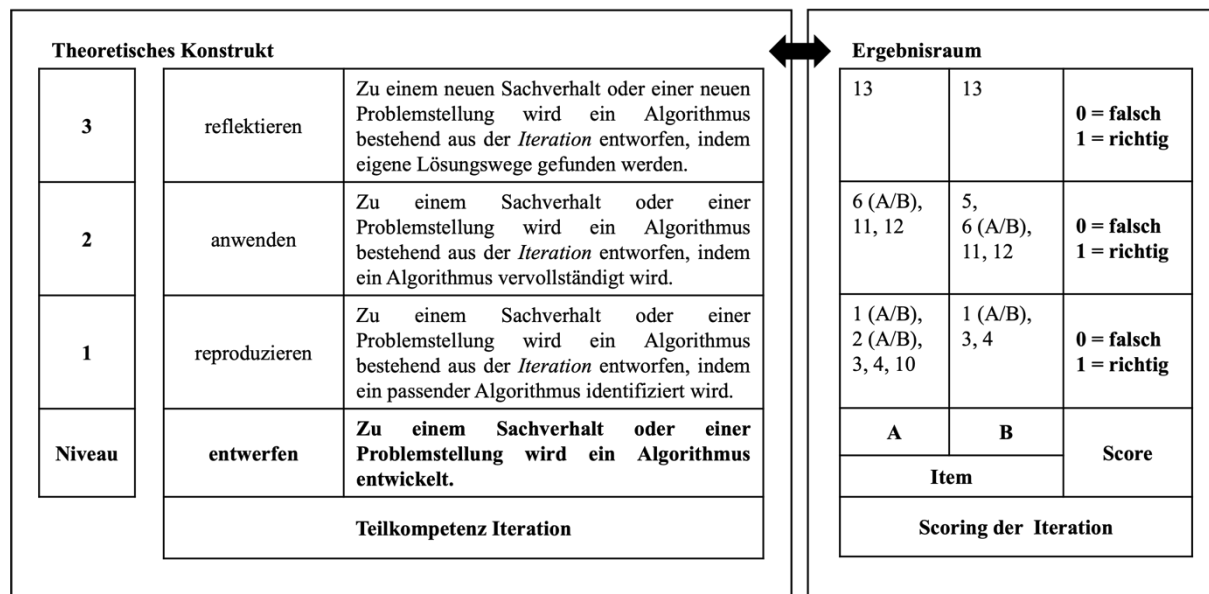
*ALTERNATIVE Beziehung zwischen dem theoretischen Konstrukt und dem Ergebnisraum*

*Anmerkung.* Angelehnt an die Visualisierung von Eggert und Bögeholz, S. 378 (2014)

Der Ergebnisraum der Iteration (Abbildung 13) wird separat für Testversion A und B betrachtet. Die Testversion A wird in der Niveaustufe 1 durch die fünf Items (1, 2, 3, 4 und 10) abgedeckt. Die Testversion B wird im Bereich des Reproduzierens durch die drei Items 1, 3 und 4 repräsentiert. Der zweiten Niveaustufe sind drei Items (6, 11 und 12) der Testversion A und vier Items (5, 6, 11 und 12) der Testversion B zuzuordnen. Die Niveaustufe des Reflektierens wird in beiden Testheften durch die Aufgabe 13 abgebildet.

Abbildung 13

*ITERATION Beziehung zwischen dem theoretischen Konstrukt und dem Ergebnisraum*



*Anmerkung.* Angelehnt an die Visualisierung von Eggert und Bögeholz, S. 378 (2014)

#### 4.4 Messmodell (Rasch-Analyse)

Die Durchführung der Rasch-Analyse wird basierend auf der Stichprobe aus der Hauptstudie durchgeführt. Zusammenfassend werden noch einmal wesentliche Daten wiederholt. Die Stichprobe wurde in verschiedenen Stadtteilen der Stadt Köln erhoben und lag bei  $n = 378$  Schülerinnen und Schülern. Innerhalb einer Klasse haben die Kinder immer ein Themenheft in der Version A oder B bearbeitet. Das Testheft der Sequenz (Version A) haben 70 Kinder, das Testheft der Sequenz (Version B) haben 64 Kinder, das Testheft der Alternative A haben 60 Kinder, das Testheft der Alternative B haben 65 Kinder, das Testheft der Iteration (Version A) haben 61 Kinder und das Testheft der Iteration (Version B) haben 58 Kinder bearbeitet.

Zur Analyse der Daten wird das Softwareprogramm MINISTEPS® genutzt. Richtige Antworten sind mit 1 und falsche Antworten sind mit 0 gescort. Es werden folgende Analyseschritte für jedes Testheft (Sequenz A, Sequenz B; Alternative A; Alternative B; Iteration A und Iteration B) durchgeführt und nachfolgend beschrieben: Die Überprüfung der Eindimensionalität, die Bewertung der Zielsetzung sowie die Boden- und Deckeneffekte und die Evaluation der Funktionsfähigkeit des Messinstruments und der Konstruktvalidität.

#### 4.4.1 Eindimensionalität

Die Überprüfungen der Testhefte zeigen, dass eine Eindimensionalität vorliegt. Die MNSQ *Infit*- und *Outfit* -Werte liegen bei allen Testheften in einem Bereich zwischen 0.5 und 1.5. Lediglich im Testheft der Sequenz (Version A) und dem Testheft der Iteration A mussten Items auf Grund eines zu hohen Wertes entfernt werden. Im Testheft der Sequenz handelt es sich um das Item 8 und im Testheft der Schleife (Version A) um das Item 6. Die Betrachtung der unerklärten Varianz mittels PCAR-Methode liegt bei zwei Testheften bei einem gewünschten Eigenwert von kleiner als 2.0 im ersten Kontrast. In diesem Kontext ist das Testheft der Sequenz B und der Schleife A zu nennen. Alle anderen Testhefte weisen eine minimale Abweichung des gewünschten Werts auf. Da es sich um minimale Abweichungen handelt, wird die Eindimensionalität durch die PCAR-Methode angenommen. Die *Point-Measure-Correlation* liegt bei allen Testheften über dem empfohlenen Mindestwert von 0.3, weshalb eine Eindimensionalität in allen Testheften angenommen werden kann.

Insgesamt ergibt sich mittels Überprüfung der drei Verfahren eine Eindimensionalität für alle Testhefte. Die Berechnungen sind dem Anhang zu entnehmen (E Berechnungen der Rasch-Analyse).

#### 4.4.2 Personen- und Itemreliabilität sowie Personen- und Itemseparation

Im Folgenden werden die Ergebnisse der Personen- und Itemreliabilität und der Personen- und Itemseparation zu den einzelnen Testheften dargelegt. Eine tabellarische Auflistung der Berechnungen sind Tabelle 9 zu entnehmen. Für die Sequenz (Version A) liegt die Personenreliabilität bei 0.66 und die Personenseparation bei 1.38. Die Itemreliabilität liegt bei 0.86 und die Itemseparation bei 2.40. Die Version B der Sequenz weist eine Personenreliabilität von 0.68 und eine Personenseparation von 1.39 auf. Die Itemreliabilität liegt bei 0.86 und die Itemseparation bei 2.43. Bei der Alternative Version A und B liegt eine Personenreliabilität und -separation von 0.00 vor. In Version A der Alternative weist die Itemreliabilität einen Wert von 0.55 und die Itemseparation einen Wert von 1.10 auf. Die Version B liefert für die Itemreliabilität einen Wert von 0.10 und für die Itemseparation einen Wert von 0.33. Für die Schleife (Version A) liegt eine Personenreliabilität von 0.50 und eine Personenseparation von 1.00 vor. Die Itemreliabilität weist einen Wert von 0.83 und die Itemseparation einen Wert von 2.23 auf. Die Version B der Schleife weist für die Personenreliabilität einen Wert von 0.55 und für die Personenseparation einen Wert von 1.11 auf. Die Itemreliabilität besitzt einen Wert von 0.91 und die Itemseparation einen Wert von 2.99. Insgesamt betrachtet sind für alle Testhefte die geforderten Werte der Personen -und Itemseparationen gegeben. Hinsichtlich der

Personenreliabilitäten sind starke Abweichungen vom gewünschten Wert (0.80) zu verzeichnen. Die Itemreliabilitäten liegen bei der Sequenz (Version A und B) und der Schleife (Version A) sehr nah am gewünschten Wert. Die Schleife (Version B) erfüllt als einziges Testheft den geforderten Personenreliabilitätswert mit 0.91. Die Alternative liefert in beiden Versionen sehr niedrige Reliabilitätswerte.

**Tabelle 9**

*Personen- und Itemreliabilität sowie Personen- und Itemseparation*

Testheft	Personen		Item	
	Reliabilität	Separation	Reliabilität	Separation
Sequenz A	0.66	1.38	0.86	2.40
Sequenz B	0.68	1.39	0.86	2.43
Alternative A	0.00	0.00	0.55	1.10
Alternative B	0.00	0.00	0.10	0.33
Iteration A	0.50	1.00	0.83	2.23
Iteration B	0.55	1.22	0.91	2.99

#### 4.4.3 Zielsetzung sowie Boden- und Deckeneffekte

Die Analyse der Zielsetzung in Bezug auf die Durchschnittspersonenfähigkeit und die Durchschnittsitemschwierigkeit ergibt bei allen Testheften einen Wert, welcher unter 1.0 Logits liegt. Somit liegt eine gute Zielsetzung zwischen Personenfähigkeit und Itemschwierigkeit vor. Die Betrachtung der Boden- und Deckeneffekte ergeben insgesamt sehr unterschiedliche Ergebnisse, weshalb eine Betrachtung nacheinander pro Testheft erfolgt. Die Sequenz (Version A) weist einen Deckeneffekt von 8.6 Prozent und einen Bodeneffekt von 12.9 Prozent bei einem maximalen Score von 8 Punkten auf. Der Deckeneffekt für die Sequenz (Version B) liegt bei 14.1 Prozent und der Bodeneffekt bei 6.3 Prozent bei einem maximalen Score von acht Punkten. In beiden Versionen wird die empfohlene Fünf-Prozent-Marke überschritten. Beide Testversionen der Alternative weisen sehr hohe Deckeneffekte (Version A: 68.3 % und Version B: 52.3 %) bei einem maximalen Score von fünf Punkten auf. Die Bodeneffekte liegen bei beiden Versionen im gewünschten Bereich (Version A: 5.0 % und Version B: 2.0 %). Innerhalb der Schleife (Version A) ist ein Deckeneffekt von 14.8 % zu verzeichnen. Die Bodeneffekte liegen bei beiden Testversionen bei 0 %. Die Version B der Schleife weist einen etwas geringeren Deckeneffekt als die Version A der Iteration auf (13.8%).

#### 4.4.4 Funktionsfähigkeit des Messinstruments und Konstruktvalidität

Bei der Evaluation der Funktionsfähigkeit des Messinstruments wird insbesondere die Person-Item-Map, auf welcher die Personenfähigkeit und die Itemschwierigkeit in einer Skala dargestellt werden, betrachtet.

Bei der Betrachtung der Person-Item-Map zur Sequenz sowohl bei Version A als auch Version B ist zu erkennen, dass gerade im oberen und unteren Schwierigkeitsbereich Items fehlen (Abbildung 14). Die Version A weist außerdem zwischen den Aufgaben 7 und 14 eine Lücke auf. Zusätzlich sind die Items 9 und 13 einer identischen Itemschwierigkeit zuzuordnen. Die Sequenz (Version B) weist insbesondere Lücken zwischen den Items 12 und 11 und 7 und 9 auf. Dopplung von Items sind in der Version B nicht vorhanden.

Die Alternative (Abbildung 15) Version A zeigt bei insgesamt 5 Items einen besonders großen Sprung von Item 1 zu Item 2 beziehungsweise Item 6 (Item 2 und 6 befinden sich auf einer Schwierigkeitsebene). Zusätzlich ist anzumerken, dass insbesondere im oberen Schwierigkeitsbereich keine Items vorhanden sind. Die Alternative Version B zeigt eine Dopplung der Items 2 und 4. Weiter fehlen schwierige Items, welche über Aufgabe 1 liegen. Auch ist eine große Lücke unterhalb des Items 5 bei leichten Itemschwierigkeiten zu verzeichnen. Zusätzlich gibt es eine Lücke zwischen den Items 5 und 6 und den Items 6 und 2.

Die Iteration (Version A) weist eine passable Verteilung der Items auf der Person-Item-Map auf, jedoch existieren großen Lücken zwischen einzelnen Items (Abbildung 16). In diesem Zusammenhang sind beispielsweise Lücken zwischen der Aufgabe 10 und 11 oder Aufgabe 11 und 4 zu nennen. Zusätzlich fehlen Items im niedrigen und hohe Itemschwierigkeitsbereich. Die Items 11 und 3 weisen eine gleiche Itemschwierigkeit auf. Die Items der Iteration (Version B) zeigen eine gute Verteilung im leichten Bereich. Jedoch sind Lücken zwischen Item 13 und Item 1 sowie zwischen Item 1 und 6 zu identifizieren. Ebenso fehlen im unteren und oberen Bereich Items zur Schwierigkeitskennzeichnung. Item 11 und 4 weisen eine gleiche Schwierigkeit auf.

Detaillierte Schlussfolgerungen aus den zuvor beschriebenen Ergebnissen der Rasch-Analyse werden im Unterkapitel Diskussion der Ergebnisse (5.1) ausgeführt.

## 4.5 Testvalidierung

Für die Berechnungen im Rahmen der Testvalidierung wird auf die Items aus der Itemanalyse zurückgegriffen und nicht wie üblich auf die Ergebnisse des Messinstruments aus der Rasch-Analyse. Dieses Vorgehen wird deshalb gewählt, weil die Ergebnisse aus der Rasch-Analyse erst nach der Validierungsstudie vorlagen.

Die Bestimmung der diskriminanten und konvergenten Validität wird mittels bivariater Korrelationsberechnungen in IBM® SPSS® durchgeführt. In diesem Kontext werden die erreichten Testwerte pro Subtest miteinander korreliert. Zwecks Einteilung der Korrelationskoeffizienten wird den Werten in der Literatur gefolgt. Korrelationswerte zwischen 0.40 und 0.60 sind als mittelmäßig und Werte über  $r \geq 0.60$  als hoch einzustufen (Bühner, 2011; Fisseni, 1990; Weise, 1975). Die Überprüfung aller Variablen (PRO SU 2, DEMAT 2+ und CFT) auf Normalverteilung mittels Shapiro-Wilk Test haben eine Signifikanz ( $p < 0.05$ ) ergeben, weshalb keine Normalverteilung angenommen werden kann (Field, 2018; Shapiro & Wilk, 1965). Ausnahmen bilden die Testhefte Sequenz (Version A und B) und das Testheft Ähnlichkeiten des CFT 1-R Tests. Diese Subtests weisen eine Normalverteilung auf. Da insgesamt nur wenige Subtests eine Normalverteilung aufweisen, werden die folgenden Korrelationen mittels Spearmans Rho berechnet (Bortz & Lienert, 2008; Field, 2018; Meyer-Bahlburg, 1969). Die Korrelationsberechnungen zwischen dem in der Dissertation entwickelten Test PRO SU 2 und dem DEMAT 2+ (Mathematiktest) bzw. CFT 1-R (Intelligenztest) werden angelehnt an die Multitrait-Multimethod Analyse dargestellt. Es gilt herauszufinden, inwiefern diskriminante und konvergente Validitäten vorliegen. Dafür werden die Korrelationsberechnungen innerhalb der schwarz umrandeten Kästen genauer analysiert und betrachtet (Tabelle 10 bis Tabelle 14). Hier gilt es über die Korrelationskoeffizienten hohe und niedrige Zusammenhänge herauszufinden. Dafür werden zuerst die Subtests zwischen dem PRO SU 2 (Testversion A und B) und DEMAT 2+ dargelegt (Kapitel 4.5.1). Danach werden die Korrelationsberechnungen zwischen den Subtests des PRO SU 2 und des CFT 1-R präsentiert (Kapitel 4.5.2). Abschließend werden die Heterotrait-Monomethod-Korrelationen (verschiedene Konstrukte mit derselben Methode) innerhalb der Subtests des PRO SU 2 betrachtet (4.5.3). Diese sind in den folgenden Tabellen als blaues Dreieck markiert. Mittlere Zusammenhänge werden als  $0.40 < r < 0.60$  und hohe Korrelation als  $r \geq 0.60$  definiert (Bortz & Döring, 2016; Bühner, 2011; Fisseni, 1990; Weise, 1975).

#### 4.5.1 PRO SU 2 und DEMAT 2+

In diesem Kapitel wird die Korrelation zwischen dem PRO SU 2 A mit dem DEMAT 2+ und anschließend zwischen dem PRO SU 2 B und dem DEMAT 2+ beschrieben. Danach werden die Korrelationsberechnungen von beiden Testversionen A und B des PRO SU 2 mit dem DEAMT 2+ in Beziehung gesetzt. Die Korrelationsberechnungen zwischen dem PRO SU 2 (Version A) und dem DEMAT 2+ werden in Tabelle 10 und zwischen dem PRO SU 2 (Version B) und dem DEMAT 2+ in Tabelle 11 dargestellt.

**Tabelle 10**

*Korrelationen zwischen dem PRO SU 2 (Version A) und dem DEMAT 2+*

	SE	AL	IT
SE	1		
AL	0,05	1	
IT	,593**	0,275	1
ZE	0,191	-0,118	0,025
LÄ	-0,01	-0,138	-0,234
AD	,449**	-0,254	0,342
SU	,483**	-0,086	,502**
DO	0,168	0,035	0,231
DI	,479**	0,11	,366*
HA	,525**	-0,006	0,295
GEL	0,263	-0,123	,428*
SA	0,281	-0,236	,449**
GEO	0,269	0,213	,530**

*Anmerkung.* SE Sequenz, AL Alternative, IT Iteration, ZE Zahleneigenschaften, LÄ Längenvergleich, AD Addition, SU Subtraktion, DO Verdoppeln, DI Division, HA Halbieren, GEL Rechnen mit Geld, SA Sachaufgaben, GEO Geometrie. \*\* Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant. \* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.

Zwischen der Sequenz und den Subtests der Addition, der Subtraktion, der Division und dem Subtest Halbieren liegen insgesamt mittelmäßige Zusammenhänge zwischen 0.45 und 0.53 vor. Zwischen der Sequenz und dem Subtest Halbieren (DEMAT 2+) liegt der höchste Korrelationswert mit  $r = 0.53$  vor. Am geringsten korrelieren die signifikanten Werte zwischen der Sequenz und der Addition mit  $r = 0.45$ . Die Betrachtung der Alternative zu den zehn Subtests des DEMAT 2+ ergeben keine signifikanten Korrelationswerte. Zwischen der Iteration und den Subtests Subtraktion, Division, Rechnen mit Geld, Sachaufgaben und Geometrie liegen

mittlere Korrelationskoeffizienten vor. Die niedrigste signifikante Korrelation besteht zwischen der Iteration und der Division, mit einem Wert von  $r = 0.37$ . Der höchste Korrelationswert liegt bei  $r = 0.53$  zwischen der Iteration und dem Subtest Geometrie. Nach der vorgeschlagenen Einteilung in der Literatur lassen sich insgesamt lediglich diskriminante Validitäten zwischen PRO SU 2 A und dem DEMAT 2+ identifizieren. Konvergente Validitäten mit Werten  $r > 0.60$  zwischen dem Programmieretest Sachunterricht (Testversion A) und den DEMAT 2+ sind nicht erkennbar. Es folgen die Auswertungen zu den Korrelationsberechnungen zwischen dem PRO SU 2 (Version B) und dem DEMAT 2+ (Tabelle 11).

**Tabelle 11**

*Korrelationen zwischen dem PRO SU 2 (Version B) und dem DEMAT 2+*

	SE	AL	IT
SE	1		
AL	0,083	1	
IT	,411*	0,181	1
ZE	0,182	0,219	,530**
LÄ	,361*	0,001	,695**
AD	,453*	0,069	,604**
SU	,428*	0,065	,589**
DO	0,291	0,051	,604**
DI	0,23	-0,134	,625**
HA	,414*	-0,183	,366*
GEL	,406*	-0,079	,570**
SA	,628**	0,152	,759**
GEO	-0,045	-0,241	0,031

*Anmerkung.* SE Sequenz, AL Alternative, IT Iteration, ZE Zahleneigenschaften, LÄ Längenvergleich, AD Addition, SU Subtraktion, DO Verdoppeln, DI Division, HA Halbieren, GEL Rechnen mit Geld, SA Sachaufgaben, GEO Geometrie. \*\* Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant. \* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.

Zwischen der Sequenz und den Subtests des DEMAT 2+ liegen mittelmäßige bis starke Korrelationswerte vor. Es bestehen Zusammenhänge zwischen der Sequenz und den Subtests Längenvergleich, Addition, Subtraktion, Halbieren, Rechnen mit Geld und Sachaufgaben. Der geringste signifikante Korrelationswert innerhalb der Sequenz liegt bei  $r = 0.36$  mit dem Subtest Längenvergleich. Der höchste Korrelationswert  $r = 0.63$  wird zwischen der Sequenz und dem Subtest Sachaufgaben erreicht. Zwischen der Alternative und dem DEAMT 2+ Test sind keine signifikanten Korrelationskoeffizienten zu identifizieren. Die Iteration weist mittlere

Korrelationen zu den Subtests Zahleneigenschaften ( $r = 0.53$ ), Addition ( $r = 0.60$ ), Subtraktion ( $r = 0.59$ ), Verdoppeln ( $r = 0.60$ ), Halbieren ( $r = 0.37$ ) und Rechnen mit Geld ( $r = 0.57$ ) auf. Ein starker Zusammenhang liegt zwischen den Subtests der Iteration und den Subtests Längenvergleich ( $r = 0.70$ ), Division ( $r = 0.63$ ) und Sachaufgaben ( $r = 0.76$ ) vor. Auffällig ist, dass bis auf den Subtest der Geometrie, mittlere bis starke Korrelationen zwischen der Iteration und den genannten Subtests des DEMAT 2+ existieren. Insgesamt ergeben sich konvergente Validitäten zwischen der Iteration und den Subtests Längenvergleich, Division und Sachaufgaben sowie zwischen der Sequenz und dem Subtest Sachaufgaben. Alle weiteren Korrelationsberechnungen ergeben laut der in der Literatur vorgeschlagenen Einteilung diskriminante Validitäten. Abschließend folgt ein Vergleich der beiden Testversionen A und B des PRO SU 2 mit dem DEAMT 2+ (Tabelle 12).

**Tabelle 12**

*Vergleichende Korrelationen zwischen dem PRO SU 2 (Version A) und PRO SU 2 (Version B) mit dem DEMAT 2+*

	SE		AL		IT	
	A	B	A	B	A	B
ZE	0,191	0,182	-0,118	0,219	0,025	,530**
LÄ	-0,01	,361*	-0,138	0,001	-0,234	,695**
AD	,449**	,453*	-0,254	0,069	0,342	,604**
SU	,483**	,428*	-0,086	0,065	,502**	,589**
DO	0,168	0,291	0,035	0,051	0,231	,604**
DI	,479**	0,23	0,11	-0,134	,366*	,625**
HA	,525**	,414*	-0,006	-0,183	0,295	,366*
GEL	0,263	,406*	-0,123	-0,079	,428*	,570**
SA	0,281	,628**	-0,236	0,152	,449**	,759**
GEO	0,269	-0,045	0,213	-0,241	,530**	0,031

*Anmerkung.* SE Sequenz, AL Alternative, IT Iteration, ZE Zahleneigenschaften, LÄ Längenvergleich, AD Addition, SU Subtraktion, DO Verdoppeln, DI Division, HA Halbieren, GEL Rechnen mit Geld, SA Sachaufgaben, GEO Geometrie. \*\* Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant. \* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.

Beide Testversionen A und B der Sequenz weisen mittlere Zusammenhänge zu den Subtests Addition, Subtraktion und Halbieren auf. Weiter ist auffällig, dass die Korrelationswerte zwischen Testversion A und B starke Diskrepanzen aufweisen, wie die Subtests

Längenvergleiche ( $r_a^{11}=-0.01$ ;  $r_b^{12}=0.36$ ) oder Sachaufgaben ( $r_a=0.28$ ;  $r_b=0.63$ ). Bei Betrachtung der Alternative sind in beiden Testversionen A und B keine signifikanten Korrelationen zu den Subtests des DEMAT 2+ vorhanden. Auch liegen die Korrelationswerte eng beieinander. Als letztes werden die Korrelationskoeffizienten der Iteration fokussiert. Zum einen weisen beide Testversionen mittlere bis starke Korrelationen zu den Subtests Subtraktion, Division, Rechnen mit Geld und Sachaufgaben auf. Zum anderen sind jeweils in der Testversion A oder Testversion B alleinige Zusammenhänge zu Subtests des DEMATs 2+ vorhanden. Auffällig sind insbesondere starke Korrelationsabweichungen zwischen der Iteration Version A und Version B zum DEMAT 2+. So zeigt der Vergleich zwischen den Korrelationskoeffizienten Zahleneigenschaften A ( $r_a=0.03$ ) zu Zahleneigenschaften B ( $r_b=0.53$ ), Längenvergleich A ( $r_a=-0.23$ ) zu Längenvergleich B ( $r_b=0.70$ ), Addition A ( $r_a=0.34$ ) zu Addition B ( $r_b=0.60$ ), Verdoppeln A ( $r_a=0.23$ ) zu Verdoppeln B ( $r_b=0.60$ ) oder Geometrie A ( $r_a=.53$ ) zu Geometrie B ( $r_b=0.03$ ) starke Abweichungen. Zusätzlich korreliert die Testversion B mit allen Subtests des DEMATs 2+, außer mit dem Subtest der Geometrie. Die Testversion A weist insgesamt weniger starke Zusammenhänge auf.

#### 4.5.2 PRO SU 2 und CFT 1-R

In diesem Kapitel werden zuerst die Korrelationen zwischen dem PRO SU 2 (Version A) mit dem CFT 1-R dargelegt. Folgend werden detailliert die Zusammenhänge zwischen dem PRO SU (Version B) und dem CFT 1-R in Bezug auf ihre Subtests beschrieben. Abschließend werden die Korrelationsberechnungen von beiden Testversionen A und B des PRO SU 2 mit dem CFT 1-R in Beziehung gesetzt. Es folgt die Auswertung der Korrelationsberechnungen für den PRO SU 2 A mit dem Intelligenztest CFT 1-R. Eine detaillierte Aufschlüsselung der Korrelationskoeffizienten ist Tabelle 13 zu entnehmen.

---

<sup>11</sup>  $r_a$ =Korrelationswert der Testversion A

<sup>12</sup>  $r_b$ = Korrelationswert der Testversion B

**Tabelle 13**

*Korrelationen zwischen dem PRO SU 2 (Version A) und dem CFT R-1*

	SE	AL	IT
SE	1		
AL	0,05	1	
IT	593**	0,275	1
SUBST	0,2	0,169	0,135
LABYR	0,157	-0,002	0,043
ÄHNL	0,261	0,064	-0,04
REIH	0,278	-0,127	0,203
KLASS	,367*	0,153	0,293
MATR	0,198	0,197	0,157

*Anmerkung.* SE Sequenz, AL Alternative, IT Iteration, SUBST Substitutionen, LABYR Labyrinth, ÄHNL Ähnlichkeiten, REIH Reihenfortsetzen, KLASS Klassifikationen, MATR Matrizen. \*\* Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant. \* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.

In diesem Zusammenhang weist lediglich die Sequenz einen Zusammenhang zu dem Subtest Klassifikationen ( $r = 0.37$ ) auf. Alle anderen Korrelationsberechnungen zwischen dem Programmieretest im Sachunterricht (Version A) und dem Intelligenztest (CFT R-1) sind als gering einzustufen bzw. es liegen keine nennenswerten Korrelationen vor. Resümierend liegen keine erwähnenswerten konvergenten Validierungen zwischen dem PRO SU 2 A und dem CFT 1-R vor. Es folgt die Darlegung der Ergebnisse des PRO SU 2 (Version B) mit CFT 1-R. Eine Übersicht der errechneten Korrelationen sind der nachstehenden Tabelle zu entnehmen.

**Tabelle 14**

*Korrelationen zwischen dem PRO SU 2 (Version B) und dem CFT R-1*

	SE	AL	IT
SE	1		
AL	0,083	1	
IT	,411*	0,181	1
SUBST	-0,198	-0,243	-0,295
LABYR	,417*	0,1	0,221
ÄHNL	0,201	0,028	0,253
REIH	,539**	0,172	0,302
KLASS	,454*	0,07	,393*
MATR	,501**	-0,015	,396*

*Anmerkung.* SE Sequenz, AL Alternative, IT Iteration, SUBST Substitutionen, LABYR Labyrinth, ÄHNL Ähnlichkeiten, REIH Reihenfortsetzen, KLASS Klassifikationen, MATR Matrizen. \*\* Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant. \* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.

Die Auswertung der Korrelationsberechnungen des Programmiertests im Sachunterricht (Version B) mit dem Intelligenztest (CFT 1-R) ergeben zwischen der Sequenz (Version B) und den Subtests Labyrinth ( $r = 0.41$ ), Reihenfortsetzen ( $r = 0.54$ ), Klassifikation ( $r = 0.45$ ) und Matrizen ( $r = 0.50$ ) mittlere Zusammenhänge. Zwischen der Alternative und den Subtest des CFTs 1-R sind nur geringe bis keine Zusammenhänge zu identifizieren. Die Iteration weist zu den Subtests der Klassifikationen und den Matrizen einen mittleren Zusammenhang von  $r = 0.39$  und  $r = 0.40$  auf. Alle anderen Korrelationsberechnungen sind als gering einzustufen. Unter Betrachtung der literarischen Interpretationen der Korrelationskoeffizienten kann von keiner nennenswerten konvergenten Validität in diesem Zusammenhang ausgegangen werden, da alle Werte unter  $r = 0.60$  liegen.

Abschließend folgt ein Vergleich zwischen der Testversion A und B des PRO SU 2 mit dem CFT 1-R. Eine Gegenüberstellung der Korrelationsberechnungen ist in der anschließenden Tabelle visualisiert.

**Tabelle 15**

*Vergleichende Korrelationen zwischen dem PRO SU 2 (Version A) und PRO SU (Version B) mit dem CFT 1-R*

	SE		AL		IT	
	A	B	A	B	A	B
SUBST	0,2	-0,198	0,169	-0,243	0,135	-0,295
LABYR	0,157	,417*	-0,002	0,1	0,043	0,221
ÄHNL	0,261	0,201	0,064	0,028	-0,04	0,253
REIH	0,278	,539**	-0,127	0,172	0,203	0,302
KLASS	,367*	,454*	0,153	0,07	0,293	,393*
MATR	0,198	,501**	0,197	-0,015	0,157	,396*

*Anmerkung.* SE Sequenz, AL Alternative, IT Iteration, SUBST Substitutionen, LABYR Labyrinth, ÄHNL Ähnlichkeiten, REIH Reihenfortsetzen, KLASS Klassifikationen, MATR Matrizen. \*\* Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant. \* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.

Auffällig in diesem Kontext ist, dass die Sequenz (Version B) zu vier Subtests des Intelligenztests mittlere Korrelationen aufweist: Labyrinth, Reihenfortsetzen, Klassifikationen und Matrizen. Die A-Version der Sequenz hingegen erreicht lediglich mit einem Subtest (Klassifikationen) eine mittlere Korrelation. Zwischen der Alternative in Bezug auf die Subtests des Intelligenztests liegen bei beiden Testversionen A und B geringe und gleichwertige Berechnungen vor. Weiter ist auffällig, dass die Iteration (Version B) zu dem Subtest der Klassifikationen und den Matrizen eine mittlere Korrelation erzielt, hingegen die Version A keine signifikanten Werte erreicht. Ein weiteres interessantes Wertepaar bildet die Sequenz (A und B) und der Subtest der Matrizen. Auch hier liegen die errechneten Korrelationen mit einem Unterschiedswert von 0.30 relativ weit auseinander. Insgesamt betrachtet liegen keine nennenswerten Korrelationen zwischen dem Programmieretest im Sachunterricht und dem Intelligenztest im Sinne der konvergenten Validität vor.

### 4.5.3 PRO SU 2

Im Folgenden werden die Korrelationsberechnungen für die Subtests der Sequenz, Alternative und Iteration dargelegt. Es wird die Heterotrait-Monometrie-Matrix ausschließlich für den PRO SU 2 betrachtet. Es handelt sich hierbei um Korrelationswerte, welche unterschiedliche Konstrukte mit einer Methode berechnen. Diese sind in der Abbildung 10 im oberen blauen Dreieck dargestellt. Die Korrelationswerte sowohl im Testheft der Iteration Version A ( $r = 0.18$ ) als auch im Testheft der Version B ( $r = 0.28$ ) zu der Alternative bestätigen keine

Zusammenhänge. Die Korrelationsberechnungen zwischen dem Testheft der Sequenz und der Alternative sind in beiden Versionen ( $r_a = 0.05$ ;  $r_b = 0.08$ ) ebenfalls gering. Somit liegt in allen vier Fällen eine diskriminante Validität vor. Die entsprechenden Berechnungen sind Tabelle 10 und Tabelle 11 zu entnehmen. Die Betrachtung der Korrelationswerte zwischen der Sequenz (Version A) und der Iteration (Version A) ergeben einen Zusammenhang von  $r = 0.59$ . Der Korrelationskoeffizient der Sequenz (Version B) und der Iteration (Version B) liegen mit einem Wert von  $r = 0.41$  etwas darunter. Beide Werte sind als mittlere Zusammenhänge zu werten. Eine mögliche Erklärung für die errechneten Zusammenhänge könnte auf die ähnlichen Aufgabentypen zurückzuführen sein.

## 5 Diskussion der Ergebnisse

In diesem Kapitel werden nacheinander die Forschungsfragen basierend auf den zuvor beschriebenen Ergebnissen interpretiert und diskutiert. Zu Beginn wird die Forschungsfrage 1 beantwortet (5.1). Dies beinhaltet eine Zusammenfassung wesentlicher Ergebnisse in Bezug auf die Zielsetzung der Arbeit. Danach wird nacheinander auf die Forschungsfragen FF2 (5.2) bis FF7 (5.7) und die damit verbundenen Hypothesen eingegangen.

### 5.1 Grundlegende Programmierkompetenz im Sachunterricht

**FF 1:** Lässt sich die *grundlegende Programmierkompetenz im Sachunterricht* von Kindern der zweiten Klasse objektiv, reliabel und valide messen?

**Objektivität:** Die Durchführungsobjektivität des Tests ist gewährleistet, da eine Testinstruktion vorliegt, welche von der Testleitung vorgelesen wird. Die Auswertungsobjektivität ist durch die entwickelten Auswertungsschablonen gegeben. Ein Beispiel zur Erfüllung der Interpretationsobjektivität kann nicht präsentiert werden, da zuerst eine abschließende Endversion des Programmiertests benötigt wird. Diese gilt es in zukünftigen Studien zu entwickeln.

**Reliabilität:** Die Reliabilitätsberechnungen nach Cronbachs Alpha liefern mit Werten von  $\alpha \geq 0.77$  gute Reliabilitäten für die Sequenz (Version A und B) und für die Alternative (Version A und B). Lediglich für die Themenhefte der Iteration liegen geringere Korrelationskoeffizienten ( $\alpha = 0.62$  und  $\alpha = 0.70$ ) vor.

**Validität:** Die Korrelationsrechnungen zwischen den Subtests PRO SU 2 (Version A) und dem DEMAT 2+ ergeben nach der vorgeschlagenen Einteilung der Literatur lediglich diskriminante Validitäten. Bei dem Programmieretest Version B und dem DEMAT 2+ ergeben sich insgesamt konvergente Validitäten zwischen der Iteration und den Subtests Längenvergleich, Division und Sachaufgaben sowie zwischen der Sequenz und dem Subtest Sachaufgaben. Alle weiteren Korrelationsberechnungen ergeben diskriminante Validitäten. Die Korrelationsberechnungen zwischen der Intelligenz und Programmierfähigkeiten liefern keine nennenswerte Ergebnisse im Sinne der konvergenten Validität. Eine detaillierte Erläuterung von möglichen Konstruktüberschneidungen zu mathematischen Fähigkeiten oder der Intelligenz werden in den

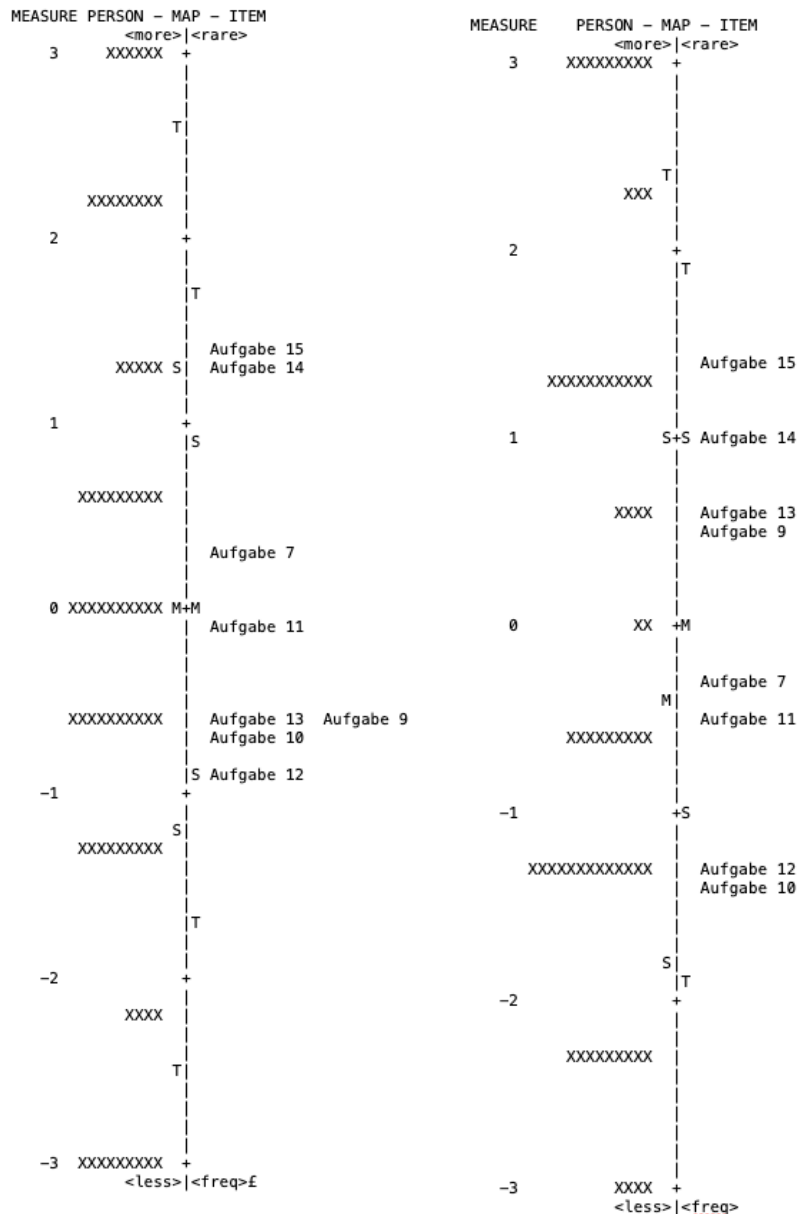
Forschungsfragen 6 und 7 weiter ausdifferenziert. Insgesamt lässt sich sagen, dass Aussagen über konvergente Validitäten hinsichtlich der Konstrukts mittels DEMAT 2+ und CFT 1-R nicht abschließend getätigt werden können.

Die Überprüfung der Konstruktvalidität in Bezug auf die einzelnen Testhefte erfolgt basierend auf den Ergebnissen der Rasch-Analyse. In diesem Kontext sind die Person-Item-Map und die damit verbundenen errechneten Kennzahlen zentral. Dies bietet eine detaillierte Betrachtung des Messinstruments und gibt Auskunft darüber, wie gut das Messinstrument funktioniert. Es folgt eine Interpretation der Testhefte Sequenz (A und B), Alternative (A und B) und Iteration (A und B).

**Sequenz:** Die Betrachtung der Person-Item-Map zum Testheft der Sequenz A (Abbildung 14) zeigt, dass eine bessere Differenzierung der Personenfähigkeit erfolgen kann, wenn weitere Items im oberen und unteren Schwierigkeitsbereich ergänzt werden. Durch diese Maßnahme kann dem Decken- und Bodeneffekt entgegengewirkt werden. Zusätzlich ist es zu empfehlen Items zu konzipieren, die zwischen Item 7 und 14 einzuordnen sind, um die Lücke zu schließen. Da es sich bei den Aufgaben 9 und 13 um gleiche Aufgabenschwierigkeiten handelt und Aufgabe 13 einem neuen Aufgabentyp entspricht, kann Aufgabe 13 bei weiteren Testungen herausgenommen werden. Aus den Ergebnissen der Person-Item-Map zu der Sequenz (Version B) lässt sich eine bessere Differenzierung der Personenfähigkeit ableiten, wenn weitere Items im oberen und unteren Schwierigkeitsbereich hinzugefügt werden. In diesem Zusammenhang kann dem Boden- und Deckeneffekt begegnet werden. Weiter werden Items benötigt, welche zwischen Item 12 und 11 sowie zwischen 7 und 9 einzuordnen sind. Eine Ergänzung von weiteren Items ermöglicht abschließend eine Verbesserung der berechneten Personen- und Itemreliabilitäten für die Sequenz (Version A und B).

**Abbildung 14**

*Person-Item-Map zur Sequenz*



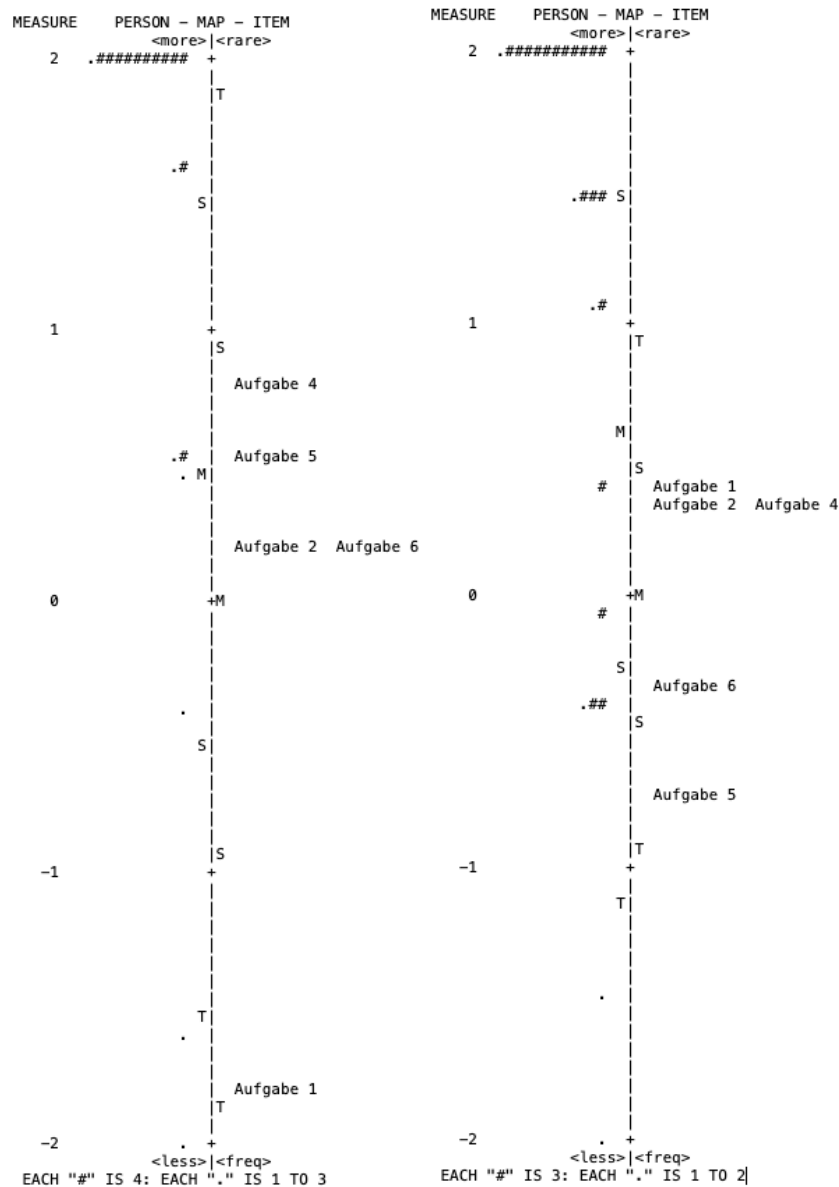
*Anmerkung.* Grafische Darstellung mittels MINISTEPS®. Sequenz A = links. Sequenz B = rechts. *M* = Mittelwert. *S* = 1 SD vom Mittelwert. *T* = 2 SD vom Mittelwert.

**Alternative:** Die Person-Item-Map der Alternative (Version A und B) weisen überwiegend einfache Aufgaben auf (Abbildung 15). Zur Differenzierung der Personenfähigkeit sind insbesondere zusätzliche schwierige Items sowie Items zwischen den großen Lücken notwendig. So können weitere Items bei der Alternative (Version A) zwischen Item 1 und 2 und bei der Alternative (Version B) zwischen Item 2 und 6 sowie zwischen Item 6 und 5 ergänzt

werden. Dadurch kann den hohen Deckeneffekten (Version A: 68.3 Prozent und Version B: 52.3 Prozent) entgegengewirkt und das Persönlichkeitsmerkmal der Alternative zuverlässiger gemessen werden. Die zuvor genannten Aspekte bestätigen zusätzlich die geringen Werte der Personen- und Itemreliabilität der Alternative (A und B).

**Abbildung 15**

*Person-Item-Map zur Alternative*



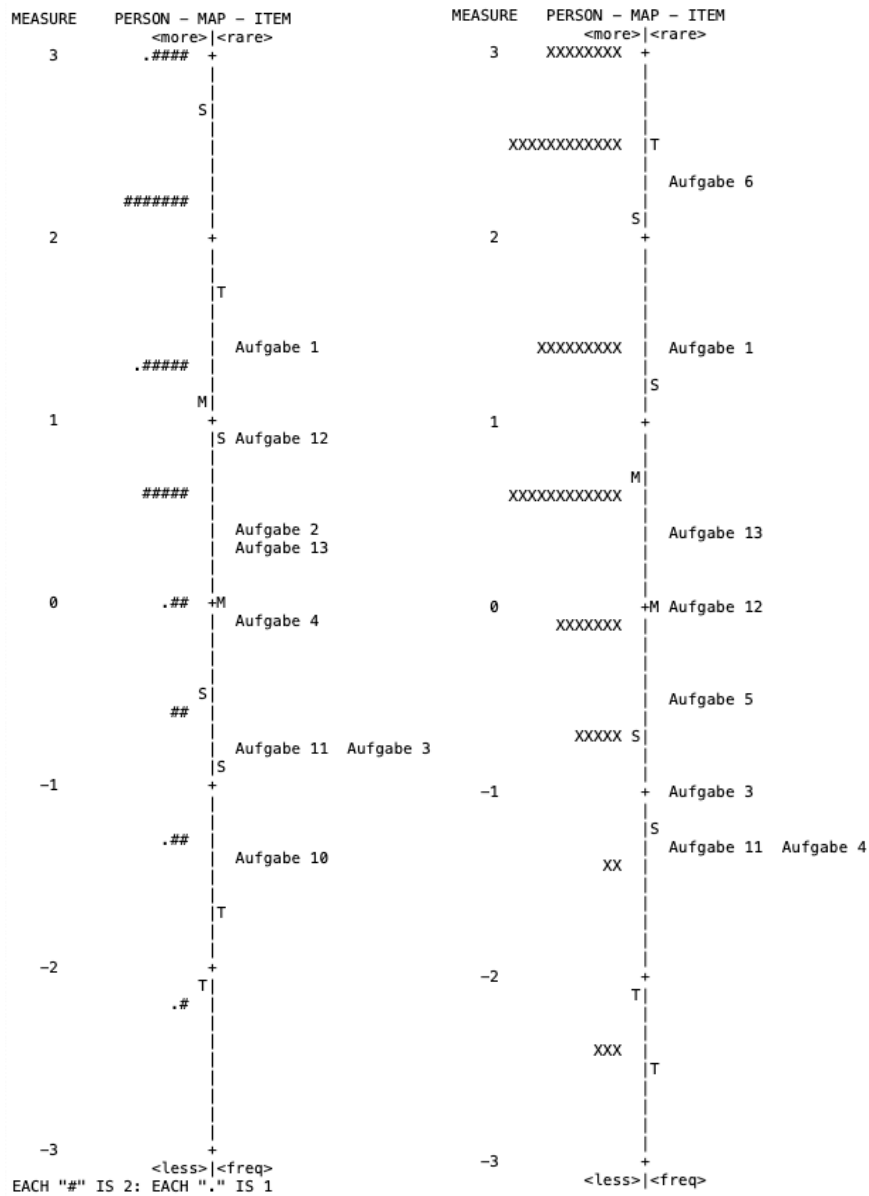
*Anmerkung.* Grafische Darstellung mittels MINISTEPS®. Alternative A = links. Alternative B = rechts. M = Mittelwert. S = 1 SD vom Mittelwert. T = 2 SD vom Mittelwert.

**Iteration:** Die Person-Item-Map zur Iteration (Version A) zeigt, dass weitere Differenzierungsmöglichkeiten zwischen den Personenfähigkeiten wünschenswert sind (Abbildung 16). Denn zwischen den Items existieren große Lücken, welche es zu füllen gilt. Zusätzlich sind weitere schwierige Items oberhalb der Aufgabe 1 zu empfehlen, um die Personenfähigkeit besser abbilden zu können und dem Deckeneffekt (14.8%)

entgegenzuwirken. Da es sich bei Aufgabe 11 um eine Aufgabe der Zählschleife handelt und bei Aufgabe 3 um eine Aufgabe der Bedingten Schleife, werden beide Aufgaben trotz gleicher Schwierigkeit beibehalten. Die Person-Item-Map der Iteration (Version B) benötigt zur besseren Differenzierung der Personenfähigkeit im oberen Bereich oberhalb des Items 6 weitere Items. Denn dort sind einige Fähigkeiten von Personen zu verzeichnen, welche eine weitere Differenzierung benötigen. Diese zeigen ebenfalls die Berechnung des Deckeneffekts. Zusätzlich sind Aufgaben zwischen Item 13 und 1 zu empfehlen. Da es sich bei Aufgabe 11 um eine Aufgabe der Zählschleife handelt und bei Aufgabe 4 um eine Aufgabe der Bedingten Schleife, werden beide Aufgaben trotz gleicher Schwierigkeit beibehalten.

**Abbildung 16**

*Person-Item-Map zur Iteration*



*Anmerkung.* Grafische Darstellung mittels MINISTEPS®. Iteration A = links. Iteration B = rechts. *M* = Mittelwert. *S* = 1 SD vom Mittelwert. *T* = 2 SD vom Mittelwert.

Zusammenfassend ist festzuhalten, dass insgesamt weitere Items benötigt werden, um das Konstrukt angemessen zu repräsentieren. Zwar können die Objektivität, die Reliabilität und die gemessenen diskriminanten Validitäten bestätigt werden. Jedoch konnte insbesondere durch die Überprüfung des Messinstruments mittels der Rasch-Analyse gezeigt werden, dass die Konstruktvalidität verbessert werden muss. Hierfür werden zu allen drei Themen (Sequenz,

Alternative und Iteration) weitere Items benötigt, die in einer weiteren Studie mittels großer Stichprobe überprüft werden sollten. Dafür bietet es sich an, die Items der Version A und B in einem Testheft zusammenzufassen und erst zum Schluss aus der Testendversion eine echte Parallelfarm zu entwickeln.

**Die Forschungsfrage, ob ein objektives, reliables und valides Messinstrument entwickelt werden konnte, lässt sich abschließend so beantworten, dass es sich bei dem entwickelten Messinstrument um ein wesentliches Fundament handelt, welches es weiter auszudifferenzieren gilt.**

## 5.2 Schwierigkeitserzeugende Kriterien

**FF2:** Lassen sich allgemeine schwierigkeitserzeugende Merkmale für Items in Bezug auf die Personenfähigkeit ableiten?

Lassen sich gruppenspezifische Merkmale für Items wiederfinden, welche den a priori postulierten Kompetenzformulierungen des KMgPS entsprechen?

Zur Beantwortung der FF2 werden die Items berücksichtigt, welche nach der Itemanalyse (3.3.2) im Itempool verblieben sind. Die Interpretation und Diskussion der Forschungsfrage werden basierend auf den Person-Item-Maps der Sequenz, Alternative und Iteration vorgenommen (siehe hierzu Abbildung 14 bis Abbildung 16). In diesem Zusammenhang werden für jeden Inhalt nacheinander schwierigkeitserzeugende Merkmale identifiziert. Zum Schluss wird auf den zweiten Teil der Forschungsfrage eingegangen.

**Sequenz:** Basierend auf der Person-Item-Map und einer Analyse der Aufgaben werden schwierigkeitserzeugende Merkmale für die Sequenz definiert. In diesem Kontext lassen sich vier Merkmale ableiten: Die Anzahl der unterschiedlichen Befehlsauswahl, die Anzahl der gesuchten Befehle für die Lösung, die Anzahl der vorkommenden Befehlswechsel und die Anzahl von möglichen Lösungen (Tabelle 16). Die Analyse hinsichtlich der Auswahl an unterschiedlichen Befehlen lässt den Rückschluss zu, dass mit einer Zunahme von Befehlen die Aufgabe schwieriger zu lösen ist. Die Betrachtung der gesuchten Befehlslänge zeigt, dass die Befehlslänge mit Komplexität der Aufgabe zunimmt. Ausnahmen bilden die Aufgabe 13, der Sequenz Version A und die Aufgabe 7 der Sequenz B. Bei der Betrachtung der Aufgabe 13 im

Vergleich zu Aufgabe 11 aus dem Testheft Version A könnte die Vorgabe des Weges eine Erklärung liefern, warum Aufgabe 11 schwieriger zu lösen ist als Aufgabe 13, obwohl für Aufgabe 13 mehr Befehle aneinandergereiht werden müssen. So ist in Aufgabe 13 (und auch Aufgabe 9) der Weg vorgegeben, wohingegen er in Aufgabe 11 selbstständig gefunden werden muss. In Bezug auf den Ausreißer Aufgabe 7 in der Version B ist in diesem Zusammenhang die Laufrichtung der Figur zu nennen. Diese könnte eine Erklärung dafür sein, dass Aufgabe 9 deutlich schwieriger im Vergleich zu Aufgabe 7 ist, da Aufgabe 9 mehr Befehle entgegen der Leserichtung enthält. Weiter fällt auf, dass die Befehlswechsel mit zunehmender Itemschwierigkeit innerhalb der Version B zunehmen. In Bezug auf das Testheft der Version A bildet die Aufgabe 11 eine Ausnahme. Die Anzahl von möglichen Lösungen erscheint insgesamt betrachtet als kein schwierigkeiterzeugendes Merkmal. Wenn hingegen Items in Gruppen zusammengefasst werden, könnte die Anzahl der Lösungsmöglichkeiten innerhalb der Gruppe eine entscheidende Rolle spielen.

**Tabelle 16***SEQUENZ schwierigkeiterzeugende Merkmale*

Aufgabe	Anzahl der Befehlsauswahl	Anzahl der gesuchten Befehle	Befehlswechsel	Anzahl von möglichen Lösungen
15	5	mind. 8	mind. 5 oder 6	4
14	5	mind. 8	mind. 5 oder 6	3
7	4	5	4	1
11	4	4	1 - 3	6
13 / 9	4 / 4	8 / 4	4 / 4	1 / 1
10	4	3	1 - 2	3
12	4	3	1 - 2	3

Aufgabe	Anzahl der Befehlsauswahl	Anzahl der gesuchten Befehle	Befehlswechsel	Anzahl von möglichen Lösungen
15	5	mind. 8	mind. 5 oder 6	4
14	5	mind. 8	mind. 5 oder 6	3
13	4	8	4	1
9	4	4	4	1
7	4	5	4	1
11	4	4	1 - 3	6
12	4	3	1 - 2	3
10	4	3	1 - 2	3

*Anmerkung.* oben = Sequenz A. unten = Sequenz B. / = Trennung von Aufgaben, welche eine identische Itemschwierigkeit aufweisen. Die Aufgaben sind nach Schwierigkeit von oben (schwierig) nach unten (leicht) sortiert.

**Alternative:** Basierend auf den Person-Item-Maps der Alternative (Version A und B) und einer Analyse der Aufgaben lassen sich vier Merkmale herausstellen: Die Bedingung ist erfüllt, die Bedingung ist nicht erfüllt, es existieren Ecken innerhalb des Weges (längere Wegstrecke) und die Anzahl von möglichen Wegen (Tabelle 17). Mit Blick auf die Ergebnisse lassen sich keine einheitlichen Muster oder Indizien erkennen, die auf eine Komplexität hinweisen. Ob eine Bedingung erfüllt ist oder nicht, erscheint nicht ausschlaggebend für die Schwierigkeit einer Aufgabe zu sein. Es ist anzumerken, dass alle Aufgaben einer leichten Itemschwierigkeit entsprechen und nur unzureichend die Personenfähigkeiten voneinander trennen. Eine Möglichkeit, die Aufgaben schwieriger zu machen, könnte sein, die Anzahl von möglichen Wegen oder die Anzahl von aneinandergereihten Bedingungen zu erhöhen. In diesem Zusammenhang sind weitere Items notwendig, um schwierigkeiterzeugende Merkmale zu identifizieren.

**Tabelle 17**

*ALTERNATIVE schwierigkeiterzeugende Merkmale*

Aufgabe	Bedingung erfüllt	Verneinung erfüllt	Ecken innerhalb des Weges	Anzahl von möglichen Wegen
4	ja	-	8	3
5	-	ja	3	3
2 / 6	- / ja	ja / -	0 / 4	3 / 3
1	ja	4	0	3

Aufgabe	Bedingung erfüllt	Verneinung erfüllt	Ecken innerhalb des Weges	Anzahl von möglichen Wegen
1	ja	-	0	3
2 / 4	- / ja	ja / -	1 / 8	3 / 3
6	ja	-	4	3
5	-	ja	3	3

*Anmerkung.* oben = Alternative A. unten = Alternative B. / = Trennung von Aufgaben, welche eine identische Itemschwierigkeit aufweisen. Die Aufgaben sind nach Schwierigkeit von oben (schwierig) nach unten (leicht) sortiert. - = trifft nicht zu.

**Iteration:** Basierend auf den Person-Item-Maps der Iteration und der Analyse der Aufgaben lassen sich vier Kategorien bilden: Die Anzahl der Befehlsauswahl, die Anzahl der gesuchten Befehle, die Anzahl der Wiederholung der Schleife und die Anzahl von möglichen Lösungen (Tabelle 18). Die Betrachtung der Bedingten Schleife (Version A) zeigt, dass ein Abnehmen der Wiederholung zu einer erhöhten Schwierigkeit führt. Version B der Iteration weist eine gegenläufiges Muster auf. Je mehr Wiederholungen durchgeführt werden desto schwieriger war

die Aufgabe zu lösen. Für die Zählschleife sind keine Regelmäßigkeiten innerhalb der gebildeten Kategorien zu erkennen. Zusätzlich liegen pro Subthema (Zählschleife und Bedingte Schleife) wenige Items vor, um adäquate Rückschlüsse auf schwierigkeiterzeugende Merkmale zu ziehen. In diesem Zusammenhang wäre es sinnvoll, zukünftig leicht abgeänderte Items zu jedem Item und weitere neue Aufgabentypen zu entwickeln, um darüber weitere schwierigkeiterzeugende Kategorien zu identifizieren.

**Tabelle 18***ITERATION schwierigkeiterzeugende Merkmale*

Aufgabe	Anzahl der Befehlsauswahl	Anzahl der gesuchten Befehle	Wiederholung der Schleife	Anzahl von möglichen Lösungen
BeSch 1	0	1	2	3
ZäSch 12	8	4	2, 4	256
BeSch 2	0	1	4	3
ZäSch 13	15	6	1, 2 und 3	> 256
BeSch 4	0 (Zielfeld wird in Aufgabenstellung genannt)	1	5	3
ZäSch 11 / BeSch 3	8 / 0 (Zielfeld wird in Aufgabenstellung genannt)	2 / 1	5 / 3	16 / 3
ZäSch 10	3	1	4	3

Aufgabe	Anzahl der Befehlsauswahl	Anzahl der gesuchten Befehle	Wiederholung der Schleife	Anzahl von möglichen Lösungen
BeSch 6	0 (Felder müssen angemalt werden)	1	7	1
BeSch 1	0 (Zielfeld muss gefunden werden)	1	3	3
ZäSch13	15	6	1, 2 und 3	> 256
ZäSch 12	8	4	2 und 4	256
BeSch 5	0 (Felder müssen angemalt werden)	1	3	1
BeSch 3	0 (Ziel wird in Aufgabenstellung genannt)	1	2	3
BeSch 4 / ZäSch 11	0 (Ziel wird in Aufgabenstellung genannt) / 8	1 / 2	4 / 5	3 / 16

*Anmerkung.* oben = Iteration A. unten = Iteration B. / = Trennung von Aufgaben, welche eine identische Itemschwierigkeit aufweisen. Die Aufgaben sind nach Schwierigkeit von oben (schwierig) nach unten (leicht) sortiert. ZäSch = Zählschleife. BeSch = Bedingte Schleife.

Insgesamt lässt sich feststellen, dass sich auf Grund der geringen Anzahl von Items lediglich innerhalb der Sequenz erste schwierigkeiterzeugende Merkmale identifizieren lassen. Zu den Themen der Alternative und Iteration lassen sich ausschließlich Tendenzen aufzeigen. Da keine ausreichend große Menge an Items vorliegt, um sie zu einer Gruppe zusammenzufassen. Somit

lassen sich keine schwierigkeiterzeugenden Merkmale für Itemgruppen identifizieren. Um zu schlüssigen Interpretationen zu gelangen, wird eine hohe Anzahl und eine Bandbreite von Items innerhalb einer Gruppe benötigt. Zusätzlich beeinflusst die geringe Anzahl von gefundenen Kategorien die Aussagekraft hinsichtlich der a priori festgesetzten Kompetenzformulierungen. Diese können in diesem Kontext nicht überprüft werden, da weitere Gruppen von schwierigkeiterzeugenden Merkmalen benötigt werden. Zukünftig können innerhalb der Gruppen Gemeinsamkeiten der Items genutzt werden, um Rückschlüsse zu den formulierten Kompetenzen zu ziehen.

### 5.3 Gedrehte oder gespiegelte Abbildungen

**FF 3:** Was bewirken gedrehte oder gespiegelte Abbildungen in den Items?

Die Ergebnisse der beiden Testversionen A und B ergeben hinsichtlich gedrehter oder gespiegelter Abbildungen unterschiedliche Itemschwierigkeiten. Die Analyse der Aufgaben hat gezeigt, dass Aufgaben, in denen die Figur von links nach rechts bewegt werden soll, einfacher von Kindern zu lösen sind als Aufgaben, in denen die Figur von rechts nach links bewegt werden soll. Zusätzlich zeigt das Ergebnis, dass eine Aufgabe, in der eine Figur von oben nach unten bewegt wird, einfacher zu lösen ist als eine Aufgabe, in der die Figur von unten nach oben gesteuert werden soll. Daher kann davon ausgegangen werden, dass durch die Veränderung der Laufrichtung der Figur (vorgenommen durch eine Drehung oder Spiegelung der Abbildung) unterschiedliche Aufgabenschwierigkeiten erzeugt werden. Somit deutet sich eine Bestätigung der Hypothese 1 an, dass die Leserichtung die Aufgabenschwierigkeit beeinflusst. Deshalb sollten Aufgaben nicht gespiegelt oder gedreht werden, um Abschreibeffekte zu verhindern, da dadurch eine Gruppe schwieriger beziehungsweise leichter Aufgaben zugeteilt bekommt. Interessant an dieser Stelle wäre es, herauszufinden, inwiefern die unterschiedlichen Laufrichtungen (unten-oben und links-rechts) in verschiedenster Weise die Aufgabenschwierigkeit beeinflussen. Hier wäre es auch spannend, zu erfahren, wie Kinder diese Aufgaben lösen, die von rechts nach links lesen.

## 5.4 Prozedurales Wissen

**FF 4:** Verfügen Kinder der zweiten Klasse über prozedurales Wissen in Bezug auf die Basisprogrammierbausteine Sequenz, Alternative und Iteration?

Die Interpretation der Häufigkeitsverteilungen nach Inhalten (siehe Kapitel 4.2.2) liefern interessante Ergebnisse hinsichtlich des prozeduralen Wissens in Bezug auf die Programmierbausteine Sequenz, Alternative und Iteration von Kindern aus der zweiten Klasse. So liegt der Mittelwert der erreichten Punktzahl innerhalb der Sequenz (Version A) bei 4.7. Insgesamt 37 von 70 Kindern erreichen einen Wert von 5 oder höher (max. 9). Lediglich sechs Kinder erlangen eine Punktzahl von Null. Innerhalb der Sequenz (B) erzielen 29 von 64 Kindern 4 oder mehr von 8 möglichen Punkten. Der Mittelwert liegt hier bei 3.9. Nur vier Kinder erzielen keine Punkte. Dass insgesamt zehn Kinder innerhalb der Testversion A und B keinen Punkt erreichen, lässt sich damit begründen, dass lediglich zwei Aufgaben einer leichten Aufgabenschwierigkeit entsprechen und weitere sehr einfache Aufgaben benötigt würden, um das Konstrukt angemessen zu repräsentieren. Zusätzlich sind nach den Schwierigkeitsberechnungen einige sehr einfache Aufgaben aussortiert worden, welche von diesen Kindern ggf. hätten gelöst werden können. Insgesamt erreichen die Kinder sowohl in Version A als auch Version B eine angemessene Punktzahl, was dazu führt, die Hypothese 2 anzunehmen: Kinder der zweiten Klasse verfügen über ein prozedurales Wissen in Bezug auf den Programmierbaustein der Sequenz.

Die Ergebnisse der Alternative zeigen, dass über 80 Kinder von 105 Kindern einen Wert von 4 oder 5 Punkten erreicht haben. Maximal konnten 5 Punkte erreicht werden. Der Mittelwert liegt bei  $M = 4.1$  und die Standardabweichung bei  $SD = 1.38$ . Nur eine sehr geringe Anzahl von Kindern haben keine Punkte erreicht. Diese Werte bestätigen, dass es sich insgesamt um einfache Aufgaben handelt, welche von der Mehrzahl der Kinder richtig gelöst werden konnten. Um eine umfangreichere Interpretation der Alternativen-Aufgaben vornehmen zu können, bedarf es weiterer schwieriger und mittel schwieriger Aufgaben. Hier wäre interessant, in welchem Schwierigkeitsumfang die Kinder die Aufgaben lösen würden. Insgesamt kann die Hypothese 3 folgendermaßen bestätigt werden: Die Kinder der zweiten Klasse verfügen über ein einfaches prozedurales Wissen in Bezug auf den Programmierbaustein der Alternative.

Die Ergebnisse der Iteration (Version A) zeigen, dass 44 von 61 Kindern einen Testwert von 5 oder höher (max. 8) erreichen. Auch im Testheft der Iteration (Version B) erreichen 41 von 58

Kindern einen Wert von 5 Punkten oder höher (max. 8). In beiden Testversionen erlangen alle Kinder mindestens einen Punkt. Dies spricht dafür, dass die Kinder über prozedurales Wissen in Bezug auf den Programmierbaustein der Iteration verfügen und die Hypothese 4 somit bestätigt werden kann. Jedoch ist anzumerken, dass die Testhefte überwiegend einfach zu lösende Aufgaben abdecken und weitere Aufgaben insbesondere im oberen und mittleren Schwierigkeitsbereich notwendig sind, um differenzierte Aussagen hinsichtlich des prozeduralen Wissens treffen zu können.

## 5.5 Geschlechterunterschiede

**FF 5:** Wie unterscheiden sich die Teilgruppen hinsichtlich des Geschlechts in Bezug auf die Basisprogrammierbausteine Sequenz, Alternative und Iteration?

In der Testversion A der Sequenz sind insbesondere die Randbereiche auffällig. So haben sechs Mädchen keine Punkte erreichen können, bei den Jungen hingegen hat nur ein Kind null Punkte. Darüber hinaus konnten vier Jungen die volle Punktzahl erreichen, wobei es bei den Mädchen nur ein Kind ist. Zusätzlich liegt der Mittelwert bei den Jungen um 1.5 Punkte höher als bei den Mädchen ( $M = 4.0$ ). In beiden Geschlechtergruppen sind 35 Kinder vertreten. Die Überprüfung der Mittelwertunterschiede mittels  $t$ -Test ergab keine signifikanten Unterschiede zwischen Jungen und Mädchen. Dies bedeutet, dass das Geschlecht nicht ausschlaggebend für die Mittelwertunterschiede ist. In der Testversion B sind die Auffälligkeiten zwischen Jungen und Mädchen extremer ausgeprägt. So erreichen deutlich mehr Mädchen 0 bis 3 Punkte im Verhältnis zu den Jungen. Weiter erreichen mehr Jungen im Vergleich zu den Mädchen einen Testwert von 5 bis 8 Punkten. Diese Beobachtungen werden durch die Mittelwerte gestützt. Die Jungen erreichen durchschnittlich 4.8 Punkte und die Mädchen hingegen nur 2.7 Punkte. Dabei ist zu beachten, dass eine ungleiche Anzahl von Teilnehmern (Jungen  $n = 37$ ) und Teilnehmerinnen (Mädchen  $n = 27$ ) vorliegt. Die Mittelwerte der Jungen ( $M = 4.8$ ) liegen deutlich über denen der Mädchen ( $M = 2.7$ ). Die Berechnung des  $t$ -Tests zeigt einen signifikanten Unterschied zwischen den Geschlechtern Jungen und Mädchen. Somit ist das Geschlecht ausschlaggebend für das Erreichen einer höheren beziehungsweise niedrigeren Punktzahl. Ein Grund für unterschiedlichen Ergebnisse in Bezug auf das Testheft Version A und Version B kann zum einen die parallele Testform sein, die unterschiedlichen Aufgaben umfasst. Ein weiterer Grund könnte sein, dass in beiden Versionen eine unterschiedlich hohe Gesamtpunktzahl erreicht werden kann. Ein tiefergehender Vergleich zwischen Version A und

Version B soll an dieser Stelle nicht erfolgen, da das Ziel zukünftiger Studien die Konzeption eines Testhefts sein sollte. Aus diesem Grund sind erneute Gruppenvergleiche in weiteren Studien zu empfehlen. Insgesamt betrachtet kann die Hypothese 5 folgendermaßen bestätigt werden: Jungen, welche das Testheft B bearbeitet haben, verfügen über ein höheres prozedurales Wissen innerhalb des Programmierbausteins der Sequenz als Mädchen. In Bezug auf das Testheft A lässt sich die Hypothese widerlegen, denn Jungen und Mädchen verfügen gleichermaßen über prozedurales Wissen innerhalb des Programmierbausteins der Sequenz.

Die Ergebnisse über das prozedurale Wissen des Programmierbausteins der Alternative zeigen, dass eine ähnliche Häufigkeitsverteilung zwischen Jungen und Mädchen bei einem Testwert von 0 bis 4 Punkten vorliegt. Lediglich den Testwert von 5 Punkten erreichen deutlich mehr Mädchen. In der Gruppe der Mädchen haben zwölf Kinder mehr an der Testung teilgenommen als in der Gruppe der Jungen. Die Mittelwerte der Jungen ( $M = 4.0$ ) und Mädchen ( $M = 4.2$ ) liegen sehr eng beieinander. Die Berechnung des  $t$ -Tests ergab keine signifikanten Unterschiede zwischen den Geschlechtern Jungen und Mädchen. Insgesamt liegt bei Jungen und Mädchen ein vergleichbares prozedurales Wissen über den Programmierbaustein der Alternative vor. Die Hypothese 6 kann somit nicht bestätigt werden. Zukünftige Arbeiten sollten untersuchen, wie sich das prozedurale Wissen zwischen Mädchen und Jungen beim Hinzufügen von schwierigen und mittel schwierigen Aufgaben verhält.

Die Ergebnisse der Iteration (Version A) zeigen, dass bei ähnlich großen Stichproben von Jungen und Mädchen insbesondere bei den hohen Punktzahlen (7 und 8) die Jungen besser abschneiden. Lediglich die Testwerte 5 und 6 erreichen deutlich mehr Mädchen. Die Betrachtung der  $t$ -Tests ergeben keine Signifikanz, weshalb das Geschlecht Junge oder Mädchen nicht ausschlaggebend für die Mittelwertunterschiede ist. In der Testversion B sind die Verteilungen zwischen Mädchen und Jungen relativ ausgeglichen, wie beispielsweise die eng aneinander liegenden Mittelwerte zwischen Jungen ( $M = 5.6$ ) und Mädchen ( $M = 5.2$ ) zeigen. Auch die Überprüfung des  $t$ -Tests auf eine vorliegende Signifikanz kann nicht bestätigt werden, weshalb die Geschlechter nicht für die Mittelwertunterschiede verantwortlich sind. Daraus ergibt sich, dass die Hypothese 7 widerlegt ist und dass Jungen und Mädchen über ein vergleichbares prozedurales Wissen im Bereich der Iteration verfügen.

## 5.6 Konstruktüberschneidung mathematische Fähigkeiten

**FF 6:** Lässt sich eine Konstruktüberschneidung zu mathematischen Fähigkeiten herstellen?

Zur Beantwortung der Forschungsfrage werden die Ergebnisse aus den Korrelationsberechnungen zwischen dem Subtest des PRO SU 2 und dem DEMAT 2+ interpretiert. Zwischen dem Subtest des PRO SU 2 (Version A) und dem DEMAT 2+ liegen diskriminante Validitäten mit Werten unter  $r < 0.60$  vor. Die Testversion B hingegen weist eine starke Korrelation zu dem Subtest Sachaufgaben auf ( $r = 0.63$ ). Eine inhaltliche Begründung des Zusammenhangs lässt sich nicht identifizieren. Dieser Zusammenhang könnte auf Grund ähnlicher Aufgabenstellungen und Aufgabenauswahlen entstanden sein. Zusätzlich sind die mittelmäßigen Korrelationen zwischen der Sequenz und weiteren einzelnen Subtests anzumerken. Die Hypothese 8, dass diskriminante Validitäten zwischen der Sequenz und mathematischen Inhalten vorliegen, kann nicht sicher bestätigt werden. Hier sind weitere Validierungsstudien auch mit anderen Mathematiktests notwendig, welche zusätzlich die mathematischen Inhalte wie Muster und Strukturen und mit einem höheren Anteil den Bereich Raum und Form abdecken. Des Weiteren hat sich mittels Korrelationsberechnungen bestätigt, dass keine Zusammenhänge zwischen der Alternative und dem DEMAT 2+ vorliegen. Somit kann die Hypothese 9, dass diskriminante Validitäten zwischen der Alternative und mathematischen Inhalten vorliegen, angenommen werden. Innerhalb der Iteration haben sich starke Zusammenhänge zu dem arithmetischen Bereich ergeben. So liegen konvergente Validitäten zwischen der Iteration und dem Subtest der Division und der Sachaufgaben vor und die Hypothese 10 kann bestätigt werden. Dieser Zusammenhang lässt sich inhaltlich begründen, da es sich bei Divisionsaufgaben um Umkehraufgaben der Multiplikation handelt und reine Multiplikationsaufgaben im DEMAT 2+ Test in den Sachaufgaben aufgegriffen werden. Ein Zusammenhang zwischen dem Subtest Längenvergleich und Iteration lässt sich inhaltlich und hinsichtlich des Aufgabentyps nicht schlüssig begründen. Ergänzende Studien sind an dieser Stelle notwendig, um weitere Aussagen treffen zu können.

Darüber hinaus sollen an dieser Stelle Gründe für die unterschiedlichen Korrelationswerte zwischen der Testversion A und B geliefert werden. Zum einen könnten die unterschiedlichen Testaufgaben innerhalb der Version A und B bei der Sequenz und der Iteration zu Korrelationsabweichungen beitragen. Zusätzlich sind die Aufgaben in Testversion A und B nicht gleichermaßen in die Schwierigkeitsbereiche einzuordnen (Tabelle 16, Tabelle 17, Tabelle 18), was eine Differenz in den Werten herbeiführen könnte. Die geringen

Abweichungen zwischen der Alternative A und B lässt sich damit erklären, dass es sich um eine Pseudoparallellform handelt und nur die Lösungen oder Aufgaben anders sortiert wurden. Alle Aufgaben sind dem gleichen Schwierigkeitsniveau zuzuordnen. Weiter könnte eine Klumpenstichprobe andere Ergebnisse liefern, da die Validierungsstudie nur in einer Schule durchgeführt wurde. Eine breitere Streuung der Probandinnen und Probanden wäre zu empfehlen. Ein weiterer Grund könnte die Stichprobengröße von 32 bzw. 31 sein. Eine größere Stichprobengröße würde aussagekräftigere Ergebnisse liefern. Interessant wäre eine weitere Testversion, welche beide Testhefte vereint betrachtet. Für den neuen Test müssten zusätzlich weitere Testitems entwickelt werden, damit alle Schwierigkeitsniveaus gleichermaßen repräsentiert sind. Der Test müsste anschließend in einer weiteren Pilotierung mit hoher Stichprobengröße erprobt werden. Danach wäre eine Validierungsstudie mit einer hohen Stichprobengröße empfehlenswert. Mit diesen Schritten sind zuverlässigere Ergebnisse zu erwarten, um aussagekräftigere Aussagen hinsichtlich überschneidender und abgrenzender Themenbereiche zwischen informatischen und mathematischen Inhalten zu erhalten.

## 5.7 Konstruktüberschneidung Intelligenz

**FF 7:** Lässt sich eine Konstruktüberschneidung zur Intelligenz herstellen?

Die Annahme, dass geringfügige Zusammenhänge zwischen dem CFT 1-R und dem PRO SU 2 vorliegen, können teilweise bestätigt werden. So sind die Korrelationswerte der Alternative mit dem Intelligenztest gering, sodass angenommen werden kann, dass der Subtest der Alternative keine fluide Intelligenz abgefragt hat und die Hypothese 12 somit bestätigt ist. Die zu erwartende Annahme, dass zwischen der Sequenz und dem CFT 1-R geringe Zusammenhänge vorliegen kann bestätigt werden (Hypothese 11). Jedoch ist anzumerken, dass mittlere Zusammenhänge zwischen der Sequenz (Version B) mit den Subtests Labyrinth, Reihenfortsetzen, Klassifikationen und Matrizen sowie der Sequenz (Version A) mit dem Subtest Klassifikationen vorliegen. Hier wären eine größere Stichprobe sowie eine Klumpenstichprobe hilfreich, um weitere Rückschlüsse ziehen zu können. Insgesamt ist die Hypothese 11 nicht abschließend zu beantworten. Die Betrachtung der Korrelationswerte der Iteration in Bezug auf die Subtests der Intelligenz lässt den Rückschluss zu, dass keine Zusammenhänge vorliegen. Denn lediglich zu dem Subtest der Klassifikationen und Matrizen weist die Iteration (Version B) mittlere Korrelationen auf. Der Rest der Korrelationswerte zeigt keine nennenswerte Beziehung zu dem CFT 1-R Test. Insgesamt lässt sich Hypothese 13 somit

bestätigen. Zusätzlich wäre ein weiterer Intelligenztest, welcher die kristalline Intelligenz abfragt, interessant. Darüber könnte herausgefunden werden, ob es sich bei den Programmierbausteinen um Fähigkeiten handelt, welche im Lauf der Zeit von den Kindern erlernt werden.

## 6 Zusammenfassung und Ausblick

In dieser Dissertation konnte ein Fundament für ein objektives, reliables und valides Messinstrument zur grundlegenden Programmierkompetenz im Sachunterricht für zweite Klassen konzipiert werden. In diesem Zusammenhang wurden mit Hilfe einer Feldstudie Programmieraufgaben mit 12 Kindern der zweiten Klasse in Eins-zu-eins-Situationen erprobt. Basierend auf den daraus resultierenden Ergebnissen wurden in einer Vorstudie ein Itempool und Testinstruktionen zu den Testheften Sequenz, Alternative und Iteration mit 125 Kindern getestet. Anschließend erfolgte die Überprüfung der modifizierten Aufgaben und Testinstruktionen in einer Hauptstudie mit 378 Kindern. Nach dem methodischen Vorgehen erfolgte die Durchführung der Itemanalyse und der Überprüfung des Messinstruments mittels der Rasch-Analyse. Das Ergebnis bildet einen Itempool von 35 unterschiedlichen Items. Für die Sequenz existieren 17 Items, für die Alternative 5 Items und für die Iteration 13 Items. Die Ergebnisse zeigen, dass sich für die Aufgaben der Sequenz drei schwierigkeits erzeugende Merkmale charakterisieren lassen. In diesem Kontext sind „die Anzahl der unterschiedlichen Befehlsauswahl“, „die Anzahl der gesuchten Befehle für die Lösung“ und „die Anzahl der vorkommenden Befehlswechsel innerhalb der gesuchten Lösung“ zu nennen. Für die Alternative konnten keine konkreten schwierigkeits erzeugende Merkmale identifiziert werden. Erste Ergebnisse lassen jedoch vermuten, dass die Zustimmung oder Ablehnung der Bedingung bezüglich einfacher Aufgaben nicht ausschlaggebend für die Schwierigkeit ist. Innerhalb der Iteration lassen sich auf Grund der widersprüchlichen Ergebnisse, der Aufteilung in Bedingte Schleife und Zählschleife und den daraus resultierenden wenigen Aufgaben keine schwierigkeits erzeugenden Merkmale ableiten. Zusätzlich konnte gezeigt werden, dass gedrehte oder gespiegelte Abbildungen und die Laufrichtung der Figur die Aufgabenschwierigkeit beeinflussen. Aufgaben in welchen die Figur der Leserichtung folgt, also von links nach rechts bewegt wird, sind für Kinder einfacher zu lösen. Weitere wichtige Erkenntnisse liefert die Betrachtung des prozeduralen Wissens zu den Programmierbausteinen Sequenz, Alternative und Iteration. Die Berechnungen zeigen, dass Kinder der zweiten Klasse über ein prozedurales Wissen in Bezug auf die Sequenz verfügen. Zusätzlich konnte begründet werden, dass zu den Programmierbausteinen Alternative und Iteration ein simples prozedurales Wissen bei den Kindern vorliegt. In diesem Zusammenhang ergab sich, dass bis auf eine Ausnahme das Geschlecht nicht ausschlaggebend für das Erreichen einer hohen beziehungsweise niedrigen Punktzahl ist. Die Ausnahme bildet das Testheft der Sequenz (Version B). Hier konnte die Mittelwertabweichung auf das Geschlecht zurückgeführt werden, Jungen erreichten einen höheren Testwert als Mädchen. Abschließend zeigen die Ergebnisse

der Konstruktüberschneidungen, dass kein Zusammenhang zwischen mathematischen Fähigkeiten und der Alternative besteht. Es existiert aber ein Zusammenhang zwischen der Iteration und den Subtests Division und Sachaufgaben, was auf den abgefragten Inhalt der Multiplikation zurückzuführen ist. Ob Zusammenhänge zwischen der Sequenz und mathematischen Fähigkeiten vorliegen ist nicht abschließend zu beantworten, da teilweise Korrelationen zu der Testversion B vorliegen zur Testversion A hingegen nicht. Eine Konstruktüberschneidung zwischen der fluiden Intelligenz und den Programmierbausteinen der Alternative und der Iteration liegt nicht vor. Allerdings finden sich in diesem Kontext Korrelationen zwischen der Sequenz und den Subtests der Intelligenz. Hier sind weitere Überprüfungen notwendig, um einen Zusammenhang zu bestätigen oder zu widerlegen.

Damit aus dem Fundament des entwickelten Messinstruments ein konstruktvalides Messinstrument entworfen wird, sind weitere Schritte zu empfehlen. Der erste Schritt umfasst eine Erweiterung des Itempools. Dafür sollte zu Beginn der Itempool aller Testhefte der Versionen A und B zu einem Itempool zusammengefasst werden. Zudem ist zu empfehlen, weitere Items zu jedem Inhalt (Sequenz, Alternative und Iteration), basierend auf den Ergebnissen der Personen-Item-Maps und den schwierigkeiterzeugenden Merkmalen, zu konzipieren. Danach ist die Entwicklung eines Pseudo-Paralleltests ratsam. Der zweite Schritt beinhaltet die Durchführung einer quantitativen Studie mit einer Klumpenstichprobe von über 300 Kindern. Anschließend sind die in der Dissertation beschriebenen Schritte der Itemanalyse und die Überprüfung des Messmodells zu empfehlen. Auf Grund der höheren Stichprobe und des angepassten Itempools sind aussagekräftige Ergebnisse hinsichtlich des prozeduralen Wissens basierend auf einem gesamten Testwert zu erwarten. Zusätzlich bietet sich in diesem Zusammenhang eine erneute Untersuchung der Testgütekriterien (Objektivität und Reliabilität) an. Darüber hinaus sollten weitere Validierungsstudien durchgeführt werden, um konvergente und diskriminante Validitäten hinsichtlich mathematischer Fähigkeiten und (kristalline) Intelligenz zuverlässiger bestimmen zu können. Des Weiteren könnte ein Test zum Hörverstehen interessante Ergebnisse bereithalten. Zum Schluss ist es ratsam eine Überprüfung der formulierten Kompetenzerwartungen vorzunehmen. Ebenso könnte eine Erweiterung des Tests für die Klassen 1, 3 und 4 sehr aufschlussreich sein. Hier wäre insbesondere interessant, herauszufinden, welche Ausprägungen das prozedurale Wissen der Kinder in unterschiedlichen Jahrgangstufen zeigt. Gibt es Inhalte, die schon recht früh oder erst spät erworben werden? Wie können die Inhalte an verschiedene Klassenstufen angepasst werden? Anschließend wäre eine Konzeption weiterer Messinstrumente basierend auf dem Prozessdimensionen des

Implementierens und des Testens interessant. Darüber hinaus könnten zukünftige Forschungsvorhaben mit Hilfe des PRO SU 2 Tests Interventionsstudien durchführen, welche konkrete Vorschläge und Ideen zur Umsetzung von Programmiermaterialien im Sachunterricht bieten. Entscheidende Fragen sind, welche Programmiermaterialien einen größeren Wissenszuwachs bieten und warum dies so ist.

Insgesamt konnte das Ziel der Dissertation, ein Fundament zur Entwicklung eines Messinstruments zur grundlegenden Programmierkompetenz im Sachunterricht für Kinder der zweiten Klasse zu entwickeln, erfüllt werden. Die Ergebnisse bilden eine wesentliche Basis für zukünftige Testentwicklungen. Eine künftige Endversion des Programmiertests gibt Lehrkräften wiederum die Möglichkeit, ihren Unterricht didaktisch individuell auf die Bedürfnisse des Kindes abzustimmen. Auf diese Weise erhalten Kinder die Gelegenheit, unerklärliche Phänomene und Zusammenhänge bestmöglich zu begreifen, wie die in dieser Arbeit untersuchten informatischen Inhalte.

## Literatur

- American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME) (2014). *Standards for Educational and Psychological Testing*. AERA. Abgerufen April 24, 2024, von [https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards\\_2014edition.pdf](https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf)
- Bergius, R. J. W. (2022, August 02). *Feldstudie, Feldforschung*. Abgerufen April 24, 2024, von <https://dorsch.hogrefe.com/stichwort/feldstudie-feldforschung>
- Bergner, N., Hubwieser, P., Köster, H., Magenheimer, J., Müller, K., Romeike, R., Schroeder, U., & Schulte, C. (2018). *Frühe informatische Bildung – Ziele und Gelingensbedingungen für den Elementar- und Primarbereich*. Barbara Budrich. <https://doi.org/10.2307/j.ctvbkk1sq>
- Bers, M. U. (2007). Project Interactions: A multigenerational robotic learning environment. *Journal of Science and Technology Education*, 16(6), 537–552. <https://doi.org/10.1007/s10956-007-9074-2>
- Boone, W., Staver, J., & Yale, M. (2014). *Rasch Analysis in the Human Sciences*. Springer. <https://doi.org/10.1007/978-94-007-6857-4>
- Boone, W., & Staver, J. (2020). *Advances in Rasch Analysis in the Human Sciences*. Springer. <https://doi.org/10.1007/978-3-030-43420-5>
- Bortz, J., & Döring, N. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Bortz, J., & Lienert, G. A. (2008). *Kurzgefasste Statistik für die Klinische Forschung* (3. Überarb. Aufl.). Springer. <https://doi.org/10.1007/978-3-540-75738-2>
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Überarb. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-12770-0>
- Borowski, C., Diethelm, I., & Stoffers, A.-M. (2010). Informatische Bildung im Sachunterricht der Grundschule. *widerstreit-sachunterricht*, 15, 1–8. <http://dx.doi.org/10.25673/92431>
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Personality and Social Psychology*, 54(1), 106–148. <https://doi.org/10.1111/j.1467-6494.1986.tb00391.x>

- Brown, R. L., Obasi, C. N., & Barrett, B. (2016). Rasch analysis of the WURSS-21 dimensional validation and assessment of invariance. *Journal of Lung, Pulmonary, and Respiratory Research*, 3(2), 46–53. <https://doi.org/10.15406/jlpr.2016.03.00076>
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* (4. Überarb. Aufl.). Pearson.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Überarb. Aufl.). Pearson.
- Bundesweite Informatikwettbewerbe (BWINF) (2023, Dezember 04). *Informatik-Biber. Aufgabensammlung*. BWINF. Abgerufen April 24, 2024, von <https://bwinf.de/biber/aufgaben/>
- Bundesministerium für Bildung und Forschung (BMBF) (2003). *Eine Expertise: Zur Entwicklung nationaler Bildungsstandards*. Referat Publikationen; Internetredaktion.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1–22. <https://doi.org/10.1037/h0046743>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Lawrence Erlbaum Associates.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Davis, D., & Boone, W. (2021). Using Rasch analysis to evaluate the psychometric functioning of the other-directed, lighthearted, intellectual, and whimsical (OLIW) adult playfulness scale. *International Journal of Educational Research Open*, 2, 1 – 12. <https://doi.org/10.1016/j.ijedro.2021.10005>
- Edelmann, W., & Wittman, S. (2012). *Lernpsychologie* (7. Aufl.). Beltz.
- Eggert, S., & Bögeholz, S. (2014). Entwicklung eines Testinstruments zur Messung von Schülerkompetenzen. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 371–384). Springer. [https://doi.org/10.1007/978-3-642-37827-0\\_2](https://doi.org/10.1007/978-3-642-37827-0_2)
- Eid, M., & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Hogrefe.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2. Aufl.). Cambridge: MIT. <https://doi.org/10.7551/mitpress/5657.001.0001>

- Ficher, P., & Hofer, P. (2011). *Lexikon der Informatik* (15. Überarb. Aufl.). Springer.  
<https://doi.org/10.1007/978-3-642-15126-2>
- Field, A. (2018). *Discovering statistics using ibm spss statistics*. Sage.
- Finger, R. P., Fenwick, E., Pesudovs, K., Marella, M., Lamoureux, E. L., & Holz, F. (2012). Rasch Analysis Reveals Problems with Multiplicative Scoring in the Macular Disease Quality of Life Questionnaire. *Ophthalmology*, *119*(11), 2351–2357.  
<https://doi.org/10.1016/j.ophtha.2012.05.031>
- Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21*(1), 1095. Abgerufen April 24, 2024, von <https://www.rasch.org/rmt/rmt211m.htm>
- Fisseni, H.-J. (1990). *Lehrbuch der psychologischen Diagnostik*. Hogrefe.
- Gesellschaft für Didaktik des Sachunterrichts (GDSU) (2021): *Positionspapier Sachunterricht und Digitalisierung*. Erarbeitet von der AG Medien & Digitalisierung der Gesellschaft für Didaktik des Sachunterrichts – GDSU (Peschel, M., Gervé, F., Gryl, I., Irion, T., Schmeinck, D., & Straube, P.). Abgerufen April 24, 2024, von [https://gdsu.de/sites/default/files/PDF/GDSU\\_2021\\_Positionspapier\\_Sachunterricht\\_und\\_Digitalisierung\\_deutsch\\_de.pdf](https://gdsu.de/sites/default/files/PDF/GDSU_2021_Positionspapier_Sachunterricht_und_Digitalisierung_deutsch_de.pdf)
- Gesellschaft für Didaktik des Sachunterrichts (GDSU) (2013). *Perspektivrahmen Sachunterricht*. Julius Klinkhardt.
- Gesellschaft für Informatik e. V. (GI) (2019). Kompetenzen für informatische Bildung im Primarbereich. *Beilage zu LOG IN*, *39*(191/192), 1–26.
- Gibson, J. P. (2012). Teaching graph algorithm to children of all ages. In Association for Computing Machinery (Hrsg.), *ITiCSE'12: Proceedings of the ACM annual conference on Innovation and technology in computer science education* (S. 34–39). ACM.
- Gläser-Zirkuda, M. (2015). Qualitative Auswertungsverfahren. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Hrsg.). *Empirische Bildungsforschung: Strukturen und Methoden* (S.109 – 119). Springer. <https://doi.org/10.1007/978-3-531-19994-8>
- Goecke, L., Stiller, J., Pech, D., & Pinkwart, N. (2017). Informatische Grundbildung: Exploration des Erstzugangs zu Lego® Wedo 2.0 und Cubelets von Dirttklässler\_innen. In I. Diethelm (Hrsg.), *Informatische Bildung zum Verstehen und Gestalten der digitalen Welt* (S. 417–418). Gesellschaft für Informatik e.V.

- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265–275. <https://doi.org/10.1037/0033-2909.103.2.265>
- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In Klieme, K., & Beck, B. (Hrsg.). *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englische Schülerleistung International)* (S. 83–99). Beltz. <https://doi.org/10.25656/01:3143>
- Hartig, J., & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In Bundesministerium für Bildung und Forschung (BMBF) (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 17 – 36).
- Heinrichs, J. (2020). Programmieren im Sachunterricht. In K. Kaspar, M. Becker-Mrotzek, S. Hofhues, J. König, & D. Schmeinck (Hrsg.), *Bildung, Schule, Digitalisierung* (S. 339–344). Waxmann. <https://doi.org/10.31244/9783830992462>
- Hoffmann, S., Wendlandt, K., & Wendlandt, M. (2017). Algorithmisieren im Grundschulalter. In I. Diethelm (Hrsg.), *Informatische Bildung zum Verstehen und Gestalten der digitalen Welt* (S. 73–82). Gesellschaft für Informatik e.V.
- Holmefur, M. M., & Krumlinde-Stundholm, L. (2016). Psychometric properties of a revised version of the Assisting Hand Assessment (Kids-AHA 5.0). *Developmental Medicine & Child Neurology*, *58*(6), 618–624. <https://doi.org/10.1111/dmcn.12939>
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270. <https://doi.org/10.1037/h0023816>
- Kauertz, A. (2014). Entwicklung eines Rasch-skalierten Leistungstests. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 341–353). Springer. [https://doi.org/10.1007/978-3-642-37827-0\\_2](https://doi.org/10.1007/978-3-642-37827-0_2)
- Kauertz, A., Fischer H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, *16*, 135–153.
- Khadka, J., Huang, J., Chen, H., Chen, C., Gao, R., Bao, F., Zhang, S., Wang, Q., & Pesudovs, K. (2016). Assessment of Cataract Surgery Outcome Using the Modified Catquest Short-Form Instrument in China. *PLoS ONE*, *11*(10), Article e0164182. <https://doi.org/10.1371/journal.pone.0164182>

- Kleickmann, T., Hardy, I., Möller, K., Pollmeier, J., Tröbst, S., & Beinbrech, C. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 265–283.
- Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen?. *Pädagogik*, 6, 10–13.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876 – 903.
- Klopp, E. (2010). *Explorative Faktorenanalyse*. Abgerufen April 24, 2024, von <https://doi.org/10.23668/psycharchives.9238>
- Krajewski, K., Liehm, S. & Schneider, W. (2004). *Deutscher Mathematiktest für die zweite Klassen*. Beltz Test GmbH.
- Levene, H. (1960). Robust tests for equality of variances. In I., Olkin, Ghurye, G. S., Hoeffding, W., Madow, G. W., & Mann, H. B. (Hrsg.), *Contributions to Probability and Statistics* (S. 278–292). Stanford University Press.
- Li, C. Y., Romero, S., Bonilha, H. S., Simpson, K. N., Simpson, A. N., Hong, I., & Velozo, C. A. (2016). Linking existing instruments to develop an activity of daily living item bank. *Evaluation & the Health Professions*, 41(1), 25–43. <https://doi.org/10.1177/0163278716676873>
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Beltz.
- Linacre, J. M. (1998). Structure in Rasch residuals: Why principal components analysis (PCA)? *Rasch Measurement Transactions*, 12(2), 636. Abgerufen April 24, 2024, von <https://www.rasch.org/rmt/rmt122m.htm>
- Malec, J. F., Torsher, L. C., Dunn, W. F., Wiegmann, D. A., Arnold, J. J., Brown, D. A., & Phatak, V. (2007). The Mayo High Performance Teamwork Scale: Reliability and Validity for Evaluating Key Crew Resource Management Skills, *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare* 2(1), 4–10. <https://doi.org/10.1097/SIH.0b013e31802b68ee>
- Magenheim, J., & Romeike, R. (2019). *Informatikunterricht und Didaktik der Informatik*. <https://computingeducation.de/publications/>
- Mayer, J., & Wellnitz, N. (2014). Die Entwicklung von Kompetenzstrukturmodellen. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der*

- naturwissenschaftsdidaktischen Forschung* (S. 19–29). Springer.  
[https://doi.org/10.1007/978-3-642-37827-0\\_2](https://doi.org/10.1007/978-3-642-37827-0_2)
- Meyer-Bahlburg, H. F. L. (1969). Spearmans rho als punktbiserialer Rangkorrelationskoeffizient. *Biometrische Zeitschrift*, *11*(1), 60–66.  
<https://doi.org/10.1002/bimj.19690110105>
- Medienberatung NRW (Hrsg.). (2018). *Broschüre Medienkompetenzrahmen NRW*. Abgerufen April 24, 2024, von <https://medienkompetenzrahmen.nrw>
- Medienpädagogischer Forschungsverbund Südwest (mpfs) (Hrsg.). (2023). *KIM-Studie 2022. Kindheit, Internet, Medien. Basisuntersuchung zum Medienumgang 6- bis 13-Jähriger*. Abgerufen April 24, 2024, von <https://www.mpfs.de/studien/kim-studie/2022/>
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (MSB) (o.J.). Abgerufen April 24, 2024, von <https://www.schulministerium.nrw/schulsozialindex>
- Moosbrugger H. & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion* (3. Überarb. Aufl.). Springer. <https://doi.org/10.1007/978-3-662-61532-4>
- Neumann, K. (2014). Rasch-Analyse naturwissenschaftsbezogener Leistungstests. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 341–353). Springer.  
[https://doi.org/10.1007/978-3-642-37827-0\\_2](https://doi.org/10.1007/978-3-642-37827-0_2)
- O'Connor, J. P., Penney, D., Alfrey, L., Phillipson, S., Phillipson, S. N., & Jeanes, R. (2016). The Development of the Stereotypical Attitudes in HPE Scale. *Australian Journal of Teacher Education*, *41*(7), 70–87. <https://doi.org/10.14221/ajte.2016v41n7.5>
- Pólya, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsberechnung und das Momentenproblem. *Mathematische Zeitschrift*, *8*, 171–181.  
<https://doi.org/10.1007/BF01206525>
- Portelance, D. J., Strawhacker, A. L., & Bers, M. U. (2016). Constructing the Scratch Jr programming language in the early childhood classroom. *International Journal of Technology and Design Education*, *26*(4), 489–504. <https://doi.org/10.1007/s10798-015-9325-0>
- Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.
- Schecker, H. (2014). Überprüfung der Konsistenz von Itemgruppen mit Cronbachs  $\alpha$ . Chapter: Online-Zusatzmaterial. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden*

- in der naturwissenschaftsdidaktischen Forschung* (S. 1–6). Springer.  
[https://doi.org/10.1007/978-3-642-37827-0\\_2](https://doi.org/10.1007/978-3-642-37827-0_2)
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45 – 66.
- Schlieben, J. (2022). Testinstrumentenentwicklung zur Messung grundlegender Programmierkompetenz im Sachunterricht. In B. Brandt, L. Böll, & H. Dausend (Hrsg.), *Digitales Lernen in der Grundschule III. Fachdidaktiken in der Diskussion* (S. 353–366). Waxmann. <https://doi.org/10.31244/9783830995913>
- Schlieben, J. (i. D.). Testinstrumentenentwicklung zur Messung grundlegender Programmierkompetenz im Sachunterricht – Eine Beschreibung des methodischen Vorgehens und der Ergebnisse aus der Vorstudie. In B. Brandt, L. Böll, L., & H. Dausend (Hrsg.), *Digitales Lernen in der Grundschule IV*. Waxmann.
- Schmeinck, D. (2023). Computational Thinking in der Grundschule am Beispiel von Lego® Education und SPIKE™ Essential. In Irion, T., Peschel, M., & Schmeinck, D. (Hrsg.), *Grundschule und Digitalität. Grundlagen, Herausforderungen, Praxisbeispiele* (S. 289–296). Grundschulverband e.V. <https://doi.org/10.25656/01:25820>
- Schmeinck, D. (2020). Akzeptanzstudie „Hands on Coding“ – ausgewählte Tools, Softwareapplikationen und Programmiersprachen aus der Sicht von Grundschullehrer\*innen. In K. Kaspar, M. Becker-Mrotzek, S. Hofhues, J. König, & D. Schmeinck (Hrsg.), *Bildung, Schule, Digitalisierung* (S. 115–120). Waxmann. <https://doi.org/10.31244/9783830992462>
- Schmeinck, D. (2018). Wenn Roboter laufen lernen – Programmieren in der Grundschule. *Weltwissen Sachunterricht*, 1, 42–44.
- Schmitt, N. (1996). Use and abuses of coefficient alpha. *American Psychological Assessment*, 8(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>
- Scratch Foundation Inc. (2014). Scratch Jr. (1.5.11) [Mobile App]. App Store.  
<https://apps.apple.com/de/app/scratchjr/id895485086>
- Schräpler, J.-P., & Jeworutzki, S. (2021). *Konstruktion des Sozialindex für Schulen in Nordrhein-Westfalen*. ZEFIR. <https://doi.org/10.46586/rub.zefir.213.186>
- Schwill, A. (2001). Ab wann kann man mit Kindern Informatik machen? – Eine Studie über informatische Fähigkeiten von Kindern. In J. Magenheimer, & R. Keil-Slawik (Hrsg.), *Informatikunterricht und Medienbildung* (S. 13–30). Gesellschaft für Informatik e.V.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591–611. <https://doi.org/10.2307/2333709>

- Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2005). *Bildungsstandards im Fach Mathematik für den Primarbereich*. Luchterhand. [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2022/2022\\_06\\_23-Bista-Primarbereich-Mathe.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-Primarbereich-Mathe.pdf)
- Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2016). *Strategie der Kultusministerkonferenz „Bildung in der digitalen Welt“*. Abgerufen April 24, 2024, von [https://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2018/Digitalstrategie\\_2017\\_mit\\_Weiterbildung.pdf](https://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2018/Digitalstrategie_2017_mit_Weiterbildung.pdf)
- Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2022). *Bildungsstandards für das Fach Mathematik. Primarbereich*. Abgerufen April 24, 2024, von [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2022/2022\\_06\\_23-Bista-Primarbereich-Mathe.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-Primarbereich-Mathe.pdf)
- Straube, P., Brämer, M., & Köster, H. (2019). Freies Explorieren und Programmieren. Erkenntnisse über die Eignung didaktischer Materialien und informatischer Fokus für offene Unterrichtskonzepte in der Grundschule. In A. Pasternak (Hrsg.), *Informatik für alle. 18. GI-Fachtagung Informatik und Schule* (S. 367–368). Gesellschaft für Informatik e.V.
- Straube, P., Mamlouk, N., M., Köster, H., Nordmeier, V., Müller-Birn, C., & Schulte, C., (2013). DoInG – Informatisches Denken und Handeln in der Grundschule. *PhyDid B – Didaktik der Physik – Beiträge Zur DPG-Frühjahrstagung*, 1–6.
- Trautmann, T. (2010). *Interviews mit Kindern: Grundlagen, Techniken, Besonderheiten, Beispiele*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.25656/01:18710>
- Wagner, W. (2018). Erstes Algorithmisieren über die Darstellung von Bewegungen mittels grafischer Symbole: Vom eigenen Handeln zum Programmieren. *Medienimpulse*, 56(4). 1–11. <https://doi.org/10.21243/mi-03-18-14>
- Weigend, M. (2009). Algorithmik in der Grundschule. In B. Koeber (Hrsg.), *Zukunft braucht Herkunft. 25 Jahre INFOS – Informatik und Schule* (S. 97–108). Gesellschaft für Informatik e.V.
- Weise, G. (1975). *Psychologische Leistungstests. Ein Handbuch für Studium und Praxis*. Hogrefe
- Weiß, R. (1969). *Die Brauchbarkeit des Culture Free Intelligence Tests Skala 3 (CFT 3) bei begabungspsychologischen Untersuchungen*. Universität Würzburg.

- Weiß, H., R., & Osterland, J. (2013). *CFT 1-R. Grundintelligenztest. Skala 1*. Hogrefe
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410611697>
- Wong, M. H. Y., Fenwick, E., Aw, A. T., Lamoureux, E. L., & Seah, L. L. (2018). Development and validation of the Singapore thyroid eye disease quality of life questionnaire. *Translational Vision Science & Technology*, 7 (5). <https://doi.org/10.1167/tvst.7.5.14>
- Wright B. D., & Linacre J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transaction*, 8(3), 370. Abgerufen Mai 04, 2024, von <https://rasch.org/rmt/rmt83b.htm>
- Wright, B., & Masters, G. (1982). *Rating scale analysis*. MESA Press.
- Wright, B. D., Mead, R. J., & Ludlow, L. H. (1980). *KIDMAP*. Statistical Laboratory, Department of Education, The University of Chicago. Abgerufen April 24, 2024, von <https://www.rasch.org/memo29.pdf>
- Wright, B. D., & Stone, M. H. (1979). *Best Test Design: Rasch Measurement*. MESA Press.
- Wright, B. D., & Stone, M. H. (1999). *Measurement Essential* (2. Aufl.) Wide Range, INC. Abgerufen April 24, 2024, von <https://www.rasch.org/measess/met-1.pdf>

## Anhang

### A Fachbeiträge

Hinweis: Die Fachbeiträge wurden aus der veröffentlichten Dissertation herausgenommen. Alle publizierten Fachbeiträge sind frei zugänglich, da sie als Open Access Formate veröffentlicht wurden.

Fachbeitrag 1: Heinrichs, J. (2020). Programmieren im Sachunterricht. In K. Kaspar, M. Becker-Mrotzek, S. Hofhues, J. König, & D. Schmeinck (Hrsg.), *Bildung, Schule, Digitalisierung* (S. 339–344). Waxmann. <https://doi.org/10.31244/9783830992462>

Fachbeitrag 2: Schlieben, J. (2022). Testinstrumentenentwicklung zur Messung grundlegender Programmierkompetenz im Sachunterricht. In B. Brandt, L. Bröll, L., & H. Dausend (Hrsg.), *Digitales Lernen in der Grundschule III. Fachdidaktiken in der Diskussion* (S. 353–366). Waxmann. <https://doi.org/10.31244/9783830995913>

Fachbeitrag 3: Schlieben, J. (i. D.). Testinstrumentenentwicklung zur Messung grundlegender Programmierkompetenz im Sachunterricht – Eine Beschreibung des methodischen Vorgehens und der Ergebnisse aus der Vorstudie. In B. Brandt & L., Bröll (Hrsg.), *Digitales Lernen in der Grundschule IV*. Waxmann.

## B Feldstudie

### Interviewfragen

Anweisung: „Deine Aufgabe ist es Progis Ziel zu finden.“

Zusatz: Befehlspool und Befehlswiederholung wird mehr.

Fragen:

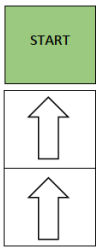
- Woher wusstest du, dass Progi dort ankommt?
- Wie bist du so schnell zum Ziel gekommen?
- Wie kommst du darauf?
- Woher weißt du das?
- Warum muss Progi so fahren?
- Wer sitzt in der Schule denn neben dir? Wie würdest Du es \_\_\_\_\_ erklären, der das nicht versteht?
- Was verwundert dich?
- Was fragst du dich?
  
- Sag bitte laut, was du denkst bzw. machst
- Wenn du eine Frage hast, dann frag einfach

Provokation: *Anderes Ziel nehmen.*


- Ich sage Progi kommt dort an. – Was meinst du?
- Ein Befehl weglassen und Wiederholung durchführen.

### Beispiel Sequenz

Befehl 1:





Befehl 1:



Befehl 1:

01: Start  
02: gehe\_hoch  
03: gehe\_hoch

Karte 3

Start:  Ziel: 

Ergebnis:

### Befehlskarten



Spielplan

Beispiel Alternative

Start:

Wenn oder von Progi ,  
dann fährt Progi bis ,  
sonst fährt Progi bis .

Ziel:

Start:

Wenn von Progi nicht ist,  
dann fährt Progi bis ,  
sonst fährt Progi .

Ziel:

Start:

Wenn von Progi ,  
dann fährt Progi ,  
sonst fährt Progi .

Ziel:

Befehlskarten: OR























NOT

AND

















































Spielfeld

Beispiel Iteration

<p>Start: </p> <p>Gehe </p> <p>Ziel: </p>	<p>Start: </p> <p>Gehe  </p> <p>Ziel: </p>	<p>Start: </p> <p>Gehe 3 mal  .</p> <p>Ziel:</p>	<p>Start: </p> <p>Gehe 5 mal  .</p> <p>Ziel:</p>		
<p>Start: </p> <p>Gehe    </p> <p>Ziel: Spielfeldrand</p>		<p>Start: </p> <p>Gehe 2 mal   .</p> <p>Ziel:</p>		<p>Start: </p> <p>Gehe 3 mal   .</p> <p>Ziel:</p>	

Befehlskarte: Bedingte Schleife

Zählschleife

Spielplan

## C Vorstudie

### Beispiel Sequenz: Aufgabe 7

Du hast nun das Blatt mit der **Aufgabe 7** vor dir liegen.

Die **Figur** möchte zum **Haus** (im Spielplan zeigen).

Zeichne die passenden Befehle **unten** (auf die Lösung zeigen) in die Kreise.

Nutze: nach unten, nach oben, nach rechts, nach links (zeigen).

Du kannst die Befehle öfter benutzen.

Wenn du fertig bist, lege deinen Stift hin und nimm die Hände hoch.

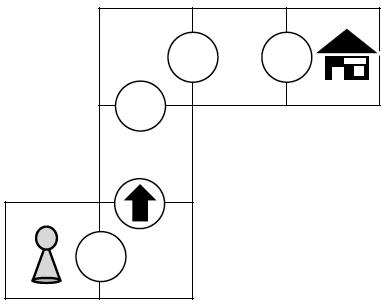
Blättere nun um.

Testinstruktion

Aufgabe 7





**Information**

Die **Figur** möchte zum **Haus**.




**Aufgabe**

Zeichne die passenden Befehle unten in die Kreise.

Nutze:     .

Du kannst die Befehle öfter benutzen.

**Lösung**



10

Testitem

## Beispiel Alternative: Aufgabe 1

Du hast nun das Blatt mit der **Aufgabe 1** vor dir liegen.

Steht die **Figur** auf einem **grauen** Feld (*im Spielfeld zeigen*)?

**Ja:** Die Figur geht nach links (*auf den Pfeil zeigen*).

**Nein:** Die Figur geht nach rechts (*auf den Pfeil zeigen*).

Bei welchem **Ziel** kommt die Figur an?

Kreuze die richtige Lösung an (*auf die Lösungsmöglichkeiten zeigen*).

Wenn du fertig bist, lege deinen Stift hin und nimm die Hände hoch.

Blättere nun um.

Testinstruktion

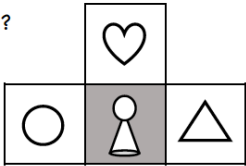
**Aufgabe 1**

**Informationen**

Steht die **Figur** auf einem **grauen** Feld?

**Ja:** Die Figur geht .

**Nein:** Die Figur geht .



**Aufgabe**

Bei welchem **Ziel** kommt die Figur an?  
Kreuze die richtige Lösung an.

**Lösung**

2

Testitem

### Beispiel Bedingte Schleife: Aufgabe 3

Du hast nun das Blatt mit der **Aufgabe 3** vor dir liegen.  
 Die Figur soll beim **Haus** ankommen.  
**Regel:** Steht die **Figur** auf einem **grauen** Feld (*im Spielfeld zeigen*)?  
**Ja:** Die Figur geht nach unten (*auf den Pfeil zeigen*).  
**Nein:** Die Figur geht nach links (*auf den Pfeil zeigen*).  
 Benutze die **Regel** (*zeigen*), bis die Figur beim Haus angekommen ist.  
 Male dafür die Felder richtig an (*auf die Lösungsmöglichkeiten zeigen*).  
 Finde eine richtige Lösung.  
 Wenn du fertig bist, lege deinen Stift hin und nimm die Hände hoch.  
 Blättere nun um.


#### Testinstruktion


**Aufgabe 3**



**Information**

Die Figur soll beim **Haus** ankommen.

**Regel:**  
 Steht die **Figur** auf einem **grauen** Feld?

**Ja:** Die Figur  .



**Nein:** Die Figur  .

**Aufgabe**

Benutze die **Regel** bis die Figur beim Haus angekommen ist.  
**Male** dafür die **Felder** richtig an.  
 Finde eine richtige Lösung.

**Lösung**

4

#### Testitem

## Beispiel Zählschleife: Aufgabe 2

Du hast nun das Blatt mit der **Aufgabe 2** vor dir liegen.

Die Figur geht 2-mal nach unten (*auf den Pfeil zeigen*).

Bei welchem **Haus** kommt die Figur an?

Kreuze die richtige Lösung an (*auf die Lösungsmöglichkeiten zeigen*).

Wenn du fertig bist, lege deinen Stift hin und nimm die Hände hoch.

Blättere nun um

## Testinstruktion

Aufgabe 2

**Information**

Die Figur geht:

.

	2	
1		
		3

**Aufgabe**

Bei welchem Haus kommt die Figur an?  
Kreuze die richtige Lösung an.

**Lösung**

1

2

3

3

## Testitem

---

Leitfragen der an die Testung anschließenden Interviews

- Was könnten andere Kinder schwer finden?
  - Was sollte ich verbessern?
  - Wie kommst du darauf?
  - Wie bist du auf das Ergebnis gekommen?
  - Nenne mir schwere Wörter.
  - Woher weißt du das?
  - Wie würdest du die Aufgabe einem anderen Kind erklären?
  - Was findest du an der Aufgabe komisch?
- 
- Sag bitte laut, was du denkst.

# D Hauptstudie

## Sequenz A

**Fragebogen - SE - A**

Alter: \_\_\_\_\_

Geschlecht:  Mädchen  
 Junge

0

**Beispielaufgabe**

Setze richtig ein.

?		
?		

7

**Aufgabe 7**

Setze richtig ein.

	?	
?	?	?
	?	

9





**Aufgabe 8**




Setze richtig ein.


	?	


10

**Aufgabe 9**

Setze richtig ein.    





		?
		?
	?	













11

**Aufgabe 10**




Setze richtig ein.    



















12

**Aufgabe 11**





Setze richtig ein.    



		
		




13

**Aufgabe 12**





Setze richtig ein.    



			
			




14

**Aufgabe 13**

Setze richtig ein.    






			








--	--	--	--	--	--	--	--

15

**Aufgabe 14**

Setze richtig ein.     






		
		
		








--	--	--	--	--	--	--	--	--

16

**Aufgabe 15**

Setze richtig ein.     



--	--	--	--	--	--	--	--	--

**ENDE!** 17

Sequenz B

**Fragebogen - SE - 8**

Alter: \_\_\_\_\_

Geschlecht:  Mädchen  
 Junge

9

**Beispielaufgabe**

Setze richtig ein.

?			
?			

--	--	--

7

**Aufgabe 7**

Setze richtig ein.

	?		
?	?	?	
?			

--	--	--	--	--	--	--

9

**Aufgabe 9**

Setze richtig ein.

		?	
	?		
	?		





--	--	--	--	--


11



**Aufgabe 14**

Setze richtig ein. 





		
		
		




○	○	○	○	○	○	○	○	○
---	---	---	---	---	---	---	---	---

**Aufgabe 15**

Setze richtig ein. 



○	○	○	○	○	○	○	○	○
---	---	---	---	---	---	---	---	---

## Testinstruktion Sequenz

Du hast nun das Blatt mit der **Beispielaufgabe** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** geht in das **Haus** (*im Spielplan zeigen*).  
Setze die richtigen Pfeile **hier** (*auf die Lösung zeigen*) ein.  
Nutze: nach unten, nach oben, nach rechts, nach links (*zeigen*).  
Du kannst die Pfeile mehrmals benutzen.  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 7** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** geht in das **Haus** (*im Spielplan zeigen*).  
Setze die richtigen Pfeile **hier** (*auf die Lösung zeigen*) ein.  
Nutze: nach unten, nach oben, nach rechts, nach links (*zeigen*).  
Du kannst die Pfeile mehrmals benutzen.  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 8** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** geht in das **Haus** (*im Spielplan zeigen*).  
Setze den richtigen Pfeil **hier** (*auf die Lösung zeigen*) ein.  
Nutze: nach unten, nach oben, nach rechts, nach links (*zeigen*).  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 9** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** geht in das **Haus** (*im Spielplan zeigen*).  
Setze die richtigen Pfeile **hier** (*auf die Lösung zeigen*) ein.  
Nutze: nach unten, nach oben, nach rechts, nach links (*zeigen*).  
Du kannst die Pfeile mehrmals benutzen.  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 10** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** sammelt alle Äpfel ein.  
Zum Schluss geht die **Figur** in das **Haus**.  
Setze die richtigen Pfeile (*auf die Lösung zeigen*) ein.  
Nutze: nach unten, nach oben, nach rechts, nach links (*zeigen*).  
Du kannst die Pfeile mehrmals benutzen.  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 15** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** geht einen schnellen Weg in das **Haus**.  
Auf dem Weg hebt die **Figur** alle **Äpfel** (*auf den Befehl zeigen*) auf.  
**Achtung:** Dafür gibt es einen neuen Befehl.  
Setze die richtigen Befehle (*auf die Lösung zeigen*) ein.  
Es können auch Kreise leer bleiben (*zeigen*).  
Nutze: nach unten, nach oben, nach rechts, nach links, aufheben (*zeigen*).  
Du kannst die Befehle mehrmals benutzen.  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Schließe nun das Heft. Du hast es geschafft. Vielen Dank für deine Hilfe (*Hefte einsammeln*)!

Du hast nun das Blatt mit der **Aufgabe 11** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** geht **einen** schnellen **Weg** in das **Haus**.  
Setze die richtigen Pfeile (*auf die Lösung zeigen*) ein.  
Es können auch Kreise leer bleiben (*zeigen*).  
Nutze: nach unten, nach oben, nach rechts, nach links (*zeigen*).  
Du kannst die Pfeile mehrmals benutzen.  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 12** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** geht **einen** schnellen **Weg** in das **Haus**.  
Setze die richtigen Pfeile (*auf die Lösung zeigen*) ein.  
Es können auch Kreise leer bleiben (*zeigen*).  
Nutze: nach unten, nach oben, nach rechts, nach links (*zeigen*).  
Du kannst die Pfeile mehrmals benutzen.  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 13** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** geht **einen** schnellen **Weg** in das **Haus**.  
Die **Figur** kann nur über die **gestrichelten Linien** gehen.  
Setze die richtigen Pfeile **hier** (*auf die Lösung zeigen*) ein.  
Es können auch Kreise leer bleiben (*zeigen*).  
Nutze: nach unten, nach oben, nach rechts, nach links (*zeigen*).  
Du kannst die Pfeile mehrmals benutzen.  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 14** vor dir liegen.  
**Achtung: Ich lese jetzt vor. Höre gut zu.**  
Die **Figur** geht einen schnellen Weg in das **Haus**.  
Auf dem Weg hebt die **Figur** alle **Äpfel** (*auf den Befehl zeigen*) auf.  
**Achtung:** Dafür gibt es einen neuen Befehl.  
Setze die richtigen Befehle (*auf die Lösung zeigen*) ein.  
Es können auch Kreise leer bleiben (*zeigen*).  
Nutze: nach unten, nach oben, nach rechts, nach links, aufheben (*zeigen*).  
Du kannst die Befehle mehrmals benutzen.  
Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Alternative A

**Fragebogen - AL - A**

Alter: \_\_\_\_\_

Geschlecht:  Mädchen  
 Junge

0

**Beispielaufgabe**

Kreuze an.

1

**Aufgabe 1**

Kreuze an.

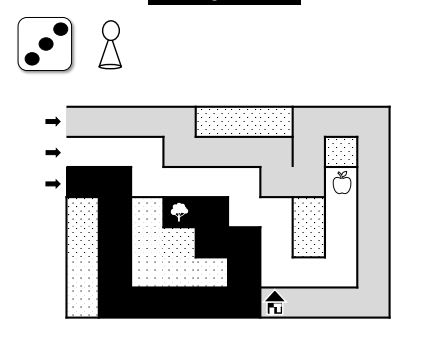
2

**Aufgabe 2**


Kreuze an.


3


**Aufgabe 4**




Maze diagram for Aufgabe 4. A person icon is at the top left. A key icon is at the top left, containing three black dots. The maze has a start area with three arrows pointing right. A path leads to a house icon at the bottom center. A tree icon is in the middle, and an apple icon is at the top right. A pencil icon is at the bottom right.

Kreuze an.  

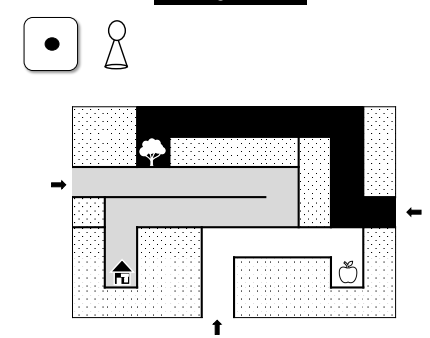








5


**Aufgabe 5**




Maze diagram for Aufgabe 5. A person icon is at the top right. A key icon is at the top right, containing one black dot. The maze has a start area with one arrow pointing right. A path leads to a house icon at the bottom left. A tree icon is at the top left, and an apple icon is at the bottom right. A pencil icon is at the bottom right.

Kreuze an.  

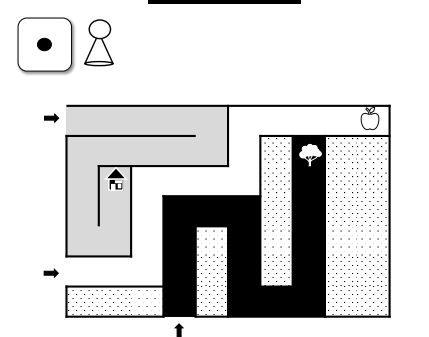








6


**Aufgabe 6**




Maze diagram for Aufgabe 6. A person icon is at the top right. A key icon is at the top right, containing one black dot. The maze has a start area with two arrows pointing right. A path leads to a house icon at the top left. A tree icon is at the top right, and an apple icon is at the top right. A pencil icon is at the bottom right.

Kreuze an.  







7

**Fragebogen - AL - 8**

Alter: \_\_\_\_\_

Geschlecht:  Mädchen  
 Junge

0

**Beispielaufgabe**

Kreuze an.

1

**Aufgabe 1**

Kreuze an.

2

**Aufgabe 2**

Kreuze an.

3



## Testinstruktion Alternative

Aufgabenteil

Du hast nun das Blatt mit der **Beispielaufgabe** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Wenn eine 3 gewürfelt wurde (*zeigen*),  
dann geht die Figur den schwarzen Weg,  
sonst geht die Figur den weißen Weg.

Wo kommt die Figur an?

Kreuze die richtige Lösung an (*auf die Lösungsmöglichkeiten zeigen*).

**Hier, wo du den Stift siehst (*zeigen*)!**

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 1** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Wenn eine 1 gewürfelt wurde,  
dann geht die Figur den grauen Weg,  
sonst geht die Figur den weißen Weg.

Wo kommt die Figur an?

Kreuze die richtige Lösung an.

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 2** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Wenn eine 1 gewürfelt wurde,  
dann geht die Figur den grauen Weg,  
sonst geht die Figur den schwarzen Weg.

Wo kommt die Figur an?

Kreuze die richtige Lösung an.

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 4** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Wenn eine 3 gewürfelt wurde,  
dann geht die Figur den grauen Weg,  
sonst geht die Figur den schwarzen Weg.

Wo kommt die Figur an?

Kreuze die richtige Lösung an.

**2. Mal lesen.**

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 5** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Wenn eine 3 gewürfelt wurde,  
dann geht die Figur den grauen Weg,  
sonst geht die Figur den schwarzen Weg.

Wo kommt die Figur an?

Kreuze die richtige Lösung an (*auf die Lösungsmöglichkeiten zeigen*).

**2. Mal lesen.**

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 6** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Wenn eine 1 gewürfelt wurde,  
dann geht die Figur den weißen Weg,  
sonst geht die Figur den grauen Weg.

Wo kommt die Figur an?

**2. Mal lesen.**

Kreuze die richtige Lösung an.

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.  
Blättere nun um.

Schleife A

**Fragebogen - SCH - A**

Alter: \_\_\_\_\_

Geschlecht:  Mädchen  
 Junge

**Beispielaufgabe**

	♀		🍏	
--	---	--	---	--

Kreuze an.

<input type="checkbox"/>		♀	🍏	
<input type="checkbox"/>		♂		
<input type="checkbox"/>		🍏	♂	

**Aufgabe 1**

🍏		🍏	♂
---	--	---	---

Kreuze an.

<input type="checkbox"/>	🍏	♂	🍏	🍏
<input type="checkbox"/>	🍏		♂	🍏
<input type="checkbox"/>	♂		🍏	🍏

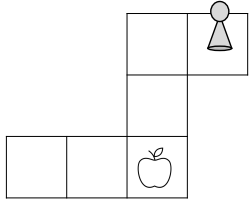
**Aufgabe 2**


♂		
		🍏
		🌸

Kreuze an.

<input type="checkbox"/>	
<input type="checkbox"/>	🍏
<input type="checkbox"/>	🌸

**Aufgabe 3**



Kreuze an. 

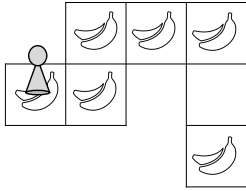
2 mal


4 mal

3 mal

4

**Aufgabe 4**



Kreuze an. 


6 mal

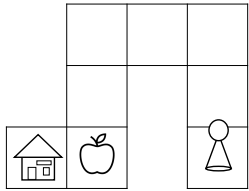
5 mal

4 mal

5

**Aufgabe 6**

Male richtig an. 

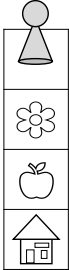



7


**Beispielaufgabe**


2


↓



Kreuze an. 







9

**Aufgabe 10**

? →

Kreuze an.

3

4

5

12

**Aufgabe 11**

Setze richtig ein.

4 5 6 7 und ↓ ↑ ← →

13

**Aufgabe 12**

Setze richtig ein.

1 2 4 6 und ↓ ↑ ← →

1.

2.

14

**Aufgabe 13**

Setze richtig ein.

0 1 2 3 4 5 6 7 8 9 10 und ↓ ↑ ← →


1.

2.

3.

15

Schleife B

**Fragebogen - SCH - B**

Alter: \_\_\_\_\_

Geschlecht:  Mädchen  
 Junge

**Beispielaufgabe**

	👤		🍏	
--	---	--	---	--

Kreuze an.

<input type="checkbox"/>		👤	🍏	
<input type="checkbox"/>		👤		
<input type="checkbox"/>		🍏	👤	

**Aufgabe 1**

🍏		🍏	👤
---	--	---	---

Kreuze an.

<input type="checkbox"/>	👤		🍏	🍏
<input type="checkbox"/>	🍏		👤	🍏
<input type="checkbox"/>	🍏	👤	🍏	🍏

**Aufgabe 2**

👤		
		🌸
		🍏

Kreuze an.

<input type="checkbox"/>	🌸
<input type="checkbox"/>	
<input type="checkbox"/>	🍏

**Aufgabe 3**

Kreuze an.

4 mal

2 mal

3 mal

4

**Aufgabe 4**

Kreuze an.

4 mal

6 mal

5 mal

5

**Aufgabe 5**

Male richtig an.

6

**Aufgabe 6**

Male richtig an.

7

**Beispielaufgabe**

2 ↓

Kreuze an.

9

**Aufgabe 11**

Setze richtig ein.

4 5 6 7 und ↓ ↑ ← →

13

**Aufgabe 12**

Setze richtig ein.

1 2 4 6 und ↓ ↑ ← →

1.

2.

14

**Aufgabe 13**

Setze richtig ein.

0 1 2 3 4 5 6 7 8 9 10 und ↓ ↑ ← →


1.

2.

3.

15

**Aufgabenteil**

Du hast nun das Blatt mit der **Beispielaufgabe** vor dir liegen.

**Achtung: Ich lese jetzt laut vor. Höre gut zu.**

Wenn die **Figur** auf einem **leeren** Feld steht, dann geht die **Figur** zum **nächsten** Feld (*im Spielplan zeigen*). Wiederhole, bis die Figur stoppt.

Auf welchem Feld kommt die Figur an?

Kreuze die richtige Lösung an (*auf die Lösungsmöglichkeiten zeigen*).

**Hier**, wo du den Stift siehst (*zeigen*)!

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

– Ja, genau. Die 2. Lösung, wo die Figur auf dem Apfel steht, ist richtig.

Hast du noch eine Frage?

Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 1** vor dir liegen.

**Achtung: Ich lese jetzt laut vor. Höre gut zu.**

Wenn die **Figur** auf einem **Apfel-** Feld steht,

dann geht die **Figur** zum **nächsten** Feld.

Wiederhole, bis die Figur stoppt.

Auf welchem Feld kommt die Figur an?

Kreuze die richtige Lösung an.

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 2** vor dir liegen.

**Achtung: Ich lese jetzt laut vor. Höre gut zu.**

Wenn die **Figur** auf einem **leeren** Feld steht,

dann geht die **Figur** zum **nächsten** Feld.

Wiederhole, bis die Figur stoppt.

Auf welchem Feld kommt die Figur an?

Kreuze die richtige Lösung an.

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 3** vor dir liegen.

**Achtung: Ich lese jetzt laut vor. Höre gut zu.**

Wenn die **Figur** auf einem **leeren** Feld steht,

dann geht die **Figur** zum **nächsten** Feld.

Wiederhole, bis die Figur auf dem Apfel-Feld steht.

Wie oft ist die Figur gegangen?

Kreuze die richtige Lösung an.

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 4** vor dir liegen.

**Achtung: Ich lese jetzt laut vor. Höre gut zu.**

Wenn die **Figur** auf einem **Bananen-** Feld steht,

dann geht die **Figur** zum **nächsten** Feld.

Wiederhole, bis die Figur auf dem leeren Feld steht.

Wie oft ist die Figur gegangen?

Kreuze die richtige Lösung an.

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 5** vor dir liegen.

**Achtung: Ich lese jetzt laut vor. Höre gut zu.**

Wenn die **Figur** auf einem **grauen** Feld steht,

dann geht die **Figur** zum **nächsten** Feld.

Wiederhole, bis die Figur im Haus ist.

**Achtung:** Male dafür die passenden Felder mit Bleistift an. Schraffieren reicht (*an der Tafel zeigen*).

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 6** vor dir liegen.

**Achtung: Ich lese jetzt laut vor. Höre gut zu.**

Wenn die **Figur** auf einem **grauen** Feld steht,

dann geht die **Figur** zum **nächsten** Feld.

Wiederhole, bis die Figur im Haus ist.

**Achtung:** Male dafür **wieder** die passenden Felder mit Bleistift an.

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Blättere nun um.

Du hast nun das Blatt mit der **Beispielaufgabe** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Die **Figur** geht **2-mal** nach unten (*auf die Karte zeigen*).

Auf welchem Bild kommt die Figur an?

Kreuze die richtige Lösung an. Hier, wo du den Stift siehst (*zeigen*).

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

– Ja, genau. Die 1. Lösung, wo die Figur auf dem Apfel steht, ist richtig.

Hast du noch eine Frage?

Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 10** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Die **Figur** geht in das Haus.

Wie oft ist die Figur gegangen?

Kreuze die richtige Lösung an.

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Blättere nun um.

Du hast es fast geschafft

Du hast nun das Blatt mit der **Aufgabe 11** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Die **Figur** geht in das Haus.

Setze dafür **hier unten** (*zeigen*) richtig ein.

Nutze: Eine Zahl 4, 5, 6 oder 7 **und** einen Pfeil nach unten, nach oben, nach links oder nach rechts.

Die Zahl kommt auf die Linie und der Pfeil kommt in den Kreis (*zeigen*).

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 12** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Die **Figur** geht auf das **Blumen-Feld** (*auf die Karte zeigen*).

Setze dafür **hier unten** (*zeigen*) richtig ein.

Nutze: Die Zahlen 1, 2, 4 oder 6 **und** die Pfeile nach unten, nach oben, nach links oder nach rechts.

Die Zahl kommt auf die Linie und der Pfeil kommt in den Kreis (*zeigen*).

Es gibt 2 Lösungsfelder zum Ausfüllen (*zeigen*).

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Blättere nun um.

Du hast nun das Blatt mit der **Aufgabe 13** vor dir liegen.

**Achtung: Ich lese jetzt vor. Höre gut zu.**

Die **Figur** sammelt **alle** Äpfel ein.

**Zum Schluss** geht die Figur in das Haus.

Setze dafür **hier unten** (*zeigen*) richtig ein.

Nutze: Die Zahlen 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, oder 10 **und die Pfeile** nach unten, nach oben, nach links oder nach rechts.

Die Zahl kommt auf die Linie und der Pfeil kommt in den Kreis (*zeigen*).

Es gibt 3 Lösungsfelder zum Ausfüllen (*zeigen*).

Wenn du fertig bist, lege deinen Stift hin und nimm beide Hände hoch.

Schließe nun dein Rätselheft.

Super, du kannst stolz auf dich sein. Du hast es geschafft. Schließe nun das Rätselheft. Du hast es geschafft. Vielen Dank für deine Hilfe (*Hefte einsammeln*)!

Lösungsschablone Beispiele

Lösungsschablonen

**B** Lösungsschablone - Sequenz

Aufgabe 11:

→	→	↑	↑
↑	↑	→	→
→	↑	→	↑
↑	→	↑	→
↑	→	→	↑
→	↑	↑	→

Aufgabe 12:

←	←	↑
↑	←	←
←	↑	←

Aufgabe 13:

↓	→	↑	↑	↑	←	←	↓	←
---	---	---	---	---	---	---	---	---

**A** Lösungsschablone - Schleife

Beispielaufgabe: 2. Lösung 

		♁	
--	--	---	--

Aufgabe 1: 

🍏	♁	🍏	🍏
---	---	---	---

Aufgabe 2:

Aufgabe 3: **3 mal**

Aufgabe 4: **5 mal**

Aufgabe 5: 

🏠			♁
---	--	--	---

Aufgabe 6: 

🏠	🍏			♁
---	---	--	--	---

**A** Lösungsschablone - Alternative

Beispielaufgabe:

Aufgabe 1:

Aufgabe 2:

Aufgabe 4:

Aufgabe 5:

Aufgabe 6:

## E Berechnungen der Rasch-Analyse

### Eindimensionalität

#### Sequenz A

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
1	31	70	.26	.31	1.23	1.64	1.29	1.37	.55	.63	61.8	72.3	Aufgabe 7
2	40	70	-.62	.32	.95	-.34	.84	-.61	.65	.63	67.3	71.8	Aufgabe 9
3	41	70	-.72	.32	1.11	.85	1.29	1.12	.58	.63	72.7	72.6	Aufgabe 10
4	35	70	-.13	.31	.90	-.73	.89	-.51	.66	.63	80.0	71.2	Aufgabe 11
5	43	70	-.93	.32	.85	-1.04	.77	-.76	.68	.63	76.4	74.2	Aufgabe 12
6	40	70	-.62	.32	1.15	1.14	1.55	2.01	.56	.63	70.9	71.8	Aufgabe 13
7	21	70	1.32	.35	.79	-1.16	.67	-1.05	.69	.62	83.6	78.9	Aufgabe 14
8	20	70	1.45	.35	.90	-.51	.75	-.69	.65	.62	78.2	79.9	Aufgabe 15
MEAN	33.9	70.0	.00	.33	.99	.0	1.01	.1			73.9	74.1	
P.SD	8.5	.0	.87	.02	.15	1.0	.30	1.1			6.7	3.2	

Anmerkung. Diagrammerstellung mittels MINISTEPS®. MNSQ und Point-Measure-Correlation

#### Sequenz B

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
1	33	64	-.26	.34	1.20	1.18	1.22	.97	.60	.66	72.5	75.9	Aufgabe 7
2	27	64	.45	.35	1.15	.82	1.24	.88	.64	.69	76.5	78.3	Aufgabe 9
3	43	64	-1.40	.34	1.11	.76	1.13	.49	.56	.59	72.5	73.9	Aufgabe 10
4	35	64	-.49	.34	1.00	.03	.96	-.10	.65	.65	72.5	74.6	Aufgabe 11
5	42	64	-1.28	.34	.77	-1.66	.60	-1.40	.68	.60	84.3	73.2	Aufgabe 12
6	26	64	.58	.36	1.03	.22	.87	-.35	.69	.69	78.4	78.4	Aufgabe 13
7	23	64	.98	.37	.85	-.77	1.02	.17	.73	.69	84.3	79.4	Aufgabe 14
8	20	64	1.41	.39	.88	-.55	.67	-.65	.73	.70	80.4	80.7	Aufgabe 15
MEAN	31.1	64.0	.00	.35	1.00	.0	.96	.0			77.7	76.8	
P.SD	8.0	.0	.96	.02	.14	.9	.22	.8			4.7	2.6	

Anmerkung. Diagrammerstellung mittels MINISTEPS®. MNSQ und Point-Measure-Correlation

#### Alternative A

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
1	46	51	-1.78	.82	.99	.15	.88	.17	.74	.74	86.7	86.4	Aufgabe 1
2	50	60	.22	.56	1.39	1.67	1.43	1.43	.71	.78	56.3	69.0	Aufgabe 2
3	48	60	.82	.55	.66	-1.97	.59	-1.42	.86	.80	87.5	67.8	Aufgabe 4
4	49	60	.52	.55	.99	.01	.94	-.13	.79	.79	68.8	65.5	Aufgabe 5
5	50	60	.22	.56	.95	-.16	1.01	.14	.79	.78	68.8	69.0	Aufgabe 6
MEAN	48.6	58.2	.00	.61	1.00	-.1	.97	.0			73.6	71.5	
P.SD	1.5	3.6	.92	.11	.23	1.2	.27	.9			11.9	7.6	

Anmerkung. Diagrammerstellung mittels MINISTEPS®. MNSQ und Point-Measure-Correlation

## Alternative

B

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
1	43	54	.38	.47	1.15	.78	1.25	1.13	.64	.70	63.6	70.0	Aufgabe 1
2	50	65	.35	.41	1.45	2.55	1.43	2.23	.54	.69	37.9	68.8	Aufgabe 2
3	50	65	.35	.41	.68	-2.26	.66	-2.19	.79	.69	86.2	68.8	Aufgabe 4
4	56	65	-.74	.46	1.00	.06	1.09	.38	.60	.61	79.3	77.0	Aufgabe 5
5	54	65	-.34	.43	.72	-1.54	.68	-1.27	.74	.64	89.7	71.6	Aufgabe 6
MEAN	50.6	62.8	.00	.44	1.00	-.1	1.02	.1			71.3	71.2	
P.SD	4.5	4.4	.46	.03	.29	1.7	.31	1.6			18.9	3.0	

Anmerkung. Diagrammerstellung mittels MINISTEPS®. MNSQ und Point-Measure-Correlation

## Iteration A

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
1	29	61	1.38	.32	1.28	1.96	1.48	1.58	.47	.60	65.4	72.1	Aufgabe 1
2	38	61	.44	.33	1.24	1.52	1.44	1.88	.46	.58	65.4	73.2	Aufgabe 2
3	48	61	-.75	.37	.78	-1.09	.56	-1.39	.63	.52	82.7	81.0	Aufgabe 3
4	43	61	-.12	.34	.90	-.52	.75	-.99	.62	.56	76.9	76.7	Aufgabe 4
5	52	61	-1.38	.42	.93	-.23	.65	-.64	.51	.46	84.6	84.9	Aufgabe 10
6	48	61	-.75	.37	.86	-.66	1.18	.61	.55	.52	90.4	81.0	Aufgabe 11
7	34	61	.86	.32	.88	-.83	.78	-1.01	.66	.59	73.1	72.1	Aufgabe 12
8	39	61	.33	.33	1.05	.36	.96	-.12	.57	.58	67.3	73.9	Aufgabe 13
MEAN	41.4	61.0	.00	.35	.99	.1	.97	.0			75.7	76.8	
P.SD	7.3	.0	.87	.03	.17	1.1	.33	1.2			8.9	4.6	

Anmerkung. Diagrammerstellung mittels MINISTEPS®. MNSQ und Point-Measure-Correlation

## Iteration B

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
1	27	58	1.43	.34	1.08	.55	1.13	.49	.58	.62	70.0	73.7	Aufgabe 1
2	47	58	-.97	.40	.78	-.99	.66	-.72	.58	.49	88.0	82.5	Aufgabe 3
3	49	58	-1.31	.42	.83	-.63	.56	-.82	.54	.46	88.0	85.2	Aufgabe 4
4	44	58	-.54	.37	1.19	1.03	1.06	.28	.45	.52	72.0	79.0	Aufgabe 5
5	20	58	2.29	.37	1.09	.52	.99	.14	.60	.63	76.0	78.4	Aufgabe 6
6	49	58	-1.31	.42	1.10	.45	.74	-.35	.45	.46	80.0	85.2	Aufgabe 11
7	40	58	-.03	.35	1.19	1.18	1.44	1.49	.45	.56	70.0	75.1	Aufgabe 12
8	36	58	.43	.34	.77	-1.68	.62	-1.73	.69	.58	78.0	73.0	Aufgabe 13
MEAN	39.0	58.0	.00	.38	1.00	.1	.90	-.2			77.8	79.0	
P.SD	10.0	.0	1.23	.03	.17	1.0	.29	.9			6.8	4.6	

Anmerkung. Diagrammerstellung mittels MINISTEPS®. MNSQ und Point-Measure-Correlation

Testheft	Unerklärte Varianz im ersten Kontrast
Sequenz A	2.02
Sequenz B	1.87
Alternative A	2.23
Alternative B	2.00
Iteration A	1.86
Iteration B	2.09

Principal Component Analysis of Residuals (PCAR) Berechnungen