



PhD-FLSHASE-2015-39
Fakultät für Sprachwissenschaften und Literatur, Geisteswissenschaften, Kunst und
Erziehungswissenschaften

DISSERTATION

verteidigt am 23/11/2015 in Esch-sur-Alzette

zur Erlangung des Titels

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN PSYCHOLOGIE

von

André Kretzschmar

geboren am 17. April 1985 in Pirna (Deutschland)

KONSTRUKTVALIDITÄT DES KOMPLEXEN PROBLEMLÖSENS UNTER BESONDERER BERÜCKSICHTIGUNG MODERNER DIAGNOSTISCHER ANSÄTZE

Prüfungskommission

Dr., Samuel Greiff
Associate Professor, Universität Luxemburg

Dr., Matthias Ziegler
Professor, Humboldt-Universität zu Berlin

Dr., Romain Martin
Professor, Universität Luxemburg

Dr., Stephan Krumm
Professor, Freie Universität Berlin

Dr., Sascha Wüstenberg
TWT Science & Innovation

Danksagung

Verschiedene Personen haben auf ganz unterschiedliche Art und Weise – subtil, stetig, substantiell und manchmal wohl auch unwissentlich – ihren Beitrag zu inhaltlichen Überlegungen, zu meiner persönlichen Weiterentwicklung und zu dieser Arbeit insgesamt geleistet. Jedem Einzelnen sei dafür gedankt!

In alphabetischer Reihenfolge:

Melanie Baumgarten, Maik Böttcher, Peer Ederer, Stefanie Erler, Andreas Fischer, Joachim Funke, Samuel Greiff, Katinka Hardt, Julia Hilse, Dirk Kretzschmar, Klaus Kretzschmar, Petra Kretzschmar, Katarina Krkovic, Jakob Mainert, Lysann Majcherek, Romain Martin, Stephan Nebe, Jonas Neubert, Ursula Pöll, Julia Rudolph, Philipp Sonnleitner, Matthias Stadler, Heinz-Martin Süß, Sebastian Thiemme, Cornelia Vogt, Werner W. Wittmann, Christian Wolff, Sascha Wüstenberg, Matthias Ziegler

sowie zahlreiche studentische Hilfskräfte und StudienteilnehmerInnen.

Zusammenfassung

Diese Dissertation beschäftigt sich mit der Frage, ob die Fähigkeit zum Lösen komplexer Probleme als ein eigenständiges Fähigkeitskonstrukt verstanden werden sollte, welches in den etablierten Konzeptualisierungen von Intelligenz bisher nicht ausreichend berücksichtigt wird.

Dazu wird im ersten Teil dieser Arbeit (Kapitel 1) eine Einführung in das Forschungsfeld des Komplexen Problemlösens (KPL) gegeben. Dabei zeigt sich, dass die KPL-Forschung seit jeher stark von der Güte der verwendeten Messinstrumente bestimmt wird, das heißt die Beantwortung der Frage nach einem eigenständigen KPL-Konstrukt insbesondere unter Berücksichtigung der Qualität der KPL-Operationalisierungen vorzunehmen ist.

Der zweite Teil dieser Arbeit widmet sich entsprechend der KPL-Diagnostik. In Kapitel 2 werden grundsätzliche Herausforderungen an die KPL-Diagnostik dargestellt, bevor die wichtigsten Vor- und Nachteile aktueller KPL-Tests am Beispiel von MicroDYN diskutiert werden. Darauf aufbauend werden in Kapitel 3 die Entwicklung sowie erste empirische Ergebnisse eines neuen KPL-Tests namens MicroFIN vorgestellt. Kapitel 4 beschäftigt sich mit einem für die grundsätzlich computerbasierte KPL-Diagnostik besonders relevanten Thema: dem Einfluss der *ICT-Literacy* (d.h. Kompetenz im Umgang mit Computern) auf spezifische KPL-Tests. Dazu wird in Studie 1 die Relevanz der *ICT-Literacy* für die Leistung bei MicroDYN an drei verschiedenen Stichproben ($N_1 = 222$, $N_2 = 341$, $N_3 = 389$) untersucht. Zusätzlich werden die Ergebnisse zum Einfluss der *ICT-Literacy* auf die Leistung bei MicroFIN auf Basis einer Pilotstichprobe ($N = 91$) dargestellt. Die Ergebnisse weisen darauf hin, dass *ICT-Literacy* weder die Leistung bei MicroDYN noch bei MicroFIN in einer validitätsgefährdenden Art und Weise beeinflusst.

Der dritte Teil dieser Arbeit beschäftigt sich mit der zentralen Fragestellung dieser Dissertation, das heißt ob KPL als eigenständiges, kognitives Fähigkeitskonstrukt in Abgrenzung zu etablierten Intelligenzkonstrukten aufgefasst werden sollte oder nicht. Dazu werden in Kapitel 5 Kriterien für die Bewertung eines eigenständigen Fähigkeitskonstrukts unter Berücksichtigung aktueller Forschungsergebnisse für KPL besprochen. Sich daraus ergebende, offene Fragestellungen für die Annahme eines eigenständigen KPL-Konstrukts werden in zwei empirischen Studien untersucht.

In Studie 2 (Kapitel 6; $N = 1908$) steht die inkrementelle Validität von KPL über verbales Reasoning bezüglich der statistischen Erklärung von Schulnoten im Mittelpunkt. Dazu werden erstmals in der neueren KPL-Forschung weitere Einflussgrößen auf Schulnoten (d.h. schulfachliche Kompetenzen) in der mathematischen und sprachlichen Domäne berücksichtigt. Die Ergebnisse deuten auf eine inkrementelle Validität von KPL in der mathematischen, nicht aber in der sprachlichen Domäne hin. Zusätzlich liefert die Studie Hinweise darauf, dass sich die inkrementelle Validität substantiell verringert, wenn zusätzlich schulfachliche Kompetenzen als relevante Prädiktoren der Schulnoten berücksichtigt werden.

In Studie 3 (Kapitel 7; $N = 227$) wird die Validität von KPL in Verbindung mit einer umfangreichen Operationalisierung verschiedener Intelligenzkonstrukte überprüft. Es zeigen sich konvergente sowie diskriminante Validitätsbelege, so dass KPL als spezifische kognitive Fähigkeit interpretiert werden könnte, die in traditionellen Intelligenztests keine Berücksichtigung findet. Hinsichtlich der inkrementellen Validität von KPL bei der statistischen Vorhersage der Schulnoten über etablierte Intelligenzkonstrukte hinaus zeigt sich, dass der Nachweis der inkrementellen Varianzaufklärung nur bei einer spezifischen, nicht jedoch bei einer konstruktrepräsentativen Operationalisierung der Intelligenz gelingt. In Kapitel 8 wird eine alternative Interpretation der in Kapitel 7 aufgezeigten, spezifischen KPL-Varianz vorgestellt, welche in der bisherigen KPL-Forschung noch keine Berücksichtigung gefunden hat. Nach dieser Sichtweise wird KPL vielmehr als eine eigene Inhaltsfacette kognitiver Aufgaben aufgefasst, nicht jedoch als eigenständige kognitive Fähigkeit.

Im letzten Teil dieser Arbeit (Kapitel 9) werden die gewonnenen Ergebnisse und Erkenntnisse zur Diagnostik und Konstruktvalidität von KPL diskutiert. Unter Berücksichtigung aktueller KPL-Tests und den Ergebnissen für MicroFIN werden noch offene Forschungsfragen und Entwicklungspotenziale für die KPL-Diagnostik

aufgezeigt. Anschließend werden die Befunde aus Studie 2 und 3 zur Konstruktvalidität hinsichtlich der Kriterien für ein eigenständiges Fähigkeitskonstrukt eingeordnet. Darauf aufbauend wird die Annahme eines KPL-Konstrukts im nomologischen Netzwerk der Intelligenz bewertet. Dabei wird deutlich, dass die Annahme eines eigenständigen KPL-Konstrukts in Abgrenzung von etablierten Intelligenzkonstrukten derzeit nicht gerechtfertigt erscheint, wobei weitere Forschungsbemühungen zur Klärung dieser Frage notwendig sind.

Abstract

This thesis addresses the question whether Complex Problem Solving (CPS) could be seen as a distinct cognitive ability construct that is not considered sufficiently in the context of established theories of intelligence.

For this purpose, an introduction into the research field of CPS is provided in the first part of this thesis (chapter 1). It is shown that CPS research has always been strongly influenced by the quality of the assessment tools. Therefore, to evaluate whether CPS is an independent construct requires a particular emphasis on the quality of the CPS assessment tools.

Consequently, the second part of this work deals with the assessment of CPS. In chapter 2, the general challenges for CPS assessment as well as important advantages and disadvantages of contemporary CPS assessment tools (using the example of MicroDYN) are discussed. In chapter 3, the development and first empirical results of a new CPS assessment tool called MicroFIN are presented. Chapter 4 covers a particular relevant topic of computer-based CPS assessment: the influence of ICT literacy (i.e. the competence in using computers) on specific CPS assessment tools. For this purpose, the impact of ICT literacy on MicroDYN performance is examined in three different samples ($N_1 = 222$, $N_2 = 341$, $N_3 = 389$). In addition, based on a sample of a pilot study ($N = 91$) the influence of ICT literacy on MicroFIN performance is presented. The results indicate that the impact of ICT Literacy appears neither for MicroDYN nor for MicroFIN as a threat of test validity.

The third part of this work covers the central question of this thesis, that is whether CPS could be seen as a distinct cognitive ability construct that is independent of established intelligence constructs. Therefore, several criteria for an independent ability construct are discussed in chapter 5. Taking current research findings for

CPS into account, open research questions for the assumption of a distinct CPS construct are examined in two empirical studies.

Study 2 (chapter 6; $N = 1908$) focuses on the incremental validity of CPS beyond verbal reasoning in explaining the variance of school grades. For the first time in current CPS research, additional predictors for school grades (i.e. school competencies) in the mathematical and the reading domain are considered. The results indicate an incremental validity of CPS in the mathematical, but not in the reading domain. In addition, the study provides evidence that the incremental validity of CPS holds only to a small degree if relevant predictors of school grades are simultaneously taken into account.

In study 3 (chapter 7; $N = 227$), the validity of CPS is examined with a comprehensive operationalisation of several intelligence constructs. The results indicate evidence for convergent and discriminant validity. Consequently, CPS could be seen as a specific cognitive ability that is less covered by traditional intelligence tests. With regard to the incremental validity of CPS, evidence for a statistical prediction of school grades beyond established intelligence constructs is only found for a specific, but not for a construct-representative operationalisation of intelligence. In chapter 8, an alternative interpretation of the specific CPS variance is discussed. Based on that interpretation, which has not been considered in CPS research yet, CPS can be seen as a specific content facet of cognitive tasks, but not as an independent cognitive ability.

In the last part of this thesis (chapter 9), the findings and considerations concerning the assessment and construct validity of CPS are summarized. Considering current CPS assessment tools and the results for MicroFIN, research questions that are still open and the potential for future CPS assessment are presented. Furthermore, the findings from study 2 and 3 are reviewed with regard to the criteria for an independent ability construct. Finally, the assumption of a distinct CPS construct in the nomological network of intelligence is evaluated. Although further research is needed to clarify this issue, currently the assumption of a distinct CPS construct independent from established intelligence constructs does not seem justified.

Inhaltsverzeichnis

Tabellenverzeichnis	x
Abbildungsverzeichnis	xi
Veröffentlichungen für die kumulative Dissertation	xii
Zusätzliche Veröffentlichungen	xiii
I Theoretischer Hintergrund	1
1 Komplexes Problemlösen in der Psychologie	2
II Diagnostik	7
2 Ein Resümee der KPL-Diagnostik	8
3 MicroFIN als neuer KPL-Test	19
3.1 Facettendiagnostik	21
3.2 Empirische Überprüfung des Messmodells	26
3.3 Item- und Skalenkennwerte	32
3.4 Zusammenfassung	35
4 ICT-Literacy als Herausforderung für die KPL-Diagnostik	39
4.1 Studie 1: ICT-Literacy und MicroDYN	42
4.2 ICT-Literacy und MicroFIN	43
4.3 Zusammenfassung	47
III Konstruktvalidität	50
5 KPL als eigenständiges Fähigkeitskonstrukt?	51
5.1 Kriterien für ein eigenständiges Fähigkeitskonstrukt	52
5.2 Zusammenfassende Bewertung und offene Forschungsfragen	62

6	Studie 2: KPL und die Bedeutung von schulfachlichen Kompetenzen	71
7	Studie 3: KPL im nomologischen Netzwerk der Intelligenz	72
8	KPL aus der Perspektive der BIS-Inhaltsmodalität	73
8.1	Alternative Interpretation zur Konstruktvalidität	73
8.2	Empirische Überprüfung	74
8.3	Diskussion und Implikation	76
IV	Synopse	80
9	Zusammenfassung der bisherigen Befunde und Implikationen	81
9.1	KPL-Diagnostik	81
9.2	Konstruktvalidität von KPL	85
9.3	Ausblick für die weitere KPL-Forschung	89
V	Anhang	95
A	Deskriptive Statistiken zur Studie <i>ICT-Literacy</i> und MicroFIN	96
B	KPL aus der Perspektive der BIS-Inhaltsmodalität	97
C	Modifizierte Parcelbildung der BIS-Aufgaben	100

Tabellenverzeichnis

3.1	Übersicht über die entwickelten MicroFIN-Aufgaben	21
3.2	Modellfit der Messmodelle für MicroFIN	30
3.3	Übersicht über die Item- und Skalenkennwerte der MicroFIN-Aufgaben bei Neubert et al. (2014)	32
3.4	Übersicht über die Item- und Skalenkennwerte der MicroFIN-Aufgaben bei Kretzschmar et al. (in 2015, Kapitel 7)	33
3.5	Übersicht über die Item- und Skalenkennwerte der MicroFIN-Aufgaben bei Mainert et al. (eingereicht)	33
4.1	Korrelationen zwischen der Leistung bei MicroFIN sowie MicroDYN und <i>ICT-Literacy</i> mit und ohne Kontrolle der Intelligenz	45
5.1	Übersicht der Studien mit KPL-Tests des MKS-Ansatzes in Verbindung mit Intelligenzmessungen sowie ggf. externe Kriterien.	57
5.2	Klassifizierung der Intelligenzmessungen der bisherigen Studien entsprechend der Inhaltsmodalitäten des BIS-Modells	59
A.1	Deskriptive Statistiken und Gruppenvergleich zur Studie <i>ICT-Literacy</i> und MicroFIN	96
C.1	Modifizierte Parcelbildung	100

Abbildungsverzeichnis

2.1	Eine MicroDYN-Aufgabe in der Explorationsphase	12
3.1	Explorationsphase bei MicroFIN	23
3.2	Wissensdiagnostikphase „Initialzustand“ bei MicroFIN	25
3.3	Wissensdiagnostikphase „Zustandsidentifikation“ bei MicroFIN . . .	26
3.4	Wissensanwendungsphase bei MicroFIN	27
3.5	Messmodelle für MicroFIN	28
4.1	Pfadmodell hinsichtlich der statistischen Vorhersage von MicroFIN und MicroDYN durch <i>ICT-Literacy</i> und Intelligenz	46
8.1	<i>Nested-factor</i> -Messmodell auf der Grundlage der BIS-Inhaltsdimen- sionen mit zusätzlichem KPL-Faktor	75
8.2	Erweitertes BIS-Modell mit die Inhaltsfacette „interaktiv“	78
B.1	<i>Nested-factor</i> -Messmodell des BIS-Tests auf der Grundlage modifi- zierter Parcelbildung und mit zusätzlichem Faktor für verbale Fä- higkeit	98

Veröffentlichungen für die kumulative Dissertation

- Greiff, S., Kretzschmar, A., Müller, J. C., Spinath, B. & Martin, R. (2014). The computer-based assessment of complex problem solving and how it is influenced by students' information and communication technology literacy. *Journal of Educational Psychology*, 106(3), 666–680. doi:10.1037/a0035426
- Kretzschmar, A., Neubert, J. C. & Greiff, S. (2014). Komplexes Problemlösen, schulfachliche Kompetenzen und ihre Relation zu Schulnoten. *Zeitschrift für Pädagogische Psychologie*, 28(4), 205–215. doi:10.1024/1010-0652/a000137
- Kretzschmar, A., Neubert, J. C., Wüstenberg, S. & Greiff, S. (2015). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence*. doi:10.1016/j.intell.2015.11.004

Zusätzliche Veröffentlichungen

Greiff, S., Kretzschmar, A. & Leutner, D. (2014). Problemlösen in der Pädagogischen Psychologie. *Zeitschrift für Pädagogische Psychologie*, 28(4), 161–166. doi:10.1024/1010-0652/a000140

Mainert, J., Baggen, Y., Kretzschmar, A., Neubert, J. C., Lans, T., Biemans, H., ... Greiff, S. (eingereicht). Perceiving Entrepreneurial Challenges as Complex Problems: The Role of Complex Problem Solving in Opportunity Identification. *Journal of Business Venturing*.

Mainert, J., Kretzschmar, A., Neubert, J. C. & Greiff, S. (2015). Linking complex problem solving and general mental ability to career advancement: Does a transversal skill reveal incremental predictive validity? *International Journal of Lifelong Education*, 34(4), 393–411. doi:10.1080/02601370.2015.1060024

Neubert, J. C., Kretzschmar, A., Wüstenberg, S. & Greiff, S. (2015). Extending the Assessment of Complex Problem Solving to Finite State Automata: Embracing Heterogeneity. *European Journal of Psychological Assessment*, 31(3), 181–194. doi:10.1027/1015-5759/a000224

Neubert, J. C., Mainert, J., Kretzschmar, A. & Greiff, S. (2015). The Assessment of 21st Century Skills in Industrial and Organizational Psychology: Complex and Collaborative Problem Solving. *Industrial and Organizational Psychology*, 8(2), 238–268. doi:10.1017/iop.2015.14

Teil I

Theoretischer Hintergrund

Kapitel 1

Komplexes Problemlösen in der Psychologie

Komplexes Problemlösen (KPL) ist in der Psychologie eine seit etwa 40 Jahren währende Bemühung, die Diagnostik und die Konzeptualisierung der kognitiven Fähigkeiten zu erweitern. Mit den Arbeiten von Dörner (z. B. Dörner, Kreuzig, Reither & Stäudel, 1983) als Ansatzpunkt dieser – vor allem in der deutschsprachigen Forschungslandschaft beheimateten – Entwicklung wurde das etablierte Intelligenzkonzept und insbesondere dessen Diagnostik insofern kritisiert, dass allenfalls nur in beschränktem Maße die kognitiven Anforderungen des alltäglichen Lebens abgebildet würden. Daraufhin entwickelte sich eine rege Forschung und Diskussion, die insbesondere die Diagnostik von KPL als auch die Verortung von KPL im nomologischen Netzwerk der Intelligenz zum Gegenstand hatten.

Das KPL-Konzept adaptiert (oder beschränkt) letztendlich die klassische Definition von Problemlösen (d. h. die Überführung eines Anfangszustandes in einen Zielzustand, wobei die Art und Weise der Überführung anfänglich unbekannt ist; Dörner, 1976; Duncker, 1935; Funke, 2003; Hussy, 1984) auf spezifische Situationen, welche sich laut Dörner et al. (1983) unter anderen durch die Eigenschaften Komplexität, Vernetztheit, Intransparenz, Eigendynamik und Polytelie (Vielzieligkeit) auszeichnen (für einen Überblick siehe Funke, 2003). Entsprechend liegt dieser Arbeit analog zu aktuellen KPL-Forschungsarbeiten Buchners Definition von KPL zugrunde, welche die charakteristischen Eigenschaften komplexer Situationen betont:

CPS [Complex Problem Solving] is the successful interaction with task environments that are dynamic (i.e.; change as a function of the user's interventions and/or as a function of time) and in which some, if not all, of the environment's regularities can only be revealed by successful exploration and integration of the information gained in that process.

(Frensch & Funke, 1995, S. 14)

Dabei wird mit dem KPL-Konzept angenommen, dass die Eigenschaften komplexer Problemsituationen kognitive Anforderungen implizieren, welche für die Lösung von alltäglichen Problemen (z. B. bei der Benutzung eines unvertrauten Mobiltelefons) relevant sein sollten, aber in den bisherigen Konzeptualisierungen als auch Operationalisierungen kognitiver Fähigkeiten nicht adäquat berücksichtigt werden (Funke, 2003).

Die KPL-Forschung hatte jedoch von Beginn an mit methodischen Unzulänglichkeiten der KPL-Tests zu kämpfen, so dass die frühen Versprechungen, eine von Intelligenz unabhängige Leistung erfassen zu können, kritisch hinterfragt werden mussten (für einen umfassenden Überblick siehe u. a. Funke, 2003; Süß, 1996). Mit dem Ende der 1990er-Jahren flachte das Interesse an der KPL-Forschung ab, weil insbesondere die diagnostischen Probleme nur schwer zu bewältigen waren. Zusätzlich schlussfolgerten einige Forscher auf Grundlage methodisch anspruchsvoller Studien, dass KPL letztendlich eher als neue Bezeichnung für ein Konglomerat etablierter (Intelligenz-)Konstrukte und nicht als neues, eigenständiges Fähigkeitskonstrukt angesehen werden sollte (z. B. Süß, 1996; für einen Überblick siehe auch Rost, 2009).

In den letzten Jahren hat das Interesse an der KPL-Forschung vor allem in den empirischen Bildungswissenschaften wieder zugenommen (siehe z. B. Greiff, Kretschmar & Leutner, 2014). Einerseits ist dies wohl auf innovative Neu- und Weiterentwicklungen von KPL-Tests zurückzuführen, andererseits wird KPL zu den *21st century skills* (National Research Council, 2012) gezählt, welchen besonders in Zeiten internationaler Schulleistungsstudien (z. B. *Programme for International Student Assessment*, PISA; OECD, 2014) große Aufmerksamkeit geschenkt wird. Konsequenterweise ist KPL mittlerweile auch Bestandteil einiger dieser Schulleistungsstudien (z. B. PISA 2012; OECD, 2014).

Durch die verbesserten Diagnostikmöglichkeiten von KPL lassen sich gleichsam auch fundierte Hinweise auf die schon seit Beginn der KPL-Forschung aufgeworfene und rege diskutierte Frage beisteuern: Lässt sich KPL von etablierten Intelligenzkonstrukten abgrenzen? Wenn dem so sein sollte, ist dementsprechend eine Erweiterung bestehender Intelligenztheorien um ein eigenständiges KPL-Konstrukt gerechtfertigt? Wenn nicht, durch welche etablierten Konstrukte kann KPL erklärt werden?

Das Ziel dieser Arbeit ist es, Evidenz für die Bewertung eines eigenständigen KPL-Konstrukts beizusteuern. Die Frage nach einem eigenständigen Konstrukt ist dabei nicht nur aus theoretischer Sicht relevant (vgl. Brocke & Beauducel, 2001; Cronbach & Meehl, 1955; Weber & Westmeyer, 2001; Wilhelm, 2009), sondern hat auch praktische Bedeutsamkeit. Beispielsweise hat sich in der über 100-jährigen Intelligenzforschung gezeigt, dass Intelligenz nicht oder nur bedingt trainierbar ist – insbesondere wenn der Transfer fernab von Testaufgaben berücksichtigt wird (für einen umfassenden Überblick siehe Rost, 2009). Wenn jedoch ein neues, eigenständiges KPL-Konstrukt postuliert wird, welches de facto eine neue Bezeichnung für etablierte Intelligenzkonstrukte darstellt, werden gegebenenfalls Forschungsanstrengungen und -finanzierungen hinsichtlich der KPL-Trainierbarkeit bemüht, deren Erfolgsaussichten von vornherein nur sehr begrenzt sind. Gelingt es jedoch nachzuweisen, in welcher Art und Weise sich ein neues KPL-Konstrukt von etablierten Intelligenzkonstrukten abgrenzen lässt, dann würde das Wissen über die distinkten KPL-Bestandteile die Erfolgsaussichten für die Trainierbarkeit von KPL entsprechend vergrößern. Die Frage der Trainierbarkeit (siehe dazu Funke, 2003) ist also eng verbunden mit der Frage nach der Konstruktvalidität von KPL.

Um der Fragestellung der Konstruktvalidität von KPL im Rahmen dieses Dissertationsprojekts nachzugehen, gliedert sich die hier dargestellte Arbeit in zwei aufeinander aufbauende Teile. Der erste Teil widmet sich der Diagnostik der komplexen Problemlöseleistung (Kapitel 2), da die Qualität der KPL-Tests substantiell die Möglichkeiten beeinflusst, die Konstruktvalidität von KPL zu erforschen. Nachdem die Herausforderungen an die KPL-Diagnostik dargestellt wurden, wird näher auf die aktuellen KPL-Operationalisierungen eingegangen. Darauf aufbauend werden die Konstruktion und erste empirische Ergebnisse eines neuen KPL-Tests namens MicroFIN beschrieben (Kapitel 3). Die Entwicklung von MicroFIN soll dem Anspruch gerecht werden, die bestehende KPL-Diagnostik zu erweitern,

so dass verschiedenartige, psychometrisch überzeugende KPL-Operationalisierungen für die Forschung zur Verfügung stehen. Im Anschluss daran wird in Studie 1 (Kapitel 4) auf ein für die grundsätzlich computerbasierte KPL-Diagnostik besonders relevantes Thema eingegangen: der Einfluss der *ICT-Literacy* (d.h. die Kompetenz im Umgang mit Computern) auf spezifische KPL-Test.

Der zweite Teil dieser Arbeit widmet sich der Konstruktvalidität von KPL, das heißt der Beziehung von KPL zu etablierten Intelligenzkonstrukten und der Frage, ob die Annahme eines distinkten KPL-Konstrukts gerechtfertigt erscheint. Dazu werden in Kapitel 5 aktuelle Forschungsbefunde in einen Kriterienkatalog für ein eigenständiges Fähigkeitskonstrukt eingeordnet. Im Anschluss wird in zwei Studien offenen Fragestellungen nachgegangen, welche bisher nur unzureichend in der aktuellen KPL-Forschung Beachtung gefunden haben. In Studie 2 (Kapitel 6) wird überprüft, ob die oft berichtete inkrementelle Erklärung der Schulnotenvarianz durch KPL über etablierte Intelligenzkonstrukte hinaus repliziert werden kann, wenn Schulkompetenzen als wichtige Prädiktoren der Schulnoten berücksichtigt werden. In Studie 3 (Kapitel 7 und 8) wird erstmals in der neueren KPL-Forschungslandschaft eine breite KPL-Operationalisierung einer umfassenden Intelligenzoperationalisierung gegenübergestellt. Als Folge können Hinweise gewonnen werden, ob eine Erweiterung bestehender Intelligenztheorien (hier am Beispiel des Berliner Intelligenzstrukturmodells; Jäger, 1982, 1984) um ein neues KPL-Konstrukt angemessen erscheint.

Diese Arbeit schließt mit einer zusammenfassenden Darstellung und Interpretation der aktuellen Befunde hinsichtlich des Status eines eigenständigen KPL-Konstrukts (Kapitel 9). Gleichzeitig werden noch offene Forschungsfragen thematisiert, welchen in zukünftigen Studien nachgegangen werden sollte, um weitere Erkenntnisse über KPL und dessen Beziehung zu etablierten Intelligenzkonstrukten zu erlangen.

Literatur

- Brocke, B. & Beauducel, A. (2001). Intelligenz als Konstrukt. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 13–42). Lengerich: Pabst.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281.

- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. Stuttgart: Kohlhammer.
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Duncker, K. (1935). *The psychology of productive thinking*. Berlin: Springer.
- Frensch, P. A. & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Hrsg.), *Complex problem solving: The European perspective*. (S. 3–25). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Greiff, S., Kretschmar, A. & Leutner, D. (2014). Problemlösen in der Pädagogischen Psychologie. *Zeitschrift für Pädagogische Psychologie*, 28(4), 161–166. doi:10.1024/1010-0652/a000140
- Hussy, W. (1984). *Denkpsychologie. Ein Lehrbuch*. Stuttgart: Kohlhammer.
- Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195–225.
- Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35(1), 21–35.
- National Research Council. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: The National Academies Press.
- OECD. (2014). *Pisa 2012 results: Creative problem solving: Students' skills in tackling real-life problems (Volume V)*. Paris: OECD Publishing.
- Rost, D. H. (2009). *Intelligenz: Fakten und Mythen*. Weinheim: Beltz PVU.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen: Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Weber, H. & Westmeyer, H. (2001). Die Inflation der Intelligenzen. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 251–266). Lengerich: Pabst.
- Wilhelm, O. (2009). Issues in Computerized Ability Measurement: Getting out of the Jingle and Jangle Jungle. In F. Scheuermann & J. Björnsson (Hrsg.), *The Transition to Computer-Based Assessment New Approaches to Skills Assessment and Implications for Large-scale Testing* (S. 145–149). Luxembourg: Office for Official Publications of the European Communities.

Teil II

Diagnostik

Kapitel 2

Ein Resümee der KPL-Diagnostik

Die Frage nach einem eigenständigen KPL-Konstrukt (siehe Kapitel 5) kann nur dann zuverlässig erforscht werden, wenn qualitativ überzeugende Operationalisierungen des Konstrukts zur Verfügung stehen (Greiff, 2012). In diesem Sinne war und ist die KPL-Forschung seit jeher intensiv von der Diskussion über die Qualität der jeweiligen Diagnostikinstrumente beeinflusst. In den Anfangsjahren kamen über einen langen Zeitraum hinweg theoretisch sowie methodisch suboptimale Erhebungsinstrumente zum Einsatz, welche selten den klassischen Gütekriterien der psychologischen Diagnostik genügten (für einen Überblick, siehe z. B. Funke, 2003; Süß, 1996). Erst durch die Weiterentwicklung der KPL-Diagnostik mit der Arbeit von Greiff (2012) wurde letztendlich die Grundlage für eine Diagnostik gelegt, welche überzeugende psychometrische Qualitäten bei einer effizienten Testdurchführung ermöglichte.

In diesem Kapitel werden daher zuerst die Herausforderungen an die KPL-Diagnostik umrissen sowie die derzeitig etablierten Lösungsansätze aufgezeigt. Da sich diesem Thema andernorts umfassend und wiederholt gewidmet wurde (siehe auch Kapitel 7), wird es an dieser Stelle nur insofern überblicksartig dargestellt, wie es für das Verständnis der darauf aufbauenden Themen notwendig ist. Anschließend wird im Kapitel 3 ein neu entwickelter KPL-Test vorgestellt, welcher zu einer erweiterten KPL-Diagnostik beitragen soll.

Entwicklung der KPL-Diagnostik hin zum Multiple-Komplexe-Systeme-Ansatz

Anfangs zeichnete sich die KPL-Diagnostik vor allem durch umfangreiche Computerszenarien aus (sogenannte Mikrowelten), die mehr oder weniger charakteristische Eigenschaften von Alltagsproblemen und damit des postulierten KPL-Konstrukts (z. B. Komplexität, Vernetztheit, (Eigen-)Dynamik, Intransparenz, Polyteylie, siehe Dörner, Kreuzig, Reither & Stäudel, 1983; Funke, 2003) widerspiegeln sollten. Trotz oder aufgrund der (angenommenen) ökologischen Validität (Funke, 2001; Süß, 1999) zeichneten sich die vorhandenen Messinstrumente jedoch durch substantielle Einschränkungen in der psychometrischen Qualität aus, welchen nur durch besonders aufwändige, aber wenig praktikable Erhebungsdesigns entgegen gewirkt werden konnte (siehe z. B. die Studien von Süß, 1996; Wittmann & Süß, 1999). Umfassend wurde auf die Limitationen hinsichtlich der psychometrischen Qualität der Mikrowelten in früheren Arbeiten (z. B. Funke, 2003; Greiff, 2012; Süß, 1996, 1999) eingegangen, so dass an dieser Stelle nur die drei wichtigsten Punkte umrissen werden sollen: der Einfluss des Vorwissens auf die KPL-Leistung, die mangelnde Objektivität sowie die unzureichende Reliabilität der Leistungsbeurteilung.

Der erste Kritikpunkt bezieht sich auf die Vorwissensabhängigkeit der KPL-Tests. Durch die möglichst realitätsnahe Aufgabengestaltung im Sinne der ökologischen Validität war die Leistung bei den KPL-Tests auch immer vom Vorwissen der Problemlösenden beeinflusst, so dass unterschiedliches Vorwissen zu Leistungsunterschieden beim eigentlichen Problemlösen führte (d. h. Problemlösende mit großem Vorwissen wurden bevorteilt). Als Resultat umfasste die Leistung in den KPL-Tests also auch situationsspezifisches Vorwissen, welches im Sinne des angenommenen, domänenübergreifenden KPL-Konstrukts jedoch als konstruktirrelevant (Messick, 1995) einzustufen war. Diesem Problem wurde in der weiteren Entwicklung der KPL-Diagnostik entweder mit der statistischen Kontrolle des Vorwissens (z. B. Süß, 1996; Süß, Kersting & Oberauer, 1991) oder aber mit einer abstrakten (z. B. in der Mikrowelt FSYS, Wagener, 1997) beziehungsweise fehlenden bedeutungsvollen, semantischen Einbettung (z. B. in der Mikrowelt MultiFlux Kröner, 2001) entgegengewirkt. In den letzten Jahren hat sich vor allem der zweite Weg innerhalb der KPL-Diagnostik durchgesetzt, bei dem die KPL-Tests (relativ) vorwissensneutral konstruiert werden.

Als zweiter Nachteil der traditionellen Mikrowelten galt die oftmals fehlende theoriegeleitete Konstruktion der KPL-Tests und der damit einhergehenden, mangelnden Objektivität. Dieser Umstand führte dazu, dass es in der Regel keine optimalen Lösungswege gab, konkrete Ziele hinsichtlich der Problemlösung nur selten definiert wurden und die Bewertung der KPL-Leistung insgesamt mehrdeutig sein konnte (siehe jedoch Engelhart, Funke & Sager, 2013, für einen Lösungsansatz). Durch die Einführung von formalen Rahmenkonzepten zur Konstruktion und Beschreibung von KPL-Aufgaben (z. B. der DYNAMIS-Ansatz auf Basis linearer Strukturgleichungen; Funke, 1985; oder der Ansatz der finiten Automaten; Funke & Buchner, 1992) konnte dieses Problem gelöst werden. In diesen Rahmenkonzepten wurden formal beschreibbare Aufgaben entwickelt, so dass sich eindeutige Lösungswege sowie Ziele definieren ließen, um die KPL-Leistung objektiv bewerten zu können. Dies war umso besser umsetzbar, je weniger sich auf realistische, das heißt vorwissensbasierte Probleme gestützt wurde. Die KPL-Tests auf Basis formaler Rahmenkonzepte gingen daher oft Hand in Hand mit dem Ansatz der vorwissensneutralen KPL-Aufgaben (siehe z. B. Kluge, 2008; Wirth & Funke, 2005).

Der letzte hier ausgeführte Kritikpunkt ließ sich allerdings auch nicht mit vorwissensneutralen, auf formalen Rahmenkonzepten basierten KPL-Aufgaben lösen. Obwohl typische KPL-Aufgaben meist Bearbeitungszeiten von bis zu einer Stunde und länger hatten, basierte die Leistungsbewertung letztendlich nur auf einem Datenpunkt einer einzelnen Messung (z. B. finales Gesamtkapital einer Firma). Die daraus folgenden Einschränkungen in der Reliabilität (vgl. *single-act-criterion*; Fishbein & Ajzen, 1975) waren offensichtlich, so dass es auch nicht überraschend erschien, dass verschiedene KPL-Tests nur gering miteinander korrelierten (Funke, 2003; Süß, 1996, 1999). Auch alternative Zugänge, bei denen mehrere abhängige Items innerhalb einer Aufgabe (z. B. Erreichen mehrerer Zielzustände auf der Basis der gleichen Problemsituation) genutzt wurden, waren aus psychometrischer Sicht als nicht befriedigend einzuschätzen (siehe Greiff, 2012).

Durch die Entwicklung des Multiple-Komplexe-Systeme (MKS) Ansatzes (Greiff, 2012; Greiff, Fischer, Stadler & Wüstenberg, 2014) wurde diesem Problem entgegengewirkt. Hinter dem MKS-Ansatz steht die grundlegende Idee, Reliabilitätssteigerungen durch eine größere Anzahl an unabhängigen Aufgaben zu erreichen (vgl. Bühner, 2006; Lienert & Raatz, 1998; Wittmann & Matt, 1986). Da es wenig praktikabel erschien, mehrere einstündige KPL-Aufgaben vorzugeben, wurden stattdessen unterschiedliche, weniger komplexe Problemsituationen mit entsprechend

geringerer Bearbeitungszeit (je etwa 5–6min) eingesetzt. Als Folge konnten in der gleichen Erhebungszeit der klassischen KPL-Tests bis zu 10 unabhängige Messungen vorgenommen werden, so dass die Einschränkungen der *single-act*-Messung (Fishbein & Ajzen, 1975) überwunden wurden.

Am Ende dieser – vereinfacht dargestellten – Entwicklung der KPL-Diagnostik, das heißt von umfangreichen, sehr komplexen und vorwissensabhängigen Mikrowelten hin zu formalisierten, multiplen und vorwissensneutralen KPL-Aufgaben, standen am Anfang dieses Dissertationsprojekts zwei KPL-Tests, die sich in der KPL-Forschung und insbesondere in den empirischen Bildungswissenschaften in den letzten Jahren etabliert hatten: MicroDYN (Greiff, 2012; Greiff, Wüstenberg & Funke, 2012) und GeneticsLab (Sonnleitner et al., 2012, 2013). Da sich beide vom Konstruktionsprinzip, dem Aufgabenformat als auch den kognitiven Anforderungen grundsätzlich sehr ähneln, wird im Folgenden nur näher auf MicroDYN eingegangen. Die Ausführungen lassen sich jedoch in der Regel ohne größere Einschränkungen auch auf GeneticsLab übertragen.

Der MicroDYN-Ansatz und dessen kritische Würdigung

Bei MicroDYN (Greiff, 2012; Greiff et al., 2012) werden die Vorteile des DYNAMIS-Konzepts (z. B. formale Beschreibbarkeit der Aufgaben auf Basis linearer Strukturgleichungen) mit denen des MKS-Ansatzes (z. B. Reliabilitätssteigerungen durch multiple Aufgaben) verbunden, so dass eine Vergleichbarkeit und systematische Schwierigkeitsskalierung der Aufgaben sowie überzeugende Reliabilitäten erreicht werden können (Greiff, 2012).

In der Praxis werden bei MicroDYN üblicherweise 6–10 Aufgaben während einer Testung präsentiert. Jede Aufgabe ist dabei durch verschiedene Eingabe- und Ausgabevariablen und deren lineare Beziehung zueinander formalisiert (siehe Kapitel 7 für die Strukturgleichungen eines typischen MicroDYN-Tests). Zum Zwecke einer Facettendiagnostik orientiert sich der Ablauf bei jeder MicroDYN-Aufgabe an den angenommenen Prozessen des Problemlösens (d. h. Exploration, Wissenserwerb und Wissensanwendung; Fischer, Greiff & Funke, 2012; Funke, 2003). In der ersten Phase (Exploration und Wissenserwerb) soll das Problem erkundet werden, um sich Wissen über die unbekannte Problemstruktur anzueignen. Dazu werden die Eingabevariablen manipuliert, um anhand der daraus resultierenden

Veränderungen der Ausgabevariablen auf Beziehungen zwischen den einzelnen Variablen schließen zu können (siehe Abbildung 2.1 für ein Aufgabenbeispiel). In der zweiten Phase (Wissensanwendung) soll das erworbene Wissen über die Problemstruktur genutzt werden, um ein vorgegebenes Ziel zu erreichen. Dazu wird die zugrundeliegende Aufgabenstruktur (d. h. die Beziehung zwischen den Variablen) präsentiert, auf deren Grundlage konkrete Zielwerte der Ausgabevariablen in einer bestimmten Anzahl an Schritten erreicht werden sollen. Am Ende kann für jede MicroDYN-Aufgabe die Leistung bezüglich des Explorationsverhaltens, des Wissenserwerbs und der Wissensanwendung bewertet werden.

Eine ausführlichere Beschreibung von MicroDYN findet sich bei Greiff et al. (2012) sowie in den jeweiligen Publikationen dieses Dissertationsprojekts (Greiff, Kretzschmar, Müller, Spinath & Martin, 2014, Kapitel 4; Kretzschmar, Neubert & Greiff, 2014, Kapitel 6; Kretzschmar, Neubert, Wüstenberg & Greiff, 2015, Kapitel 7).

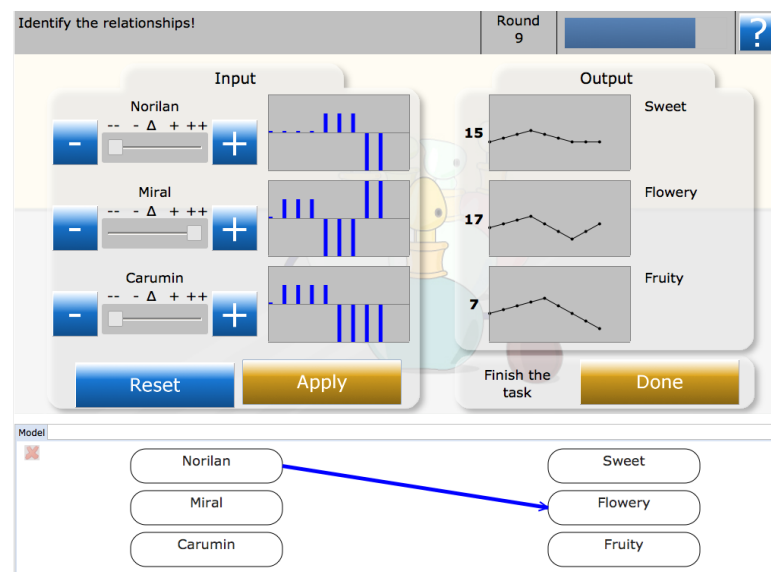


ABBILDUNG 2.1: Eine MicroDYN-Aufgabe in der Explorationsphase nach der im Rahmen dieses Dissertations- und des LLLight'in'Europe-Projekts umgesetzten Optimierung der Darstellung und Bedienbarkeit.

Obwohl umfangreiche Forschungen zu den psychometrischen Eigenschaften von MicroDYN für die KPL-Diagnostik bemerkenswerte Befunde liefern konnten (z. B. hohe interne Konsistenzen und ein theoretisch fundiertes, mehrdimensionales Messmodell; siehe Greiff et al., 2012; Schweizer, Wüstenberg & Greiff, 2013; Wüstenberg, Greiff & Funke, 2012; oder Messinvarianz hinsichtlich verschiedener Gruppen; siehe Greiff & Wüstenberg, 2014; für einen Überblick siehe auch Greiff, Fischer

et al., 2014), ist der MicroDYN-Ansatz nicht frei von Kritik (z. B. Funke, 2010, 2014; siehe aber Greiff & Martin, 2014).

Im Grunde zielt die Kritik am MicroDYN-Ansatz darauf ab, dass der Test nur einen begrenzten Ausschnitt einer potenziellen Problemlösefähigkeit erfasse. Tatsächlich gibt es bei den derzeitig verfügbaren MicroDYN-Aufgaben eine spezifische Strategie (d. h. eine erweiterte Form¹ von *Vary-one-thing-at-a-time*, VOTAT; Vollmeyer, Burns & Holyoak, 1996; sowie Tschirgi, 1980), die für alle Aufgaben optimal ist. Die Bewertungsvorschrift für die Explorationsleistung bei MicroDYN zielt daher auch nur darauf ab, ob ein Problemlöser die erweiterte VOTAT-Strategie angewendet hat oder nicht (siehe Greiff et al., 2012). Berücksichtigt man weiterhin, dass die Korrelation der Explorationsleistung mit der Wissenserwerbsleistung meist so hoch ist, dass eine empirische Trennung beider Facetten nicht gerechtfertigt erscheint (siehe z. B. Schweizer et al., 2013; Wüstenberg et al., 2012), das heißt die erweiterte VOTAT-Strategie auch für die Wissenserwerbsleistung außerordentlich relevant ist, dann wird die herausragende Rolle dieser einen Explorationsstrategie für MicroDYN umso deutlicher. Obwohl die Fokussierung auf die erweiterte VOTAT-Strategie – welche als guter Indikator für wissenschaftliches Denken angesehen werden kann (vgl. Chen & Klahr, 1999; Vollmeyer et al., 1996) – und die dadurch resultierende Homogenität der Aufgaben vor allem psychometrische Vorteile hat, steht damit letztendlich die Frage im Raum, ob die Konzentration auf derart spezifische, kognitive Anforderungen eine ausreichend inhaltsvalide Operationalisierung (vgl. Bühner, 2006; Kubinger, 2009; Messick, 1995; Schmidt-Atzert & Amelang, 2012) eines potenziellen KPL-Konstrukts darstellt. Es wird also hinterfragt, ob bei MicroDYN möglicherweise kognitive Anforderungen außer Acht gelassen werden, die als inhärent für die Lösung komplexer Systeme angesehen werden können (Funke, 2014). Tatsächlich ist MicroDYN mit den derzeitigen Aufgabenversionen kaum noch vergleichbar mit der Komplexität traditioneller Mikrowelten (siehe z. B. Dörner et al., 1983), weswegen ursprünglich wohl auch vom Minimal-Komplexe-Systeme-Ansatz hinsichtlich MicroDYN gesprochen wurde (siehe Greiff, 2012; Greiff & Funke, 2010).

Theoretische Überlegungen, wie diese Einschränkungen bei MicroDYN teilweise überwunden werden könnten, finden sich bei Greiff und Fischer (2013), wobei die

¹Eine erweiterte Form in dem Sinne, dass bei der Bewertung der Explorationsleistung die ursprüngliche VOTAT-Strategie um Nullrunden – das heißt Runden, in denen gar keine Manipulation der Eingabevariablen seitens des Problemlösenden stattfindet – ergänzt wurde (für Details siehe Greiff et al., 2012).

empirische Überprüfung noch aussteht. Kritisch anzumerken bleibt dabei, dass bestimmte Problemlöseanforderungen (z. B. qualitative Veränderungen nach Schwellenwertüberschreitungen; verschiedene Variablenbeziehungen in Subsystemen) in MicroDYN nur schwer oder gar nicht umgesetzt werden können (Neubert, Kretschmar, Wüstenberg & Greiff, 2015). Im Rahmen dieser Arbeit wurde deshalb ein anderer Ansatz gewählt, und zwar die Neuentwicklung des KPL-Tests MicroFIN (siehe Kapitel 3).

Zusammenfassend kann festhalten werden, dass mit MicroDYN und GeneticsLab in der langen Reihe der unterschiedlichen KPL-Messinstrumente (für einen Überblick siehe Funke, 2006) die bisher wohl psychometrisch überzeugendsten KPL-Tests vorliegen. Trotz der Einschränkungen ist damit ein wichtiger Schritt in der KPL-Forschung gelungen, um – in Kombination mit weiteren KPL-Operationalisierungen – die Frage nach der Konstruktvalidität von KPL (siehe Kapitel 5) in angemessener Art und Weise zu erforschen.

Literatur

- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Chen, Z. & Klahr, D. (1999). All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy. *Child Development*, 70, 1098–1120.
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Engelhart, M., Funke, J. & Sager, S. (2013). A decomposition approach for a new test-scenario in complex problem solving. *Journal of Computational Science*, 4(4), 245–254. doi:10.1016/j.jocs.2012.06.005
- Fischer, A., Greiff, S. & Funke, J. (2012). The process of solving complex problems. *The Journal of Problem Solving*, 4(1), 19–41. doi:10.7771/1932-6246.1118
- Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: an introduction to theory and research*. Addison-Wesley series in social psychology. Reading, Mass: Addison-Wesley Pub. Co.
- Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle. *Zeitschrift für Psychologie*, 193, 435–457.

- Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning*, 7(1), 69–89. doi:10.1080/13546780042000046
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Funke, J. (2006). Komplexes Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen* (S. 375–446). Göttingen: Hogrefe.
- Funke, J. (2010). Complex problem solving: a case for complex cognition? *Cognitive Processing*, 11, 133–142. doi:10.1007/s10339-009-0345-0
- Funke, J. (2014). Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.00739
- Funke, J. & Buchner, A. (1992). Finite Automaten als Instrumente für die Analyse von wissensgeleiteten Problemlöseprozessen: Vorstellung eines neuen Untersuchungsparadigmas. *Sprache & Kognition*, 11, 27–37.
- Greiff, S. (2012). *Individualdiagnostik der komplexen Problemlösefähigkeit*. Münster: Waxmann.
- Greiff, S. & Fischer, A. (2013). Measuring Complex Problem Solving: An educational application of psychological theories. *Journal for educational research online*, 5(1), 38–58.
- Greiff, S., Fischer, A., Stadler, M. & Wüstenberg, S. (2014). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning*, 21(3), 356–382. doi:10.1080/13546783.2014.989263
- Greiff, S. & Funke, J. (2010). Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme. Projekt Dynamisches Problemlösen. *Zeitschrift für Pädagogik. Beiheft*. 216–227.
- Greiff, S., Kretschmar, A., Müller, J. C., Spinath, B. & Martin, R. (2014). The computer-based assessment of complex problem solving and how it is influenced by students' information and communication technology literacy. *Journal of Educational Psychology*, 106(3), 666–680. doi:10.1037/a0035426
- Greiff, S. & Martin, R. (2014). What you see is what you (don't) get: a comment on Funke's (2014) opinion paper. *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.01120
- Greiff, S. & Wüstenberg, S. (2014). Assessment With Microworlds using Micro-DYN: Measurement Invariance and Latent Mean Comparisons: Psychometric Properties Across Several Student Samples and Blue-Collar Workers. *European Journal of Psychological Assessment*, 30(4), 1–11. doi:10.1027/1015-5759/a000194

- Greiff, S., Wüstenberg, S. & Funke, J. (2012). Dynamic Problem Solving: A New Assessment Perspective. *Applied Psychological Measurement*, 36(3), 189–213. doi:10.1177/0146621612439620
- Kluge, A. (2008). Performance Assessments With Microworlds and Their Difficulty. *Applied Psychological Measurement*, 32(2), 156–180. doi:10.1177/0146621607300015
- Kretzschmar, A., Neubert, J. C. & Greiff, S. (2014). Komplexes Problemlösen, schulfachliche Kompetenzen und ihre Relation zu Schulnoten. *Zeitschrift für Pädagogische Psychologie*, 28(4), 205–215. doi:10.1024/1010-0652/a000137
- Kretzschmar, A., Neubert, J. C., Wüstenberg, S. & Greiff, S. (2015). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence*. doi:10.1016/j.intell.2015.11.004
- Kröner, S. (2001). *Intelligenzdiagnostik per Computersimulation*. Münster: Waxmann.
- Kubinger, K. D. (2009). *Psychologische Diagnostik Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Weinheim: Beltz, Psychologie-Verl.-Union.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.
- Neubert, J. C., Kretzschmar, A., Wüstenberg, S. & Greiff, S. (2015). Extending the Assessment of Complex Problem Solving to Finite State Automata: Embracing Heterogeneity. *European Journal of Psychological Assessment*, 31(3), 181–194. doi:10.1027/1015-5759/a000224
- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik mit 82 Tabellen*. Berlin: Springer.
- Schweizer, F., Wüstenberg, S. & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*, 24, 42–52. doi:10.1016/j.lindif.2012.12.011
- Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin, R., ... Latour, T. (2012). The Genetics Lab: Acceptance and psychometric characteristics of a computer-based microworld assessing complex problem solving.

- Psychological Test and Assessment Modeling*, 54(1), 54–72. doi:10.1037/e578442014-045
- Sonnleitner, P., Keller, U., Martin, R. & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, 41(5), 289–305. doi:10.1016/j.intell.2013.05.002
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen: Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Süß, H.-M. (1999). Intelligenz und komplexes Problemlösen: Perspektiven für eine Kooperation zwischen differentiell-psychometrischer und kognitionspsychologischer Forschung. *Psychologische Rundschau*, 50(4), 220–228. doi:10.1026/0033-3042.50.4.220
- Süß, H.-M., Kersting, M. & Oberauer, K. (1991). Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen. *Diagnostica*, 37, 334–352.
- Tschirgi, J. E. (1980). Sensible Reasoning: A Hypothesis about Hypotheses. *Child Development*, 51(1), 1. doi:10.2307/1129583. JSTOR: 1129583?origin=crossref
- Vollmeyer, R., Burns, B. D. & Holyoak, K. J. (1996). The Impact of Goal Specificity on Strategy Use and the Acquisition of Problem Structure. *Cognitive Science*, 20(1), 75–100. doi:10.1207/s15516709cog2001_3
- Wagener, D. (1997). *Das komplexe Szenario FSYS*. Vortrag gehalten auf der.
- Wirth, J. & Funke, J. (2005). Dynamisches Problemlösen: Entwicklung und Evaluation eines neuen Messverfahrens zum Steuern komplexer Systeme. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern: Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 55–72). Wiesbaden: Verlag für Sozialwissenschaften.
- Wittmann, W. W. & Matt, G. E. (1986). Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts- und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des Berliner Intelligenzstrukturmodells. *Diagnostica*, 32, 309–329.
- Wittmann, W. W. & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In P. L. Ackerman, P. C. Kyllonen & R. Roberts

(Hrsg.), *Learning and individual differences: Process, trait and content determinants* (S. 77–104). Washington: APA.

Wüstenberg, S., Greiff, S. & Funke, J. (2012). Complex problem solving — More than reasoning? *Intelligence*, 40(1), 1–14. doi:10.1016/j.intell.2011.11.003

Kapitel 3

MicroFIN als neuer KPL-Test

Um die in Kapitel 2 dargestellten Einschränkungen von MicroDYN und GeneticsLab hinsichtlich der Aufgabenhomogenität zu überwinden und gleichzeitig die Operationalisierung eines potentiellen KPL-Konstrukts zu erweitern, wurde im Rahmen des LLLight'in'Europe- und dieses Dissertationsprojekts der KPL-Test MicroFIN entwickelt.

Im MicroFIN-Konzept (siehe ausführlich dazu Neubert, Kretzschmar, Wüstenberg & Greiff, 2015; sowie Greiff, Fischer, Stadler & Wüstenberg, 2014) wird der Ansatz finiter Automaten (Buchner & Funke, 1993; Funke & Buchner, 1992) mit dem MKS-Ansatz (siehe Kapitel 2) verbunden. Der Ansatz der finiten Automaten, so wie er hier genutzt wird, stellt eine bestimmte Formalisierung von Problemsituationen dar. Ein DVD-Player, ein Mobiltelefon oder ein Bahnfahrkartenautomat lassen sich beispielsweise als finite Automaten begreifen. All diesen Beispielen und damit grundsätzlich den finiten Automaten ist gemein, dass sie eine begrenzte Anzahl von Zuständen (z. B. ein/aus-Zustand des DVD-Players), Interventionsmöglichkeiten (z. B. verschiedene Tasten auf der Fernbedienung des DVD-Players) und Zustandsübergangsfunktionen (z. B. Einschalten durch Tastenbetätigung; automatisches Zurücksetzen des DVD-Players nach einer bestimmten Zeitdauer) besitzen. Durch die Anzahl an Zuständen, Interventionsmöglichkeiten und insbesondere verschiedenartigen Zustandsübergangsfunktionen lassen sich damit grundsätzlich eine Vielzahl von Problemsituationen beschreiben und Aufgaben konstruieren, die heterogene Anforderungen und Alltagssituationen abbilden.

Dementsprechend zielt das MicroFIN-Konzept darauf ab, das Potenzial der finiten Automaten (d. h. insbesondere die größere Aufgabenheterogenität) mit den

Vorteilen des MKS-Ansatzes (z. B. hohe Reliabilität durch multiple Aufgaben) zu kombinieren. Um dies zu gewährleisten, basiert die Konstruktion der MicroFIN-Aufgaben auf folgenden Richtlinien, welche unter Berücksichtigung der Entwicklung in der KPL-Diagnostik (siehe Kapitel 2) formuliert wurden:

1. Die Lösung der Aufgaben soll nicht hauptsächlich auf der erweiterten VOTAT-Strategie basieren, da mit MicroDYN und GeneticsLab zwei etablierte KPL-Tests für diese Strategie zur Verfügung stehen.
2. Die Aufgaben sollen verschiedenartige Anforderungen abbilden, so dass es keine beste Strategie zur Lösung über alle Aufgaben hinweg gibt.
3. Um eine Konfundierung von domänenspezifischem Vorwissen und Problemlöseleistung zu reduzieren, soll die Lösung der Aufgaben weitgehend unabhängig vom Sachwissen des Problemlösenden sein.
4. Die Aufgaben sollen die theoretisch angenommenen und in den etablierten KPL-Tests berücksichtigten KPL-Prozesse (d. h. Exploration zum Wissenserwerb, sowie Wissensanwendung zur Zielerreichung) zum Zwecke einer Facettendiagnostik abbilden.
5. Die Bearbeitung einer einzelnen Aufgabe soll entsprechend des MKS-Ansatzes nicht länger als 10 Minuten dauern, damit eine ausreichend große Anzahl an Aufgaben während einer Testsitzung vorgegeben werden kann.
6. Zusätzlich sollen die Bedienbarkeit und Darstellungsart eine Erhebung mit Tabletcomputern ermöglichen.

Auf der Basis dieser Richtlinien wurde die Konstruktion von 14 Aufgaben vorgenommen, welche in Tabelle 3.1 dargestellt sind.¹

Während der Aufgabenentwicklung fand ein mehrstufiger Begutachtungsprozess mit Experten aus dem Bereich Komplexes Problemlösen statt, um die oben genannten Richtlinien zur Aufgabenkonstruktion sicherzustellen. Durch diesen Evaluationsprozess konnte schrittweise eine Selektion von Aufgaben sowie gleichzeitig von Items innerhalb der einzelnen Aufgaben (siehe Abschnitt 3.1 zur Facettendiagnostik) vorgenommen werden. Neben der Steigerung der psychometrischen Qualität sollte damit auch die Erhebungszeit für MicroFIN signifikant reduziert werden

¹Die Entwicklung der MicroFIN-Aufgaben wurde von Jonas Neubert und André Kretzschmar, das heißt den beiden ersten Autoren von Neubert et al. (2015), zu gleichen Teilen vorgenommen.

TABELLE 3.1: Übersicht über die entwickelten MicroFIN-Aufgaben und deren Einsatz in verschiedenen Studien.

Aufgaben	Pilotierung	Studie 1	Studie 2	MicroFIN-Test
Circle-o-maton	x			
Concert-o-maton	x	x	x	x
Cook-o-maton	x		x	
Exchange-o-maton	x		x	
Fish-o-maton	x	x	x	x
Flooz-o-maton	x	x		
Green-o-maton	x		x	x
Mix-o-maton	x			
Navigate-o-maton	x			
Plant-o-maton	x	x		x
Plan-o-maton	x	x	x	x
Sluice-o-maton	x			
Wash-o-maton	x			x
4-Field-o-maton	x			

Anmerkungen: Studie 1: Neubert et al. (2014); Studie 2: Kretzschmar et al. (2015, Kapitel 7); MicroFIN-Test: Version des MicroFIN-Tests, wie er unter anderem im LLLight'in'Europe-Projekt oder bei Mainert et al. (eingereicht) eingesetzt wurde.

(von ca. 120 Minuten bei der Pilotierung auf ca. 50 Minuten beim MicroFIN-Test mit sechs Aufgaben). Als Resultat dieses Prozesses steht eine aktuelle Testversion mit sechs Aufgaben (sowie einer zusätzlichen Instruktionsaufgabe) für die KPL-Diagnostik zur Verfügung. Eine detaillierte Beschreibung einzelner MicroFIN-Aufgaben findet sich in Kapitel 7 sowie bei Neubert et al. (2015).

Nachfolgend sind die wichtigsten Eigenschaften der MicroFIN-Aufgaben sowie ein Überblick über die psychometrischen Befunde dargestellt, soweit sie für die hier vorliegende Arbeit relevant sind.

3.1 Facettendiagnostik

Ein Ziel der Aufgabenentwicklung war die bei der KPL-Diagnostik mittlerweile etablierte, differenzierte Leistungsbewertung der einzelnen Problemlösephasen (d. h. Exploration, Wissenserwerb und Wissensanwendung; Greiff, Wüstenberg & Funke, 2012; Sonnleitner, Keller, Martin, Latour & Brunner, in Druck) auch bei

MicroFIN umzusetzen. Im Gegensatz zu den MicroDYN- und GeneticsLab-Aufgaben, bei denen die Wissensabfrage parallel zur Exploration stattfindet (vgl. Abbildung 2.1), wurden bei MicroFIN alle drei Phasen getrennt voneinander umgesetzt. Dementsprechend ist zuerst das System zu explorieren, anschließend sind Fragen bezüglich des erworbenen Wissens zu beantworten und zum Schluss ist das Wissen zur Zielerreichung anzuwenden.

Explorationsphase

Innerhalb der Explorationsphase ist das unbekannte System in einer vorgegeben Zeitspanne (300 Sekunden) mit dem Ziel zu erkunden, sich Wissen über das System (d. h. über mögliche Zustände, Interventionsmöglichkeiten und Zustandsübergangsfunktionen) anzueignen, welches zur späteren Zielerreichung notwendig ist. Da die nachfolgenden Ziele während der Explorationsphase unbekannt sind (zur Notwendigkeit der Zielunspezifität siehe Vollmeyer & Funke, 1999), ist der Problemlösende also angehalten, das System möglichst vollständig zu erkunden.

Anhand des Fish-o-matons (siehe Abbildung 3.1) soll die Explorationsphase beispielhaft veranschaulicht werden. Bei dem Fish-o-maton sind die Füllstände der drei unteren Behälter (= Eingabevariablen) zu verändern und die Auswirkungen auf die Fische im Aquarium (= Ausgabevariable) zu beobachten. Bei dieser Aufgabe ist die entscheidende Regel, dass nur dann Fische im Aquarium zu sehen sind (= relevanter Zustand), wenn alle drei Behälter den gleichen Füllstand haben. Als zusätzliche Regel gilt, dass ein höherer Füllstand der Behälter (bei gleicher Füllhöhe) zu mehr Fischen im Aquarium führt. Alle Zustände, die auf ein Ungleichgewicht der Eingabevariablen beruhen, haben ein Aquarium ohne Fische zur Folge.

Um diese Regeln im Fish-o-maton zu erkennen, kann der Problemlösende alle möglichen Zustände der Eingabevariablen (64 Zustände) durchprobieren und sich die Variablenkombinationen der relevanten Zustände merken. Zwar kann damit im konkreten Fall das gesamte relevante Wissen über die Systemzustände erworben werden, allerdings verlangt dieses Vorgehen ein hohes Maß an Merkfähigkeit und stellt eine wenig effiziente Problemlösung dar.

Eine andere Möglichkeit ist es, äquivalent zu MicroDYN die VOTAT-Strategie (Vollmeyer, Burns & Holyoak, 1996) anzuwenden, das heißt die einzelnen Eingabevariable in allen Stufen nacheinander durchzuprobieren. Durch diese systematische

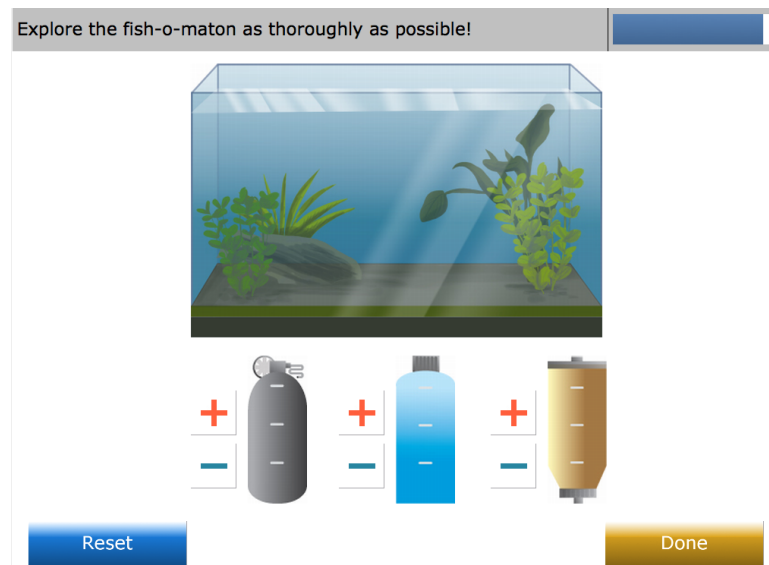


ABBILDUNG 3.1: Explorationsphase der MicroFIN-Aufgabe Fish-o-maton

Variation der Eingabemöglichkeiten kommt es jedoch nicht zu einem Gleichheitszustand der Eingabevariablen, so dass die relevanten Zustände (d. h. Fische im Aquarium) nicht beobachtet werden können (veranschaulicht in Abbildung A1 bei Neubert et al., 2015). Die VOTAT-Strategie ist (entsprechend der Konstruktionsrichtlinien der MicroFIN-Aufgaben) hier also nicht zielführend.

Hingegen ist ein sinnvolles Explorationsverhalten, wenn der Problemlösende nach dem Entdecken eines relevanten Zustands (vgl. Kluwe, Misiak & Heider, 1989) die damit einhergehende Hypothese „Gleichgewicht = Fische im Aquarium“ systematisch testet. Dies kann beispielsweise geschehen, indem gezielt in den relevanten Zustand ein- und ausgetreten wird (d. h. durch Manipulation der einzelnen Eingabevariablen). Müller, Kretzschmar und Greiff (2013) haben diese Explorationsstrategie *nested-VOTAT* genannt und konnten erste empirische Hinweise hinsichtlich der Nützlichkeit dieses Explorationsverhaltens aufzeigen. Entsprechend den oben dargestellten Richtlinien zur Aufgabenentwicklung scheint die *nested-VOTAT*-Strategie jedoch nicht für alle MicroFIN-Aufgaben sinnvoll zu sein.

Anzumerken ist, dass die Bewertung der Explorationsleistung bei MicroFIN noch am Anfang der Forschungsbemühungen steht. Aufgrund der Heterogenität der Aufgaben wird angenommen, dass in den verschiedenen Aufgaben auch unterschiedliche Explorationsstrategien als effektiv einzustufen sind. Demnach müsste für jede Aufgabe eine spezifische Bewertungsvorschrift entwickelt werden. Dazu wäre es beispielsweise möglich, äquivalent zu dem Vorgehen bei Müller et al. (2013) eine

potenziell beste Explorationsstrategie auf Basis der implementierten Systemregeln je Aufgabe festzulegen und zu bewerten. Dabei besteht jedoch grundsätzlich die Gefahr, dass weniger offensichtliche, aber trotzdem zielführende Explorationsstrategien nicht erkannt und entsprechend nicht bewertet werden. Eine andere Möglichkeit wäre es, verschiedene Explorationsstrategien auf der Basis der Logfiles² und *data-mining*-Analysen zu identifizieren (vgl. Gobert, Kim, Pedro, Kennedy & Betts, 2015), diese Verhaltensmuster mit der Problemlöseleistung in der Wissenserwerbs- oder Wissensanwendungsphase in Verbindung zu setzen und entsprechend eine (möglicherweise auch differenzierte) Bewertungsvorschrift für das Explorationsverhalten zu entwickeln.

Ob und in welcher Art die beiden unterschiedlichen Ansätze erfolgreich sind, ist derzeit noch nicht abzusehen. Die Bewertung des Explorationsverhaltens ist damit als zentraler Weiterentwicklungsschritt für MicroFIN anzusehen, so dass das Explorationsverhalten derzeit auch noch nicht bei der Bewertung der MicroFIN-Leistung berücksichtigt wird.

Wissensdiagnostik

Nach der Systemexploration wird in der zweiten Phase (ohne Zeitbegrenzung) Wissen abgefragt, welches sich der Problemlösende über das System angeeignet hat. Die Auswahl des abzufragenden Wissens orientiert sich dabei grundsätzlich an der Relevanz des Wissens für das System. Dementsprechend sind vor allem kritische Systemmerkmale für die spätere Zielerreichung (z. B. bestimmte Zustandsübergänge) in der Wissensdiagnostik enthalten (vgl. auch den KPL-Prozess der Informationsreduktion; Fischer, Greiff & Funke, 2012).

Grundsätzlich sind dafür verschiedene Aufgabenformate geeignet, wobei sich in den bisherigen Studien auf das Aufgabenformat „Initialzustand“ gestützt wurde. Bei diesem an die *retrognostic tasks* von Buchner und Funke (1993) angelehntem Aufgabenformat wird der Endzustand des Systems sowie eine direkt vorher ausgeführte Ein-Schritt-Intervention (z. B. einmaliges Betätigen eines bestimmten Schalters) dargestellt. Die Aufgabe des Problemlösenden ist es, den Anfangszustand mit Hilfe verschiedener Bausteine zu rekonstruieren (siehe Abbildung 3.2). Die Anzahl

²Ein Vorteil der computerbasierten Testadministration ist die Protokollierung einer Vielzahl von Informationen in den Logdateien, wie beispielsweise Bearbeitungszeiten oder die Anzahl der aktiven Interaktionen des Problemlösenden mit den Aufgaben.

der auszuwählenden Elemente variiert dabei je nach Aufgabe von einem bis zu acht Bausteinen. Bei dieser Art der Wissensdiagnostik wird also vornehmlich Regelwissen über das System abgeprüft, das heißt welche Auswirkungen (entsprechend der Zustandsübergangsfunktionen) bestimmte Interventionselemente auf das System haben. Üblicherweise werden zwei unterschiedliche „Initialzustand“-Items je MicroFIN-Aufgabe präsentiert.

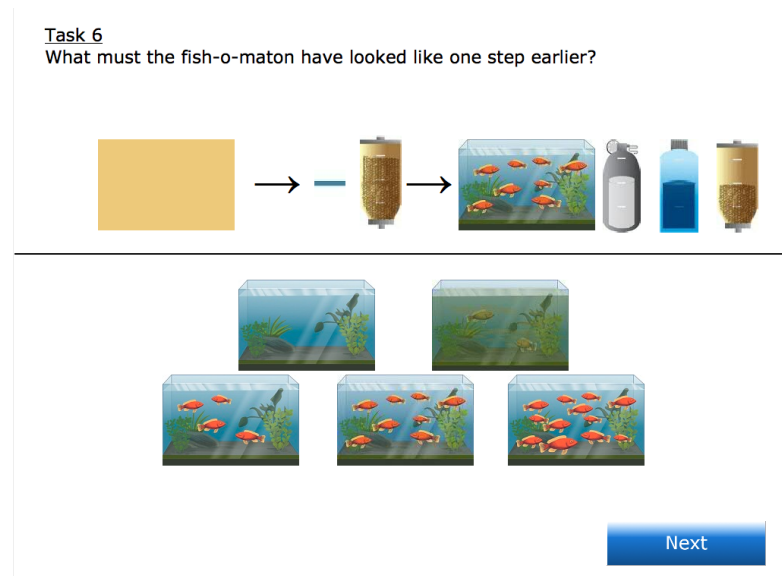


ABBILDUNG 3.2: Wissensdiagnostikphase „Initialzustand“ der MicroFIN-Aufgabe Fish-o-maton

Im Laufe des Entwicklungsprozesses für MicroFIN wurde auch ein zweites Aufgabenformat „Zustandsidentifikation“ implementiert, welches sich an die *verification tasks* von Buchner und Funke (1993) orientiert. Bei diesem Aufgabenformat werden potenzielle Zustände des Systems dargestellt und der Problemlösende soll entscheiden, ob dieser Zustand auftreten kann oder nicht (siehe Abbildung 3.3). Durch diese Art der Wissensdiagnostik wird vor allem Faktenwissen über das System abgefragt, in dem die Items sich beispielsweise auf Variablenbereiche (z.B. mögliche Anzahl der Fische im Aquarium) oder generelle Systemelemente (z.B. mögliche Zustände im Aquarium) beziehen. Die Bewertung folgt dem Konzept des sequentiellen Testens (Kubinger, 2009) zur Reduktion der Ratewahrscheinlichkeit bei üblicherweise vier Items. Das Aufgabenformat „Zustandsidentifikation“ fand in den bisherigen Analysen noch keine Berücksichtigung, so dass in zukünftigen Studien zu überprüfen ist, ob diese Art der Wissensdiagnostik einen zusätzlichen Nutzen zu der bisher eingesetzten Aufgabenart „Initialzustand“ leisten kann.

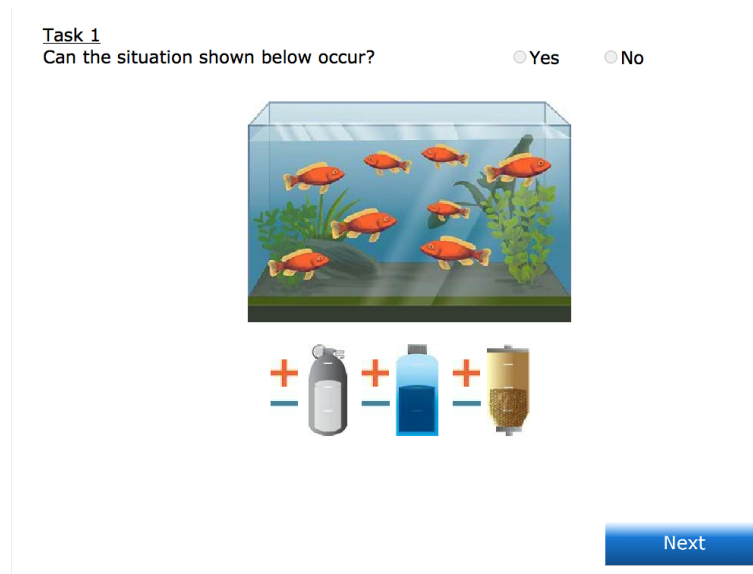


ABBILDUNG 3.3: Wissensdiagnostikphase „Zustandsidentifikation“ der MicroFIN-Aufgabe Fish-o-maton

Wissensanwendung

In der letzten Phase werden in der Regel zwei Items mit unterschiedlichen Zielzuständen vorgegeben, bei dem der Problemlösende das vorher durch die Exploration erworbene Wissen anwenden soll. Dazu ist der jeweilige Zielzustand in einer vorgegeben Zeitspanne (60 Sekunden) und in möglichst wenigen Schritten zu erreichen (siehe Abbildung 3.4). Für die Aufgabe Fish-o-maton wäre eine typische Aufgabenstellung: „Stellen Sie die Füllhöhe der Regler so ein, dass 8 Fische im Aquarium schwimmen!“ Die Bewertung der Wissensanwendung basiert derzeit darauf, ob der Zielzustand erreicht wurde oder nicht. Anzumerken ist, dass im Gegensatz zu MicroDYN dem Problemlösenden nicht die korrekte Systemstruktur präsentiert wird, also nur das vorher erworbene Wissen genutzt werden kann.

3.2 Empirische Überprüfung des Messmodells

Im vorherigen Abschnitt wurden die verschiedenen Aufgabenarten bei MicroFIN entsprechend der angestrebten Facettendiagnostik dargestellt. Ob sich diese angenommene Struktur auch in einem empirisch überprüfbareren Messmodell niederschlägt, wird im Sinne der faktoriellen Validität als eine zentrale Fragestellung der Testentwicklung angesehen (Bühner, 2006; Greiff, 2012). Im Folgenden werden die entsprechenden Ergebnisse für MicroFIN dargestellt.

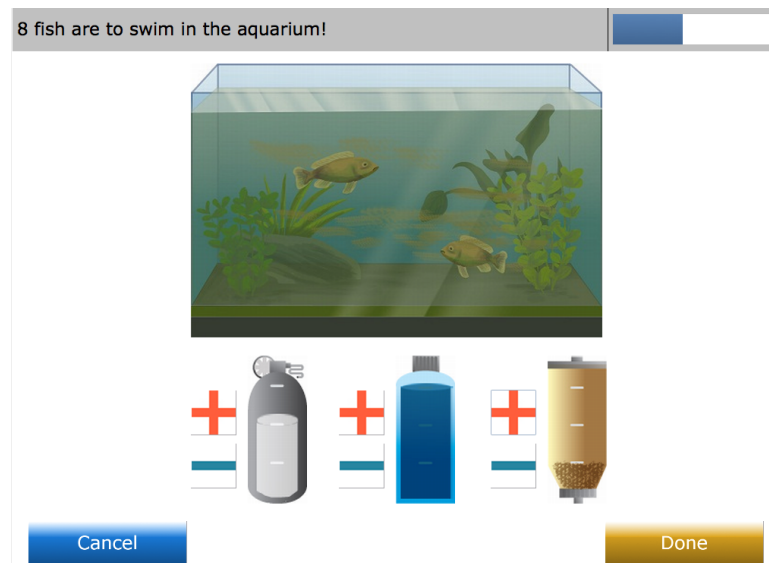


ABBILDUNG 3.4: Wissensanwendungsphase der MicroFIN-Aufgabe Fish-o-maton

Aufgrund der noch laufenden Entwicklungsarbeiten hinsichtlich der Bewertung des Explorationsverhaltens wurden in den bisherigen Studien mit MicroFIN nur die Phasen Wissenserwerb (d. h. die Aufgabenart „Initialzustand“) sowie Wissensanwendung (d. h. die Zielerreichungsaufgaben) berücksichtigt. Dies deckt sich mit dem Vorgehen bei MicroDYN, bei dem in der Regel die Analysen nur auf Basis von Wissenserwerb und -anwendung durchgeführt werden.³ Das daraus resultierende, zweidimensionale Messmodell von MicroDYN (d. h. zwei korrelierende Faktoren erster Ordnung für Wissenserwerb bzw. -anwendung) hat sich in einer Vielzahl von Studien auch im Vergleich zu einem eindimensionalen Messmodell (d. h. ein gemeinsamer Faktor erster Ordnung für Wissenserwerb und -anwendung) als sinnvoll erwiesen (siehe auch die im Rahmen dieser Arbeit durchgeführten Analysen; Kapitel 4.1, 6 und 7). Für MicroFIN wurde daher angenommen, dass dieses etablierte, zweidimensionale Messmodell ebenfalls eine zutreffende Repräsentation der KPL-Leistung darstellt (siehe Abbildung 3.5a).

In Tabelle 3.2 sind die Befunde hinsichtlich der Überprüfung des Messmodells von MicroFIN auf der Basis bisher verfügbarer Daten dargestellt.⁴ In der Arbeit von Neubert et al. (2015) konnte das zweidimensionale Modell bestätigt werden

³Bei MicroDYN wird in der Regel auf eine Berücksichtigung der Explorationsleistung verzichtet, da die hohe Korrelation zwischen Exploration und Wissenserwerb (z. B. $r = .97$ bei Wüstenberg, Greiff & Funke, 2012) eine empirische Trennung verhindert.

⁴Die Details zur statistischen Auswertung sowie zur verwendeten Software können Kapitel 7, Abschnitt „Data Analyses“ entnommen werden.

und zeigte sich dem eindimensionalen Modell überlegen. In der Arbeit von Kretzschmar, Neubert, Wüstenberg und Greiff (2015, Kapitel 7) wurde deutlich, dass weder für das zweidimensionale noch für das eindimensionale Modell ein ausreichend guter Modellfit erzielt werden konnte. Basierend auf den Daten von Mainert et al. (eingereicht) zeigte sich, dass das eindimensionale Messmodell gegenüber dem etablierten zweidimensionalen Messmodell vorzuziehen ist. Dabei hatten zusätzliche Modifikationen im Sinne von Aufgaben- oder Itemselektionen keinen Einfluss auf das Befundmuster.

Insgesamt ergibt sich dadurch ein uneinheitliches Bild bezüglich des Messmodells von MicroFIN. Auch wenn zu beachten ist, dass in den Studien jeweils unterschiedliche Aufgaben (siehe Tabelle 3.1) mit unterschiedlichen Stichproben (d. h. SchülerInnen, Studierende) verwendet wurden, spricht dies letztendlich nicht dafür, dass das in der bisherigen KPL-Forschung eingesetzte, zweidimensionale Messmodell ohne weiteres auch für MicroFIN genutzt werden kann.

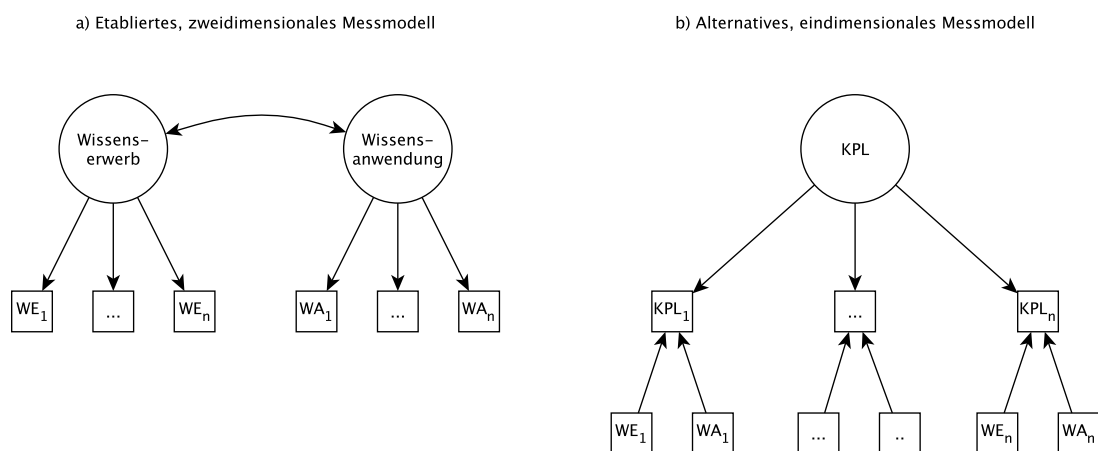


ABBILDUNG 3.5: Messmodelle für MicroFIN. Manifeste Indikatoren in Rechtecken, latente Faktoren in Ellipsen dargestellt. WE = Wissenswerb; WA = Wissensanwendung. Index entsprechend der n KPL-Aufgaben.

Als Konsequenz wurde ein alternatives Messmodell vorgeschlagen (siehe Kretzschmar et al., 2015, Kapitel 7), bei dem die einzelnen Items für Wissenserwerb und -anwendung je MicroFIN-Aufgabe zu einem Gesamtwert verrechnet werden. Diese aufgabenspezifischen Gesamtwerte werden anschließend als manifeste Indikatoren für einen globalen KPL-Faktor genutzt (siehe Abbildung 3.5b).

Das alternative Messmodell basiert dabei auf zwei Annahmen. (1) Die Wissensdiagnostik ist grundsätzlich reaktiv, das heißt durch die Wissensdiagnostik wird direkt oder indirekt Wissen über das System vermittelt (Süß, 1996; Süß, Kersting

& Oberauer, 1993). Dies trifft besonders für MicroFIN zu, indem beispielsweise Zustände in der Wissensdiagnostik präsentiert werden, die der Problemlösende womöglich bei der Exploration des Systems nicht erkundet hat. Durch die Präsentation von unbekannten Zuständen kann die Hypothesenbildung über die jeweilige Systemstruktur angeregt werden, so dass am Ende der Wissensdiagnostik durch schlussfolgerndes Denken zusätzliches Wissen für die anschließende Wissensanwendungsphase zur Verfügung stehen kann. Als Folge ist anzunehmen, dass die aufgabenspezifischen Indikatoren für Wissenserwerb und Wissensanwendung nicht unabhängig voneinander sind und jene zu einem gemeinsamen Indikator zusammengefasst werden können. (2) Spezifische Eigenschaften der MicroFIN-Aufgaben werden durch die Verwendung der Aufgabengesamtwerte betont, was insbesondere bei der Heterogenität der MicroFIN-Aufgaben nützlich sein kann (für unterschiedliche Repräsentationen eines Konstrukts auf Basis der Indikatoren siehe Bagozzi & Edwards, 1998). Wenn beispielsweise bei einer größeren Anzahl an MicroFIN-Aufgaben spezifische, kognitive Prozesse eine übergeordnete Rolle spielen sollten (z. B. Informationsreduktion bei sehr komplexen Problemsituationen), dann sollte sich dies in dem alternativen Messmodell durch die Verwendung der Aufgabengesamtwerte als Indikatoren deutlicher zeigen (z. B. durch Residualkorrelationen). Als Konsequenz könnten diese spezifischen, kognitiven Prozesse dann auch explizit und als adäquate Repräsentation der KPL-Leistung modelliert werden.

Die empirische Überprüfung des alternativen, eindimensionalen Messmodells auf der Datenbasis der drei Studien zeigt einen in der Regel guten Modellfit (siehe Tabelle 3.2) entsprechend den üblichen Kriterien nach Hu und Bentler (1999) sowie durchweg signifikante und substantielle Faktorladungen ($\lambda > .30$).

TABELLE 3.2: Modellfit der Messmodelle für MicroFIN in unterschiedlichen Studien.

Messmodell	χ^2	df	p	CFI	RMSEA 90% CI	WRMR/ SRMR *	Modellvergleich (2-dim vs. 1-dim)
Neubert et al. (2014)							
2-dimensional	52.30	34	.02	.99	.03 [.01 – .05]	.74	$\Delta\chi^2 = 24.50, df = 1, p = .00$
1-dimensional	89.22	35	.00	.97	.05 [.04 – .07]	.98	
alternativ	10.31	5	.07	.99	.04 [.00 – .08]	.02 *	
Kretzschmar et al. (2015)							
2-dimensional	78.26	53	.01	.89	.05 [.02 – .07]	.80	$\Delta\chi^2 = 0.33, df = 1, p = .57$
1-dimensional	78.68	54	.02	.89	.05 [.02 – .07]	.80	
alternativ	13.48	9	.14	.96	.05 [.00 – .10]	.04 *	
Mainert et al. (eingereicht)							
2-dimensional	53.97	34	.02	.91	.06 [.03 – .09]	.75	$\Delta\chi^2 = 0.13, df = 1, p = .71$
1-dimensional	53.91	35	.02	.92	.06 [.02 – .09]	.76	
alternativ	11.19	5	.05	.92	.09 [.01 – .16]	.05 *	

Anmerkungen: CFI=Comparative Fit Index; RMSEA=Root-Mean-Square Error of Approximation; WRMR=Weighted Root-Mean-Square Residual; SRMR=Standardized Root-Mean-Square Residual. Bei dem Modellvergleich wird das eindimensionale mit dem zweidimensionalen Messmodell verglichen. Das alternative Messmodell ist hinsichtlich der anderen Messmodelle nicht *nested*, kann also nicht mittels χ^2 -Differenzentest verglichen werden. WRMR für *Weighted Least Squares Means and Variance Adjusted* (WLSMV) basierte Schätzungen, SRMR für *Maximum Likelihood* (ML) basierte Schätzungen. Kleinere Unterschiede zwischen den in den Studien berichteten und den hier dargestellten Werten basieren auf der Verwendung unterschiedlicher Analysesoftware.

Als Kritikpunkt an dem alternativen, eindimensionalen Messmodell könnte eingewendet werden, dass spezifische Informationen entsprechend der postulierten Problemlöseprozesse (d. h. Wissenserwerb und -anwendung; Fischer et al., 2012; Funke, 2003) nicht mehr explizit für die weiteren Analysen zur Verfügung stehen (z. B. um den unterschiedlichen Einfluss der Problemlöseprozesse auf externe Kriterien zu untersuchen). Dem kann einerseits entgegnet werden, dass es bisher nicht gelungen ist, differentielle Effekte der einzelnen Problemlösefacetten nachzuweisen (z. B. inkrementelle Varianzaufklärung von Wissensanwendung über Wissenserwerb bzgl. externen Kriterien; vgl. Greiff & Fischer, 2013a, 2013b; Greiff et al., 2013; Greiff & Neubert, 2014; Schweizer, Wüstenberg & Greiff, 2013), so dass teilweise schon dazu übergegangen wurde, einen globalen KPL-Faktor zweiter Ordnung zu verwenden (z. B. Greiff & Fischer, 2013a; Greiff & Neubert, 2014; Kretzschmar, Neubert & Greiff, 2014). Andererseits zeigen sich regelmäßig hohe Korrelationen zwischen den latenten Faktoren des zweidimensionalen Messmodells (z. B. $r = .83$ bei Greiff & Neubert, 2014; $r = .88$ bei Greiff & Fischer, 2013a; oder $r = .93$ bei Berücksichtigung von Reasoning bei Greiff et al., 2013), so dass zu hinterfragen ist, ob die empirische Trennung zwischen Wissenserwerb und -anwendung tatsächlich nützlich ist. Insgesamt erscheint beim derzeitigen Forschungsstand der Verzicht der Nutzung von spezifischen Informationen bezüglich Wissenserwerb und -anwendung keinen Nachteil darzustellen.

Basierend auf dem aktuellen Forschungsstand kann festgehalten werden, dass das alternative, eindimensionale Messmodell im Gegensatz zum etablierten, zweidimensionalen Messmodell konsistent in den verschiedenen Studien bestätigt werden konnte. Auch wenn die hier dargestellten Befunde an weiteren Stichproben repliziert werden müssen, kann das alternative, eindimensionale Messmodell also derzeit als das bevorzugte Messmodell für MicroFIN angesehen werden.⁵

⁵Tatsächlich könnte das alternative, eindimensionale Messmodell – auch unter Berücksichtigung der oben dargestellten Vorbehalte gegenüber dem bisherigen, zweidimensionalen Messmodell – eine lohnenswerte Alternative auch für andere KPL-Tests darstellen. Eine exemplarische Überprüfung des alternativen, eindimensionalen Messmodell-Ansatzes für MicroDYN (basierend auf den Daten von Kretzschmar et al., 2015; und Neubert et al., 2015) ergab einen guten bis sehr guten Modellfit (entsprechend den üblichen Kriterien nach Hu & Bentler, 1999).

3.3 Item- und Skalenkennwerte

Zentrale Bewertungskriterien für psychologische Testitems im Sinne der klassischen Testtheorie sind unter anderem die Schwierigkeit, die Trennschärfe und die Reliabilität der Testaufgaben und -skalen (vgl. Bühner, 2006; Lienert & Raatz, 1998; Moosbrugger & Kelava, 2012). Nachfolgend werden die empirische Ergebnisse für MicroFIN auf Basis der bisher verfügbaren Studien zusammenfassend dargestellt. Zu beachten ist dabei, dass sich die Studien hinsichtlich der eingesetzten Aufgaben (siehe Tabelle 3.1) und der Stichprobeneigenschaften unterscheiden, so dass vergleichende Aussagen nur bedingt möglich sind.

Die Tabellen 3.3, 3.4 und 3.5 geben den Schwierigkeitsindex (nach Lienert & Raatz, 1998) sowie die Trennschärfe für die bei Neubert et al. (2015, $N = 576$ SchülerInnen), Kretzschmar et al. (2015, Kapitel 7; $N = 227$ Studierende) beziehungsweise Mainert et al. (eingereicht, $N = 154$ Studierende) eingesetzten Aufgaben wieder. Je Studie sind jeweils die Itemkennwerte für die Subskalen (d. h. Wissenserwerb und -anwendung nach dem etablierten, zweidimensionalen Messmodell) sowie für den Gesamtwert (basierend auf dem alternativen, eindimensionalen Messmodell, siehe Abschnitt 3.2) dargestellt. Aufgrund der methodischen Vorbehalte gegenüber Cronbachs α wurde für die interne Konsistenz der Skalen McDonalds ω als Reliabilitätsschätzung berechnet (siehe Dunn, Baguley & Brunsden, 2013; Revelle & Zinbarg, 2009).

TABELLE 3.3: Übersicht über die Item- und Skalenkennwerte der MicroFIN-Aufgaben bei Neubert et al. (2014); Schülerstichprobe

Aufgabe	W-Erwerb		W-Anwendung		Gesamtwert	
	P	r_{it}	P	r_{it}	P	r_{it}
1. Concert-o-maton	49	.49	77	.48	63	.62
2. Plan-o-maton	18	.40	70	.45	43	.55
3. Plant-o-maton	20	.47	67	.49	41	.57
4. Fish-o-maton	46	.49	76	.47	61	.56
5. Flooz-o-maton	35	.36	63	.43	49	.51
M (SD)	.34 (.14)		.71 (.06)		.51 (.10)	
McDonalds ω	.79		.78		.79	

Anmerkungen: W-Erwerb=Wissenserwerb, W-Anwendung=Wissensanwendung, P =Schwierigkeitsindex, r_{it} =Trennschärfe, M =Mittelwert der Skala, SD =Standardabweichung der Skala. Gesamtwert als Mittelwert von Wissenserwerb und Wissensanwendung entsprechend dem alternativen, eindimensionalen Messmodell (siehe Abschnitt 3.2).

TABELLE 3.4: Übersicht über die Item- und Skalenkennwerte der MicroFIN-Aufgaben bei Kretzschmar et al. (2015, Kapitel 7); Studierendenstichprobe

Aufgabe	W-Erwerb		W-Anwendung		Gesamtwert	
	P	r_{it}	P	r_{it}	P	r_{it}
1. Concert-o-maton	79	.36	95	.29	87	.41
2. Exchange-o-maton	64	.30	87	.24	76	.39
3. Fish-o-maton	76	.30	94	.27	85	.38
4. Plan-o-maton	69	.28	93	.26	81	.38
5. Green-o-maton	76	.35	89	.22	82	.42
6. Cook-o-maton	59	.16	82	.20	71	.25
M (SD)	.71 (.08)		.90 (.05)		.80 (.06)	
McDonalds ω	.66		.70		.65	

Anmerkungen: W-Erwerb=Wissenserwerb, W-Anwendung=Wissensanwendung, P =Schwierigkeitsindex, r_{it} =Trennschärfe, M =Mittelwert der Skala, SD =Standardabweichung der Skala. Gesamtwert als Mittelwert von Wissenserwerb und Wissensanwendung entsprechend dem alternativen, eindimensionalen Messmodell (siehe Abschnitt 3.2).

TABELLE 3.5: Übersicht über die Item- und Skalenkennwerte der MicroFIN-Aufgaben bei Mainert et al. (eingereicht); Studierendenstichprobe

Aufgabe	W-Erwerb		W-Anwendung		Gesamtwert	
	P	r_{it}	P	r_{it}	P	r_{it}
1. Green-o-maton	72	.31	77	.38	75	.34
2. Fish-o-maton	82	.27	81	.39	82	.41
3. Plan-o-maton	85	.39	70	.36	77	.56
4. Plant-o-maton	35	.25	69	.29	48	.39
5. Wash-o-maton	37	.25	56	.18	40	.36
M (SD)	.62 (.24)		.71 (.10)		.64 (.20)	
McDonalds ω	.58		.62		.60	

Anmerkungen: W-Erwerb=Wissenserwerb, W-Anwendung=Wissensanwendung, P =Schwierigkeitsindex, r_{it} =Trennschärfe, M =Mittelwert der Skala, SD =Standardabweichung der Skala. Gesamtwert als Mittelwert von Wissenserwerb und Wissensanwendung entsprechend dem alternativen, eindimensionalen Messmodell (siehe Abschnitt 3.2).

Hinsichtlich der Schwierigkeiten zeigt sich, dass die eingesetzten Aufgaben eher zu leicht sind. Eine Ausnahme bildet dabei die Skala Wissenserwerb bei der Schülerstichprobe von Neubert et al. (2015). Weiterhin wird deutlich, dass die Aufgaben das Fähigkeitsspektrum ungleichmäßig abbilden, das heißt es gibt mehrere leichte und wenige schwierige Aufgaben (siehe insbesondere die Skala Wissenserwerb bei Mainert et al., eingereicht). Der Vergleich der Wissenserwerbs- mit den Wissensanwendungsaufgaben zeigt, dass die Zielerreichung den Problemlösenden leichter fällt. Die Trennschärfen liegen eher am unteren Rand des akzeptablen Bereichs (vgl. Bühner, 2006; Moosbrugger & Kelava, 2012), wobei jedoch die Abhängigkeit des Trennschärfekoeffizienten unter anderem von der Itemschwierigkeit (d. h. über- oder unterdurchschnittliche Schwierigkeiten gehen in der Regel mit einer verringerten Trennschärfe einher) und der Homogenität der Stichproben (d. h. homogenere Stichproben führen aufgrund eingeschränkter Varianzen zu geringeren Trennschärfen) zu berücksichtigen ist (Bühner, 2006; Lienert & Raatz, 1998).

Die Reliabilitäten variieren je nach Skala und genutzten Aufgaben zwischen $\omega = .58$ und $\omega = .79$. Damit fallen die internen Konsistenzen wesentlich niedriger aus als beispielsweise bei MicroDYN und liegen in der Regel auch unter der Grenze eines gemeinhin als akzeptabel angesehenen Bereichs ($> .70$, vgl. Wilhelm & Kunina-Habenicht, 2015). Andererseits sollte jedoch die Schätzung der internen Konsistenz bei der relativ geringen Anzahl an Indikatoren nicht überbewertet werden, insbesondere dann nicht, wenn heterogene Aufgaben mit dem Ziel der Validitätssteigerung wie bei MicroFIN eingesetzt werden (vgl. Reliabilitäts-Validitäts-Dilemma; Bühner, 2006; Cronbach & Meehl, 1955; Lienert & Raatz, 1998; Messick, 1995; Wilhelm & Kunina-Habenicht, 2015).

Zusammenfassend lassen die Item- und Skalenkennwerte hinsichtlich der psychometrischen Qualitäten den Schluss zu, dass MicroFIN ähnlich wie MicroDYN und GeneticsLab zu Forschungszwecken geeignet erscheint. Bevor MicroFIN jedoch zur Individualdiagnostik (z. B. zur Personalauswahl) eingesetzt werden kann, sind zwingend Weiterentwicklungen mit dem Ziel verbesserter psychometrischer Itemkennwerte (z. B. ein breiteres Schwierigkeitsspektrum) notwendig.

3.4 Zusammenfassung

MicroFIN wurde mit dem Ziel entwickelt, die KPL-Diagnostik zu erweitern, so dass zusätzlich zu MicroDYN und GeneticsLab ein KPL-Test auf der Basis des MKS-Ansatzes zur Verfügung steht, welcher jedoch heterogenere Anforderungen als die bisherigen KPL-Tests an den Problemlösenden stellt. Die auf dem Ansatz der finiten Automaten bisher entwickelten Aufgaben zeigten in unterschiedlichen Studien zu Forschungszwecken ausreichend gute psychometrische Eigenschaften, so dass MicroFIN in Kombination mit MicroDYN für die im zweiten Teil dieser Arbeit im Mittelpunkt stehenden Fragestellungen der Konstruktvalidität von KPL geeignet erscheint. Befunde zur Validität von MicroFIN sind entsprechend im Kapitel 5 dargestellt und werden an dieser Stelle nicht weiter ausgeführt.

Abschließend bleibt festzuhalten, dass die Entwicklung von MicroFIN zum jetzigen Zeitpunkt noch nicht abgeschlossen ist. Insbesondere die Bewertung der Explorationsleistung, die Integration zusätzlicher Aufgabenformate für die Wissensdiagnostik (z. B. die „Zustandsidentifikation“-Aufgaben) sowie die Skalierung der Aufgabenschwierigkeit anhand spezifischer Aufgabeneigenschaften (vgl. Greiff, 2012; Greiff, Krkovic & Nagy, 2014) sind Punkte mit Entwicklungs- und Verbesserungspotenzial.

Literatur

- Bagozzi, R. P. & Edwards, J. R. (1998). A General Approach for Representing Constructs in Organizational Research. *Organizational Research Methods*, 1(1), 45–87. doi:10.1177/109442819800100104
- Buchner, A. & Funke, J. (1993). Finite-state automata: Dynamic task environments in problem-solving research. *The Quarterly Journal of Experimental Psychology Section A*, 46(1), 83–118. doi:10.1080/14640749308401068
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281.
- Dunn, T. J., Baguley, T. & Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 1–14. doi:10.1111/bjop.12046

- Fischer, A., Greiff, S. & Funke, J. (2012). The process of solving complex problems. *The Journal of Problem Solving*, 4(1), 19–41. doi:10.7771/1932-6246.1118
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Funke, J. & Buchner, A. (1992). Finite Automaten als Instrumente für die Analyse von wissensgeleiteten Problemlöseprozessen: Vorstellung eines neuen Untersuchungsparadigmas. *Sprache & Kognition*, 11, 27–37.
- Gobert, J. D., Kim, Y. J., Pedro, M. A. S., Kennedy, M. & Betts, C. G. (2015). Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld. *Thinking Skills and Creativity*. doi:10.1016/j.tsc.2015.04.008
- Greiff, S. (2012). *Individualdiagnostik der komplexen Problemlösefähigkeit*. Münster: Waxmann.
- Greiff, S. & Fischer, A. (2013a). Der Nutzen einer komplexen Problemlösekompetenz: Theoretische Überlegungen und empirische Befunde. *Zeitschrift für Pädagogische Psychologie*, 27(1), 27–39. doi:10.1024/1010-0652/a000086
- Greiff, S. & Fischer, A. (2013b). Measuring Complex Problem Solving: An educational application of psychological theories. *Journal for educational research online*, 5(1), 38–58.
- Greiff, S., Fischer, A., Stadler, M. & Wüstenberg, S. (2014). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning*, 21(3), 356–382. doi:10.1080/13546783.2014.989263
- Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M. & Martin, R. (2013). A multitrait-multimethod study of assessment instruments for complex problem solving. *Intelligence*, 41(5), 579–596. doi:10.1016/j.intell.2013.07.012
- Greiff, S., Krkovic, K. & Nagy, G. (2014). The systematic variation of task characteristics facilitates the understanding of task difficulty: A cognitive diagnostic modeling approach to complex problem solving. *Psychological Test and Assessment Modeling*, 56(1).
- Greiff, S. & Neubert, J. C. (2014). On the relation of complex problem solving, personality, fluid intelligence, and academic achievement. *Learning and Individual Differences*, 36, 37–48. doi:10.1016/j.lindif.2014.08.003
- Greiff, S., Wüstenberg, S. & Funke, J. (2012). Dynamic Problem Solving: A New Assessment Perspective. *Applied Psychological Measurement*, 36(3), 189–213. doi:10.1177/0146621612439620

- Hu, L.-t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi:10.1080/10705519909540118
- Kluwe, R. H., Misiak, C. & Heider, H. (1989). Erste Ergebnisse zu einem Modell der Steuerung eines komplexen Systems. In D. Dörner & H. Wegener (Hrsg.), *Idola fori et idola theatri* (S. 101–119). Göttingen: Verlag für Psychologie.
- Kretzschmar, A., Neubert, J. C. & Greiff, S. (2014). Komplexes Problemlösen, schulfachliche Kompetenzen und ihre Relation zu Schulnoten. *Zeitschrift für Pädagogische Psychologie*, 28(4), 205–215. doi:10.1024/1010-0652/a000137
- Kretzschmar, A., Neubert, J. C., Wüstenberg, S. & Greiff, S. (2015). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence*. doi:10.1016/j.intell.2015.11.004
- Kubinger, K. D. (2009). Psychologische Computerdiagnostik. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 57(1), 23–32.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Weinheim: Beltz, Psychologie-Verl.-Union.
- Mainert, J., Baggen, Y., Kretzschmar, A., Neubert, J. C., Lans, T., Biemans, H., ... Greiff, S. (eingereicht). Perceiving Entrepreneurial Challenges as Complex Problems: The Role of Complex Problem Solving in Opportunity Identification. *Journal of Business Venturing*.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. Berlin: Springer.
- Müller, J. C., Kretzschmar, A. & Greiff, S. (2013). Exploring exploration: inquiries into exploration behavior in Complex Problem Solving assessment. In *Proceedings of the 6th International Conference on Educational Data Mining* (S. 336–337).
- Neubert, J. C., Kretzschmar, A., Wüstenberg, S. & Greiff, S. (2015). Extending the Assessment of Complex Problem Solving to Finite State Automata: Embracing Heterogeneity. *European Journal of Psychological Assessment*, 31(3), 181–194. doi:10.1027/1015-5759/a000224

- Revelle, W. & Zinbarg, R. E. (2009). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145–154. doi:10.1007/s11336-008-9102-z
- Schweizer, F., Wüstenberg, S. & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*, 24, 42–52. doi:10.1016/j.lindif.2012.12.011
- Sonnleitner, P., Keller, U., Martin, R., Latour, T. & Brunner, M. (in Druck). Assessing Complex Problem Solving in the Classroom: Meeting Challenges and Opportunities. In B. Csapo & J. Funke (Hrsg.), *The Nature of Problem Solving*. Paris: OECD.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen: Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Süß, H.-M., Kersting, M. & Oberauer, K. (1993). Zur Vorhersage von Steuerungsleistungen an computersimulierten Systemen durch Wissen und Intelligenz. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 14, 189–203.
- Vollmeyer, R., Burns, B. D. & Holyoak, K. J. (1996). The Impact of Goal Specificity on Strategy Use and the Acquisition of Problem Structure. *Cognitive Science*, 20(1), 75–100. doi:10.1207/s15516709cog2001_3
- Vollmeyer, R. & Funke, J. (1999). Personen- und Aufgabenmerkmale beim komplexen Problemlösen. *Psychologische Rundschau*, 50, 213–219.
- Wilhelm, O. & Kunina-Habenicht, O. (2015). Pädagogisch-psychologische Diagnostik. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 305–328). Berlin: Springer.
- Wüstenberg, S., Greiff, S. & Funke, J. (2012). Complex problem solving — More than reasoning? *Intelligence*, 40(1), 1–14. doi:10.1016/j.intell.2011.11.003

Kapitel 4

ICT-Literacy als Herausforderung für die KPL-Diagnostik

Aufgrund der besonderen Aufgabenanforderungen innerhalb der KPL-Diagnostik (z. B. Interaktivität, Dynamik; vgl. Funke, 2001, 2003), sind die KPL-Tests durchweg computerbasiert (Wirth & Klieme, 2003). Dabei haben schon frühere Studien gezeigt, dass die Leistung bei der KPL-Leistungsmessung auch von der *ICT-Literacy* (d. h. Kompetenz im Umgang mit Computern) der Problemlöser abhängig sein kann (vgl. Süß, 1996; Wagener, 2001). Dies ist insofern problematisch, da eine systematische Benachteiligung von ProbandenInnen mit geringerer *ICT-Literacy* einer Verletzung des Testgütekriteriums der Testfairness entspräche (Bühner, 2006; Kubinger, 2009). Letztendlich kann aber auch argumentiert werden, dass, solange *ICT-Literacy* nicht als Bestandteil eines potentiellen KPL-Konstrukts angesehen wird, die KPL-Messung bei einem substantiellen Einfluss von *ICT-Literacy* nicht intendierte Leistungen umfasst, so dass die Konstruktvalidität der jeweiligen KPL-Tests gefährdet wäre (Messick, 1995; siehe umfassend dazu auch bei Greiff, Kretzschmar, Müller, Spinath & Martin, 2014, Abschnitt 4.1).

Das Ausmaß des Einflusses der *ICT-Literacy* ist dabei ganz offensichtlich messinstrumentenabhängig (vgl. Kleinmann & Strauß, 1995; Kyllonen, 2009; Mead & Drasgow, 1993). Je komplexer die Interaktionen mit dem computerbasierten Messinstrument sind (z. B. Umschalten zwischen verschiedenen Programmoberflächen, unterschiedliche Eingabeoptionen oder im Sinne der *Usability* suboptimale Bedienfunktionen), desto größer ist die Wahrscheinlichkeit, dass der Einfluss der *ICT-Literacy* zunimmt. Sind hingegen nur sehr einfache Interaktionen mit dem

Computer zur Bearbeitung der Aufgaben notwendig (z. B. einfacher Tastendruck bei Reaktionszeitaufgaben), dann sollte nur ein vernachlässigbarer Einfluss der *ICT-Literacy* vorliegen. In diesem Sinne ist zu überprüfen, welche Rolle die *ICT-Literacy* bei spezifischen KPL-Tests einnimmt. In Studie 1 (Greiff et al., 2014, Abschnitt 4.1) wurde daher diese Fragestellung für MicroDYN aufgegriffen und analysiert, ob es einen unterschiedlichen Einfluss von *ICT-Literacy* bei MicroDYN im Vergleich zu computerbasierten Tests etablierter, kognitiver Fähigkeiten gibt. Anschließend werden im Abschnitt 4.2 erste Ergebnisse hinsichtlich des Einfluss von *ICT-Literacy* bei MicroFIN dargestellt.

Literatur

- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning*, 7(1), 69–89. doi:10.1080/13546780042000046
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Greiff, S., Kretzschmar, A., Müller, J. C., Spinath, B. & Martin, R. (2014). The computer-based assessment of complex problem solving and how it is influenced by students' information and communication technology literacy. *Journal of Educational Psychology*, 106(3), 666–680. doi:10.1037/a0035426
- Kleinmann, M. & Strauß, B. (1995). Softwareergonomische Voraussetzungen computersimulierter Szenarien. In B. Strauß & M. Kleinmann (Hrsg.), *Computersimierte Szenarien in der Personalarbeit* (S. 127–141). Göttingen: Verlag für Angewandte Psychologie.
- Kubinger, K. D. (2009). *Psychologische Diagnostik Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kyllonen, P. (2009). New constructs, methods, and directions for computer-based assessment. In F. Scheuermann & J. Björnsson (Hrsg.), *The transition to computer-based assessment* (S. 151–156). Luxembourg: Office for Official Publications of the European Communities.
- Mead, A. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449–458.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.

- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen: Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Wagener, D. (2001). *Psychologische Diagnostik mit komplexen Szenarios - Taxonomie, Entwicklung, Evaluation*. Lengerich: Pabst Science Publishers.
- Wirth, J. & Klieme, E. (2003). Computer-based Assessment of Problem Solving Competence. *Assessment in Education: Principles, Policy & Practice*, 10(3), 329–345. doi:10.1080/0969594032000148172

4.1 Studie 1: ICT-Literacy und MicroDYN

Greiff, S., Kretschmar, A., Müller, J. C., Spinath, B., & Martin, R. (2014). The computer-based assessment of complex problem solving and how it is influenced by students' information and communication technology literacy. *Journal of Educational Psychology*, 106(3), 666–680. <http://doi.org/10.1037/a0035426>

4.2 ICT-Literacy und MicroFIN

In Studie 1 (Abschnitt 4.1) wurde der Einfluss von *ICT-Literacy* auf die KPL-Diagnostik am Beispiel von MicroDYN aufgezeigt. Allerdings sollte die Befundlage nicht unreflektiert auf andere KPL-Tests generalisiert werden, da unterschiedliche Bedienungsweisen, Aufgabenformate usw. den Einfluss von *ICT-Literacy* verstärken oder verringern können.

Daher stellt sich die Frage, inwieweit die Ergebnisse der Studie 1 auch für MicroFIN gelten, welches durch andere Bedienkonzepte und Aufgabenimplementierungen gekennzeichnet ist (zum Vergleich siehe Abbildung 2.1 für MicroDYN und Abbildung 3.1 für MicroFIN). Da bisher keine Befunde bezüglich des Einflusses von *ICT-Literacy* auf die Leistung bei MicroFIN vorliegen, wurden Informationen zur *ICT-Literacy* für eine Teilstichprobe von Kretzschmar, Neubert, Wüstenberg und Greiff (2015, Kapitel 7) im Rahmen einer Zusatzbefragung erhoben. Obwohl diese Nachbefragung als Pilotstudie anzusehen ist, lassen sich doch erste Hinweise zur Relevanz von *ICT-Literacy* bei MicroFIN ableiten. Nachfolgend wird die Studie beschrieben, wobei hauptsächlich Informationen dargestellt werden, die nicht bei Kretzschmar et al. (2015, siehe Kapitel 7) berichtet werden.

Stichprobe

Insgesamt konnten die Daten von 91 Personen (Teilnehmerquote: 40%) im Rahmen der Nacherhebung gewonnen werden, in der Informationen zur *ICT-Literacy* der Teilnehmenden erhoben wurden. Ein Vergleich der Nacherhebungsteilnehmenden mit den Nicht-Teilnehmenden zeigte zum Teil signifikante Unterschiede bei einigen kognitiven Leistungsmaßen in dem Sinne, dass die Nacherhebungsteilnehmenden höhere Leistungen zeigten (siehe Tabelle A.1 im Anhang). Bei der Teilstichprobe handelt es sich damit um eine selektierte Stichprobe, was den Pilotcharakter dieser Studie unterstreicht.

Instrumente

Zur Erhebung der *ICT-Literacy* wurden drei unterschiedliche, etablierte Wissensfragebögen eingesetzt: die Subskalen Praktisches Computerwissen (PRACOWI; 20

Aufgaben) und Theoretisches Computerwissen (TECOWI, 20 Aufgaben) des Inventars zur Computerbildung INCOBI-R (Richter, Naumann & Horz, 2010) sowie die Kurzform A (schwer) der Testbatterie für Berufseinsteiger – START-C (24 Aufgaben; Wagener, 2007). Für jede Skala wurde entsprechend dem Manual der Skalensummenwert gebildet, wobei zufriedenstellende interne Konsistenzen (McDonalds ω ; siehe Dunn, Baguley & Brunsden, 2013; Revelle & Zinbarg, 2009) erreicht wurden: $\omega_{PRACOWI} = .84$, $\omega_{TECOWI} = .87$ und $\omega_{START-C} = .86$. Die manifesten Korrelationen zwischen den drei Skalen waren signifikant und von großer Effektstärke: $r_{START-C.PRACOWI} = .52$, $r_{START-C.TECOWI} = .68$ und $r_{TECOWI.PRACOWI} = .74$. Da keine differentiellen Effekte hinsichtlich der unterschiedlichen *ICT-Literacy*-Skalen erwartet wurden, wurden die drei z-standardisierten Skalenwerte auch im Sinne der Reliabilitätssteigerung durch Aggregation (Wittmann & Matt, 1986) zu einem Gesamtwert zusammengefasst.

Die weiteren Erhebungsinstrumente MicroFIN, MicroDYN¹, BIS-Test und BO-WIT entsprechen vollständig den Beschreibungen bei Kretzschmar et al. (2015) und sind im Kapitel 7 dargestellt.

Durchführung

Die Teilnehmenden der Hauptstudie wurden nach Beendigung der Erhebung via E-Mail kontaktiert und zu der Nacherhebung eingeladen. Dazu wurde ein Onlinefragebogen zur Verfügung gestellt, welcher ohne Zeitlimit ausgefüllt werden konnte. Die Teilnehmenden nahmen freiwillig teil, wobei jeder Studienteilnehmende für das erfolgreiche Ausfüllen des Fragebogens einen Amazon-Gutschein im Wert von 5 Euro erhalten und zusätzlich an einer Verlosung von 20-Euro-Amazon-Gutscheinen teilgenommen hat.

Ergebnisse

Die deskriptiven Statistiken und manifesten Korrelationen finden sich in Tabelle A.1 (Anhang). Die Korrelationen weisen auf einen signifikanten Zusammenhang mittlerer Stärke zwischen *ICT-Literacy* und der Leistung bei MicroFIN hin. Da *ICT-Literacy*, insbesondere wenn wie hier über Wissenstests operationalisiert, substantielle Überlappungen mit Intelligenz aufweist (vgl. Investmenttheorie; Cattell,

¹MicroDYN wurde zum Zwecke der Vergleichbarkeit in die Analysen mit eingeschlossen.

1963), wurde zusätzlich um den Einfluss der Intelligenz bei den Analysen kontrolliert. Die Partialkorrelationen in Tabelle 4.1 zeigen an, dass nach Kontrolle der Intelligenzfähigkeiten die Korrelation zwischen der Leistung bei MicroFIN und der *ICT-Literacy* abfällt, die Signifikanzschwelle jedoch nur knapp verfehlt wird ($p = .05$). Das in etwa gleiche Befundmuster zeigt sich für MicroDYN und *ICT-Literacy*. Die Korrelationen zwischen der MicroFIN-Leistung und *ICT-Literacy* sowie der MicroDYN-Leistung und *ICT-Literacy* unterscheiden sich nicht signifikant voneinander ($z = .51, p > .10$).

TABELLE 4.1: Korrelationen zwischen der Leistung bei MicroFIN sowie MicroDYN und *ICT-Literacy* mit und ohne Kontrolle der Intelligenz

	FIN	DYN	ICT	BIS-K	BIS-B	BIS-M	BIS-E	BOWIT
FIN	–	.51*	.31*	.40*	.28*	.35*	.06	.26*
DYN	.49*	–	.38*	.52*	.28*	.24*	.12	.25*
ICT	.22	.29*	–	.29*	.10	.03	.08	.36*

Anmerkungen: FIN = MicroFIN; DYN = MicroDYN; ICT = ICT-Literacy; BIS-K = Reasoning; BIS-B = Bearbeitungsgeschwindigkeit; BIS-M = Merkfähigkeit; BIS-E = Einfallsreichtum; BOWIT = Bochumer Wissenstest. Oberhalb der Hauptdiagonalen bivariate Pearson-Korrelationen, unterhalb Partialkorrelationen nach Kontrolle von BIS- und BOWIT-Leistungen. * $p < .05$

Um die dargestellten Befunde abzusichern, wurden die Analysen in einem Pfadmodell bei gleichzeitiger Betrachtung aller Einflussgrößen erweitert. Zusätzlich wurde auf das *Full-Information-Maximum-Likelihood-Verfahren* (FIML) zum Umgang mit fehlenden Werten zurückgegriffen (siehe Lüdtke, Robitzsch, Trautwein & Köller, 2007), so dass der vollständige Datensatz von $n = 227$ genutzt werden konnte (siehe Kapitel 7). In Abbildung 4.1 sind die Ergebnisse der Analyse zusammengefasst: *ICT-Literacy* ist neben der Intelligenz ein substantieller Prädiktor für MicroFIN und MicroDYN. Im Vergleich zu einem Modell (hier nicht näher dargestellt), in dem die Varianz der KPL-Leistung bei MicroFIN nur durch Intelligenz erklärt wird, kann durch die Hinzunahme von *ICT-Literacy* als weiteren Prädiktor 5.0% zusätzliche Varianz erklärt werden: $F(1, 220) = 15.56, p < .01$. Die Ergebnisse für MicroDYN zeigen ein ähnliches Befundmuster (d. h. signifikante inkrementelle Varianzaufklärung von 5.6%, $F(1, 220) = 18.87, p < .01$).

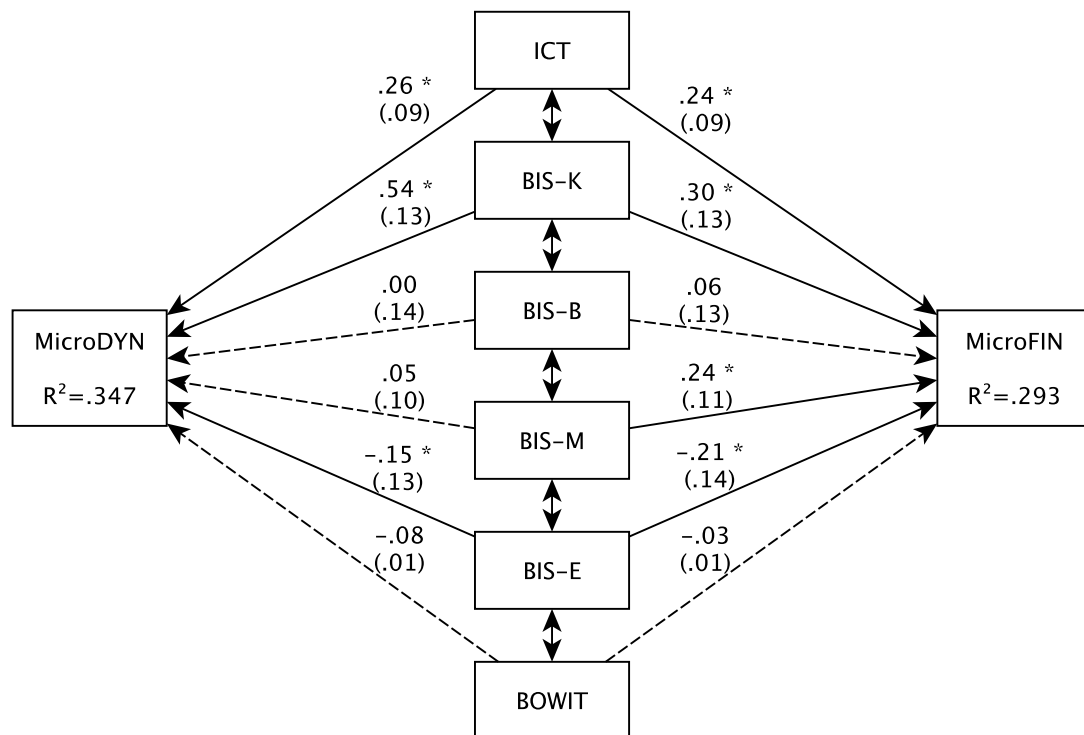


ABBILDUNG 4.1: Pfadmodell hinsichtlich der statistischen Vorhersage von MicroFIN und MicroDYN durch *ICT-Literacy* und Intelligenz. BIS-K = Reasoning; BIS-B = Bearbeitungsgeschwindigkeit; BIS-M = Merkfähigkeit; BIS-E = Einfallsreichtum; BOWIT = Bochumer Wissenstest. Die Pfadmodelle für MicroFIN und MicroDYN wurden separat berechnet und hier gemeinsam in einer Abbildung dargestellt. Die Korrelationen zwischen *ICT-Literacy* und jeweils den einzelnen Intelligenzfacetten wurden der Übersichtlichkeit halber nicht eingezeichnet. * $p < .05$

Diskussion

Insgesamt lassen die hier dargestellten Ergebnisse darauf schließen, dass aufgrund der computerbasierten Erhebungsart auch bei MicroFIN die *ICT-Literacy* eine signifikante Rolle spielt, welche jedoch eher zu vernachlässigen ist, wenn zusätzliche allgemeine kognitive Fähigkeiten berücksichtigt werden. Damit konnten Teile des bei Greiff, Kretschmar, Müller, Spinath und Martin (2014) dargestellten Befundmusters für MicroDYN konzeptionell auch für MicroFIN repliziert werden. Weiterhin kann vorläufig festgehalten werden, dass trotz der verschiedenartigen Aufgabenpräsentationen bei MicroFIN und MicroDYN der Einfluss der *ICT-Literacy* bei beiden vergleichbar ist.

Einschränkend für die Befunde ist natürlich die relativ kleine sowie selektive Stichprobe (d.h. durchweg junge Studierende, die als *digital natives* charakterisiert

werden können; Prensky, 2001) dieser Studie zu berücksichtigen. Auch hat sich in Studie 1 (Abschnitt 4.1) gezeigt, dass neben dem Sachwissen auch das Handlungswissen sowie die affektive Komponente von *ICT-Literacy* (siehe Tsai, 2002) einen Einfluss haben kann. Daher sind weitere Forschungsarbeiten zu MicroFIN und *ICT-Literacy* notwendig, in denen größere und heterogenere Stichproben (d. h. insbesondere mit einer größeren Varianz hinsichtlich der *ICT-Literacy*) sowie weitere Facetten der *ICT-Literacy* berücksichtigt werden sollten.

4.3 Zusammenfassung

Die Forschung zum Einfluss der computerbasierten, psychologischen Diagnostik hat immer wieder gezeigt, dass die computerbasierte Darbietung von Tests mit besonderen Herausforderungen einhergeht (vgl. Clariana & Wallace, 2002; Kyllonen, 2009; Mead & Drasgow, 1993; Russell, Goldberg & O’connor, 2003). Die KPL-Diagnostik, welche aufgrund der besonderen Aufgabeneigenschaften (z. B. Intransparenz, Dynamik) grundsätzlich computerbasiert umgesetzt wird, ist davon nicht ausgeschlossen. Wenn jedoch die Leistungen in den KPL-Tests substantiell durch *ICT-Literacy* beeinflusst werden, dann werden Testteilnehmende mit geringerer *ICT-Literacy* systematisch benachteiligt – unter der Voraussetzung, dass *ICT-Literacy* nicht als Bestandteil eines potentiellen KPL-Konstrukt angesehen wird.

Die hier dargestellten Befunde zeigen, dass es einen Einfluss der *ICT-Literacy* auf die Leistung bei MicroDYN und MicroFIN gibt. Dieser Einfluss war jedoch für MicroDYN in zwei von drei Studien vergleichbar mit dem Einfluss der *ICT-Literacy* bei computerbasierten Darbietungen etablierter, kognitiver Leistungstest (siehe Abschnitt 4.1). Hinsichtlich MicroFIN konnten erste Hinweise aufgezeigt werden, dass der Einfluss von *ICT-Literacy* zu einem substantiellen Teil auf allgemeine, kognitive Leistungen zurückzuführen ist (siehe Abschnitt 4.2). Zusammenfassend kann also derzeit festgehalten werden, dass für die hier untersuchten KPL-Tests und hinsichtlich des Einflusses von *ICT-Literacy* keine Gefährdung der Testvalidität beziehungsweise keine Verletzung des Testfairnesskriteriums (Kubinger, 2009; Messick, 1995) anzunehmen ist.

Literatur

- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22. doi:10.1037/h0046743
- Clariana, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5), 593–602.
- Dunn, T. J., Baguley, T. & Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 1–14. doi:10.1111/bjop.12046
- Greiff, S., Kretzschmar, A., Müller, J. C., Spinath, B. & Martin, R. (2014). The computer-based assessment of complex problem solving and how it is influenced by students' information and communication technology literacy. *Journal of Educational Psychology*, 106(3), 666–680. doi:10.1037/a0035426
- Kretzschmar, A., Neubert, J. C., Wüstenberg, S. & Greiff, S. (2015). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence*. doi:10.1016/j.intell.2015.11.004
- Kubinger, K. D. (2009). *Psychologische Diagnostik Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kyllonen, P. (2009). New constructs, methods, and directions for computer-based assessment. In F. Scheuermann & J. Björnsson (Hrsg.), *The transition to computer-based assessment* (S. 151–156). Luxembourg: Office for Official Publications of the European Communities.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58, 103–117. doi:10.1026/0033-3042.58.2.103
- Mead, A. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449–458.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.
- Prensky, M. (2001). Digital natives, digital immigrants. Part 1. *On the Horizon*, 9, 1–6. doi:10.1108/10748120110424816
- Revelle, W. & Zinbarg, R. E. (2009). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijsma. *Psychometrika*, 74(1), 145–154. doi:10.1007/s11336-008-9102-z

- Richter, T., Naumann, J. & Horz, H. (2010). Eine revidierte Fassung des Inventars zur Computerbildung (INCOBI-R). *Zeitschrift für Pädagogische Psychologie*, 24, 23–37. doi:10.1024/1010-0652/a000002
- Russell, M., Goldberg, A. & O’connor, K. (2003). Computer-based testing and validity: A look back into the future. *Assessment in Education: Principles, Policy & Practice*, 10, 279–293. doi:10.1080/0969594032000148145
- Tsai, M.-J. (2002). Do male and female students often perform better than female students when learning computers? A study of Taiwanese eight graders’ computer education through strategic and cooperative learning. *Journal of Educational Computing Research*, 26, 67–85. doi:10.2190/9JW6-VV1P-FAX8-CGE0
- Wagener, D. (2007). *START-C. Testbatterie für Berufseinsteiger - Computer*. Göttingen: Hogrefe.
- Wittmann, W. W. & Matt, G. E. (1986). Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts-und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des Berliner Intelligenzstrukturmodells. *Diagnostica*, 32, 309–329.

Teil III

Konstruktvalidität

Kapitel 5

KPL als eigenständiges Fähigkeitskonstrukt?

Seit Beginn der KPL-Forschung wird darüber diskutiert, ob KPL als ein von Intelligenz unabhängiges Konstrukt, (z. B. Dörner, Kreuzig, Reither & Stäudel, 1983), als ein Subkonstrukt (z. B. Wüstenberg, Greiff & Funke, 2012) oder aber als ein redundantes Konstrukt (z. B. Süß, 1996) im nomologischen Netzwerk bestehender Intelligenzkonstrukte anzusehen ist. In Anlehnung an die Kriterien für die Eigenständigkeit von Intelligenzkonstrukten wurden von Süß (1999) Kriterien für die Überprüfung eines KPL-Fähigkeitskonstrukts postuliert (siehe auch Cronbach & Meehl, 1955; Funke, 2003; Jäger, Süß & Beauducel, 1997; Rost, 2009; Süß, 1996, 2006; Weber & Westmeyer, 2001; Wilhelm, 2009; Ziegler, Booth & Bensch, 2013). Demnach sollte das Konstrukt ...

1. ... sich durch ein hohes Maß an Generalität auszeichnen.
2. ... auf Leistungsdaten basieren.
3. ... mittelfristig zeitlich stabil sein.
4. ... teilweise unabhängig von etablierten Fähigkeitskonstrukten sein, die aufgrund ihrer Ähnlichkeit ebenfalls zur Beschreibung der Effekte herangezogen werden können (diskriminante Validität).
5. ... durch unterschiedliche Aufgabentypen operationalisiert werden (konvergente Validität), damit der Einfluss von aufgabenspezifischen, nicht beabsichtigten Varianzquellen reduziert werden kann.

6. ... inkrementelle Validität in der Erklärung von externen Kriterien gegenüber etablierten Fähigkeitskonstrukten aufweisen.

Im Nachfolgenden wird auf die einzelnen Punkte spezifischer eingegangen, um einerseits den aktuellen Forschungsstand darzustellen, aber andererseits auch offene Punkte hinsichtlich der Konstruktvalidität von KPL aufzuzeigen. Besonderes Augenmerk wird dabei auf die letzten drei Kriterien gelegt, da jene üblicherweise in der Forschungsliteratur zur Verortung von KPL im nomologischen Netzwerk der Intelligenz herangezogen werden (siehe z. B. Greiff & Fischer, 2013; Greiff, Fischer, Wüstenberg et al., 2013; Greiff & Neubert, 2014; Greiff, Wüstenberg, Molnar et al., 2013; Neubert, Kretzschmar, Wüstenberg & Greiff, 2015; Sonnleitner, Keller, Martin & Brunner, 2013; Wüstenberg et al., 2012)

5.1 Kriterien für ein eigenständiges Fähigkeitskonstrukt

5.1.1 Generalität von KPL

Die Generalität eines Konstruktes bezieht sich auf die Breite der Anforderungen, bei denen die entsprechende Fähigkeit relevant ist. Sehr spezifische Anforderungen (z. B. das Lösen von Zahlenreihenaufgaben) haben nur ein geringes, während breite Anforderungen (z. B. schlussfolgerndes Denken bei sehr unterschiedlichen Aufgaben) ein hohes Generalitätsniveau besitzen. Laut dem Generalitätskriterium haben nur jene Fähigkeiten das Potenzial, einem eigenständigen Intelligenzkonstrukt zugeschrieben zu werden, welche eine hohe Generalität aufweisen (d. h. Zahlenreihenvervollständigen wäre nicht als eigenständiges Konstrukt anzusehen, schlussfolgerndes Denken hingegen schon). Aus dem Generalitätskriterium ergeben sich zwei wichtige Implikationen: (a) Das Konstrukt sollte durch eine Vielzahl von unterschiedlichen Aufgaben abgebildet werden und (b) die dem potenziellen Konstrukt zugeschriebenen Leistungen sollten für verschiedenartige externe Kriterien relevant sein. Der erste Punkt wird im Abschnitt zur konvergenten Validität näher beleuchtet, während nachfolgend die Bedeutsamkeit von KPL für externe Kriterien dargestellt wird.

Für KPL wird grundsätzlich angenommen, dass es eine bereichsunspezifische Fähigkeit und somit für eine Vielzahl an Lebenssituationen relevant ist (vgl. Funke, 2003, 2010; Greiff, 2012; Neubert, Mainert, Kretzschmar & Greiff, 2015; OECD, 2014). Da die postulierten Prozesse des Problemlösens (z. B. Exploration der Problemsituation zur Wissensaneignung, Wissensanwendung zur Problemlösung; Fischer, Greiff & Funke, 2012) unabhängig von spezifischen Problemstellungen als relevant angesehen werden können, scheint dieses Kriterium augenscheinlich erfüllt zu sein. Empirische Evidenz wird darin gesehen, dass die KPL-Leistung mit verschiedenen Außenkriterien korreliert, wie zum Beispiel Noten als Indikator von Schulleistungen (z. B. Greiff & Fischer, 2013; Greiff, Wüstenberg, Molnar et al., 2013) oder Studienleistungen (Stadler, Becker, Greiff & Spinath, 2015) sowie Vorgesetztenurteile oder Einkommen als Indikatoren für Berufserfolg (Danner, Hagemann, Schankin, Hager & Funke, 2011; Ederer, Nedelkoska, Patt & Castellazzi, 2015). Aktuelle Arbeiten, welche im Rahmen einer Kooperation innerhalb dieses Dissertationsprojekts entstanden sind, konnten aufzeigen, dass KPL ebenfalls mit Fortbildungstagen als Indikator für lebenslanges Lernen (Mainert, Kretzschmar, Neubert & Greiff, 2015) sowie mit kreativem, unternehmerischem Denken (Mainert et al., eingereicht) substantiell korreliert.

Andererseits gibt es jedoch auch Befunde, auf deren Grundlage die Generalität von KPL hinterfragt werden kann. Im pädagogisch-psychologischen Kontext zeigt sich beispielsweise regelmäßig, dass Schulnoten vor allem der naturwissenschaftlichen und mathematischen Fächer im Zusammenhang mit KPL stehen, jedoch Noten der sprachlichen oder geisteswissenschaftlichen Fächer nur in geringerem oder gar keinem Maße (siehe z. B. Greiff & Fischer, 2013; Greiff, Fischer, Stadler & Wüstenberg, 2014; Greiff, Fischer, Wüstenberg et al., 2013; Greiff & Neubert, 2014; Schweizer, Wüstenberg & Greiff, 2013). Als mögliche Ursache für diese differentiellen Effekte werden oft die unterschiedlichen kognitiven Anforderungen in den sprachlichen oder geisteswissenschaftlichen Fächern genannt, welche sich von Anforderungen des wissenschaftlichen und systematischen Ansatzes des KPL-Konzepts unterscheiden (z. B. Greiff, Fischer, Wüstenberg et al., 2013; Schweizer et al., 2013). Dies widerspräche allerdings dem bereichsunspezifischen Charakter eines potenziellen KPL-Konstrukts, insbesondere wenn man berücksichtigt, dass auch die Leistungen in den sprachlichen oder geisteswissenschaftlichen Fächern durchaus durch etablierte Intelligenzkonstrukte erklärt werden können (wenn auch in geringeren Maße als bei naturwissenschaftlichen Fächern, siehe z. B. Wittmann & Matt, 1986; Wittmann & Süß, 1996).

Insgesamt liegen also durchaus erste Hinweise für die Generalität eines potenziellen KPL-Konstrukts vor. Akademische Leistungen (d. h. vor allem Schulnoten) sind dabei das bisher am meisten genutzte externe Kriterium in der KPL-Forschung (siehe Tabelle 5.1). Dementsprechend wäre es angebracht, verschiedenartige Kriterien in zukünftigen Studien zu berücksichtigen, um die Annahme einer generellen, bereichsunspezifischen Fähigkeit zu untermauern. Wichtig ist dabei, dass aufgrund der Überlappung von KPL mit etablierten Intelligenzkonstrukten (siehe Abschnitt 5.1.4 zur diskriminanten Validität) überprüft wird, ob der Zusammenhang von KPL mit externen Kriterien auch nach Kontrolle relevanter Prädiktoren bestehen bleibt (siehe Abschnitt 5.1.6 zur inkrementellen Validität).

5.1.2 Leistungsdaten

Die Erforschung kognitiver Fähigkeitskonstrukte sollte grundsätzlich auf objektiven und reliablen Leistungsdaten (T-Daten im Sinne von Cattell, 1965) basieren, wohingegen Selbst- oder Fremdeinschätzungen erst nach der Etablierung eines Fähigkeitskonstrukts genutzt werden sollten (Süß, 2006; Weber & Westmeyer, 2001).

Wie in Kapitel 2 dargestellt, beruhen die neueren KPL-Tests auf formalen Rahmenkonzepten (z. B. lineare Strukturgleichungen oder finite Automaten), so dass Ziele und optimale Lösungswege eindeutig definiert werden können und auch das Explorationsverhalten objektiv bewertet werden kann (Greiff, Wüstenberg & Funke, 2012; Sonnleitner, Keller, Martin, Latour & Brunner, in Druck). In Verbindung mit dem MKS-Ansatz können zusätzlich hohe Reliabilitäten erreicht werden, so dass insgesamt festzuhalten ist, dass das Kriterium der leistungsdatenbasierten Diagnostik bei den aktuellen KPL-Tests des MKS-Ansatzes erfüllt ist.

5.1.3 Zeitliche Stabilität

Die zeitliche Stabilität wird als eine wichtige Voraussetzung für Fähigkeitskonstrukte angesehen (Wilhelm & Kunina-Habenicht, 2015), wobei langfristige, entwicklungspsychologische Veränderungen nicht ausgeschlossen werden (vgl. Schaie, 1994). So konnten für Intelligenzkonstrukte wie im BIS-Modell spezifiziert Retestkorrelationen von $r = .70$ bis $r = .88$ über einen Zeitraum von 4 Jahren nachgewiesen werden, was als starke Evidenz für diese Intelligenzkonstrukte gewertet werden kann (Schmidt, 1983).

In der KPL-Forschung existieren so gut wie keine längsschnittlichen Untersuchungen; Forschungsarbeiten zur Retestreliabilität neuerer KPL-Tests liegen bisher nicht vor. Die Arbeit von Süß (1996) auf der Basis der traditionellen Mikrowelt Schneiderwerkstatt (mit den entsprechenden Nachteilen hinsichtlich der psychometrischen Qualität, siehe Kapitel 2) stellt dabei eine Ausnahme dar. In dieser Studie konnte eine Stabilität der KPL-Leistung in der Höhe von $r = .46$ über eine Zeitspanne von etwa einem Jahr nachgewiesen werden, wobei die gesamte zeitlich stabile KPL-Varianz durch etablierte Intelligenzkonstrukte erklärt werden konnte. Entsprechend wurde geschlussfolgert, dass die Annahme eines eigenständigen KPL-Konstrukts nicht gerechtfertigt erscheine.

Eine weitere Ausnahme bildet die Arbeit von Frischkorn, Greiff und Wüstenberg (2014), bei der KPL durch MicroDYN operationalisiert zu drei verschiedenen Messzeitpunkten (Abstand jeweils ca. 1 Jahr) erhoben wurde. Leider wurden in der Arbeit keine direkten Korrelationen zwischen den KPL-Leistungen der verschiedenen Messzeitpunkte berichtet, so dass der Studie auch nicht der von Reasoning unabhängige, zeitlich stabile Varianzanteil von KPL entnommen werden kann.

Somit bleibt für das Kriterium der zeitlichen Stabilität festzuhalten, dass es bisher noch keine empirische Evidenz gibt, die die Annahme eines eigenständigen KPL-Konstrukt stützt. Da durch die Einführung der neuen KPL-Tests psychometrisch überzeugende Erhebungsinstrumente zur Verfügung stehen, sollten sich zukünftige Forschungsarbeiten verstärkt darum bemühen, diese Lücke zu schließen. Zu berücksichtigen ist dabei, dass für ein eigenständiges KPL-Konstrukts der zeitlich stabile Varianzanteil der KPL-Leistungen nicht ausschließlich auf etablierte Intelligenzkonstrukte zurückzuführen sein sollte (vgl. Süß, 1996).

5.1.4 Diskriminante Validität: Das Verhältnis von Intelligenz und KPL

Leistungen eines potenziellen Konstrukts sollten sich nicht durch schon etablierte Konstrukte erklären lassen, sondern es sollte eine partielle Eigenständigkeit im nomologischen Netzwerk verwandter, etablierter Konstrukte nachweisbar sein (Wilhelm, 2009). Ansonsten entstünden eine Vielzahl unterschiedlicher Bezeichnungen für gleiche Phänomene, die die wissenschaftliche Auseinandersetzung und Theoriebildung unnötig erschweren (vgl. *Occam's Razor*).

Der Zusammenhang zwischen KPL und etablierten Intelligenzkonstrukten im Sinne der Überprüfung eines eigenständigen KPL-Fähigkeitskonstrukts ist wahrscheinlich die am meisten untersuchte Forschungsfrage in der KPL-Forschung (z. B. Greiff, Fischer, Wüstenberg et al., 2013; Kröner, Plass & Leutner, 2005; Sonnleitner et al., 2013; Süß, 1996; Wüstenberg et al., 2012). Dies ist insofern nicht verwunderlich, da KPL ursprünglich als Abgrenzung zu bestehenden Intelligenzkonstrukten in die psychologische Forschung eingeführt wurde (Dörner et al., 1983), wobei jüngere Studien durchweg einen substantiellen Zusammenhang zwischen KPL und etablierten Intelligenzkonstrukten aufzeigen konnten (z. B. Kröner et al., 2005; Süß, 1996; Wüstenberg et al., 2012).

Tabelle 5.1 gibt einen Überblick über die bisherigen Studien, in denen KPL mit den KPL-Tests des MKS-Ansatzes operationalisiert und gleichzeitig der Zusammenhang zu Intelligenzkonstrukten untersucht wurde. Über die verschiedenen Studien hinweg finden sich Zusammenhänge zwischen $r = .27$ (Fischer et al., 2015) bis $r = .69$ (Wüstenberg, Stadler, Hautamäki & Greiff, 2014), wobei der Median bei $r = .52$ liegt. Korrelationen in dieser Größenordnung können je nach Sichtweise als ausreichend (z. B. Wüstenberg et al., 2012) beziehungsweise unzureichend (z. B. Kröner et al., 2005) hinsichtlich der Interpretation im Sinne der diskriminanten Validität eingestuft werden. Eine detaillierte Darstellung der unterschiedlichen Perspektiven wird in Kapitel 7 gegeben, wobei festzuhalten ist, dass in den neueren Forschungsarbeiten (siehe Tabelle 5.1) die Korrelationen zwischen KPL und etablierten Intelligenzkonstrukten durchweg im Sinne der diskriminanten Validität interpretiert werden.

TABELLE 5.1: Übersicht der Studien mit KPL-Tests des MKS-Ansatzes in Verbindung mit Intelligenzmessungen sowie ggf. externe Kriterien.

Studie	Stichprobe	KPL-Test	Intelligenz-Test	Konvergente Validität	Diskriminante Validität	Externe Kriterien	Inkrementelle Validität
1	217 Auszubildende	MicroDYN	CFT / Wissen ^a	-	$r = .48$ / n.b.	DPS	n.s. bis 7% ^b
2	399 ArbeiterInnen	MicroDYN, MicroFIN	SPM	-	$r = .62$	Berufsleistung ^c	n.s. bis 1% ^d
3	577 SchülerInnen	MicroDYN	KFT	-	$r = .52$	Schulnoten	3% ^e
3	339 Studierende	MicroDYN	IST-2000-R ^f	-	$r = .27$	Schulnoten	n.s. ^e
4	140 Studierende	MicroDYN	SPM	-	$r = .50$	Schulnoten	6% ^g
5 / 6	339 Studierende	MicroDYN, MicroFIN, GeneticsLab, Tailorshop	IST-2000-R ^f	$r = .25 - .73$	$r = .53$ ^h	Schulnoten	n.s. bis 5% ⁱ
7	486 SchülerInnen	MicroDYN	CFT	-	$r = .47$ ^k	Schulnoten	n.s. bis 4% ^l
8	490 SchülerInnen	MicroDYN	CFT	-	$r = .62$ ^k	Schulnoten	n.s. bis 10%
9	254 ArbeiterInnen	MicroDYN	CFT	-	$r = .68$	Berufsleistung ^m	n.s. ⁿ
10	788 SchülerInnen	MicroDYN	IR ^o	-	$r = .44$	DPS	7% ^p
11	576 SchülerInnen	MicroDYN MicroFIN	KFT	$r = .73$	$r = .63$	-	-
12	2000 SchülerInnen	MicroDYN	DR / SR	-	$r = .44/.15$	Schulnoten	n.b.
13	393 SchülerInnen	MicroDYN	MUN	-	$r = .46$ ^k	Schulnoten	n.s. bis 11%
14	61 SchülerInnen	GeneticsLab	IST-2000-R ^q	-	$r = .40$	Schulnoten	n.b.
15	299 SchülerInnen	GeneticsLab	IST-2000-R ^r	-	$r = .59$	-	-
16	563 SchülerInnen	GeneticsLab	IST-2000-R ^s	-	$r = .62$	Schulnoten Schulkompetenzen PISA-Kompetenzen	n.s. ^t 14% ^u n.s.
17	222 Studierende	MicroDYN	APM	-	$r = .62$ ^k	Schulnoten	6%
18	3191 SchülerInnen	MicroDYN	(DR/QR) ^w / SR	-	$r = .69/.56$	-	-

Abkürzungen: APM=Advanced progressive matrices (Raven, 1958). CFT=CFT-20-R (Weiß, 2006). DPS=Domänenspezifisches Problemlösen. DR=Deductive reasoning (J. Ross & C. Ross, 1976). IST=IST-2000-R (Amthauer, Brocke, Liepmann & Beauducel, 2001). IR=Inductive Reasoning (Csapo, 1997). KFT=KFT 4-12+R (Heller & Perleth, 2000). MUN=Memory Updating Numeric (K. Oberauer, Süß, Schulze, Wilhelm & Wittmann, 2000). QR=Quantitative reasoning (Sternberg, Castejón, Prieto, Hautamäki & Grigorenko, 2001). SR=Scientific reasoning (Hautamäki et al., 2002). SPM=Standard Progressive Matrices (Horn, 2009). n.b.=nicht berichtet. n.s.=nicht signifikant.

Erläuterungen: ^a=Technisches Fachwissen. ^b=Stichprobenspezifisch. Bei Berücksichtigung von technischem Fachwissen: generell n.s. ^c=Monatliches Nettoeinkommen. ^d=Abhängig von den zusätzlich berücksichtigten Prädiktoren. ^e=Nur durch KPL aufgeklärte Varianz (Kommunalitätenanalyse). ^f=Subtest: Matrizenest. ^g=Für Fachnoten: n.s. bis 8% . ^h=Globaler KPL-Faktor. Einzelne KPL-Tests: $r = .24 - .56$. ⁱ=Globaler KPL-Faktor, fachspezifisch. Einzelne KPL-Tests: n.s. bis 4%, fachspezifisch. ^k=Aus multipler Regression abgeleitet. ^l=Klassenstufenspezifisch. ^m=Berufsgruppen und Fortbildungstage. ⁿ=Die berichtete inkrementelle Varianzaufklärung konnte durch einen F-Test nicht abgesichert werden. ^o=Subtest: Zahlenanalogien, Zahlenserien, Wortanalogien. ^p=Basierend auf Partialkorrelation für die Gesamtstichprobe (Klassenspezifisch: n.s. bis 13%). ^q=Subtests: Figurenauswahl, Zahlenserien. ^r=Subtests: Matrizenest, Zahlenserien. ^s=Subtests: Matrizenest, Zahlenserien, Würfelaufgaben. ^t=Für Fachnoten: n.s. bis 1%. ^u=Fachspezifisch: 5 bis 14%. ^w=Globaler Faktor. Bei mehrdeutigen Angaben wird sich bezogen auf: latente Werte (vor manifesten Werten), Generalfaktoren (vor spezifischen Subfaktoren, z. B. KPL-Gesamtwert vor KPL-Facetten), KPL-Facette mit höchstem Zusammenhang, Inkrement für globaleres Kriterium (vor spezifischeren Kriterium, z. B. Abitur- vor Fachnote) sowie Kriterium mit der größten inkrementellen Varianzaufklärung (bei gleichartigen Kriterien).

Studien: 1=Abele et al. (2012). 2=Ederer, Nedelkoska, Patt und Castellazzi (2015) 3=Fischer et al. (2015). 4=Greiff und Fischer (2013). 5=Greiff, Fischer, Wüstenberg et al. (2013). 6=Greiff, Stadler, Sonnleitner, Wolff und Martin (2015). 7=Greiff, Wüstenberg, Molnar et al. (2013). 8=Greiff und Neubert (2014). 9=Mainert, Kretzschmar, Neubert und Greiff (2015). 10=Molnár, Greiff und Csapo (2013). 11=Neubert, Kretzschmar, Wüstenberg und Greiff (2015). 12=Scherer, Greiff und Hautamäki (2015). 13=Schweizer, Wüstenberg und Greiff (2013). 14=Sonnleitner et al. (2012). 15=Sonnleitner, Brunner, Keller und Martin (2014). 16=Sonnleitner, Keller, Martin und Brunner (2013). 17=Wüstenberg, Greiff und Funke (2012). 18=Wüstenberg, Stadler, Hautamäki und Greiff (2014).

Hinsichtlich der aktuellen KPL-Forschungsarbeiten sind dabei mehrere Punkte beachtenswert. Wird beispielsweise das Berliner Intelligenzstrukturmodell (BIS; Jäger, 1984, siehe Abbildung 1 in Kapitel 7) als Rahmenkonzept für die Klassifikation der Intelligenztestleistung herangezogen, wird deutlich, dass die Studien basierend auf den KPL-Tests des MKS-Ansatzes eine sehr spezifische Intelligenzoperationalisierung genutzt haben. Im BIS Modell lassen sich Intelligenztestleistungen hinsichtlich der zwei Modalitäten Operationen und Inhalte unterscheiden. Hinsichtlich der Operationen sind in der aktuellen Version des BIS Modells Verarbeitungskapazität (*reasoning*), Bearbeitungsgeschwindigkeit (*mental speed*), Merkfähigkeit (*short-term memory*) und Einfallsreichtum (*fluency, creativity*) vertreten. Die Inhaltsmodalität umfasst die Kategorien figural, numerisch und verbal. Beide Dimensionen gekreuzt ergeben damit eine feingliedrige Klassifikationsmöglichkeit sowie eine umfassende Konzeptualisierung der Intelligenzleistungen.

Die Einordnung der in den Studien aus Tabelle 5.1 verwendeten Intelligenztests in das BIS-Modell veranschaulicht, dass in den genannten Studien ausschließlich Reasoning-Tests verwendet wurden (siehe Abbildung 3 in Kapitel 7).¹ Dies ist insofern nachvollziehbar, da in früherer Forschung wiederholt aufgezeigt wurde, dass insbesondere Reasoning der wichtigste Einflussfaktor für KPL-Leistungen ist (z. B. Süß, 1996; Wittmann & Hattrup, 2004). Zu beachten ist jedoch, dass auch andere kognitive Leistungen beim Lösen komplexer Probleme eine wichtige Rolle spielen können. Am Beispiel des Bahnfahrkartenautomaten, auf welchen oft stellvertretend für alltägliche komplexe Problemsituationen Bezug genommen wird (z. B. Fischer et al., 2012; Greiff et al., 2012; OECD, 2014), kann dies verdeutlicht werden. Oftmals steht nur ein beschränktes Zeitfenster für Nutzereingaben zur Verfügung (vgl. Bearbeitungsgeschwindigkeit), je nach Komplexität des Tarifsystems sind eine Fülle an Informationen zu beachten (vgl. Merkfähigkeit) und die Benutzerführung erfordert teilweise kreative Herangehensweisen, um Produkte in der verzweigten Menüführung aufzufinden (vgl. Einfallsreichtum). Eine Fokussierung auf Reasoning kann somit als bedeutsame Limitation der neueren KPL-Studien angesehen werden (insbesondere auch, wenn Zusammenhänge zu externen Kriterien untersucht werden, siehe Abschnitt 5.1.6 zur inkrementellen Validität).

¹Die Studie von Schweizer et al. (2013) stellt eine Ausnahme dar, da in jener Arbeitsgedächtniskapazität erfasst wurde. Arbeitsgedächtniskapazität, wie mit der Aufgabe MUN erfasst, ist nahe dem Reasoning-Faktor verortet (Klaus Oberauer, Schulze, Wilhelm & Süß, 2005, 2008; Rost, 2009; Süß, Oberauer, Wittmann, Wilhelm & Schulze, 2002), so dass jene für die Darstellung vereinfacht Reasoning zugeordnet wurde.

Einschränkend kommt hinzu, dass auch Reasoning meist nur sehr spezifisch operationalisiert wurde. Eine konstruktrepräsentative Operationalisierung (Gustafsson & Åberg-Bengtsson, 2010; Messick, 1995; Shadish, Cook & Campbell, 2001) wäre erst dann gegeben, wenn die Inhalte figural, numerisch und verbal in der entsprechenden Operationalisierung abgebildet sind (Jensen & Wang, 1994; Reeve & Blacksmith, 2009; Rost, 2009). Tabelle 5.2 zeigt jedoch, dass sich die neueren KPL-Studien zu einem sehr großen Teil auf einseitige Operationalisierungen von Reasoning (d. h. meist figural) gestützt haben. Ähnlich sieht es hinsichtlich der KPL-Operationalisierungen aus. Auch hier wurden in den Studien bis auf wenige Ausnahmen einzelne KPL-Tests genutzt (siehe Tabelle 5.1), um den Zusammenhang auf Konstruktebene zu untersuchen. Als Folge könnten in den Analysen zur diskriminanten Validität daher aufgaben- und inhaltspezifische Varianzanteile betont worden sein, die als unerwünschte Varianz den Zusammenhang zwischen etablierten Intelligenzkonstrukten und KPL beeinflusst haben (vgl. Wittmann & Matt, 1986; Wittmann & Schmidt, 1983).

TABELLE 5.2: Klassifizierung der Intelligenzmessungen der bisherigen Studien entsprechend der Inhaltsmodalitäten des BIS-Modells

Modalität	Studien																
figural	1	2	3	4	5/6	7	8	9	-	11	-	-	14	15	16	17	-
numerisch	-	-	-	-	-	-	-	-	10	-	-	13	14	15	16	-	18
verbal	-	-	-	-	-	-	-	-	10	-	12	-	-	-	-	-	18

Anmerkungen: Studienbezeichnungen entsprechend Tabelle 5.1.

Durch diese Unterrepräsentation des Intelligenzkonstrukts (d. h. ausschließlich Reasoning mit überwiegend figuralen Aufgaben) und den Einsatz von einzelnen, spezifischen KPL-Tests in den neueren KPL-Studien sind die Befunde bezüglich eines eigenständigen KPL-Konstrukts also nur eingeschränkt interpretierbar. Für das Kriterium diskriminante Validität wären in zukünftigen Studien entsprechend umfassendere Operationalisierungen von Intelligenz als auch KPL wünschenswert.

5.1.5 Konvergente Validität: Der Zusammenhang verschiedener KPL-Maße

Die konvergente Validität in der KPL-Forschung wurde lange Zeit vernachlässigt (Süß, 1999) und spiegelt sich auch nur langsam in der neueren Forschungsliteratur wider (siehe Tabelle 5.1). Studien auf Basis traditioneller Mikrowelten konnten

aufzeigen, dass unterschiedliche KPL-Tests nur wenig bis gar keine systematische Varianz teilten, insbesondere wenn für etablierte Intelligenzkonstrukte kontrolliert wurde (z. B. Süß, 1999). Dies widerspricht offensichtlich der Annahme, dass hinter der KPL-Leistung ein stabiles Fähigkeitskonstrukt steht (Funke, 2003).

Eine neuere Studie (Greiff et al., 2014; Greiff, Fischer, Wüstenberg et al., 2013), welche auch KPL-Tests des MKS-Ansatzes berücksichtigte, konnte jedoch aufzeigen, dass auch nach Kontrolle von Intelligenzkonstrukten signifikant und substantiell Varianz zwischen den KPL-Operationalisierungen des MKS-Ansatzes geteilt wurde ($r = .49 - .64$). Dies traf jedoch nur noch eingeschränkt für die klassische Mikrowelt Schneiderwerkstatt zu ($r = .15 - .36$), was als weiterer Beleg für die psychometrische Überlegenheit der neueren KPL-Tests interpretiert wurde (vgl. Greiff, Stadler, Sonnleitner, Wolff & Martin, 2015). Beachtet werden sollte bei dieser Studie, dass bei der Analyse der konvergenten Validität mit MicroDYN und GeneticsLab zwei Operationalisierungen verwendet wurden, die sich in einem hohen Maße gleichen (z. B. basieren beide auf dem DYNAMIS-Ansatz, erfassen die gleichen Anforderungen hinsichtlich der Explorationsstrategie, folgen einer sehr ähnlichen Testpräsentation usw.).² Wünschenswert für die Analyse der konvergenten Validität von KPL wäre demnach die Berücksichtigung von psychometrisch überzeugenden, aber dennoch unterschiedlichen Operationalisierungen.

Mit der Entwicklung von MicroFIN (siehe Kapitel 3) konnte dieser Forderung nachgekommen werden. Entsprechend wurden in einer weiteren Studie zur konvergenten Validität (Neubert, Kretzschmar et al., 2015) mit MicroDYN und MicroFIN zwei heterogene KPL-Tests innerhalb des MKS-Ansatzes eingesetzt, welche hohe konvergente Zusammenhänge (je nach Facette $r = .56 - .73$) zeigten, die auch nach Kontrolle von (figuralem) Reasoning signifikant und substantiell (je nach Facette $r = .33 - .52$) blieben.

Insgesamt konnten also auf Basis der KPL-Tests des MKS-Ansatzes die früheren Befunde (z. B. Süß, 1999) hinsichtlich der geringen (beziehungsweise nach Berücksichtigung von Intelligenz fehlenden) konvergenten Validität nicht repliziert werden. Einschränkend für die KPL-Studien zur konvergenten Validität (Greiff et al., 2014; Greiff, Fischer, Wüstenberg et al., 2013; Neubert, Kretzschmar et al.,

²Die beiden zusätzlich eingesetzten MicroFIN-Aufgaben befanden sich in einer frühen Phase der Entwicklung, mit einer reduzierten Anzahl an Items je Aufgabe, vergleichsweise niedrigen Reliabilitäten und einem empirisch nicht bestätigten, theoretisch erwarteten Messmodell (siehe Greiff, Fischer, Wüstenberg et al., 2013). Die Ergebnisse hinsichtlich MicroFIN sollten daher mit Zurückhaltung interpretiert werden und finden entsprechend hier keine Berücksichtigung.

2015) ist jedoch zu erwähnen, dass die Kritikpunkte hinsichtlich zu enger Intelligenzoperationalisierungen auch hier in vollem Umfang zutreffen (siehe Abschnitt 5.1.4). Damit liegen zwar wichtige Hinweise bezüglich des Kriteriums konvergente Validität vor, welche aber mit einer konstruktrepräsentativen Intelligenzoperationalisierung repliziert werden sollten. Erst dann kann angenommen werden, dass die gemeinsame Varianz verschiedener KPL-Tests nicht auf etablierte Intelligenzkonstrukte zurückzuführen ist.

5.1.6 Inkrementelle Validität: Nützlichkeit eines zusätzlichen Konstrukts

Die Nützlichkeit eines neuen Konstrukts zeigt sich insbesondere dann, wenn externe Kriterien wie beispielsweise Ausbildungs- und Berufserfolg besser vorhergesagt werden können als mit etablierten Fähigkeitskonstrukten (Greiff, 2012; Kersting, 2001; Süß, 1999; Wilhelm, 2009). Da wiederholt ein großer Überlappungsbereich zwischen etablierten Intelligenzkonstrukten und KPL aufgezeigt werden konnte (z. B. Greiff, Fischer, Wüstenberg et al., 2013; Sonnleitner et al., 2013; Süß, 1996; Wüstenberg et al., 2012), ist damit die inkrementelle Validität von KPL über bestehende Intelligenzkonstrukte hinaus interessant für die Bewertung eines KPL-Konstrukts. Sollte sich der Mehrwert eines KPL-Konstrukts (im Sinne von zusätzlicher Varianzaufklärung) auch nach Kontrolle etablierter, kognitiver Fähigkeitskonstrukte zeigen, kann dies als starke Evidenz für ein eigenständiges KPL-Konstrukt interpretiert werden. Wenn sich jedoch kein Mehrwert aufzeigen lassen sollte, dann wäre im Sinne des Sparsamkeitsprinzips (vgl. *Occam's Razor*) zu hinterfragen, ob ein eigenständiges KPL-Konstrukt gerechtfertigt erscheint. Sinnvoller wäre in diesem Fall eher eine Zuordnung zu etablierten Fähigkeitskonstrukten im Sinne einer anders- oder neuartigen Operationalisierung.

Die neueren Forschungsarbeiten konnten in einer Vielzahl von Studien nachweisen, dass KPL über Reasoning hinaus Varianz an externen Kriterien aufklären konnte (siehe Tabelle 5.1). Dabei wurden überwiegend Schulnoten als externe Kriterien genutzt (zur Bedeutung von KPL im Schulkontext, siehe Kapitel 6). Insgesamt konnten dabei bis zu 11% an zusätzlicher Varianz durch KPL über etablierte Intelligenzkonstrukte (d. h. Reasoning) erklärt werden. Aber auch für domänenspezifisches Problemlösen, Schulkompetenzen (erhoben mit objektiven Leistungstests)

und Berufsleistungskriterien konnte die inkrementelle Validität von KPL nachgewiesen werden.

Allerdings zeigt eine nähere Betrachtung der Studien in Tabelle 5.1 auch, dass sich das Befundmuster nicht einheitlich darstellt. Beispielsweise zeigte sich bei einigen Studien die inkrementelle Validität nicht oder nur eingeschränkt, wenn fachspezifische Schulnoten (z. B. sprachliche oder geisteswissenschaftliche Fächer; Greiff & Fischer, 2013; Greiff et al., 2015; Schweizer et al., 2013), Subgruppen (z. B. niedrigere Klassenstufen; Greiff, Wüstenberg, Molnar et al., 2013) oder weitere für das Kriterium relevante Prädiktoren (z. B. fachspezifisches Wissen; Abele et al., 2012) berücksichtigt wurden. In zukünftigen Studien sollten diese Faktoren entsprechend stärker beachtet werden, um deren Einfluss auf die inkrementelle Validität von KPL detaillierter erforschen zu können.

Insgesamt spricht die Befundlage hinsichtlich des Kriteriums inkrementeller Validität zumindest teilweise für ein eigenständiges KPL-Konstrukt. Nachteilig ist jedoch wieder anzumerken, dass die eingeschränkten Operationalisierungen von Intelligenz und KPL die Aussagekraft zur inkrementellen Validität mindern.

5.2 Zusammenfassende Bewertung und offene Forschungsfragen

Hinsichtlich der vorher ausgeführten Kriterien zur Bewertung eines eigenständigen kognitiven Fähigkeitskonstrukts sind die Befunde uneinheitlich. Während Kriterien wie beispielsweise eine auf Leistungsdaten basierte Messung zu vollem Umfang erfüllt werden, sind die anderen Kriterien durch die neueren KPL-Studien nur teilweise oder nicht abgedeckt. Weiterer Forschungsbedarf ist also bei den Kriterien zur zeitlichen Stabilität, für die bisher keine Befunde berichtet wurden, sowie zur diskriminanten, konvergenten und inkrementellen Validität angebracht.

Die diskriminante Validität von KPL zu etablierten Intelligenzkonstrukten wurde bisher vor allem auf der Basis spezifischer Operationalisierungen von Intelligenz (d. h. figurales, vereinzelt auch numerisches oder verbales Reasoning) und KPL untersucht. Diese Konstruktunterrepräsentation kann dabei zu verzerrten Zusammenhängen zwischen KPL und Intelligenz geführt haben, indem konstruktirrelevante, aufgabenspezifische Varianzanteile in den Analysen betont wurden (vgl.

Wittmann & Matt, 1986; Wittmann & Schmidt, 1983). Daher ist es angeraten, die bisherigen Befunde mit weiteren, potentiell relevanten Intelligenzkonstrukten und entsprechenden konstruktrepräsentativen Operationalisierungen zu replizieren.

Zur konvergenten Validität liegen bisher nur wenige Ergebnisse vor, wobei einerseits die Homogenität der KPL-Tests und andererseits die verwendeten, eingeschränkten Intelligenzoperationalisierungen die Bewertung hinsichtlich der Eigenständigkeit eines potenziellen KPL-Konstrukts erschweren. Notwendig sind daher Studien, die sowohl heterogene KPL-Tests als auch eine breite Erfassung etablierter Intelligenzkonstrukte berücksichtigen.

Die inkrementelle Validität von KPL wurde bisher überwiegend bei Schulleistungen nachgewiesen. Evidenz für ein potenzielles KPL-Konstrukt fand sich dabei regelmäßig bei Schulnoten in der mathematisch-naturwissenschaftlichen, aber nicht oder nur begrenzt in der sprachlich-geisteswissenschaftlichen Domäne. Zusätzlich zeigten sich erste Hinweise, dass die inkrementelle Varianzaufklärung durch KPL nicht bestätigt werden konnte, wenn zusätzlich für das Kriterium relevante kognitive Prädiktoren berücksichtigt wurden. Kritisch anzumerken ist auch hier wieder die Verwendung spezifischer Operationalisierungen der Intelligenz als auch von KPL. So zeigte sich beispielsweise, dass einzelne KPL-Tests eine geringere inkrementelle Validität als aggregierte Leistungen über verschiedene KPL-Operationalisierungen hinweg aufwiesen (siehe Greiff & Fischer, 2013; Greiff et al., 2014). Zusammenfassend sollte daher in der weiteren Forschung zur inkrementellen Validität der differentielle Einfluss von KPL auf verschiedene Kriterien (auch hinsichtlich des Generalitätskriteriums) unter Berücksichtigung weiterer relevanter Prädiktoren, konstruktangemessenen Intelligenzoperationalisierungen sowie möglichst verschiedenartigen KPL-Tests untersucht werden.

Im Rahmen dieses Dissertationsprojekts wurde versucht, einige dieser offenen Punkte hinsichtlich der Eigenständigkeit eines potentiellen KPL-Konstrukts zu berücksichtigen. Dazu wurde in Studie 2 (Kapitel 6) das Untersuchungsdesign bisheriger Studien hinsichtlich zweier Punkte adaptiert. Einerseits wurden neben Schulnoten jeweils in der mathematischen sowie in der sprachlichen Domäne auch schulfachliche Kompetenzen (d. h. basale Kompetenzen in Mathematik und Lesen) als externe Kriterien für die Überprüfung der inkrementellen Validität von KPL genutzt. Andererseits wurden die schulfachlichen Kompetenzen in einem zweiten Schritt als relevante Prädiktoren zusätzlich zu einer spezifischen Intelligenzoperationalisierung (d. h. verbales Reasoning) in den Analysen zur inkrementellen

Validität von KPL berücksichtigt. Damit liefert Studie 2 Hinweise hinsichtlich des differentiellen Einflusses von KPL bezüglich verschiedener, externer Kriterien (vgl. Generalitätskriterium) und überprüft gleichzeitig, ob die bisher berichteten Befunde zur inkrementellen Validität von KPL hinsichtlich der Schulnoten auf die fehlende Berücksichtigung von relevanten Prädiktoren zurückzuführen ist (vgl. Wilhelm, 2009).

In Studie 3 (Kapitel 7) wurden insbesondere die Einschränkungen der Konstruktunterrepräsentationen berücksichtigt. Dazu wurde eine breite Intelligenzoperationalisierung (d. h. Reasoning, Bearbeitungsgeschwindigkeit, Merkfähigkeit und Einfallsreichtum sowie kristalline Intelligenz) gemeinsam mit einer breiten KPL-Operationalisierung (d. h. MicroDYN und MicroFIN) eingesetzt. Zusätzlich wurde die inkrementelle Validität hinsichtlich der globalen Schulleistungen unter Berücksichtigung unterschiedlicher, statistischer Analysestrategien untersucht. Studie 3 überprüft damit insbesondere, ob die bisher berichteten Befunde zur Konstruktvalidität von KPL durch die spezifischen Operationalisierungen von Intelligenz und KPL beeinflusst wurden. Gleichzeitig liefern die Befunde der Studie 3 die bisher belastbarsten Hinweise, ob und gegebenenfalls in welcher Art und Weise ein eigenständiges KPL-Konstrukt im nomologischen Netzwerk der Intelligenz angebracht erscheint (Kapitel 8).

Literatur

- Abele, S., Greiff, S., Gschwendtner, T., Wüstenberg, S., Nickolaus, R., Nitzschke, A. & Funke, J. (2012). Dynamische Problemlösekompetenz: Ein bedeutsamer Prädiktor von Problemlöseleistungen in technischen Anforderungskontexten? *Zeitschrift für Erziehungswissenschaft*, 15(2), 363–391. doi:10.1007/s11618-012-0277-9
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *I-S-T 2000 R - Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.
- Cattell, R. B. (1965). *The scientific analysis of personality*. Harmondsworth: Penguin.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281.

- Csapo, B. (1997). The development of inductive reasoning: Cross-sectional assessments in an educational context. *International Journal of Behavioral Development*, 20(4), 609–626.
- Danner, D., Hagemann, D., Schankin, A., Hager, M. & Funke, J. (2011). Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence*, 39, 323–334. doi:10.1016/j.intell.2011.06.004
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Ederer, P., Nedelkoska, L., Patt, A. & Castellazzi, S. (2015). What Do Employers Pay for Employees' Complex Problem Solving Skills? *International Journal of Lifelong Education*, 34(4), 430–447. doi:10.1080/02601370.2015.1060026
- Fischer, A., Greiff, S. & Funke, J. (2012). The process of solving complex problems. *The Journal of Problem Solving*, 4(1), 19–41. doi:10.7771/1932-6246.1118
- Fischer, A., Greiff, S., Wüstenberg, S., Fleischer, J., Buchwald, F. & Funke, J. (2015). Assessing analytic and interactive aspects of problem solving competency. *Learning and Individual Differences*, 39, 172–179. doi:10.1016/j.lindif.2015.02.008
- Frischkorn, G. T., Greiff, S. & Wüstenberg, S. (2014). The development of complex problem solving in adolescence: A latent growth curve analysis. *Journal of Educational Psychology*, 106(4), 1007–1020. doi:10.1037/a0037114
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Funke, J. (2010). Complex problem solving: a case for complex cognition? *Cognitive Processing*, 11, 133–142. doi:10.1007/s10339-009-0345-0
- Greiff, S. (2012). *Individualdiagnostik der komplexen Problemlösefähigkeit*. Münster: Waxmann.
- Greiff, S. & Fischer, A. (2013). Der Nutzen einer komplexen Problemlösekompetenz: Theoretische Überlegungen und empirische Befunde. *Zeitschrift für Pädagogische Psychologie*, 27(1), 27–39. doi:10.1024/1010-0652/a000086
- Greiff, S., Fischer, A., Stadler, M. & Wüstenberg, S. (2014). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning*, 21(3), 356–382. doi:10.1080/13546783.2014.989263
- Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M. & Martin, R. (2013). A multitrait-multimethod study of assessment instruments for complex problem solving. *Intelligence*, 41(5), 579–596. doi:10.1016/j.intell.2013.07.012

- Greiff, S. & Neubert, J. C. (2014). On the relation of complex problem solving, personality, fluid intelligence, and academic achievement. *Learning and Individual Differences*, 36, 37–48. doi:10.1016/j.lindif.2014.08.003
- Greiff, S., Stadler, M., Sonnleitner, P., Wolff, C. & Martin, R. (2015). Sometimes less is more: Comparing the validity of complex problem solving measures. *Intelligence*, 50, 100–113. doi:10.1016/j.intell.2015.02.007
- Greiff, S., Wüstenberg, S. & Funke, J. (2012). Dynamic Problem Solving: A New Assessment Perspective. *Applied Psychological Measurement*, 36(3), 189–213. doi:10.1177/0146621612439620
- Greiff, S., Wüstenberg, S., Molnar, G., Fischer, A., Funke, J. & Csapo, B. (2013). Complex problem solving in educational contexts—something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, 105(2), 364–379. doi:10.1037/a0031856
- Gustafsson, J.-E. & Åberg-Bengtsson, L. (2010). Unidimensionality and interpretability of psychological instruments. In S. E. Embretson (Hrsg.), *Measuring psychological constructs: Advances in model-based approaches*. (S. 97–121). Washington: American Psychological Association.
- Hautamäki, J., Arinen, P., Eronen, S., Hautamäki, A., Kupiainen, S., Lindblom, B., ... Scheinin, P. (2002). *Assessing learning-to-learn: a framework*. Helsinki: National Board of Education.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Beltz.
- Horn, R. (2009). *Standard Progressive Matrices (SPM). Deutsche Bearbeitung und Normierung nach J. C. Raven*. Frankfurt am Main: Pearson Assessment.
- Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35(1), 21–35.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. Form 4*. Göttingen: Hogrefe.
- Jensen, A. R. & Wang, L.-J. (1994). What is a good g? *Intelligence*, 18, 231–258.
- Kersting, M. (2001). Zur Konstrukt- und Kriteriumsvalidität von Problemlösenszenarien anhand der Vorhersage von Vorgesetztenurteilen über die berufliche Bewährung. *Diagnostica*, 47(2), 67–76. doi:10.1026//0012-1924.47.2.67
- Kröner, S., Plass, J. L. & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence*, 33(4), 347–368. doi:10.1016/j.intell.2005.03.002

- Mainert, J., Baggen, Y., Kretzschmar, A., Neubert, J. C., Lans, T., Biemans, H., ... Greiff, S. (eingereicht). Perceiving Entrepreneurial Challenges as Complex Problems: The Role of Complex Problem Solving in Opportunity Identification. *Journal of Business Venturing*.
- Mainert, J., Kretzschmar, A., Neubert, J. C. & Greiff, S. (2015). Linking complex problem solving and general mental ability to career advancement: Does a transversal skill reveal incremental predictive validity? *International Journal of Lifelong Education*, 34(4), 393–411. doi:10.1080/02601370.2015.1060024
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.
- Molnár, G., Greiff, S. & Csapo, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, 9, 35–45. doi:10.1016/j.tsc.2013.03.002
- Neubert, J. C., Kretzschmar, A., Wüstenberg, S. & Greiff, S. (2015). Extending the Assessment of Complex Problem Solving to Finite State Automata: Embracing Heterogeneity. *European Journal of Psychological Assessment*, 31(3), 181–194. doi:10.1027/1015-5759/a000224
- Neubert, J. C., Mainert, J., Kretzschmar, A. & Greiff, S. (2015). The Assessment of 21st Century Skills in Industrial and Organizational Psychology: Complex and Collaborative Problem Solving. *Industrial and Organizational Psychology*, 8(2), 238–268. doi:10.1017/iop.2015.14
- Oberauer, K. [K.], Süß, H.-M., Schulze, R., Wilhelm, O. & Wittmann, W. W. (2000). Working memory capacity — facets of a cognitive ability construct. *Personality and Individual Differences*, 29, 1017–1045. doi:10.1016/S0191-8869(99)00251-2
- Oberauer, K. [Klaus], Schulze, R., Wilhelm, O. & Süß, H.-M. (2005). Working Memory and Intelligence - Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131, 61–65. doi:10.1037/0033-2909.131.1.61
- Oberauer, K. [Klaus], Süß, H.-M., Wilhelm, O. & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, 36, 641–652. doi:10.1016/j.intell.2008.01.007
- OECD. (2014). *Pisa 2012 results: Creative problem solving: Students' skills in tackling real-life problems (Volume V)*. Paris: OECD Publishing.
- Raven, J. (1958). *Advanced progressive matrices*. London: Lewis.

- Reeve, C. L. & Blacksmith, N. (2009). Identifying g: A review of current factor analytic practices in the science of mental abilities. *Intelligence*, 37(5), 487–494. doi:10.1016/j.intell.2009.06.002
- Ross, J. & Ross, C. (1976). *Ross Test of Higher Cognitive Processes*. Novata, CA.: Academic Therapy Publications.
- Rost, D. H. (2009). *Intelligenz: Fakten und Mythen*. Weinheim: Beltz PVU.
- Schaie, K. W. (1994). The course of adult intellectual development. *American psychologist*, 49(4), 304–313.
- Scherer, R., Greiff, S. & Hautamäki, J. (2015). Exploring the Relation between Time on Task and Ability in Complex Problem Solving. *Intelligence*, 48, 37–50. doi:10.1016/j.intell.2014.10.003
- Schmidt, J. (1983). Stabilität und Veränderungen von Intelligenzleistungen und -strukturen. In G. Lüer (Hrsg.), *Deutsche Gesellschaft für Psychologie: Bericht über den 33. Kongress der Deutschen Gesellschaft für Psychologie - 33,1 : Mainz 1982* (S. 371–374). Göttingen: Hogrefe.
- Schweizer, F., Wüstenberg, S. & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*, 24, 42–52. doi:10.1016/j.lindif.2012.12.011
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin, R., ... Lator, T. (2012). The Genetics Lab: Acceptance and psychometric characteristics of a computer-based microworld assessing complex problem solving. *Psychological Test and Assessment Modeling*, 54(1), 54–72. doi:10.1037/e578442014-045
- Sonnleitner, P., Brunner, M., Keller, U. & Martin, R. (2014). Differential relations between facets of complex problem solving and students' immigration background. *Journal of Educational Psychology*, 106(3), 681–695. doi:10.1037/a0035506
- Sonnleitner, P., Keller, U., Martin, R. & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, 41(5), 289–305. doi:10.1016/j.intell.2013.05.002

- Sonnleitner, P., Keller, U., Martin, R., Latour, T. & Brunner, M. (in Druck). Assessing Complex Problem Solving in the Classroom: Meeting Challenges and Opportunities. In B. Csapo & J. Funke (Hrsg.), *The Nature of Problem Solving*. Paris: OECD.
- Stadler, M., Becker, N., Greiff, S. & Spinath, F. M. (2015). The Complex Route to Success: Complex Problem Solving Skills in the Prediction of University Success. *Higher Education Research and Development*. doi:10.1080/07294360.2015.1087387
- Sternberg, R. J., Castejón, J. L., Prieto, M. D., Hautamäki, J. & Grigorenko, E. L. (2001). Confirmatory factor analysis of the Sternberg Triarchic Abilities Test in three international samples: An empirical test of the triarchic theory of intelligence. *European Journal of Psychological Assessment*, 17(1), 1.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen: Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Süß, H.-M. (1999). Intelligenz und komplexes Problemlösen: Perspektiven für eine Kooperation zwischen differentiell-psychometrischer und kognitionspsychologischer Forschung. *Psychologische Rundschau*, 50(4), 220–228. doi:10.1026/0033-3042.50.4.220
- Süß, H.-M. (2006). Eine Intelligenz – viele Intelligenzen? Neuere Intelligenztheorien im Widerstreit. In H. Wagner (Hrsg.), *Intellektuelle Hochbegabung: Aspekte der Diagnostik und Beratung ; Tagungsbericht* (S. 7–39). Bad Honnef: Bock.
- Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O. & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, 30, 261–288. doi:10.1016/S0160-2896(01)00100-3
- Weber, H. & Westmeyer, H. (2001). Die Inflation der Intelligenzen. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 251–266). Lengerich: Pabst.
- Weiß, R. (2006). *CFT 20-R - Grundintelligenztest Skala 2 - Revision*. Göttingen: Hogrefe.
- Wilhelm, O. (2009). Issues in Computerized Ability Measurement: Getting out of the Jingle and Jangle Jungle. In F. Scheuermann & J. Björnsson (Hrsg.), *The Transition to Computer-Based Assessment New Approaches to Skills Assessment and Implications for Large-scale Testing* (S. 145–149). Luxembourg: Office for Official Publications of the European Communities.

- Wilhelm, O. & Kunina-Habenicht, O. (2015). Pädagogisch-psychologische Diagnostik. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 305–328). Berlin: Springer.
- Wittmann, W. W. & Hatstrup, K. (2004). The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioral Science*, 21(4), 393–409. doi:10.1002/sres.653
- Wittmann, W. W. & Matt, G. E. (1986). Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts- und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des Berliner Intelligenzstrukturmodells. *Diagnostica*, 32, 309–329.
- Wittmann, W. W. & Schmidt, J. (1983). *Die Vorhersagbarkeit des Verhaltens aus Trait-Inventaren. Theoretische Grundlagen und empirische Ergebnisse mit dem Freiburger Persönlichkeitsinventar (FPI). Forschungsbericht 10*. Freiburg: Psychologisches Institut der Albert-Ludwigs-Universität Freiburg i. Br.
- Wittmann, W. W. & Süß, H.-M. (1996). Vorhersage und Erklärung von Schulnoten durch das Berliner Intelligenzstrukturmodell. In B.-J. Ertelt & M. Hofer (Hrsg.), *Theorie und Praxis der Beratung. Beratung in Schule, Familie, Beruf und Betrieb. Beiträge zur Arbeitsmarkt- und Berufsforschung*. (203, S. 161–183). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit.
- Wüstenberg, S., Greiff, S. & Funke, J. (2012). Complex problem solving — More than reasoning? *Intelligence*, 40(1), 1–14. doi:10.1016/j.intell.2011.11.003
- Wüstenberg, S., Stadler, M., Hautamäki, J. & Greiff, S. (2014). The Role of Strategy Knowledge for the Application of Strategies in Complex Problem Solving Tasks. *Technology, Knowledge and Learning*. doi:10.1007/s10758-014-9222-8
- Ziegler, M., Booth, T. & Bensch, D. (2013). Getting Entangled in the Nomological Net: Thoughts on Validity and Conceptual Overlap. *European Journal of Psychological Assessment*, 29(3), 157–161. doi:10.1027/1015-5759/a000173

Kapitel 6

Studie 2: KPL und die Bedeutung von schulfachlichen Kompetenzen

Kretzschmar, A., Neubert, J. C., & Greiff, S. (2014). Komplexes Problemlösen, schulfachliche Kompetenzen und ihre Relation zu Schulnoten. *Zeitschrift für Pädagogische Psychologie*, 28(4), 205–215. <http://doi.org/10.1024/1010-0652/a000137>

Kapitel 7

Studie 3: KPL im nomologischen Netzwerk der Intelligenz

Kretzschmar, A., Neubert, J. C., Wüstenberg, S., & Greiff, S. (2015). Construct Validity of Complex Problem Solving: A Comprehensive View on Different Facets of Intelligence and School Grades. *Intelligence*. doi:10.1016/j.intell.2015.11.004

Kapitel 8

KPL aus der Perspektive der BIS-Inhaltsmodalität

Die Ergebnisse der Studie 3 (Kapitel 7) hinsichtlich der diskriminanten Validität zeigen deutlich, dass (1) neben Reasoning die weiteren untersuchten Intelligenzkonstrukte Bearbeitungsgeschwindigkeit, Merkfähigkeit, Einfallsreichtum oder kristalline Intelligenz¹ keinen signifikanten Einfluss auf KPL haben, und (2) sich neben den etablierten Intelligenzkonstrukten ein zusätzlicher latenter Faktor für KPL nachweisen lässt.

8.1 Alternative Interpretation zur Konstruktvalidität

Der zusätzliche latente KPL-Faktor neben den Faktoren der etablierten Intelligenzkonstrukte (siehe Abbildung 3 in Kapitel 7) stützt augenscheinlich die Annahme, dass KPL etwas anderes erfasst als traditionelle Intelligenztests (Funke, 2003; Greiff, 2012; Sonnleitner, Keller, Martin & Brunner, 2013). Dies führt wiederum direkt zu der Frage, ob bestehende Intelligenztheorien um einen eigenen KPL-Faktor erweitert werden sollten. Tatsächlich wurde seit jeher die Offenheit des BIS-Modells betont (Jäger et al., 1997), so dass sich in der Forschungsliteratur auch entsprechende Überlegungen finden, das BIS-Modell um praktische (Sperber,

¹In diesem Kapitel wird die kristalline Intelligenz nicht weiter beachtet, sondern nur die Intelligenzfähigkeiten des BIS-Modells (Jäger, Süß & Beauducel, 1997) als Grundlage genommen. Die nachfolgende Argumentation wird dadurch jedoch nicht beeinflusst.

Wörpel, Jäger & Pfister, 1985) und auditive Intelligenzfähigkeiten (Conzelmann & Süß, 2015) zu ergänzen.

Für die praktische und für die auditive Intelligenzfähigkeit wird jedoch eine Erweiterung hinsichtlich der Inhalts- und nicht der operativen Modalität postuliert, so dass als Resultat beispielsweise auditives Reasoning analog zu figuralem, numerischem oder verbalem Reasoning in einem erweiterten BIS-Modell stünde. Dies ist insofern nachvollziehbar, da die kognitiven Anforderungen von Reasoning (entsprechend der BIS-Modalität Operationen) in gleichem Maße bei Papier-und-Bleistift-basierten und auditiven Intelligenzaufgaben relevant sind.

Bemerkenswerterweise wurde in der bisherigen KPL-Forschung jedoch stets im Sinne einer Abgrenzung zu den operativen Fähigkeiten argumentiert. Während also die partielle Unabhängigkeit von Reasoning wiederholt diskutiert wurde (z. B. Greiff et al., 2013; Sonnleitner et al., 2013; Wüstenberg, Greiff & Funke, 2012), blieb die Inhaltsmodalität bisher unberücksichtigt. Dies ist umso erstaunlicher, da die zur Beschreibung eines potentiellen KPL-Konstrukts genutzten Konzepte wie Interaktivität oder Dynamik vielmehr Aufgaben- und weniger Kognitionsbeschreibungen entsprechen (Funke, 2003).²

Eine alternative Erklärungsweise der bisherigen Befundlage könnte also auch sein, dass die spezifische KPL-Varianz nicht im Sinne einer neuartigen operativen Fähigkeit zu interpretieren ist, sondern im Sinne einer im BIS-Modell bisher nicht berücksichtigten Inhaltsfacette (ähnlich wie auditive und praktische Intelligenz). KPL könnte also auch als „komplexes“ oder „interaktives“ Reasoning neben figuralem, numerischem und verbalem Reasoning im BIS-Modell aufgefasst werden.

8.2 Empirische Überprüfung

Die Überprüfung dieser Annahme lässt sich durch modifizierte Analysen der im Kapitel 7 genutzten Daten (Studie 3) durchführen. Die Details zur Methodik sowie zu dem zugrundeliegenden Messmodell von KPL können den entsprechenden Abschnitten des vorherigen Kapitels entnommen werden.

²Dies gilt besonders für die oft genutzte KPL-Definition von Buchner: „CPS is the successful interaction with task environments that are dynamic (i.e.; change as a function of the user's interventions and/or as a function of time) and in which some, if not all, of the environment's regularities can only be revealed by successful exploration and integration of the information gained in that process.“ (Frensch & Funke, 1995, S. 14)

Die Aufgaben des BIS-Tests wurden für die nachfolgend dargestellte Analyse entsprechend der Inhaltsmodalität – und nicht wie im Kapitel 7 entsprechend der operativen Fähigkeiten – ausgewertet (siehe Jäger et al., 1997). Dazu wurden den Berechnungsvorschriften des Testmanuals folgend inhaltshomogene Parcels gebildet (d. h. die operativen Varianzanteile wurden unterdrückt), um die inhaltsgebundenen figuralen, numerischen und verbalen Fähigkeiten modellieren zu können. Das resultierende *nested-factor*-Messmodell mit einem *g*-Faktor und spezifischen Faktoren für figurale, numerische und verbale Fähigkeit zeigte einen guten Fit auf die Daten ($\chi^2 = 22.59$, $df = 18$, $p = .21$; CFI = 0.995; Gamma = .993; SRMR = .025; BIC = 2521); vergleichbar mit Befunden in früheren Studien (z. B. Süß, 2013; Süß & Beauducel, 2005, 2015).

Dieses Modell wurde analog zu dem Vorgehen im Kapitel 7 durch die Integration der KPL-Aufgaben (d. h. MicroFIN und MicroDYN) um einen spezifischen KPL-Faktor erweitert (siehe Abbildung 8.1). Auch dieses Messmodell zeigte einen guten Modellfit ($\chi^2 = 184.85$, $df = 122$, $p = .00$; CFI = 0.966; Gamma = .971; SRMR = .052; BIC = 4187); alle Faktorladungen inklusive die des KPL-Faktors waren signifikant und substantiell.

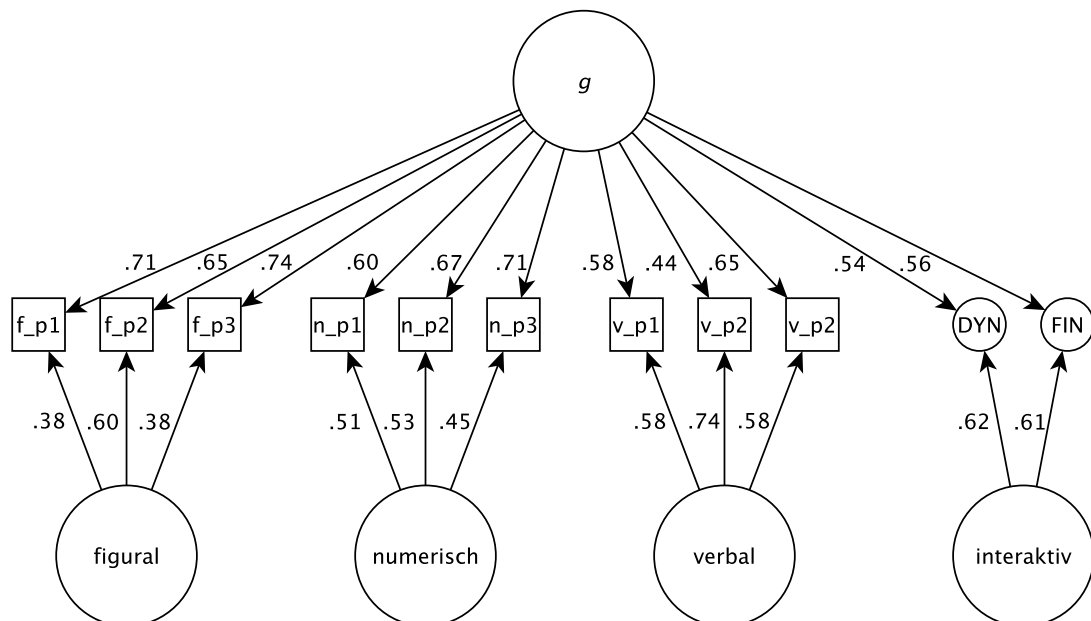


ABBILDUNG 8.1: *Nested-factor*-Messmodell auf der Grundlage der BIS-Inhaltsdimensionen mit zusätzlichem KPL-Faktor „interaktiv“. Manifeste Indikatoren (Parcels) in Rechtecken, latente Faktoren in Ellipsen dargestellt. DYN = MicroDYN; FIN = MicroFIN. Alle Faktorladungen sind signifikant ($p < .05$).

Zu beachten ist bei dem in Abbildung 8.1 dargestellten Messmodell, dass der orthogonale KPL-Faktor „interaktiv“ nur die spezifische Varianz der KPL-Aufgaben in Abgrenzung zu den Inhaltsfaktoren figural, numerisch, verbal sowie dem *g*-Faktor enthält. Vergleichbar zu der Interpretation im Kapitel 7, bei dem der spezifische KPL-Faktor als eigenständige operative Fähigkeit interpretiert wurde, kann der spezifische KPL-Faktor in diesem Modell also als eigenständige inhaltsgebundene Fähigkeit interpretiert werden.

8.3 Diskussion und Implikation

Die dargestellten Überlegungen und Ergebnisse legen den Schluss nahe, dass KPL nicht zwangsläufig als eine neuartige, operative Fähigkeit angesehen werden muss, sondern auch als eine eigenständige Inhaltskomponente (hier mit „interaktiv“ beschrieben) aufgefasst werden kann. Letztendlich liegen also zwei mögliche Erklärungsansätze für die spezifische KPL-Varianz vor, nämlich KPL als eigenständige Fähigkeit in Abgrenzung zu operativen Fähigkeiten wie beispielsweise Reasoning (siehe Abbildung 3 in Kapitel 7), sowie KPL als eigenständige Inhaltsfacette in Abgrenzung zu inhaltsgebundenen Fähigkeiten wie beispielsweise verbalen Fähigkeiten (siehe Abbildung 8.1).

Welcher der beiden Erklärungsansätze zutreffender ist, wird alleine anhand der Empirie nicht entschieden werden können. Vielmehr ist zu berücksichtigen, dass Korrelationen $r < 1.00$ und eigenständige latente Faktoren in Messmodellen verschiedenartige Ursachen haben können. Beispielsweise wäre auch eine Interpretation der Ergebnisse aus Kapitel 7 in der Art möglich, dass der spezifische KPL-Faktor neben den Faktoren der operativen Fähigkeiten auf eine inhaltspezifische KPL-Varianz „interaktiv“ zurückzuführen ist. Im Anhang B sind entsprechende Analysen auf der Grundlage des BIS-Tests dargestellt, welche aufzeigen, dass sich auch spezifische Inhaltsvarianzen neben operativen Fähigkeiten als latente Faktoren in einem gemeinsamen Messmodell modellieren lassen.³

Daher sind theoretische Argumente notwendig, die erklären können, warum KPL eine in den bisherigen Intelligenzkonstrukten nicht berücksichtigte Kognition und nicht vielmehr eine andersartige Inhaltsfacette darstellen soll (vgl. Wilhelm, 2009).

³Analog gilt, dass sich ein latenter Faktor entsprechend einer spezifischen operativen Fähigkeit in einem Messmodell mit inhaltspezifischen Faktoren aufzeigen lässt (vgl. Abbildung 8.1).

In der Tat sind die konzeptionellen Überlappungen hinsichtlich der kognitiven Anforderungen (d. h. im Sinne der operativen Fähigkeiten) von KPL mit etablierten Intelligenzkonstrukten beachtenswert (siehe Kapitel 7). Auch weist Funke (2003) darauf hin, dass bisherige Abgrenzungsversuche vor allem auf der Ebene der Aufgabenanforderungen stattgefunden haben, eine Unterscheidung auf der Ebene der kognitiven Prozesse bis dato aber nicht überzeugend dargelegt wurde.

Offensichtlich sind dahingehend die Unterschiede in der Testdarbietung beziehungsweise in dem Aufgabenformat (d. h. statisch vs. interaktiv; zusätzlich aber auch oftmals noch Papier-und-Bleistift vs. computerbasiert) von KPL-Tests und Operationalisierungen von etablierten Intelligenzkonstrukten. Eine Interpretation von KPL im Sinne einer neuen Inhaltsfacette „interaktiv“ scheint daher unter der Berücksichtigung der aktuellen Forschungslage mindestens genauso gerechtfertigt, wenn nicht sogar überzeugender.

Was würde diese Sichtweise für die Frage nach einem eigenständigen KPL-Konstrukt und der Erweiterung bestehender Intelligenztheorien bedeuten? Wie anfangs erwähnt, ist hinsichtlich der kognitiven Fähigkeiten vor allem Reasoning, nicht jedoch Bearbeitungsgeschwindigkeit, Merkfähigkeit und Einfallsreichtum, für die KPL-Leistung relevant. Eine Ergänzung des BIS-Modells um eine Inhaltsfacette „interaktiv“ wäre auf Basis der bisherigen KPL-Operationalisierungen derzeit also nur für Reasoning, nicht jedoch für die weiteren operativen Fähigkeiten angebracht (Abbildung 8.2 veranschaulicht eine derartige Modellerweiterung). Entsprechend der ursprünglichen Konzeption des BIS-Modells sind auf Zellenebene (z. B. „interaktives“ Reasoning) jedoch keine Fähigkeiten, sondern nur Leistungen in spezifischen Aufgaben angesiedelt (Jäger et al., 1997; Süß & Beauducel, 2011). Diese entspricht letztendlich den Anforderungen einer inhaltsvaliden Operationalisierung von Intelligenzfähigkeiten über verschiedene Aufgabenformate hinweg (vgl. Jensen & Wang, 1994; Reeve & Blacksmith, 2009) sowie im weiteren Sinne auch dem in Kapitel 5.1.1 dargestellten Generalitätskriterium für kognitive Fähigkeitskonstrukte. Um also das BIS-Modell um eine Inhaltsfacette „interaktiv“ zu erweitern, bedarf es – sofern man der ursprünglichen Konzeption des BIS-Modells folgt — auch Operationalisierungen, die die kognitiven Fähigkeiten Bearbeitungsgeschwindigkeit, Merkfähigkeit und Einfallsreichtum „interaktiv“ abbilden. Ob dies möglich ist, kann entsprechend nur in zukünftigen Forschungsarbeiten gezeigt werden.

- Jensen, A. R. & Wang, L.-J. (1994). What is a good g? *Intelligence*, 18, 231–258.
- Reeve, C. L. & Blacksmith, N. (2009). Identifying g: A review of current factor analytic practices in the science of mental abilities. *Intelligence*, 37(5), 487–494. doi:10.1016/j.intell.2009.06.002
- Sonnleitner, P., Keller, U., Martin, R. & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, 41(5), 289–305. doi:10.1016/j.intell.2013.05.002
- Sperber, W., Wörpel, S., Jäger, A. O. & Pfister, R. (1985). *Praktische Intelligenz: Untersuchungsbericht und erste Ergebnisse*. Arbeitsberichte des Forschungsschwerpunktes "Produktives Denken, intelligentes Verhalten". Berlin: Institut für Psychologie im Fachbereich Erziehungs- und Unterrichtswissenschaften der Freien Universität Berlin.
- Süß, H.-M. (2013). The construct validity of the Berlin Intelligence Structure Model (BIS). In A. Roazzi, B. Campello & W. Bilsky (Hrsg.), *Proceedings of the 14th Facet Theory Conference. Searching for structure in complex social, cultural & psychological phenomena* (S. 140–156). Recife, Brazil: Facet Theory Organization.
- Süß, H.-M. & Beauducel, A. (2005). Faceted Models of Intelligence. In O. Wilhelm & R. W. Engle (Hrsg.), *Handbook of understanding and measuring intelligence* (S. 313–332). Thousand Oaks, Calif.: Sage Publications.
- Süß, H.-M. & Beauducel, A. (2011). Intelligenztests und ihre Bezüge zu Intelligenztheorien. In L. Hornke, M. Amelang & M. Kersting (Hrsg.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik* (Bd. 3, S. 97–234). Enzyklopädie der Psychologie, Serie Psychologische Diagnostik. Göttingen: Hogrefe.
- Süß, H.-M. & Beauducel, A. (2015). Modeling the construct validity of the Berlin Intelligence Structure Model. *Estudos de Psicologia (Campinas)*, 32(1), 13–25. doi:10.1590/0103-166X2015000100002
- Wilhelm, O. (2009). Issues in Computerized Ability Measurement: Getting out of the Jingle and Jangle Jungle. In F. Scheuermann & J. Björnsson (Hrsg.), *The Transition to Computer-Based Assessment New Approaches to Skills Assessment and Implications for Large-scale Testing* (S. 145–149). Luxembourg: Office for Official Publications of the European Communities.
- Wüstenberg, S., Greiff, S. & Funke, J. (2012). Complex problem solving — More than reasoning? *Intelligence*, 40(1), 1–14. doi:10.1016/j.intell.2011.11.003

Teil IV

Synopse

Kapitel 9

Zusammenfassung der bisherigen Befunde und Implikationen

Das Hauptziel dieser Arbeit war die Überprüfung der Konstruktvalidität eines potenziellen KPL-Konstrukts. Dazu wurde im ersten Teil dieses Dissertationsprojekts die KPL-Diagnostik durch die Entwicklung von MicroFIN erweitert, so dass im zweiten Teil der Frage nach einem eigenständigen KPL-Konstrukt im nomologischen Netzwerk der Intelligenz auf Basis einer breiten KPL-Operationalisierung nachgegangen werden konnte. In diesem abschließenden Kapitel wird diese Struktur beibehalten, so dass sich die Besprechung der Befunde sowie die Implikationen zuerst auf die KPL-Diagnostik bezieht, bevor anschließend auf die Konstruktvalidität eingegangen wird.

9.1 KPL-Diagnostik

Die Güte der KPL-Tests bestimmte seit jeher die KPL-Forschung, so dass viele anfängliche Befunde zur Konstruktvalidität von KPL auf die spezifischen Eigenschaften und Defizite der KPL-Tests zurückzuführen waren (z. B. Funke, 2003; Greiff, 2012; Süß, 1996, 1999). Erst in jüngerer Zeit stehen durch innovative Weiterentwicklungen und eine Fokussierung auf psychometrische Testeigenschaften mit MicroDYN (Greiff, 2012; Greiff, Wüstenberg & Funke, 2012) und GeneticsLab (Sonnleitner et al., 2012, 2013) KPL-Tests zur Verfügung, welche die methodischen Unzulänglichkeiten früherer Messinstrumente überwunden haben. Allerdings ging

diese Entwicklung einher mit einer Einengung der kognitiven Anforderungen, weshalb die Inhaltsvalidität der neuen KPL-Tests zum Teil kritisch hinterfragt wurde (Funke, 2010, 2014, sowie Kapitel 2).

Als Konsequenz wurde mit MicroFIN ein neuer KPL-Test entwickelt, welcher die psychometrischen Vorteile aktueller KPL-Tests mit den heterogeneren Anforderungen früherer Messinstrumente verbinden sollte. Die bisherigen Studien zeigten zufriedenstellende psychometrische Qualitäten (siehe Kapitel 3) sowie erwartungskonforme konvergente Zusammenhänge (siehe Kapitel 5). Dies ist insofern begrüßenswert, da nun mit MicroFIN, MicroDYN sowie GeneticsLab unterschiedliche KPL-Tests genutzt werden können, um die Erforschung der Konstruktvalidität von KPL auf verschiedenartige, psychometrisch akzeptable Operationalisierungen stützen zu können.

Dabei darf nicht außer acht gelassen werden, dass für MicroFIN, aber teilweise auch für die anderen KPL-Tests im Allgemeinen, noch ein zum Teil erhebliches Weiterentwicklungs- und Forschungspotenzial besteht. Einige ausgewählte Punkte, die für die weitere KPL-Forschung besonders relevant erscheinen, sollen abschließend noch einmal umrissen werden.

Für KPL-Tests auf Basis linearer Strukturgleichungen liegen Arbeiten vor, in denen der systematische Einfluss von Systemmerkmalen auf die Aufgabenschwierigkeit untersucht wurde (z.B. Funke, 1985, 1992; Greiff, 2012; Greiff, Krkovic & Nagy, 2014). Theoretisch können so die Aufgaben über das gesamte Schwierigkeitsspektrum skaliert werden, wobei dies praktisch nur teilweise gelingt (vgl. die Aufgabenschwierigkeiten beispielsweise bei Greiff, 2012; Schweizer, Wüstenberg & Greiff, 2013; Wüstenberg, Greiff & Funke, 2012). Für MicroFIN, basierend auf dem Ansatz der finiten Automaten (Buchner & Funke, 1993; Funke & Buchner, 1992), liegen bisher keine derartigen Untersuchungen vor. Dies mag insbesondere auch daran liegen, dass es aufgrund der Heterogenität der Aufgaben schwieriger ist, spezifische und vergleichbare Systemmerkmale für eine Schwierigkeitsskalierung zu identifizieren. Erschwerend kommt hinzu, dass für die Aufgabenschwierigkeiten nicht nur Einzelmerkmale, sondern vielmehr die Interaktionen verschiedener Merkmale relevant sind (Greiff, 2012). In zukünftigen Studien ist daher zu untersuchen, ob beispielsweise naheliegende Indikatoren wie die Anzahl der Zustände oder der (unterschiedlichen) Zustandsübergangsfunktionen zur Schwierigkeitsskalierung der MicroFIN-Aufgaben nützlich sein können. Dabei haben die bisherigen Befunde verdeutlicht, dass derartige Forschungsarbeiten notwendig sind, wenn das

gesamte Schwierigkeitsspektrum bei MicroFIN abgedeckt oder der Test gar für die Individualdiagnostik (z. B. für die Personalauswahl) eingesetzt werden soll.

Eine charakteristische Eigenschaft der neuen KPL-Tests ist die Vorwissensneutralität, die aufgrund der Konstruktionsprinzipien (d. h. abstrakte bzw. fehlende bedeutungsvolle, semantischen Einbettung) angenommen wird. Eine empirische Überprüfung dieser Annahme steht allerdings noch aus. Zwar konnte in Studie 3 (Kapitel 7) aufgezeigt werden, dass Allgemeinwissen keinen Einfluss auf die KPL-Leistung über die Intelligenzkonstrukte des BIS-Modells (Jäger, Süß & Beauducel, 1997) hatte, allerdings kann dieser Befund nur als erster Schritt zur Überprüfung der Vorwissensneutralität gewertet werden. Vielmehr sind differenziertere Wissenstests in zukünftigen Studien zu berücksichtigen, damit die angenommene und charakteristische Eigenschaft der Vorwissensneutralität überprüft werden kann. Tatsächlich haben frühere Studien auf Basis traditioneller Mikrowelten gezeigt, dass die Leistung in bereichsspezifischen Wissenstests zusätzlich zu Intelligenzkonstrukten Varianz der KPL-Leistung aufklären kann (z. B. Süß, 1996). Von besonderer Bedeutung für die neuen KPL-Tests sind dabei vor allem Wissenstests, die auf spezifische Problemlösestrategien abzielen (für einen Überblick siehe Funke, 2003; sowie Wüstenberg, Stadler, Hautamäki & Greiff, 2014, für einen ersten Ansatz bzgl. MicroDYN). Strategien können letztendlich als Handlungswissen aufgefasst werden (Süß, 1996), so dass die Eigenschaft der Vorwissensneutralität nicht mehr gegeben wäre, wenn die Leistung bei den KPL-Tests auch substantiell darauf zurückzuführen ist, dass einige Problemlösende vorhandenes Wissen über spezifische Problemlösestrategien zu ihrem Vorteil nutzen können (für einen Überblick über die Experten-Novizen-Forschung im KPL-Kontext siehe Funke, 2003; sowie Süß, 1996). Dies ist umso relevanter, je größer die Rolle von einzelnen, spezifischen Strategien bei bestimmten KPL-Operationalisierungen ist.

Weitere Forschungsarbeiten sind auch hinsichtlich der Leistungsbewertung bei der KPL-Messung notwendig. Beispielsweise befindet sich die Bewertung der Explorationsleistung bei MicroFIN derzeit noch in einer frühen Entwicklungsphase. Aufgrund der in Kapitel 3 dargestellten Richtlinien zur Konstruktion der aktuellen MicroFIN-Aufgaben kann davon ausgegangen werden, dass es nicht eine einzige, beste Strategie gibt, welche bei allen MicroFIN-Aufgaben zum Erfolg führt. Die Herausforderung besteht letztendlich darin, eine Bewertungsvorschrift zu entwickeln, die die Leistung zum Lösen komplexer Probleme adäquat abbildet. Dabei wird es nicht ausreichend sein, für jede MicroFIN-Aufgabe ad-hoc definierte oder

mittels *data mining* gewonnene Explorationsstrategien zu bewerten (siehe Kapitel 3). Vielmehr besteht die Leistung bei MicroFIN auch darin, zwischen verschiedenen Strategien je Aufgabe wechseln zu können. Nur wenn ein Problemlöser in der Lage ist, eine einmal als sinnvoll identifizierte Strategie wieder aufzugeben und andere Explorationsstrategien bei einer neuen Aufgabe anzuwenden, können die MicroFIN-Aufgaben erfolgreich gelöst werden. Damit umfasst die Explorationsleistung bei MicroFIN also mehr als die reine Strategieranwendung, sondern sie bildet vielmehr auch die Fähigkeit ab, adaptiv auf verschiedene Problemsituationen einzugehen (strategische Adaptivität nach Schoppek & Putz-Osterloh, 2003).¹ Dies ist insofern bedeutsam, da nach der domänenübergreifenden Konzeptualisierung (vgl. Funke, 2003; Greiff, 2012; OECD, 2014) die Fähigkeit zum Lösen komplexer Probleme für sehr verschiedenartige Problemsituationen relevant sein soll, die adaptive Anpassung des Problemlöseverhaltens also als eine charakteristische Eigenschaft des angenommenen KPL-Konstrukts angesehen werden kann.

Den letzten an dieser Stelle näher ausgeführten Punkt stellt der Einfluss der *ICT-Literacy* als besondere Herausforderung an die KPL-Diagnostik dar. Die Leistung, komplexe Probleme zu lösen, wird als domänenunspezifisch und daher explizit nicht auf computerbasierte Probleme eingeschränkt verstanden. Die computerbasierte Aufgabendarstellung ist also eine Konsequenz der spezifischen Eigenschaften von komplexen Problemen (z. B. Interaktivität, Dynamik), aber letztendlich nur eine Testdarbietungsmethode. Ein substantieller Einfluss der *ICT-Literacy* auf die KPL-Leistung wäre also als Gefährdung der Testfairness beziehungsweise der Validität des KPL-Tests zu interpretieren. Die hier dargestellten Forschungsbefunde (Studie 1; Kapitel 4) lassen darauf schließen, dass dies weder für MicroDYN noch für MicroFIN der Fall ist. Zwar konnte analog zu früheren Studien (z. B. Süß, 1996; Wagener, 2001) ein signifikanter Einfluss von *ICT-Literacy* aufgezeigt werden, jedoch war dieser (1) nicht beziehungsweise nur teilweise größer ausgeprägt als bei computerbasierten Darstellungen traditioneller, kognitiver Fähigkeitstests, sowie (2) zu einem großen Teil auf die gemeinsame Varianz von *ICT-Literacy* und Intelligenzleistungen zurückzuführen. Zu betonen bleibt allerdings, dass die Befunde nicht ohne Weiteres auf neue oder modifizierte KPL-Tests generalisiert werden dürfen. Tatsächlich geht die computerbasierte Testdarbietung mit einer

¹Hierdurch wird gleichzeitig auch eine Brücke zwischen der KPL-Forschung und dem Forschungsfeld der exekutiven Funktionen (z. B. Carlson, 2005) geschlagen, welches in der neueren Forschung zur Konstruktvalidität von KPL (siehe Kapitel 5) noch keine große Berücksichtigung gefunden hat (für einen ersten Ansatz siehe Schmuck, 1992).

Vielzahl an unterschiedlichen Darstellungs- und Bedienmöglichkeiten einher, welche den Einfluss der *ICT-Literacy* verringern oder vergrößern können. Sorgfältige Testkonstruktionen in Verbindung mit – noch immer zu selten durchgeführten – Usability-Studien (vgl. Sonnleitner, Keller, Martin, Latour & Brunner, in Druck) sowie technischen Weiterentwicklungen (z. B. intuitiv bedienbare Tabletcomputer) können dabei den nicht intendierten Einfluss der *ICT-Literacy* auf die jeweilige Testleistung substantiell reduzieren.

Trotz dieses Entwicklungs- und Forschungsbedarfs hinsichtlich der KPL-Diagnostik kann zusammenfassend festgehalten werden, dass der KPL-Forschung mit MicroDYN, GeneticsLab und MicroFIN im Sinne der psychometrischen Qualität die bisher überzeugendsten Messinstrumente zur Verfügung stehen. Dabei ergänzen sich die KPL-Tests insofern, dass jedes in Abhängigkeit von der Forschungsfrage seine spezifischen Stärken hat. MicroDYN und GeneticsLab sind innovative Messinstrumente, die sich mit der derzeitigen Fokussierung auf die erweiterte VOTAT-Strategie (Vollmeyer, Burns & Holyoak, 1996) in die Tradition der Erforschung des wissenschaftlichen Denkens besonders im naturwissenschaftlichen Bereich stellen (vgl. Chen & Klahr, 1999). GeneticsLab lässt sich durch die bei der Entwicklung durchgeführten, umfangreichen Usability-Studien besonders im Kontext mit jüngeren Schülern einsetzen, während bei Fragestellungen mit Studierenden und älteren Erwachsenen mit Hinblick auf die Darstellung der Aufgaben eher MicroDYN zu bevorzugen ist. Mit MicroFIN ist mit dem Ansatz der finiten Automaten (Buchner & Funke, 1993; Funke & Buchner, 1992) ein KPL-Test entwickelt worden, welcher an den Problemlösenden heterogenere Anforderungen im Vergleich zu MicroDYN und GeneticsLab stellt. Damit sollte sich MicroFIN besonders eignen, um eine breitere Erfassung der Problemlösefähigkeit wieder etwas mehr im Sinne der ursprünglichen KPL-Forschung (vgl. Dörner, Kreuzig, Reither & Stäudel, 1983; Funke, 2014) zu ermöglichen. Ob die jeweiligen KPL-Tests aufgrund der verschiedenartigen Anforderungen unterschiedlich nützlich sind (bspw. im Sinne der inkrementellen Varianzaufklärung von externen Kriterien), ist dabei noch eine offene Fragestellungen für zukünftige Studien.

9.2 Konstruktvalidität von KPL

Der zweite Teil der Arbeit widmete sich der Konstruktvalidität von KPL und damit der Frage, ob ein eigenständiges KPL-Fähigkeitskonstrukt im nomologischen

Netzwerk der Intelligenz angemessen erscheint oder nicht. Anhand von Kriterien, die sich als hilfreich bei der Beurteilung von kognitiven Fähigkeitskonstrukten erwiesen haben, wurden die aktuelle Forschungslage und noch offene Fragestellungen dargestellt (Kapitel 5).

Befürworter eines eigenständigen KPL-Konstrukts stützen sich neben den postulierten Anforderungen, welche komplexe Probleme im Vergleich zu Intelligenztestaufgaben an den Problemlösenden stellen (Funke, 2003, 2010), auf die Vielzahl an empirischen Ergebnissen, in denen KPL über etablierte Intelligenzkonstrukte hinaus Varianz an externen Kriterien – meist Schulnoten – erklären konnte. Eine genauere Betrachtung der Forschungsbefunde zeigte allerdings, dass die inkrementelle Varianzaufklärung unter anderem (1) eher bei der mathematisch-naturwissenschaftlichen und weniger bei der sprachlich-geisteswissenschaftlichen Domäne gelang, (2) in der Regel auf der Basis spezifischer Intelligenz- und KPL-Operationalisierungen und (3) unter Ausschluss weiterer, für die externen Kriterien relevanter Prädiktoren gezeigt wurde.

In Studie 2 (Kapitel 6) wurde daher den Punkten (1) und (3) besondere Berücksichtigung eingeräumt. Die Studienergebnisse können als Replikation der Befundlage angesehen werden, dass der Nachweis der inkrementellen Varianzaufklärung von KPL in der sprachlichen Domäne nicht oder nur selten zu gelingen scheint. Die zusätzliche Varianzaufklärung der Schulnoten in der mathematischen Domäne sank ebenfalls auf nur 1%, wenn für die schulfachliche, mathematische Kompetenz kontrolliert wurde. Auch unter Berücksichtigung von Punkt (2) muss die inkrementelle Validität von KPL daher kritisch hinterfragt werden, da in dieser Studie analog zu früheren Arbeiten nur eine eingeschränkte Intelligenzoperationalisierung zum Einsatz kam. Zu beachten ist dahingehend auch, dass die Unterscheidung zwischen Intelligenztestaufgaben und schulfachlichen Kompetenztestaufgaben nicht unumstritten ist (vgl. Baumert, Brunner, Lüdtke & Trautwein, 2007; Rindermann, 2006; Wilhelm & Nickolaus, 2013). Typische Aufgaben beispielsweise für die mathematische Kompetenz finden sich auch bei numerischen Intelligenztestaufgaben. Nimmt man vereinfachend an, dass die hier eingesetzten schulfachlichen Kompetenztestaufgaben dem Grunde nach Reasoning-Aufgaben seien, so ließen sich die Studienergebnisse auch dahingehend interpretieren, dass sich bei einer breiteren Intelligenzoperationalisierung die inkrementelle Validität von KPL merklich reduziert (vgl. Studie 3).

Neben der Konstruktunterrepräsentation von Intelligenz zeichnet sich der Großteil der KPL-Studien auch dadurch aus, dass nur eine spezifische KPL-Operationalisierung eingesetzt wurde. In Studie 3 (Kapitel 7) zur Konstruktvalidität von KPL wurde daher Punkt (2) besonders berücksichtigt, indem eine breite Intelligenzoperationalisierung sowie zwei unterschiedliche KPL-Operationalisierungen (MicroDYN und MicroFIN) genutzt wurden, um die diskriminante, konvergente und inkrementelle Validität von KPL zu überprüfen.

Hinsichtlich der konvergenten Validität zeigte sich, dass MicroDYN und MicroFIN gemeinsame Varianz teilten (entsprechend dem latenten, globalen KPL-Faktor in einem gemeinsamen Messmodell), auch wenn für etablierte Intelligenzkonstrukte kontrolliert wurde (entsprechend dem gemeinsamen Messmodell für den BIS-Test und den KPL-Tests, bei dem sich signifikante und substantielle Faktorladungen für den spezifischen KPL-Faktor zeigten; siehe Abbildung 3 in Kapitel 7). Damit konnten die Ergebnisse von Neubert, Kretzschmar, Wüstenberg und Greiff (2015) hinsichtlich der konvergenten Validität in Bezug auf MicroDYN und MicroFIN bestätigt werden.

Die Befunde bezüglich der inkrementellen Validität veranschaulichten, dass KPL über eine breite Intelligenzoperationalisierung hinaus keine zusätzliche Varianz an Schulleistungen erklären konnte. Bemerkenswert ist dabei, dass zusätzliche Varianz der Schulnoten durch KPL aufgeklärt werden konnte, wenn analog zu den bisherigen KPL-Studien nur eine spezifische Intelligenzoperationalisierung (d. h. figuralles Reasoning) berücksichtigt wurde. Dies unterstreicht damit erneut die Wichtigkeit, Intelligenzkonstrukte repräsentativ zu operationalisieren (Gustafsson & Åberg-Bengtsson, 2010; Messick, 1995; Shadish, Cook & Campbell, 2001). Gleichzeitig rückt damit die Frage in den Mittelpunkt, ob die in den neueren Studien berichtete inkrementelle Validität von KPL über Intelligenz hinaus letztendlich nur als methodisches Artefakt zu interpretieren ist, da in jenen Studien nur sehr spezifische Intelligenzoperationalisierung eingesetzt wurden. In zukünftigen Forschungsarbeiten sollte daher Intelligenz dem Konstrukt angemessen operationalisiert werden (vgl. Jensen & Wang, 1994; Reeve & Blacksmith, 2009; Rost, 2009), um dieser Fragestellung nachgehen zu können.

Besonderes Augenmerk sollte dabei auch auf die Auswahl der externen Kriterien gelegt werden. Bisherige Forschungsarbeiten haben größtenteils Schulnoten als externes Kriterium genutzt. Dies mag aufgrund der aktuellen Popularität des KPL-Konzepts in den empirischen Bildungswissenschaften (vgl. OECD, 2014), der

leichten Verfügbarkeit von selbstberichteten Schulnoten sowie der offensichtlichen Relevanz von Problemlöseprozessen im Schulkontext (siehe Kapitel 6) nicht verwundern. Überzeugender wären jedoch umfassendere und vielseitigere externe Kriterien, die die Alltagsrelevanz sowie die Domänenunspezifität von KPL untermauern können (vgl. Generalitätskriterium). Neuere Studien, in denen beispielsweise Kriterien des Berufserfolgs (z. B. Ederer, Nedelkoska, Patt & Castellazzi, 2015) oder des lebenslangen Lernens (Mainert, Kretzschmar, Neubert & Greiff, 2015) berücksichtigt wurden, sind daher begrüßenswert. Gänzlich ausstehend sind allerdings Zusammenhänge zu alltäglichen Problemlöseleistungen, wie beispielsweise die erstmalige Benutzung eines bisher unbekannten Mobiltelefons. Sollte die inkrementelle Validität von KPL bezüglich der Problemlöseleistungen in derartigen Situationen aufgezeigt werden können, wäre das ein starkes Argument für ein eigenständiges KPL-Konstrukt.

Im Hinblick auf die diskriminante Validität verdeutlichen die Studienergebnisse, dass die KPL-Leistung spezifische Varianz enthält, welche nicht vollständig durch Operationalisierungen etablierter Intelligenzkonstrukte abgedeckt ist. Ungeklärt bleibt die Frage, ob diese spezifische Varianz auf spezielle, kognitive Anforderungen (operative Fähigkeiten im BIS-Modell; Jäger et al., 1997) oder aber auf eine interaktive Art der Aufgabendarstellung (Inhaltsmodalität im BIS-Modell; Jäger et al., 1997) zurückzuführen ist. Die hier dargestellten Analysen und Überlegungen stützen beide Sichtweisen und zeigen damit gleichzeitig auf, dass diese Frage allein empirisch nicht zu lösen ist (siehe Kapitel 8). Vielmehr sollten in zukünftigen Forschungsarbeiten die distinkten, kognitiven Prozesse bei der KPL-Leistung herausgearbeitet werden, auf deren Grundlage eine eigenständige, operative Fähigkeit angenommen und überprüft werden kann (Funke, 2003). Die offensichtlichen Unterschiede zwischen KPL-Tests und traditionellen Intelligenztests hinsichtlich der Aufgabendarbietung lassen hingegen die Interpretation der spezifischen KPL-Varianz als neue Inhaltsfacette „interaktiv“ plausibel erscheinen. In diesem Sinne wäre KPL als „interaktives“ Reasoning neben figuralem, numerischem und verbalem Reasoning aufzufassen.

Bezugnehmend auf die in Kapitel 5 dargestellten Kriterien zur Bewertung eines eigenständigen KPL-Konstrukts lässt sich unter Berücksichtigung der bisherigen Befunde festhalten, dass die Kriterien leistungsdatenbasierte Diagnostik sowie konvergente Validität als zufriedenstellend erfüllt angesehen werden können. Auch das

Kriterium der diskriminanten Validität kann als potenziell erfüllt betrachtet werden, wobei unterschiedliche Erklärungen hinsichtlich der spezifischen KPL-Varianz vorliegen. Die Kriterien Generalität und inkrementelle Validität sind allenfalls als teilweise erfüllt zu betrachten. Bezüglich des Kriteriums zeitliche Stabilität konnten im Rahmen dieses Dissertationsprojekts keine Befunde beigesteuert werden, so dass das Kriterium auch unter Berücksichtigung der gesamten KPL-Forschung als noch nicht ausreichend beforscht angesehen werden muss.

Soll nun in Anbetracht der aktuellen Forschungslage ein potenziell eigenständiges KPL-Konstrukt im nomologischen Netzwerk der Intelligenz bewertet werden, so ist festzuhalten, dass die Befunde zum jetzigen Zeitpunkt wenig Anlass geben, KPL als ein eigenständiges Fähigkeitskonstrukt anzusehen. Während für bestimmte Kriterien (z. B. zeitliche Stabilität) noch keine ausreichenden, empirischen Ergebnisse vorliegen, sind für andere Kriterien (z. B. diskriminante und inkrementelle Validität) Alternativerklärungen der Ergebnisse verfügbar, die gegen ein KPL-Konstrukt in Abgrenzung von etablierten Intelligenzkonstrukten wie Reasoning sprechen. Dabei sollte die Idee eines eigenständigen KPL-Konstrukts nicht voreilig verworfen werden, allerdings haben die hier dargestellten Befunde, Interpretationen und offenen Forschungsfragen gezeigt, dass noch umfangreiche Forschungsarbeiten notwendig sind.

9.3 Ausblick für die weitere KPL-Forschung

Wissenschaft ist ein Prozess der kumulativen Wissens- und Erkenntnisgewinnung, dies trifft insbesondere auch für Validitätsfragestellungen zu (Cronbach & Meehl, 1955; Messick, 1995; Schmidt-Atzert & Amelang, 2012). Deshalb kann am Ende dieser Arbeit auch keine abschließende Antwort auf die Frage nach der Konstruktvalidität eines potenziellen KPL-Konstrukts stehen. Im Gegenteil ist die Arbeit vielmehr auch in der Art zu verstehen, dass sie Erkenntnisse beisteuern möchte, um weitere Forschung anzuregen. Dies ist natürlich auch vor dem Hintergrund der Limitationen dieser Arbeit relevant, wie sie an den entsprechenden Stellen in den jeweiligen Kapiteln ausgeführt wurden.

Auch wenn zukünftige Forschungsarbeiten den ursprünglichen Gedanken nicht stützen sollten, mit dem KPL-Konzept eine eigenständige, in den Intelligenztheorien bisher nicht berücksichtigte kognitive Fähigkeit zu erfassen, wurden durch

die KPL-Forschung eine Reihe von interessanten Fragestellungen angestoßen, diagnostische Innovationen gefördert und Erkenntnisse über kognitive Fähigkeiten gewonnen. So können beispielsweise die seit Beginn der KPL-Forschung betonten Verhaltensanalysen (Dörner, 1986) nun unter Zuhilfenahme der Logfiles aktueller KPL-Tests Einblicke in kognitive Prozesse ermöglichen (z. B. Wüstenberg et al., 2014), die über die bloße Analyse von Reaktionszeiten bei etablierten Intelligenztests (z. B. Goldhammer, Naumann & Greiff, 2015; siehe auch Carroll, 1993) hinausgehen können (z. B. Scherer, Greiff & Hautamäki, 2015).

Abschließend soll ein nicht zu unterschätzender Aspekt der aktuellen KPL-Forschung gewürdigt werden. Obwohl im bildungswissenschaftlichen Kontext starke Ressentiments gegen intelligenzbasierte Forschung anzutreffen sind (Adey, Csapo, Demetriou, Hautamäki & Shayer, 2007), ist es mit Hilfe der KPL-Forschung gelungen, psychologische Fragestellungen bezüglich der kognitiven Fähigkeiten in den empirischen Bildungswissenschaften zu etablieren (z. B. PISA-Studie; OECD, 2014). Durch die Verfügbarkeit dieser bis dato einmaligen und umfangreichen Daten der internationalen Schulleistungsstudien eröffnen sich nun Möglichkeiten, eine Vielzahl an beispielsweise kulturvergleichenden oder entwicklungspsychologischen Fragestellungen hinsichtlich der kognitiven Fähigkeiten zu erforschen. Die Befunde und Überlegungen der hier vorliegenden Arbeit können dabei hoffentlich als wertvolle Anregung und Hilfestellung genutzt werden, um spannende und fundierte Erkenntnisse zu gewinnen.

Literatur

- Adey, P., Csapo, B., Demetriou, A., Hautamäki, J. & Shayer, M. (2007). Can we be intelligent about intelligence? *Educational Research Review*, 2(2), 75–97. doi:10.1016/j.edurev.2007.05.001
- Baumert, J., Brunner, M., Lüdtke, O. & Trautwein, U. (2007). Was messen internationale Schulleistungsstudien? - Resultate kumulativer Wissenserwerbsprozesse: Eine Antwort auf Heiner Rindermann. *Psychologische Rundschau*, 58, 118–128. doi:10.1026/0033-3042.58.2.118
- Buchner, A. & Funke, J. (1993). Finite-state automata: Dynamic task environments in problem-solving research. *The Quarterly Journal of Experimental Psychology Section A*, 46(1), 83–118. doi:10.1080/14640749308401068

- Carlson, S. M. (2005). Developmentally Sensitive Measures of Executive Function in Preschool Children. *Developmental Neuropsychology*, 28(2), 595–616. doi:10.1207/s15326942dn2802_3
- Carroll, J. (1993). *Human cognitive abilities. A survey of factoranalytic studies*. Cambridge: Cambridge University Press.
- Chen, Z. & Klahr, D. (1999). All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy. *Child Development*, 70, 1098–1120.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281.
- Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*, 32, 290–208.
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Ederer, P., Nedelkoska, L., Patt, A. & Castellazzi, S. (2015). What Do Employers Pay for Employees' Complex Problem Solving Skills? *International Journal of Lifelong Education*, 34(4), 430–447. doi:10.1080/02601370.2015.1060026
- Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle. *Zeitschrift für Psychologie*, 193, 435–457.
- Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung*. Berlin; New York: Springer.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Funke, J. (2010). Complex problem solving: a case for complex cognition? *Cognitive Processing*, 11, 133–142. doi:10.1007/s10339-009-0345-0
- Funke, J. (2014). Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.00739
- Funke, J. & Buchner, A. (1992). Finite Automaten als Instrumente für die Analyse von wissensgeleiteten Problemlöseprozessen: Vorstellung eines neuen Untersuchungsparadigmas. *Sprache & Kognition*, 11, 27–37.
- Goldhammer, F., Naumann, J. & Greiff, S. (2015). More is not Always Better: The Relation between Item Response and Item Response Time in Raven's Matrices. *Journal of Intelligence*, 3(1), 21–40. doi:10.3390/jintelligence3010021
- Greiff, S. (2012). *Individualdiagnostik der komplexen Problemlösefähigkeit*. Münster: Waxmann.

- Greiff, S., Krkovic, K. & Nagy, G. (2014). The systematic variation of task characteristics facilitates the understanding of task difficulty: A cognitive diagnostic modeling approach to complex problem solving. *Psychological Test and Assessment Modeling*, 56(1).
- Greiff, S., Wüstenberg, S. & Funke, J. (2012). Dynamic Problem Solving: A New Assessment Perspective. *Applied Psychological Measurement*, 36(3), 189–213. doi:10.1177/0146621612439620
- Gustafsson, J.-E. & Åberg-Bengtsson, L. (2010). Unidimensionality and interpretability of psychological instruments. In S. E. Embretson (Hrsg.), *Measuring psychological constructs: Advances in model-based approaches*. (S. 97–121). Washington: American Psychological Association.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. Form 4*. Göttingen: Hogrefe.
- Jensen, A. R. & Wang, L.-J. (1994). What is a good g? *Intelligence*, 18, 231–258.
- Mainert, J., Kretschmar, A., Neubert, J. C. & Greiff, S. (2015). Linking complex problem solving and general mental ability to career advancement: Does a transversal skill reveal incremental predictive validity? *International Journal of Lifelong Education*, 34(4), 393–411. doi:10.1080/02601370.2015.1060024
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.
- Neubert, J. C., Kretschmar, A., Wüstenberg, S. & Greiff, S. (2015). Extending the Assessment of Complex Problem Solving to Finite State Automata: Embracing Heterogeneity. *European Journal of Psychological Assessment*, 31(3), 181–194. doi:10.1027/1015-5759/a000224
- OECD. (2014). *Pisa 2012 results: Creative problem solving: Students' skills in tackling real-life problems (Volume V)*. Paris: OECD Publishing.
- Reeve, C. L. & Blacksmith, N. (2009). Identifying g: A review of current factor analytic practices in the science of mental abilities. *Intelligence*, 37(5), 487–494. doi:10.1016/j.intell.2009.06.002
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? *Psychologische Rundschau*, 57, 69–86. doi:10.1026/0033-3042.57.2.69
- Rost, D. H. (2009). *Intelligenz: Fakten und Mythen*. Weinheim: Beltz PVU.
- Scherer, R., Greiff, S. & Hautamäki, J. (2015). Exploring the Relation between Time on Task and Ability in Complex Problem Solving. *Intelligence*, 48, 37–50. doi:10.1016/j.intell.2014.10.003

- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik mit 82 Tabellen*. Berlin: Springer.
- Schmuck, P. (1992). Zum Zusammenhang zwischen der Effizienz exekutiver Kontrolle und dem mehrfachen Lösen eines komplexen Problems. *Sprache & Kognition*, 11(4), 193–207.
- Schoppek, W. & Putz-Osterloh, W. (2003). Individuelle Unterschiede und die Bearbeitung komplexer Probleme. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24(3), 163–173.
- Schweizer, F., Wüstenberg, S. & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*, 24, 42–52. doi:10.1016/j.lindif.2012.12.011
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin, R., ... Latour, T. (2012). The Genetics Lab: Acceptance and psychometric characteristics of a computer-based microworld assessing complex problem solving. *Psychological Test and Assessment Modeling*, 54(1), 54–72. doi:10.1037/e578442014-045
- Sonnleitner, P., Keller, U., Martin, R. & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, 41(5), 289–305. doi:10.1016/j.intell.2013.05.002
- Sonnleitner, P., Keller, U., Martin, R., Latour, T. & Brunner, M. (in Druck). Assessing Complex Problem Solving in the Classroom: Meeting Challenges and Opportunities. In B. Csapo & J. Funke (Hrsg.), *The Nature of Problem Solving*. Paris: OECD.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen: Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Süß, H.-M. (1999). Intelligenz und komplexes Problemlösen: Perspektiven für eine Kooperation zwischen differentiell-psychometrischer und kognitionspsychologischer Forschung. *Psychologische Rundschau*, 50(4), 220–228. doi:10.1026/0033-3042.50.4.220

- Vollmeyer, R., Burns, B. D. & Holyoak, K. J. (1996). The Impact of Goal Specificity on Strategy Use and the Acquisition of Problem Structure. *Cognitive Science*, 20(1), 75–100. doi:10.1207/s15516709cog2001_3
- Wagener, D. (2001). *Psychologische Diagnostik mit komplexen Szenarios - Taxonomie, Entwicklung, Evaluation*. Lengerich: Pabst Science Publishers.
- Wilhelm, O. & Nickolaus, R. (2013). Was grenzt das Kompetenzkonzept von etablierten Kategorien wie Fähigkeit, Fertigkeit oder Intelligenz ab? *Zeitschrift für Erziehungswissenschaft*, 16, 23–26. doi:10.1007/s11618-013-0380-6
- Wüstenberg, S., Greiff, S. & Funke, J. (2012). Complex problem solving — More than reasoning? *Intelligence*, 40(1), 1–14. doi:10.1016/j.intell.2011.11.003
- Wüstenberg, S., Stadler, M., Hautamäki, J. & Greiff, S. (2014). The Role of Strategy Knowledge for the Application of Strategies in Complex Problem Solving Tasks. *Technology, Knowledge and Learning*. doi:10.1007/s10758-014-9222-8

Teil V

Anhang

Anhang A

Deskriptive Statistiken zur Studie *ICT-Literacy* und MicroFIN

TABELLE A.1: Deskriptive Statistiken und Gruppenvergleich zur Studie *ICT-Literacy* und MicroFIN

	Teilnehmer		Nicht-Teilnehmer		Unterschied
Geschlecht M/W	28	/ 63	33	/ 102	$\chi^2 = .81, df = 1, p > .05$
Alter	23.15	(4.04)	22.70	(4.42)	$t(204) = -0.80, p > .05$
MicroFIN	0.05	(0.75)	-0.03	(0.78)	$t(180) = -0.68, p > .05$
MicroDYN	0.13	(0.96)	-0.09	(0.88)	$t(183) = -1.73, p > .05$
BIS-K	0.12	(0.53)	-0.10	(0.61)	$t(195) = -2.77, p = .01$
BIS-B	0.11	(0.63)	-0.10	(0.60)	$t(185) = -2.41, p = .02$
BIS-M	0.11	(0.63)	-0.08	(0.55)	$t(178) = -2.26, p = .03$
BIS-E	0.06	(0.54)	-0.04	(0.54)	$t(189) = -1.31, p > .05$
BOWIT	21.51	(6.31)	20.09	(6.03)	$t(186) = -1.60, p > .05$
PRACOWI	12.51	(4.01)			
TECOWI	10.53	(4.39)			
START-C	7.37	(4.64)			

Anmerkungen: M = Männlich; W = Weiblich; BIS-K = Reasoning; BIS-B = Bearbeitungsgeschwindigkeit; BIS-M = Merkfähigkeit; BIS-E = Einfallsreichtum; BOWIT = Bochumer Wissenstest; PRACOWI = Praktisches Computerwissen; TECOWI = Theoretisches Computerwissen; START-C = Allgemeines Computerwissen. BIS-Leistungen entsprechend der Berechnungen der Skalenwerte im Testmanual (Jäger, Süß & Beauducel, 1997); z-standardisierte Mittelwerte für MicroFIN und MicroDYN; Summenwert für BOWIT, PRACOWI, TECOWI und START-C.

Anhang B

KPL aus der Perspektive der BIS-Inhaltsmodalität

Ein inhaltsspezifischer Faktor in einem auf operativen Fähigkeiten basierendem BIS-Messmodell

Durch die Symmetrie der Inhalts- und operativen Facetten im BIS-Modell sowie mit Hilfe der theoriegeleiteten Aggregation ist es möglich, inhalts- oder aber operationsunabhängige Faktoren auf Basis der BIS-Aufgaben zu modellieren (siehe Jäger, 1982; Jäger, Süß & Beauducel, 1997). Dieses Vorgehen ist für KPL jedoch nicht anwendbar, da keine den einzelnen Modalitäten eindeutig zugehörigen Aufgaben vorliegen, die analog zu der BIS-Parcelbildung derart aggregiert werden können, dass die Inhalts- oder operativen Varianzen unterdrückt werden können. Im Gegensatz zu den latenten BIS-Faktoren, welche entweder auf den operativen oder den inhaltsspezifischen Varianzen basieren, sind bei einem KPL-Faktor also immer sowohl operative als auch inhaltsspezifische Varianzanteile relevant.

Dieser Umstand kann dazu führen, dass sich in einem Modell mit latenten Faktoren der operativen BIS-Fähigkeiten (d. h. Reasoning, Bearbeitungsgeschwindigkeit, Merkfähigkeit und Einfallsreichtum) auch ein eigenständiger KPL-Faktor modellieren ließe, der jedoch auf einer spezifischen Inhaltsvarianz – und nicht auf einer spezifischen Varianz einer operativen Fähigkeit – basiert.

Abbildung B.1 veranschaulicht diese Überlegung anhand der BIS-Aufgaben. Für das dargestellte Messmodell wurde die theoriegeleitete Aggregation der BIS-Aufgaben derart modifiziert, dass die Parcels für die latenten Faktoren der operativen Fähigkeiten (d. h. Reasoning, Bearbeitungsgeschwindigkeit, Merkfähigkeit und Einfallsreichtum) nur auf figuralen und numerischen Aufgaben basieren und die verbalen Aufgaben folglich nicht berücksichtigt wurden. Die verbalen Aufgaben wurden hingegen derart aggregiert, dass die Varianz der operativen Fähigkeiten unterdrückt wurde, so dass ein inhaltspezifischer, von den operativen Fähigkeiten unabhängiger, latenter Faktor für verbale Fähigkeit modelliert werden konnte (siehe Anhang C zur Parcelbildung). Das daraus resultierende *nested-factor*-Messmodell mit einem *g*-Faktor, spezifischen Faktoren für die operativen Fähigkeiten Reasoning, Bearbeitungsgeschwindigkeit, Merkfähigkeit und Einfallsreichtum sowie einem spezifischen Faktor für die inhaltsgebundene verbale Fähigkeit zeigte einen guten Modellfit ($\chi^2 = 147.01$, $df = 102$, $p = .00$; CFI = 0.97; Gamma = .97; SRMR = .045; BIC = 12381).

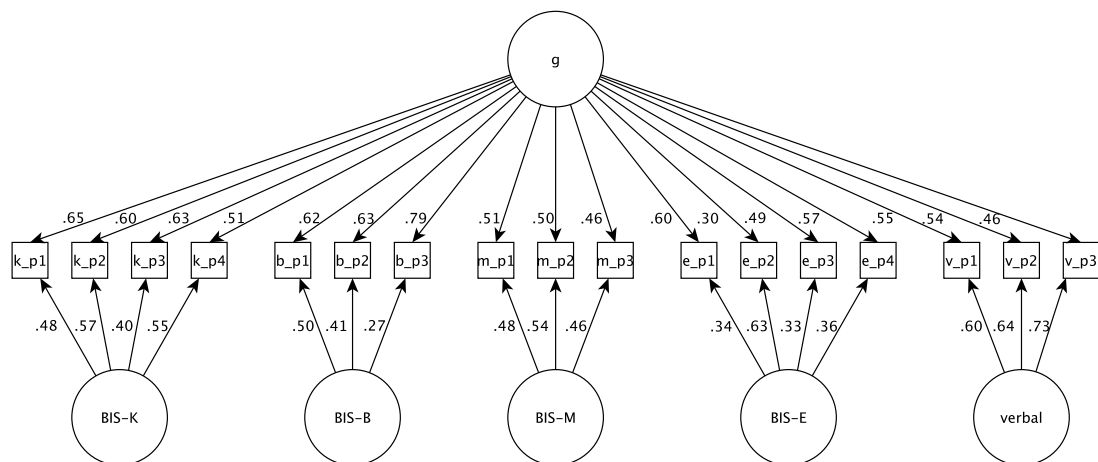


ABBILDUNG B.1: *Nested-factor*-Messmodell des BIS-Tests auf der Grundlage modifizierter Parcelbildung und mit zusätzlichem Faktor für verbale Fähigkeit. Manifeste Indikatoren (Parcels) in Rechtecken, latente Faktoren in Ellipsen dargestellt. BIS-K = Reasoning; BIS-B = Bearbeitungsgeschwindigkeit; BIS-M = Merkfähigkeit; BIS-E = Einfallsreichtum. Alle Faktorladungen sind signifikant ($p < .05$)

Damit konnte verdeutlicht werden, dass sich eine spezifische Inhaltsvarianz durch einen eigenständigen Faktor modellieren lässt, wenn die anderen Faktoren auf operativen, inhaltsunabhängigen Fähigkeiten basieren und in jenen diese spezifische Inhaltsvarianz nicht berücksichtigt wurde. Ein eigenständiger KPL-Faktor, wie er

in Kapitel 7, Abbildung 3, aufgezeigt werden konnte, könnte also auch auf spezifische Inhaltsvarianz von KPL zurückzuführen sein (vgl. Kapitel 8).

Anhang C

Modifizierte Parcelbildung der BIS-Aufgaben

TABELLE C.1: Modifizierte Parcelbildung

Faktor	Parcels						
Reasoning	Parcel 1	= an	+ zn				
	Parcel 2	= ch	+ tl	+ sc			
	Parcel 3	= bg	+ br	+ aw			
	Parcel 4	= fa	+ rd				
Bearbeitungsgeschwindigkeit	Parcel 1	= bd	+ si				
	Parcel 2	= oe	+ rz				
	Parcel 3	= zs	+ xg				
Merkfähigkeit	Parcel 1	= fm	+ zp				
	Parcel 2	= og	+ zz				
	Parcel 3	= we	+ zw				
Einfallsreichtum	Parcel 1	= lo	+ dr				
	Parcel 2	= zf	+ tn				
	Parcel 3	= oj	+ zg				
	Parcel 4	= zk	+ zr				
Verbale Fähigkeit	Parcel 1	= wa	+ tg	+ st	+ ef	+ ws	
	Parcel 2	= tm	+ uw	+ wm	+ ma	+ am	
	Parcel 3	= sv	+ kw	+ ps	+ it		

Anmerkungen: Aufgabenbezeichnung nach Jäger, Süß und Beauducel (1997, Tabelle A 4.2)

Literatur

- Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195–225.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. Form 4*. Göttingen: Hogrefe.