

Paul Dörbecker

Matrikelnummer: 

paul.doerrbecker@student.uni-tuebingen.de

**Optimierung des Mathematiktests aus TIMSS 2003 mit Methoden
der kompetenzbasierten Wissensstrukturtheorie**

Masterthesis

Fachbereich: Forschungsmethoden und Mathematische Psychologie

Betreuung: Prof. Dr. Jürgen Heller

Zweitkorrektur: Dr. Florian Wickelmaier

Sommersemester 2024

Eberhard Karls Universität Tübingen

01.08.2024

Inhaltsverzeichnis

1. Einleitung	5
2. Theoretischer Hintergrund	15
2.1 Klassische Wissensstrukturtheorie	15
2.2 Kompetenzbasierte Wissensstrukturtheorie	18
2.3 Wissensstruktur aus Kompetenzstruktur ableiten	20
2.4 Testoptimierung	22
2.4.1 Voraussetzungen und Vorannahmen für Shortening und Improvement	22
2.4.2 Shortening	25
2.4.3 Improvement	27
2.4.4 Gesamter Optimierungsprozess auf Kompetenz-Ebene	28
2.4.5 Gesamter Optimierungsprozess auf Item-Ebene	30
3. Methode	31
3.1 Software	31
3.2 Stichprobe und Grundlagen	32
3.3 Material, Ablauf und Design	32
3.4 Optimierungsprozess: Kompetenz-Ebene	33
3.5 Optimierungsprozess: Item-Ebene	35
3.6 Analyse der optimierten Tests	37
4. Ergebnisse	39
4.1 Optimierungsergebnisse: Kompetenz-Ebene	39
4.2 Optimierungsergebnisse: Item-Ebene	42
4.3 Optimierungsergebnisse: Vergleich der optimierten Booklets	45
4.4 Optimierungsergebnisse: Kompetenz-/ Wissensstruktur-Ebene	46
4.5 Optimierungsergebnisse: Fully informative Verbesserung	48
5. Diskussion	48
5.1 Interpretation der Ergebnisse	49
5.2 Vergleich der optimierten Booklets	53
5.3 Der Weg zum fully informative Test	54
5.4 Limitationen	59
5.5 Verdienst dieser Arbeit	62
6. Literatur	64
7. Appendix	69

Zusammenfassung

In dieser Arbeit werden aus vier Itemsammlungen, die dasselbe erfassen sollen („Booklets“), des Mathematiktests der TIMSS 2003 die Items betrachtet, welche die Fähigkeiten der Testpersonen in den Bereichen „Algebra“ und „Number and Operation“ erfassen sollten. Jedes dieser Booklets wird zweimal anhand der von Anselmi und Kollegen (2022) entwickelten Testoptimierungsmethoden des „Shortenings“ und des „Improvements“ aus der kompetenzbasierten Wissensstrukturtheorie optimiert. Hierbei wurden jeweils andere Grundannahmen vorausgesetzt. Die hierfür nötigen theoretischen Grundannahmen sind einem Paper von Su und Kollegen (2013) entnommen. Nach der Optimierung kann jedes Booklet mehr unterschiedliche potentielle Wissenszustände von Testpersonen erfassen als zuvor. Außerdem benötigen die meisten Booklets durch die Optimierung auch weniger Items trotz dieser nun präziseren Wissensdiagnose. Die Booklets zeichnen sich nach der Optimierung also durch ihre Befähigung zu einer möglichst ökonomischen und zugleich möglichst präzisen Wissensdiagnose aus. Ein Vergleich der optimierten Booklets zeigt, dass sich die Booklets durch den Optimierungsprozess sehr ähnlich geworden sind, somit ist die Annahme von unterschiedlichen Itemsammlungen, welche dasselbe messen sollen, für die optimierten Booklets fraglich. Außerdem wird gezeigt wie man die optimierten Booklets zu „fully informative“ Tests aufwerten könnte, was eine, gemäß der theoretischen Grundannahmen, maximal präzise Wissensdiagnose bedeutet.

Optimierung des Mathematiktests aus TIMSS 2003 mit Methoden der kompetenzbasierten Wissensstrukturtheorie

Die Psychologie hat das Ziel den Verstand und die Fähigkeiten von Menschen zu analysieren, zu kategorisieren und zu verstehen. Um dies zu erreichen ist es nötig möglichst präzise Informationen über die untersuchten Personen zu erlangen. Außerdem sollen diese Informationen in einer möglichst kurzen Zeit generiert werden, um zeitliche, kognitive, sowie finanzielle Ressourcen zu sparen. Klassisch werden, um Informationen über die Fähigkeiten einer Person zu erhalten, Tests eingesetzt. Je nach untersuchter Fähigkeit kommen unterschiedliche Tests zum Einsatz. Gerade im Schulkontext ist es oft sinnvoll Tests anzuwenden um die Fähigkeiten von Schülern bezüglich der Fächer, in welchen sie unterrichtet werden, zu erfassen. Ein Test, der in eben einem solchen Kontext zum Einsatz kommt, ist die „Trends in International Mathematics and Science Study“ (TIMSS) (Arora et al., 2005). Die TIMSS wird seit 1995 einmal in vier Jahren durchgeführt und soll einen internationalen Vergleich von Schulperformanz ermöglichen. Genauer wird die TIMSS bei Kindern in der vierten und in der achten Schulklasse durchgeführt und untersucht deren Fähigkeiten in den Bereichen Mathematik und Naturwissenschaften. Die Durchführung erfolgt in unterschiedlichen Ländern weltweit. Zu diesen zählen beispielsweise die USA, England und Japan (Martin & Mullis, 2003). Nach der Durchführung der Studie und der Analyse der Ergebnisse werden die Ergebnisse, wie auch die Rohdaten und die Hälfte der verwendeten Items veröffentlicht. Dies ermöglicht zum einen den Vergleich zwischen Schulleistungen unterschiedlicher Länder, stellt zum anderen aber auch viele Möglichkeiten zur Reanalyse dar. Schließlich sind Rohdaten mit vielen Versuchspersonen, sowie die Hälfte der zur Generation dieser Daten verwendeten Items, frei verfügbar. Eben dies macht sich auch diese Arbeit zunutze, in welcher Teile des Mathematiktests aus TIMSS 2003 analysiert und mit Methoden der kompetenzbasierten Wissensstrukturtheorie optimiert werden.

Diese Arbeit baut unter anderem auf einem Paper von Su und Kollegen aus dem Jahr 2013 auf. In diesem Paper wurden Teile der TIMSS 2003 bereits auf eine Weise analysiert, welche die Voraussetzung für die in dieser Arbeit angewandten Methoden der kompetenzbasierten Wissensstrukturtheorie mit einschließt. Auf das genaue Vorgehen hierbei soll später noch genauer eingegangen werden. An dieser Stelle soll zunächst die TIMSS 2003 und vor allem der Mathematikanteil der TIMSS 2003 einmal genauer vorgestellt werden. Die

Items der TIMSS 2003 (Mullis et al., 2004) sind aufgeteilt in 12 Booklets, von welchen jeweils eines einem Schüler/ einer Schülerin bei der Untersuchung vorgelegt wurde. Dabei setzt sich jedes der Booklets aus sechs Sammlungen von Items zusammen. Von diesen Itemsammlungen gibt es insgesamt 28, von denen 14 sich auf naturwissenschaftliche (S1-S14) und 14 sich auf den mathematische (M1-M14) Themenbereiche beziehen. Jedes der Booklets setzt sich entweder aus vier Itemsammlungen aus dem mathematischen und zwei aus dem naturwissenschaftlichen Themenbereich zusammen oder umgekehrt. Die genaue Aufteilung kann überblicksweise Abbildung 1 (Exhibit 2.3 aus Martin et al., 2005) entnommen werden.

Abbildung 1

Zusammensetzung der 12 Booklets aus der TIMSS 2003 (Exhibit 2.3 entnommen aus Martin et al., 2005)

Exhibit 2.3 Booklet Design for TIMSS 2003 – Grade 4 and Grade 8

Student Booklet	Assessment Blocks					
	Part I			Part II		
Booklet 1	M01	M02	S06	S07	M05	M07
Booklet 2	M02	M03	S05	S08	M06	M08
Booklet 3	M03	M04	S04	S09	M13	M11
Booklet 4	M04	M05	S03	S10	M14	M12
Booklet 5	M05	M06	S02	S11	M09	M13
Booklet 6	M06	M01	S01	S12	M10	M14
Booklet 7	S01	S02	M06	M07	S05	S07
Booklet 8	S02	S03	M05	M08	S06	S08
Booklet 9	S03	S04	M04	M09	S13	S11
Booklet 10	S04	S05	M03	M10	S14	S12
Booklet 11	S05	S06	M02	M11	S09	S13
Booklet 12	S06	S01	M01	M12	S10	S14

Die Items des Mathematikteils der TIMSS 2003 sind gemäß einem Framework von Mullis und Kollegen (2003) unterteilt in fünf Domänen einer Inhalts- und vier Domänen einer kognitiven Dimension. Die fünf Inhaltsdimensionen sind „Number and Operation“, „Algebra“, „Measurement“, „Geometry“ und „Data“ und die vier kognitiven Dimensionen sind „Knowing Facts and Procedures“, „Using Concepts“, „Solving Routine Problems“, und

„Reasoning“. Hierbei ist jedem Item stets sowohl eine Inhalts- als auch eine kognitiven Dimension zugeordnet.

Su und Kollegen haben 2013 die Items der Inhaltsdimensionen „Number and Operation“ (im Folgenden: „Number“) und „Algebra“ für Schüler der achten Schulklasse aus den Booklets 1 und 2 im Rahmen von „Cognitive diagnostic models“ (CDMs; Bolt, 2007; Junker & Sijtsma, 2001; Tatsuoka, 1990, 2002; de la Torre, 2009; Williamson, 2023) analysiert. CDMs sind „latent class“-Modelle. Als solche nehmen sie nicht an, dass man nur einen Faktor untersucht und mittels einer Untersuchung überprüft wie gut die Versuchsperson in diesem Faktor ist (wie es bei „latent trait“-Modellen der Fall wäre). Stattdessen wird das Untersuchte so unterteilt, dass angenommen wird, dass für die Lösung jedes Items bestimmte Fähigkeiten (genannt „Attributes“) benötigt werden. Eine Person, die ein bestimmtes Item löst, verfügt demnach tendenziell über eben die Attributes, die für die Lösung des Items nötig sind. Somit erhält man am Ende einer Untersuchung nicht einen Wert zwischen beispielsweise 1 und 100, der die Fähigkeit einer Person im Bereich Mathematik angeben soll. Stattdessen erhält man eine Liste von Attributes, über welche die Person verfügt. Umgekehrt erhält man außerdem eine Liste der übrigen Attributes, zu welchen es Items in der Untersuchung gab, über welche die Person aber nicht verfügt (Williamson, 2023). Außerdem arbeiten CDMs mit Annahmen bezüglich Voraussetzungsbeziehungen zwischen den untersuchten Attributes. Man nimmt hierbei an, dass einzelne Attributes nur dann vorliegen können, wenn ein anderes Attribute, welches deren direkte Voraussetzung ist, auch vorliegt. So kann man vom Vorliegen der Attributes, welche andere Attributes als Voraussetzung haben, darauf schließen, dass auch mindestens eines der Attributes vorliegen muss, welches eine direkte Voraussetzung für das Attribute darstellt. An dieser Stelle soll nicht weiter auf CDMs eingegangen werden, da diese nicht das Thema dieser Arbeit sein werden. Die kurze Einführung, sowie auch das Thema CDMs ist hier nur deshalb relevant, weil gezeigt wurde, dass CDMs sehr eng mit der kompetenzbasierten Wissensstrukturtheorie verwandt sind (Heller, 2022; Heller et al., 2015 & 2016). Auf kompetenzbasierte Wissensstrukturen soll später im Abschnitt „2. Theoretischer Hintergrund“ nochmal genauer eingegangen werden. An dieser Stelle sei nur gesagt, dass kompetenzbasierte Wissensstrukturen und CDMs beide einige Voraussetzungen benötigen um angewandt werden zu können. Also um entweder das CDM zu erstellen (im Falle von CDMs) oder eine klassische Wissensstruktur abzuleiten (im Falle von kompetenzbasierten Wissensstrukturen). Da Su und Kollegen (2013) mit CDMs gearbeitet haben, haben sie eben diese Voraussetzungen bereits entwickelt. Hier soll nun

zunächst weiterhin die Terminologie der CDMs verwendet werden, die Überführung in die Terminologie der kompetenzbasierten Wissensstrukturtheorie soll später im Abschnitt „2. Theoretischer Hintergrund“ erfolgen.

Benötigt wird zunächst eine Menge an Attributes, sodass inhaltlich definiert ist welche Fähigkeiten im Test untersucht werden. Diese Attributes können Abbildung 2 (Table 2 (leicht modifiziert) aus Su et al., 2013) entnommen werden.

Abbildung 2

Die in dieser Arbeit betrachteten Attributes und deren Benennung (leicht modifizierte Version von Table 2 aus Su et al., 2013)

this work	Su et al.
	Attribute
a	1. Understand concepts of a ratio and a unit rate and use language appropriately.
b	2. Use ratio and rate reasoning to solve real-world and mathematical problems
c	3. Compute fluently with multi-digit numbers and find common factors and multiples.
d	4. Apply and extend previous understandings of numbers to the system of rational numbers.
e	5. Apply and extend previous understandings of arithmetic to algebraic expressions.
f	6. Reason about and solve one-variable equations and inequalities.
g	7. Recognize and represent proportional relationships between quantities.
h	8. Use proportional relationships to solve multi-step ratio and percent problems.
i	9. Apply and extend previous understandings of operations with fractions to add, subtract, multiply, and divide rational numbers.
j	10. Solve real-world and mathematical problems involving the four operations with rational numbers.
k	11. Solve real-life and mathematical problems using numerical and algebraic expressions and equations.
l	12. Know and apply the properties of integer exponents to generate equivalent numerical expressions.
m	13. Compare two fractions with different numerators and different denominators; Understand a fraction a/b with $a > 1$ as a sum of fractions $1/b$.
n	14. Solve multi-step word problems posed with whole numbers and having whole-number answers using the four operations, including problems in which remainders must be interpreted. Represent these problems using equations with a letter standing for the unknown quantity; Generate a number or shape pattern that follows a given rule. Identify apparent features of the pattern that were not explicit in the rule itself.
o	15. Use equivalent fraction as a strategy to add and subtract fractions.

Als nächstes ist es nötig alle Items zu betrachten und jedem Item zuzuordnen welche Attributes für seine Lösung benötigt werden. An dieser Stelle sei erwähnt, dass in dieser Arbeit davon ausgegangen wird, dass zur Lösung eines Items stets alle Attributes gleichzeitig vorliegen müssen, die mit ihm assoziiert sind. Diese Annahme entspricht der Annahme, dass es sich bei den untersuchten Booklets um sogenannte „conjunctive“ Tests handelt, was gerade im Bildungsbereich und im Bereich Mathematik eine sinnvolle Annahme ist (Su et al., 2003; Anselmi et al., 2022). Wie genau conjunctive Tests definiert sind, ist im Abschnitt „2. Theoretischer Hintergrund“ beschrieben. Abbildungen 3 und 4 (Table 4 und Table 5 aus Su et al., 2013) zeigen die von Su und Kollegen (2013) vorgenommene und auch in dieser Arbeit verwendete Zuordnung von Items zu den Attributes. Diese Zuordnungstabellen werden im Rahmen der CDMs als „Q-Matrizen“ bezeichnet.

Abbildung 3

Q-Matrix von Booklet 1 (Table 4 entnommen aus Su et al., 2013)

Item\Attribute	1	2	3	4	5	6	7	8	9	11	12	13	14	15	Sum
1	M012001	1	0	0	0	0	0	0	0	0	0	1	0	1	3
2	M012002	0	0	0	0	0	1	0	0	0	0	0	0	0	1
3	M012004	0	1	0	0	0	0	1	0	0	0	0	0	0	2
4	M012040	0	0	0	0	0	1	1	0	0	0	0	0	0	2
5	M012041	1	0	0	0	0	0	0	0	0	0	0	0	0	1
6	M012042	0	0	0	0	1	0	0	0	1	0	0	0	0	2
7	M032570	0	1	0	0	0	0	0	0	0	0	0	0	0	1
8	M032643	0	0	0	1	0	0	0	0	1	0	0	0	0	2
9	M012016	0	0	0	1	0	0	0	0	1	0	0	0	0	2
10	M012017	0	0	0	0	0	1	0	0	0	1	0	0	1	3
11	M022251	0	0	0	0	1	0	0	0	0	0	0	0	0	1
12	M022185	0	0	0	0	1	0	0	0	0	0	0	0	0	1
13	M022191	0	0	0	0	0	1	0	0	0	0	0	0	0	1
14	M022194	0	1	0	0	0	0	0	0	0	0	0	1	0	2
15	M022196	0	0	0	0	1	0	0	0	0	0	0	0	0	1
16	M022198	0	0	0	1	0	0	0	0	0	0	0	0	0	1
17	M022199	0	0	1	0	0	0	0	0	1	0	0	1	0	4
18	M022043	0	0	0	0	0	0	0	0	0	0	1	0	0	1
19	M022046	0	0	1	0	0	0	0	0	0	0	0	0	0	1
20	M022050	0	0	0	0	0	1	0	0	0	0	0	0	0	1
21	M022057	0	1	0	0	0	0	0	1	0	0	0	0	0	2
22	M022066	0	0	1	0	0	0	0	0	0	0	0	0	1	2
23	M022232	0	1	0	0	0	0	0	0	0	0	0	0	0	1
24	M022234B	1	0	0	0	0	0	0	0	0	0	0	0	0	1
25	M032142	0	1	0	0	0	0	1	0	0	0	0	0	0	2
26	M032198	0	0	0	0	0	0	0	0	0	1	0	0	0	1
27	M032640	0	0	0	0	0	1	0	0	0	1	0	0	1	3
28	M032755	0	1	0	0	0	0	1	1	0	1	0	0	0	4
29	M032163	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Sum		3	7	3	3	5	6	4	2	4	3	1	3	3	3

Abbildung 4

Q-Matrix von Booklet 2 (Table 5 entnommen aus Su et al., 2013)

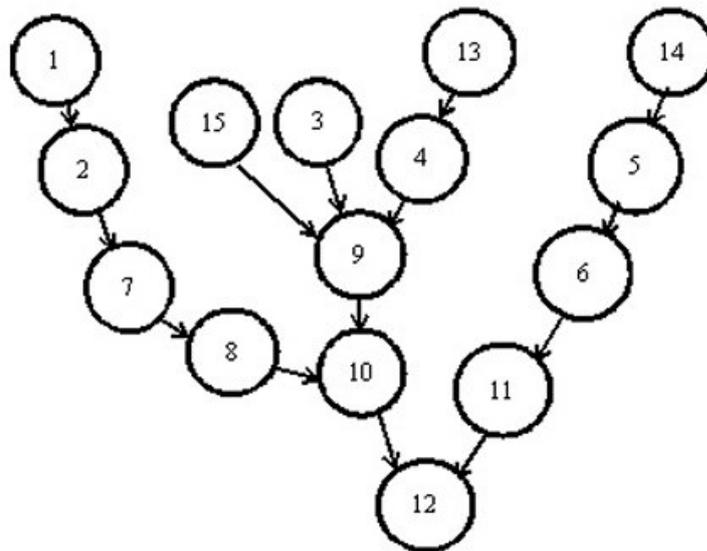
Item\Attribute	1	2	3	4	5	6	7	8	9	10	11	13	14	15	Sum
1 M012016	0	0	0	1	0	0	0	0	1	0	0	0	0	0	2
2 M012017	0	0	0	0	0	1	0	0	0	0	1	0	1	0	3
3 M022251	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
4 M022185	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
5 M022191	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
6 M022194	0	1	0	0	0	0	0	0	0	0	0	0	1	0	2
7 M022196	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
8 M022198	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
9 M022199	0	0	1	0	0	0	0	0	1	0	0	1	0	1	4
10 M012025	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
11 M012027	1	1	0	0	0	0	0	1	0	0	0	0	0	0	3
12 M012029	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
13 M022139	1	1	0	0	0	0	0	1	0	0	0	0	0	0	3
14 M022144	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
15 M022253	0	0	0	0	1	1	0	0	0	0	0	0	0	0	2
16 M022156	0	1	0	0	0	1	0	0	0	0	0	0	0	0	2
17 M022104	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
18 M022106	1	0	0	0	0	1	1	0	0	0	0	0	0	0	3
19 M022110	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
20 M032307	0	0	0	0	0	0	0	0	1	1	0	0	0	1	3
21 M032523	0	0	0	0	0	0	0	0	0	1	0	0	0	1	2
22 M032701	0	1	0	0	0	1	0	0	0	0	0	0	0	0	2
23 M032704	1	1	0	0	0	0	1	0	0	0	0	0	0	0	3
24 M032525	0	0	0	0	0	1	0	0	1	0	0	0	0	0	2
25 M032381	0	0	1	0	0	0	1	0	0	0	0	0	0	0	2
26 M032416	0	0	1	0	0	0	0	0	0	0	0	0	0	1	2
27 M032160	0	1	0	0	0	0	1	0	0	0	0	0	0	0	2
28 M032540	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
29 M032698	0	0	0	0	1	1	0	0	0	0	0	0	0	0	2
30 M032529	0	1	0	0	0	0	0	1	0	0	0	0	0	0	2
Sum	4	8	4	3	5	9	5	3	4	2	1	1	4	4	

Zuletzt ist es noch nötig die Voraussetzungsbeziehungen zwischen den Attributes zu bestimmen. Hierbei wird zu jedem Attribute untersucht, welche anderen Attributes Voraussetzungen für es darstellen. Su und Kollegen (2013) haben diese Voraussetzungsbeziehung in einem Und-Oder-Graph dargestellt, sodass die Voraussetzungsbeziehungen grafisch ersichtlich werden. Dieser Graph ist in Abbildung 5 (Figure 5 aus Su et al., 2013) dargestellt. Der Graph ist so zu lesen, dass ein Attribute am Anfang eines Pfeils stets die direkte Voraussetzung ist für das Attribute, auf welches der Pfeil zeigt. Zudem sind alle Attributes, welche an den Ursprüngen vorhergehender Pfeile stehen bei einer Verkettung, die auf ein Attribute zeigt, auch alle Voraussetzungen für das Attribute am Ende der Kette, also das, auf welches der Pfeil zeigt. Diese Regel gilt für jedes der Kettenglieder. Wenn auf ein Attribute mehrere Pfeile zeigen, dann müssen nicht alle Voraussetzungen von dem Attribute gleichzeitig gegeben sein, damit es vorliegen kann.

Stattdessen genügt es, wenn eine der Voraussetzungen, beziehungsweise eine der Ketten, die dieser Voraussetzung wiederum als Voraussetzung vorausgeht, vorliegt. Diese Ansammlung an Voraussetzungsbeziehung, wie auch ihre grafische Darstellung in Abbildung 5 (Figure 5 aus Su et al., 2013), wird im Rahmen der CDMs als „Attribute Hierarchy“ bezeichnet.

Abbildung 5

Attribute Hierarchy aller betrachteten Attributes dargestellt als Und-Oder-Graph (Figure 5 entnommen aus Su et al., 2013)



Die beschriebenen drei Grundlagen stellen auch für mein Arbeit die Grundlagen dar, weshalb an dieser Stelle noch einmal kurz ihr Ursprung erläutert werden soll. Su und Kollegen (2013) haben die drei Grundlagen in ihrem Paper Folgendermaßen erstellt: Die Auswahl der Attributes erfolgte durch eine Anpassung der Fähigkeiten für Schüler der sechsten bis achten Klasse im Bereich Mathematik vom CCSS 2010 (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). Die Q-Matrizen wurden von zwei Doktoranden aus dem Bereich „secondary teaching and learning“ erstellt, welche beide zuvor Mathelehrer an amerikanischen middle und high schools waren. Die beiden wurden instruiert jedem Item die für seine Lösung nötigen Attributes zuzuordnen. Die Attribute Hierarchy wurde von denselben zwei Doktoranden erstellt. Hierbei wurden sie instruiert für jedes Attribute die Attributes herauszusuchen, welche Voraussetzungen gemäß der Definition des CDMs für es darstellen, insofern es welche gebe. Eine Veränderung im

Vergleich zur Arbeit von Su und Kollegen (2013), die in dieser Arbeit vorgenommen wurde, ist, dass die Attributes, hier statt mit den Zahlen 1-15, mit den Buchstaben a–o benannt sind. Diese Änderung kann auch Abbildung 2 entnommen werden. Sie wurde vorgenommen, da so einige Bezeichnungen eindeutiger werden und weil das Programmieren so erleichtert wurde.

Diese Arbeit befasst sich nicht nur mit den Booklets 1 und 2, welche Su und Kollegen (2013) bereits analysiert haben, sondern zusätzlich auch mit den Booklets 7 und 8 der TIMSS 2003. Da für alle weiteren Untersuchungen stets die oben beschriebenen drei Grundlagen gegeben sein mussten, waren diese zwei Booklets die einzigen beiden, die sich anhand der Arbeit von Su und Kollegen (2013) problemlos hinzu nehmen ließen. Die Items zum Thema Mathematik aus den Booklets 7 und 8 sind nämlich ausschließlich Items, welche bereits entweder in Booklet 1 oder Booklet 2 vorlagen. Folglich konnten für die Items aller genannter Booklets dieselben Attributes, dieselben Q-Matrizen wie auch dieselbe Attribute Hierarchy angenommen werden konnte. Abbildung 6 (modifizierte Version von Exhibit 2.3 entnommen aus Martin et al., 2005) illustriert dies noch einmal. Auf der rechten Seite der Abbildung sind wie auch schon in Abbildung 1 die Itemsammlungen abgebildet, aus welchen sich die jeweiligen Booklets auf der linken Seite der Abbildung zusammensetzen. Mit einem grünen Kasten versehen sind auf der rechten Seite die Itemsammlungen aus dem Bereich Mathematik, welche in Booklets 1 und 2 vorkommen. Nur wenn in einer Zeile alle Itemsammlungen aus dem Bereich Mathematik auch in Booklet 1 und/ oder Booklet 2 enthalten sind, dann sind alle drei Grundlagen für dieses Booklet gegeben. In solchen Fällen erhielt das zugehörige Booklet auf der linken Seite ebenfalls einen grünen Kasten. Damit zeigt Abbildung 6, dass für Booklets 1, 2, 7 und 8 gilt, dass alle drei Grundlagen in Form der Attributes, der Q-Matrix und der Attribute Hierarchy aus dem Paper von Su und Kollegen (2013) entnommen werden können. Daher fokussiert sich diese Arbeit ausschließlich auf diese vier Booklets der TIMSS 2003.

Abbildung 6

Zusammensetzung der 12 Booklets aus der TIMSS 2003 mit Zuordnung der Itemsammlungen, welche bereits Teile von Booklet 1 oder 2 waren (hier grün markiert) zu allen andern Booklets (modifizierte Version von Exhibit 2.3 entnommen aus Martin et al., 2005)

Exhibit 2.3 Booklet Design for TIMSS 2003 – Grade 4 and Grade 8

Student Booklet	Assessment Blocks					
	Part I		Part II			
Booklet 1	M01	M02	S06	S07	M05	M07
Booklet 2	M02	M03	S05	S08	M06	M08
Booklet 3	M03	M04	S04	S09	M13	M11
Booklet 4	M04	M05	S03	S10	M14	M12
Booklet 5	M05	M06	S02	S11	M09	M13
Booklet 6	M06	M01	S01	S12	M10	M14
Booklet 7	S01	S02	M06	M07	S05	S07
Booklet 8	S02	S03	M05	M08	S06	S08
Booklet 9	S03	S04	M04	M09	S13	S11
Booklet 10	S04	S05	M03	M10	S14	S12
Booklet 11	S05	S06	M02	M11	S09	S13
Booklet 12	S06	S01	M01	M12	S10	S14

Basierend auf diesen Grundlagen werde ich im Folgenden Methoden verwenden, um die Booklets bezüglich ihrer Itemanzahl und ihres Informationsgehalts zu optimieren. Die angewandten Methoden entstammen der kompetenzbasierten Wissensstrukturtheorie. Um im Rahmen dieser Theorie zu arbeiten sind ebenso wie bei den CDMs die drei Grundlagen nötig, sie werden lediglich neu benannt. Im hieran anschließenden Abschnitt „2. Theoretischer Hintergrund“ sollen die grundlegenden Theorien sowie die verwendeten Methoden noch einmal genauer beleuchtet werden. An dieser Stelle soll jedoch schon einmal eine kurze Beschreibung dieser Aspekte gegeben werden, sowie die Einführung einiger Grundbegriffe der kompetenzbasierten Wissensstrukturtheorie. Die drei Grundlagen sind inhaltlich auch in der kompetenzbasierten Wissensstrukturtheorie nahezu dieselben wie bei den CDMs, sie tragen jedoch andere Namen. Die Attributes heißen in diesem Kontext nun „Skills“, die Q-Matrix trägt den Namen „Skillmap“ und der Prozess des Zuschreibens der zur Lösung eines Items hinreichenden Skills zu eben diesem Item nennt sich „Skillfunktion“. Zuletzt lässt sich aus der Attribute Hierarchy eine „Kompetenzstruktur“ ableiten, welche den Platz der Attribute Hierarchy einnimmt und so mit Skillmap und Skills die drei Grundlagen der

kompetenzbasierten Wissensstrukturtheorie bildet. Diese drei Grundlagen sind nun sowohl für das im Folgenden beschriebene Verfahren, wie auch für das Ableiten einer klassischen Wissensstruktur, notwendig.

Die vier Booklets sollen bezüglich Itemanzahl und Informationsgehalt alle auf dieselbe Weise optimiert werden. Daher wird im Folgenden überbegriffsweise nur von „dem Booklet“ gesprochen, um das Verfahren kurz zu erläutern: Zunächst wird untersucht wie gut das vorliegende Booklet bereits Informationen (bezüglich der Kompetenzstruktur) über die zu testende Person erfassen kann und welche Informationen es noch nicht mit seinen Items erfassen kann. Dann werden alle Items, die bisher noch nicht Teil des Booklets sind, aber dennoch in der Skillmap auftauchen, dahingehend untersucht ob sie Aufschluss über die Informationen geben können, die das Booklet bisher noch nicht erfasst. Im nächsten Schritt wird eine Methode namens „Competency Deletion Procedure“ (Anselmi et al., 2022, 2024a, 2024b) angewandt. Mittels dieser lässt sich die Anzahl der Items reduzieren, indem Items, die nicht dabei helfen zusätzliche Informationen über eine Person zu erfassen, ausgeschlossen werden. Die verbleibenden neuen Items werden dem Booklet hinzugefügt („Improvement“) und in einem letzten Schritt wird erneut das „Competency Deletion Procedure“ angewandt, um das Booklet möglichst stark zu kürzen („Shortening“). Am Ende entsteht so ein neues Booklet, welches tendenziell mehr Informationen erfassen kann als das ursprüngliche und dafür möglichst wenige zusätzliche und im besten Fall sogar weniger Items als das ursprüngliche Booklet beinhaltet. Dieses Ergebnis soll also genau die beiden Kriterien vom Anfang erfüllen, indem es erstens eine möglichst präzise Erfassung von Informationen über die Versuchsperson und das zweitens in möglichst kurzer Zeit ermöglicht. Es soll sozusagen eine möglichst ökonomische Wissensdiagnose ermöglicht werden. Ersteres soll durch das Improvement und das zweite soll durch das Shortening erreicht werden. Im letzten Schritt sollen die durch das Optimieren erhaltenen Ergebnisse im Rahmen dieser Arbeit noch einmal genauer analysiert werden. Zunächst sollen durch die optimierten Booklets Vorschläge für Tests entstehen, die mit den untersuchten Items der TIMSS 2003 möglichst optimal Informationen über die untersuchten Inhaltsdimensionen Algebra und Number erfassen können. Hierbei sollen pro Booklet je zwei unterschiedliche Ansätze verfolgt werden. Außerdem soll ein Vergleich der einzelnen neu entstandenen Booklets und einem leicht anderen Ansatz angestrebt werden. Da alle Booklets anhand derselben Items optimiert wurden und lediglich von unterschiedlichen Grunditems ausgingen, erscheint es mir interessant zu analysieren inwiefern sich die Ergebnisse entsprechen oder auch nicht. Außerdem soll

betrachtet werden wie stark man die Booklets tatsächlich optimieren kann, also wie viel mehr Information die neuen Booklets erfassen als die alten und wie stark sich die Itemanzahl optimieren lässt. Zuletzt soll noch betrachtet werden ob es möglich wäre die Booklets maximal informativ bezüglich der zugrunde gelegten Kompetenzstruktur zu machen.

2. Theoretischer Hintergrund

2.1 Klassische Wissensstrukturtheorie

Erstmals 1985 von Doignon und Falmagne entwickelt und seitdem mehrfach optimiert (Doignon & Falmagne, 1999; Falmagne & Doignon, 2010) stellt die klassische Wissensstrukturtheorie eine latent class-Theorie dar, welche sich damit befasst wie Wissensdiagnose möglichst effizient, also möglichst präzise und dabei mit möglichst wenigen Items, ablaufen kann. Hierfür werden Zusammenhänge zwischen Items auf inhaltlicher oder statistischer Ebene angenommen und getestet, um so eine möglichst präzise, nicht-numerische Erfassung des Wissenstands einer Person zu erreichen. Grundsätzlich wird angenommen, dass alle untersuchten Items dichotom lösbar sind: Ein Item wird immer entweder vollständig korrekt oder falsch beantwortet. Da diese Arbeit sich auf die kompetenzbasierte Wissensstrukturtheorie konzentriert, soll an dieser Stelle statt einer umfassenden Einführung in die gesamte klassische Wissensstrukturtheorie nur eine Einführung in ihre Grundlagen und in die Anteile, die für diese Arbeit relevant sind, erfolgen.

Um nun in die Theorie einzusteigen wird zunächst einiges an Grundterminologie eingeführt: Eine „Wissensstruktur“ $(\langle Q, \mathcal{K} \rangle)$ ist ein Relativ aus einem „Wissensbereich“ (Q) und einer Familie von Teilmengen von Q , welche stets mindestens \emptyset und Q enthält (\mathcal{K}) . Q setzt sich aus allen für die aktuelle Untersuchung relevanten Items zusammen und \mathcal{K} setzt sich aus seinen Teilmengen zusammen. Die Teilmengen, aus welchen sich \mathcal{K} zusammensetzt, nennt man „Wissenszustände“ (K) . Unter ihnen sind stets \emptyset und Q . Dabei stellt \emptyset den Zustand der vollkommenen Unwissenheit dar, also einen Zustand in dem sich eine Person befinden würde, die keines der Items aus Q lösen kann und Q stellt den Zustand des vollkommenen Wissens dar, also einen Zustand in dem sich eine Person befinden würde, die jedes der Items aus Q lösen kann. Alle anderen Wissenszustände in \mathcal{K} sind zwischen \emptyset und Q zu verorten und beschreiben Zustände, in welchen sich Personen befinden würden, die einige aber nicht alle Items aus Q lösen können. Allgemein kann man sagen, dass eine Person der ein bestimmter

Wissenszustand zugeschrieben wurde, der deterministischen Theorie zufolge immer dazu in der Lage ist eben die Items aus Q zu lösen, die in ihrem Wissenszustand enthalten sind.

Der nächste Grundbegriff ist die „Präzedenzrelation“. Sie beschreibt die Voraussetzungsbeziehungen zwischen Items. So gibt die „Präzedenzrelation“ Aufschluss darüber, welche Items Voraussetzungen für welche anderen Items sind. Damit erfüllt sie für Items eine ähnliche Funktion wie es die Attribute Hierarchy für Attributes bei den CDMs tut. Der Unterschied liegt darin, dass im Rahmen der Präzedenzrelation die Annahme gilt, dass wenn Items Voraussetzungen für ein Item sind, dann muss eine Person, die das Item lösen kann stets auch dazu in der Lage sein alle Items zu lösen, welche Voraussetzungen für es darstellen. Im Rahmen einer Attribute Hierarchy jedoch gilt die Annahme, dass wenn Attributes Voraussetzungen für ein Attribute sind, dann muss eine Person, die über das Attribute verfügt, stets nur über mindestens eines der Attributes verfügen, welches eine direkte Voraussetzung für es darstellt. Im Falle einer Darstellung via Und-Oder-Graph wie in Abbildung 5 sind direkte Voraussetzungsbeziehungen dadurch kenntlich, dass Attributes optisch direkt über eine Pfeilverbindung verbunden sind. Es sei noch erwähnt, dass \mathcal{K} nicht die Potenzmenge von Q sein muss, also muss \mathcal{K} nicht alle möglichen Kombinationen von Items, beziehungsweise jeden theoretisch möglichen Zustand, als ein K beinhalten. Genaugenommen sorgen allein die Voraussetzungsbeziehungen meist schon dafür, dass dies nicht der Fall ist, da durch sie definiert ist, dass einige Items nie ohne andere in einem K auftreten können. Somit ermöglicht die klassische Wissensstrukturtheorie beispielsweise auch, dass adaptives Testen (Heller, 2023; Falmagne et al., 1990, 2006) möglich wird. Dank der Voraussetzungsbeziehungen in der Wissensstruktur muss man einer Versuchsperson nämlich nicht mehr alle Items vorlegen, um ihren Wissenszustand bezüglich aller Items zu ermitteln. An dieser Stelle ist noch eine grundlegende Unterscheidung einzuführen: Eine Wissensstruktur, welche abgeschlossen ist bezüglich Mengenvereinigung, wird als „Wissensraum“ bezeichnet und eine Wissensstruktur, welche abgeschlossen ist bezüglich Mengenvereinigung und -schnitt nennt man „quasi-ordinalen Wissensraum“. Ein quasi-ordinaler Wissensraum hat viele wünschenswerte Eigenschaften, so korrespondiert er beispielsweise eins zu eins mit einer zugehörigen Präzedenzrelation, was auf einen Wissensraum oder auch Wissensstrukturen im Allgemeinen nicht zutrifft. Zwar kann man auch von einem Wissensraum eine Präzedenzrelation ableiten, von dieser käme man jedoch bei einem inversen Vorgehen nicht wieder zurück auf den ursprünglichen Wissensraum,

sondern auf einen quasi-ordinalen Wissensraum, der dem ursprünglichen Wissensraum lediglich ähnelt. Der quasi-ordinale Wissensraum hat im Kontext von adaptivem Testen zudem insbesondere eine Eigenschaft: Man kann, wenn man einen quasi-ordinalen Wissensraum vorliegen hat von der korrekten Beantwortung eines Items darauf schließen, dass die Versuchsperson auch alle Items beherrscht, welche gemäß der Präzedenzrelation Voraussetzungen für das Item darstellen. Da wir auf Basis einer Attribute Hierarchy arbeiten, welche sich in ihren grundlegenden Eigenschaften wie bereits beschrieben von einer Präzedenzrelation unterscheidet, können wir nur Wissensräume ableiten. Folglich lässt sich zwar nach wie vor ein adaptives Testverfahren ableiten, dieses ließe aber keine so simplen Schlüsse durch die korrekte Beantwortung eines Items zu. Stattdessen sähen die Rückschlüsse im Falle eines Wissensraums wie folgt aus: Wenn wir der Person zuerst ein Item vorlegen, welches andere Items als Voraussetzung hat, dann können wir, sollte sie das Item lösen können darauf schließen, dass sie das Item wie auch mindestens eine seiner Voraussetzungen lösen kann. Durch das Vorlegen weiterer Items kann man nun ermitteln, welche der Voraussetzungen des Items dies ist und sobald man dies ermittelt hat, kann man so wiederum darauf schließen, dass die Versuchsperson auch über mindestens eine der Voraussetzungen dieser Items verfügen muss. Sobald Items nur noch eine Voraussetzung haben und dies sich vielleicht sogar über eine längere Kette erstreckt, kann man so auf die Lösungsfähigkeit der Person bezüglich Items schließen, welche man ihr nie vorgelegt hat. So würden eine effizientere Wissensdiagnose möglich, welche im Rahmen einer „adaptive Testung“ genutzt werden könnte.

Diese Beschreibung ist hierbei rein deterministisch und lässt nicht zu, dass eine Versuchsperson auch mal ein Item zufällig korrekt oder durch einen versehentlichen Fehler falsch beantwortet. Diesen Möglichkeiten wird in der „probabilistischen Wissensstrukturtheorie“ Rechnung getragen, diese soll allerdings nicht der primäre Gegenstand dieser Arbeit sein. Stattdessen bleiben wir im Rahmen dieser Arbeit erst einmal im rein deterministischen Bereich. Lediglich an einer Stelle wird sich einer Methode der probabilistischen Wissensstrukturtheorie bedient. Der Grund hierfür ist, dass die probabilistische Wissensstrukturtheorie Identifizierbarkeitsprobleme hat, also keine verlässliche Parameterschätzung zulässt (Stefanutti, & Spoto, 2020; Heller, 2017). Da diese Arbeit außerdem primär durch eine konkrete Anwendung einen Einblick in die Möglichkeiten der kompetenzbasierten Wissensstrukturtheorie geben möchte, wird hier der Fokus auf die

vielleicht weniger realistische, aber dafür verlässlichere und leichter verständliche deterministische Wissensstrukturtheorie gelegt.

2.2 Kompetenzbasierte Wissensstrukturtheorie

Die kompetenzbasierte Wissensstrukturtheorie ist eine Erweiterung der klassischen Wissensstrukturtheorie und wurde erstmals Ende des zwanzigsten Jahrhunderts eingeführt (Doignon, 1994; Falmagne et al., 1990). Für einen guten Überblick über die im Folgenden beschriebenen Grundlagen und noch einiges mehr sei an dieser Stelle auf das Paper von Heller, Ünlü und Albert aus dem Jahr 2013 verwiesen. Grundsätzlich ist die kompetenzbasierte Wissensstrukturtheorie eine Art Metatheorie in Relation zur klassischen Wissensstrukturtheorie. Hier steht als grundlegender Begriff statt der Wissensstruktur die Kompetenzstruktur und mit den drei in der Einleitung bereits im Kontext von CDMs beschriebenen Grundlagen lässt sich aus der Kompetenzstruktur eine Wissensstruktur ableiten. Man kann also sagen, dass man in der kompetenzbasierten Wissensstrukturtheorie von einer übergeordneten Struktur ausgeht, von welcher man allerdings wieder auf eine Wissensstruktur schließen kann und so dann mit der Wissensstruktur wie gewohnt arbeiten kann. Zusätzlich dazu bietet die kompetenzbasierte Wissensstrukturtheorie einige Möglichkeiten, die wir mit der klassischen Wissensstrukturtheorie nicht hätten. Bevor jedoch auf diese weiter eingegangen werden soll, sollen an dieser Stelle zunächst einige Grundbegriffe eingeführt werden. Zunächst der wohl zentralste Begriff: Eine Kompetenzstruktur ähnelt stark einer Wissensstruktur, nur dass sie als Grundlage nicht die Items des Tests verwendet, sondern die Fähigkeiten (Skills), welche für die Lösung eben dieser Items benötigt werden. Diese Skills sind inhaltlich dasselbe wie die Attributes bei den CDMs. Wie auch eine Wissensstruktur beinhaltet eine Kompetenzstruktur Zustände, welche hier nun statt der Fähigkeit einer Person bestimmte Items zu lösen angeben, ob die Person über bestimmte allgemeinere Fähigkeiten (die Skills) verfügt. Diese Zustände heißen „Kompetenzzustände“. Anhand eines Beispiels könnte man den Unterschied zur Wissensstruktur folgendermaßen beschreiben: Wenn eine Person x fähig ist die Aufgabe a („ $12 + 7$ “) korrekt zu lösen, würde dies im Wissenszustand der Person in der Art vermerkt, dass dieser aussagen würde „Person x kann Aufgabe a lösen“. Wenn man nun jedoch im Kontext der Kompetenzstruktur annehmen würde, dass eine Person zur Lösung von Aufgabe a den Skill „Addition“ beherrschen muss, würde die korrekte Lösung der Aufgabe a im

Kompetenzzustand der Person in der Art vermerkt, dass dieser aussagen würde „Person x beherrscht den Skill Addition“. Diese Aussage ist nun natürlich um einiges allgemeiner und für Außenstehende auch besser verständlich als die Aussage aus der klassischen Wissensstrukturtheorie. Dafür liegt die Aussage weniger nah an den Daten. Um nun solche Aussagen überhaupt machen zu können ist es nötig jedem Item zuzuweisen, welche Skills für seine Lösung notwendig sind. Diesen Job können beispielsweise wie bei Su und Kollegen (2013) Experten übernehmen. Die Zuweisung wird in einer sogenannten Skillmap festgehalten, welche eng verwandt ist mit der Q-Matrix bei den CDMs. Die Skills sind den Aufgaben über eine Skillfunktion zugewiesen in einer Weise, dass man ein Item in diese Skillfunktion geben kann und als Ergebnis kommen eben die Skills in Form einer sogenannten „Kompetenz“ heraus, welche für die Lösung des Items hinreichend sind. Hier sei noch erwähnt, dass diese Skillfunktion nicht wirklich eine mathematische Berechnung vornimmt, sondern lediglich die, der Aufgabe als hinreichend für ihre Lösung zugeordneten, Skills aus der Skillmap ausgibt. Somit ist die Kompetenz eines Items immer eine Menge der Skills, die gerade hinreichend ist, um es korrekt zu beantworten. Wie auch bei einer Wissensstruktur werden in einer Kompetenzstruktur Voraussetzungsbeziehungen zwischen Items festgehalten, sodass manche Skills Voraussetzungen für andere darstellen und somit teilweise vorliegen müssen, wenn die Skills für welche sie Voraussetzungen sind auch vorliegen. Hierbei gelten dieselben Regeln wie bei den Voraussetzungsbeziehungen des Wissensraums, welche im Abschnitt „2.1 Klassische Wissensstrukturtheorie“ erklärt wurden. Inhaltlich beinhaltet die Kompetenzstruktur also all die Kompetenzzustände, welche gemäß dieser Voraussetzungsbeziehungen möglich sind. Genauso wie eine Wissensstruktur beinhaltet auch sie stets einen Zustand der vollkommenen Unfähigkeit, in welchem keiner der Skills vorliegt, und einen Zustand des vollkommenen Befähigung, in dem alle Skills gleichzeitig vorliegen. Auf die Kompetenzstruktur mit all ihren möglichen Zuständen kommt man beispielsweise, indem man die Attribute Hierarchy aus den CDMs betrachtet und daraus alle den Voraussetzungsbeziehungen entsprechend möglichen Kompetenzzustände ableitet. Man kann also sagen, dass die Kompetenzstruktur der Attribute Hierarchy inhaltlich sehr ähnlich ist. Hiermit haben wir erneut unsere drei Grundlagen versammelt, im Falle der kompetenzbasierten Wissensstrukturtheorie sind dies nun also die Skills, die Skillmap und zuletzt die Kompetenzstruktur.

2.3 Wissensstruktur aus Kompetenzstruktur ableiten

Mittels der drei beschriebenen Grundlagen können wir nun eine Wissensstruktur von unserer Kompetenzstruktur ableiten (Heller, 2023; Anselmi et al., 2024a, 2024b). Hierbei induziert einer (oder auch mehrere) Kompetenzzustand(/-stände) einen Wissenszustand. Den Wissenszustand, der davon ausgeht, dass die Person die Items lösen kann, für welche in der hier betrachteten Teilmenge von Skills die nötigen Kompetenzen vorliegen. Folglich hat die induzierte Wissensstruktur maximal genauso viele Wissenszustände wie die Kompetenzstruktur Kompetenzzustände hat. Für den Induktionsprozess wird eine „problem function“ verwendet, welche wie oben beschrieben jedem Kompetenzzustand einen Wissenszustand zuweist. Man kann also aus einer Kompetenzstruktur, solange man weiß welche Skills für die Lösung welches Items hinreichend sind (also man eine Skillfunktion hat), stets eine Wissensstruktur ableiten. Wie bereits erwähnt können auch mehrere Kompetenzzustände auf denselben Wissenszustand abbilden. Dies ist dann der Fall, wenn die Items des Tests nicht ausreichend sind, um zwischen allen Kompetenzzuständen des Tests zu unterscheiden. Als sehr simples Beispiel zur Veranschaulichung können wir einen Test mit nur einem Item „a“ betrachten, welchem die Kompetenz $\{s\}$ zugewiesen ist. Nehmen wir an unsere Kompetenzstruktur beinhalte sowohl den Kompetenzzustand $\{s\}$, als auch den Kompetenzzustand $\{s, t\}$. Wir könnten von einer Person, die sich im Kompetenzzustand $\{s\}$ befindet wie auch von einer im Kompetenzzustand $\{s, t\}$ durch unser eines Item „a“ nur auf eines schließen. Darauf, dass sich die Person im Wissenszustand $\{a\}$ befindet, da es kein Item gibt, welches eine Unterscheidung zwischen den beiden Kompetenzzuständen ermöglicht. Wir hätten nun also eine Kompetenzstruktur mit zwei Kompetenzzuständen und könnten daraus aber nur eine Wissensstruktur mit einem Wissenszustand ableiten. Verantwortlich für diese Situation wäre der Mangel an Items, die uns diese Unterscheidung ermöglichen könnten. Gäbe es ein weiteres Item „b“ im Test, welchem die Kompetenz $\{s, t\}$ zugeordnet wäre, dann könnten wir durch die nun zwei Items die nötige Unterscheidung durchführen. Eine Person im Kompetenzzustand $\{s\}$ könnte nur Item „a“ lösen und würde folglich dem Wissenszustand $\{a\}$ zugewiesen während eine Person im Kompetenzzustand $\{s, t\}$ beide Items lösen könnte und somit dem Wissenszustand $\{a, b\}$ zugewiesen werden würde. Man kann also allgemein sagen, dass je mehr Kompetenzzustände durch die Items unterschieden werden können, desto mehr Wissenszustände hat die Wissensstruktur, die von der Kompetenzstruktur abgeleitet werden kann. Dementsprechend gilt auch das Folgende: Wenn sich anhand der Items eines Tests manche Kompetenzzustände nicht unterscheiden lassen,

dann führt das dazu, dass die Kompetenzstruktur eine Wissensstruktur induziert, welche weniger Wissenszustände hat als die Kompetenzstruktur Kompetenzzustände hat. Das liegt daran, dass durch die Ununterscheidbarkeit einiger Kompetenzzustände, diese auf denselben Wissenszustand abgebildet werden. Man spricht an dieser Stelle von Identifizierbarkeitsproblemen. Wenn ein Test zwischen allen Kompetenzzuständen der angenommenen Kompetenzstruktur unterscheiden kann, spricht man von einem „fully informative“ Test bezüglich der angenommenen Kompetenzstruktur (Anselmi et al., 2022, 2024a, 2024b). Dies ist erstrebenswert, jedoch aufgrund teils sehr großer Kompetenzstrukturen und begrenzter Itemmengen oft nicht erreichbar. Dennoch ist es erstrebenswert einen Test um Items mit Kompetenzen zu erweitern, die mehr Unterscheidungen zulassen, da der Test so informativer wird. Genauer kann man sagen, dass wenn ein Test zwischen denselben und mehr Kompetenzzuständen differenzieren kann als ein anderer, dann ist er „more informative“ als der andere. In diesem Fall hätte die induzierte Wissensstruktur des differenzierenden Tests mehr Wissenszustände als die induzierte Wissensstruktur des anderen Tests. Im optimalen Fall eines fully informative Tests hätte die induzierte Wissensstruktur genauso viele Wissenszustände wie die Kompetenzstruktur, von welcher sie abgeleitet wurde Kompetenzzustände hat. Eine Wissensstruktur mit möglichst vielen Wissenszuständen aus einer Kompetenzstruktur abzuleiten ist deshalb so erstrebenswert, weil die Kompetenzstruktur nur auf theoretischen Annahmen bezüglich der untersuchten Skills beruht. Die Skillmap basiert genauso nur auf theoretischen Annahmen, diesmal bezüglich der Eigenschaften der Items. Wenn wir nun jedoch empirisch etwas in Erfahrung bringen wollen, oder auch unsere Annahmen testen wollen, dann benötigen wir dafür den Test und seine Items. An dieser Stelle kommen wir nun mit einer Wissensstruktur, welche weniger Wissenszustände als unsere theoretisch zu ihr korrespondierende Kompetenzstruktur Kompetenzzustände hat, an ein Problem. Wir können durch unseren Test nur direkt auf den Wissenszustand einer Person schließen und auch das nur, wenn wir die deterministische Annahme in den Test stecken, dass eine Person, die bestimmte Items lösen kann, auch immer genau diese und keine anderen Items lösen wird. An die Kompetenzzustände, welche wir mit der Untersuchung potentiell gerade validieren wollen, kommen wir nur indem wir von unserer Wissensstruktur eine Kompetenzstruktur ableiten. Wenn nun beispielsweise zwei Kompetenzzustände denselben Wissenszustand induzieren, dann bedeutet das umgekehrt, dass wir vielleicht bei unserer Erhebung einen Wissenszustand vertreten haben, welcher beim Versuch unsere Kompetenzstruktur aus der Wissensstruktur

abzuleiten zwei Kompetenzzustände zugleich induziert. Dabei kann sich eine Person per Definition zu jedem Zeitpunkt nur in einem Kompetenzzustand befinden. Ob sich die Annahmen unserer theoretischen Kompetenzstruktur also bekräftigen lassen, ließe sich hier nicht eindeutig sagen und verantwortlich dafür wäre der Test, welcher keine Unterscheidung der beiden erwähnten Kompetenzzustände zulässt. Je weniger derartige Fälle wir haben, desto verlässlicher lässt sich unsere Kompetenzstruktur durch Erhebungen untersuchen, deshalb streben wir Tests an, die more informative sind. Außerdem hat eine Wissensstruktur mit möglichst vielen Wissenszuständen auch den Vorteil, dass sie den tatsächlichen aktuellen Wissensstand einer Person tendenziell präziser erfassen kann als eine Wissensstruktur mit weniger Wissenszuständen.

2.4 Testoptimierung

Mittels der drei Grundlagen kann außerdem eine umfassende Testoptimierung erfolgen. Der im Folgenden beschriebene Testoptimierungsprozess basiert auf zwei Methoden aus Anselmi und Kollegen (2022), nämlich dem Shortening und dem Improvement eines Tests durch Methoden der kompetenzbasierten Wissensstrukturtheorie. Daher sollen diese zwei Methoden und die Voraussetzungen und Vorannahmen um sie anzuwenden im Folgenden zunächst eingeführt werden.

2.4.1 Voraussetzungen und Vorannahmen für Shortening und Improvement

Bevor wir die erwähnten Methoden anwenden können, muss zunächst eine zentrale Annahme bezüglich des Tests und der Skills getroffen werden. Für die Anwendung der Methoden ist es nötig anzunehmen, dass es sich um entweder einen conjunctive oder einen „disjunctive“ Test handelt, da die Methoden bisher nur in diesen beiden Fällen und nicht in einem allgemeinen Fall anwendbar sind (Anselmi et al., 2022). Diese Annahme treffen auch schon Su und Kollegen (2013) genauso. Theoretisch ist es möglich, dass es mehrere Wege geben kann um ein Item zu lösen. So könnte ein Item beispielsweise vollständig korrekt beantwortet werden, wenn entweder Skills „s“ und „t“ gleichzeitig, oder aber Skills „u“ und „v“ gleichzeitig vorliegen. In diesem Fall gäbe es sozusagen zwei unterschiedliche Herangehensweisen an das Item, welche aber beide zum selben korrekten Ergebnis führen. Weder conjunctive noch disjunctive Tests lassen einen derart allgemeinen Fall zu, stattdessen schränken sie die Regeln, nach welchen entschieden wird, welche Skills für die Lösung der

Items nötig sind, wie folgt ein: Ein conjunctive Test ist dadurch definiert, dass alle Skills, die mit einem Item assoziiert sind, stets gleichzeitig vorliegen müssen um es zu lösen. Folglich gibt es zu jedem Item nur genau eine Kompetenz, die hinreichend für seine Lösung ist und diese Kompetenz ist die, die alle Skills beinhaltet, die mit dem Item assoziiert sind. Eine Person, die das Item lösen kann, muss also über all diese Skills zugleich verfügen. Ein disjunctive Test ist dadurch definiert, dass von den Skills, die mit einem Item assoziiert sind, stets nur einer vorliegen muss, um das Item zu lösen. Folglich gibt es zu jedem Item genauso viele Kompetenzen, die hinreichend für seine Lösung sind, wie es Skills gibt, die mit diesem Item assoziiert sind. Diese Kompetenzen sind die, die je einen unterschiedlichen der Skills beinhaltet, die mit dem Item assoziiert sind. Eine Person, die das Item lösen kann, muss also über nur einen dieser Skills verfügen. Anselmi und Kollegen (2022) und auch Su und Kollegen (2013) schreiben davon, dass die Annahme eines conjunctive Tests für Tests sinnvoller sind, welche im Bildungskontext Anwendung finden, da unser Bildungssystem, insbesondere das schulische, darauf ausgelegt ist, dass Schülern immer eine Fähigkeit nach der anderen in einer durchdachten Reihenfolge vermittelt wird. So ist es durchaus sinnvoll anzunehmen, dass die Skills eines bestimmten Bildungsabschnitts stets zugleich vorliegen müssen. Andersherum schreiben Anselmi und Kollegen (2022), dass die Annahme von disjunctive Tests sinnvoller im medizinischen/ diagnostischen Kontext ist, da hier oft bereits das Vorliegen einiger Symptome, die durch einen Test erfasst wurden, zur Diagnose einer bestimmten Krankheit ausreicht. Außerdem ist es diagnostisch nur selten nötig, dass die Symptome alle zugleich vorliegen. Aufgrund dieser Annahmen wird im Folgenden, da es sich bei der TIMSS 2003 um einen Test aus dem Bildungskontext handelt, davon ausgegangen, dass die untersuchten Teile der TIMSS 2003 conjunctive Tests darstellen.

Um den Prozess des Shortenings und des Improvements zu erklären, ist es zudem nötig, den Begriff des „discernibility systems“ einzuführen (Anselmi et al., 2022, 2024a, 2024b). Ein discernibility system ist ein Quadrupel der Art $(S, \mathcal{D}, T, \delta)$. Bevor die einzelnen Bestandteile des Quadrupels eingeführt werden, soll an dieser Stelle noch einmal erwähnt werden, dass für das Erstellen eines und die Arbeit mit einem discernibility system stets drei Grundlagen benötigt werden. Erstens eine endliche und bekannte Menge an Skills (S). Zweitens eine Kompetenzstruktur (C), welche auf S definiert ist. Drittens eine endliche und bekannte Menge an Items, für die eine Skillfunktion existiert, sodass jedem dieser Items Teilmengen von Skills (Kompetenzen), die hinreichend für deren Lösung sind, aus S

zugewiesen werden können. Da Shortening und Improvement beide nur auf conjunctive oder disjunctive Tests anwendbar sind, muss die erwähnte Skillfunktion immer entweder jedem Item genau eine Teilmenge aus S , die hinreichend für die Lösung des Items ist zuweisen (conjunctive Test) oder jedem Item eine oder mehrere Teilmengen aus S zuweisen, die jeweils aus nur einem Skill bestehen (disjunctive Test). In unserem Fall wird wie bereits erwähnt ein conjunctive Test angenommen, folglich trifft hier der erstere Fall zu. Die drei hier beschriebenen Voraussetzungen (S , C , Skillfunktion), sind nun genau die drei Grundlagen, die zuvor bereits eingeführt und beschrieben wurden und die sich aus der Arbeit von Su und Kollegen (2013) entnehmen lassen. Folglich können sie im Fall dieser Arbeit auch als gegeben angenommen werden. Nun kommen wir zurück zum discernibility system ($(S, \mathcal{D}, T, \delta)$). S wurde bereits zuvor eingeführt und beschreibt eine nichtleere, endliche und bekannte Menge von Skills. \mathcal{D} beschreibt eine nichtleere Sammlung ungeordneter Paare von Kompetenzzuständen aus der auf S definierten Kompetenzstruktur C . \mathcal{D} muss dabei nicht aus allen möglichen ungeordneten Paarkombinationen aus C bestehen, wenn man jedoch bei einem Test ein umfassendes Shortening oder Improvement durchführen möchte, dann sollte \mathcal{D} aus allen möglichen ungeordneten Paarkombinationen aus C bestehen. T beschreibt eine nichtleere Sammlung nichtleerer Kompetenzen, welche mit den Items aus der Skillfunktion korrespondieren. Da wir einen conjunctive Test annehmen, gibt es je eine für seine Lösung hinreichende Kompetenz pro Item. T beinhaltet alle verschiedenen dieser Kompetenzen je einmal, weshalb T maximal so viele Elemente beinhaltet wie es Items gibt. Es können auch weniger sein, wenn mehrere Items dieselben Skills für ihre Lösung benötigen, also derselben Kompetenz zugeordnet werden. Zuletzt gibt es noch δ . δ nennt sich auch „discernibility function“ und nimmt als Argument stets zwei Kompetenzzustände aus C , also die Elemente aus \mathcal{D} . δ weißt jedem seiner Argumente alle Kompetenzen aus T zu, anhand derer man das Paar unterscheiden kann. Also enthält das Ergebnis der discernibility function genau die Kompetenzen aus T , die eine Unterscheidung der beiden in die Funktion gegebenen Kompetenzzustände ermöglichen. Wenn man nun also die Items in einem Test hat, welche eine Unterscheidung zwischen den zwei Kompetenzzuständen ermöglichen, kann man über die problem function eine Wissensstruktur mit mehr Wissenszuständen ableiten als zuvor. Die Sammlungen von Kompetenzen, welche als Ergebnis von der discernibility function ausgegeben werden, nennen sich auch „competency bags“. Es gilt, dass die beiden

Kompetenzzustände, die man in δ gegeben hat nur unterscheidbar sind, wenn das zu ihnen gehörige competency bag nichtleer ist. An dieser Stelle sei auf den Abschnitt „5.

Discernibility systems“ aus dem Paper von Anselmi und Kollegen (2022) verwiesen. Im Rahmen der Einführung dieses Konzepts wird das discernibility system in deren Arbeit ausführlich und anhand eines Beispiels erläutert.

2.4.2 Shortening

Beim Shortening wird davon ausgegangen, dass es nur sinnvoll ist Items mit bestimmten Kompetenzen im Test zu haben, die auch einen individuellen Beitrag dazu leisten zwischen Kompetenzzuständen der grundlegenden Kompetenzstruktur zu unterscheiden. Folglich sollen Items mit Kompetenzen, die nur zwischen denselben Kompetenzzuständen unterscheiden können wie andere Items mit anderen Kompetenzen des Tests, aus dem Test ausgeschlossen werden. Dadurch wird der Test kürzer und beinhaltet letztendlich nur noch die Items, die auch einen Beitrag dazu leisten eine Wissensstruktur mit möglichst vielen Wissenszuständen zu generieren. Dadurch kann der Wissensstand einer Person nach wie vor möglichst exakt bestimmt werden. Das Vorgehen hierfür sieht wie folgt aus: Dank δ wissen wir durch welche Kompetenzen aus \mathcal{T} sich die Kompetenzzustände unserer Struktur \mathcal{C} unterscheiden lassen. Wenn man nun permanent die competency bags, welche δ ausgibt, monitoren würde, könnte man Schritt für Schritt aus \mathcal{T} Kompetenzen streichen und dabei darauf achten keine competency bags zu leeren. Als Resultat könnte ein Test, der nur noch aus Items besteht, welche mit eben diesen resultierenden Kompetenzen assoziiert sind, immer noch zwischen denselben Zuständen unterscheiden, benötigte dafür jedoch weniger Items als der ursprüngliche Test, da das neue \mathcal{T} weniger Kompetenzen beinhalten würde. Für dieses Ziel wurde ein Verfahren entwickelt, welches schrittweise eine Kompetenz nach der anderen löscht, das „Competency Deletion Procedure“ (CDP) (Anselmi et al., 2022, 2024a, 2024b). Das Verfahren läuft in vier Schritten ab: Erstens werden anhand der discernibility function alle competency bags bestimmt und es werden alle nichtleeren competency bags ermittelt, welche mehr als eine Kompetenz beinhalten.

Zweitens werden nun alle Kompetenzen bestimmt, die theoretisch gelöscht werden können. Diese müssen in nichtleeren competency bags auftauchen, welche mehr als eine Kompetenz beinhalten und sie dürfen nirgends in einem nichtleeren competency bag stehen,

welches nur sie als Kompetenz beinhaltet. Wäre letzteres der Fall, würde ihre Löschung aus \mathcal{T} dafür sorgen, dass competency bags leer würden und somit könnten die Kompetenzen in \mathcal{T} nicht mehr zwischen denselben und genauso vielen Kompetenzzuständen unterscheiden wie zuvor.

Drittens muss nun entschieden werden welche der theoretisch löszbaren Kompetenzen in diesem Schritt gelöscht werden soll. Hierfür gibt es mehrere Ansätze (Anselmi et al., 2024a), von denen beispielsweise der simpelste eine rein zufällige Auswahl der Kompetenz vorschlägt. Da das Ziel dieser Arbeit jedoch ist möglichst kurze Tests zu erschaffen, wird hier eine Methode mit dem Namen „Many bags preferred“ (MBP) gewählt. Diese Methode erzielt als Ergebnis tendenziell ein \mathcal{T} , das die geringstmöglichen Kompetenzen beinhaltet (Anselmi et al., 2024a). Für MBP wird bei jedem Löschungsschritt die Kompetenz gelöscht, die zum Zeitpunkt des Löschungsschritts in den wenigsten competency bags enthalten ist. Sollten mehrere Kompetenzen zugleich in der geringsten Anzahl von competency bags enthalten sein, wird durch eine Zufallsziehung entschieden, welche davon in diesem Schritt gelöscht wird. So werden die competency bags über möglichst viele Löschungsschritte hinweg nicht leer, sodass man tendenziell mehr Löschungsschritte durchführt und so mehr Kompetenzen löschen kann. Somit werden tendenziell möglichst viele Kompetenzen gelöscht und Tests mit möglichst wenigen Items entstehen.

Viertens wird die ausgewählte Kompetenz aus \mathcal{T} gelöscht und eine neue Version von \mathcal{T} entsteht. Nachdem nun alle vier Schritte durchlaufen wurden, beginnt der Prozess beim ersten Schritt aufs Neue, nur wird dieses Mal statt dem alten, das neue \mathcal{T} zum Erstellen der competency bags gewählt. Das Verfahren durchläuft so viele Iterationen, bis irgendwann im zweiten Schritt keine löszbare Kompetenz mehr identifiziert werden kann. An diesem Punkt wird die zu diesem Zeitpunkt aktuelle Version von \mathcal{T} als finales und vollständig gekürztes neues \mathcal{T} verwendet. Derartige vollständig gekürzte Versionen eines ursprünglichen \mathcal{T} s, welche der Theorie entsprechend mit weniger Kompetenzen zwischen denselben Kompetenzzuständen unterscheiden können wie das ursprüngliche \mathcal{T} , werden als „reducts“ des ursprünglichen \mathcal{T} s bezeichnet.

An dieser Stelle sei auf den Abschnitt „7. Competency deletion procedure“ aus dem Paper von Anselmi und Kollegen (2022) verwiesen. Im Rahmen der Einführung dieses Konzepts wird das CDP in deren Arbeit ausführlich und anhand eines Beispiels erläutert. Aus demselben Grund, hier jedoch auf das Konzept des Shortenings, sei auf die Abschnitte „8.3. Shortening an existing test“ aus dem Paper von Anselmi und Kollegen (2022) und „5 Shortening of an existing test“ aus dem Paper von Anselmi und Kollegen (2024b) verwiesen. Während ersteres wieder im Rahmen der Einführung des Konzepts das Shortening an einem simplifizierten Beispiel ausführlich erläutert, gibt letzteres ein Beispiel aus der echten Welt für das Shortening an. Bei letzterem wird das Shortening auf den „fraction subtraction test“ (Tatsuoka, 1990) angewandt, um so die Anwendbarkeit in der Praxis zu veranschaulichen.

2.4.3 Improvement

Beim Improvement wird davon ausgegangen, dass man einen Test bezüglich seines Informationsgehalts verbessern kann. Hierfür wird er um Items mit Kompetenzen verlängert, die dafür sorgen, dass der neu entstehende Test dazu in der Lage ist mehr Kompetenzzustände zu unterscheiden als der bisherige. Es sollen nach Möglichkeit Items mit Kompetenzen, die zwischen Kompetenzzustandsparen unterscheiden können, zwischen denen bisher nicht unterschieden werden konnte, zum Test hinzugefügt werden. Dadurch wird der Test zwar länger, ermöglicht jedoch auch das Erfassen einer Wissensstruktur mit mehr Wissenszuständen und somit kann der Wissensstand der Versuchsperson genauer bestimmt werden. Der Test soll durch dieses Verfahren more informative (Anselmi et al., 2022) werden. Das Vorgehen, nach welchem Improvement abläuft um einen Test more informative zu machen, besteht grundsätzlich aus drei Schritten und sieht wie folgt aus: Erstens müssen möglichst viele Items gesammelt werden, die für ihre Lösung dieselben Skills benötigen wie die anderen Items des Tests. Es werden also Items gesucht, die mit Kompetenzen assoziiert sind, welche ausschließlich Skills aus S beinhalten (im Folgenden „Improvement Kompetenzen“).

Zweitens werden durch δ die competency bags bezüglich der „Improvement Kompetenzen“ erstellt. Allerdings werden sie nicht basierend auf allen Kompetenzzustandsparen der Kompetenzstruktur erstellt wie zuvor. Stattdessen werden lediglich die competency bags von den Kompetenzzustandsparen erstellt, die sich durch den ursprünglichen Test nicht unterscheiden lassen. Auf diese competency bags und die

„Improvement Kompetenzen“ wird nun das CDP angewandt, um so eine neue, gekürzte Menge von Kompetenzen zu erhalten, die noch immer zwischen möglichst vielen Kompetenzzuständen unterscheiden kann, zwischen denen der ursprüngliche Test nicht unterscheiden konnte.

Drittens wird das so entstandene reduct (im Folgenden „Improvement reduct“) den Kompetenzen \mathcal{T} des ursprünglichen Tests hinzugefügt. Dadurch entsteht ein neues \mathcal{T} , welches die alten und möglichst wenige neue Kompetenzen beinhaltet, nämlich genau die, die für einen more informative Test nötig sind. Wenn man nun im letzten Schritt noch den ursprünglichen Items des Tests welche hinzufügt, die mit den Kompetenzen des neu hinzugefügten „Improvement reducts“ assoziiert sind, dann erhält man einen möglichst knapp verlängerten und gemäß der verfügbaren Items maximal informativen more informative Test im Vergleich zum ursprünglichen Test. An dieser Stelle sei auf die Abschnitte „8.2. Improving an existing test“ aus dem Paper von Anselmi und Kollegen (2022) und „Improvement of an existing test“ aus dem Paper von Anselmi und Kollegen (2024b) verwiesen. In deren Arbeiten wird das Improvement ausführlich und anhand jeweils eines Beispiels erläutert. Während ersteres im Rahmen der Einführung des Konzepts das Improvement an einem simplifizierten Beispiel ausführlich erläutert, gibt letzteres ein Beispiel aus der echten Welt für das das Improvement an. Bei letzterem wird das Improvement auf das, auch in dieser Arbeit verwendete, Booklet 2 des „TIMSS 2003“ (Arora et al., 2005) angewandt, um so die Anwendbarkeit in der Praxis zu veranschaulichen.

2.4.4 Gesamter Optimierungsprozess auf Kompetenz-Ebene

Der gesamte Optimierungsprozess erfolgt zunächst auf der Kompetenz-Ebene. Dies erfolgt durch die Anwendung der zuvor beschriebenen Methoden. Das vollständige Vorgehen besteht hierbei aus vier Schritten. Erstens werden mittels der Kompetenzstruktur und der Kompetenzen, die mit den Items des ursprünglichen Tests assoziiert sind, nicht unterscheidbare Kompetenzzustandspaare identifiziert.

Im zweiten Schritt werden die Kompetenzen aller verfügbaren Items gesammelt, die nicht den Kompetenzen des ursprünglichen Tests entsprechen und aber auf derselben Menge von Skills definiert sind. Nun wird das CDP auf die Kompetenzen der neuen Items angewandt. Dies erfolgt nur bezüglich der competency bags der Kompetenzzustandspaare, die mit dem ursprünglichen Test nicht unterscheidbar waren.

Im dritten Schritt werden die Kompetenzen des so entstandenen „Improvement reducts“ den Kompetenzen des ursprünglichen Tests hinzugefügt. Schritte zwei und drei sind somit der Improvement-Anteil des Vorgehens.

Im vierten Schritt wird das CDP auf alle möglichen Kompetenzzustandspaare und die neu entstandene Kompetenzsammlung angewandt. Schritt vier ist somit der Shortening-Anteil des Vorgehens. Letzten Endes erhält man mit diesem Vorgehen eine Kompetenzsammlung, die zu einem maximal informativen Test gemäß der Kompetenzstruktur und der verfügbaren Items bzw. Kompetenzen führen kann. Dieses so entstandene finale reduct wird von mir im Folgenden als „optimiertes reduct“ bezeichnet, zudem wird der hier beschriebene vollständige Optimierungsprozess als „Optimierung“ bezeichnet. Relevant zu erwähnen sei an dieser Stelle noch, dass es nötig ist den Improvement-Anteil vor dem Shortening-Anteil durchzuführen, weil bei einem inversen Vorgehen potentiell ein „optimiertes reduct“ entstehen könnte, welches noch weiter gekürzt werden könnte, da das Improvement am Ende Kompetenzen hinzugefügt hat, welche den Job von bereits vorliegenden Kompetenzen bezüglich des Unterscheidens noch mit erledigen. Das „optimierte reduct“ wäre folglich potentiell nicht einmal ein reduct. Bei dieser inversen Reihenfolge wäre es also nochmals nötig Shortening zu betreiben und da das weniger ökonomisch wäre, wird der vorgestellte Ablauf verwendet. Grundsätzlich könnte man auch statt der Optimierung auch einfach alle verfügbaren Kompetenzen sammeln und auf diese dann Shortening bezüglich aller Kompetenzzustandspaare anwenden. In diesem Fall sei von einem „Shortening auf alle Kompetenzen“ die Rede. Dieses Vorgehen hat jedoch folgende Probleme: Es würde nicht mehr dem Vorgehen aus Anselmi und Kollegen (2022, 2024b) entsprechen. Es wäre keine Methode mehr zur Optimierung eines bestimmten Tests, da alle möglichen Items zugleich und als gleichwertig betrachten würde. Es ist aufgrund der potentiell sehr großen Menge an Kompetenzen auch wegen einer sehr hohen benötigten Rechenleistung nicht immer anwendbar. Da dieser Arbeit diese drei Punkte wichtig sind, wird stattdessen die zuvor vorgestellte Methode verwendet. Allerdings gibt es in dieser Arbeit auch ein Beispiel für die „Shortening auf alle Kompetenzen“-Methode, dieses soll allerdings lediglich als Vergleichswert dienen.

2.4.5 Gesamter Optimierungsprozess auf Item-Ebene

Bis hierhin lief der gesamte Optimierungsprozess lediglich auf der Kompetenz-Ebene ab. Zwar wurde des Öfteren erwähnt, dass zu den Kompetenzen Items korrespondieren, allerdings ist diese Information nicht ausreichend für einen möglichst erfolgreichen Optimierungsprozess. Das grundlegende Problem an dieser Stelle ist nämlich, dass man, selbst wenn man nun weiß welche Kompetenzen man durch einen Test abfragen möchte nachdem man ihn einer Optimierung unterzogen hat, noch nicht weiß aus welchen Items man den Test nun zusammensetzen sollte. Das liegt daran, dass jede Kompetenz potentiell mit mehreren Items assoziiert ist, da mehrere Items genau dieselben Skills erfordern um korrekt gelöst zu werden. Wenn dieser Fall bei einer Anwendung nie auftritt, man also von jeder Kompetenz auf nur genau ein Item schließen kann, dann könnte man so direkt von der Kompetenz-Ebene auf die Item-Ebene verallgemeinern. Da dies aber im allgemeinen nicht der Fall ist, ist es nötig ein Vorgehen für diesen allgemeinen Fall zu entwickeln. In dieser Arbeit wird ein Vorgehen gewählt, welches sich Methoden aus der probabilistischen Wissensstrukturtheorie (Doignon & Falmagne, 1999) zunutze macht. Grundsätzlich sind probabilistische Wissensstrukturen so wie klassische Wissensstrukturen. Der Unterschied liegt darin, dass sie anstatt der deterministischen Annahme, dass sich der Wissenszustand einer Person direkt in ihrem Antwortmuster widerspiegelt, eine probabilistische Annahme machen. Diese ist, dass es für jedes Item eine Wahrscheinlichkeit dafür gibt das Item durch Zufall korrekt zu lösen („lucky guess“) und eine Wahrscheinlichkeit das Item trotz des nötigen Wissens durch einen Fehler falsch zu beantworten („careless error“). Außerdem gibt es in der probabilistischen Wissensstrukturtheorie noch die Möglichkeit den einzelnen Wissenszuständen einer Wissensstruktur jeweils eine Auftretenswahrscheinlichkeit zuzuweisen, dies ist für diese Arbeit jedoch irrelevant. Diese Wahrscheinlichkeitswerte können den Items über die Anwendung eines Modells der probabilistischen Wissensstrukturtheorie zugewiesen werden. Genauer wird hierfür klassischer Weise das „basic local independence model“ (BLIM) verwendet. Lucky guess- und careless error-Werte werden auch als die beiden Fehlerwahrscheinlichkeiten eines Items bezeichnet, denn hohe Werte in diesen beiden Wahrscheinlichkeitswerten bedeuten inhaltlich, dass das Item nicht optimal seine Aufgabe erfüllt das Wissen einer Person zu erfassen. Dies kann beispielsweise zu Stande kommen, wenn die Person entweder durch Raten leicht Wissen vorspielen kann, oder durch eine unklare Aufgabenstellung trotz vorliegendem Wissen eine Aufgabe oft falsch beantwortet. Wenn man die beiden Fehlerwahrscheinlichkeiten aufsummiert, erhält man die

„Fehlersumme“. Diese stellt einen Wert dar, der einen Vergleich mit Fehlersummen-Werten anderer Items erlaubt, um so inhaltlich zu überprüfen welches Item eher das tut was es soll, nämlich überprüfen ob eine Person tatsächlich über das nötige Wissen verfügt um ein Item korrekt zu beantworten. Diese Fehlersummen-Werte stellen nun für die Auswahl der bestmöglichen Items für eine Kompetenz ein sehr gutes Maß dar. Um das Ziel eines möglichst kurzen Tests zu erreichen, sollen nun die Items identischer Kompetenzen verglichen werden und so Items mit möglichst geringen Fehlersummenwerten gewählt werden. So entsteht letztendlich ein optimierter Test, der zu jeder Kompetenz aus dem „optimierten reduct“ genau ein Item beinhaltet. Abschließend sei noch ein Problem des BLIMs erwähnt, nämlich hat das BLIM im allgemeinen Identifizierbarkeitsprobleme bezüglich aller Wahrscheinlichkeitsparameter, die es schätzt. Mit anderen Worten gibt es eine Art trade-off zwischen den Wahrscheinlichkeitswerten für die lucky guess-, careless error- und den Zustandswahrscheinlichkeit der einzelnen Wissenszuständen. Das bedeutet, man kann nie sicher sein ob die einzelnen Parameter tatsächlich der Realität entsprechend geschätzt werden, oder ob einzelne dieser Wahrscheinlichkeitswerte zu hoch und andere dafür zu niedrig geschätzt werden. Dieses Problem ist sehr grundlegender Natur und schränkt im Normalfall den Einsatzbereich von BLIMs stark ein. In dem hier betrachteten Fall, gibt es dieses Problem jedoch mit an Sicherheit grenzender Wahrscheinlichkeit nicht, da Spoto, Stefanutti und Vidotto (2013) festgestellt haben, dass es beim Vergleich von „equally informative“ Items, welche eben solche Items sind, die derselben Kompetenz zugeordnet sind, keine Identifizierbarkeitsprobleme beim BLIM gibt. Diese Postulation wurde zwar in einem Paper von Heller (2017) abgeschwächt, aber nach wie vor gilt, dass es in dem hier betrachteten Fall mit an Sicherheit grenzender Wahrscheinlichkeit keine Identifizierbarkeitsprobleme bei den geschätzten BLIMs geben sollte.

3. Methode

3.1 Software

Für das Erstellen der Kompetenz- und Wissensstrukturen und auch die sonstige Auswertung wurde die Software „R“ (R Core Team, 2024) verwendet. In diesem Rahmen fanden die Zusatzpakete „pks“ (Heller & Wickelmaier, 2013), „relations“ (Meyer & Hornik, 2023), „foreign“ (R Core Team, 2023), „Rgraphviz“ (Hansen et al., 2023) und „IRdisplay“

(Kluyver, Angerer P. & Schulz J., 2022) Verwendung. Außerdem wurden digitale R-Skripte des Fachbereichs „Forschungsmethoden und Mathematische Psychologie“ der Universität Tübingen und aus der Bachelor Arbeit von Frederik Panse (2023) verwendet.

3.2 Stichprobe und Grundlagen

Der Großteil dieser Arbeit basiert nicht direkt auf einer Stichprobe, sondern lediglich auf der Kompetenzstruktur, den Skills und der Skillmap, welche aus dem Paper von Su und Kollegen (2013) entnommen wurden. Die Skills wurden durch eine Anpassung der Fähigkeiten für Schüler der sechsten bis achten Klasse im Bereich Mathematik vom CCSS 2010 (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) generiert. Die Skillmap und die Kompetenzstruktur wurden von zwei Doktoranden aus dem Bereich „secondary teaching and learning“ erstellt, welche beide zuvor Mathelehrer an amerikanischen middle und high schools waren. Mit diesen drei Grundlagen ließ sich der gesamte Optimierungsprozess auf Kompetenzebene ohne erhobene Daten durchführen. Da es für den Teil des Optimierungsprozesses, welcher die Ergebnisse von der Kompetenz- auf die Item-Ebene überführt, nötig ist für jedes der untersuchten Booklets ein BLIM zu schätzen, waren an dieser Stelle auch erhobene Daten nötig. Hierfür wurden wie auch schon bei Su und Kollegen (2013) und Anselmi (2024a) die Daten der US-Stichprobe der TIMSS 2003 verwendet. Die Stichprobengrößen sahen dabei wie folgt aus: Booklet 1: $n = 757$ Personen, Booklet 2: $n = 740$ Personen, Booklet 7: $n = 728$ Personen und Booklet 8: $n = 749$ Personen. Die Versuchspersonen waren allesamt Schüler und Schülerinnen der achten Schulklasse amerikanischer middle schools.

3.3 Material, Ablauf und Design

Der Prozess der Generierung der drei Grundlagen durch die beiden Doktoranden kann dem Abschnitt Q-Matrix aus dem Paper von Su und Kollegen (2013) entnommen werden. Eine genaue Beschreibung des Ablaufs der TIMSS 2003 kann dem Abschnitt „Appendix A Overview of TIMSS 2003 Procedures: Mathematics Achievement“ aus Mullis und Kollegen (2004) entnommen werden.

3.4 Optimierungsprozess: Kompetenz-Ebene

Der im Abschnitt „theoretische Grundlagen“ eingeführte Optimierungsprozess wurde auf jedes der untersuchten Booklets angewandt. Als Grundmenge der Kompetenzen wurden die Kompetenzen gewählt, die mit den Items korrespondierten, aus welchen sich das entsprechende Booklet zusammensetzte. Die „Improvement Kompetenzen“ waren stets Kompetenzen, die nicht den bereits verwendeten entsprachen und für die es Items in irgendeinem der anderen Booklets gab, die mit ihnen korrespondierten. Da Su und Kollegen (2013) nur die Booklets 1 und 2 analysiert haben, waren alle verfügbaren Items für den Optimierungsprozess stets eben die Items aus diesen zwei Booklets. Booklet 1 durchlief das Improvement also mit den Kompetenzen der Items aus Booklet 2 und Booklet 2 mit denen aus Booklet 1. Booklets 7 und 8 durchliefen das Improvement, da sie nur aus Items bestanden, die entweder aus Booklet 1 oder Booklet 2 stammten immer mit den Kompetenzen der Items, die sie bisher noch nicht beinhalteten. Es wurden einmal alle Booklets basierend auf der vollständigen Kompetenzstruktur und allen 15 Skills optimiert, aber dann wurden alle Booklets auch nochmal basierend auf folgender anderer Grundlage optimiert: Es wurde überprüft ob die Booklets in ihren ursprünglichen Zuständen dazu in der Lage waren Informationen über alle 15 Skills zu generieren, wobei sich bei jedem der untersuchten Booklets herausstellte, dass es immer mindestens einen der 15 Skills gibt, den das Booklet mit keinem seiner ursprünglichen Items erfasst. Daher wurde sich dafür entschieden die Booklets auch einmal nur anhand der Skills zu optimieren, über die ihre Items von Beginn an Aufschluss geben konnten. Hierfür wurden die entsprechenden Skills für diesen Optimierungsprozess nicht mehr berücksichtigt. Dadurch waren sie nichtmehr Teil der Skills. Die Skillmap wurde dahingehend angepasst, dass den „Improvement Kompetenzen“, die fürs Improvement genutzt wurden, nun keine Kompetenzen mehr zugewiesen wurden, die einen der ausgeschlossenen Skills beinhalteten. Stattdessen wurden die Skills einfach aus den zugewiesenen Kompetenzen gelöscht. So wurde beispielsweise in einem Fall, in welchem Skill „j“ nie im ursprünglichen Booklet abgefragt wurde, einem Item, welchem ursprünglich die Kompetenz {j, o} zugewiesen war nun die Kompetenz {o} zugewiesen. Zuletzt wurde der Skill auch aus der Kompetenzstruktur gelöscht, indem er aus allen möglichen Kompetenzzuständen gelöscht wurde und seine Voraussetzungen nun zu Voraussetzungen für die Skills wurden, für welche er zuvor eine Voraussetzung war. Der Gedanke hinter dieser Entscheidung war, dass man argumentieren könnte, dass das ursprüngliche Booklet nie dafür gedacht war Informationen über diese so gelöschten Skills zu generieren. Daher sollte seine

Optimierung auch nur anhand der Skills erfolgen, bei denen es auch dafür gedacht war Informationen über sie zu generieren. Zwar wurden die Booklets vollkommen unabhängig von den Skills erstellt, aber da die Skills in dieser Arbeit als Grundlage der Items angenommen werden, drängt sich diese Argumentation auf. Auf der anderen Seite wird mindestens implizit eine Äquivalenz zwischen den Booklets angenommen (Mullis et al., 2004; Martin et al., 2005), weshalb das vorherig beschriebene Vorgehen basierend auf allen Skills und der vollständigen Kompetenzstruktur ebenso sinnvoll ist. Zusätzlich zu diesen beiden Verfahren, wurden auch einmal alle verfügbaren Kompetenzen gesammelt und dem „Shortening auf alle Kompetenzen“-Prozess unterzogen, um das Ergebnis mit den Ergebnissen des Optimierungsprozesses anhand der kompletten Kompetenzstruktur und allen Skills zu vergleichen. Alles in allem wurden also insgesamt acht vollständige Optimierungsprozesse durchgeführt (zwei pro Booklet) und ein „Shortening auf alle Kompetenzen“-Prozess, was am Ende neun Kompetenzsammlungen ergab. Nun soll noch einmal der Optimierungsprozess in vier Schritten kurz beschrieben werden: Erstens wurden via δ mittels der Kompetenzstruktur und der Kompetenzen (jeweils entweder anhand der vollständigen oder der um die ursprünglich nicht erfassten Skills gekürzten) nicht unterscheidbare Kompetenzzustandspaare identifiziert. Zweitens wurden alle möglichen „Improvement Kompetenzen“ gesammelt. Mit ihnen wurde das CDP auf die competency bags angewendet, die nur die Kompetenzzustandspaare enthielten, die bisher nicht unterscheidbar waren. Drittens wurden die so entstandenen „Improvement reduces“ zu der Kompetenzsammlung des ursprünglichen Booklets hinzugefügt. Viertens wurde das CDP auf die neue, nun durch Improvement verlängerte, Kompetenzsammlung angewendet. So ergab sich ein optimiertes, also ein möglichst informatives, aber auch möglichst kurzes reduct. Diesen vierschrittigen Prozess durchlief jedes Booklet zweimal: Einmal für alle möglichen Skills mit vollständiger Skillmap und allen Skills. Dann noch einmal für die, um die initial nicht vom Booklet erfassten Skills, gekürzten Versionen von Kompetenzstruktur, Skillmap und Skills. So entstanden acht optimierte reduces. Das „Shortening auf alle Kompetenzen“ lief so ab, dass eine Kompetenzsammlung, bestehend aus allen verfügbaren unterschiedlichen Kompetenzen, den vierten Schritt des Optimierungsprozesses durchlief um so eine Art optimiertes reduct zu erhalten. Um später die Ergebnisse besser vergleichen zu können zeigt Tabelle 1 welche Kompetenzen ursprünglich in den einzelnen Booklets enthalten waren und gibt deren Anzahl an.

Tabelle 1

Kompetenzen, die vor der Optimierung in den einzelnen Booklets enthalten waren und deren Anzahl

Booklets	Kompetenzen	Anzahl von Kompetenzen
Booklet 1	$\{\{a\}, \{a, m, o\}, \{b\}, \{b, g\}, \{b, g, h, k\}, \{b, h\}, \{b, n\}, \{c\},$ $\{c, i, m, o\}, \{c, o\}, \{d\}, \{d, i\}, \{e\}, \{e, i\}, \{f\}, \{f, g\},$ $\{f, k, n\}, \{l\}, \{m\}\}$	19
Booklet 2	$\{\{a, b, g\}, \{a, b, h\}, \{a, f, g\}, \{b, f\}, \{b, g\}, \{b, h\}, \{b, n\},$ $\{c\}, \{c, g\}, \{c, i, m, o\}, \{c, o\}, \{d\}, \{d, i\}, \{e\}, \{e, f\}, \{f\},$ $\{f, i\}, \{f, k, n\}, \{g\}, \{i, j, o\}, \{j, o\}, \{n\}\}$	22
Booklet 7	$\{\{a, b, g\}, \{a, f, g\}, \{b, f\}, \{b, g\}, \{b, g, h, k\}, \{c\}, \{d\}, \{e\},$ $\{f, i\}, \{f, k, n\}, \{i, j, o\}, \{j, o\}, \{l\}\}$	13
Booklet 8	$\{\{a\}, \{b, g\}, \{b, h\}, \{c\}, \{c, g\}, \{c, o\}, \{e, f\}, \{f\}, \{m\}\}$	9

3.5 Optimierungsprozess: Item-Ebene

Um von den optimierten reducts auf Kompetenz-Ebene auf die optimierten Booklets auf Item-Ebene zu kommen, war es nötig für jedes der ursprünglichen Booklets ein BLIM zu schätzen. Nachdem die BLIMs geschätzt waren, wurden die careless error- und lucky guess-Parameter entnommen und für alle Items aus einem Booklet die Fehlersummen berechnet. Hierbei wurden zunächst alle Fehlersummen aller Items Booklet-spezifisch erstellt. So entstanden auch teilweise mehrere Fehlersummenwerte für ein und dasselbe Item, wenn dieses Item in mehr als nur einem Booklet auftauchte. Es wurde folgende Methode zur Auswahl des bestmöglichen Items für ein Booklet gewählt: Es wurde immer das Item einer Kompetenz gewählt, das in dem Booklet die niedrigste Fehlersumme aufwies. Es wurde immer ein Booklet nach dem anderen durchgegangen. Sollten Items identischer Kompetenz in anderen Booklets, aber auch im aktuell untersuchten Booklet, vorliegen, wurden die Vergleiche nur bezüglich der, im aktuell betrachteten Booklet vorliegenden, Items durchgeführt. Wenn eine Kompetenz mit mehreren Items assoziiert war, die alle nicht im

aktuell betrachteten Booklet vorkamen, also durch Improvement zum Booklet dazu kamen, dann wurde das Item gewählt, das über alle Booklets hinweg den geringsten Fehlersummenwert hatte. Für dieses unterschiedliche Vorgehen bezüglich Kompetenzen, die bereits von Beginn an im Booklet auftauchten, wurde sich entschieden, da nicht davon auszugehen war, dass ein Item über alle Booklets hinweg dieselben Fehlersumme zugewiesen bekommt. Das liegt daran, dass die anderen Items des Booklets, wie auch die Position des Items im Booklet die Schätzung der Fehlerparameter beeinflussen könnten. Durch einen Vergleich von ausschließlich den Items, die ohnehin schon im untersuchten Booklet vertreten waren, sollte die Wahrscheinlichkeit erhöht werden, dass das bestmögliche Item für das untersuchte Booklet ausgewählt wird. Letzten Endes sollten so für alle vier Booklets in ihren je zwei Ausführungen sowie dem einen „Shortening auf alle Kompetenzen“ optimierte Itemsammlungen mit genau so vielen Items wie Kompetenzen in den reducts entstehen. Hierbei sollten die Itemsammlungen stets aus den Items bestehen, welche möglichst gut für das entsprechende Booklet geeignet waren. Abschließend sei noch erwähnt, dass im Falle des „Shortenings auf alle Kompetenzen“ stets das Item mit der insgesamt niedrigsten Fehlersumme gewählt wurde, da dieses „optimierte reduct“ nicht auf einem Booklet basiert und daher hier kein Booklet-spezifischer Vergleich möglich war. Um später die Ergebnisse besser vergleichen zu können, zeigt Tabelle 2 aus welchen Items die einzelnen Booklets ursprünglich zusammengesetzt waren und gibt deren Anzahl an.

Tabelle 2

Items aus welchen sich die Booklets vor der Optimierung zusammengesetzt haben und deren Anzahl

Booklets	Items	Itemanzahl
Booklet 1	M012001, M012002, M012004, M012016, M012017, M012040, M012041, M012042, M022043, M022046, M022050, M022057, M022066, M022185, M022191, M022194, M022196, M022198, M022199, M022232, M022234B, M022251, M032142, M032163, M032198, M032570, M032640, M032643, M032755	29
Booklet 2	M012016, M012017, M012025, M012027, M012029, M022104, M022106, M022110, M022139, M022144, M022156, M022185, M022191, M022194, M022196, M022198, M022199, M022251, M022253, M032160, M032307, M032381, M032416, M032523, M032525, M032529, M032540, M032698, M032701, M032704	30
Booklet 7	M022104, M022106, M022110, M032142, M032163, M032198, M032307, M032523, M032525, M032640, M032701, M032704, M032755	13
Booklet 8	M022043, M022046, M022050, M022057, M022066, M022234B, M032160, M032381, M032416, M032529, M032540, M032698	12

3.6 Analyse der optimierten Tests

Zunächst wird betrachtet, wie sehr sich die Booklets durch den Optimierungsprozess verändert haben und was sich genau verändert hat. Hierfür werden zwei Aspekte besonders betrachtet. Erstens soll bei allen optimierten Booklets verglichen werden wie sich die Itemanzahl im Vergleich zu den ursprünglichen Booklets verändert hat. Zweitens werden aus allen verfügbaren „optimierten reducts“ wie aber auch den ursprünglichen Kompetenzsammlungen die zugehörigen Wissensstrukturen abgeleitet. Anhand dieser wird

ein Vergleich der Anzahl der Wissenszustände angestrebt um zu analysieren wie viel genauer die Booklets inzwischen die Wissenszustände von Personen erfassen können. Die Anzahl der Wissenszustände soll auch mit der Anzahl der Kompetenzzustände in der jeweiligen Kompetenzstruktur verglichen werden. So soll analysiert werden wieviel näher der Optimierungsprozess die Booklets an eine perfekte Aufklärung des Wissenszustandes einer Person gemäß der zugrunde gelegten Kompetenzstruktur gebracht hat. Daraufhin sollen die „optimierten reducts“, welche anhand der vollständigen Kompetenzstruktur, Skillmap und Skills erstellt wurden, sowie die Kompetenzsammlung, die durch „Shortening auf alle Kompetenzen“ entstand, miteinander verglichen werden. Dieser Vergleich soll zunächst bezüglich der Kompetenzen, die sie beinhalteten, erfolgen. Der Vergleich ist interessant, da diese „optimierten reducts“ alle basierend auf denselben Kompetenzen optimiert wurden. Dadurch können sie alle zwischen denselben Kompetenzzuständen unterscheiden und induzieren alle dieselbe Wissensstruktur. Man könnte also erwarten, dass sie auch dieselben Kompetenzen beinhalten, aber vielleicht ist dem nicht so, da sie von unterschiedlichen Ausgangspunkten (entweder einem der Booklets oder auch von allen verfügbaren Kompetenzen) starten. Zusätzlich zu dem Vergleich auf der Kompetenz-Ebene soll auch ein Vergleich nach einem ähnlichen Prinzip auf der Item-Ebene erfolgen.

Zuletzt sollte noch betrachtet werden ob es möglich wäre, und wenn ja mit welchen zusätzlichen Items, die einzelnen Ergebnisse zu fully informative Tests zu verbessern. Hierfür sollten zunächst die Ergebnisse, welche alle Skills berücksichtigen betrachtet werden. Danach sollten die Ergebnisse, die nur die Skills berücksichtigten, zu deren Erfassung von vorn herein Items in den Booklets waren, betrachtet werden. Grundsätzlich wurde eine sehr basale Methode zu diesem Zweck gewählt. Um diese zu verstehen ist es zunächst relevant zu wissen, dass ein Test immer zwangsläufig fully informative ist, wenn er für jeden Skill aus der Kompetenzstruktur ein Item hat, welches explizit nur diesen Skill und keine anderen abfragt. Folglich wäre diesem Item nur dieser Skill als Kompetenz zugewiesen. Das liegt daran, dass sich so jeder Skill einzeln erfassen lassen würde und somit auch jede möglich Kombination von Skills erfasst werden könnte. In der Realität ist dies oft nicht möglich, da es nicht möglich ist ein Item zu entwerfen, das nur einen Skill einer Menge an Skills abfragt, ohne auch andere Skills derselben Menge von Skills mitabzufragen. In unserem Fall habe ich daher zunächst alle Skills identifiziert, zu welchen noch kein Item mit einer Kompetenz, welches nur sie allein abfragt, vorlag. Die Skills auf welche dies zutraf waren die Skills „h“, „i“, „j“, „k“ und „o“. Im nächsten Schritt wurde das CDP angewandt auf diese Skills als Menge von

Kompetenzen \mathcal{T} und eine discernibility function δ , welche nur die Kompetenzzustandspaare betrachtete, die mit den anhand aller Skills optimierten Booklets noch nicht unterschieden werden konnten. Für die Analyse bezüglich der Ergebnisse, die nur die Skills berücksichtigten, zu deren Erfassung von vorn herein Items in den Booklets waren, wurde Folgendes angenommen: Die Kompetenzstrukturen dieser Booklets sind insofern Teilmengen der Ergebnisse, welche alle Skills berücksichtigen, dass sie nur Kompetenzzustände beinhalten, die auch die übergeordnete Kompetenzstruktur mit allen Skills beinhaltet. Daher ist es sicher, dass eine Hinzunahme von Items, die den Kompetenzen zugeordnet sind, die bei der letzten Untersuchung herauskamen, genügen würde um sie zu fully informative Tests zu verbessern. Allerdings könnte dies potentiell auch mit weniger Items gelingen. Es wird dieselbe Methode wie zuvor angewandt, indem das CDP auf die Paare von Kompetenzzuständen angewandt wird, die das jeweilige Booklet noch nicht unterscheiden kann. Die Kompetenzen, die für die Booklets, welche alle Skills berücksichtigen, als Ergebnis herausgekommen waren, wurden hier als Kompetenzsammlung \mathcal{T} für das CDP verwendet.

4. Ergebnisse

4.1 Optimierungsergebnisse: Kompetenz-Ebene

Zuerst wurden die Booklets nur bezüglich der Skills, zu welchen sie ursprünglich auch Items mit entsprechenden Kompetenzen beinhalteten, optimiert. Tabelle 3 zeigt die Ergebnisse dieses Optimierungsprozesses, indem sie die bezüglich Informationsgehalt und Kompetenzanzahl „optimierten reducts“ und die Anzahl der Kompetenzen, die sie beinhalten für jedes Booklet darlegt.

Tabelle 3

„Optimierte reducts“ des Optimierungsprozesses der Booklets mit nur den Skills, zu welchen die Booklets ursprünglich auch Items mit Kompetenzen, die Sie erfassen konnten, beinhalteten und die Anzahl der Kompetenzen in den „optimierten reducts“

Booklets	Kompetenzen	Anzahl von Kompetenzen
Booklet 1 ohne Skill „j“	{a}, {b}, {b, g}, {b, h}, {c}, {d}, {d, i}, {e}, {e, i}, {f}, {f, k, n}, {i, o}, {l}, {m}, {n}, {o}}	16
Booklet 2 ohne Skill „l“	{a}, {a, m, o}, {b}, {b, g}, {b, h}, {c}, {c, i, m, o}, {c, o}, {d}, {d, i}, {e}, {e, i}, {f}, {f, k, n}, {i, j, o}, {j, o}, {m}, {n}}	18
Booklet 7 ohne Skill „m“	{a}, {a, b, h}, {a, o}, {b}, {b, g}, {c}, {c, i, o}, {c, o}, {d}, {d, i}, {e}, {e, i}, {f}, {f, k, n}, {i, j, o}, {j, o}, {l}, {n}}	18
Booklet 8 ohne Skills „d“, „i“, „j“, „k“, „l“, „n“	{a}, {b}, {b, g}, {b, h}, {c}, {e}, {e, f}, {m}, {o}}	9

Danach wurden die Booklets bezüglich aller Skills optimiert. Tabelle 4 zeigt die Ergebnisse dieses Optimierungsprozesses. Hier werden die bezüglich Informationsgehalt und Kompetenzanzahl „optimierten reducts“ und die Anzahl der Kompetenzen, die sie beinhalten für jedes Booklet darlegt. Zudem stellt Tabelle 4 diese Ergebnisse auch bezüglich des „Shortenings auf alle Kompetenzen“ dar.

Tabelle 4

„Optimierte reducts“ des Optimierungsprozesses der Booklets mit allen Skills sowie das Ergebnis des „Shortenings auf alle Kompetenzen“ und die Anzahl der Kompetenzen in den „optimierten reducts“

Booklets	Kompetenzen	Anzahl von Kompetenzen
Booklet 1 mit allen Skills	{a}, {a, m, o}, {b}, {b, g}, {b, h}, {c}, {c, i, m, o}, {c, o}, {d}, {d, i}, {e}, {e, i}, {f}, {f, k, n}, {i, j, o}, {j, o}, {l}, {m}, {n}}	19
Booklet 2 mit allen Skills	{a}, {a, b, h}, {a, m, o}, {b}, {c}, {c, i, m, o}, {c, o}, {d}, {d, i}, {e}, {e, i}, {f}, {f, k, n}, {g}, {i, j, o}, {j, o}, {l}, {m}, {n}}	19
Booklet 7 mit allen Skills	{a}, {a, b, h}, {a, m, o}, {b}, {b, g}, {c}, {c, i, m, o}, {c, o}, {d}, {d, i}, {e}, {e, i}, {f}, {f, k, n}, {i, j, o}, {j, o}, {l}, {m}, {n}}	19
Booklet 8 mit allen Skills	{a}, {a, m, o}, {b}, {b, g}, {b, h}, {c}, {c, i, m, o}, {c, o}, {d}, {d, i}, {e}, {e, i}, {f}, {f, k, n}, {i, j, o}, {j, o}, {l}, {m}, {n}}	19
„Shortening auf alle Kompetenzen“	{a}, {a, m, o}, {b}, {b, h}, {c}, {c, i, m, o}, {c, o}, {d}, {d, i}, {e}, {e, i}, {f}, {f, k, n}, {g}, {i, j, o}, {j, o}, {l}, {m}, {n}}	19

An dieser Stelle sei noch erwähnt, dass obwohl für die Auswahl der zu löschenden Kompetenzen beim CDP stets die MBP-Methode gewählt wurde, die Ergebnisse nie eindeutig waren. Das bedeutet genauso gut wie die Kompetenzsammlungen aus Tabellen 3 und 4, hätten auch andere Kompetenzsammlungen als Ergebnisse herauskommen können. Es gab des Öfteren mehrere Kompetenzen zum Zeitpunkt der Entscheidung, welche als nächstes beim CDP gelöscht werden könnten, weil sie zugleich in der geringsten Anzahl von

competency bags enthalten waren. Folglich musste die zu löschende Kompetenz zufällig gewählt werden, weshalb die Ergebnisse bei erneuter Durchführung des CDP potentiell anders aussehen würden. Dies gilt gemäß einer Testung meinerseits für alle Kompetenzsammlungen aus Tabellen 3 und 4. Die alternativen Ergebnisse bestanden jedoch stets aus gleichvielen Kompetenzen wie die aus den Tabellen 3 und 4 und ähnelten ihnen auch bezüglich der enthaltenen Kompetenzen.

4.2 Optimierungsergebnisse: Item-Ebene

Daraufhin wurde für jedes der Booklets ein BLIM geschätzt, welche Schätzungen für die Fehlerparameter jedes der enthaltenen Items ermöglichen. Tabellen A1 – A4 des Appendix zeigen die Ergebnisse dieser BLIMs. Pro Tabelle wird ein Booklet abgedeckt indem zu jedem der Items des Booklets seine beiden Fehlerparameter, die Fehlersumme und die Kompetenz, der das Item zugeordnet ist, angegeben ist. Hierbei beinhaltet Tabelle A1 die Ergebnisse von Booklet 1, Tabelle A2 die von Booklet 2, Tabelle A3 die von Booklet 7 und Tabelle A4 die Ergebnisse von Booklet 8.

Aus den Tabellen A1 – A4 wurden die Fehlersummenwerte aller Items entnommen, welchen Kompetenzen zugewiesen sind, die mehr als einem Item zugewiesen ist. Daraufhin erfolgten Kompetenz für Kompetenz möglichst Booklet-spezifische Itemvergleiche. Die Ergebnisse dieser Vergleiche sollten dann Booklet-spezifisch für die Auswahl optimaler Items für jede Kompetenz genutzt werden. Die zu vergleichenden Items und deren Fehlersummen anhand derer sie verglichen werden, können Tabellen A5-A20 des Appendix entnommen werden.

Die Booklet-spezifischen Vergleiche ermöglichen mithilfe der optimierten Kompetenzen die Auswahl der bestmöglichen Items für jedes Booklet nach dem oben beschriebenen Vorgehen. Dementsprechend wurden neue, optimierte Booklets erstellt, welche auch neue Items beinhalten. Wie auch schon oben, sollen im Folgenden zunächst die Ergebnisse für die Booklets mit nur den Skills, zu welchen sie ursprünglich entsprechende Items innehatten, dargestellt werden. Tabelle 5 zeigt die Ergebnisse des finalen Optimierungsprozesses, indem sie die bezüglich Informationsgehalt, Itemart und Itemanzahl optimierten Booklets und die Anzahl der Items, die sie beinhalten für jedes Booklet darlegt.

Tabelle 5

Bezüglich Informationsgehalt, Itemart und Itemanzahl optimierte Booklets mit nur den Skills, zu welchen sie ursprünglich auch Items mit Kompetenzen, die sie erfassen konnten, innehatten und die Anzahl der Items, die sie beinhalten

Booklets	Items	Anzahl von Items
Booklet 1 ohne Skill „j“	M012004, M012017, M012041, M012042, M022043, M022046, M022057, M022144, M022191, M022196, M022198, M022232, M032198, M032307, M032523, M032643	16
Booklet 2 ohne Skill „l“	M012001, M012016, M012017, M012042, M022043, M022104, M022110, M022144, M022191, M022196, M022199, M022232, M022234B, M032160, M032307, M032416, M032523, M032529	18
Booklet 7 ohne Skill „m“	M012001, M012027, M012042, M022050, M022066, M022104, M022110, M022144, M022199, M022232, M022234B, M032142, M032163, M032198, M032307, M032523, M032640, M032643	18
Booklet 8 ohne Skills „d“, „i“, „j“, „k“, „l“, „n“	M022043, M022046, M022232, M022234B, M032160, M032163, M032307, M032529, M032698	9

Wie auch schon oben, sollen daraufhin die Ergebnisse für die Booklets nach Optimierung anhand aller Skills dargestellt werden. Tabelle 6 zeigt die Ergebnisse des finalen Optimierungsprozesses, indem sie die bezüglich Informationsgehalt, Itemart und Itemanzahl optimierten Booklets und die Anzahl der Items, die sie beinhalten für jedes Booklet darlegt. Zudem stellt Tabelle 6 diese Ergebnisse auch bezüglich des „Shortenings auf alle Kompetenzen“ dar.

Tabelle 6

Bezüglich Informationsgehalt, Itemart und Itemanzahl optimierte Booklets mit allen Skills und die Anzahl der Items, die sie beinhalten. Selbige Ergebnisse sind auch für das „Shortenings auf alle Kompetenzen“ angegeben

Booklets	Items	Anzahl von Items
Booklet 1 mit allen Skills	M012001, M012004, M012017, M012041, M012042, M022043, M022046, M022057, M022066, M022144, M022191, M022196, M022198, M022199, M022232, M032198, M032307, M032523, M032643	19
Booklet 2 mit allen Skills	M012001, M012016, M012017, M012025, M012027, M012042, M022043, M022104, M022110, M022144, M022191, M022196, M022199, M022232, M022234B, M032198, M032307, M032416, M032523	19
Booklet 7 mit allen Skills	M012001, M012027, M012042, M022043, M022050, M022066, M022104, M022110, M022144, M022199, M022232, M022234B, M032142, M032163, M032198, M032307, M032523, M032640, M032643	19
Booklet 8 mit allen Skills	M012001, M012017, M012042, M022043, M022046, M022050, M022066, M022104, M022144, M022199, M022232, M022234B, M032160, M032163, M032198, M032307, M032523, M032529, M032643	19

„Shortening auf alle Kompetenzen“	M012001, M012017, M012025, M012042, M022043, M022050, M022066, M022104, M022110, M022144, M022199, M022232, M022234B, M032163, M032198, M032307, M032523, M032529, M032643	19
--------------------------------------	--	----

4.3 Optimierungsergebnisse: Vergleich der optimierten Booklets

Für den Vergleich der optimierten Booklets, welche alle Skills berücksichtigen, wurden zwei Tabellen erstellt, eine für den Vergleich auf Kompetenz- und eine für den Vergleich auf Item-Ebene. Um Kompetenzen, die in allen Optimierungsergebnissen auftauchen zu identifizieren, sollen nun alle Kompetenzen aus Tabelle 4, in zwei Kategorien eingeteilt werden: Eine Kategorie von den Kompetenzen, die in jedem Optimierungsergebnis auftauchen und eine Restkategorie. Die Ergebnisse dieser Kategorisierung können Tabelle 7 entnommen werden.

Tabelle 7

Kategorisierung aller Kompetenzen aus denen sich die Optimierungsergebnisse zusammensetzen: Eine Kategorie von Kompetenzen, die in jedem Optimierungsergebnis auftauchen und eine Restkategorie

In allen Ergebnissen	Restkategorie
{a}, {a, m, o}, {b}, {c}, {c, i, m, o}, {c, o}, {d}, {d, i}, {e}, {e, i}, {f}, {f, k, n}, {i, j, o}, {j, o}, {l}, {m}, {n}	{b, g}, {b, h}, {a, b, h}, {g}

Um Items, die in allen Optimierungsergebnissen auftauchen zu identifizieren, sollen nun alle Items aus Tabelle 6, in drei Kategorien eingeteilt werden. Zusätzlich zu den beiden Kategorien, die es auch schon bei der Kategorisierung in Tabelle 7 gab, gibt es hier nun noch eine Kategorie für Items, die in allen außer einem Ergebnis auftauchen. Hierfür wurde sich entschieden um genauere Aussagen bezüglich der Ähnlichkeiten treffen zu können. Die Kategorisierung kann Tabelle 8 entnommen werden.

Tabelle 7

Kategorisierung aller Items aus denen sich die Optimierungsergebnisse zusammensetzen: Eine Kategorie von Items, die in jedem Optimierungsergebnis auftauchen, eine Kategorie von Items, die in jedem außer einem der Optimierungsergebnisse auftauchen und eine Restkategorie

In allen Ergebnissen	In allen Ergebnissen außer einem	Restkategorie
M012001, M012042, M022043, M022144, M022199, M022232, M032198, M032307, M032523	M012017, M022066, M022104, M022234B, M032643	M012004, M012016, M012025, M012027, M012041, M022046, M022050, M022057, M022110, M022191, M022196, M022198, M032142, M032160, M032163, M032416, M032529, M032640

4.4 Optimierungsergebnisse: Kompetenz-/ Wissensstruktur-Ebene

Im nächsten Schritt soll die Anzahl der Wissenszustände, die die einzelnen Booklets erfassen können, vor und nach der Optimierung verglichen werden. Außerdem sollte die Anzahl der Wissenszustände mit der Anzahl der Kompetenzzustände der jeweiligen Kompetenzstruktur verglichen werden. Die Anzahl der Kompetenzzustände in den Kompetenzstrukturen der ursprünglichen Booklets, sowie die Anzahl der Kompetenzzustände der allgemeinen Kompetenzstruktur können Tabelle 9 entnommen werden. Die Anzahl der Wissenszustände der abgeleiteten Wissensstrukturen vor und nach der Optimierung können Tabelle 10 entnommen werden.

Tabelle 9

Anzahl der Kompetenzzustände in den Kompetenzstrukturen der ursprünglichen Booklets und Anzahl der Kompetenzzustände der allgemeinen Kompetenzstruktur

Booklets	Anzahl Kompetenzzustände in Kompetenzstruktur
Booklet 1 ohne Skill „j“	908
Booklet 2 ohne Skill „l“	860
Booklet 7 ohne Skill „m“	880
Booklet 8 ohne Skills „d“, „i“, „j“, „k“, „l“ und „n“	120
Kompetenzstruktur die alle Skills berücksichtigt	1280

Tabelle 10

Anzahl der Wissenszustände der abgeleiteten Wissensstrukturen vor und nach der Optimierung

Booklets	Anzahl Wissenszustände der abgeleiteten Wissensstruktur	
	Vor Optimierung	Nach Optimierung
Booklet 1 ohne Skill „j“	744	884
Booklet 2 ohne Skill „l“	329	640
Booklet 7 ohne Skill „m“	247	726
Booklet 8 ohne Skills „d“, „i“, „j“, „k“, „l“ und „n“	48	120
Booklet 1 mit allen Skills	744	1022
Booklet 2 mit allen Skills	329	1022

Booklet 7 mit allen Skills	247	1022
Booklet 8 mit allen Skills	48	1022
„Shortening auf alle Kompetenzen“	-	1022

4.5 Optimierungsergebnisse: Fully informative Verbesserung

Für die Ergebnisse, welche sich durch die Optimierung unter Berücksichtigung aller Skills ergeben hatten, ergab eine Anwendung des CDP folgendes Ergebnis: Eine zeitgleiche Hinzunahme von Items, denen die Kompetenzen $\{i\}$, $\{j\}$ und $\{o\}$ zugeordnet sind, würde diese Ergebnisse zu fully informative Tests machen. Danach erfolgte eine Anwendung auf die Ergebnisse, welche sich durch die Optimierung unter Berücksichtigung der Skills, zu denen die Booklets ursprünglich entsprechende Items hatten, ergeben hatten. Hier wurde das CDP auf die Ergebnisse von zuvor angewandt und es ergaben sich die folgenden Ergebnisse: Das optimierte Booklet 1, welches Skill „j“ nicht in seiner Kompetenzstruktur berücksichtigt, benötigt nur ein Item. Dieses müsste als Kompetenz $\{i\}$ haben um das optimierte Booklet zu einem fully informative Test zu verbessern. Booklets 2, welches Skill „l“ nicht in seiner Kompetenzstruktur berücksichtigt, sowie Booklet 7, welches Skill „m“ nicht in seinem Kompetenzstruktur berücksichtigt, benötigen beide alle drei neuen Items. Sie benötigen also eines mit Kompetenz $\{i\}$, eines mit Kompetenz $\{j\}$ und eines mit Kompetenz $\{o\}$. Booklet 8, welches die Skills „d“, „i“, „j“, „k“, „l“ und „n“ nicht berücksichtigt, benötigt keine weiteren Items, da es bereits ein fully informative Test ist.

5. Diskussion

Das Ziel dieser Abschlussarbeit war es basierend auf Items und Booklets des TIMSS 2003 Tests zu erstellen, die eine ökonomische Wissensdiagnose in den Bereichen Number und Algebra ermöglichen sollten. Hierfür wurden Methoden aus der kompetenzbasierten Wissensstrukturtheorie verwendet. Für die Erreichung dieses Ziels wurden Ergebnisse erstellt. Die Ergebnisse lassen einige interessante Schlüsse zu, sowohl bezüglich der Optimierung der TIMSS 2003, als auch der Testoptimierung im Allgemeinen. Ebenso ergeben sich relevante

Schlussfolgerungen in Bezug auf die theoretischen Grundlagen, welche der Testoptimierung im Rahmen der kompetenzbasierten Wissensstrukturtheorie zugrunde liegen. Zuletzt zeigen sich auch relevante Punkte in Bezug auf die Annahmen, auf denen die Zuweisung von Kompetenzen zu Items basiert. Im Folgenden werde ich zunächst allgemein einige der oben genannten Ergebnisse interpretieren. Danach werde ich auf einige Ergebnisse genauer eingehen. Ich werde zuerst einen Vergleich der anhand aller Skills optimierten Booklets anstreben. Daraufhin werde ich mich mit der Frage beschäftigen, ob man die untersuchten Booklets durch das Dazunehmen einzelner Items zu fully informative Tests aufrüsten könnte und wenn ja welche Items dies wären. Zum Abschluss werde ich dann die Limitationen dieser Arbeit beleuchten und den Verdienst dieser Arbeit elaborieren.

5.1 Interpretation der Ergebnisse

Zuerst soll es um die Kompetenzen der Booklets ohne Berücksichtigung der Skills gehen, die das jeweilige Booklet vor der Optimierung nicht abgefragt hat. Daraufhin soll der allgemeine Fall mit allen Skills betrachtet werden. Bei der Interpretation sollen stets zunächst die Optimierungsergebnisse auf Kompetenz-Ebene betrachtet werden, dann soll die Item-Ebene behandelt werden und zuletzt soll die Kompetenz- und Wissensstrukturebene betrachtet werden.

Vor der Optimierung umfasste Booklet 1 19, Booklet 2 22, Booklet 7 13 und Booklet 8 9 Kompetenzen (Tabelle 1). Nach der Optimierung umfasst Booklet 1 16, Booklet 2 18, Booklet 7 18 und Booklet 8 9 Kompetenzen (Tabelle 3). Durch den Prozess wurde die Kompetenzanzahl der Booklets 1 und 2 also reduziert, die von Booklet 7 erhöht und die von Booklet 8 blieb gleich. Vor der Optimierung umfasste Booklet 1 29, Booklet 2 30, Booklet 7 13 und Booklet 8 12 Items (Tabelle 2). Nach der Optimierung umfasste Booklet 1 16, Booklet 2 18, Booklet 7 18 und Booklet 8 9 Items (Tabelle 5). Durch den Prozess wurde die Itemanzahl der Booklets 1, 2 und 8 also reduziert und die von Booklet 7 wurde erhöht. Alles in allem konnten also drei von vier Booklets, insofern sie weiterhin nur Aufschluss über die Skills geben, welche sie von vorn herein auch abfragen konnten bezüglich ihrer Itemanzahl gekürzt werden. Gleichzeitig sind sie durch das Improvement nun besser dafür geeignet die Wissenszustände von Testpersonen aus der zugehörigen Kompetenzstruktur abzuleiten. Booklet 7 ist als einziges in der Kompetenz- und Itemanzahl gewachsen. Dies ist auch darauf zurückzuführen, dass es, obwohl es von vorn herein aus weniger als halb so vielen Items

bestand wie die Booklets 1 und 2, dennoch genauso viele Skills abgefragt hat wie die beiden. Booklet 8 hat beispielsweise mit ähnlich vielen Items wie Booklet 7 fünf Skills weniger abgefragt als Booklets 1 und 2. Das hat es sehr viel wahrscheinlicher gemacht, dass Kompetenzen und so auch Items dazukommen, als bei Booklet 8. Aber dafür kann Booklet 7 nach der Optimierung eine sehr viel größere Wissensstruktur erfassen als Booklet 8 und folglich mehr über die Wissenszustände der Testpersonen aussagen.

Nun soll betrachtet werden wie stark eben diese Unterschiede bei allen Booklets sind. Die maximale Anzahl von Wissenszuständen, die die Wissensstruktur unter der jeweiligen vorgegebenen Kompetenzstruktur haben kann, ist gleich der Anzahl der Kompetenzzustände in dieser Kompetenzstruktur (Anselmi et al., 2022). Die Kompetenzstruktur von Booklet 1, welches zu Beginn Items mit zugeordneten Kompetenzen beinhaltete, von denen je mindestens eine einen von allen Skills außer Skill „j“ beinhaltete, hat 908 Kompetenzzustände (Tabelle 9). Von diesen 908 theoretisch möglichen Wissenszuständen konnte Booklet 1 vor der Optimierung 744 aus der Kompetenzstruktur anhand der Items des Booklets ableiten, nach der Optimierung sind es nun 884 (Tabelle 10). Die Kompetenzstruktur von Booklet 2, welches zu Beginn Items mit zugeordneten Kompetenzen beinhaltete, von denen je mindestens eine einen von allen Skills außer Skill „l“ beinhaltete, hat 860 Kompetenzzustände (Tabelle 9). Von diesen 860 theoretisch möglichen Wissenszuständen konnte Booklet 2 vor der Optimierung 329 aus der Kompetenzstruktur anhand der Items des Booklets ableiten, nach der Optimierung sind es nun 640 (Tabelle 10). Die Kompetenzstruktur von Booklet 7, welches zu Beginn Items mit zugeordneten Kompetenzen beinhaltete, von denen je mindestens eine einen von allen Skills außer Skill „m“ beinhaltete, hat 880 Kompetenzzustände (Tabelle 9). Von diesen 880 theoretisch möglichen Wissenszuständen konnte Booklet 7 vor der Optimierung 247 aus der Kompetenzstruktur anhand der Items des Booklets ableiten, nach der Optimierung sind es nun 726 (Tabelle 10). Die Kompetenzstruktur von Booklet 8, welches zu Beginn Items mit zugeordneten Kompetenzen beinhaltete, von denen je mindestens eine einen von allen Skills außer den Skills „d“, „i“, „j“, „k“, „l“ und „n“ beinhaltete, hat 120 Kompetenzzustände (Tabelle 9). Von diesen 120 theoretisch möglichen Wissenszuständen konnte Booklet 8 vor der Optimierung 48 aus der Kompetenzstruktur anhand der Items des Booklets ableiten, nach der Optimierung sind es nun 120 (Tabelle 10).

Jedes Booklet sticht nach der Optimierung durch andere besonders positive Auffälligkeiten heraus, auf die im Folgenden genauer eingegangen werden soll. Booklet 1 ist das am stärksten gekürzte Booklet mit nun 13 Items weniger als zuvor und trotz dieser starken Kürzung kann es nun 140 Wissenszustände mehr erfassen als zuvor. Booklet 2 ist das Booklet, welches die stärkste Kombination aus starker Kürzung und großem Zuwachs an aus seiner Kompetenzstruktur ableitbaren Wissenszuständen darstellt. Es ist um 12 Items kürzer und kann 311 Wissenszustände mehr erfassen als zuvor. Booklet 7 ist das Booklet, welches die meisten neuen Wissenszustände erfassen kann, es ist zwar um fünf Items länger als zuvor, kann dafür aber nun auch 479 neue Wissenszustände erfassen. Booklet 8 ist durch die Optimierung fully informative geworden und kann nun alle 120 möglichen Wissenszustände erfassen und das mit drei Items weniger als zuvor. Alle Booklets sind more informative geworden durch die Optimierung und die meisten zusätzlich sogar noch kürzer, der Prozess war damit ein Erfolg.

Wenn wir nun die optimierten Booklets mit allen Skills als Grundlage betrachten, sowie das „Shortening auf alle Kompetenzen“, dann zeichnet sich ein ganz ähnliches Bild ab. Vor der Optimierung sah alles genauso aus wie auch im Fall zuvor: Booklet 1 umfasste 19, Booklet 2 22, Booklet 7 13 und Booklet 8 9 Kompetenzen (Tabelle 1). Nach der Optimierung umfasste Booklet 1 19, Booklet 2 19, Booklet 7 19 und Booklet 8 auch 19 Kompetenzen (Tabelle 4). Zudem kam noch das „Shortening auf alle Kompetenzen“ hinzu, welches ebenfalls als Ergebnis aus 19 Kompetenzen bestand. Da alle Booklets anhand derselben Kompetenzen optimiert wurden und das CDP dank der MBP-Methode tendenziell möglichst kurze reducts als Ergebnisse ausgab, erfassen die Booklets nun dieselbe Anzahl von Kompetenzen. Durch den Optimierungsprozess wurde die Kompetenzanzahl der Booklets 1 und 2 reduziert und die von Booklets 7 und 8 erhöht. Vor der Optimierung umfasste Booklet 1 29, Booklet 2 30, Booklet 7 13 und Booklet 8 12 Items (Tabelle 2). Nach der Optimierung umfasste Booklet 1 19, Booklet 2 19, Booklet 7 19 und Booklet 8 19 Items. Auch das „Shortening auf alle Kompetenzen“, führte zu einem Ergebnis bestehend aus 19 Items (Tabelle 6).

Die Itemanzahlen entsprechen den Kompetenzanzahlen, da jede Kompetenz durch exakt ein Item abgefragt werden soll, dementsprechend bestehen auch alle Booklets nun auch alle aus derselben Anzahl von Items. Durch den Optimierungsprozess wurde die Itemanzahl der Booklets 1 und 2 reduziert und die Itemanzahl der Booklets 7 und 8 wurde erhöht. Alles

in allem konnten also zwei von vier Booklets, insofern alle Skills berücksichtigt wurden, bezüglich ihrer Itemanzahl gekürzt werden, während sie gleichzeitig durch das Improvement nun besser dafür geeignet sind die Wissenszustände von Testpersonen möglichst präzise zu erfassen. Booklets 7 und 8 sind in ihrer Kompetenz- und Itemanzahl gewachsen. Das ist auch darauf zurückzuführen, dass sie von vornherein aus weniger als halb so vielen Items bestanden wie Booklets 1 und 2 und nun nach der Optimierung aber über dieselbe Menge an Skills Aufschluss geben können sollten. Insbesondere bei Booklet 8 zeigt sich dieser Zuwachs an untersuchten Skills auf eine auffällige Weise.

Nun soll betrachtet werden wie stark die Unterschiede bezüglich der abgeleiteten Wissensstrukturen basierend auf der einheitlichen Kompetenzstruktur bei allen Booklets wie auch beim „Shortening auf alle Kompetenzen“ sind. Die maximale Anzahl von Wissenszuständen, die die Wissensstruktur haben kann, die von der Kompetenzstruktur mit allen Skills abgeleitet wurde, ist gleich der Anzahl der Kompetenzzustände in dieser Kompetenzstruktur (Anselmi et al., 2022). Die Kompetenzstruktur aller Booklets ist hier dieselbe, da alle Skills zugleich betrachtet werden. Diese Kompetenzstruktur hat 1280 Kompetenzzustände (Tabelle 9). Von den 1280 theoretisch möglichen Wissenszuständen einer korrespondierenden Wissensstruktur konnte Booklet 1 vor der Optimierung 744 anhand der Items des Booklets erfassen, Booklet 2 konnte 329, Booklet 7 247 und Booklet 8 48 erfassen (Tabelle 10). Diese Zahlen decken sich mit denen aus dem vorherigen Abschnitt, denn nur weil eine größere Kompetenzstruktur betrachtet wird, können die Booklets nicht plötzlich mehr Wissenszustände erfassen. Stattdessen sehen die abgeleiteten Wissensstrukturen nun kleiner im Vergleich zu der nun größeren Kompetenzstruktur aus. Es ist hier relevant im Gedächtnis zu behalten, dass die neuen Kompetenzzustände der nun größeren Kompetenzstruktur sich bei jedem Booklet durch Skills auszeichnen, zu welchen das Booklet vor der Optimierung ohnehin kein Item mit einer passenden Kompetenz zugeordnet hatte. Nach der Optimierung konnten nun alle Booklets und auch die Kompetenzsammlung, die durch das „Shortening auf alle Kompetenzen“ entstanden war, von der Kompetenzstruktur eine Wissensstruktur mit 1022 Wissenszuständen ableiten (Tabelle 10). Diese Wissensstruktur war auch bei allen Ergebnissen exakt dieselbe, was daraus folgt, dass stets anhand derselben verfügbaren Kompetenzen optimiert wurde. Die Steigerung in der Anzahl von aus der zugehörigen Kompetenzstruktur ableitbaren Wissenszuständen ist bei jedem Booklet bemerkenswert. Bei Booklets 1 und 2 ist dies besonders auffällig dadurch, dass sie um 10, respektive 11 Items gekürzt wurden und dennoch nun eine Wissensstruktur mit

wesentlich mehr Wissenszuständen (278 respektive 693) erfassen können. Booklets 7 und 8 zeichnen sich wiederum besonders durch ihre extremen Sprünge bezüglich der Menge an Wissenszuständen, welche sie nun von der zugehörigen Kompetenzstruktur erfassen können, aus. So kann Booklet 7 nach der Optimierung 775 Wissenszustände mehr erfassen als noch zuvor und Booklet 8 macht den mit Abstände extremsten Sprung, indem es nun statt der vorherigen 48 von 1280 möglichen Wissenszuständen nun 974 mehr, also 1022 von 1280 möglichen Wissenszuständen erfassen kann. Diese extremen Sprünge erreichen Booklets 7 und 8 zwar nur auf Kosten einer Aufnahme zusätzlicher Items, aber dafür sind sie wirklich auffällig. Auch hier im Falle der Betrachtung aller Skills sind alle Booklets durch die Optimierung more informativ geworden und zwei Booklets wurden zusätzlich noch gekürzt. Auch hier kann man den Optimierungsprozess also als Erfolg bezeichnen.

5.2 Vergleich der optimierten Booklets

Der Vergleich der optimierten Booklets erfolgt an dieser Stelle nur anhand der Booklets, welche basierend auf der Kompetenzstruktur mit allen Skills erstellt wurden und dem Ergebnis des „Shortenings auf alle Kompetenzen“. Das liegt daran, dass diese Ergebnisse, da sie alle auf derselben Kompetenzstruktur basieren, vergleichbar sind. Da die anderen Ergebnisse auf unterschiedlichen Kompetenzstrukturen basieren, die weder dieselben Skills berücksichtigen noch gleichviele Kompetenzzustände haben, wäre anhand ihrer kein sinnvoller Vergleich möglich. Der Vergleich soll an dieser Stelle auch wieder zuerst auf der Kompetenz-Ebene beginnen und dann auf die Item-Ebene übergehen. Zunächst ist jedoch natürlich die Ähnlichkeit zu erwähnen, dass alle Ergebnisse dieselbe Wissensstruktur induzieren. Das ist darauf zurückzuführen, dass Sie auf denselben Kompetenzen basieren. Aus diesen wurden jeweils diejenigen genommen, die das Ableiten neuer Wissenszustände aus der Kompetenzstruktur ermöglichen, um das jeweilige Booklet zu optimieren. Bei allen Ergebnissen wurden alle anderen gelöscht. Daher entstanden am Ende nur Ergebnisse, die dieselbe Wissensstruktur induzieren. Als ersten Punkt gilt es zu erwähnen, dass die Ergebnisse in einem Fall identisch sind. Die optimierten Booklets 1 und 8 sind tatsächlich vollständig identisch. Das lässt die Vermutung zu, dass es leicht passieren kann, dass bei dieser Optimierungsmethode durch das Verwenden derselben Kompetenzen auch identische Ergebnisse erzielt werden. Dafür spricht auch, dass es viele Kompetenzen gibt, die in allen Optimierungsergebnissen auftauchen. Um diese zu identifizieren, soll nun Tabelle 7

betrachtet werden. Wie man Tabelle 7 entnehmen kann, tauchen mit Abstand die meisten Kompetenzen tatsächlich in allen „optimierten reducts“ auf. Was wir daraus mitnehmen können ist, dass es bei der hier angewendeten Optimierungsmethode zwar einen kleinen Unterschied macht von welchem Booklet man bei der Optimierung ausgeht, jedoch sind sich die Booklets und auch das Ergebnis des „Shortenings auf alle Kompetenzen“ nach der Optimierung auf der Kompetenz-Ebene so ähnlich, dass man sie kaum noch als unterschiedliche Booklets bezeichnen kann. Für die Unterschiedlichkeit der Booklets ist aber natürlich die Item-Ebene noch wichtiger als die Kompetenz-Ebene. Man könnte sogar sagen, dass es gut ist, dass die Booklets sich auf der Kompetenz-Ebene nach der Optimierung nun so ähnlich sind. Immerhin sollen sie offiziell etwas vergleichbares messen (Mullis et al., 2004, Martin et al., 2005) und das tun sie nach der Optimierung eher als zuvor.

Auf der Item-Ebene sollten nun jedoch Unterschiede auffallen. Wenn nicht wäre es in der Praxis eher nicht mehr sinnvoll die optimierten Booklets als unterschiedliche Itemsammlungen, zu betrachten oder zu verwenden. An dieser Stelle soll Tabelle 8 betrachtet werden. Es tauchen 14 der verwendeten Items in allen Ergebnissen außer maximal einem auf. Da es sich hierbei um 14 von 19 Items in den Booklets handelt, die oft identisch sind und 9 von 19, die stets identisch sind, können die Booklets meines Erachtens eher nicht mehr als unterschiedliche Itemsammlungen verwendet werden. Die neuen optimierten Itemsammlungen lassen nun also Aussagen über dieselben zugrundeliegenden Fähigkeiten der Testpersonen zu, sind sich dafür aber auch sehr ähnlich geworden. Ein Abgleich der Ergebnisse in allen möglichen Zweier-Kombinationen zeigt zunächst, dass keine zwei Ergebnisse vollständig identisch sind. Bei genauerer Betrachtung fällt die starke Ähnlichkeit der Ergebnisse jedoch schnell auf. Sollte man also weiterhin so viele verschiedene Booklets verwenden wollen, müsste man entweder neue Items entwickeln oder sich mit den Items zufriedengeben, die nicht optimal abschneiden bezüglich careless error- und lucky guess-Wahrscheinlichkeiten.

5.3 Der Weg zum fully informative Test

Die optimierten Booklets sind zwar nun alle more informative im Vergleich zu zuvor, allerdings sind die meisten nicht fully informative. Folglich können nicht alle möglichen Kompetenzzustände in der Art von dem Test erfasst werden, dass man vom Kompetenzzustand der Person auf ihren Wissenszustand schließen kann und umgekehrt. Man

kann also aus der Kompetenzstruktur keine Wissensstruktur mit so vielen Wissenszuständen wie die Kompetenzstruktur Kompetenzzustände hat, ableiten. Da der Test aber der einzige Weg für uns als Tester ist um Informationen über die Fähigkeiten und das Wissen der untersuchten Person zu erlangen, sind wir in dieser Hinsicht ohne einen fully informative Test limitiert. Um das Ziel einer möglichst präzisen Wissensdiagnose zu erreichen, müsste man nun also Items zu dem Test hinzunehmen, um diesen zu einem fully informative Test zu verbessern. Es gibt mehrere Möglichkeiten zu entscheiden welche Items man zu diesem Zweck zum Test hinzunehmen könnte. In dieser Arbeit wurde die beschriebene, sehr basale Methode zu diesem Zweck gewählt. Die Anwendung des CDP auf die Kompetenzen $\{h\}$, $\{i\}$, $\{j\}$, $\{k\}$ und $\{o\}$ ergab ein \mathcal{T} , dass nur noch die Kompetenzen $\{i\}$, $\{j\}$ und $\{o\}$ beinhaltet. Folglich würde es für einen fully informative Test genügen, wenn man zu diesen drei Kompetenzen noch je ein Item finden könnte und diese Items dann zu den anhand aller Skills optimierten Booklets hinzunehmen würde. Diese drei Skills haben inhaltlich folgende Bedeutungen: Skill „i“ entspricht „Apply and extend previous understandings of operations with fractions to add, subtract, multiply, and divide rational numbers.“; Skill „j“ entspricht „Solve real-world and mathematical problems involving the four operations with rational numbers.“ und Skill „o“ entspricht „Use equivalent fraction as a strategy to add and subtract fractions.“. Die Zuweisungen aller Skills zu ihren inhaltlichen Bedeutungen können Abbildung 2 aus der Einleitung dieser Arbeit entnommen werden. Zunächst wurden theoretische Überlegungen angestellt ob diese drei Skills überhaupt einzeln durch je ein Item erfasst werden können. Nachdem diese positiv ausfielen, wurden Items gesucht, welche diesen Ansprüchen genügen. Hierfür wurden die veröffentlichten Items der Kategorien Algebra und Number der TIMSS 2007 und der TIMSS 2011 für Kinder in der achten Schulklasse betrachtet (TIMSS 2007: Study, T. I. I. M. A. S. & Achievement, I. A. F. T. E. O. E., 2009; TIMSS 2011: Foy et al., 2013). Hierbei wurden drei Items gefunden. Skill „i“, kann durch das Item M032166 aus der TIMSS 2011 (Abbildung 7) erfasst werden. Skill „j“, kann durch das Item M042304A aus der TIMSS 2007 (Abbildung 8) erfasst werden. Skill „o“, kann durch das Item M032094 aus der TIMSS 2011 (Abbildung 9) erfasst werden. Somit wäre das Ziel des fully informative Tests bezüglich aller Skills erreicht.

Abbildung 7

Item M032166 aus der TIMSS 2011, welchem von mir die Kompetenz {i} zugeordnet wurde

ID: M032166	Mathematics Grade 8	Block_Seq: M01_01
<p>Which of these is the BEST estimate of $\frac{7.21 \times 3.86}{10.09}$?</p> <p>(A) $\frac{7 \times 3}{10}$</p> <p>(B) $\frac{7 \times 4}{10}$</p> <p>(C) $\frac{7 \times 3}{11}$</p> <p>(D) $\frac{7 \times 4}{11}$</p>		<p>Content Domain Number</p> <p>Topic Area Fractions and Decimals</p> <p>Cognitive Domain Knowing</p> <p>Maximum Points 1</p> <p>Key B</p>

Abbildung 8

Item M042304A aus der TIMSS 2007, welchem von mir die Kompetenz {j} zugeordnet wurde

Item ID	M042304A	Subject	M	Grade	8	Block	M04	Block Seq	05
---------	----------	---------	---	-------	---	-------	-----	-----------	----

Triathlon

A triathlon is a race in which athletes swim, then cycle, then run set distances. The first person to complete the whole course is the winner.

Kathy, Barbara, and Sue competed with each other in a triathlon. The course they covered consisted of a 1 kilometer swim, followed by a 40 kilometer cycle ride, and then a 15 kilometer run.

A. Barbara was the fastest swimmer and completed the 1 km distance in 25 minutes. Kathy took 10 minutes longer than Barbara, and Sue took 5 minutes longer than Kathy.

Use this information to complete the table for swimming:

Swimming	Kathy	Barbara	Sue
Time taken (minutes)		25	

B. Kathy was the fastest cyclist. She averaged 30 kilometers per hour for the 40 km ride. Barbara took 10 minutes more than Kathy, and Sue took 15 minutes more than Kathy.

Use this information to complete the table for cycling:

Cycling	Kathy	Barbara	Sue
Time taken (minutes)			

TIMSS 2007

Mathematics

Eighth Grade

Content Domain

Number

Cognitive Domain

Applying

Maximum Points

1

Key

See scoring guide

Copyright © 2008 International Association for the Evaluation of Educational Achievement (IEA). All rights reserved.

Code	Response	Item: M042304A
	Correct Response	
10	Kathy 35, Sue 40	
	Incorrect Response	
70	Kathy 35, Sue 30	
79	Other incorrect (including crossed out, erased, stray marks, illegible, or off task)	
	Nonresponse	
99	Blank	

Abbildung 9

Item M032094 aus der TIMSS 2011, welchem von mir die Kompetenz {o} zugeordnet wurde

ID: M032094	Mathematics Grade 8	Block_Seq: M05_01
$\frac{4}{100} + \frac{3}{1000} =$ <p> (A) 0.043 (B) 0.1043 (C) 0.403 (D) 0.43 </p>		<p>Content Domain</p> <p>Number</p> <p>Topic Area</p> <p>Fractions and Decimals</p> <p>Cognitive Domain</p> <p>Knowing</p> <p>Maximum Points</p> <p>1</p> <p>Key</p> <p>A</p>

Die Zuordnung der neuen Items zu den Kompetenzen ist strittig, so wie es meistens der Fall ist bei derartigen Zuordnungen. In diesem Fall aber sicherlich noch einmal mehr, da die Zuordnung allein meiner Einschätzung entspringt und nicht beispielsweise durch ein Expertenurteil erstellt oder bekräftigt wurde. Sie soll aber ohnehin eher die Möglichkeit veranschaulichen die Booklets zu fully informative Tests zu verbessern, als eine gesicherte Lösung darzulegen. Im nächsten Schritt sollten auch die Booklets, welche nur anhand der Kompetenzstrukturen mit den Skills zu welchen sie von vorn herein entsprechende Items beinhaltet haben, betrachtet werden. Die Kompetenzstrukturen dieser Booklets sind

Teilmengen der Ergebnisse, welche alle Skills berücksichtigen, da sie nur Kompetenzzustände beinhalten, die auch die übergeordnete Kompetenzstruktur mit allen Skills beinhaltet. Daher ist es sicher, dass eine Hinzunahme von Items, die den Kompetenzen $\{i\}$, $\{j\}$, und $\{o\}$ zugeordnet sind genügen würde um sie zu fully informative Tests zu verbessern. Allerdings könnte dies potentiell auch mit weniger Items gelingen. Es erfolgte eine erneute Anwendung des CDP. Als Ergebnis erhält man, dass das optimierte Booklet 1, welches Skill „j“ nicht in seiner Kompetenzstruktur berücksichtigt, nur ein Item benötigt. Dieses müsste als Kompetenz $\{i\}$ haben um das optimierte Booklet zu einem fully informative Test zu verbessern. Dies ist darauf zurückzuführen, dass es zum einen Skill „j“ nicht erfasst und dadurch zum anderen das Item mit der Kompetenz $\{j, o\}$ für Booklet 1 schon so funktioniert wie ein Item, welches der Kompetenz $\{o\}$ zugeordnet wäre. Somit fehlt Booklet 1 von den drei möglichen Erweiterungsisems lediglich eines mit der Kompetenz $\{i\}$. Booklets 2, welches Skill „l“ nicht in seiner Kompetenzstruktur berücksichtigt, sowie Booklet 7, welches Skill „m“ nicht in seinem Kompetenzstruktur berücksichtigt, benötigen beide alle drei neuen Items. Sie benötigen also eines mit Kompetenz $\{i\}$, eines mit Kompetenz $\{j\}$ und eines mit Kompetenz $\{o\}$. Booklet 8, welches die Skills „d“, „i“, „j“, „k“, „l“ und „n“ nicht berücksichtigt, benötigt keine weiteren Items, da es bereits ein fully informative Test ist. Abschließend kann man sagen, dass alle Ergebnisse das Potential haben zu fully informative Tests zu werden, indem man maximal drei neue Items aufnimmt. Wenn man nun die drei von mir gewählten Items nochmal von Experten validieren lassen würde oder Experten bitten würde andere Items für diesen Zweck zu entwickeln, stünde einer, der Kompetenzstruktur entsprechend optimalen, Wissensdiagnose nichts mehr im Wege.

5.4 Limitationen

Limitierend gilt es zunächst zu erwähnen, dass es sich bei dem gesamten Verfahren um ein rein deterministisches Vorgehen handelt. Sobald eine Versuchsperson einen Flüchtigkeitsfehler macht oder korrekt rät, ist die ökonomische Wissensdiagnose nicht mehr auf diese Weise möglich. Zudem liegen auch keine Zustandswahrscheinlichkeiten der Kompetenz- oder Wissenszustände der entsprechenden Kompetenz- oder Wissensstruktur vor. Daher lässt sich die Auswahl der zuerst vorzulegenden Items ebenfalls nur deterministisch durchführen. Beispielsweise durch das Vorlegen der Items, mit denen man möglichst viele Wissenszustände ausschließen kann. Deterministische Ansätze können die

Realität nahezu niemals korrekt abbilden, da es in der Realität zu viele Einfluss- und Störfaktoren gibt. Daher wäre es in zukünftigen Arbeitsschritten nötig zumindest teilweise in den probabilistischen Bereich überzugehen. Sollte sich dies allerdings nicht umsetzen lassen, durch beispielsweise die Identifizierbarkeitsprobleme des BLIM (Stefanutti, & Spoto, 2020; Heller, 2017), sollte zumindest eine Validierung der Kompetenzstruktur und der Skillmap anhand einer Studie mit möglichst vielen Probanden erfolgen. Sollte sich der rein deterministische Ansatz dort nicht bekräftigen lassen, wäre es vielleicht sinnvoll zu überlegen ob man jede Kompetenz mit mehreren Items abfragen sollte, um so potentiellen Antwortfehlern entgegenzuwirken. Dies würde die Tests zwar verlängern, aber dafür ein wenig die Probleme des deterministischen Ansatzes kompensieren.

Auch die Auswahl der Skills sollte potentiell grundsätzlich überdacht werden. Die Skills sind mit einer Anzahl von 15 Skills auf insgesamt 50 Items scheinbar relativ weit gefächert. Allerdings wäre es bei 50 Items durchaus möglich noch einige Skills mehr anzunehmen. Selbst wenn man sich dagegen entscheiden sollte die Anzahl der Skills zu erhöhen, wäre es vermutlich sinnvoll die Auswahl der Skills, welche man erheben möchte, nochmal zu überdenken. Von den aktuell betrachteten Skills erheben mit den Skills „i“, „m“ und „o“ drei Skills die Rechenarbeit mit Brüchen und mit den Skills „a“, „b“ und „h“ erheben drei Skills Fähigkeiten zur Rechenarbeit mit Verhältnissen, was eng verwandt ist mit der Rechenarbeit mit Brüchen. Im Allgemeinen erfassen die Skills primär eben die Fähigkeiten, welche Schüler in der achten Schulklasse vor kurzem neu gelernt haben und vermutlich gerade zu diesem Zeitpunkt in deren Lehrplan vorkommen. Hierzu gehört neben dem zuvor beschriebenen Rechnen mit Brüchen und dem Umgang mit Verhältnissen (Skills „a“, „b“, „h“, „i“, „m“ und „o“) auch das Lösen von und Arbeiten mit Gleichungen (Skill „f“, „l“ und „n“) und der Umgang mit mathematischen Problemen aus der echten Welt anhand von Textaufgaben (Skills „j“, „k“ und „n“). Für eine Klassenarbeit oder auch den internationalen Vergleich, für welchen die TIMSS 2003 gedacht war, mag diese Menge von Skills dementsprechend auch sinnvoll sein. Allerdings sollte als Grundgedanke hinter derartigen Tests stets auch stehen wie man den Schülern helfen kann sich zu verbessern. Dafür sind die Skills zu wenig feingliedrig und zu fokussiert auf einige wenige Bereiche der Algebra und der Number Inhaltsdimensionen, welche sie erfassen sollen. Als Vorschlag könnte man einige dieser doch sehr ähnlichen Skills streichen. Man könnte zudem einige der Skills, welche oft nur erfassen ob in bestimmten Kontexten alle vier Grundrechenarten beherrscht werden (Skills „i“, „j“ und „n“), so abändern, dass daraus vier Skills werden, von welchen je einer

eine der Grundrechenarten erfasst. Das durch einzelne Items abzufragen wäre in der Praxis definitiv realistischer als die aktuelle Version. Außerdem sollten einige Skills in ihrer Formulierung nochmals durchdacht werden. So sind beispielsweise Skills „e“ („Apply and extend previous understandings of arithmetic to algebraic expressions.“) und „f“ („Reason about and solve one-variable equations and inequalities.“) so allgemein formuliert, dass auch sie kaum durch ein einzelnes Item erfasst werden können. Sie beschreiben eine derart große Vielzahl an Operationen und dementsprechend auch unterschiedlichen Fähigkeiten, dass sie vermutlich auch nach dem Vorlegen von drei Items noch nicht als eindeutig vorliegend oder nicht vorliegend angenommen werden können. An dieser Stelle wäre es also vielleicht sinnvoller jeden von ihnen in drei Skills zu unterteilen, deren Vorliegen sich tatsächlich durch einzelne Items prüfen ließe. Selbst wenn dann noch nicht der gesamte Umfang des ursprünglichen Skills erfasst wäre, könnte man zumindest Aussagen über die neuen, tatsächlich geprüften Skills tätigen. Auch das Hinzunehmen von Skills, welche bei Kindern in der achten Schulklasse bereits als gegeben angenommen werden, wie beispielsweise ein Skill bezüglich des Lösen simpler Multiplikationsaufgaben, sollte überdacht werden. Denn auch wenn die meisten Kinder diese Skills haben, könnte man so genauer bestimmen ob dem so ist oder nicht. So könnte Kindern besser geholfen werden, von welchen man unter Annahme der ursprünglichen Skills, nur gewusst hätte, dass sie nicht über diverse der höheren Skills verfügen, aber nicht dass der Ursprung hierfür wesentlich tiefer, nämlich beispielsweise bei dem Fehlen von Grundkenntnissen im Bereich Multiplikation liegt. Alles in allem würde ich empfehlen die Skills feingliedriger zu gestalten, da dies eher eine wirklich präzise Wissensdiagnose ermöglichen könnte als es bisher der Fall ist.

Ein weiteres Problem dieser Arbeit ist die Zuordnung von Kompetenzen zu Items. Wie immer in der Praxis ist diese auch hier nicht optimal erfolgt. So verlangt beispielsweise Item M012001 von der Testperson, dass diese entscheidet wie viele von 10 Würfeln dunkel gefärbt werden sollen, damit nach der Färbung $\frac{4}{5}$ der Würfel dunkel sind. Hierfür wird Skill „o“ („Use equivalent fraction as a strategy to add and subtract fractions.“) angenommen. Allerdings lässt sich diese Aufgabe auch durch bloßes Gruppieren der Würfel in fünf Gruppen lösen und so würde sie rein geometrisch gelöst werden und das Kürzen/ Erweitern von Brüchen auf denselben Nenner wäre nicht nötig. Gleichzeitig nimmt die Aufgabe Skill „g“ („Recognize and represent proportional relationships between quantities“) nicht als notwendig an, obwohl hier durchaus mit verhältnismäßigen Zusammenhängen gearbeitet wird, weshalb diese

entsprechend Skill „g“ auch erkannt werden müssen (Su et al., 2013; Martin et al., 2005). Eine derartige Argumentation über nicht passende oder zu wenig gewählte Skills lässt sich für nahezu jedes Item führen. Dies ist auch nicht überraschend, da die Skills basierend auf dem bereits fertigen Test entwickelt wurden, anstatt dass man sich zuerst Skills überlegt hat um von diesen dann einen Test abzuleiten. Da Su und Kollegen (2013) lediglich den Test reanalysieren wollten, ist dies auch das korrekte Vorgehen. Allerdings sollte man in Zukunft eher ein inverses Verfahren anwenden. Die Entwicklung von interessierenden Skills und daraufhin von Items, welche diese Skills mit erwünschten Kompetenzen abfragen, stellt meines Erachtens die sinnvollere Reihenfolge dar. Der hier verwendete Optimierungsprozess könnte auch nach wie vor bei der Erstellung eines solchen Tests helfen, es wäre nun nur eher gesichert, dass die Items auch auf das ausgelegt sind, wofür man sie gern nutzen würde.

Würde man zuerst die Skills und die Kompetenzstruktur entwerfen und danach erst den Test, könnte man bei der Auswahl der Skills darauf achten, dass die aus der Kompetenzstruktur abgeleitete Wissensstruktur einen quasi ordinalen Wissensraum (Heller, 2023) darstellt. Dies hätte neben diversen anderen auch den Vorteil, dass man so von der korrekten Beantwortung eines Items darauf schließen könnte, dass eine Versuchsperson alle Voraussetzungen des Items ebenfalls beherrscht. So könnte die Wissensdiagnose gerade im Rahmen von adaptivem Testen noch ökonomischer gestaltet werden (Heller, 2023; Falmagne et al., 1990, 2006). Die von Su und Kollegen (2013) entwickelte Attribute Hierarchy leitet leider nur einen gewöhnlichen Wissensraum ab. Deshalb kann man hier nur darauf schließen, dass eine Testperson irgendeinen der Pfade, welcher eine Voraussetzung für ein vorgegebenes Item darstellt, beherrschen muss, wenn die Testperson das Item lösen kann. Daher würde man hier für die vollständige Bestimmung des Wissenszustands einer Person im Falle einer adaptiven Testung wesentlich mehr Items benötigen, abhängig davon wie verzweigt der Voraussetzungspfad zu einem Item ist.

5.5 Verdienst dieser Arbeit

Neben den Limitationen bringt das hier vorgestellte Vorgehen und auch schon die Grundlagen dafür von Su und Kollegen (2013) diverse Vorteile im Vergleich zur ursprünglichen TIMSS 2003. Einmal ist es durch die Zuordnung von Kompetenzen zu Items nun wesentlich besser und kleinteiliger möglich zu bestimmen über welche Fähigkeiten eine Versuchsperson verfügt. Zudem ermöglicht eben diese Zuordnung ein besseres Verständnis

der Items, da sie nicht mehr nur einer von fünf Inhalts- und einer von vier kognitiven Dimensionen zugeordnet werden. Stattdessen werden den Items der zwei hier betrachteten Inhaltsdimensionen Algebra und Number bereits 15 unterschiedliche Skills zugeordnet. Dies ließe sich sicherlich auch für alle anderen Items mit noch mehr verschiedenen Skills umsetzen, sodass am Ende ein wesentlich besseres Verständnis der Vor- und Nachteile, sowie des Zwecks jedes Items bestimmbar wäre. Außerdem sind wir dank des Vorgehens dieser Arbeit dazu in der Lage Items, welche redundant bezüglich der Kompetenzen sind, welche ihnen zugewiesen sind, miteinander zu vergleichen und das Item mit der geringsten Fehlerwahrscheinlichkeitssumme zu identifizieren. Außerdem können nun Items identifiziert werden, die aufgrund ihrer Kompetenz nichts zur Unterscheidung von Kompetenzzuständen, also zu einer möglichst differenzierten Wissensdiagnose anhand von Wissenszuständen, beitragen. Diese Items auszuschließen kann den Test kürzer machen, ohne ihn seinen Qualitäten zu berauben. Final ist natürlich die größte Errungenschaft dieser Arbeit, dass man nun mit, in vielen Fällen weniger Items, eine präzisere Wissensdiagnose bezüglich der Inhaltsdimensionen Algebra und Number der Booklets 1, 2, 7 und 8 des Mathematiktest-Teils aus der TIMSS 2003 für die achte Schulklasse durchführen kann. Zudem liegen nun neue Booklets vor, welche im Gegensatz zu allen vorherigen Booklets Aufschluss über alle angenommenen Skills zugleich geben können. Zuletzt kann diese nun optimierte Wissensdiagnose vermutlich sogar gemäß der angenommenen Kompetenzstruktur optimal durchgeführt werden, wenn man nur maximal drei Items zu den Booklets hinzunimmt. Wenn man nun noch Methoden aus dem Bereich des adaptiven Testens (Heller, 2023; Falmagne et al., 1990, 2006) anwenden würde, wäre eine optimale Wissensdiagnose in einem Bruchteil der zuvor benötigten Zeit möglich. Selbstverständlich wäre es hiervoor noch nötig die neuen Booklets zunächst anhand einer Erhebung zu validieren und so zu überprüfen ob die von Su und Kollegen (2013) getroffenen Abhängigkeitsbeziehungen, sowie die Auswahl der Skills und die Zuordnung dieser zu Items über eine Skillmap via Kompetenzen, sich so auch in der Praxis finden lässt. Sollte sich dies so bekräftigen lassen, stünde einer ökonomischeren und zugleich präziseren Wissensdiagnose nichts mehr im Weg.

6. Literatur

- Anselmi, P., Heller, J., Stefanutti, L. & Robusto, E. (2022). Constructing, improving, and shortening tests for skill assessment. *Journal Of Mathematical Psychology*, *106*, 102621. <https://doi.org/10.1016/j.jmp.2021.102621>
- Anselmi, P., Heller, J., Stefanutti, L., & Robusto, E. (2024a). Constructing tests for skill assessment with competence-based test development. *British Journal of Mathematical and Statistical Psychology*. <https://doi.org/10.1111/bmsp.12335>
- Anselmi, P., Heller, J., Stefanutti, L., & Robusto, E. (2024b). *Real-Life Applications of Competence-Based Test Development to the Construction, Improvement, and Shortening of Tests* [Unpublished manuscript].
- Arora, A., Achievement, I. A. F. T. E. O. E., Study, T. I. I. M. A. S. & Center, B. C. I. S. (2005). *TIMSS 2003 User Guide for the International Database*.
- Bolt, D. (2007). The Present and Future of IRT-Based Cognitive Diagnostic Models (ICDMs) and Related Methods. *Journal Of Educational Measurement*, *44*(4), 377–383. <https://doi.org/10.1111/j.1745-3984.2007.00045.x>
- De La Torre, J. (2009). DINA Model and Parameter Estimation: A Didactic. *Journal Of Educational And Behavioral Statistics*, *34*(1), 115–130. <https://doi.org/10.3102/1076998607309474>
- Doignon, J.-P. (1994). Knowledge spaces and skill assignments. In G. H. Fisher & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (pp. 111-121). Berlin, Heidelberg, and New York: Springer-Verlag.
- Doignon, J.-P., & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, *23*(2), 175–196. [https://doi.org/10.1016/S0020-7373\(85\)80031-6](https://doi.org/10.1016/S0020-7373(85)80031-6)

Doignon, J. & Falmagne, J. (1999). Knowledge spaces. In *Springer eBooks*.

<https://doi.org/10.1007/978-3-642-58625-5>

Falmagne, J., Cosyn, E., Doignon, J. & Thiéry, N. (2006). The Assessment of Knowledge, in Theory and in Practice. In *Lecture notes in computer science* (S. 61–

79). https://doi.org/10.1007/11671404_4

Falmagne, J. & Doignon, J. (2010). Learning Spaces: Interdisciplinary Applied Mathematics.

<http://ci.nii.ac.jp/ncid/BB1576842X>

Falmagne, J., Koppen, M., Villano, M., Doignon, J. & Johannesen, L. (1990). Introduction to knowledge spaces: How to build, test, and search them. *Psychological Review*, 97(2),

201–224. <https://doi.org/10.1037/0033-295x.97.2.201>

Foy, P., Arora, A., & Stanco, G. M. (2013). TIMSS 2011 User Guide for the International Database. (released Items 8th Grade Mathematics) *International Association for the Evaluation of Educational Achievement*.

Hansen K. D., Gentry J., Long L., Gentleman R., Falcon S., Hahne F. & Sarkar D. (2023).

Rgraphviz: Provides plotting capabilities for R graph objects.

doi:10.18129/B9.bioc.Rgraphviz <https://doi.org/10.18129/B9.bioc.Rgraphviz>, R package version 2.46.0, <https://bioconductor.org/packages/Rgraphviz>.

Heller, J. (2017). Identifiability in probabilistic knowledge structures. *Journal Of*

Mathematical Psychology, 77, 46–57. <https://doi.org/10.1016/j.jmp.2016.07.008>

Heller, J. (2022). Complete Q-matrices in conjunctive models on general attribute structures.

British Journal Of Mathematical & Statistical Psychology/British Journal Of

Mathematical And Statistical Psychology, 75(3), 522–549.

<https://doi.org/10.1111/bmsp.12266>

Heller, J. (2023). Vorlesung psychometrische Theorien und Testentwicklung. *Vorlesung an der Eberhard Karls Universität Tübingen*, Foliensätze 11-13.

Heller, J., Stefanutti, L., Anselmi, P. & Robusto, E. (2015). On the Link between Cognitive Diagnostic Models and Knowledge Space Theory. *Psychometrika*, 80(4), 995–1019. <https://doi.org/10.1007/s11336-015-9457-x>

Heller, J., Stefanutti, L., Anselmi, P. & Robusto, E. (2016). Erratum to: On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, 81, 250–251.

Heller, J., Ünlü, A., & Albert, D. (2013). Skills, competencies and knowledge structures. *Knowledge spaces: Applications in education*, 229-242.

Heller J. & Wickelmaier F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, 42, 49-56. doi: 10.1016/j.endm.2013.05.145 <<https://doi.org/10.1016/j.endm.2013.05.145>>.

Junker, B. W. & Sijtsma, K. (2001). Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric Item Response Theory. *Applied Psychological Measurement*, 25(3), 258–272. <https://doi.org/10.1177/01466210122032064>

Kluyver T., Angerer P. & Schulz J. (2022). IRdisplay: 'Jupyter' Display Machinery. *R package version 1.1*, <https://CRAN.R-project.org/package=IRdisplay>.

Martin, M., Arora A., Barth J., Chrostowski S., Diaconu D., Erberber E., Foy P., Galia J., Gonzalez E., Johansone I., Joncas M., Li I., Mullis I. V., Neuschmidt O., Ramirez M. & J., Taneva M. (2005). TIMSS 2003 User Guide for the International Database / Edited by Michael O. Martin. *TIMSS & PIRLS International Study Center Lynch School of Education Boston College*

Martin, M. O., & Mullis, I. V. (2003). Overview of TIMSS 2003. *TIMSS*, 3-20.

Meyer D. & Hornik K. (2023). relations: Data Structures and Algorithms for Relations. *R package version 0.6-13*, <https://CRAN.R-project.org/package=relations>.

- Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004). Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. *Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.*
- Mullis, I. V., Martin, M. O., Smith, T. A., & International Association for the Evaluation of Educational Achievement. (2003). TIMSS: Assessment frameworks and specifications 2003.
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C.
- Panse, F. (2023). Building Knowledge Structures for the TIMSS 2003 Mathematics Test: Data-Driven versus Competence-Based Approaches. Unveröffentlichte Bachelorarbeit der Universität Tübingen.
- R Core Team (2023). foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', *R package version 0.8-86*, <https://CRAN.R-project.org/package=foreign>.
- R Core Team (2024). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. URL: <https://www.R-project.org/>.
- Spoto, A., Stefanutti, L. & Vidotto, G. (2013). Considerations about the identification of forward- and backward-graded knowledge structures. *Journal Of Mathematical Psychology*, 57(5), 249–254. <https://doi.org/10.1016/j.jmp.2013.09.002>
- Stefanutti, L. & Spoto, A. (2020). BLIM's identifiability and parameter invariance under backward and forward transformations. *Journal Of Mathematical Psychology*, 95, 102314. <https://doi.org/10.1016/j.jmp.2019.102314>

Study, T. I. I. M. A. S. & Achievement, I. A. F. T. E. O. E. (2009). TIMSS 2007 International Database and User Guide. (released Items 8th Grade Mathematics)

Su, Y.-L., Choi, K., Lee, W., Choi, T., & McAninch, M. (2013). *Hierarchical cognitive diagnostic analysis for TIMSS 2003 mathematics*. Centre for Advanced Studies in Measurement and Assessment, 35, 1–71.

Tatsuoka, C. (2002). Data Analytic Methods for Latent Partially Ordered Classification Models. *Journal Of The Royal Statistical Society Series C (Applied Statistics)*, 51(3), 337–350. <https://doi.org/10.1111/1467-9876.00272>

Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Eds.), *Monitoring skills and knowledge acquisition* (pp. 453-488). Hillsdale: Lawrence Erlbaum Associates.

Williamson, J. (2023). Cognitive Diagnostic Models and How They Can Be Useful. Research Report. *Cambridge University Press & Assessment*.

7. Appendix

Tabelle A1

Ergebnisse des BLIMs von Booklet 1: Fehlerparameter, Fehlersumme und die Kompetenz, die dem Item zugeordnet ist, für jedes im ursprünglichen Booklet enthaltene Item

Item	Erfasste Kompetenz	careless error β	lucky guess η	Fehlersumme
M012001	{a, m, o}	0,016	0,285	0,301
M012002	{f}	0,102	0,569	0,671
M012004	{b, g}	0,163	0,221	0,384
M012040	{f, g}	0,000	0,722	0,722
M012041	{a}	0,084	0,553	0,637
M012042	{e, i}	0,030	0,339	0,369
M032570	{b}	0,082	0,511	0,593
M032643	{d, i}	0,111	0,248	0,359
M012016	{d, i}	0,193	0,436	0,629
M012017	{f, k, n}	0,102	0,448	0,550
M022251	{e}	0,664	0,153	0,817
M022185	{e}	0,348	0,277	0,625
M022191	{f}	0,269	0,300	0,569
M022194	{b, n}	0,153	0,371	0,524
M022196	{e}	0,112	0,245	0,357
M022198	{d}	0,266	0,278	0,544
M022199	{c, i, m, o}	0,324	0,205	0,529
M022043	{m}	0,099	0,622	0,721
M022046	{c}	0,081	0,515	0,596
M022050	{f}	0,421	0,171	0,592
M022057	{b, h}	0,054	0,555	0,609
M022066	{c, o}	0,036	0,171	0,207
M022232	{b}	0,555	0,033	0,588
M022234B	{a}	0,719	0,007	0,726
M032142	{b, g}	0,355	0,377	0,732
M032198	{l}	0,000	0,000	0,000
M032640	{f, k, n}	0,651	0,008	0,659

M032755	{b, g, h, k}	0,567	0,007	0,574
M032163	{e}	0,370	0,305	0,675

Tabelle A2

Ergebnisse des BLIMs von Booklet 2: Fehlerparameter, Fehlersumme und die Kompetenz, die dem Item zugeordnet ist, für jedes im ursprünglichen Booklet enthaltene Item

Item	Erfasste Kompetenz	careless error β	lucky guess η	Fehlersumme
M012025	{g}	0,059	0,634	0,693
M012027	{a, b, h}	0,055	0,459	0,514
M012029	{n}	0,128	0,347	0,475
M022104	{d}	0,016	0,608	0,624
M022106	{a, f, g}	0,489	0,134	0,623
M022110	{c}	0,038	0,783	0,821
M022139	{a, b, h}	0,329	0,222	0,551
M022144	{n}	0,123	0,350	0,473
M022156	{b, f}	0,201	0,131	0,332
M022253	{e, f}	0,197	0,248	0,445
M032160	{b, g}	0,691	0,098	0,789
M032307	{i, j, o}	0,262	0,047	0,309
M032381	{c, g}	0,118	0,250	0,368
M032416	{c, o}	0,517	0,000	0,517
M032523	{j, o}	0,427	0,215	0,642
M032525	{f, i}	0,086	0,378	0,464
M032529	{b, h}	0,364	0,272	0,636
M032540	{f}	0,272	0,531	0,803
M032701	{b, f}	0,025	0,811	0,836
M032704	{a, b, g}	0,075	0,438	0,513
M032698	{e, f}	0,406	0,135	0,540
M012016	{d, i}	0,250	0,347	0,597
M012017	{f, k, n}	0,000	0,345	0,345
M022185	{e}	0,353	0,347	0,700
M022191	{f}	0,319	0,347	0,666

M022194	{b, n}	0,242	0,403	0,645
M022196	{e}	0,107	0,299	0,406
M022198	{d}	0,330	0,326	0,656
M022199	{c, i, m, o}	0,334	0,186	0,520
M022251	{e}	0,695	0,207	0,902

Tabelle A3

Ergebnisse des BLIMs von Booklet 7: Fehlerparameter, Fehlersumme und die Kompetenz, die dem Item zugeordnet ist, für jedes im ursprünglichen Booklet enthaltene Item

Item	Erfasste Kompetenz	careless error β	lucky guess η	Fehlersumme
M032142	{b, g}	0,299	0,353	0,652
M032198	{l}	0,000	0,103	0,103
M032640	{f, k, n}	0,676	0,014	0,690
M032755	{b, g, h, k}	0,000	0,010	0,010
M032163	{e}	0,178	0,000	0,178
M022104	{d}	0,000	0,000	0,000
M022106	{a, f, g}	0,356	0,135	0,491
M022110	{c}	0,000	0,000	0,000
M032307	{i, j, o}	0,085	0,056	0,141
M032523	{j, o}	0,478	0,247	0,725
M032701	{b, f}	0,013	0,673	0,686
M032704	{a, b, g}	0,060	0,354	0,414
M032525	{f, i}	0,091	0,253	0,344

Tabelle A4

Ergebnisse des BLIMs von Booklet 1: Fehlerparameter, Fehlersumme und die Kompetenz, die dem Item zugeordnet ist, für jedes im ursprünglichen Booklet enthaltene Item

Item	Erfasste Kompetenz	careless error β	lucky guess η	Fehlersumme
M022043	{m}	0,000	0,000	0,000
M022046	{c}	0,058	0,375	0,433
M022050	{f}	0,265	0,244	0,509

M022057	{b, h}	0,053	0,683	0,736
M022066	{c, o}	0,000	0,225	0,225
M022234B	{a}	0,586	0,000	0,586
M032381	{c, g}	0,063	0,442	0,505
M032416	{c, o}	0,581	0,105	0,686
M032160	{b, g}	0,604	0,074	0,678
M032540	{f}	0,139	0,569	0,709
M032698	{e, f}	0,190	0,266	0,457
M032529	{b, h}	0,091	0,229	0,320

Tabelle A5

Die zu vergleichenden Items, welchen Kompetenz {e} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {e}		Items			
		M022251	M022185	M022196	M032163
Fehlersummen	Booklet 1	0,817	0,625	0,357	0,675
	Booklet 2	0,902	0,700	0,406	-
	Booklet 7	-	-	-	0,178
	Booklet 8	-	-	-	-

Tabelle A6

Die zu vergleichenden Items, welchen Kompetenz {f} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {f}		Items			
		M012002	M022191	M022050	M032540
Fehlersummen	Booklet 1	0,671	0,569	0,592	-
	Booklet 2	-	0,666	-	0,803
	Booklet 7	-	-	-	-
	Booklet 8	-	-	0,509	0,709

Tabelle A7

Die zu vergleichenden Items, welchen Kompetenz {b, g} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {b, g}		Items		
		M012004	M032142	M032160
Fehlersummen	Booklet 1	0,384	0,732	-
	Booklet 2	-	-	0,789
	Booklet 7	-	0,652	-
	Booklet 8	-	-	0,678

Tabelle A8

Die zu vergleichenden Items, welchen Kompetenz {a} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {a}		Items	
		M012041	M022234B
Fehlersummen	Booklet 1	0,637	0,726
	Booklet 2	-	-
	Booklet 7	-	-
	Booklet 8	-	0,586

Tabelle A9

Die zu vergleichenden Items, welchen Kompetenz {b} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {b}		Items	
		M032570	M022232
Fehlersummen	Booklet 1	0,593	0,588
	Booklet 2	-	-
	Booklet 7	-	-
	Booklet 8	-	-

Tabelle A10

Die zu vergleichenden Items, welchen Kompetenz {d, i} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {d, i}		Items	
		M032643	M012016
Fehlersummen	Booklet 1	0,359	0,629
	Booklet 2	-	0,597
	Booklet 7	-	-
	Booklet 8	-	-

Tabelle A11

Die zu vergleichenden Items, welchen Kompetenz {f, k, n} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {f, k, n}		Items	
		M012017	M032640
Fehlersummen	Booklet 1	0,550	0,659
	Booklet 2	0,345	-
	Booklet 7	-	0,690
	Booklet 8	-	-

Tabelle A12

Die zu vergleichenden Items, welchen Kompetenz {d} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {d}		Items	
		M022198	M022104
Fehlersummen	Booklet 1	0,544	-
	Booklet 2	0,656	0,624
	Booklet 7	-	0,000
	Booklet 8	-	-

Tabelle A13

Die zu vergleichenden Items, welchen Kompetenz {c} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {c}		Items	
		M022046	M022110
Fehlersummen	Booklet 1	0,596	-
	Booklet 2	-	0,821
	Booklet 7	-	0,000
	Booklet 8	0,433	-

Tabelle A14

Die zu vergleichenden Items, welchen Kompetenz {b, h} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {b, h}		Items	
		M022057	M032529
Fehlersummen	Booklet 1	0,609	-
	Booklet 2	-	0,636
	Booklet 7	-	-
	Booklet 8	0,736	0,320

Tabelle A15

Die zu vergleichenden Items, welchen Kompetenz {c, o} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {c, o}		Items	
		M022066	M032416
Fehlersummen	Booklet 1	0,207	-
	Booklet 2	-	0,517
	Booklet 7	-	-
	Booklet 8	0,225	0,686

Tabelle A16

Die zu vergleichenden Items, welchen Kompetenz {a, b, h} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {a, b, h}		Items	
		M012027	M022139
Fehlersummen	Booklet 1	-	-
	Booklet 2	0,514	0,551
	Booklet 7	-	-
	Booklet 8	-	-

Tabelle A17

Die zu vergleichenden Items, welchen Kompetenz {n} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {n}		Items	
		M012029	M022144
Fehlersummen	Booklet 1	-	-
	Booklet 2	0,475	0,473
	Booklet 7	-	-
	Booklet 8	-	-

Tabelle A18

Die zu vergleichenden Items, welchen Kompetenz {e, f} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {e, f}		Items	
		M022253	M032698
Fehlersummen	Booklet 1	-	-
	Booklet 2	0,445	0,540
	Booklet 7	-	-
	Booklet 8	-	0,457

Tabelle A19

Die zu vergleichenden Items, welchen Kompetenz {b, f} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

Kompetenz: {b, f}		Items	
		M022156	M032701
Fehlersummen	Booklet 1	-	-
	Booklet 2	0,332	0,836
	Booklet 7	-	0,686
	Booklet 8	-	-

Tabelle A20

Die zu vergleichenden Items, welchen Kompetenz {o} zugeordnet ist und deren Fehlersummen anhand derer sie verglichen werden

(Es gibt eigentlich keine Items, denen Kompetenz {o} zugeordnet ist, aber da beim Improvement von Booklet 8 die beiden Items, denen die Kompetenzen {j, o} und {i, j, o} zugeordnet sind, beide zu Items, denen die Kompetenz {o} zugeordnet ist werden, weil Booklet 8 die Skills „j“ und „i“ ursprünglich mit keinem Item erfasst hat, erfolgt hier dieser Vergleich)

Kompetenz: {o} (bzw. {i, j, o} und {j, o})		Items	
		M032307	M032523
Fehlersummen	Booklet 1	-	-
	Booklet 2	0,309	0,642
	Booklet 7	0,141	0,725
	Booklet 8	-	-