

# Naturwissenschaftliche Wettbewerbe für Schülerinnen und Schüler

Charakterisierung der Anforderungen und  
Teilnehmenden hinsichtlich spezifischer  
Leistungsmerkmale

Dissertation  
zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von  
Christine Köhler

Kiel, 2017



Erste Gutachterin : Prof. Dr. Ilka Parchmann  
Zweiter Gutachter : Prof. Dr. Knut Neumann  
Tag der mündlichen Prüfung : 22.05.2017  
Zum Druck genehmigt : 22.05.2017

gez. Prof. Dr. Natascha Oppelt, Dekanin



Für Jan,  
Linus und Jonas



## Danksagung

Zahlreiche Personen haben zum Gelingen der vorliegenden Arbeit beigetragen, bei denen ich mich herzlich bedanken möchte. An erster Stelle danke ich Frau Prof. Dr. Ilka Parchmann, die mir die Möglichkeit gab meine Dissertation am Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik anzufertigen. Ich fühlte mich stets gut beraten und danke für das Vertrauen in meine Arbeit.

Für die fachliche Unterstützung und Beratung sowie für das offene Ohr zu jeder Zeit danke ich im Besonderen Dr. Sascha Bernholt und Dr. Tim Höffler. Für alle Fragen aus dem Bereich der naturwissenschaftlichen Schülerwettbewerbe konnte ich stets PD Dr. Heide Peters und Dr. Stefan Petersen zu Rate ziehen, wofür ich mich ebenso herzlich bedanke.

Ohne all die Schülerinnen und Schüler, die geduldig an den zehrenden Erhebungen teilnahmen, hätte diese Arbeit nicht entstehen können. Vielen Dank! Neben den Schülerinnen und Schülern danke ich allen Lehrkräften und Schulleitungen, die die Erhebungen erst ermöglicht haben, für ihre Unterstützung. Für die Ermöglichung der Erhebungen im Rahmen von Schülerwettbewerben und jegliche Unterstützung möchte ich besonders PD Dr. Heide Peters und Claudia Körper danken. Des Weiteren danke ich Dr. Daniel Osewold und der Stiftung Jugend forscht e.V. für ihre Unterstützung.

Für die unermüdliche Unterstützung bei den Erhebungen quer durch Deutschland, bei der Datenauswertung und Literaturbeschaffung und für die gute Zusammenarbeit danke ich Kristina Schwarzlow, Vanessa Vogt, Daniela Ingwersen und Jeannot Gersner.

Allen Kolleginnen und Kollegen der Didaktik der Chemie danke ich für das konstruktive Arbeitsklima und die gute Zusammenarbeit. Insbesondere möchte ich hier Gerda Ledwig erwähnen, die mich überhaupt erst auf den Gedanken brachte, eine Promotion in Erwägung zu ziehen. Meiner Mitbewohnerin der Büro-WG Dr. Janet Blankenburg gilt mein besonderer Dank für die zahlreichen fachlichen und nicht-fachlichen Gespräche, für die kreativen Kaffeepausen und produktiven Konferenzen. Meiner kürzlich zugezogenen zweiten Mitbewohnerin der Büro-WG, Anneke Steegh, danke ich besonders für die Unterstützung und Versorgung mit Nervennahrung in der letzten Phase meiner Promotion. Für aufheiternde und humorvolle Emails, Gespräche, Ideen und Reisen danke ich zudem Dr. Pay Dierks, Dr. Stefan Schwarzer und Wilfried Wentorf.

Abschließend möchte ich mich bei meinen Liebsten bedanken: Bei Inge für die zahlreichen konstruktiven Diskussionen in den Mittagspausen und darüber hinaus. Bei meiner Familie für ihren Stolz und die Unterstützung in allen Belangen. Bei Jan, ohne den ich eine Promotion nie begonnen hätte, und unseren wundervollen Kindern Linus und Jonas - fürs Diskutieren, Motivieren, Unterstützen und Ertragen.



## Zusammenfassung

Die nachhaltige und zielgerichtete Förderung naturwissenschaftlich interessierter und talentierter Schülerinnen und Schüler stellt ein wichtiges bildungspolitisches Ziel dar, um der stetig wachsenden Nachfrage an qualifizierten Fachkräften im naturwissenschaftlich-technischen Bereich nachzukommen. Als eine geeignete Fördermaßnahme haben sich naturwissenschaftliche Schülerwettbewerbe erwiesen, über deren Anforderungen und Merkmale ihrer Teilnehmenden nur wenige Erkenntnisse vorliegen, die auf Daten basieren, die während oder kurz nach einer Wettbewerbsteilnahme erfasst wurden. Die Charakterisierung der Anforderungen und Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe anhand spezifischer Leistungsmerkmale ist daher Gegenstand der vorliegenden Dissertation.

Zunächst wurde ein Überblick über Forschung im Bereich naturwissenschaftlicher Schülerwettbewerbe erstellt. Um theoriebasiert ein Testinstrument entwickeln zu können, das auch die Anforderungen naturwissenschaftlicher Schülerwettbewerbe berücksichtigt, wurden die Anforderungen zweier exemplarisch gewählter Wettbewerbe (Internationale JuniorScienceOlympiade (IJSO) sowie Jugend forscht bzw. Schüler experimentieren) analysiert. Für die Analyse von Wettbewerbsaufgaben wurde ein Klassifikationsschema entwickelt und validiert. Aus den Erkenntnissen, z. B. zu in den Wettbewerben schwerpunktmäßig vertretenen Kompetenzbereichen, konnten Anforderungsprofile der untersuchten Wettbewerbe erstellt und für die Entwicklung eines Testinstruments zur Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten genutzt werden. Die Konzeption des Testinstruments zielte darauf ab, in den Fächern Biologie, Chemie und Physik und den Kompetenzbereichen Fachwissen und Erkenntnisgewinnung die naturwissenschaftlichen Kenntnisse und Fähigkeiten von Wettbewerbsteilnehmenden im Vergleich zu Schülerinnen und Schülern, die noch nie an einem entsprechenden Wettbewerb teilgenommen hatten, zu erfassen.

Das Testinstrument wurde in einer ersten Studie mit  $n = 195$  Wettbewerbsteilnehmenden (u. a. aus dem Bundesfinale der IJSO 2013) und Schülerinnen und Schülern, die noch nie an einem entsprechenden Wettbewerb teilgenommen hatten, validiert. Die anhand von Methoden der probabilistischen Testtheorie erhaltenen Ergebnisse bestätigten die Validität des Testinstruments. In weiteren Analysen fanden sich für die Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb die Schulnoten in den Naturwissenschaften, das Alter sowie das Interesse an Tätigkeiten, die in naturwissenschaftlichen Fördermaßnahmen ausgeübt werden, als signifikante Prädiktoren. Als signifikante Prädiktoren für Wettbewerbserfolg wurden die naturwissenschaftlichen Kenntnisse und Fähigkeiten, die Schulnoten in den Naturwissenschaften sowie die Anzahl der Wettbewerbsteilnahmen identifiziert. Weitere Ergebnisse zeigten, dass Teilnehmende des IJSO-Bundesfinales 2013 signifikant höhere naturwissenschaftliche Kenntnisse und Fähigkeiten aufwiesen als Schülerinnen und Schülern, die mindestens einmal oder noch nie an einem entsprechenden Wettbewerb teilgenommen hatten. Auch der Unterschied zwischen Wettbewerbsteilnehmenden, die mindestens einmal in einem mathematisch-naturwissenschaftlichem Wettbewerb teilgenommen hatten, und wettbewerbsunerfahrenen Schülerinnen und Schülern war signifikant.

Da in der Wettbewerbsforschung der Zugang zu Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe zeitnah zum Wettbewerb nur eingeschränkt und in knappen Zeitfenstern möglich ist, jedoch viele Faktoren zur Charakterisierung von Wettbewerbsteilnehmenden er-

forderlich sind, wurde in einer zweiten Studie mit  $n = 955$  Wettbewerbsteilnehmenden und wettbewerbsunerfahrenen Schülerinnen und Schülern ein reduziertes Testinstrument erprobt, dessen Güte jedoch nicht den in der ersten Studie erreichten Qualitätskriterien entsprach. Das Vorliegen von Differential Item Functioning (DIF) gab jedoch Hinweise auf existierende spezifische Charakteristika der untersuchten Subgruppen.

Zur Identifikation möglicher spezifischer Charakteristika der untersuchten Subgruppen wurden die im reduzierten Testinstrument eingesetzten Aufgaben in einer dritten Studie qualitativ und varianzanalytisch untersucht. Die Ergebnisse zeigten signifikante Mittelwertsunterschiede sowie Unterschiede in den jeweils erreichten Antwortkategorien, z. B. zwischen Teilnehmenden der IJSO-Bundesfinale und Schülerinnen und Schülern, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten. Hieraus konnten erste Hinweise auf spezifische Charakteristika einzelner Subgruppen gewonnen werden. So schnitten bspw. Teilnehmende der IJSO-Bundesfinale besser in Aufgaben ab, in denen fachlich begründete Aussagen gefordert waren, während Teilnehmende an Jugend forscht bzw. Schüler experimentieren eine Stärke im Planen von Versuchen aufwiesen.

Mit dem in der ersten Studie entwickelten Instrument zur Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten können in Kombination mit Instrumenten zur Erfassung affektiver Merkmale umfassende Charakterisierungen von Wettbewerbsteilnehmenden und Schülerinnen und Schülern, die noch nie an einem entsprechenden Wettbewerb teilgenommen hatten, erfolgen. Diese Erkenntnisse können in einem nächsten Schritt zur Identifizierung und, in Kombination mit den Hinweisen zu spezifischen Charakteristika von Teilnehmenden unterschiedlicher Wettbewerbe und den erstellten Anforderungsprofilen der Wettbewerbe, zur zielgerichteten Förderung naturwissenschaftlich interessierter Schülerinnen und Schüler genutzt werden. Das entwickelte Schema zur Klassifikation und Analyse von Wettbewerbsaufgaben kann zudem zur Charakterisierung weiterer aufgabenorientierter Wettbewerbe sowie zur Entwicklung von bspw. wettbewerbsergänzenden Übungsmaterialien genutzt werden. Die Erkenntnisse über Einflussfaktoren einer Wettbewerbsteilnahme und Wettbewerbserfolg können zudem zur Konzeption von Maßnahmen genutzt werden, die gerade junge Schülerinnen und Schüler zu einer Wettbewerbsteilnahme motivieren bzw. Zugangsbarrieren reduzieren. Auf lange Sicht könnte so eine nachhaltige Förderung Jugendlicher in den Naturwissenschaften erfolgen, die möglichst viele ermutigt, einen Beruf im naturwissenschaftlich-technischen Bereich zu ergreifen.

## Abstract

An important educational goal is to foster interested and talented students in science in a lasting and purposeful way in order to satisfy the constantly growing need of qualified workforce in the field of science, technology, engineering and mathematics (STEM). Student competitions in science have proved to be a suitable measure, but there are few research findings in this field about requirements of competitions or participants which are usually not based on data which has been collected during or shortly after the competition. The characterization of the requirements and participants in scientific student competitions with regard to specific attributes of performance is the object of investigation of this dissertation.

First, an overview of research in the field of scientific student competitions was given. In order to develop a theory-based test instrument which also takes into account the requirements of scientific student competitions, the requirements of two selected competitions (IJSO and *Jugend forscht* or *Schüler experimentieren*) were analysed. For this purpose, a scheme to classify and analyse competition tasks was developed and validated. Based on these findings, for example regarding the main areas of competence covered by the competitions, it was possible to specify the requirements of the competitions which were then used to develop an instrument to measure scientific knowledge and abilities. The test aimed at covering the scientific knowledge and abilities in biology, chemistry and physics as well as in two areas of competence (use and acquisition of knowledge). Furthermore, the test aimed at measuring the scientific knowledge and abilities of participants of scientific student competitions compared to students who had never participated in a scientific competition.

The test was validated in a first study with  $n = 195$  participants of scientific student competitions (among others from the final German round of the IJSO) and students who had never participated in a scientific competition. The results obtained by using item response theory analyses confirmed the validity of the instrument. In further analyses, the grades in science, the age and the interest in science-related enrichment activities were found to be significant predictors for the participation in a scientific students competition. For a successful participation in a scientific students competition, the scientific knowledge and abilities, the grades in science and the number of previous participations were identified as significant predictors. Further results showed that participants of the final German round of the IJSO 2013 had significantly higher scientific knowledge and abilities than students who had participated at least once or never in a corresponding competition. The difference between those who took at least once part in a scientific competition and students who never did was also significant.

In competition research, access to participants is only possible to a limited extent and in tight time windows. Yet, many factors are needed to characterize participants. Therefore, a second study with a reduced test instrument with  $n = 955$  participants and students who had never participated in a scientific competition was conducted, but the quality of the reduced instrument did not reach the criteria of quality which were achieved in the first study. However, the presence of DIF gave evidence of existing specific characteristics of the investigated subgroups.

In order to identify possible specific characteristics of different subgroups, the items which were used in the reduced test were investigated qualitatively and by analyses of variance in

a third study. The results showed significant differences in the mean values achieved in the different items as well as differences in the response categories of the items, e. g. between participants of the final German round of the IJSO and students who had at least once participated in a scientific students competition. From this, first indications of specific characteristics of individual subgroups could be obtained. Participants in the final German round of the IJSO, for example, performed better in items that required specialist explanation, while participants of *Jugend forscht* or *Schüler experimentieren* showed a strength in the planning of experiments.

In combination with the test for scientific knowledge and abilities developed in the first study and instruments for the assessment of affective characteristics, a comprehensive characterization of participants in scientific student competitions and students who had never participated in a corresponding competition will be possible. These findings can then be used in a next step to identify and, in combination with the indications of specific characteristics of participants in different competitions and the identified requirements of the competitions, to foster interested students in science in a purposeful way. Additionally, the developed scheme to classify and analyse competition tasks can be used to characterize further task-oriented competitions and to develop, for example, competitive training materials. The findings on influencing factors of competition participation and of success in a competition can also be used to conceive measures that are especially motivating young students to participate in competitions or reduce access barriers. In the long term, this could be a lasting support for young people in the natural sciences, which may encourage as many students as possible to enter a career in the field of STEM.

# Inhaltsverzeichnis

Abbildungsverzeichnis	v
Tabellenverzeichnis	vii
Abkürzungsverzeichnis	ix
<b>1 Einleitung</b>	<b>1</b>
<b>2 Theoretischer Hintergrund</b>	<b>5</b>
2.1 Das Modell IKoN	5
2.1.1 Motivation und Zielsetzung	5
2.1.2 Theoretischer Hintergrund zum Modell IKoN	6
2.2 Naturwissenschaftliche Schülerwettbewerbe	8
2.2.1 Schülerwettbewerbe und ihre Funktion im Bildungswesen	8
2.2.2 Naturwissenschaftliche Schülerwettbewerbe - ein Überblick	10
2.2.3 Charakteristika von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe	12
2.2.4 Gründe für die Teilnahme an einem naturwissenschaftlichen Wettbewerb	19
2.2.5 Einflussfaktoren für eine erfolgreiche Wettbewerbsteilnahme	20
2.2.6 Zusammenfassung und Fazit	21
2.3 Naturwissenschaftliche Kenntnisse und Fähigkeiten	24
2.3.1 Definition: Naturwissenschaftliche Kenntnisse und Fähigkeiten	24
2.3.2 Zum Zusammenhang von Intelligenz, (domänenspezifischem) Wissen und Leistung	26
2.3.3 Zusammenfassung und Fazit	28
<b>3 Zielsetzungen und Forschungsfragen</b>	<b>29</b>
<b>4 Analyse von Wettbewerbsanforderungen</b>	<b>31</b>
4.1 Klassifikation und Analyse von Aufgaben: ein Überblick	31
4.2 Analyse der Anforderungen in der Internationalen JuniorScienceOlympiade	36
4.2.1 Entwicklung eines Schemas zur Klassifikation und Analyse von Wettbewerbsaufgaben	37
4.2.2 Validierung des Schemas zur Klassifikation und Analyse von Wettbewerbsaufgaben	42
4.2.3 Beispiele: Klassifikation ausgewählter Aufgaben der ersten Runde der Internationalen JuniorScienceOlympiade aus 2012 und 2014	44

4.2.4	Deskriptive Ergebnisse der Analyse von Wettbewerbsaufgaben der Internationalen JuniorScienceOlympiade . . . . .	46
4.2.5	Interpretation und Diskussion der deskriptiven Ergebnisse . . . . .	50
4.3	Analyse der Anforderungen im Wettbewerb Jugend forscht bzw. Schüler experimentieren . . . . .	55
4.3.1	Analyse der Jury-Arbeit und wettbewerbsinterner Richtlinien . . . . .	55
4.3.2	Qualitative Befragung von Jurorinnen und Juroren und deskriptive Ergebnisse . . . . .	57
4.3.3	Interpretation und Diskussion der deskriptiven Ergebnisse . . . . .	58
4.4	Zusammenfassung und Fazit . . . . .	62
<b>5</b>	<b>Entwicklung eines Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten von Wettbewerbsteilnehmenden</b>	<b>63</b>
5.1	Erstellung eines Testentwurfs . . . . .	63
5.1.1	Festlegung der Zielgruppe . . . . .	63
5.1.2	Rahmenbedingungen . . . . .	65
5.1.3	Testziel und Konstruktionsstrategie . . . . .	65
5.1.4	Generieren von Indikatoren und Eingrenzen des Konstrukts . . . . .	65
5.1.5	Definition des Messgegenstands . . . . .	65
5.1.6	Wahl des Itemformats . . . . .	66
5.2	Entwicklung und Auswahl der Aufgaben für den ersten Testentwurf . . . . .	66
5.2.1	Entwicklung von Aufgaben . . . . .	66
5.2.2	Auswahl der Aufgaben für den ersten Testentwurf . . . . .	69
5.3	Prä-Pilotierungen des ersten Testentwurfs . . . . .	71
5.4	Pilotierung des zweiten Testentwurfs . . . . .	74
5.4.1	Schwierigkeitsanalyse . . . . .	75
5.4.2	Reliabilitätsanalyse . . . . .	76
5.4.3	Weitere Auswertungen . . . . .	76
5.5	Zusammenfassung und Fazit . . . . .	77
<b>6</b>	<b>Validierung des Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten von Wettbewerbsteilnehmenden (Studie 1)</b>	<b>79</b>
6.1	Stichprobe . . . . .	79
6.2	Erhebungsinstrumente . . . . .	80
6.2.1	Fragebogen . . . . .	80
6.2.2	Kognitiver Fähigkeitstest . . . . .	80
6.3	Ablauf der Erhebungen . . . . .	82
6.4	Auswertung . . . . .	84
6.4.1	Beurteilerreliabilität . . . . .	84
6.4.2	Umgang mit fehlenden Werten in Itemantworten . . . . .	84
6.4.3	Skalierung der Itemantworten . . . . .	85
6.4.4	Testqualität . . . . .	87
6.5	Ergebnisse . . . . .	90
6.5.1	Analyse fehlender Werte . . . . .	90
6.5.2	Skalierung der Itemantworten . . . . .	90
6.5.3	Testqualität . . . . .	92
6.6	Interpretation und Diskussion der Ergebnisse . . . . .	102

6.6.1	Analyse fehlender Werte . . . . .	102
6.6.2	Skalierung der Itemantworten . . . . .	102
6.6.3	Testqualität . . . . .	103
6.7	Zusammenfassung und Fazit . . . . .	108
<b>7</b>	<b>Erprobung eines reduzierten Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten (Studie 2)</b>	<b>111</b>
7.1	Folgerungen aus Studie 1 . . . . .	111
7.1.1	Folgerungen für die Rekrutierung der Stichprobe . . . . .	112
7.1.2	Folgerungen für den Ablauf der Erhebungen und die eingesetzten Erhebungsinstrumente . . . . .	113
7.1.3	Kürzung des Testinstruments . . . . .	114
7.2	Stichprobe . . . . .	115
7.2.1	Teilnehmende des Bundesfinales der Internationalen JuniorScienceOlympiade 2013 und 2014 . . . . .	116
7.2.2	Wettbewerbsteilnehmende . . . . .	116
7.2.3	Nichtteilnehmende . . . . .	117
7.3	Erhebungsinstrumente . . . . .	117
7.3.1	Fragebogen . . . . .	117
7.3.2	Kreativitätstest . . . . .	117
7.3.3	Kognitiver Fähigkeitstest . . . . .	118
7.4	Ablauf der Erhebungen . . . . .	119
7.5	Auswertung . . . . .	119
7.5.1	Beurteilerreliabilität . . . . .	119
7.5.2	Umgang mit fehlenden Werten in Itemantworten . . . . .	120
7.5.3	Ausschluss von Fällen . . . . .	120
7.5.4	Skalierung der Itemantworten . . . . .	121
7.5.5	Testqualität . . . . .	121
7.6	Ergebnisse . . . . .	121
7.6.1	Skalierung der Itemantworten . . . . .	121
7.6.2	Testqualität . . . . .	122
7.7	Interpretation und Diskussion der Ergebnisse . . . . .	126
7.8	Zusammenfassung und Fazit . . . . .	129
<b>8</b>	<b>Varianzanalytischer Vergleich und qualitative Analyse der Verteilungen der Antwortkategorien für einzelne Aufgaben und Subgruppen (Studie 3)</b>	<b>131</b>
8.1	Varianzanalytischer Vergleich der Mittelwerte . . . . .	132
8.2	Vergleich der nach Gruppen getrennten Verteilung der Kategorien . . . . .	132
8.2.1	Vergleich der Wettbewerbsgruppen . . . . .	132
8.2.2	Vergleich von Teilnehmenden der Internationalen JuniorScienceOlympiade und Jugend forscht bzw. Schüler experimentieren . . . . .	139
8.3	Interpretation der Ergebnisse des Vergleichs der Verteilungen der Antwortkategorien anhand des Kategoriensystems . . . . .	143
8.3.1	Aufgabe EB10 . . . . .	143
8.3.2	Aufgabe EC02 . . . . .	144
8.3.3	Aufgabe EP11 . . . . .	145
8.3.4	Aufgabe FB04 . . . . .	146

8.3.5	Aufgabe FB14 . . . . .	146
8.3.6	Aufgabe FC10 . . . . .	147
8.3.7	Aufgabe FC12 . . . . .	148
8.3.8	Aufgabe FP05 . . . . .	149
8.3.9	Aufgabe FP07 . . . . .	150
8.3.10	Aufgabe KP01 . . . . .	150
8.4	Spezifische Charakteristika und erkennbare Trends einzelner Gruppen . . . . .	152
8.5	Diskussion . . . . .	155
8.6	Zusammenfassung und Fazit . . . . .	156
<b>9</b>	<b>Diskussion und Ausblick</b>	<b>157</b>
9.1	Zusammenfassung und Diskussion . . . . .	157
9.2	Theoretische und praktische Implikationen . . . . .	162
9.3	Limitationen der Arbeit . . . . .	164
9.4	Ausblick auf weiterführende Forschung . . . . .	165
	<b>Anhang</b>	<b>167</b>
<b>A</b>	<b>Sonderfälle bei der Klassifikation von Aufgaben der Internationalen Junior- ScienceOlympiade</b>	<b>167</b>
A.1	Sonderfälle bei der Lösungskomplexität . . . . .	167
A.2	Sonderfälle bei den Kompetenzbereichen . . . . .	168
A.3	Sonderfälle auf der Ebene einzelner Aufgaben . . . . .	172
<b>B</b>	<b>Musterlösungen zu den Beispielen von klassifizierten Aufgaben</b>	<b>175</b>
<b>C</b>	<b>Fragebogen für Jury-Mitglieder des Wettbewerbs Jugend forscht</b>	<b>181</b>
<b>D</b>	<b>Beispielaufgaben aus Studie 1</b>	<b>185</b>
<b>E</b>	<b>Fragebogen aus Studie 1</b>	<b>189</b>
<b>F</b>	<b>Itemparameter, Itemfit und Itemstatistiken (Studie 1)</b>	<b>203</b>
<b>G</b>	<b>Fragebogen aus Studie 2</b>	<b>207</b>
<b>H</b>	<b>Itemparameter, Itemfit und Itemstatistiken (Studie 2)</b>	<b>215</b>
<b>I</b>	<b>Itemparameter der DIF-Analysen (Studie 2)</b>	<b>217</b>
	<b>Literaturverzeichnis</b>	<b>221</b>

# Abbildungsverzeichnis

2.1	Das Modell „Individuelles Konzept über die Naturwissenschaften“ . . . . .	7
4.1	Schema zur Klassifikation und Analyse von Wettbewerbsaufgaben . . . . .	38
4.2	Aufgabe 4 der ersten Runde der IJSO 2012 und Aufgabe 1 der ersten Runde der IJSO 2014 . . . . .	45
4.3	Verteilung der Aufgabenbeantwortung über die IJSO-Wettbewerbsrunden . . .	49
4.4	Ergebnisse der Analyse von Wettbewerbsaufgaben der IJSO . . . . .	51
5.1	Beispiele aus der Pilotierung einer erwartungskonformen und einer unüblichen Verteilung der Häufigkeiten auf die verschiedenen Kategorien bzw. erreichten Punkte . . . . .	75
5.2	Finales Testinstrument nach Abschluss und Auswertung der Pilotierungen . . .	77
6.1	Ergebnisse der Analyse fehlender Werte für Studie 1 . . . . .	91
6.2	Wright Map . . . . .	93
6.3	Ergebnisse der varianzanalytischen Auswertung der Daten aus Studie 1 . . . . .	101
7.1	Wright Map . . . . .	123
7.2	<i>Expected Score Curves</i> für Aufgaben EB10 und EP11 . . . . .	127
8.1	Vergleich der nach Wettbewerbsgruppe getrennten Verteilung der Kategorien für einzelne Aufgaben . . . . .	136
8.0	Vergleich der nach Wettbewerbsgruppe getrennten Verteilung der Kategorien .	137
8.1	Vergleich der nach Teilnehmenden der IJSO bzw. Jufo-Teilnehmenden getrennten Verteilung der Kategorien für einzelne Aufgabe . . . . .	141
8.0	Vergleich der nach Teilnehmenden der IJSO bzw. Jufo-Teilnehmenden getrennten Verteilung der Kategorien für einzelne Aufgabe . . . . .	142
8.1	Übersicht über das Abschneiden der Wettbewerbsgruppen in den verschiedenen Aufgabenkategorien . . . . .	154
8.2	Übersicht über das Abschneiden der IJSO- und Jufo-Teilnehmenden in den verschiedenen Aufgabenkategorien . . . . .	154



# Tabellenverzeichnis

2.1	Übersicht über die von der Arbeitsgemeinschaft bundesweiter Schülerwettbewerbe gelisteten mathematisch-naturwissenschaftlichen Wettbewerbe und ihre Charakteristika . . . . .	11
4.1	Übersicht über allgemeindidaktische Klassifikationsschemata . . . . .	34
4.2	Übersicht über naturwissenschaftsdidaktische Klassifikationsschemata . . . . .	35
4.3	Übersicht über die Themengebiete des Schemas zur Klassifikation und Analyse von Wettbewerbsaufgaben . . . . .	41
4.4	Schema der Lösungskomplexität für die Klassifikation und Analyse von Wettbewerbsaufgaben . . . . .	42
4.5	Deskriptive Ergebnisse der Oberflächenmerkmale der klassifizierten Aufgaben der IJSO . . . . .	48
4.6	Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Für wie wichtig erachten Sie die folgenden Aspekte für eine gute Leistung im Wettbewerb Jugend forscht? . . . . .	57
4.7	Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Für wie wichtig erachten Sie die folgenden Kompetenzbereiche für eine gute Leistung im Wettbewerb Jugend forscht? . . . . .	58
4.8	Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Verteilen Sie auf die folgenden Aspekte entsprechend ihrer Bedeutung für eine gute Leistung im Wettbewerb Jugend forscht die Plätze 1 bis 7 . . . . .	59
4.9	Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Verteilen Sie auf die folgenden Teilleistungen entsprechend ihrer Bedeutung für eine gute Leistung im Wettbewerb Jugend forscht die Plätze 1 bis 4 . . . . .	59
4.10	Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Bitte beurteilen Sie die folgenden Aussagen . . . . .	60
4.11	Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Für wie wichtig erachten Sie die folgenden Aspekte der Präsentation? . . . . .	60
5.1	Auswahl und Verteilung der für den ersten Testentwurf ausgewählten Aufgaben	70
5.2	Verteilung der Aufgaben für die zweite Prä-Pilotierung auf die verschiedenen Kompetenzbereiche und Fächer . . . . .	72
5.3	Beispiel (FB11) aus dem mehrstufigen Kategoriensystem zur Bewertung der Schülerantworten . . . . .	73
6.1	In Studie 1 erhobene Informationen und Konstrukte im Fragebogen zum IKoN	81
6.2	Vergleich zwischen Modellen mit und ohne DIF . . . . .	94

6.3	Vergleich des eindimensionalen PCM (1-dim) mit einem zwei- (2-dim) und einem dreidimensionalen (3-dim) Modell . . . . .	94
6.4	Korrelationen zwischen der Personenfähigkeit und erhobenen Begleitvariablen . . . . .	95
6.5	Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten . . . . .	96
6.6	Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für Mädchen . . . . .	98
6.7	Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für Jungen . . . . .	98
6.8	Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für die Teilnehmenden des IJSO-Bundesfinales . . . . .	98
6.9	Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für Schülerinnen und Schüler, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten . . . . .	98
6.10	Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für Schülerinnen und Schüler, die noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten . . . . .	98
6.11	Korrelationen zwischen der Teilnahme an einem mathematisch-naturwissenschaftlichen Wettbewerb bzw. am Bundesfinale der IJSO 2013 und erhobenen Begleitvariablen . . . . .	99
6.12	Binäre logistische Regression zur Ermittlung von Prädiktoren für die Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb . . . . .	100
6.13	Binäre logistische Regression zur Ermittlung von Prädiktoren für Wettbewerbserfolg (Teilnahme am Bundesfinale der IJSO 2013) . . . . .	100
7.1	In Studie 2 erhobene Informationen und Konstrukte im Fragebogen zum IKoN . . . . .	118
7.2	Vergleich zwischen Modellen mit und ohne DIF . . . . .	124
7.3	Analyse möglicher von DIF betroffener Aufgaben . . . . .	125
8.1	Varianzanalytischer Vergleich der in den jeweiligen Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten erreichten Mittelwerte . . . . .	133
F.1	Schwellenparameter . . . . .	203
F.2	Mean-Squared-Infit und Infit- $t$ . . . . .	204
F.3	Itemstatistiken . . . . .	205
H.1	Schwellenparameter . . . . .	215
H.2	Mean-Squared-Infit und Infit- $t$ . . . . .	215
H.3	Itemstatistiken . . . . .	216
I.1	Itemparameter der DIF-Analysen für das Merkmal Geschlecht . . . . .	217
I.2	Itemparameter der DIF-Analysen für das Merkmal Teilnahme am IJSO-Bundesfinale . . . . .	218
I.3	Itemparameter der DIF-Analysen für das Merkmal Wettbewerbserfahrung . . . . .	219

# Abkürzungsverzeichnis

- AIC** Akaike Information Criterion
- BIC** Bayesian Information Criterion
- BUW** BundesUmweltWettbewerb
- DFG** Deutsche Forschungsgemeinschaft
- DIF** Differential Item Functioning
- IBO** Internationale BiologieOlympiade
- ICC** Item-Characteristic-Curves
- ICHo** Internationale ChemieOlympiade
- IJSO** Internationale JuniorScienceOlympiade
- IKoN** Individuelles Konzept über die Naturwissenschaften
- IQB** Institut zur Qualitätsentwicklung im Bildungswesen
- IMO** Internationale Mathematik-Olympiade
- IRT** Item Response Theory
- IPhO** Internationale PhysikOlympiade
- Jufo** Jugend forscht bzw. Schüler experimentieren
- KFT** Kognitiver Fähigkeitstest
- KMK** Kultusministerkonferenz
- MINT** Mathematik, Informatik, Naturwissenschaften und Technik
- NEPS** Nationales Bildungspanel
- NoS** Nature of Science
- NoSt** Nature of Scientists
- OECD** Organisation for Economic Co-operation and Development
- PISA** Programme for International Student Assessment

**PCM** Partial-Credit-Modell

**TAM** Test Analysis Modules

**TIMSS** Trends in International Mathematics and Science Study

**WLE** Weighted-Likelihood-Estimation

# Kapitel 1

## Einleitung

„Am Anfang sind wir alle MINT.“

(Stefan Schmitt in einem Kommentar über Kinder als geborene Naturforscher  
aus DIE ZEIT Nr. 29/2015)

Bereits im frühen Alter versuchen Kinder, Zusammenhänge ihrer Umgebung zu ergründen. Eine Vielzahl an Untersuchungen zeigte, dass die Voraussetzungen für einen Zugang zu naturwissenschaftlichen Phänomenen bereits im Alter zwischen drei und fünf Jahren angelegt sind (Lück, 2004). Das Interesse an naturwissenschaftlichen Themen und Phänomenen ist bei Kindern im Kindergarten- und Grundschulalter groß, nimmt allerdings im Verlauf der Schulzeit kontinuierlich ab (z. B. Baumert & Köller, 1998; Hoffmann, Krapp, Renninger & Baumert, 1998; K. Möller, 2014). Die daraus resultierende Problematik, dass nur wenige junge Menschen einen Beruf im Bereich von Mathematik, Informatik, Naturwissenschaften und Technik (MINT) anstreben, wodurch der Bedarf an qualifizierten MINT-Fachkräften nicht gedeckt werden kann, ist nicht neu (z. B. Hetze, 2011). So sahen bspw. die Wirtschaftsprogramme der Europäischen Union, die Lissabon-Strategie aus dem Jahr 2000 und ihr Nachfolger Europa 2020 aus dem Jahr 2010, in der Beseitigung des Fachkräftemangels im MINT-Bereich eine wesentliche Voraussetzung dafür, dass Europa wettbewerbsfähig bleiben kann (Europäische Kommission, 2004, 2010). In den letzten Jahren sind darauf hin diverse Initiativen ins Leben gerufen worden, die den Nachwuchs im MINT-Bereich fördern und für Ausbildungsberufe und Studienfächer gewinnen möchte (z. B. „MINT Zukunft schaffen“, „Komm mach MINT“, „MINT EC“, „plus-MINT“).

Trotz dieser Initiativen besteht nach wie vor ein Bedarf an qualifizierten Fachkräften im MINT-Bereich (z. B. Bundesministerium für Bildung und Forschung, 2012). Dass die Schwierigkeit, naturwissenschaftlich-technischen Nachwuchs auch für entsprechende Berufe zu begeistern, aktueller denn je ist, zeigen die Ergebnisse der jüngsten Erhebung des *Programme for International Student Assessment* (PISA) 2015. So zeigte sich für Deutschland, dass die Freude und das Interesse an Naturwissenschaften sowie die Bedeutung und Relevanz, die Jugendliche den Naturwissenschaften für ihre Zukunft zuschreiben (instrumentelle Motivation), im Vergleich der *Organisation for Economic Co-operation and Development* (OECD) unterdurchschnittlich ausgeprägt sind (Schiepe-Tiska, Rönnebeck et al., 2016). Anlass zur Sorge

gibt im Besonderen, dass gegenüber den Ergebnissen von PISA 2006 die Freude und das Interesse an Naturwissenschaften, die instrumentelle Motivation sowie die Selbstwirksamkeitserwartung abnahmen (Schiepe-Tiska, Rönnebeck et al., 2016). Für Deutschland besonders auffällig waren die ausgeprägten Geschlechterdifferenzen zugunsten der Jungen in den genannten Konstrukten sowie erstmals auch in der naturwissenschaftlichen Kompetenz (Schiepe-Tiska, Simm & Schmidtner, 2016). Was eine mögliche Berufswahl im MINT-Bereich betrifft, gaben in PISA 2015 weniger als ein Viertel der Jugendlichen an, später einen naturwissenschaftlichen Beruf ergreifen zu wollen. Im Vergleich zu PISA 2006 stieg dieser Anteil an, wobei mehr Jungen einen Beruf im MINT-Bereich in Erwägung zogen als Mädchen (Schiepe-Tiska, Rönnebeck et al., 2016). Als mögliche Gründe für den Anstieg wurden die gestiegene Medienaufmerksamkeit sowie die zahlreichen Initiativen und Netzwerke zur Förderung von MINT genannt (Schiepe-Tiska, Rönnebeck et al., 2016).

Ein Frage, die die Ergebnisse von PISA 2015 aufwarfen, war, „inwieweit es dem differenzierten Schulsystem in Deutschland gelingt, eine breite naturwissenschaftliche Grundbildung anzulegen, Talente zu fördern und die Leistungsspitze auszubauen“ (Schiepe-Tiska, Rönnebeck et al., 2016, S. 92). Als kritisch anzusehen ist, dass die naturwissenschaftliche Kompetenz von Gymnasiastinnen und Gymnasiasten im Vergleich zu den Ergebnissen aus PISA 2006 gesunken ist, auch wenn der Anteil an Schülerinnen und Schülern auf den Kompetenzstufen V und VI mit 10.6 % über dem OECD-Durchschnitt lag (Schiepe-Tiska, Simm & Schmidtner, 2016). Dieser Befund macht deutlich, dass nicht nur die Förderung leistungsschwacher, sondern auch leistungsstarker Schülerinnen und Schüler von Bedeutung ist. Die Kultusministerkonferenz (KMK) hat hierzu in 2015 eine Förderstrategie initiiert (Kultusministerkonferenz, 2015) sowie in 2016 eine gemeinsame Initiative von Bund und Ländern zur Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler gestartet (Kultusministerkonferenz, 2016). Neben Maßnahmen aus dem Bereich der Diagnostik werden vor allem schulische (Enrichment, Akzeleration, usw.) und außerschulische Maßnahmen (z. B. Schülerlabore) zur Förderung und Begleitung von Talenten empfohlen. In der Förderstrategie der KMK ausdrücklich erwähnt werden als Enrichment-Maßnahme Schülerwettbewerbe (Kultusministerkonferenz, 2015), da diese dazu beitragen,

frühzeitig das Interesse an bestimmten Fachrichtungen zu wecken, Talente zu finden und von der Breitenmotivation zur Förderung besonderer Fähigkeiten zu kommen. Wettbewerbe leisten einen wichtigen Beitrag zur Gewinnung von Fachkräften und akademischem Nachwuchs, zum Beispiel in den MINT-Fächern und in den Ingenieurwissenschaften. (Kultusministerkonferenz, 2015, S. 6)

Dass Schülerwettbewerbe Talente fördern und zur Entwicklung von Begabungen beitragen können ist weitgehend anerkannt (z. B. Fauser & Messner, 2007; Goldstein & Wagner, 1993). Wenig erforscht sind bislang jedoch Wirkmechanismen, spezifische Anforderungen der Schülerwettbewerbe oder Charakteristika von Wettbewerbsteilnehmenden. Kenntnisse über naturwissenschaftliche Schülerwettbewerbe und ihre Teilnehmenden sind jedoch erforderlich, um naturwissenschaftlich interessierte und talentierte Schülerinnen und Schüler gezielt und nachhaltig fördern zu können.

Diesem Desiderat nimmt sich das durch die Deutsche Forschungsgemeinschaft (DFG) geförderte Projekt „Charakterisierung individueller Interessen und Überzeugungen von Schülerinnen und Schülern als Grundlage einer naturwissenschaftlichen Talentförderung“ an, in dessen

Rahmen die vorliegende Arbeit entstanden ist. Anhand verschiedener Merkmale sollten Teilnehmende unterschiedlicher naturwissenschaftlicher Schülerwettbewerbe charakterisiert werden, was wiederum die Basis für eine gezielte Talentförderung und die Charakterisierung und Konzeption wirksamer Fördermaßnahmen darstellen sollte. Der Hintergrund des Projekts und des zugehörigen Modells Individuelles Konzept über die Naturwissenschaften (IKoN) werden in Kapitel 2, Abschnitt 2.1, vorgestellt. Des Weiteren wird der Forschungsstand zu naturwissenschaftlichen Schülerwettbewerben und ihren Teilnehmenden in Kapitel 2, Abschnitt 2.2, dargestellt.

Im Rahmen dieser Arbeit sollten naturwissenschaftliche Kenntnisse und Fähigkeiten von Teilnehmenden verschiedener naturwissenschaftlicher Schülerwettbewerbe erfasst werden. Eine Definition und eine Erläuterung des Zusammenhangs von Intelligenz, Wissen und Leistung sowie, damit einhergehend, dem Aufzeigen des Forschungsdesiderats folgt in Kapitel 2, Abschnitt 2.3.

Nachdem in Kapitel 3 die Zielsetzungen und Forschungsfragen dieser Arbeit erläutert werden, wird in Kapitel 4 eine Analyse der Anforderungen der naturwissenschaftlichen Schülerwettbewerbe IJSO (als ein aufgabenorientierter Wettbewerb, der sich mit Aufgaben in Biologie, Chemie und Physik an Jugendliche bis 15 Jahren richtet) und Jugend forscht (bzw. Schüler experimentieren als ein naturwissenschaftlicher und projektorientierter Wettbewerb bis 14 bzw. 21 Jahren) vorgestellt. Die Analyse der Anforderungen diente der Entwicklung eines Testinstruments zur Erfassung der naturwissenschaftlichen Kenntnisse und Fähigkeiten von Wettbewerbsteilnehmenden. Zusätzlich konnten aus den Ergebnissen Anforderungsprofile der untersuchten Wettbewerbe erstellt werden. Die Entwicklung des Testinstruments wird in Kapitel 5 dargelegt, die Validierung erfolgte quantitativ im Rahmen von Studie 1 und wird in Kapitel 6 erläutert. Ein den Anforderungen der Wettbewerbsforschung angepasstes reduziertes Testinstrument wurde quantitativ in Studie 2, siehe Kapitel 7 erprobt. Eine qualitative Analyse der eingesetzten Aufgaben zur Erfassung der naturwissenschaftlichen Kenntnisse und Fähigkeiten ergänzte Studie 1 und Studie 2 (Kapitel 8). Die Ergebnisse der vorliegenden Arbeit werden in Kapitel 9 zusammengetragen und diskutiert, bevor theoretische und praktische Implikationen erläutert werden sowie ein Ausblick auf weiterführende Forschungsdesiderata gegeben wird.



## Kapitel 2

# Theoretischer Hintergrund

Im folgenden Kapitel werden die theoretischen Hintergründe der vorliegenden Arbeit erläutert. Zunächst wird das Modell IKoN, das dieser Arbeit zugrunde liegt, vorgestellt und der theoretische Rahmen des Modells sowie des Projektes, in dessen Rahmen das Modell IKoN entwickelt wurde, vorgestellt. Im Anschluss wird ein Überblick über naturwissenschaftliche Schülerwettbewerbe gegeben und der Forschungsstand zu naturwissenschaftlichen Schülerwettbewerben und ihren Teilnehmenden resümiert. Im Anschluss wird das in dieser Arbeit zu messende Konstrukt, die naturwissenschaftlichen Kenntnisse und Fähigkeiten, definiert und abschließend der Zusammenhang zwischen Intelligenz, (domänenspezifischem) Wissen (zu dem man die naturwissenschaftlichen Kenntnisse und Fähigkeiten zählen kann) und Leistung erläutert.

### 2.1 Das Modell „Individuelles Konzept über die Naturwissenschaften“

Das Individuelle Konzept über die Naturwissenschaften (IKoN) ist ein Modell, welches im Rahmen des durch die DFG geförderten Projekts „Charakterisierung individueller Interessen und Überzeugungen von Schülerinnen und Schülern als Grundlage einer naturwissenschaftlichen Talentförderung“ entwickelt wurde. Das Projekt wurde 2010 gestartet, an ihm beteiligt sind neben drei Doktorierenden ein Psychologe, die Wettbewerbsleitung der IJSO sowie mehrere Didaktikerinnen und Didaktiker des Fachbereichs Chemie. Im Folgenden wird der theoretische Hintergrund des Projekts, in das die vorliegende Dissertation eingebettet ist, vorgestellt, sowie seine Zielsetzungen beschrieben.

#### 2.1.1 Motivation und Zielsetzung

Die Grundlage für die Entwicklung des IKoN und somit den Start des DFG-Projekts war die Tatsache, dass in wissensgestützten Wirtschaftssystemen ein hoher Bedarf an naturwissenschaftlichen Fachkräften besteht, der oft nicht gedeckt werden kann (Osborne & Dillon, 2008). Zwar wird der naturwissenschaftlichen Bildung in Europa prinzipiell ein großer Wert beigemessen (z. B. Europäische Kommission, 2004, 2007; Osborne & Dillon, 2008), jedoch existieren in Deutschland kaum Konzepte für eine gezielte Förderung naturwissenschaftlicher Talente. Hinzu kommt, dass die Einstellungen von Schülerinnen und Schülern zu den Naturwissenschaften sowie die Vorstellungen über Naturwissenschaften, Tätigkeiten und Berufsfelder im Rahmen der Naturwissenschaften sowie über Naturwissenschaftlerinnen bzw. Naturwis-

senschaftler tendenziell negativ, wenig differenziert und oftmals stereotyp ausgeprägt sind (z. B. Lederman, 1992; Höttecke, Höhle & Kircher, 2004; Osborne & Dillon, 2008; Subotnik, Olszewski-Kubilius & Worrell, 2011; Zilker, Kauertz & Fischer, 2009). Konsequenterweise sollten daher die Förderung junger naturwissenschaftlicher Talente und ihre Ermutigung, eine berufliche Laufbahn im Bereich der Naturwissenschaften zu verfolgen, wesentliche Ziele naturwissenschaftlicher Bildung sein. Die Auswahl von Schülerinnen und Schülern für eine bestimmte (naturwissenschaftliche) Fördermaßnahme erfolgt jedoch bislang meist auf Basis von schulischen Noten, erzielten Testwerten und Empfehlungen durch Lehrkräfte (Subotnik et al., 2009), so dass an dieser Stelle weiterer Forschungsbedarf zu Charakteristika von naturwissenschaftlich interessierten und talentierten Schülerinnen und Schülern und von naturwissenschaftlichen Fördermaßnahmen besteht.

Das langfristige Ziel des Projekts ist daher, Grundlagen für eine forschungsbasierte, systematische und zielgerichtete Konzeption zur nachhaltigen Förderung naturwissenschaftlich interessierter Schülerinnen und Schüler zu entwickeln, die auf einer differenzierten und umfassenden Charakterisierung von Schülerinnen und Schülern basiert. Erkenntnisse, die hieraus gewonnen werden können, sollen zudem zur differenzierten Identifizierung naturwissenschaftlicher Talente genutzt werden. Des Weiteren sollen langfristig auf Basis der erlangten Erkenntnisse Merkmale und Anforderungen naturwissenschaftlicher Fördermaßnahmen beschrieben und zielgerichtet konzipiert werden.

### 2.1.2 Theoretischer Hintergrund zum Modell IKoN

Das Vorhaben, naturwissenschaftlich interessierte bzw. talentierte Schülerinnen und Schüler zu charakterisieren um dies als Basis für eine systematische und zielgerichtete Konzeption zur nachhaltigen Förderung von naturwissenschaftlichen Talenten nutzen zu können, verlangt zunächst eine Eingrenzung des Begriffes „Talent“. Viele bestehende Konzepte definieren Talente ausschließlich über ihre gemessene Intelligenz oder ihre spezifische Leistung (z. B. C. Fischer, Mönks & Grindel, 2004; Renzulli, 2005; D. H. Rost, 2009). Im Rahmen des Projekts und des Modells IKoN werden im Gegensatz hierzu neben der Intelligenz und weiteren (domänenspezifischen) Leistungsmerkmalen zusätzlich affektive Merkmale berücksichtigt (vgl. Ackerman & Heggestad, 1997; C. Fischer, 2006; F. Gagné, 2005; Heller, Perleth & Tock, 2005). Genutzt werden hier Persönlichkeits- und Umweltfaktoren wie Überzeugungen, Interessen und das Fähigkeitsselbstkonzept, um Talente in den Naturwissenschaften möglichst umfassend zu beschreiben. Neben der auf Leistungsmerkmalen ausgerichteten Definition des Talentbegriffs ist ein weiterer Nachteil existierender Modelle, dass die Naturwissenschaften kaum als ein eigenständiger Bereich betrachtet werden. Das IKoN strebt im Gegensatz hierzu an, Personenmerkmale und Umweltfaktoren differenziert und domänenspezifisch zu betrachten, d. h. es werden spezifisch naturwissenschaftliche Überzeugungen, Interessen, usw. erfasst. Das Modell des IKoN, das im Wesentlichen ausgehend von Erkenntnissen aus Arbeiten von Daniels (2008), Friege (2002), Gräber (1992), Höttecke et al. (2004), Klieme, Lucius und Zabel (1999), Köller, Trautwein, Lüdtke und Baumert (2006) sowie Lederman (1992) entwickelt wurde, ist in Abbildung 2.1 dargestellt.

Ausgehend von den genannten Modellen und Konzepten wurde das IKoN als ganzheitlicher Ansatz zur Charakterisierung und Identifizierung junger Talente in den Naturwissenschaften entwickelt. Es berücksichtigt in Anlehnung an wissenschaftliche Erkenntnisse aus der For-

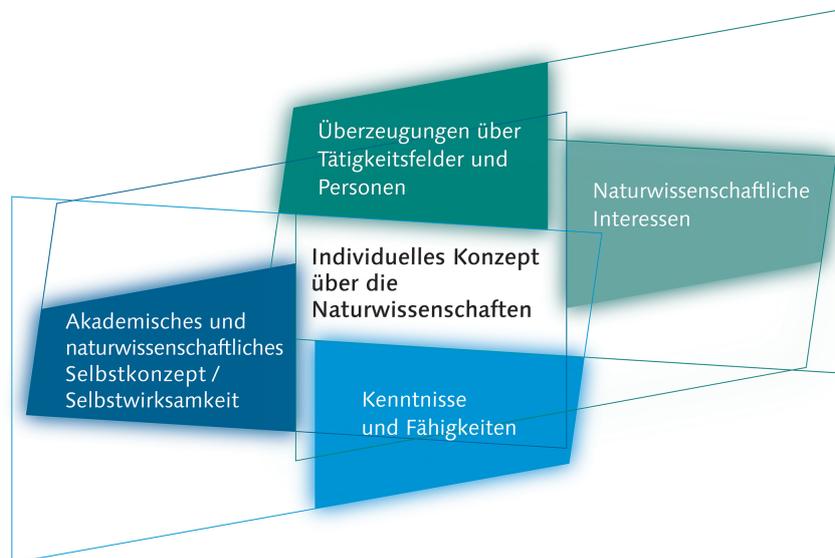


Abbildung 2.1: **Das Modell „Individuelles Konzept über die Naturwissenschaften“**

schung zu hochbegabten und hochleistenden Schülerinnen und Schülern die Überzeugungen von Schülerinnen und Schülern über Tätigkeitsfelder und Personen in den Naturwissenschaften, ihre naturwissenschaftlichen Interessen, Kenntnisse und Fähigkeiten sowie das Fähigkeitsselbstkonzept der Schülerinnen und Schüler. Das IKoN basiert auf dem RIASEC-Modell (J. L. Holland & Gottfredson, 1992; J. L. Holland, 1997), das ursprünglich zur Beschreibung und Erfassung beruflicher Interessen konzipiert wurde. Das RIASEC-Modell nimmt sechs Dimensionen beruflicher Interessen an: *Realistic* (R), *Investigative* (I), *Artistic* (A), *Social* (S), *Enterprising* (E) und *Conventional* (C). Diese Dimensionen können zudem mit bestimmten Berufen assoziiert werden, so wird die Dimension *Realistic* klassischerweise mit handwerklichen Berufen verknüpft, während die Dimension *Social* mit z. B. medizinischen Berufen verbunden wird.

Zunächst wurden Instrumente zur Erfassung affektiver Faktoren der Schülerinnen und Schüler entwickelt, die die Überzeugungen über Tätigkeitsfelder und Personen in den Naturwissenschaften (Wentorf, Höffler & Parchmann, 2015) sowie ihre naturwissenschaftlichen Interessen (Dierks, Höffler & Parchmann, 2014) erfassen sollten. Für das Instrument zur Erfassung von Überzeugungen über naturwissenschaftliche Tätigkeitsfelder und Personen wurde das RIASEC-Modell adaptiert und um die Dimension *Networking* (N) zum RIASEC+N-Modell ergänzt. Dabei wurde angenommen, dass das Tätigkeitsspektrum und die persönlichen Merkmale von Naturwissenschaftlerinnen und Naturwissenschaftlern auf alle Dimensionen des RIASEC-Modells übertragen werden können. Hiervon ausgehend wurde das Konstrukt *Nature of Scientists* (NoSt) entwickelt, das sich darauf bezieht, welche Vorstellungen Schülerinnen und Schülern über erforderliche Eigenschaften und regelmäßig ausgeübte Tätigkeiten von Naturwissenschaftlerinnen und Naturwissenschaftlern haben (Wentorf et al., 2015). Für die Entwicklung eines Instruments zur Erfassung naturwissenschaftlicher Interessen wurden

die Dimensionen des RIASEC-Modells auf verschiedene Bereiche naturwissenschaftlicher Interessen übertragen, was eine differenzierte Analyse derselben erlaubt (Dierks et al., 2014). Zur Erfassung des akademischen sowie des naturwissenschaftlichen Selbstkonzepts wurden standardisierte Erhebungsinstrumente verwendet (Daniels, 2008; Köller, Daniels, Schnabel & Baumert, 2000; Schöne, Dickhäuser, Spinath & Stiensmeier-Pelster, 2002). Die Erprobung der Erhebungsinstrumente fand mit Teilnehmenden der IJSO statt, da unter den Schülerinnen und Schülern, die erfolgreich an diesem Wettbewerb teilnehmen, naturwissenschaftlich in der Breite interessierte und talentierte Schülerinnen und Schüler vermutet wurden. Im Vergleich zu Teilnehmenden der IJSO wurden nichtteilnehmende Schülerinnen und Schüler, teilweise aus den Klassen bzw. Schulen der IJSO-Teilnehmenden, untersucht. Für den weiteren Verlauf des Projekts war vorgesehen, Teilnehmende unterschiedlicher naturwissenschaftlicher Schülerwettbewerbe im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern zu befragen.

Arbeiten von Ackerman und Heggstad (1997), F. Gagné (2005) oder Heller et al. (2005) betonen das Zusammenspiel affektiver und kognitiver Faktoren bei der Beschreibung von Talenten. Die Erfassung kognitiver Faktoren als Beitrag zur ganzheitlichen Charakterisierung von naturwissenschaftlich interessierten und talentierten Schülerinnen und Schülern im Rahmen des IKoN stellt den Hauptfokus der vorliegenden Arbeit dar. Eine Definition der naturwissenschaftlichen Kenntnisse und Fähigkeiten wird in Abschnitt 2.3 gegeben.

## 2.2 Naturwissenschaftliche Schülerwettbewerbe

Für die Studien, die im Rahmen des IKoN entstanden bzw. entstehen sollten, wurde vorgesehen, Teilnehmende (unterschiedlicher) naturwissenschaftlicher Schülerwettbewerbe im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern zu befragen. Dies sollte erfolgen, um neben spezifischen Charakteristika von Wettbewerbsteilnehmenden potentielle Unterschiede zwischen den einzelnen Gruppen im Hinblick auf die Anforderungen des jeweiligen Wettbewerbs (welche unterschiedlichen Charakteristika weisen Teilnehmende unterschiedlicher Wettbewerbe im Vergleich zu Nichtteilnehmenden auf) und die Passung als individuelle Fördermaßnahme (welcher Wettbewerb fördert bzw. spricht welche Schülerinnen und Schüler an) aufzudecken. Im Folgenden wird daher mit einem kurzen Überblick über Schülerwettbewerbe im Allgemeinen und ihre Funktion im Bildungswesen begonnen, bevor näher auf naturwissenschaftliche Schülerwettbewerbe eingegangen wird. Im Anschluss wird der Forschungsstand zu naturwissenschaftlichen Schülerwettbewerben und ihren Teilnehmenden zusammengefasst.

### 2.2.1 Schülerwettbewerbe und ihre Funktion im Bildungswesen

Wettbewerbe im schulischen Kontext stehen in einer langen Tradition. Die in den letzten Jahren stark zugenommene Relevanz von Schülerwettbewerben lässt sich durch einen kurzen Blick in die Wettbewerbsdatenbank des Deutschen Bildungsservers bestätigen, die für Schülerinnen und Schüler als Adressatengruppe und das Stichwort „Wettbewerb“ 782 Ergebnisse aufführt (Deutscher Bildungsserver, 2016, Zugriffsdatum 12.08.2016). Hierbei ist anzumerken, dass die Wettbewerbsdatenbank keinen Anspruch auf Vollständigkeit erhebt, so dass davon ausgegangen werden kann, dass die tatsächliche Anzahl an angebotenen Schülerwettbewerben sogar

höher liegt.

Die Schülerwettbewerbe übernehmen im schulischen Kontext verschiedene Funktionen. Sie dienen der Begabungsförderung sowie der Förderung besonderer Lernleistungen (Beutel & Tetzlaff, 2007), wodurch sie den schulischen Unterricht auf sinnvolle Weise ergänzen können (Messner, 2007). Des Weiteren können Schülerwettbewerbe einen wertvollen Beitrag zur Schul- und Unterrichtsentwicklung leisten (Racherbäumer & Boltz, 2012). Neben dem eigentlichen Wettbewerb bieten viele Schülerwettbewerbe ergänzende Programme wie Sommerakademien oder Jugendforen an (Tacke, 2007). Strunck (2011) nennt darüber hinaus auch die Möglichkeit der Professionalisierung von Lehrkräften durch Wettbewerbe, u. a. durch im Rahmen des Wettbewerbs angebotene Fortbildungen. Gezielte Untersuchungen hierzu liegen bislang allerdings nicht vor (Strunck, 2011).

Man unterscheidet vier verschiedene Wettbewerbsformate im Bildungssystem (vgl. Racherbäumer & Boltz, 2012; Strunck, 2011):

- Leistungswettbewerbe
- Auszeichnungswettbewerbe
- Best-Practice-Wettbewerbe
- Förderwettbewerbe

Sie unterscheiden sich im Wesentlichen in den zu erbringenden Leistungen, ihren Zielen und Funktionen. Für eine ausführliche und überblickende Darstellung der verschiedenen Wettbewerbsformate sei an dieser Stelle auf Racherbäumer und Boltz (2012) oder Strunck (2011) verwiesen.

Insbesondere die Leistungswettbewerbe, wie im naturwissenschaftlichen Bereich beispielsweise die ScienceOlympiaden oder Jugend forscht bzw. Schüler experimentieren oder im sprachlichen Bereich der Bundeswettbewerb Fremdsprachen, haben sich als talentfördernde Maßnahmen bewährt (Goldstein & Wagner, 1993; Wagner & Neber, 2007a, 2007b). Leistungswettbewerbe adressieren nach Racherbäumer und Boltz (2012) sowie Strunck (2011) Schülerinnen und Schüler, die im Rahmen des Wettbewerbs eine besondere Leistung erbringen müssen. Übliche Ziele der Leistungswettbewerbe sind die (individuelle) Förderung und Forderung besonderer Leistungen, die Förderung und Forderung innovativer Ideen sowie die Stärkung des Interesses und der Motivation (Racherbäumer & Boltz, 2012; Strunck, 2011). Wagner und Neber (2007b) unterscheiden darüber hinaus zwischen drei Typen von Leistungswettbewerben:

- „Ausführen und Darbieten“ (z. B. Jugend musiziert, Vorlesewettbewerb)
- „Entdecken und Herausfinden“ (z. B. die Internationale ChemieOlympiade (IChO) oder der Bundeswettbewerb Informatik)
- „Erfinden und Konstruieren“ (z. B. Jugend forscht bzw. Schüler experimentieren oder der BundesUmweltWettbewerb, BUW)

Gemeinsam ist den Wettbewerben der offene, freiwillige und kostenlose Zugang (Wagner & Neber, 2007b). Die Wettbewerbe lassen sich zudem nach ihren Zielgruppen (Primarstufe, Sekundarstufe 1, Sekundarstufe 2) einteilen. An die Primarstufe richtet sich bspw. der Schreib-

und Malwettbewerb der Deutschen Post oder der ZEIT LEO Weltretter Wettbewerb. Vorwiegend an die Sekundarstufe 1 richten sich z. B. im naturwissenschaftlichen Bereich die IJSO oder im sprachlichen Bereich der Vorlesewettbewerb, hauptsächlich an der Sekundarstufe 2 orientieren sich z. B. alle Senior-Olympiaden (z. B. IChO) im naturwissenschaftlichen Bereich oder der Geschichtswettbewerb des Bundespräsidenten im politisch-gesellschaftlichen Bereich. Die Aufgabenstellungen gehen in der Regel über den schulischen Horizont hinaus (Wagner & Neber, 2007b). Eine Einteilung lässt sich zudem in aufgabenorientierte (bspw. die IJSO, Dechemax Schülerwettbewerb) und projektorientierte Wettbewerbe (Jugend forscht bzw. Schüler experimentieren, Jugend präsentiert) vornehmen. Die Beurteilung der Leistungen erfolgt je nach Aufgabenstellung über eine Musterlösung bzw. einen Lösungsschlüssel (aufgabenorientierte Wettbewerbe) oder durch eine Jury anhand festgelegter Bewertungskriterien (z. B. Jugend forscht bzw. Schüler experimentieren).

### 2.2.2 Naturwissenschaftliche Schülerwettbewerbe - ein Überblick

Einen schnellen Überblick über die Vielfalt naturwissenschaftlicher Wettbewerbe liefert auch hier die Wettbewerbsdatenbank des Deutschen Bildungsservers, in der man für Schülerinnen und Schüler als Adressatengruppe mit dem Stichwort „Naturwissenschaft“ 50 Einträge, mit dem Stichwort „Biologie“ 26 Einträge, mit dem Stichwort „Chemie“ 30 Einträge und mit dem Stichwort „Physik“ 33 Einträge findet (Deutscher Bildungsserver, 2016, Zugriffsdatum 12.08.2016). Die Suchergebnisse enthalten jedoch Doppelungen, so dass die Ergebnisse nicht einfach aufsummiert werden können. Unter Ausschluss der Doppelungen ergibt sich ein Anteil von ca. einem Zehntel naturwissenschaftlicher Wettbewerbe in der Summe der von der Wettbewerbsdatenbank des Deutschen Bildungsservers aufgeführten Schülerwettbewerbe. Die Arbeitsgemeinschaft bundesweiter Schülerwettbewerbe, die bestimmte Qualitätsmerkmale eines Wettbewerbs erwartet, listet zwölf bundesweit ausgetragene und staatlich anerkannte Schülerwettbewerbe im Bereich der Naturwissenschaften auf, wobei hier auch Wettbewerbe aus den Bereichen Mathematik, Informatik und Entwicklungspolitik hinzu gezählt werden (Arbeitsgemeinschaft bundesweiter Schülerwettbewerbe, 2016).

Internationale Schülerwettbewerbe aus dem MINT-Bereich gehören zu den am längsten etablierten Wettbewerben und gehen meist auf Initiativen ehemals sozialistischer Länder zurück (Wagner & Neber, 2007b). Die Internationale Mathematik-Olympiade (IMO) wurde 1959 gegründet, die internationalen Schülerolympiaden finden seit 1967 (Internationale Physik-Olympiade (IPhO)), 1968 (IChO) und 1990 (Internationale BiologieOlympiade (IBO)) statt. Jugend forscht wurde 1965 ins Leben gerufen.

Naturwissenschaftliche Schülerwettbewerbe sind vorwiegend im Bereich der Leistungswettbewerbe anzusiedeln. Betrachtet man die Liste der Arbeitsgemeinschaft bundesweiter Schülerwettbewerbe findet man unter den Leistungswettbewerben hauptsächlich Wettbewerbe der Typen „Entdecken und Herausfinden“ sowie „Erfinden und Konstruieren“. Eine Einteilung kann zudem in Wettbewerbe mit definierten Aufgabenstellungen (aufgabenorientiert) und Wettbewerbe mit Projektcharakter (projektorientiert) vorgenommen werden. Des Weiteren sprechen die verschiedenen Wettbewerbe unterschiedliche Altersgruppen an. Eine Übersicht über die Wettbewerbe der Arbeitsgemeinschaft bundesweiter Schülerwettbewerbe und ausgewählte Charakteristika ist in Tabelle 2.1 dargestellt.

Tabelle 2.1: **Übersicht über die von der Arbeitsgemeinschaft bundesweiter Schülerwettbewerbe gelisteten mathematisch-naturwissenschaftlichen Wettbewerbe und ihre Charakteristika** AD = Ausführen und Darbieten, EH = Entdecken und Herausfinden, EK = Erfinden und Konstruieren, A = aufgabenorientiert, P = projektorientiert.

<i>Wettbewerb</i>	<i>Altersgruppe</i>	<i>Typ</i>	<i>Anlage</i>
Internationale BiologieOlympiade (Auswahlwettbewerb)	bis 20 Jahre	EH	A
Internationale ChemieOlympiade (Auswahlwettbewerb)	bis 20 Jahre	EH	A
Internationale PhysikOlympiade (Auswahlwettbewerb)	bis 20 Jahre	EH	A
BundesUmweltWettbewerb	10 bis 20 Jahre	EK	P
Bundesweiter Wettbewerb Physik	Klasse 5 bis 10	EH	A
Bundeswettbewerb Informatik	bis 21 Jahre	EH	A
Bundeswettbewerb Mathematik	Klasse 9 bis 13	EH	A
Europäische ScienceOlympiade	bis 17 Jahre	EH	A
Internationale JuniorScienceOlympiade (Auswahlwettbewerb)	bis 15 Jahre	EH	A
Jugend forscht bzw. Schüler experimentieren	Klasse 4 bis 21 Jahre	EK	P
Mathematik-Olympiaden	Klasse 3 bis 12	EH	A
Schulwettbewerb des Bundespräsidenten zur Entwicklungspolitik	Klasse 1 bis 13	AD	P

Neben den genannten Zielen von Leistungswettbewerben verfolgen die naturwissenschaftlichen Schülerwettbewerbe weitere Ziele. Neben der Spitzenförderung steht gerade in den ersten Runden der Wettbewerbe oftmals auch die Breitenförderung sowie die Förderung generell an MINT interessierter Schülerinnen und Schüler im Fokus (Peters, Giese & Müller-Balhorn, 2013). Auch das Heranführen an das wissenschaftliche Arbeiten und Forschen ist als Ziel naturwissenschaftlicher Schülerwettbewerbe zu nennen (Peters et al., 2013; Messner, 2007; Wagner & Neber, 2007b).

Die Wirkung spezifisch naturwissenschaftlicher Schülerwettbewerbe ist bislang nur wenig erforscht. Bekannt ist, dass auch naturwissenschaftliche Schülerwettbewerbe zur Entwicklung von Talenten beitragen (z.B. Campbell, 1996; Campbell, Wagner & Walberg, 2000; Heller, 2008). Untersucht wurden jedoch meist nur die Schülerolympiaden, zu Projektwettbewerben wie Jugend forscht bzw. Schüler experimentieren (oder im englischsprachigen Raum die *science fairs*) liegen nur relativ wenige Erkenntnisse vor (z. B. Abernathy & Vineyard, 2001; Dionne et al., 2012). Bekannt ist zudem, dass naturwissenschaftliche Schülerwettbewerbe positiv den Ausbildungsverlauf und die berufliche Laufbahn der Teilnehmenden beeinflussen (z. B. Campbell & Wu, 1996; Heilmann, 1999; Heller, 2008; Meißner, 2011; Subotnik, Duschl & Selmon, 1993). Auf die genannten Studien wird im nächsten Abschnitt noch genauer eingegangen.

Von den oben genannten Erkenntnissen abgesehen wurden außerdem Gründe für die Teilnahme an einem naturwissenschaftlichen Wettbewerb, Einflussfaktoren für eine erfolgreiche Wettbewerbsteilnahme bzw. die Qualifikation für eine höhere Runde oder Geschlechterun-

terschiede untersucht (z. B. Blankenburg, Höffler & Parchmann, 2015; Czerniak & Lumpe, 1996; Greenfield, 1995; Jones, 1991; Lengfelder & Heller, 2002; Urhahne, Ho, Parchmann & Nick, 2012). Affektive Faktoren wie das Interesse oder Zielorientierungen von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe wurden bislang kaum begleitend, zumal mit spezifischem Blick auf eine Charakterisierung der Teilnehmenden, in wissenschaftlichen Studien berücksichtigt. Ein Überblick über den Stand der Forschung zu naturwissenschaftlichen Schülerwettbewerben wird im Folgenden gegeben.

### 2.2.3 Charakteristika von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe

Teilnehmende naturwissenschaftlicher Schülerwettbewerbe wurden bislang meist in ihrer Entwicklung nach der Teilnahme an einem Wettbewerb mit Fokus auf ihren Ausbildungsverlauf und die berufliche Laufbahn betrachtet (z. B. Campbell, 1996; Heilmann, 1999; Heller, 2008; Meißner, 2011; Subotnik et al., 1993). Einige Studien betrachteten jedoch auch affektive Merkmale der Teilnehmenden wie das Interesse, das Fähigkeitsselbstkonzept, die Selbstwirksamkeitserwartung und weitere Konstrukte. Des Weiteren standen Unterschiede zwischen Mädchen und Jungen bei einigen Studien im Fokus. Eine der umfangreichsten Studien in diesem Feld ist die retrospektive Befragung ehemaliger Teilnehmender der internationalen Mathematik-, Chemie- und Physik-Olympiade und ihrer Eltern (Campbell & Wu, 1996). Zwischen 1995 und 2006 wurden mehr als 700 ehemalige Teilnehmende und ihre Eltern in den USA, Taiwan, China, Deutschland, Finnland und Südkorea befragt (Campbell, Tirri, Ruohotie & Walberg, 2004; Campbell & Walberg, 2010). Im Folgenden werden die Kenntnisse zu den verschiedenen Charakteristika von Wettbewerbsteilnehmenden zusammengefasst.

Zur besseren Verständlichkeit werden kurz die wesentlichen Konzepte definiert, die häufig Gegenstand von Studien im Kontext naturwissenschaftlicher Schülerwettbewerbe waren:

- Nach der Person-Gegenstands-Theorie beschreibt das **Interesse** eine zeitlich mehr oder weniger stabile, spezifische Beziehung zwischen einer Person und einem Gegenstand ihrer Umwelt (Krapp, 2002). Inhaltlich wird zwischen einer emotionalen Komponente (die positive Wahrnehmung des Gegenstandes und der auf ihn bezogenen Handlungen) und einer wertbezogenen Komponente (die hohe Wertschätzung des Gegenstandes) unterschieden (Krapp, 2002; Schiefele, 1996). Wesentliches Merkmal des Interesses ist zudem der intrinsische Charakter (Schiefele, Krapp, Wild & Winteler, 1993).
- **Attributionen** sind Ursachenzuschreibungen einer Person von Ereignissen und Sachverhalten, im Besonderen von Leistungssituationen (Urhahne, Kremer & Mayer, 2008). Man unterscheidet hierbei Fähigkeit, Anstrengung, Aufgabenschwierigkeit und Zufall zur Erklärung von Ereignissen in Leistungssituationen (Heider, 1958).
- Unter dem **Fähigkeitsselbstkonzept** versteht man fachspezifische oder allgemein akademische (schulische), durch Kompetenzerfahrungen in Schulfächern erworbene Fähigkeitseinschätzungen (J. Möller & Köller, 2004).
- Die **Selbstwirksamkeitserwartung** „wird definiert als die subjektive Gewissheit, neue oder schwierige Anforderungssituationen aufgrund eigener Kompetenz bewältigen zu können“ (Schwarzer & Jerusalem, 2002, S. 35).

- Die motivationalen Ausrichtungen einer Person in Lern- und Leistungssituationen bezeichnet man als ihre **Zielorientierungen** (Schöne, Dickhäuser, Spinath & Stiensmeier-Pelster, 2004). Dabei unterscheidet man vier Typen: Lernzielorientierte Personen werden durch das Streben nach Kompetenzsteigerung charakterisiert, während leistungszielorientierte Personen (auch: Annäherungs-Leistungszielorientierung) ihre Kompetenzen demonstrieren möchten. Von Vermeidungs-Leistungszielorientierung spricht man, wenn eine Person fehlende Kompetenzen verstecken möchte. Bestehen weder Lern- noch Leistungsanreize, spricht man von Arbeitsvermeidung (Schöne et al., 2002).

### **Familiärer Hintergrund, elterliche Einflüsse und schulische Faktoren**

Konform mit Erkenntnissen aus der Forschung zu hochbegabten und hochleistenden Jugendlichen (vgl. Olszewski, Kulieke & Buescher, 1987) fand Campbell (1996) in einer retrospektiven Befragung ehemaliger Olympioniken der *American Math Olympiad* ( $N = 72$ ), dass die Eltern von Olympiade-Teilnehmenden meist finanziell gut gestellt waren und eine überdurchschnittliche Bildung besaßen, viele von ihnen hatten einen Doktorgrad. Ähnliche Ergebnisse fanden Nokelainen, Tirri und Campbell (2004) in einer Stichprobe aus finnischen ( $n = 72$ ) und amerikanischen ( $n = 80$ ) Mathematik-Olympioniken. Zudem stellten sie fest, dass die Olympioniken eine lernförderliche Atmosphäre zu Hause als bedeutendsten Faktor in der Entwicklung ihres Talents ansahen (Nokelainen et al., 2004). Die von Campbell (1996) befragten Eltern übten bei einer großen Unterstützung ihrer Kinder wenig Leistungsdruck aus, zudem boten sie ihren Kindern viele geistige Ressourcen, kontrollierten sie dabei aber kaum. Weitere Studien, die die hier berichteten Erkenntnisse unterstützen, sind bspw. von Feng, Campbell und Verna (2001), Heller (2008), Nokelainen et al. (2004) oder W.-T. Wu und Chen (2001).

Viele Olympioniken erbrachten bereits vor der Teilnahme am Wettbewerb herausragende (schulische), zum Teil prämierte Leistungen (Campbell, 1996; Subotnik, Miserandino & Olszewski-Kubilius, 1996). Das Talent der Olympioniken wurde meist früh erkannt, zudem nahmen viele an Hochbegabten-Förderprogrammen teil (Campbell, 1996; Subotnik et al., 1996). In der Untersuchung von Verna und Feng (2002) stellte sich heraus, dass viele der Olympioniken ( $N = 93$ ) ihren Unterricht als langweilig und Schule in ihrer Entwicklung als eher hinderlich empfanden, einige, im Besonderen Mädchen, hatten sogar das Gefühl, ihre Intelligenz verstecken zu müssen.

### **Attribution von Erfolg und Misserfolg**

Als günstig wird angesehen, wenn Erfolg stabilen internalen Faktoren (z. B. die eigene Fähigkeit) und Misserfolg variablen externalen Faktoren (z. B. Pech) zugeschrieben wird (Weiner, 1986). Mädchen tendieren dazu, Erfolg externalen und Misserfolg internalen Faktoren zuzuschreiben, während dies bei Jungen genau umgekehrt verläuft (z. B. Wodzinski, 2007). Mehrere Studien zeigten für Wettbewerbsteilnehmende, dass zumeist der Erfolg der eigenen Anstrengung und Misserfolg einem Mangel an Anstrengungsbereitschaft zugeschrieben wurde (Nokelainen, Tirri, Campbell & Walberg, 2007). Die Frage, ob die eigenen Fähigkeiten oder die Anstrengung den Erfolg erklärte, wurde meist unterschiedlich beantwortet. Hochbegabte Jugendliche führten ihre Erfolge meist gleichermaßen auf ihre eigene Fähigkeit und die Anstrengung zurück (Dai, Moon & Feldhusen, 1998), ebenso wie finnische Olympioniken (Tirri & Nokelainen, 2010; Tirri, 2010). Deutsche Olympiade-Teilnehmende bewerteten die eigene

Fähigkeit wichtiger als die Anstrengung (Heller, 2008; Campbell et al., 2004). Andere Studien zeigten wiederum, dass amerikanische, chinesische und taiwanesischen Olympioniken ihren Erfolg eher auf die Anstrengung als auf ihre Fähigkeiten zurückführten (Campbell, 1996; Feng et al., 2001; W.-T. Wu & Chen, 2001). Unterschiede zwischen den Geschlechtern wurden in den genannten Studien nicht untersucht.

### **Interesse an naturwissenschaftlichen Fächern, Fähigkeitsselbstkonzept und Selbstwirksamkeitserwartung**

Diverse Studien konnten einen Zusammenhang zwischen Interesse und Fähigkeitsselbstkonzept, Interesse und Selbstwirksamkeitserwartung und dem Einfluss von Interesse, Fähigkeitsselbstkonzept und Selbstwirksamkeitserwartung auf die (schulische) Leistung zeigen, wobei die (schulische) Leistung i. d. R. positiv von den genannten Faktoren beeinflusst wurde (z. B. Köller, Baumert & Schnabel, 2001; Köller et al., 2006; Krapp, Schiefele & Schreyer, 1993; Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005; Marsh & Martin, 2011; J. Rost, 2004). Auch der Einfluss der individuellen Leistung auf Interesse, Fähigkeitsselbstkonzept und Selbstwirksamkeitserwartung wurde in den genannten Studien nachgewiesen. Mit diesen Befunden einhergehend wurden in mehreren Studien Unterschiede in den genannten Merkmalen zwischen Hochleistenden und nicht hochleistenden sowie hochbegabten und nicht hochbegabten Schülerinnen und Schülern gefunden, die i. d. R. zugunsten der Hochleistenden oder Hochbegabten ausfielen (z. B. Heller, 2008; Köller et al., 2006; Pajares, 1996; D. H. Rost, 2009; Vock, Köller & Nagy, 2013).

Nur wenige Studien untersuchten bislang das (spezifische) Interesse, das Fähigkeitsselbstkonzept oder die Selbstwirksamkeit von an (naturwissenschaftlichen) Wettbewerben teilnehmenden Schülerinnen und Schülern. Mund (2007) fand beim Vergleich von Teilnehmenden an Jugend musiziert ( $n = 93$ ), Teilnehmenden an Jugend forscht ( $n = 144$ ) und Nichtteilnehmenden ( $n = 464$ ) eine größere Anzahl an Interessen und Interessensbereichen für Teilnehmende an Jugend forscht im Vergleich zu den anderen beiden Gruppen. Das spezifisch naturwissenschaftliche Interesse nach dem RIASEC+N-Modell (siehe Absatz 2.1.2) von an Wettbewerben teilnehmenden Schülerinnen und Schülern, hier Teilnehmenden der IJSO ( $n = 130$ ), wurde von Dierks et al. (2014) im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern ( $n = 305$ ) betrachtet. Zwar fanden sich hier keine wesentlichen Unterschiede in den Interessensprofilen der Schülerinnen und Schüler, jedoch erreichten Teilnehmende der IJSO signifikant höhere Werte und wiesen dementsprechend ein höheres naturwissenschaftliches Interesse auf als Nichtteilnehmende.

Im Bereich des Fähigkeitsselbstkonzepts untersuchten Höffler, Bonin und Parchmann (2016) IJSO-Teilnehmende und Teilnehmende des Wettbewerbs Jugend trainiert für Olympia und verglichen sie mit einer Gruppe nichtteilnehmender Schülerinnen und Schüler ( $N = 574$ ). Die Ergebnisse zeigten, dass die Teilnehmenden der IJSO ein signifikant höheres schulisches Selbstkonzept - sozial - (Itembeispiel „Ich bin intelligenter als meine Mitschüler(innen)“, vgl. Schöne et al., 2002) als die beiden anderen Gruppen besaßen. Unter Berücksichtigung der letzten Zeugnisnoten der Schülerinnen und Schüler als Kovariate verschwanden die Effekte jedoch. Zwischen Teilnehmenden des Wettbewerbs Jugend trainiert für Olympia und der Vergleichsgruppe wurden keine signifikanten Unterschiede gefunden. Auch im Bereich des schulischen Selbstkonzepts - individuell - (Itembeispiel „Ich kann in der Schule mehr als früher“, vgl. Schöne

et al., 2002) wurden keine signifikanten Unterschiede gefunden, erst unter Berücksichtigung der letzten Zeugnisnoten der Schülerinnen und Schüler als Kovariate fanden sich signifikante Effekte. So wiesen IJSO-Teilnehmende ein signifikant niedrigeres schulisches Selbstkonzept - individuell - auf als Teilnehmende des Wettbewerbs Jugend trainiert für Olympia und die Vergleichsgruppe. Der Vergleich zwischen Mädchen und Jungen innerhalb der einzelnen Gruppen zeigte ein signifikant niedrigeres Selbstkonzept der IJSO-Teilnehmerinnen gegenüber den IJSO-Teilnehmern.

Forrester (2010) untersuchte u. a. den Zusammenhang zwischen der Teilnahme an *competitive science experiences* wie z. B. Schülerolympiaden oder *science fairs* sowie der Selbstwirksamkeitserwartung in den Naturwissenschaften und der Wahl eines College-Hauptfaches im naturwissenschaftlich-technischem Bereich ( $N = 1488$ ). Die Ergebnisse zeigten eine höhere Selbstwirksamkeitserwartung in den Naturwissenschaften für Studierende, die an einer *competitive science experience* teilgenommen hatten, als für Studierende, die nicht teilgenommen hatten. Einschränkend anzumerken ist hier allerdings, dass es sich bei der Befragung um eine retrospektive Studie handelte (die Teilnahme an der *competitive science experience* fand vor dem College statt).

### Schülervorstellungen über die Naturwissenschaften und Forschende

Die Vorstellungen, die Schülerinnen und Schüler über Naturwissenschaften und Forschende haben, sind äußerst divers und kaum angemessen (Höttecke, 2004). Das Bild von Naturwissenschaftlerinnen bzw. Naturwissenschaftlern ist dabei meist stereotyp (z. B. Chambers, 1983; Hannover & Kessels, 2002; Mason, Kahle & Gardner, 1991; Solomon, Scott & Duveen, 1996). Die epistemologischen Überzeugungen von Schülerinnen und Schülern sind ebenso uneindeutig und zum Teil sogar widersprüchlich (z. B. Larochelle & Désautels, 1991; Lederman, 2007; Meyling, 1997; Urhahne & Hopf, 2004). Ergebnisse einer Studie von Urhahne et al. (2008) zeigten, dass Schülerinnen und Schüler höherer Klassenstufen, mit einem höheren domänenspezifischen Selbstkonzept und besseren Lernleistungen ein höher entwickeltes Verständnis von der Natur der Naturwissenschaften aufwiesen. Differenzierte Schülervorstellungen über Tätigkeiten von Naturwissenschaftlerinnen und Naturwissenschaftlern anhand des RIASEC+N-Modells wurden von Wentorf et al. (2015) untersucht ( $N = 100$ ). Hier zeigte sich ein hauptsächlich von den (forschungsbezogenen) handwerklichen, investigativen und kreativen Komponenten geprägtes Bild vom Tätigkeitsspektrum einer Naturwissenschaftlerin bzw. eines Naturwissenschaftlers. Die Bedeutung kooperativer Aktivitäten der Dimension *Networking* wurden aus Sicht der Schülerinnen und Schüler als recht hoch eingeschätzt, wohingegen die Bedeutung unternehmerischer Tätigkeiten der Dimension *Enterprising* als eher gering eingeschätzt wurde.

Zu den Vorstellungen, die Teilnehmende naturwissenschaftlicher Schülerwettbewerbe über die Naturwissenschaften und Forschende haben, führten Welch und Huffman (2011) eine Studie in einem Robotik-Wettbewerb durch und befragten Teilnehmende ( $n = 58$ ) sowie Nichtteilnehmende ( $n = 41$ ) anhand des *test of science related attitudes*. Die Ergebnisse zeigten, dass die Teilnehmenden des Robotik-Wettbewerbs gegenüber den Nichtteilnehmenden in vier von sieben Kategorien eine positivere Einstellung gegenüber Naturwissenschaften und Forschenden aufwiesen. Finnerty (2013) fand ebenso in einer Studie mit  $N = 258$  teilnehmenden und nichtteilnehmenden Schülerinnen und Schülern einer schulinternen *science fair* sowohl im Prä- als auch im Post-Test signifikante Unterschiede in den Einstellungen gegenüber Natur-

wissenschaften. Paul und Groß (2016) untersuchten 57 Teilnehmende des Wettbewerbs Jugend forscht hinsichtlich der vorhandenen Vorstellungen zu *Nature of Science* (NoS) und inwieweit bzw. wodurch sich diese im Verlaufe des Wettbewerbs ändern. Die qualitative Studie konnte zeigen, dass Teilnehmende in den NoS-Kategorien Kreativität, Methoden, Experimentieren und Vernetzung Lernfortschritte im Verlauf des Wettbewerbs aufwiesen. Als Ursachen wurde die Wettbewerbsstruktur (intensiver Austausch am Wettbewerbstag und Betreuung durch begleitende Lehrkräfte, was zur Reflexion der eigenen Arbeit führt) sowie das projektorientierte praktische Arbeiten identifiziert (Paul & Groß, 2016, S. 68-69).

### **Zielorientierungen**

Mehrfach konnte ein Zusammenhang zwischen den Zielorientierungen einer Person und ihrem leistungsthematischen Verhalten hergestellt werden (z. B. Elliot, 2005; Köller, 1998; Köller et al., 2006; Sparfeldt, Buch, Wirthwein & Rost, 2007), weshalb die Zielorientierungen auch im Zusammenhang mit Schülerwettbewerben von Interesse sind. Forschungsergebnisse haben gezeigt, dass Lernziele meist mit positiven und Leistungsziele eher mit neutralen oder negativen Effekten in Verbindung stehen (z. B. Bell & Kozlowski, 2002; Elliott & Dweck, 1988; Elliot & Harackiewicz, 1994; Fisher & Ford, 1998; Greene & Miller, 1996). Steinmayr, Bipp und Spinath (2011) konnten bspw. zeigen, dass Lernziele und Annäherungsleistungsziele meist positiv mit der schulischen Leistung korrelierten. Vermeidungslernziele und Arbeitsvermeidung korrelierten dagegen meist negativ mit der schulischen Leistung. Bezogen auf Unterschiede zwischen Hochbegabten oder Hochleistenden sowie Schülerinnen und Schülern fanden sich zu meist Unterschiede zugunsten der Hochbegabten bzw. Hochleistenden (z. B. Bell & Kozlowski, 2002).

Die Zielorientierungen von Teilnehmenden (naturwissenschaftlicher) Schülerwettbewerbe sind bislang kaum untersucht worden. Höffler et al. (2016) untersuchten neben dem Fähigkeitsselbstkonzept die Zielorientierungen von Teilnehmenden der Schülerwettbewerbe Jugend trainiert für Olympia und IJSO. Ergebnisse zeigten signifikant höhere Werte für IJSO-Teilnehmende gegenüber den Teilnehmenden an Jugend trainiert für Olympia und Nichtteilnehmenden in den Lernzielen, die allerdings zwischen IJSO-Teilnehmenden und Teilnehmenden an Jugend trainiert für Olympia verschwanden, sobald die letzten Zeugnisnoten der Schülerinnen und Schüler berücksichtigt wurden. Keine signifikanten Unterschiede fanden sich in den Annäherungs-Leistungszielen. In den Vermeidungs-Leistungszielen fanden sich signifikante Unterschiede zwischen IJSO-Teilnehmenden und Nichtteilnehmenden mit einer höheren Vermeidungs-Leistungszielorientierung für Nichtteilnehmende. Auch für die Arbeitsvermeidung fanden sich signifikante Unterschiede: Eine höhere Arbeitsvermeidung wurde für Nichtteilnehmende und Teilnehmende an Jugend trainiert für Olympia gefunden (Höffler et al., 2016).

### **Fachspezifische kognitive Fähigkeiten**

In Studien mit (ehemaligen) Wettbewerbsteilnehmenden wurden bislang kaum fachspezifische kognitive Fähigkeiten untersucht. In Studien aus dem Bereich der Expertiseforschung, im Speziellen Problemlöseleistungen in Physik, waren Wettbewerbsteilnehmende zwar oftmals Teilnehmende der Studien, jedoch wurden sie nicht als eigenständige Gruppe betrachtet (z. B. Friege, 2003; Friege & Lind, 2006). Als Maß für fachspezifische kognitive Fähigkeiten von

Wettbewerbsteilnehmenden wurden meist die Zeugnisnoten der Wettbewerbsteilnehmenden erhoben, wobei sich meist gute bis sehr gute Noten fanden (z. B. Campbell, 1996; Lengfelder & Heller, 2002; Urhahne et al., 2012). Stang, Urhahne, Nick und Parchmann (2014) fanden in einer Stichprobe mit  $N = 77$  Wettbewerbsteilnehmenden zudem eine signifikante Korrelation zwischen den mathematisch-naturwissenschaftlichen Schulnoten sowie der Qualifikation für die vierte Runde der IChO (allerdings wurden die mathematisch-naturwissenschaftlichen Schulnoten nicht signifikant prädiktiv für das Erreichen der vierten Runde). Neben den Schulnoten erfassten einige Studien auch die allgemeinen kognitiven Fähigkeiten der an Wettbewerben teilnehmenden Schülerinnen und Schüler über standardisierte Testverfahren wie z. B. den Kognitiven Fähigkeitstest (KFT, Heller, Mönks, Subotnik & Sternberg, 2000; Studien, die die allgemeinen kognitiven Fähigkeiten der Wettbewerbsteilnehmenden erfassten, sind z. B. Mund, 2007; Urhahne et al., 2012). So fanden Urhahne et al. (2012) mit  $N = 52$  Wettbewerbsteilnehmenden einen höheren nonverbalen Intelligenzquotienten für Teilnehmende, die sich für die vierte Runde der IChO qualifiziert hatten, im Vergleich zu nicht für die vierte Runde qualifizierten Teilnehmenden. Dieses Ergebnis konnten Stang et al. (2014) jedoch nicht replizieren. Mund (2007) fand beim Vergleich von Teilnehmenden des Wettbewerbs Jugend forscht ( $n = 144$ ) und Jugend musiziert ( $n = 93$ ) mit einer Vergleichsgruppe ( $n = 464$ ) keine höhere Intelligenz von Teilnehmenden der Wettbewerbe Jugend forscht und Jugend musiziert gegenüber der Vergleichsgruppe (gemessen über den Untertest „Matrizen“ des Intelligenz-Struktur-Tests 2000 R; Liepmann, Beauducel, Brocke & Amthauer, 2007). Im Bereich fachspezifischer Tests (hier: Klausuren der dritten Auswahlrunde der IChO) fanden Urhahne et al. (2012) im Vergleich zwischen weiblichen und männlichen Teilnehmenden, dass männliche Teilnehmende signifikant besser abschnitten als weibliche Teilnehmende.

### Unterschiede zwischen Mädchen und Jungen

Mögliche Unterschiede zwischen Mädchen und Jungen im Wettbewerbskontext waren Gegenstand einiger Studien. Auf die Naturwissenschaften im Allgemeinen bezogen fanden sich geschlechtsspezifische Unterschiede vor allem im Bereich des Interesses und des (fachspezifischen) Selbstkonzepts, wobei die Jungen meist ein höheres Interesse und ein höheres (fachspezifisches) Selbstkonzept aufwiesen (z. B. Holstermann & Bögeholz, 2007; Kessels & Hannover, 2004; Köller et al., 2000).

Im Bereich der naturwissenschaftlichen Schülerwettbewerbe untersuchten Adamson, Foster, Roark und Reed (1998)  $N = 268$  Teilnehmende einer *science fair* im Elementarbereich und konnten zeigen, dass Teilnehmerinnen vorwiegend in sozialen und biologischen Bereichen arbeiteten, wohingegen Teilnehmer eher in physikalischen Bereichen Projekte wählten. Während Adamson et al. (1998) Teilnehmende über zwei Jahre beobachteten, wertete Greenfield (1995) Daten der letzten 20 Jahre eines naturwissenschaftlichen Schülerwettbewerbs aus. Auch hier zeigte sich, dass Jungen eher als Mädchen Themen im physikalischen, geowissenschaftlichen, mathematischen und computerbasierten Bereichen wählten. Zudem fand Greenfield (1995), dass Mädchen eher Themen wählten, die eine Literaturrecherche statt einer experimentellen Arbeit erforderten. Ähnliche Ergebnisse bezogen auf die Themenwahl von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe sind auch den Studien von Jones (1991) und Sonnert, Sadler und Michaels (2013) zu entnehmen.

Was die Frage der Verteilung von Jungen und Mädchen in den naturwissenschaftlichen

Schülerwettbewerben betrifft, finden sich unterschiedliche Zahlen. Im Vergleich zwischen *science fair* und *Science Olympiad* fanden Abernathy und Vineyard (2001) mit einer Gesamtstichprobe von  $N = 943$ , dass signifikant mehr Mädchen an der *science fair* teilnahmen. Jones (1991) fand, dass mehr Jungen als Mädchen an naturwissenschaftlichen Schülerwettbewerben teilnahmen, mit Ausnahme einer *senior regional science fair* und einer *junior-level state science fair* ( $N = 1085$ ). Sonnert et al. (2013) fanden, dass im *Massachusetts State Science and Engineering Fair* im Jahre 2009 mehr Mädchen als Jungen, nämlich 62% ( $N = 321$ ), teilnahmen. Auch in einer von Greenfield (1995) untersuchten *state science fair* nahmen in den letzten 20 Jahren mehr Mädchen als Jungen teil, wenn auch weniger häufig. Speziell auf deutsche Teilnehmende bezogen fanden Lengfelder und Heller (2002) in ihrer Stichprobe ( $N = 235$ ), dass Mädchen in den deutschen Olympiade-Mannschaften mit einem Verhältnis von männlich zu weiblich von 28:1 stark unterrepräsentiert waren. Im Gegensatz hierzu fanden Verna und Feng (2002) bei amerikanischen Olympioniken ein Verhältnis von 6:1. Im Vergleich zu anderen Ländern war dies die niedrigste Rate überhaupt, obwohl in den Fähigkeiten und den erreichten Leistungen keine Unterschiede zwischen Jungen und Mädchen gefunden werden konnten (Lengfelder & Heller, 2002). Vielmehr wurden Unterschiede in affektiven Bereichen gefunden, zudem wurde das lehrergesteuerte Auswahlverfahren als mögliche Ursache für den geringen Anteil an Mädchen unter den Olympioniken benannt. Betrachtet man den Anteil von Teilnehmerinnen in den deutschen Olympiade-Mannschaften der ScienceOlympiaden in den letzten Jahren finden sich deutliche Unterschiede, weiterhin sind weibliche Teilnehmende in einigen naturwissenschaftlichen Wettbewerben stark unterrepräsentiert: In der IJSO (Internationale JuniorScienceOlympiade, 2016) betrug das Verhältnis von männlich zu weiblich von 2004 bis 2016 ca. 4:1, in der IBO (Internationale BiologieOlympiade, 2016) von 2005 bis 2015 ca. 3:2, in der IChO (Internationale ChemieOlympiade, 2016) von 2006 bis 2016 43:1 und in der IPhO (Internationale PhysikOlympiade, 2016) entsprach das Verhältnis von männlich zu weiblich von 2005 bis 2016 29:1.

Bei der Frage, was für die Entwicklung des eigenen Talents am bedeutendsten gewesen war, nannten Jungen eine lernförderliche Atmosphäre zu Hause sowie die Verfügbarkeit von Büchern, während die Mädchen vorwiegend Lehrkräfte und schulische Einflüsse nannten (Lengfelder & Heller, 2002). Auch in der qualitativen Studie von Tirri (2002) mit einer Stichprobe von  $N = 26$ , davon 13 Mädchen, bewerteten die Mädchen den Einfluss der Lehrkräfte auf die Entwicklung ihres Talents höher als die Jungen.

Unterschiede im elterlichen Einfluss sowie in den Personen- und Leistungsmerkmalen zwischen weiblichen und männlichen Wettbewerbsteilnehmenden konnten von Urhahne et al. (2012) aufgedeckt werden ( $N = 52$ , davon 14 weibliche Teilnehmende). Ergebnisse zeigten, dass männliche Teilnehmende der dritten Runde des deutschen Auswahlverfahrens der IChO ein förderlicheres Elternhaus besaßen als weibliche Teilnehmende. Unterschiede fanden Urhahne et al. (2012) auch im familiären Hintergrund: So hatten die Eltern von Teilnehmern mehr Kenntnisse in Chemie und bewerteten die Chemie und ihre Bedeutung höher als die Eltern von Teilnehmerinnen. Zudem besaßen männliche Teilnehmende ein höheres fachspezifisches Selbstkonzept, eine höhere Selbstwirksamkeitserwartung und eine geringere Versagensangst. Teilnehmer setzten sich selbst höhere Ziele im Wettbewerb als Teilnehmerinnen, des Weiteren schnitten Teilnehmerinnen in den Klausuren der dritten Runde des Auswahlverfahrens signifikant schlechter als Teilnehmer ab (Urhahne et al., 2012).

### Schulische und berufliche Laufbahn von Wettbewerbsteilnehmenden

Untersuchungen über die Entwicklung von Wettbewerbsteilnehmenden nach ihrer Teilnahme erfolgten meist mit Teilnehmenden nationaler und internationaler Aufgabenwettbewerbe wie die Schülerolympiaden oder die Bundeswettbewerbe Mathematik und Informatik. Wenige Studien berücksichtigen auch Teilnehmende von Projektwettbewerben. Campbell (1996) untersuchte Teilnehmende der amerikanischen Mathematik-Olympiade und ihre Eltern,  $N = 72$  bzw.  $N = 55$ , anhand einer Fragebogen-Studie und konnte zeigen, dass die meisten der Teilnehmenden nach ihrer schulischen Laufbahn einen akademischen Werdegang wählten. 55% der Befragten arbeiteten an einem oder hatten bereits einen Doktorgrad erreicht und konnten eine hohe Zahl an Publikationen vorweisen. Die Befunde konnten später von Campbell und Walberg (2010) anhand einer größeren Stichprobe,  $N = 345$ , und Teilnehmenden der amerikanischen Mathematik-, Chemie- und Physik-Olympiade repliziert werden. Heller (2008) konnte auch für deutsche Teilnehmende der Schülerolympiaden in Mathematik, Physik und Chemie,  $N = 235$ , ähnliche Ergebnisse zeigen. Zudem fanden Lengfelder und Heller (2002) in derselben Stichprobe, dass die Wettbewerbsteilnehmenden ein hohes schulisches Leistungsniveau aufwiesen und in ihrer Olympiadisziplin bessere Abiturnoten erreichten als in anderen Fächern. Später entschieden sich viele Teilnehmende für einen Beruf innerhalb ihrer „Olympiadedisziplin“. Heilmann (1999) untersuchte u. a. Bundessieger des Bundeswettbewerbs Mathematik ( $N = 184$ ) zwischen 1970 und 1994 und fand auch hier, dass viele ehemalige Bundessieger Mathematik studierten (rund 70%), universitäre Abschlüsse meist mit sehr guten Noten erzielten und zu 80% eine Promotion planten, promovierten oder bereits promoviert waren. Ungefähr die Hälfte der Bundessieger plante zu habilitieren oder war bereits habilitiert, damit einhergehend fand Heilmann (1999), dass über 60% der Bundessieger während ihrer ersten Anstellung im akademischen Bereich tätig war. Im Vergleich mit Endrunden-Teilnehmenden ( $N = 479$ ) des Bundeswettbewerbs Mathematik fanden sich zwar einige, aber sehr geringfügige Unterschiede, die sich vorwiegend in der Studienfachwahl und der Habilitationsrate zeigten, wobei die Bundessieger häufiger Mathematik studierten und häufiger eine Habilitation planten oder bereits abgeschlossen hatten (Heilmann, 1999). Meißner (2011, 2013) verglich Teilnehmende des Bundeswettbewerbs Informatik,  $n = 80$ , mit Teilnehmenden von Jugend forscht,  $n = 61$ , in der Disziplin Mathematik/ Informatik und fand auch hier, dass ehemalige Teilnehmende allgemein und in ihrer Olympiadisziplin sehr gute Noten erzielten, meist einen Studiengang in ihrer Olympiadisziplin aufnahmen, eine Promotion anstrebten, promovierten oder promoviert waren. Interview- und Fragebogenstudien zeigten darüber hinaus, dass ehemalige Teilnehmende und ihre Eltern ihren beruflichen Erfolg bzw. den ihrer Kinder auf die Wettbewerbsteilnahme zurückführten (z. B. Campbell, 1996; Sahin, Gulacar & Stuessy, 2015).

#### 2.2.4 Gründe für die Teilnahme an einem naturwissenschaftlichen Wettbewerb

Sowohl für Wettbewerbsleitungen als auch für unterstützende Lehrkräfte oder Eltern spielt die Frage, aus welchem Grund oder aus welchem Grund gerade nicht jemand an einem (naturwissenschaftlichen) Wettbewerb teilnimmt, eine entscheidende Rolle. Im Vorfeld einer *science fair* befragte Schülerinnen und Schüler ( $N = 303$ ) nannten als Teilnahme Gründe vorwiegend die Möglichkeit, etwas Neues zu lernen, die Aussicht, ihre Schulnote zu verbessern, die Chance etwas zu gewinnen, die Verbesserung ihrer Präsentationsfähigkeiten und das Kennenlernen neuer Menschen (Czeraniak & Lumpe, 1996). Einschränkend ist bei dieser Studie jedoch zu be-

achten, dass rund 81% der Befragten nicht freiwillig an der *science fair* teilnahmen. Abernathy und Vineyard (2001) untersuchten sowohl Teilnehmende der *state science fair* ( $N = 490$ ) als auch Teilnehmende der *state Science Olympiad* ( $N = 453$ ) in Utah. Von vielen Schülerinnen und Schülern wurde die Teilnahme an der *science fair* verlangt, oder die Leistung wurde ihnen in der Schule angerechnet. Letzteres betraf fast die Hälfte der Teilnehmenden der *state science fair*, 48%, in der Stichprobe der *high school*. In der Stichprobe der *junior high* waren es sogar knapp zwei Drittel der Teilnehmenden (*science fair*) bzw. 59% (*Science Olympiad*). In der *state science fair* wurden signifikant mehr Schülerinnen und Schüler von ihren Lehrkräften aufgefordert, teilzunehmen. Auf die Frage, was die Teilnehmenden der Wettbewerbe für sich aus dem Wettbewerb mitnahmen, nannten die Teilnehmenden beider Wettbewerbe den Spaß und das Lernen neuer Dinge. Unterschiede fanden sich darin, dass Teilnehmende der *science fair* darüber hinaus das Antreten gegen andere Schülerinnen und Schüler sowie das Erlernen wissenschaftlichen Arbeitens (*learning the scientific process*) nannten. Teilnehmende der *Science Olympiad* nannten hingegen das Arbeiten mit Freunden und das Gewinnen von Preisen (Abernathy & Vineyard, 2001). Dionne et al. (2012) befragten  $N = 36$  Teilnehmende der *Canada-Wide Science Fair* zu ihren Teilnahmegründen und identifizierten das Interesse am naturwissenschaftlichen Inhalt, die Selbstwirksamkeitserwartung, materielle Gewinne, soziale Aspekte und die Verbesserung von Lernstrategien als wichtigste Einflussfaktoren.

Welche Merkmale prädiktiv auf das Interesse an einer Wettbewerbsteilnahme wirken können, untersuchten Blankenburg et al. (2015) in einer Studie mit  $N = 474$  Schülerinnen und Schülern der Jahrgangsstufe 6, die zum größten Teil noch keine Wettbewerbs Erfahrung besaßen. Das Interesse an dem zu bearbeitenden Thema und die Möglichkeit in einem Team zu arbeiten und den Bearbeitungsprozess autonom zu steuern, stellten rein deskriptiv die am höchsten bewerteten Teilnahmegründe für die Schülerinnen und Schüler dar. Geschlechterunterschiede mit höheren Werten für die Jungen fanden sich in der Annäherungs-Leistungszielorientierung, in der Nützlichkeit der Teilnahme für die Schule sowie in der Aussicht auf das Gewinnen von Auszeichnungen und Preisen (Blankenburg et al., 2015). Als signifikante Prädiktoren für die Bereitschaft zur Teilnahme an naturwissenschaftlichen Schülerwettbewerben fanden sich das Interesse an naturwissenschaftlichen Fächern, die vorherige Teilnahme an einem naturwissenschaftlichen Wettbewerb, das Wettbewerbsselbstkonzept (die Einschätzung der Eignung, an einem Wettbewerb erfolgreich teilzunehmen zu können, Blankenburg et al., 2015) sowie der Grund teilzunehmen, um etwas Neues zu lernen (intrinsisches Kompetenzerleben). Eine nach Geschlecht getrennte Analyse ergab unterschiedliche Prädiktoren für Mädchen und Jungen: Während für die Mädchen das intrinsische Kompetenzerleben, eine vorherige Wettbewerbsteilnahme und ein hohes Interesse an naturwissenschaftlichen Fächern prädiktiv war, waren für die Jungen neben dem intrinsischen Kompetenzerleben die hohe Nützlichkeitseinschätzung der Naturwissenschaften, das Wettbewerbsselbstkonzept und die Vermeidungs-Leistungszielorientierung (hier: je höher die Vermeidungs-Leistungszielorientierung, desto geringer die Bereitschaft zur Teilnahme) prädiktiv für die Teilnahmebereitschaft an einem naturwissenschaftlichen Wettbewerb.

### 2.2.5 Einflussfaktoren für eine erfolgreiche Wettbewerbsteilnahme

Nur wenige Studien widmeten sich bislang der Frage, welche affektiven oder kognitiven Merkmale einer Person eine erfolgreiche Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb beeinflussen können. Czerniak (1996) befragte  $N = 142$  Schülerinnen und

Schüler, die an einem örtlichen *science day* teilnahmen, und konnte fünf Faktoren identifizieren, die eine hohe Bewertung durch die Wettbewerbsjury bedingten: die Unterstützung durch die Eltern, den elterlichen Druck, das fachspezifische Selbstkonzept, dass das Projekt Teil der Kursnote war sowie dass die Bewertung durch die Wettbewerbsjury die Kursnote bestimmte (Czerniak, 1996). Urhahne et al. (2012) untersuchten Teilnehmende der dritten (von vier) Runden des deutschen Auswahlverfahrens der IChO ( $N = 52$ ) auf Basis des Erwartungs-Wert-Modells der Leistungsmotivation (Eccles, 1983). Als einziger Prädiktor für die Qualifikation zur vierten Runde fand sich hier die vorherige Teilnahme am deutschen Auswahlverfahren der IChO (Urhahne et al., 2012). Bereits Abernathy und Vineyard (2001) hatten deskriptiv festgestellt, dass sowohl Teilnehmende der *state science fair* als auch Teilnehmende der *state Science Olympiad* in Utah wiederholt am Wettbewerb teilgenommen hatten. Weitere mögliche, in der Stichprobe von Urhahne et al. (2012) aber vermutlich aufgrund der Stichprobengröße nicht signifikante, Einflussfaktoren für eine erfolgreiche Wettbewerbsteilnahme waren zudem die Schulnoten in den Naturwissenschaften und der intrinsische Wert, den die Familien der Wettbewerbsteilnehmenden der Chemie beimaßen. Im Vergleich affektiver Merkmale von für die vierte Auswahlrunde der IChO qualifizierten und nicht qualifizierten Teilnehmenden fanden Urhahne et al. (2012), dass qualifizierte Teilnehmende signifikant niedrigere Versagensängste und eine höhere Erfolgserwartung hatten. Zudem maßen sie der Nützlichkeit der Chemie eine größere Bedeutung bei.

Auf der Studie von Urhahne et al. (2012) aufbauend untersuchten Stang et al. (2014) Teilnehmende der dritten Runde des deutschen Auswahlverfahrens der IBO ( $n = 43$ ) und der IChO ( $n = 44$ ). Signifikant positive Korrelationen zum Wettbewerbserfolg fanden sich in beiden Stichproben für die Anzahl erhaltener Auszeichnungen, die erlebte Langeweile in den Unterrichtsstunden, die Erfolgserwartung und die wahrgenommenen relativen Kosten (z. B. „Ich muss eine Menge aufgeben, um in Biologie/ Chemie gut zu sein“, nach Berger & Karabenick, 2011). Für Teilnehmende der IBO fanden Stang et al. (2014) zusätzlich signifikante Korrelationen zwischen Wettbewerbserfolg und vorheriger Teilnahme, während für Teilnehmende der IChO die kurz- und langfristigen Ziele, das fachliche Interesse, der Zielerreichungswert (z. B. „Es ist mir wichtig, eine Person zu sein, die gut in Biologie/ Chemie ist“, nach Berger & Karabenick, 2011) und die mathematisch-naturwissenschaftliche Note ebenso positiv mit dem Wettbewerbserfolg korrelierten. Als einzige Prädiktoren für das Erreichen der vierten Runde des jeweiligen deutschen Auswahlverfahrens ergaben sich für die IBO jedoch nur die erlebte Langeweile im Schulunterricht und für die IChO die Erfolgserwartung (Stang et al., 2014).

### 2.2.6 Zusammenfassung und Fazit

Die wesentlichen Erkenntnisse aus den oben dargestellten Studien werden im folgenden Abschnitt noch einmal zusammengetragen.

Faktoren, die eine Teilnahme am Wettbewerb beeinflussen, wurden in Studien von Czerniak und Lumpe (1996), Abernathy und Vineyard (2001) und Dionne et al. (2012) identifiziert. Schülerinnen und Schüler nannten als Teilnahmegründe die Aussicht auf das Erlernen und Verbessern neuer Inhalte und Fähigkeiten, die Möglichkeit der Verbesserung der Schulnote, die Aussicht auf (materielle) Gewinne, Spaß und soziale Aspekte (z. B. die Möglichkeit, neue Menschen kennenzulernen und gegen sie anzutreten). Des Weiteren beeinflussten das naturwissenschaftliche Interesse und die Selbstwirksamkeitserwartung eine (mögliche) Wettbewerbs-

teilnahme von Schülerinnen und Schülern. Als Prädiktoren für die Bereitschaft zur Teilnahme an naturwissenschaftlichen Wettbewerben wurde das Interesse an naturwissenschaftlichen Fächern, die vorherige Teilnahme an einem naturwissenschaftlichen Wettbewerb, das Wettbewerbsselbstkonzept und das intrinsische Kompetenzerleben identifiziert (Blankenburg et al., 2015). In der Studie von Blankenburg et al. (2015) fanden sich zudem unterschiedliche Prädiktoren für Jungen und Mädchen. Prädiktiv für die Bereitschaft zur Teilnahme an naturwissenschaftlichen Wettbewerben waren für Mädchen das intrinsische Kompetenzerleben, eine vorherige Wettbewerbsteilnahme und ein hohes Interesse an naturwissenschaftlichen Fächern. Für Jungen waren das intrinsische Kompetenzerleben, die hohe Nützlichkeits einschätzung der Naturwissenschaften, das Wettbewerbsselbstkonzept und Vermeidungs-Leistungszielorientierung prädiktiv für die Teilnahmebereitschaft.

Als Einflussfaktoren für eine hohe Bewertung durch die Wettbewerbsjury eines *science day* fanden Czerniak (1996) die Unterstützung durch die Eltern, den elterlichen Druck, das fachspezifische Selbstkonzept, dass das Projekt Teil der Kursnote war und dass die Bewertung durch die Wettbewerbsjury die Kursnote bestimmte. Urhahne et al. (2012) und Stang et al. (2014) fanden als Einflussfaktoren für eine erfolgreiche Wettbewerbsteilnahme, hier das Erreichen der vierten Runde der deutschen Auswahlverfahren der IChO bzw. IBO, die vorherige Wettbewerbsteilnahme, die erlebte Langeweile im Schulunterricht und die Erfolgserwartung.

Über die Teilnehmenden an naturwissenschaftlichen Schülerwettbewerben ist z. B. durch die Studien von Campbell (1996), Feng et al. (2001), Heller (2008), Nokelainen et al. (2004) und W.-T. Wu und Chen (2001) bekannt, dass sie meist aus finanziell und sozial gut gestellten Elternhäusern stammten. Ihre Eltern, die oftmals eine überdurchschnittliche Bildung besaßen, boten den Wettbewerbsteilnehmenden eine unterstützende und lernförderliche Atmosphäre ohne Leistungsdruck. Das Talent vieler Wettbewerbsteilnehmender wurde früh erkannt und zeigte sich schon vor ihrer Wettbewerbsteilnahme in guten bis sehr guten Noten und weiteren Leistungen, sie besuchten oftmals Hochbegabten-Förderprogramme. Schule empfanden sie häufig als langweilig und hinderlich in ihrer Entwicklung (Campbell, 1996; Subotnik et al., 1996; Verna & Feng, 2002). Ihren Erfolg führten sie auf ihre eigenen Fähigkeiten oder ihre Anstrengung zurück (Campbell, 1996; Campbell et al., 2004; Feng et al., 2001; Heller, 2008; Nokelainen, Tirri & Merenti-Valimaki, 2007; Tirri, 2010; Tirri & Nokelainen, 2010; W.-T. Wu & Chen, 2001).

Teilnehmende naturwissenschaftlicher Schülerwettbewerbe wiesen im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern eine große Anzahl an Interessen und Interessensbereichen (Mund, 2007) sowie ein hohes naturwissenschaftliches Interesse auf (Dierks et al., 2014). Höffler et al. (2016) zeigten, dass Teilnehmende der IJSO ein höheres schulisches Selbstkonzept - sozial - zeigten als Teilnehmende an Jugend trainiert für Olympia oder Nichtteilnehmende, einschränkend wirkte hier allerdings der Einfluss der Schulnoten. IJSO-Teilnehmerinnen hatten ein signifikant niedrigeres Selbstkonzept als IJSO-Teilnehmer. Die Selbstwirksamkeitserwartung von Studierenden, die an einer *competitive science experience* teilgenommen hatten, war höher als für Studierende, die nicht teilgenommen hatten (Forrester, 2010). Auch die Vorstellungen von Wettbewerbsteilnehmenden über Naturwissenschaften und Forschende waren positiver als die von Nichtteilnehmenden (Welch & Huffman, 2011; Finnerty, 2013).

Was die Zielorientierungen von Wettbewerbsteilnehmenden betrifft fanden Höffler et al. (2016), dass IJSO-Teilnehmende eine signifikant höhere Lernzielorientierung und eine niedrigere Vermeidungs-Leistungszielorientierung und Arbeitsvermeidung als Nichtteilnehmende besaßen.

Über die fach- bzw. domänenspezifischen Kenntnisse und kognitiven Fähigkeiten von Wettbewerbsteilnehmenden ist bekannt, dass die Wettbewerbsteilnehmenden meist gute bis sehr gute schulische Noten aufwiesen (Campbell, 1996; Lengfelder & Heller, 2002; Urhahne et al., 2012). Zu Unterschieden in der Intelligenz bestehen widersprüchliche Aussagen (Urhahne et al., 2012; Stang et al., 2014). Geschlechterunterschiede in der Testleistung fanden sich zugunsten der Jungen ebenso bei Urhahne et al. (2012).

Darüber hinaus fanden sich Geschlechterunterschiede in den Themen, die Schülerinnen und Schüler für ihre Teilnahme an einem naturwissenschaftlichen Schülerwettbewerb wählten (Adamson et al., 1998; Greenfield, 1995; Jones, 1991; Sonnert et al., 2013). Zur Verteilung von Mädchen und Jungen in den einzelnen Wettbewerben fanden sich unterschiedliche Ergebnisse, es nahmen mal mehr Mädchen und mal mehr Jungen an den untersuchten Wettbewerben teil (Greenfield, 1995; Lengfelder & Heller, 2002; Jones, 1991; Sonnert et al., 2013; Verna & Feng, 2002). Mädchen bewerteten den Einfluss von Lehrkräften auf die Entwicklung höher als die Jungen (Lengfelder & Heller, 2002; Tirri, 2002). Unterschiede in affektiven Merkmalen von Wettbewerbsteilnehmenden fanden Urhahne et al. (2012) zugunsten der Jungen im fachspezifischen Selbstkonzept, in der Selbstwirksamkeitserwartung und in der Versagensangst. Zudem setzten sich Teilnehmer selbst höhere Ziele im Wettbewerb und schnitten in den Klausuren der dritten Runde des Auswahlverfahrens signifikant besser ab als Teilnehmerinnen. Zudem hatten die Teilnehmer ein förderlicheres Elternhaus als Teilnehmerinnen.

Nach der Teilnahme am naturwissenschaftlichen Wettbewerb schlugen die ehemaligen Teilnehmenden meist eine erfolgreiche akademische Karriere ein und führten dies auch auf ihre Wettbewerbsteilnahme zurück (Campbell, 1996; Campbell & Walberg, 2010; Heilmann, 1999; Heller, 2008; Lengfelder & Heller, 2002; Meißner, 2011, 2013; Sahin et al., 2015).

Insgesamt ist festzustellen, dass viele der dargestellten Studien meist mit großem zeitlichen Abstand nach der Teilnahme am jeweiligen Wettbewerb stattfanden. Zwar wurden bislang viele unterschiedliche affektive Merkmale von Wettbewerbsteilnehmenden untersucht, jedoch liegen zu den jeweiligen Konstrukten nur wenige einzelne Studien, aufgrund ihrer qualitativen Anlage oftmals mit einer relativ geringen Stichprobengröße, vor. Auch mangelt es an Studien, die unterschiedliche Wettbewerbe betrachten bzw. vergleichen.

Fachspezifische sowie kognitive Fähigkeiten von Teilnehmenden naturwissenschaftlicher Wettbewerbe wurden bisher kaum untersucht, dabei deuten die widersprüchlichen Ergebnisse hinsichtlich der Intelligenzunterschiede von qualifizierten versus nicht qualifizierten Teilnehmenden von Urhahne et al. (2012) und Stang et al. (2014) darauf hin, dass im Bereich kognitiver Fähigkeiten nicht nur die allgemeine Intelligenz der Wettbewerbsteilnehmenden entscheidend für ihren Erfolg ist. Darüber hinaus wurden in keiner der genannten Studien die kognitiven Fähigkeiten bzw. die Intelligenz oder auch die schulischen Noten prädiktiv für einen Wettbewerbserfolg. Zweifelsohne spielt das domänenspezifische Wissen jedoch eine Rolle für die erfolgreiche Teilnahme an einem naturwissenschaftlichen Schülerwettbewerb.

Auf den Zusammenhang von domänenspezifischem Wissen, Intelligenz und Leistung wird daher näher in Abschnitt 2.3.2 eingegangen.

Demzufolge sind weitere fundierte Kenntnisse über (potentiell) an naturwissenschaftlichen Wettbewerben teilnehmende Schülerinnen und Schüler nötig, um interessierte und talentierte Schülerinnen und Schüler gezielt identifizieren und fördern zu können. Im Besonderen quantitative Studien, die die Teilnehmenden im Quer- und Längsschnitt und möglichst zeitnah am Wettbewerb untersuchen, wären an dieser Stelle wünschenswert, um Aussagen für die Breite treffen zu können. Darüber hinaus scheinen für eine ideale Passung zwischen Teilnehmenden und Wettbewerb auch Informationen über die einzelnen Wettbewerbe und ihre Anforderungen erstrebenswert. Im Zusammenspiel affektiver Merkmale und fachspezifischer kognitiver Fähigkeiten von (potentiell) teilnehmenden Schülerinnen und Schülern sowie den Charakteristika der einzelnen Wettbewerbe kann so eine individuelle Förderung von interessierten und talentierten Schülerinnen und Schülern mithilfe naturwissenschaftlicher Schülerwettbewerbe gelingen.

## 2.3 Naturwissenschaftliche Kenntnisse und Fähigkeiten

Im folgenden Abschnitt wird eine Definition der im Rahmen dieser Arbeit zu untersuchenden naturwissenschaftlichen Kenntnisse und Fähigkeiten gegeben. Abschließend wird der Zusammenhang zwischen Intelligenz, (domänenspezifischem) Wissen und Leistung erläutert.

### 2.3.1 Definition: Naturwissenschaftliche Kenntnisse und Fähigkeiten

In der vorliegenden Arbeit werden naturwissenschaftliche Kenntnisse und Fähigkeiten über die Kompetenzbereiche der Bildungsstandards der KMK definiert (Klieme et al., 2007; Kultusministerkonferenz, 2005a, 2005b, 2005c). Begründet ist dieses Vorgehen zum einen darin, dass als Teilnehmende für die Erhebungen Schülerinnen und Schüler vorgesehen wurden. Zum anderen wurde der Bezug zu den Kenntnissen und Fähigkeiten, die in der Schule gelehrt und gelernt werden sollen, als wichtig erachtet, da aus den Ergebnissen des Projekts u. a. Empfehlungen für Lehrkräfte zur Identifizierung und Förderung naturwissenschaftlich talentierter und interessierter Jugendlicher generiert werden sollten. Darüber hinaus stellen die Evaluation der Bildungsstandards sowie die Ländervergleiche des Institut zur Qualitätsentwicklung im Bildungswesen (IQB) bildungspolitisch ein aktuelles Thema dar (z. B. Wellnitz et al., 2012; Walpuski, Kampa, Kauertz & Wellnitz, 2008; Walpuski et al., 2010; Pant, Böhme & Köller, 2013), so dass an dieser Stelle eine Vielzahl an Ansatzpunkten vermutet wurde.

Unter Kompetenzen versteht man im Rahmen der Bildungsstandards

die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können. (Weinert, 2001, S. 27-28)

Neben den Fachinhalten bzw. der inhaltlichen Dimension berücksichtigen die Bildungsstandards in den drei Naturwissenschaften Biologie, Chemie und Physik auch immer die Handlungsdimension. Die Handlungsdimension bezieht sich dabei auf die fundamentalen Elemente

der naturwissenschaftlichen Erkenntnisgewinnung (experimentelles und theoretisches Arbeiten), auf Kommunikation sowie auf die Anwendung und Bewertung naturwissenschaftlicher Sachverhalte in fachlichen und gesellschaftlichen Kontexten (Kultusministerkonferenz, 2005a, 2005b, 2005c). Daraus resultieren vier Kompetenzbereiche für die Bildungsstandards in den Naturwissenschaften:

- Fachwissen
- Erkenntnisgewinnung
- Kommunikation
- Bewertung

Der Kompetenzbereich Fachwissen stellt dabei die Inhaltsdimension und die Kompetenzbereiche Erkenntnisgewinnung, Kommunikation und Bewertung stellen die Handlungsdimensionen dar (Kultusministerkonferenz, 2005a, 2005b, 2005c).

Für die Kompetenzbereiche bestehen fachspezifische Definitionen, die sich vor allem im Kompetenzbereich Fachwissen voneinander abgrenzen:

- **Fachwissen Biologie** „Lebewesen, biologische Phänomene, Begriffe, Prinzipien, Fakten kennen und den Basiskonzepten [System, Struktur und Funktion, Entwicklung] zuordnen.“ (Kultusministerkonferenz, 2005a, S. 7)
- **Fachwissen Chemie** „Chemische Phänomene, Begriffe, Gesetzmäßigkeiten kennen und den Basiskonzepten [Stoff-Teilchen-Beziehungen, Struktur-Eigenschafts-Beziehungen, chemische Reaktion, energetische Betrachtung bei Stoffumwandlungen] zuordnen.“ (Kultusministerkonferenz, 2005b, S. 7)
- **Fachwissen Physik** „Physikalische Phänomene, Begriffe, Prinzipien, Fakten, Gesetzmäßigkeiten kennen und den Basiskonzepten [Materie, Wechselwirkung, System, Energie] zuordnen.“ (Kultusministerkonferenz, 2005c, S. 7)

Für den Kompetenzbereich Erkenntnisgewinnung bestehen hingegen nur Unterschiede zwischen der Biologie und den anderen beiden Fächern:

- **Erkenntnisgewinnung Biologie** „Beobachten, vergleichen, experimentieren, Modelle nutzen und Arbeitstechniken anwenden.“ (Kultusministerkonferenz, 2005a, S. 7)
- **Erkenntnisgewinnung Chemie, Physik** „Experimentelle und andere Untersuchungsmethoden sowie Modelle nutzen.“ (z. B. Kultusministerkonferenz, 2005b, S. 7)

Die Definitionen der Kompetenzbereiche Kommunikation und Bewertung sind hingegen für alle drei Naturwissenschaften gleich:

- **Kommunikation** „Informationen sach- und fachbezogen erschließen und austauschen.“ (z. B. Kultusministerkonferenz, 2005a).
- **Bewertung** Biologische, chemische, physikalische „Sachverhalte in verschiedenen Kontexten erkennen und bewerten.“ (z. B. Kultusministerkonferenz, 2005b, S. 7)

### 2.3.2 Zum Zusammenhang von Intelligenz, (domänenspezifischem) Wissen und Leistung

Bevor die Zusammenhänge zwischen Intelligenz, domänenspezifischem Wissen und Leistung erläutert werden, werden die genannten Konstrukte zur besseren Verständlichkeit zunächst eingegrenzt:

- **Intelligenz** wird verstanden als die Fähigkeit, „Informationen aufzunehmen und effizient zu verarbeiten, Zusammenhänge schnell und sicher zu erkennen, aus Beobachtungen und Erfahrungen Regeln und Gesetzmäßigkeiten abzuleiten und Regeln und Informationen zu kombinieren“ (Schmitt & Altstötter-Gleich, 2010, S. 127).
- **Wissen** im Allgemeinen wird verstanden als ein relativ dauerhafter Inhalt des Gedächtnisses, „dessen Bedeutung durch soziale Übereinkunft festgelegt ist“ (Gruber & Stamouli, 2015, S. 32). Zudem wird es in Zusammenhang mit entsprechenden Denkprozessen als eine Menge mentaler Repräsentationen angesehen, die es ermöglicht, Aufgaben zu bewältigen (Gruber & Stamouli, 2015). Unter **domänenspezifischem Wissen** versteht man ein „Netzwerk von Beziehungen verschiedener Wissens Elemente“ (Schrader, 2006, S. 572), das Resultat eines langfristigen Lernprozesses in einem bestimmten Fachgebiet ist (Sandmann, 2007). Es besteht aus deklarativem und prozeduralem Wissen, d. h. domänenspezifischen Kenntnissen und dem Wissen über bspw. das Lösen von Aufgaben (Schrader, 2006).
- Der Begriff **Leistung** ist hier allgemein zu verstehen, d. h. er umfasst verschiedene Domänen, Indikatoren und Vergleichsnormen zur Bewertung sowie verschiedene Grade von Leistung und Leistungsexzellenz (vgl. Harder, 2012).

Intelligenz und Wissen werden in der Forschung meist voneinander abgegrenzt, obgleich sie eng miteinander verbunden sind (Gruber & Stamouli, 2015; Schmitt & Altstötter-Gleich, 2010). Erwerb und Qualität des Wissens hängen von den intellektuellen Fähigkeiten ab (Schmitt & Altstötter-Gleich, 2010; Schrader, 2006; Weinert, 1996). So kann fehlendes Wissen nicht durch hohe intellektuelle Fähigkeiten kompensiert werden, allerdings können niedrige intellektuelle Fähigkeiten bis zu einem bestimmten Grad durch fundiertes Wissen ausgeglichen werden (Dochy, de Rijdt & Dyck, 2002). Ackerman (1987, 1992, 1996) beschreibt in seiner Theorie der *Determinants of Skilled Performance* die Rolle der Intelligenz beim Kompetenzerwerb derart, dass die Intelligenz zu Beginn eine große Rolle spielt, aber im Laufe des Entstehens von prozeduralem Wissen zunehmend in den Hintergrund tritt (Gruber & Stamouli, 2015).

Eine Vielzahl wissenschaftlicher Studien untersuchte Intelligenz und Wissen als bestimmende oder prädiktive Faktoren von schulischer oder auch beruflicher Leistung. Mit mittleren Korrelationen von  $r = .50$  bis  $r = .60$  stellte die Intelligenz bzw. stellten die kognitiven Fähigkeiten von Schülerinnen und Schülern eine der wichtigsten Determinanten der Schulleistung dar (vgl. Gustafsson & Undheim, 1996; Helmke & Schrader, 2010). Das domänenspezifische Wissen erwies sich jedoch in Studien, in denen sowohl die Intelligenz als auch das domänenspezifische Wissen als Prädiktoren domänenspezifischer Leistungen untersucht wurden, vorhersagestärker als die Intelligenz (z. B. Ceci & Liker, 1986; Helmke & Schrader, 2010; Schneider & Bjorklund, 1992; Weinert & Helmke, 1995). Auch Schrader (2006) führt an, dass

Leistungsunterschiede vor allem durch Unterschiede im domänenspezifischen Wissen (30 bis 60%), und damit mehr als durch Intelligenzunterschiede (ca. 25 bis 36%), vorhergesagt werden können, wobei es Überschneidungen der Erklärungsbeiträge von Intelligenz und Vorwissen gibt. Studien im schulischen Kontext zeigten, dass das zu Beginn des Schuljahres vorhandene Wissen den größten Einfluss auf den Lernfortschritt von Schülerinnen und Schülern hatte (Gruber & Stamouli, 2015). Für die Mathematik zeigten Weinert und Helmke (1995) für den Elementarbereich und Stern (1997) für die Sekundarstufe 2, dass die Mathematikleistung in einer höheren Jahrgangsstufe eng mit der Leistung in Mathematik in einer niedrigeren Jahrgangsstufe, und zwar mehr als mit der Intelligenz, zusammenhing.

Aus dem Bereich der Expertiseforschung ist bekannt, dass der Unterschied zwischen Experten und Novizen maßgeblich durch die Quantität und Qualität ihres domänenspezifischen Wissens bestimmt wird (Schrader, 2006) und dass domänenspezifisches Wissen der beste Prädiktor für domänenspezifische Problemlöseleistungen ist (Friege, 2002). Die Vorhersage von Leistung im Bereich des Problemlösens mithilfe von Maßen allgemeiner Intelligenz gelingt nur äußerst unzureichend (Dörner & Kreuzig, 1983; Friege, 2002). Auch im Vergleich zu fächerübergreifenden Kompetenzen wie z. B. dem schlussfolgernden Denken stellte sich das domänenspezifische Wissen als bedeutender heraus (Friege, 2003). Rist (1989) erklärte den Erfolg von Computerexperten beim Problemlösen ebenso durch ihr größeres domänenspezifisches Wissen.

Wenngleich die Forschung Intelligenz und Wissen voneinander unterscheidet, ist die Messung der beiden Konstrukte, vor allem die Messung von Wissen in Abgrenzung zur Intelligenz, als anspruchsvoll zu bezeichnen. So wurde teilweise den großen Schulleistungsstudien wie Trends in International Mathematics and Science Study (TIMSS) und PISA vorgeworfen, dass die meisten Aufgaben lediglich Leseverständnis oder kognitive Grundfähigkeiten messen würden (Hagemester, 1999) bzw. dass die Messkonstrukte der Schulleistungsstudien den Messkonstrukten aus Intelligenztests ähnlicher und die Skalen im Vergleich zu gängigen Intelligenztests auch auf Individualdatenebene deutlich homogener seien als bislang angenommen (Rindermann, 2006). Für die Chemie- und Physik-Aufgaben im IQB-Ländervergleich 2012 des Kompetenzbereichs Fachwissen, der gemäß der Definition aus Abschnitt 2.3.1 auch Bestandteil der in dieser Arbeit zu untersuchenden naturwissenschaftlichen Kenntnisse und Fähigkeiten ist, ist kritisch anzumerken, dass sich herausstellte, dass die sprachlichen Fähigkeiten der Schülerinnen und Schüler die Leistung beeinflussten (Härtig, Heitmann & Retelsdorf, 2015; Ropohl, Walpuski & Sumfleth, 2015). Dies resultierte im Besonderen daraus, dass die Aufgaben alle für die Lösung relevanten Fachinformationen enthielten (Sumfleth et al., 2013). (Begründet wurde dieses Aufgabendesign damit, dass im Rahmen des IQB-Ländervergleichs 2012 im Sinne des Kompetenzbegriffs nach Weinert (2001) nicht das Fachwissen selbst, sondern der Umgang mit Fachwissen erhoben werden sollte, vgl. Sumfleth et al., 2013.)

Neben der Intelligenz und (domänenspezifischem) Wissen als Determinanten von Schulleistung existieren noch eine Reihe weiterer Faktoren, die als mögliche Determinanten untersucht wurden. Vermehrt ins Interesse der Forschung gerückt sind volitionale Determinanten des Lernens und der Schulleistung (Helmke & Schrader, 2010). Hierzu zählen Lernstile und Lernstrategien, metakognitive Kompetenzen, Lerngewohnheiten, Arbeitstechniken und Handlungskontrolle, die bislang jedoch nur einen schwachen Zusammenhang zur schulischen Leistung zeigten (Helmke & Schrader, 2010). Stärkere Zusammenhänge fanden sich zu motivationalen Determinanten wie dem Fähigkeitsselbstkonzept (z. B. Köller et al., 2006; Marsh & Martin, 2011;

D. H. Rost, Dickhäuser, Sparfeldt & Schilling, 2004), der Lernmotivation (z. B. Paulick, Wattermann & Nückles, 2013; Steinmayr et al., 2011; Wirthwein, Sparfeldt, Pinquart, Wegerer & Steinmayr, 2013), dem Interesse (z. B. Köller et al., 2001; Krapp et al., 1993; Marsh et al., 2005) und der Einstellung zum Lernen (z. B. Goetz et al., 2004), wobei das Interesse die Schulleistung am stärksten von den genannten Variablen beeinflusste (Helmke & Schrader, 2010). Auch familiäre Einflussfaktoren spielten eine Rolle (z. B. Zimmermann & Spangler, 2001), vor allem die soziale Herkunft zeigte einen Zusammenhang zur schulischen Leistung (Helmke & Schrader, 2010; Prenzel, Heidemeier, Ramm, Hohensee & Ehmke, 2004). Schulische Faktoren wie der Klassenkontext, die Unterrichtsqualität und Lehrermerkmale konnten ebenso als Determinanten von schulischer Leistung identifiziert werden (Helmke & Schrader, 2010), jedoch ist ihr Erklärungsbeitrag geringer als derjenige von Intelligenz und (domänenspezifischem) Wissen.

Für die Wettbewerbsforschung und die Vorhersage von Leistung im Wettbewerbskontext ist festzustellen, dass das alleinige Erheben von (domänenspezifischem) Wissen und Intelligenz kaum genügen würde, um Leistung zu erklären. Der hohe Beitrag, den das (domänenspezifische) Wissen zur Erklärung von Leistung liefert, spricht jedoch trotz der Kritik, die bspw. an den Aufgaben der Schulleistungsstudien TIMSS, PISA und dem IQB-Ländervergleich verübt wurde, für das Erfassen von (domänenspezifischem) Wissen, da volitionale und motivationale Merkmale alleine nicht genügen, um Leistung zu prognostizieren. Zudem kann von komplexen Interaktionen der einzelnen Faktoren untereinander ausgegangen werden (vgl. Helmke & Schrader, 2010). Betrachtet man die Vorhersage von Leistung im Wettbewerbskontext, versucht das Modell IKoN dem Anspruch zu genügen, affektive und kognitive Merkmale zu erfassen, um u. a. die Vorhersage von Leistung im Wettbewerbskontext zu ermöglichen.

### 2.3.3 Zusammenfassung und Fazit

Die in der vorliegenden Arbeit zu messenden naturwissenschaftlichen Kenntnisse und Fähigkeiten werden über die Kompetenzbereiche der Bildungsstandards (Kultusministerkonferenz, 2005a, 2005b, 2005c) und die zugehörigen Kompetenzbereiche Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung definiert. Die Notwendigkeit der Erfassung (domänenspezifischen) Wissens, hier neben Maßen der Intelligenz über naturwissenschaftliche Kenntnisse und Fähigkeiten konkretisiert, resultiert aus der Erkenntnis, dass (domänenspezifisches) Wissen mehr noch als Intelligenz bzw. kognitive Fähigkeiten schulische und berufliche Leistung determiniert (z. B. Ceci & Liker, 1986; Helmke & Schrader, 2010; Schneider & Bjorklund, 1992; Weinert & Helmke, 1995). Der Schluss liegt nahe, dass (domänenspezifisches) Wissen auch als Prädiktor von Leistung im Wettbewerbskontext dienen kann. Weitere Merkmale, die aufgrund von Forschungsergebnissen aus dem schulischen Kontext als Prädiktoren schulischer Leistung identifiziert wurden, werden im Rahmen des IKoN erhoben und in Kombination mit den Ergebnissen zu naturwissenschaftlichen Kenntnissen und Fähigkeiten auf ihren potentiellen Einfluss als Prädiktor auf Leistung im Wettbewerbskontext untersucht.

## Kapitel 3

# Zielsetzungen und Forschungsfragen

Wie in Kapitel 2 erläutert sind im Bereich der Wettbewerbsforschung weitere Kenntnisse über Teilnehmende naturwissenschaftlicher Schülerwettbewerbe erforderlich, um naturwissenschaftlich interessierte und talentierte Schülerinnen und Schüler gezielt identifizieren und fördern zu können. Ein Forschungsdesiderat wurde im Besonderen für quantitative Quer- und Längsschnittstudien festgestellt, die neben Informationen zu Teilnehmenden auch Informationen zu den Wettbewerben und ihren Anforderungen erfassen. Des Weiteren deuten Erkenntnisse aus der Forschung darauf hin, dass gerade das domänenspezifische Wissen von an naturwissenschaftlichen Wettbewerben teilnehmenden Schülerinnen und Schülern prädiktiv für Leistung im Wettbewerbskontext sein könnte.

Für Studien in der Wettbewerbsforschung gibt es eine Reihe von Herausforderungen, die es zu überwinden gilt, um Erhebungen im Wettbewerbskontext praktikabel und ohne Einschränkungen für die Wettbewerbe selbst umsetzen zu können. So sollten Erhebungen möglichst während oder unmittelbar nach dem entsprechenden Wettbewerb durchgeführt werden, da die Teilnehmenden außerhalb des Wettbewerbs z. B. aufgrund unterschiedlicher Schulen schwer zu erreichen sind. Dies hat zur Folge, dass die Erhebungen den Rahmenbedingungen des Wettbewerbs angepasst werden müssen. Des Weiteren sind meist eine Vielfalt verschiedener affektiver und kognitiver Erhebungsinstrumente notwendig, um Teilnehmende umfassend zu beschreiben. Die Testzeit ist daher gerade in der Wettbewerbsforschung meist äußerst knapp bemessen. Im Vergleich zwischen Wettbewerbsteilnehmenden und wettbewerbsunerfahrenen Schülerinnen und Schülern, der besonders von Interesse ist, um z. B. anhand der hieraus gewonnenen Erkenntnisse potentielle Wettbewerbsteilnehmende in der Gruppe wettbewerbsunerfahrener Schülerinnen und Schüler zu identifizieren, sind Erhebungsinstrumente notwendig, die zwischen den genannten Gruppen ausreichend differenzieren können. Besonders im Bereich der Leistung, v. a. zwischen erfolgreichen Wettbewerbsteilnehmenden und wettbewerbsunerfahrenen Schülerinnen und Schülern, sind große Differenzen zu vermuten, so dass Erhebungsinstrumente benötigt werden, die sowohl im oberen als auch im unteren Leistungsspektrum differenzieren können. Für die direkte Vergleichbarkeit von Gruppen sowie aus testökonomischen Gründen sollte daher sowohl im Bereich affektiver als auch im Bereich kognitiver Merkmale angestrebt werden, gemeinsame Instrumente für die Zielpopulation zu entwickeln.

Aus den genannten Gründen ergeben sich für die vorliegende Arbeit zu den spezifischen

Leistungsmerkmalen von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe die folgenden Zielsetzungen und Forschungsfragen:

1. Lassen sich die naturwissenschaftlichen Kenntnisse und Fähigkeiten von erfolgreichen Wettbewerbsteilnehmenden im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern mit einem gemeinsamen Testinstrument erfassen? - Entwicklung und Validierung eines Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten von Wettbewerbsteilnehmenden.

Im Fokus von Studie 1 stand die Validierung eines Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten von Wettbewerbsteilnehmenden (Kapitel 6), der eine Analyse von Wettbewerbsanforderungen zugrunde lag (Kapitel 4). Die Entwicklung des Testinstruments wird in Kapitel 5 erläutert. Mit Teilnehmenden des Bundesfinales der IJSO sowie Schülerinnen und Schülern, die mindestens einmal oder gar nicht an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten, wurde das Testinstrument erprobt.

2. Gelingt die Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten mit einem an die Bedürfnisse der Wettbewerbsforschung angepassten Testinstrument? - Erprobung eines reduzierten Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten an einer umfangreichen Stichprobe.

Aufgrund der geschilderten Herausforderungen in der Wettbewerbsforschung wurde in Studie 2 eine reduzierte Version des Testinstruments mit wettbewerbsunerfahrenen Schülerinnen und Schülern und Teilnehmenden verschiedener naturwissenschaftlicher Schülerwettbewerbe erprobt (Kapitel 7).

3. Zeigen sich qualitative Unterschiede in den Antworten von Teilnehmenden unterschiedlicher naturwissenschaftlicher Schülerwettbewerbe, auch im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern? - Varianzanalytischer Vergleich und qualitative Analyse der Verteilungen der Antwortkategorien für einzelne Aufgaben und Subgruppen.

Neben quantitativen Merkmalen, die in Studie 1 und 2 untersucht wurden, wurden für Studie 3 - neben der varianzanalytischen Auswertung der erreichten Mittelwerte - qualitative Merkmale der Aufgaben und ihrer Antwortkategorien aus dem Testinstrument aus Studie 2 näher betrachtet und für verschiedene Subgruppen analysiert (Kapitel 8). Von Interesse waren hier mögliche Muster im Antwortverhalten bzw. in der Verteilung der Antwortkategorien der unterschiedlichen Subgruppen sowie Unterschiede in den verschiedenen Aufgabenkategorien, die möglicherweise auf spezifische Charakteristika der einzelnen Gruppen zurückgeführt werden konnten.

## Kapitel 4

# Analyse von Wettbewerbsanforderungen

Für eine gezielte Förderung von naturwissenschaftlich interessierten und talentierten Schülerinnen und Schülern ist neben der Charakterisierung von Wettbewerbsteilnehmenden auch eine Analyse der Anforderungen, die die jeweiligen Wettbewerbe an ihre Teilnehmenden stellen, erforderlich. Auch im Rahmen von Vorüberlegungen zur Erstellung eines ersten Testentwurfs stellte sich die Frage nach den Anforderungen von Wettbewerben, deren Teilnehmende sich eventuell an den im Rahmen dieser Arbeit durchzuführenden Studien beteiligen würden. Aus der Forschung ist hierzu kaum etwas bekannt. Lediglich im Bereich der Informatik wurden bislang Untersuchungen zu den spezifischen Leistungsanforderungen, die der Bundeswettbewerb Informatik sowie Jugend forscht an ihre Teilnehmenden stellen, durchgeführt (Meißner, 2013). Parallel zu dieser Arbeit entstand zudem eine Analyse der Inhalte und Anforderungen der Aufgaben des deutschen Auswahlverfahrens der IChO (Deussing, 2014, siehe Abschnitt 4.2.5). Da die Kenntnis der spezifischen Anforderungen der zu untersuchenden Schülerwettbewerbe als bedeutsames Kriterium für die Eingrenzung des zu messenden Konstruktes erachtet wurde, wurde eine Analyse der Anforderungen zweier Wettbewerbe vorgenommen. Ausgewählt wurden die IJSO als aufgabenorientierter Wettbewerb sowie der Wettbewerb Jugend forscht bzw. Schüler experimentieren als projektorientierter Wettbewerb (zu den Beschreibungen der Wettbewerbe siehe Abschnitt 4.2 und 4.3). Die Erkenntnisse stellten die Basis für die spätere Erstellung des Testentwurfes dar und dienten in diesem Zusammenhang der Eingrenzung des zu messenden Konstrukts und der Definition des Messgegenstandes. Im folgenden Kapitel werden nach einem kurzen Überblick zum Stand der Forschung zur Klassifikation und Analyse von Aufgaben die Methoden und Ergebnisse der Analysen der Anforderungen des deutschen Auswahlverfahrens der IJSO und des Wettbewerbs Jugend forscht bzw. Schüler experimentieren vorgestellt.

### 4.1 Klassifikation und Analyse von Aufgaben: ein Überblick

Die Klassifikation und Analyse von Aufgaben hat vor allem seit der Veröffentlichung der ersten Ergebnisse der internationalen Schulleistungsstudien TIMSS und PISA an Bedeutung gewonnen. So deckten bspw. die Ergebnisse der dritten TIMS-Studie (Baumert, Bos & Lehmann, 2000b) Defizite deutscher Schülerinnen und Schüler auf, die vor allem bei der Bearbeitung von Aufgaben lagen, „die konzeptuelles Verständnis, das Übertragen des Gelernten

auf neue Kontexte, das selbstständige Anwenden des Gelernten und flexibles Umstrukturieren von Problemkonstellationen erforderten“ (Nerdel, Neumann, Stäudel & Rehm, 2013, S. 96). Als Grund hierfür wurde vor allem der in Mathematik und den Naturwissenschaften vorherrschende fragend-entwickelnde Unterricht gesehen, der stark von Lehrkräften gesteuert und vorwiegend auf das Einüben von Routinen ausgerichtet gewesen sei (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, 1997; Klieme & Baumert, 2001). Die Befunde spiegelten sich auch in der Aufgabenkultur wider. Aufgaben seien auf das Einüben von Wissen ausgerichtet gewesen, berücksichtigten kaum verschiedene Kontexte und verbanden selten das neue Wissen mit bereits Erlerntem (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, 1997). Als Konsequenz aus diesen Befunden und um den mathematisch-naturwissenschaftlichen Unterricht nachhaltig zu verbessern, wurde u. a. beschlossen, die Aufgabenkultur im mathematisch-naturwissenschaftlichen Unterricht weiterzuentwickeln. So sollten neue Aufgabentypen, die auch mehrere Lösungswege zulassen, entwickelt und erprobt sowie anwendungsorientierte Aufgaben in variierenden Strukturen, Inhalten und Kontexten erstellt werden (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, 1997; Häußler & Lind, 1998). Auch existierende Aufgaben sollten hinsichtlich der neuen Forderungen überarbeitet werden (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, 1997). Aus der Diskussion um die Befunde der Schulleistungsstudien resultierten neben Forderungen nach der Weiterentwicklung der Aufgabenkultur ebenso Forderungen nach einer veränderten Zielsetzung naturwissenschaftlichen Unterrichts (Nerdel et al., 2013). In der Folge wurden schließlich die Bildungsstandards definiert (Kultusministerkonferenz, 2005a, 2005b, 2005c), die für das Ziel naturwissenschaftlicher Kompetenz, aufgeteilt in die vier Kompetenzbereiche Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung, als Standard setzten (Klieme et al., 2007; Nerdel et al., 2013). Schülerinnen und Schüler sollten, in Anlehnung an den Kompetenzbegriff nach Weinert (2001), durch den kompetenzorientierten naturwissenschaftlichen Unterricht in unterschiedlichen Situationen zum erfolgreichen Lösen von realen und lebensweltlichen Problemen befähigt werden (Klieme et al., 2007; Nerdel et al., 2013). Die Umsetzung der Bildungsstandards in der Unterrichtspraxis, in Aufgaben- und Lernmaterialien, aber auch die Evaluation ebendieser sowie der landes- und bundesweite Vergleich von Kompetenzen der Schülerinnen und Schüler wurde Gegenstand unterrichtspraktischer und fachdidaktischer Forschung. Die Entwicklung, die sich dabei in der Aufgabenkultur vollzogen hat, ist heute in Schulbüchern, Unterrichts- und Lehrermaterialien sowie entsprechenden Fachzeitschriften direkt ablesbar und zeigt sich im Aufgabendesign (z. B. Einsatz multipler Repräsentationen), der Orientierung an verschiedenen Anwendungen und Kontexten sowie der Entwicklung experimenteller Aufgaben (Nerdel et al., 2013).

Auch wenn Aufgaben bislang eher aus fachdidaktischer Perspektive erforscht wurden (Maier, Bohl, Kleinknecht & Metz, 2013), existieren eine Reihe allgemeindidaktischer Konzepte, die zur Klassifikation und Analyse von Aufgaben verwendet werden können. Eine Übersicht der einflussreichsten Klassifikationsschemata, basierend auf den Darstellungen von Maier et al. (2013) und Moseley et al. (2005), ist in Tabelle 4.1 dargestellt. Für überblickende Darstellungen siehe z. B. Moseley et al. (2005) oder Kleinknecht, Bohl, Maier und Metz (2013). Zudem ist anzumerken, dass die vorgestellten Klassifikationsschemata vorwiegend der Unterrichtsforschung entstammen und nicht immer speziell für die Klassifikation und Analyse von Aufgaben erstellt wurden. Der Fokus lag meist auf der gezielten Entwicklung und Modifikation von Aufgaben, basierend auf der Analyse und Strukturierung des zu erlernenden

bzw. vermittelnden Wissens (Maier et al., 2013).

Vor allem die Arbeiten von Blömeke et al. (2006) und Kleinknecht et al. (2011) liefern Ansätze, Aufgaben sowohl aus allgemeindidaktischer als auch aus fachdidaktischer Sicht zu betrachten. Eine Vielzahl an Klassifikationsschemata, die für die Analyse von Aufgaben genutzt werden können, liefern die mathematisch-naturwissenschaftlichen Fachdidaktiken. Ihre Klassifikationsschemata entstanden meist im Kontext von Schulleistungsstudien oder im Kontext von Studien zur Erfassung mathematisch-naturwissenschaftlicher Kompetenzen bzw. zur Erfassung mathematisch-naturwissenschaftlichen Wissens. Eine Übersicht über naturwissenschaftsdidaktische Klassifikationsschemata, angelehnt an jene von Maier et al. (2013), ist in Tabelle 4.2 dargestellt.

Zusammenfassend ist festzustellen, dass mittlerweile eine Vielzahl wissenschaftlicher Arbeiten zu dem Bereich der allgemeindidaktischen und der fachdidaktischen Aufgabenanalyse existieren, die in ihren Zielsetzungen jedoch teilweise stark variieren. Dennoch gibt es einige Kriterien zur Analyse von Aufgaben, die in nahezu allen hier vorgestellten Klassifikationsschemata berücksichtigt werden. So umfassen die Klassifikationsschemata meist die Beschreibung kognitiver Prozesse, die Komplexität und die Art des Wissens. Die vorgestellten Arbeiten beziehen sich jedoch, wie bereits erwähnt, meist auf die Analyse von Lern- und Testaufgaben im Unterricht oder für den Unterricht. Zu Aufgaben aus Schülerwettbewerben, insbesondere zu den Aufgaben aus den naturwissenschaftlichen Olympiaden (nationale Auswahlwettbewerbe und internationale Wettbewerbe), ist hingegen wenig bekannt. Vorhandene Arbeiten konzentrierten sich oftmals auf Informatik-Wettbewerbe. So untersuchte bspw. Schlüter (2009, 2010) Aufgaben aus dem Schülerwettbewerb Informatik-Biber hinsichtlich der Kriterien Erfahrungsnähe, Abstraktionsgrad, Komplexität, Formalisierungsgrad, Redundanz, Anforderungsbereich, Prozessbereich, Lernzielstufe, Art des Wissens, Aufgabenschwierigkeit und Trennschärfe. Aus dem Bereich der ScienceOlympiaden ist der Bericht von Klieme et al. (1999) zu nennen, in dem für die 9. IBO die Aufgaben des internationalen Wettbewerbs hinsichtlich verschiedener Kriterien wie z. B. Schwierigkeit, Reliabilität und Validität verschiedener Subtests oder Zusammenhängen zwischen verschiedenen Faktoren evaluiert wurden. Das Ziel dieser Evaluation bestand dabei vor allem darin, die Aufgabenentwicklung für künftige Olympiaden zu verbessern und die diagnostische Güte der Aufgaben zu untersuchen (Klieme et al., 1999, S.32-33). Für die IPhO untersuchten Gau und Gau (1990) Wettbewerbsaufgaben hinsichtlich inhaltlicher Schwankungen sowie in Bezug auf das Anforderungsniveau der Aufgaben des internationalen Wettbewerbs. Das Anforderungsniveau wurde darüber erfasst, welche Schritte Teilnehmende durchlaufen und welche Prinzipien, Gesetze, etc. sie kennen oder anwenden müssen, um eine Aufgabe zu lösen. Diese *requirement determining characteristics* wurden dann aufsummiert. Dabei wurde angenommen, dass je höher die Anzahl der *requirement determining characteristics* ist, desto höher seien auch die an die Teilnehmenden gestellten Anforderungen. Beide erwähnten Studien dienten jedoch nicht der Generierung systematischen Wissens über die Aufgaben von Schülerolympiaden, sondern vor allem der Beantwortung wettbewerbsinterner Fragestellungen, z. B. wie man künftig Aufgaben für den Wettbewerb entwickeln könnte. Ein erster systematischer Ansatz zur Analyse von Wettbewerbsaufgaben findet sich bei Petersen (2010). Hier wurde ein Klassifikationsschema für die IPhO zur Analyse von Wettbewerbsaufgaben sowohl des nationalen Auswahlverfahrens als auch für Aufgaben des internationalen Wettbewerbs entwickelt. Da dieses Klassifikationsschema im Gegensatz zu den genannten Studien eher übergeordnete Aufgabenmerkmale berücksichtigte und hinsichtlich der Anforderungen der

Tabelle 4.1: Übersicht über allgemeindidaktische Klassifikationsschemata

<i>Klassifikationsschema</i>	<i>Zielsetzung</i>	<i>Kriterien</i>
Bloom's taxonomy of educational objectives, cognitive domain (Bloom et al., 1956)	Einordnung kognitiver Lernziele in aufeinander aufbauende Lernstufen	Definition von sechs hierarchisch angeordneten Lernstufen, die kognitive Prozesse abbilden: know/ledge, comprehension, application, analysis, synthesis, evaluation
Gagné's eight types of learning and five types of learned capability (R. M. Gagné, 1965; R. M. Gagné & Briggs, 1974)	Analyse des zu erlernenden Wissens und Ableitung von Lernzielen	Definition von acht aufeinander aufbauenden Arten des Lernens (signal learning, stimulus/ response learning, chaining, verbal chaining, discrimination learning, concept learning, rule learning, problem-solving) und von fünf kognitiven Bereichen des Lernens (intellectual skills, cognitive strategies, verbal information, motor skills, attitudes)
Biggs and Collis' SOLO (structure of the observed learning outcome) taxonomy (Biggs & Collis, 1982)	Systematische Beschreibung der Komplexität und des Abstraktionsgrades von Lernergebnissen	Definition von fünf hierarchisch geordneten Stufen des Wissens: pre-structural, unistructural, multistructural, relational, extended abstract
Anderson and Krathwohl's revision of Bloom's taxonomy (Anderson & Krathwohl, 2001)	Weiterentwicklung der Bloom'schen Lernzieltaxonomie zu einem zweidimensionalen Konstrukt	Definition von sechs kognitiven Prozessdimensionen (remember, understand, apply, analyse, evaluate, create) und vier Wissensdimensionen (factual, conceptual, procedural, metacognitive)
Anforderungsbereiche der Bildungsstandards (z. B. Kultusministerkonferenz, 2004a, 2004b)	Einschätzung von in Aufgaben gestellten Anforderungen	Einteilung in drei hierarchisch geordnete Anforderungsbereiche: Reproduzieren, Zusammenhänge herstellen sowie Verallgemeinern und Reflektieren
Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht (Blömeke et al., 2006)	Ermittlung der Qualität von Aufgaben über ein fachübergreifendes Modell: Berücksichtigung der drei Unterrichtsperspektiven Inhalt, Lehren und Lernen sowie Entwicklung eines dreistufigen Analyseverfahrens	Exemplarische Erschließung eines gesellschaftlich relevanten Bildungsinhaltes, Ansprache eines Bedürfnisses der Schillerinnen und Schüler, Förderung genereller intellektueller Fähigkeiten, Neuigkeitswert in Bezug auf den bereichsspezifischen Wissens- und Erfahrungsstand, Chance auf Bewältigung, Potenzial zur inneren Differenzierung, Repräsentation einer authentischen Situation, Förderung von Problemlösefähigkeit, Erfordernis sozialer Interaktionen
Analyse des kognitiven Aufgabenpotenzials (Kleinmecht et al., 2011)	Entwicklung eines allgemeindidaktischen Auswertungsmanuals zur Analyse des kognitiven Potenzials von Aufgaben anhand von verschiedenen Analysedimensionen	Definition von sieben Analysedimensionen (Wissensart, kognitiver Prozess, Wissensseinheiten, Offenheit, Lebensweltbezug, sprachlogische Komplexität, Repräsentationsformen), weitere Unterteilung der Analysedimensionen in jeweils drei bis fünf Kategorien

Tabelle 4.2: Übersicht über naturwissenschaftsdidaktische Klassifikationsschemata

<i>Klassifikationsschema</i>	<i>Zielsetzung</i>	<i>Kriterien</i>
Analyse der Merkmale und des didaktisch-methodischen Einsatzes von Aufgaben im Biologieunterricht (Jatzwauk, 2007)	Analyse der durch die Aufgaben gestellten kognitiven Anforderungen und Analyse der Integration der Aufgaben in den Unterrichtsprozess	Kognitive Anforderungen: Niveau des kognitiven Prozesses (achtstufig), Antwortformat (sieben Kategorien), geforderte Kompetenzen in Anlehnung an die Bildungsstandards. Integration in den Unterricht: z. B. Funktion der Aufgabe, Adressat der Aufgabe
Instruktionspsychologische Analysen von Lernaufgaben im Physik-Unterricht (Schabram, 2007)	Überprüfung der Auswirkungen von schwierigkeitsbestimmenden Aufgabenmerkmalen auf die Lernerleistungen von Schülerinnen und Schülern	Beschreibung der kognitiven Prozesse durch die drei Bereiche Wissen, Verstehen, Anwendung
Modellierung physikalischer Kompetenz und ihrer Entwicklung (Neumann et al., 2007)	Entwicklung eines dreidimensionalen Strukturmodells physikalischer Kompetenz	Drei Dimensionen: Kognitive Aktivität (Erinnern, Strukturieren, Explorieren), Komplexität (Fakt, Mehrere Fakten, Zusammenhang, Mehrere unverbundene Zusammenhänge, Mehrere verbundene Zusammenhänge, Übergeordnetes Konzept) und Leitidee (Energie, Materie, System, Wechselwirkung)
Adaption des „Model of hierarchical complexity“ für den Fachbereich Chemie (MHC-C, Bernholt et al., 2009)	Beschreibung fachspezifischer Kompetenz in Chemie	Sechsstufiges hierarchisch angeordnetes Komplexitätsschema (Stufe 1: Unreflektiertes Erfahrungswissen, Stufe 2: Fakten, Stufe 3: Prozessbeschreibungen, Stufe 4: Lineare Kausalität, Stufe 5: Multivariate Interdependenz)
Dreidimensionales Kompetenzstrukturmodell der naturwissenschaftlichen Fächer (Sumfleth et al., 2013)	Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe 1 (IQB-Ländervergleich, Pant, Böhme & Köller, 2013)	Drei Dimensionen: Kognitive Prozesse (integrieren, organisieren, selektieren, reproduzieren), Komplexität (Niveau I - ein Fakt, Niveau II - zwei Fakten, Niveau III - ein Zusammenhang, Niveau IV - zwei Zusammenhänge, Niveau V - übergeordnetes Konzept) und Kompetenzbereiche (Umgang mit Fachwissen, Erkenntnisgewinnung, Kommunikation, Bewertung)

IJSO und der IPhO große Schnittmengen erwartet wurden, diente das Klassifikationsschema nach Petersen (2010) als Basis für die Entwicklung eines Klassifikations- und Analyseschemas für die Aufgaben der IJSO. Eine Erläuterung des Klassifikationsschemas nach Petersen (2010) findet sich in Abschnitt 4.2.1.

## 4.2 Analyse der Anforderungen in der Internationalen Junior-ScienceOlympiade

Die IJSO besteht seit 2004 und richtet sich an naturwissenschaftlich in der Breite interessierte Schülerinnen und Schüler, die im Wettbewerbsjahr höchstens 15 Jahre alt sind. Das deutsche Auswahlverfahren für die IJSO fand erstmals 2008 statt. Die Teilnehmenden bearbeiten über mehrere Wettbewerbsrunden thematisch breit gefächerte Aufgaben aus den Bereichen Biologie, Chemie und Physik. Über die Auswahlrunden verteilt verfolgt der Wettbewerb unterschiedliche Zielsetzungen, die von der Breitenförderung in der ersten Runde bis zu Enrichment und Spitzenförderung in der zweiten und dritten Runde sowie letztendlich der Auswahl der sechs besten Schülerinnen und Schüler für den internationalen Wettbewerb im Bundesfinale reichen (Peters et al., 2013). Wird in der Folge von Runden oder (Wettbewerbs-)Aufgaben der IJSO gesprochen, sind immer die Runden bzw. die Aufgaben des deutschen Auswahlverfahrens gemeint.

Das deutsche Auswahlverfahren für die IJSO, an dem jährlich ca. 3500 Schülerinnen und Schüler teilnehmen (Bundesministerium für Bildung und Forschung, 2015, Zugriffsdatum 18.05.2015), bestand ursprünglich aus drei Runden, wurde jedoch im Jahr 2015 um eine vierte Runde ergänzt (Internationale JuniorScienceOlympiade, 2016, Zugriffsdatum 09.12.2016). Da der Wettbewerb zum Zeitpunkt der Analyse der Wettbewerbsanforderungen noch drei Runden umfasste, werden diese im Folgenden erläutert: Die erste Wettbewerbsrunde der IJSO bestand aus einer Hausaufgabenrunde, die theoretische und experimentelle Aufgaben beinhaltete. In der ersten Runde war es auch möglich, in Gruppen von maximal drei Schülerinnen und/ oder Schülern die Aufgaben zu bearbeiten. In der zweiten Runde bearbeiteten ca. 500 Schülerinnen und Schüler eine Klausur mit theoretischen Aufgaben (Multiple-Choice und offene Aufgaben), die an der entsprechenden Schule der Schülerinnen und Schüler stattfand und 90 Minuten dauerte. Die dritte Runde, das Bundesfinale, fand mit 45 Schülerinnen und Schülern an einem bundeseinheitlichen Austragungsort statt. Innerhalb einer Woche schrieben die Schülerinnen und Schüler mehrere theoretische und praktische Klausuren. Die sechs besten Schülerinnen und Schüler qualifizierten sich für den internationalen Wettbewerb (Peters et al., 2013). Das neue Auswahlverfahren sieht nun eine weitere Runde nach der ersten Hausaufgabenrunde vor. So nehmen ca. 1000 Schülerinnen und Schüler in der aktuell zweiten Runde an einem Multiple-Choice-Test teil, der ebenso wie die aktuell dritte Runde an den entsprechenden Schulen der Bundesländer durchgeführt wird. Die dritte Runde, für die sich ca. 350 Schülerinnen und Schüler qualifizieren können, sieht wie die vormals zweite Runde eine Klausur mit theoretischen Aufgaben vor. Die 39 besten Schülerinnen und Schüler nehmen an der aktuell vierten Runde, dem Bundesfinale, teil, das wie gehabt durchgeführt wird (Internationale JuniorScienceOlympiade, 2016, Zugriffsdatum 09.12.2016).

Anhand der dargestellten Struktur des deutschen Auswahlverfahrens der IJSO ist zu erkennen, dass es sich hier um einen aufgabenorientierten Wettbewerb handelt. Die Beurteilung

der teilnehmenden Schülerinnen und Schüler erfolgt ausschließlich über die Bewertung ihrer Antworten, die wiederum über ein Punktesystem erfolgt. Um die Aufgabenstruktur des Wettbewerbs und somit die Anforderungen, die an die Teilnehmenden gestellt werden, herauszuarbeiten, wurde ein Schema für die Klassifikation und Analyse der IJSO-Wettbewerbsaufgaben entwickelt, das im Folgenden vorgestellt wird.

#### 4.2.1 Entwicklung eines Schemas zur Klassifikation und Analyse von Wettbewerbsaufgaben

Als Basis für die Entwicklung eines Schemas zur Klassifikation und Analyse von Wettbewerbsaufgaben diente das Klassifikationsschema nach Petersen (2010), da dies als einziges der dargestellten Schemata für Aufgaben eines Schülerwettbewerbs entwickelt wurde. Es sieht eine Einteilung in Oberflächen- und Tiefenmerkmale vor und enthält neben gängigen Merkmalen zur Charakterisierung von Aufgaben schwierigkeiterzeugende sowie wettbewerbsspezifische Merkmale. Zu den Oberflächenmerkmalen zählen hier das erwartete Ergebnis, die Aufgabenstellung und das Stoffgebiet, während die Tiefenmerkmale die Lösungskomplexität und die fachlichen Anforderungsmerkmale berücksichtigen (Petersen, 2010). Hintergrund der Erstellung dieses Klassifikationsschemas war, im Gegensatz zur vorliegenden Arbeit, der Wunsch nach der Untersuchung der diagnostischen Güte der Aufgaben der IPhO sowie die Suche nach Möglichkeiten zur Beurteilung der Qualität von Aufgaben in dem Wettbewerb (Petersen, 2010). Das Vorhaben wurde allerdings nach der Validierung des Schemas (Kröger, 2011) nicht weiter verfolgt.

Da die Klassifikation und Analyse von Wettbewerbsaufgaben der IJSO der Ermittlung der Anforderungen dieses Wettbewerbs dienen sollte, wurden lediglich ausgewählte Kriterien des Klassifikationsschemas von Petersen (2010) in der Entwicklung eines adaptierten Klassifikationsschemas berücksichtigt. Die Einteilung in Oberflächen- und Tiefenmerkmale wurde übernommen.

- Die **Oberflächenmerkmale** beschreiben formale sowie inhaltliche Merkmale einer Aufgabe und haben eine rein deskriptive Funktion. Berücksichtigt wurden die drei Kriterien Aufgabenbeantwortung, Fach und Themengebiet.
- Die **Tiefenmerkmale** beschreiben die Anforderungen, die die jeweilige Aufgabe an die Teilnehmenden stellt. Hierunter fallen die Lösungskomplexität, die von mehreren Autoren (z. B. H. E. Fischer, 1994; Kauertz, 2008; Williams & Clarke, 1997) als schwierigkeiterzeugendes Merkmal von Aufgaben identifiziert wurde, sowie der Kompetenzbereich gemäß der Bildungsstandards (z. B. Kultusministerkonferenz, 2005b).

Das im Rahmen dieser Arbeit entwickelte Schema zur Klassifikation und Analyse von Wettbewerbsaufgaben ist in Abbildung 4.1 dargestellt. Die einzelnen Kriterien der Oberflächen- und Tiefenmerkmale werden im Folgenden erläutert.

##### Kriterien zur Analyse der Oberflächenmerkmale

Die Oberflächenmerkmale beinhalten die drei Kriterien Aufgabenbeantwortung, Fach und Themengebiet. Die Klassifizierung erfolgte für jede Teilaufgabe anhand ihrer durch den Wettbewerb vorgegebenen Musterlösung.

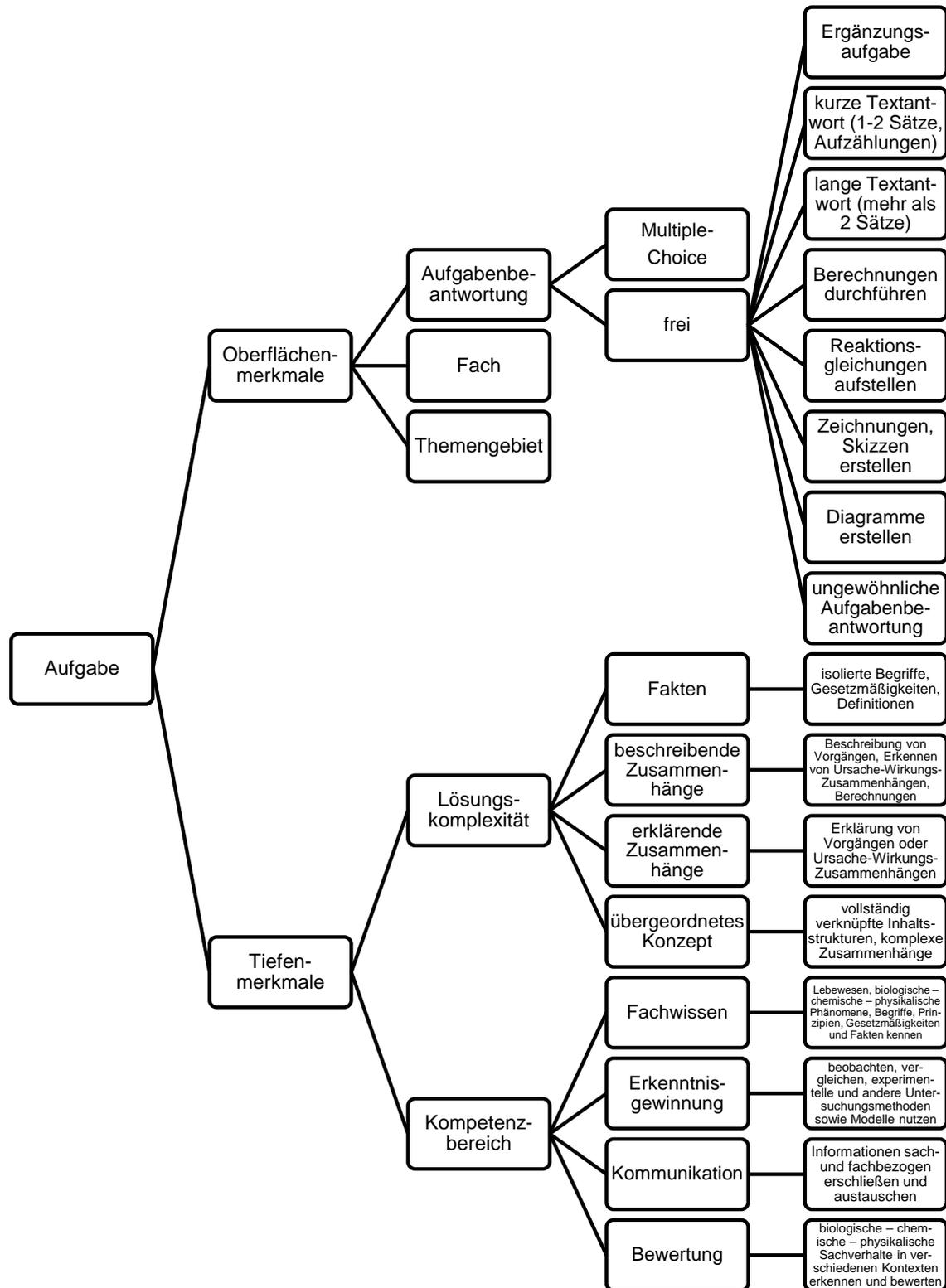


Abbildung 4.1: Schema zur Klassifikation und Analyse von Wettbewerbsaufgaben

Das Kriterium **Aufgabenbeantwortung** beschreibt, in welcher Form die Antwort auf eine Aufgabe erfolgen soll, wobei zunächst eine Unterscheidung zwischen gebundenem, d. h. Multiple-Choice, und freien Formaten vorgenommen wurde. Die freien Formate wurden noch einmal in acht verschiedene Kategorien aufgegliedert. Diese acht Kategorien orientierten sich zunächst an Bühner (2011). Im Laufe der Entwicklung des Schemas wurden die verschiedenen Möglichkeiten der Aufgabenbeantwortung jedoch an die Wettbewerbsaufgaben angepasst. Die Zuordnung einer Aufgabe zu einer Möglichkeit der Aufgabenbeantwortung orientierte sich dabei immer an der Musterlösung. Die Wahlmöglichkeiten für das Kriterium Aufgabenbeantwortung werden im Folgenden kurz erläutert.

- **Multiple-Choice:** Hierbei handelt es sich um ein gebundenes Format, bei dem aus mehreren Antwortalternativen eine oder mehrere richtige Antworten ausgewählt werden mussten. Es war unter den unterschiedlichen Möglichkeiten der Aufgabenbeantwortung das einzige gebundene Format.
- **Ergänzungsaufgabe:** Bei Ergänzungsaufgaben mussten Begriffe, Definitionen u.ä. in Tabellen oder Abbildungen ergänzt werden.
- **Kurze Textantwort:** Bei einer kurzen Textantwort handelte es sich um eine Antwort, die nicht mehr als zwei Sätze umfasste.
- **Lange Textantwort:** Lange Textantworten umfassten mehr als zwei Sätze.
- **Berechnungen durchführen:** Alle Aufgaben, in denen Schülerinnen und Schüler rechnerisch die Lösung der Aufgabe ermitteln sowie den Rechenweg darstellen sollten, wurden dieser Aufgabenbeantwortung zugeordnet.
- **Reaktionsgleichungen aufstellen:** Diese Art der Aufgabenbeantwortung beinhaltete das Aufstellen, Ergänzen oder Ausgleichen von Wort- und Reaktionsgleichungen.
- **Zeichnungen/ Skizzen erstellen:** War das Anfertigen einer Abbildung, Zeichnung oder Skizze verlangt, wurde dieser Typ der Aufgabenbeantwortung gewählt.
- **Diagramme erstellen:** Hier war das Anfertigen eines Diagramms gefordert.
- **Ungewöhnliche Aufgabenbeantwortung:** Hierunter fielen alle Aufgaben, die sich den anderen Möglichkeiten der Aufgabenbeantwortung nicht eindeutig zuordnen ließen. Zudem erfolgte eine Zuordnung hierzu, wenn eine Aufgabe mehrere Arten der Aufgabenbeantwortung umfasste. Auch bei Lückentexten, Richtig-Falsch-Aufgaben und dem Entwerfen von Nahrungsnetzen oder Stammbäumen wurde diese Aufgabenbeantwortung gewählt.

Das Kriterium **Fach** bezeichnet das (naturwissenschaftliche) Fach, dem die Aufgabe zuzuordnen ist. Im Falle der IJSO waren die Wettbewerbsaufgaben ab der zweiten Runde bereits für die Teilnehmenden fachlich gekennzeichnet und in die Fächer Biologie, Chemie und Physik eingeteilt. Aus diesem Grund wurde keine weitere fachliche Prüfung der Zuordnung vorgenommen, sondern die existierende Einteilung übernommen. Sofern keine eindeutige Fächerzuordnung vorlag, bspw. bei Aufgaben der ersten Wettbewerbsrunde, wurde eine doppelte Zuordnung (zu zwei Fächern) vorgenommen.

Das Kriterium **Themengebiet** bezeichnet das fachliche Gebiet, dem der Inhalt der Aufgabe zuzuordnen ist. Die Zuordnung zu einem Themengebiet erfolgte nach der konkreten Aufgabenstellung, unabhängig vom Thema der übergeordneten Aufgabe. Umfasste eine Aufgabe mehrere Themen (z. B. Allgemeine Chemie und Biochemie), so wurde diese Aufgabe auch mit beiden Themen kodiert. Um eine einheitliche Zuordnung zu gewährleisten, wurde im Laufe der Entwicklung des Klassifikationsschemas ein Themenkatalog entwickelt. Die Einteilung der Themengebiete basierte im Wesentlichen auf Schwarzlow (2014). Die Biologie-Themen wurden leicht verändert nach Markl und Sadava (2011). Die Chemie-Themen sind an Themengebiete gängiger Fachliteratur und Schulbücher wie Arnold et al. (2010), Atkins und de Paula (2013), Bitter, Sommer und Tompert (2010), Holleman, Wiberg und Wiberg (2007) sowie Vollhardt und Schore (2012) angelehnt. Die Physik-Themen wurden in Anlehnung an Kuchling (2011) zugeordnet, die Themenübersicht über alle Fächer ist in Tabelle 4.3 dargestellt.

### Kriterien zur Analyse der Tiefenmerkmale

Die Tiefenmerkmale beinhalten die zwei Kriterien Lösungskomplexität und Kompetenzbereich. Das Kriterium **Lösungskomplexität** kennzeichnet die Komplexität der in der Aufgabe geforderten Lösung. Die Komplexität als Maß für die inhaltliche Struktur einer Aufgabe betrachtet nach Kauertz (2008) zwei Komponenten, die Anzahl der Elemente (quantitative Komponente) sowie die Art der Beziehung (qualitative Komponente). Die Komplexität steigt, je mehr Elemente eine Aufgabe enthält und je verknüpfter diese sind (Kauertz, 2008, S. 37). Um ein Komplexitätsschema für die zugrundeliegende Arbeit zu entwickeln, wurde zunächst ein vierstufiges Komplexitätsschema angenommen, dessen Stufen „ein Fakt - mehrere Fakten - Zusammenhänge - übergeordnetes Konzept“ zunächst aus vorliegenden Arbeiten (Bernholt et al., 2009; Kauertz, 2008; Neumann et al., 2007; Sumfleth et al., 2013) generiert wurden. Nach einer ersten Durchsicht klassifizierter Wettbewerbsaufgaben zeigte sich jedoch, dass die erste Stufe „ein Fakt“ kaum in den Wettbewerbsaufgaben enthalten war, so dass diese Stufe aus dem Komplexitätsschema entfernt wurde. Des Weiteren zeigte sich, dass die Stufe „Zusammenhänge“ weiter ausdifferenziert werden musste, da sich andeutete, dass die Lösungskomplexität vieler Aufgaben in diesem Bereich lag. Im Gegensatz zu den von Kauertz, Fischer, Lau und Neumann (2008) und Neumann et al. (2007) vorgeschlagenen drei Komplexitätsstufen „Zusammenhang“, „mehrere unverbundene Zusammenhänge“ und „mehrere verbundene Zusammenhänge“ wurden jedoch nur zwei Komplexitätsstufen, „beschreibende Zusammenhänge“ und „erklärende Zusammenhänge“, in das Komplexitätsschema aufgenommen, da diese Ausdifferenzierung der Stufe „Zusammenhänge“ eine qualitative Komponente berücksichtigte. Das für die Klassifikation und Analyse der Wettbewerbsaufgaben berücksichtigte Komplexitätsschema ist mit Erläuterungen in Tabelle 4.4 abgebildet.

Beinhaltete eine Aufgabe mehrere Arbeitsaufträge, wurde sie nach dem Arbeitsauftrag, der die höchste Komplexität verlangte, klassifiziert. Multiple-Choice-Aufgaben wurden ausschließlich den beiden Lösungskomplexitäten „Fakten“ oder „beschreibende Zusammenhänge“ zugeordnet. Die Stufen „erklärende Zusammenhänge“ und „übergeordnetes Konzept“ wurden für Multiple-Choice-Aufgaben grundsätzlich ausgeschlossen, da die Antwort der Aufgabe nicht von der Person, die die Aufgabe beantworten sollte, selbst generiert werden musste. Es wurde hier lediglich die Auswahl einer korrekten Antwort gefordert, wobei zwischen den Antwortmöglichkeiten abgewogen werden musste.

Tabelle 4.3: **Übersicht über die Themengebiete des Schemas zur Klassifikation und Analyse von Wettbewerbsaufgaben** Die Einteilung der Themengebiete basiert im Wesentlichen auf Schwarzlow (2014).

	<i>Oberthema</i>	<i>Zugehörige Unterthemen</i>
Biologie	Zellbiologie	Moleküle (Aminosäuren, Fette, Kohlenhydrate etc.), Zellen (tierische und pflanzliche), Organellen, Membranen, Enzyme und Stoffwechsel, Fotosynthese
	Information und Vererbung (Genetik)	Chromosomen, Zellteilung, Genetik (Mendel etc.), DNA, Genetik von Bakterien und Viren, eukaryotisches Genom und seine Expression, Signalübertragung und Kommunikation zwischen Zellen, Gentechnik und Biotechnologie, Immunbiologie
	Entwicklungsbiologie	Menschliche und tierische Embryonalentwicklung
	Evolution	Geschichte des Lebens auf der Erde, Mechanismen der Evolution, Arten und ihre Entstehung, Stammbäume
	Botanik	Pflanzenkörper, Transport in Pflanzen, Mineralstoffhaushalt, Regulation des Wachstums, Fortpflanzung bei Blütenpflanzen, Reaktion von Pflanzen auf Umweltstress
	Tierphysiologie	Physiologie, Homöostase, Thermoregulation, Hormone von Tieren, Gewebe, Fortpflanzung bei Tieren, Nervensystem, Gasaustausch bei Tieren, Kreislaufsysteme, Ernährung, Verdauung, Resorption, Salzhaushalt, Wasserhaushalt, Stickstoffausscheidung
	Ökologie	Verhaltensökologie, Populationsökologie, Lebensgemeinschaften und Ökosysteme, Biogeografie, Naturschutzgebiete
	Mikrobiologie	Bakterien, Viren, Pilze
Chemie	Allgemeine Chemie	Atome und Atombau, Periodensystem der Elemente, Elemente, Verbindungen, Bindungen, Säuren und Basen, Salze, chemische Reaktion, Berechnungen, Chemiehistorie
	Anorganische Chemie	Minerale, Metalle, Komplexe, Stoffkreisläufe, Kristalle
	Organische Chemie	Kohlenwasserstoffe (Alkane, Alkene, Alkine), Organische Sauerstoffverbindungen, Oxidationszahlen in organischen Molekülen, Reaktionsmechanismen (Eliminierung, Substitution, ...) Aromaten, Kunststoffe, Waschmittel und Tenside, Farbstoffe
	Physikalische Chemie	Thermodynamik, Elektrochemie, Reaktionskinetik, Katalysatoren, ideale Gase
	Analytische Chemie	Nachweise und Nachweisreaktionen, Experimente, Analysen von Stoffen
	Biochemie	Stoffwechselforgänge, Vorgänge in Zellen (z. B. Osmose)
	Technische Chemie	Großtechnische Verfahren, Herstellungsverfahren von Lebensmitteln, Reaktionen und Reaktionsgleichungen in technischen Verfahren
Physik	Größen und Einheiten	Physikalische Größen, Gleichungen physikalischer Größen, Internationale Einheiten (SI)
	Mechanik	Statik des starren Körpers, Kinematik, Dynamik, Hydrostatik (ruhende Flüssigkeiten), Aerostatik (ruhende Gase), Strömungen, Moleküle, Osmose, Elastizität fester Körper, Planeten
	Wärmeenergie	Temperatur, Wärmeenergie, Aggregatzustände, Zustandsänderung idealer Gase, Kinetische Wärmetheorie, Wärmetransport
	Schwingungen und Wellen	Mechanische Schwingungen, Mechanische Wellen
	Akustik	Schallerzeugung, Schallausbreitung, Schallmessung, Ultraschall
	Optik	Geometrische Optik, Wellenoptik, Lichtstrahlung
	Elektrizitätslehre	Gleichstromkreis, Elektrisches Feld, Magnetisches Feld, Elektrische Maschinen, Wechselstromkreis, Elektrische Leitung, Elektrische Schwingungen und Wellen

Tabelle 4.4: Schema der Lösungskomplexität für die Klassifikation und Analyse von Wettbewerbsaufgaben

<i>Komplexitätsstufe</i>	<i>Erläuterung</i>
4 übergeordnetes Konzept	Vollständig verknüpfte Inhaltsstrukturen, komplexe Zusammenhänge
3 erklärende Zusammenhänge	Erklärung von Vorgängen oder Ursache-Wirkungs-Zusammenhängen
2 beschreibende Zusammenhänge	Beschreibung von Vorgängen, Erkennen von Ursache-Wirkungs-Zusammenhängen, Berechnungen
1 Fakten	Isolierte Begriffe, Gesetzmäßigkeiten, Definitionen

Das Kriterium **Kompetenzbereich** stellt die Zuordnung der Aufgabe zu einem Kompetenzbereich gemäß den Bildungsstandards des jeweiligen Faches dar (Kultusministerkonferenz, 2005a, 2005b, 2005c). Die Zuordnung der Kompetenzbereiche zu den Tiefen- statt zu den Oberflächenmerkmalen lag darin begründet, dass nach Hammann (2006) die einzelnen Kompetenzbereiche auch mit verschiedenen Anforderungen an das Lösen der Aufgabe verbunden werden können. Für die Zuordnung der Tiefenmerkmale war maßgebend, welche Anforderungen die Aufgabe stellte, und nicht, welche Voraussetzungen zum Lösen der Aufgabe gegeben sein mussten. Die fachspezifischen Definitionen der einzelnen Kompetenzbereiche für die Fächer Biologie, Chemie und Physik sind Abschnitt 2.3 zu entnehmen. Wie in Abschnitt 2.3 bereits erläutert, sind die fachspezifischen Definitionen der einzelnen Kompetenzbereiche für die drei Naturwissenschaften nicht identisch. Für die Klassifikation und Analyse der Wettbewerbsaufgaben wurde jedoch angestrebt, eine vereinheitlichte Klassifikation vorzunehmen, weshalb die Beschreibungen der Kompetenzbereiche für die einzelnen Fächer, wie in Abbildung 4.1 ersichtlich, zusammengefasst und somit vereinheitlicht wurden. Deckte eine Aufgabe mehrere Kompetenzbereiche ab, wurde sie dem Kompetenzbereich zugeordnet, der die Aufgabe schwerpunktmäßig repräsentierte. Sofern kein Schwerpunkt erkennbar war, wurde die Aufgabe in weitere Teilaufgaben unterteilt.

Sonderfälle in der Zuordnung zu einer Komplexitätsstufe oder zu einem Kompetenzbereich sowie Sonderfälle einzelner Aufgaben, die im Laufe der Klassifikation der Aufgaben dokumentiert wurden, finden sich mit begründeter Zuordnung in Anhang A.

#### 4.2.2 Validierung des Schemas zur Klassifikation und Analyse von Wettbewerbsaufgaben

In einem iterativen Verfahren wurde das Schema zur Klassifikation und Analyse von Wettbewerbsaufgaben validiert, d. h. die einzelnen Kriterien und Beispiele bereits klassifizierter Aufgaben wurden von Expertinnen und Experten wiederholt geprüft. Darüber hinaus wurden die Zuordnungen der Tiefenmerkmale hinsichtlich Reliabilität und Objektivität untersucht.

Oberflächen- und Tiefenmerkmale wurden zunächst in Zusammenarbeit mit Fachdidaktikern sowie der Wettbewerbsleitung der IJSO und der IPhO diskutiert und bewertet. Erste Klassifikationen wurden, wiederum in Zusammenarbeit mit Fachdidaktikern sowie der Wettbewerbsleitung der IJSO, überprüft und Defizite in den Kriterien und ihren Beschreibungen

identifiziert. Daraufhin wurde das Schema überarbeitet und ergänzt. Des Weiteren wurde ein Leitfaden entwickelt, in dem besondere Klassifikationen und Erläuterungen fortlaufend dokumentiert wurden.

Für die Zuordnungen zu einem Kompetenzbereich wurde die Beurteilerübereinstimmung berechnet. Die Beurteilerübereinstimmung gibt an, inwieweit verschiedene Objekte durch verschiedene Kodierer exakt gleich beurteilt werden (Wirtz & Caspar, 2002). Sind die Unterschiede zwischen den Urteilen verschiedener Kodierer vernachlässigbar klein, kann von einer reliablen und objektiven Beurteilung ausgegangen werden (Wirtz & Caspar, 2002). Dabei gibt es verschiedene Möglichkeiten, die Beurteilerübereinstimmung zu berechnen. Für polytome Kategoriensysteme, wie sie hier vorliegen, sind die gebräuchlichsten Maße die prozentuale Übereinstimmung sowie Cohens Kappa. Die prozentuale Übereinstimmung gibt dabei lediglich an, zu welchem Anteil in Prozent die Urteile exakt übereinstimmen (vgl. Wirtz & Caspar, 2002). Dabei wird nicht berücksichtigt, inwiefern die berechnete Häufigkeit der Übereinstimmung von einer zu erwartenden Häufigkeit von Zufallsübereinstimmungen beeinflusst wird, weshalb die Angabe eines zufallskorrigierten Maßes, in diesem Fall von Cohens Kappa, ergänzt werden sollte. Cohens Kappa kann Werte zwischen  $-1$  (völlig unterschiedliche Einschätzung von Objekten) und  $+1$  (perfekte Übereinstimmung) annehmen. Nach Wirtz und Caspar (2002, S. 59) können Werte für Cohens Kappa von  $\kappa > .75$  als sehr gute, Werte zwischen  $\kappa = .60$  und  $\kappa = .75$  als gute und Werte zwischen  $\kappa = .40$  und  $\kappa = .60$  als akzeptable Werte angesehen werden. Dabei ist jedoch zu beachten, dass der Wert für Cohens Kappa nicht nur von der Übereinstimmung, sondern auch von Merkmalen wie z. B. dem Rateverhalten oder der Anzahl der Skalenkategorien abhängen kann. Die Beurteilung von Cohens Kappa sollte daher immer in Abhängigkeit von den zu beurteilenden Objekten und Merkmalen vorgenommen werden (vgl. Wirtz & Caspar, 2002). Für das Kriterium Kompetenzbereich wurde für 182 doppelt kodierte Aufgaben (13.9%) eine prozentuale Übereinstimmung von 91.8% ermittelt, für Cohens Kappa ergab sich ein Wert von  $\kappa = .75$ . Nach Wirtz und Caspar (2002) ist dies als eine gute bis sehr gute Übereinstimmung zu sehen.

Zur Prüfung von Reliabilität und Objektivität der Zuordnungen zum Kriterium Lösungskomplexität wurde die Beurteilerreliabilität ermittelt. Der Unterschied zur Beurteilerübereinstimmung, wie sie für die Zuordnungen zu einem Kompetenzbereich verwendet wurde, liegt darin, dass Maße für die Beurteilerreliabilität nicht bestimmen, inwieweit verschiedene Objekte durch verschiedene Kodierer exakt gleich beurteilt werden, sondern inwieweit verschiedene Objekte durch verschiedene Kodierer ähnlich weit unter bzw. über dem Durchschnitt der untersuchten Stichprobe liegen (Wirtz & Caspar, 2002). Dies bedeutet ferner, dass die Beurteilerübereinstimmung die zu beurteilenden Daten grundsätzlich als nominalskaliert ansieht, wodurch bei nicht nominalskalierten Daten Informationen verloren gehen (Wirtz & Caspar, 2002, S. 36). Aus den o. a. Gründen und der Tatsache, dass für die Lösungskomplexität ordinal skalierte Daten vorlagen, wurde die Beurteilerreliabilität berechnet. Als Maß für die Beurteilerreliabilität wurde die Spearman-Rangkorrelation  $\rho$  als eines der gebräuchlichsten Maße für die Korrelation von mindestens ordinalskalierten Urteilen von zwei Ratern gewählt (Wirtz & Caspar, 2002, S. 133). Die Spearman-Rangkorrelation kann Werte zwischen  $-1$  (vollständiger negativer Zusammenhang) über  $.00$  (kein Zusammenhang) bis  $+1$  (vollständiger positiver Zusammenhang) annehmen. Nach Heller und Rosemann (1981, S. 122) werden Korrelationskoeffizienten zwischen  $.00$  bis  $.40$  als niedriger Zusammenhang, zwischen  $.40$  bis  $.70$  als mittlerer Zusammenhang und Korrelationskoeffizienten größer  $.70$  als hoher Zusammenhang interpretiert.

tiert. Für das Kriterium Lösungskomplexität zeigte sich für 182 doppelt kodierte Aufgaben (13.9%) eine signifikante Spearman-Rangkorrelation von  $\rho = .88$ ,  $p < .001$ . Nach Heller und Rosemann (1981, S. 122) ist dieser Wert als ein hoher Zusammenhang zu interpretieren.

### 4.2.3 Beispiele: Klassifikation ausgewählter Aufgaben der ersten Runde der Internationalen JuniorScienceOlympiade aus 2012 und 2014

Als repräsentative Beispiele wurden die vierte Aufgabe der ersten Runde der IJSO 2012 sowie die erste Aufgabe der IJSO 2014 ausgewählt, die in Abbildung 4.2 dargestellt und online unter <http://wettbewerbe.ipn.uni-kiel.de/ijso> verfügbar sind (Zugriffsdatum 09.12.2016). Die Musterlösungen zu Aufgabe 4 aus 2012 und Aufgabe 1 aus 2014, die wie die Aufgaben unter oben aufgeführter Adresse online verfügbar sind (Zugriffsdatum 09.12.2016), sind in Anhang B einzusehen.

#### Aufgabe 4 der ersten Runde 2012

Da Aufgabe 4 vier Teilaufgaben umfasste, wurde jede Teilaufgabe für sich klassifiziert. Zunächst wurden die Oberflächenmerkmale zugeordnet:

- **Aufgabenbeantwortung** Die Aufgaben 4a und 4c sind frei mit einer langen Textantwort, d. h. in mehr als zwei Sätzen, zu beantworten. Die Aufgaben 4b und 4d wurden der Kategorie ungewöhnliche Aufgabenbeantwortung zugeordnet: Aufgabe 4b erfordert das Notieren der Beobachtungen des durchgeführten Experiments und das Erklären der Ergebnisse. Da das Notieren der Beobachtungen aber in jeglicher Form erfolgen kann (Stichworte, in tabellarischer Form, Fotodokumentation) wurde die Kategorie „ungewöhnliche Aufgabenbeantwortung“ gewählt. Bei 4d liegt die Wahl dieser Form der Aufgabenbeantwortung darin begründet, dass neben dem Finden einer Begründung und dem Erläutern der chemischen Prozesse eine Reaktionsgleichung formuliert werden soll. Somit werden hier mehrere Kategorien bedient.
- **Fach** Als Fach wurde für alle Teilaufgaben Chemie gewählt.
- **Themengebiet** Alle Aufgaben wurden dem Themengebiet Organische Chemie zugeordnet.

Für die Tiefenmerkmale wurden die folgenden Zuordnungen getroffen:

- **Lösungskomplexität** Alle Aufgaben wurden der Lösungskomplexität „erklärende Zusammenhänge“ zugeordnet, da jede Teilaufgabe eine Erklärung, Begründung oder Erläuterung einfordert.
- **Kompetenzbereich** Aufgabe 4a wurde dem Kompetenzbereich Erkenntnisgewinnung zugeordnet, da diese Aufgabe eine experimentelle Komponente enthält. Aufgabe 4b wurde dem Kompetenzbereich Kommunikation zugeordnet, da sich primär Informationen erschlossen werden müssen, mit denen begründet werden kann, weshalb Stärke für die Herstellung von Papier ungeeignet ist. Die Aufgaben 4c und 4d wurden beide dem Kompetenzbereich Fachwissen zugeordnet, da für die Beantwortung der Aufgaben chemische Phänomene, Begriffe und Gesetzmäßigkeiten bekannt sein müssen.

**EXPERIMENT – „VOM HOLZ ZUM PAPIER“**

Zellwände von Pflanzen und Hölzern bestehen aus Cellulose. Der aus Holz gewonnene Zellstoff ist ein wichtiger Rohstoff für die Papierherstellung. Schreibpapier wird im Gegensatz zu Filter- oder Toilettenpapier geleimt.

Miss genau die Länge und Breite eines DIN A<sub>4</sub>-Blattes. Weiche das Blatt für einige Minuten in Wasser ein und bestimme die Größe des noch nassen Blattes. Wiederhole deine Messung, nachdem das Papier getrocknet ist. Schreibe mit Tinte einige Worte auf das so behandelte Blatt sowie auf ein neues DIN A<sub>4</sub>-Blatt.

**AUFGABE 4**

- Notiere deine Beobachtungen und erkläre die Veränderungen.
- Cellulose besteht wie Stärke aus dem Grundbaustein Glucose. Informiere dich über die Struktur von Cellulose und Stärke. Begründe, weshalb sich Stärke für die Papierherstellung nicht eignet.
- Verschiedene Prozesse tragen zum „Altern“ von Papier bei. Bibliotheken schützen ihre wertvollen Bücher durch verschiedene Maßnahmen: Sie werden in dunklen und klimatisierten Räumen aufbewahrt und dürfen nur mit Handschuhen berührt werden. Nach dem verheerenden Brand in der Herzogin-Anna-Amalia-Bibliothek in Weimar wurden die nach den Löscharbeiten geretteten Bücher gefriergetrocknet. Nenne vier Faktoren, vor denen diese Maßnahmen schützen sollen, und erläutere die zugehörigen Alterungsprozesse.
- Das Papier mit einem Selbstporträt Dürers aus dem 15. Jahrhundert ist einwandfrei erhalten. Jedoch „sterben“ massenweise Bücher und Handschriften aus der zweiten Hälfte des 19. Jahrhunderts. Finde eine Begründung und erläutere die dafür verantwortlichen chemischen Prozesse. Formuliere dazu eine chemische Reaktionsgleichung.



*Selbstbildnis des Dreizehnjährigen, Silberstift auf weiß grundiertem Papier (1484), Selbstporträt Albrecht Dürers.*

© IPN • Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik an der Universität Kiel

**AUFGABE 1**

Häufig wird gereinigte und gekämmte Schafswolle zu Wollgarnen und -stoffen versponnen und verwebt. Es werden aber auch andere tierische und pflanzliche Fasern für die Herstellung von Textilien genutzt.

Nimm ein Garn aus Baumwolle und eines aus reiner Schurwolle. Ziehe die Fäden des Garns auseinander und zupfe vorsichtig mit einer Pinzette einzelne Fasern heraus. [Tipp: Lege die zu untersuchenden Fasern auf schwarzes Tonpapier oder Millimeterpapier.]

Materialien: Garne aus reiner Baumwolle, Schurwolle, Lupe oder Binokular, Schere, Pinzette.

- Nenne drei tierische und drei pflanzliche Fasern. Gib in einer Tabelle jeweils an, von welchen Tieren, Pflanzen und Pflanzenteilen (z. B. Wurzel, Stängel, Blatt oder Frucht) sie gewonnen werden.
- Führe das Experiment in Aufgabe 1 durch und fertige jeweils eine Skizze der Fäden und Fasern von Baumwoll- und Schurwollgarn an. Nutze dazu Lupe oder Binokular. Gib je eine Probe der von dir untersuchten Garne zusammen mit deinen Ausarbeitungen ab.
- Bestimme in beiden Garnen die Anzahl der Fäden. Miss die mittlere Faserlänge und vergleiche Oberflächenbeschaffenheit, Dehnbarkeit und Elastizität beider Faserarten. Wo notwendig, nutze Lupe oder Binokular. Fertige eine Vergleichstabelle mit deinen Untersuchungsergebnissen an.

© IPN • Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik an der Universität Kiel

Abbildung 4.2: **Aufgabe 4 der ersten Runde der IJSO 2012 und Aufgabe 1 der ersten Runde der IJSO 2014** Die Aufgaben sind online verfügbar unter <http://wettbewerbe.ipn.uni-kiel.de/ijso> (Zugriffsdatum 09.12.2016) sowie Aufgabe 4 zusätzlich in Peters (2014).

### Aufgabe 1 der ersten Runde 2014

Da Aufgabe 1 drei Teilaufgaben umfasste, wurde jede Teilaufgabe für sich klassifiziert. Zunächst wurden die Oberflächenmerkmale zugeordnet:

- **Aufgabenbeantwortung** Die Aufgaben 1a, 1b und 1c wurden alle der Kategorie „ungewöhnliche Aufgabenbeantwortung“ zugeordnet. Bei Aufgabe 1a wird das Erstellen einer Tabelle gefordert, was zu selten vorkam, als dass eine eigene Kategorie gebildet wurde. Bei Aufgabe 1b wird das Durchführen eines Experiments sowie das Erstellen einer Skizze verlangt. Zur Beantwortung von Aufgabe 1c wird eine experimentelle Untersuchung und das Erstellen einer Tabelle gefordert. Die Aufgaben 1b und 1c lassen sich somit nicht eindeutig einer Kategorie der Aufgabenbeantwortung zuordnen.
- **Fach** Als Fach wurde für alle Teilaufgaben Biologie gewählt.
- **Themengebiet** Alle Aufgaben wurden dem Themengebiet Zellbiologie zugeordnet.

Für die Tiefenmerkmale wurden die folgenden Zuordnungen getroffen:

- **Lösungskomplexität** Aufgabe 1a wurde der Lösungskomplexität „Fakten“ zugeordnet, da hier das Nennen von Faktenwissen im Fokus steht. Aufgabe 1b und Aufgabe 1c wurden der Lösungskomplexität „beschreibende Zusammenhänge“ zugeordnet. In Aufgabe 1b war zwar das Erstellen einer Skizze bzw. Zeichnung gefordert, was im Normalfall der Lösungskomplexität „erklärende Zusammenhänge“ zugeordnet wurde, jedoch wurde hier aufgrund der Einfachheit der Zeichnung im Vergleich zu anderen Aufgaben ähnlichen Inhalts (die Zeichnung wurde nach Vorlage erstellt, es musste kein Sachverhalt o. ä. repräsentiert werden) die niedrigere Stufe der Lösungskomplexität gewählt. Für Aufgabe 1c mussten verschiedene Fasern anhand unterschiedlicher Kriterien beschrieben und verglichen, mögliche Unterschiede jedoch nicht erklärt werden, weshalb hier ebenso die zweite Stufe der Lösungskomplexität gewählt wurde.
- **Kompetenzbereich** Aufgabe 1a wurde dem Kompetenzbereich Fachwissen zugeordnet, da diese Aufgabe das Nennen von Informationen erfordert. Die Aufgaben 1b und 1c wurden dem Kompetenzbereich Erkenntnisgewinnung zugeteilt, da beide Aufgaben eine experimentelle Komponente enthalten.

#### 4.2.4 Deskriptive Ergebnisse der Analyse von Wettbewerbsaufgaben der Internationalen JuniorScienceOlympiade

Für die Analyse der Anforderungen in der IJSO wurden insgesamt 1305 Aufgaben aus den Wettbewerbsjahren 2008 (erstes deutsches Auswahlverfahren) bis einschließlich 2014 klassifiziert. Die Aufgaben entstammten sowohl den drei Runden des deutschen Auswahlverfahrens als auch dem internationalen Wettbewerb. Lediglich für das Wettbewerbsjahr 2014 lagen nur die Aufgaben der ersten und zweiten Wettbewerbsrunde vor. Die experimentellen Klausuren der dritten und (internationalen) vierten Wettbewerbsrunden wurden nicht klassifiziert, da der Fokus zunächst auf den nationalen theoretischen Aufgaben des Wettbewerbs liegen sollte (rein experimentelle Klausuren kommen lediglich in der dritten und vierten Runde des Wettbewerbs vor).

Die Auswertung der Oberflächen- und Tiefenmerkmale erfolgte deskriptiv. Hierzu wurden die relativen Häufigkeiten der einzelnen Merkmale bestimmt. Da die IJSO im Vergleich zu den anderen ScienceOlympiaden ein recht junger Schülerwettbewerb ist und die Aufgaben aus den Anfangsjahren daher möglicherweise nicht repräsentativ waren, wurde die Entwicklung der relativen Häufigkeiten einiger Merkmale zusätzlich über die einzelnen IJSO-Wettbewerbsjahre betrachtet. Hierbei wurden die Wettbewerbsjahre 2008 bis 2013 berücksichtigt, da zum Wettbewerbsjahr 2014 keine vollständigen Daten vorlagen. Da die IJSO-Wettbewerbsrunden unterschiedlich aufgebaut waren und unterschiedliche Funktionen hatten, wurden auch die relativen Häufigkeiten einiger Merkmale über die Wettbewerbsrunden hinweg betrachtet. Dabei wurden nur die ersten drei Wettbewerbsrunden des deutschen Auswahlverfahrens und nicht die internationale Runde betrachtet, da die Anforderungen des deutschen Auswahlverfahrens im Zentrum des Interesses standen.

Für die Oberflächenmerkmale der klassifizierten Aufgaben wurden die relativen Häufigkeiten der Merkmale Aufgabenbeantwortung, Fach und Themengebiet bestimmt. Die Ergebnisse über alle IJSO-Wettbewerbsrunden und IJSO-Wettbewerbsjahre sind in Tabelle 4.5 zusammengefasst.

Was die Aufgabenbeantwortung betrifft, siehe Tabelle 4.5, kamen zum größten Teil (56.3%) Multiple-Choice-Aufgaben vor. Von den freien Formaten her wurde das Durchführen von Berechnungen am häufigsten (11.2%) in den Aufgaben verlangt, gefolgt von der ungewöhnlichen Aufgabenbeantwortung (9.6%) und der langen Textantwort (8.0%). Betrachtet man die Möglichkeiten der Aufgabenbeantwortung über die IJSO-Runden hinweg, ist festzustellen, dass der Anteil an Multiple-Choice-Aufgaben von 0.0% in der ersten Runde über 48.5% in der zweiten und 66.1% in der dritten Wettbewerbsrunde anstieg. Der Anteil an Aufgaben, die das Formulieren langer Textantworten oder eine ungewöhnliche Aufgabenbeantwortung verlangten, fiel hingegen ab: So erforderten in den ersten Wettbewerbsrunden 83.6% der Aufgaben eine lange Textantwort oder eine ungewöhnliche Aufgabenbeantwortung, wohingegen die Anteile in den zweiten und dritten Runden der IJSO bei 23.9% bzw. 13.0% lagen. Das häufigste freie Format der Aufgabenbeantwortung in den ersten Runden war die lange Textantwort, in den zweiten und dritten Runden die ungewöhnliche Aufgabenbeantwortung, gefolgt von der langen Textantwort in den zweiten und der Ergänzungsaufgabe in den dritten Runden. Eine Übersicht ist Abbildung 4.3 zu entnehmen. Betrachtet man die Möglichkeiten der Aufgabenbeantwortung über die Wettbewerbsjahre, ist festzustellen, dass der Anteil der Multiple-Choice-Aufgaben von 61.3% in 2008 zu 48.0% in 2013 abnahm, was mit einer Kürzung der Klausuren einherging. Die Anteile der übrigen freien Antwortformate war hingegen relativ konstant, lediglich in 2013 waren in 17.8% der Aufgaben Berechnungen verlangt, wohingegen in den Vorjahren im Mittel 10.1% der Aufgaben das Durchführen von Berechnungen erforderte.

Der Anteil der Fächer war insgesamt relativ gleichmäßig auf die Aufgaben verteilt (Biologie 32.7%, Chemie 35.1%, Physik 32.2%), was so allerdings auch intendiert war, da die Fächer ab der zweiten Wettbewerbsrunde in der Bewertung gleich gewichtet werden. (In den ersten Wettbewerbsrunden gibt es noch keine strikte Trennung nach Fächern, gleichwohl ist beabsichtigt, die drei Naturwissenschaften in gleichen Anteilen abzubilden.) Aus diesem Grund wurde dieses Merkmal weder über die Wettbewerbsrunden noch über die Wettbewerbsjahre im Detail betrachtet, da Unterschiede in den Anteilen der Aufgaben vor allem durch das Vorhandensein einer unterschiedlichen Anzahl an Teilaufgaben entstanden.

Tabelle 4.5: **Deskriptive Ergebnisse der Oberflächenmerkmale der klassifizierten Aufgaben der IJSO** Die Angaben sind über alle Wettbewerbsrunden und Wettbewerbsjahre zusammengefasst.

<i>Art der Aufgabenbeantwortung</i>	<i>Anteil in Prozent</i>
Multiple-Choice	56.3
Ergänzungsaufgabe	5.2
kurze Textantwort	5.5
lange Textantwort	8.0
Berechnungen durchführen	11.2
Reaktionsgleichungen aufstellen	2.8
Zeichnungen/ Skizzen erstellen	0.9
Diagramme erstellen	0.5
ungewöhnliche Aufgabenbeantwortung	9.6

<i>Fach</i>	<i>Anteil in Prozent</i>
Biologie	32.7
Chemie	35.1
Physik	32.2

<i>Themengebiet</i>	<i>Anteil in Prozent</i>
Akustik	0.6
Akustik, Optik	0.1
Allgemeine Chemie	18.3
Analytische Chemie	4.2
Anorganische Chemie	3.2
Atomphysik	0.9
Biochemie	1.1
Botanik	4.5
Elektrizitätslehre	5.7
Entwicklungsbiologie	0.8
Evolution	1.0
Größen und Einheiten	1.0
Information und Vererbung	3.3
Mechanik	14.6
Mechanik, Wärmelehre	2.1
Ökologie	3.7
Optik	3.2
Organische Chemie	1.7
Physikalische Chemie	5.3
Schwingungen und Wellen	0.8
Technische Chemie	1.3
Tierphysiologie	13.1
Wärmelehre	5.1
Zellbiologie	4.2

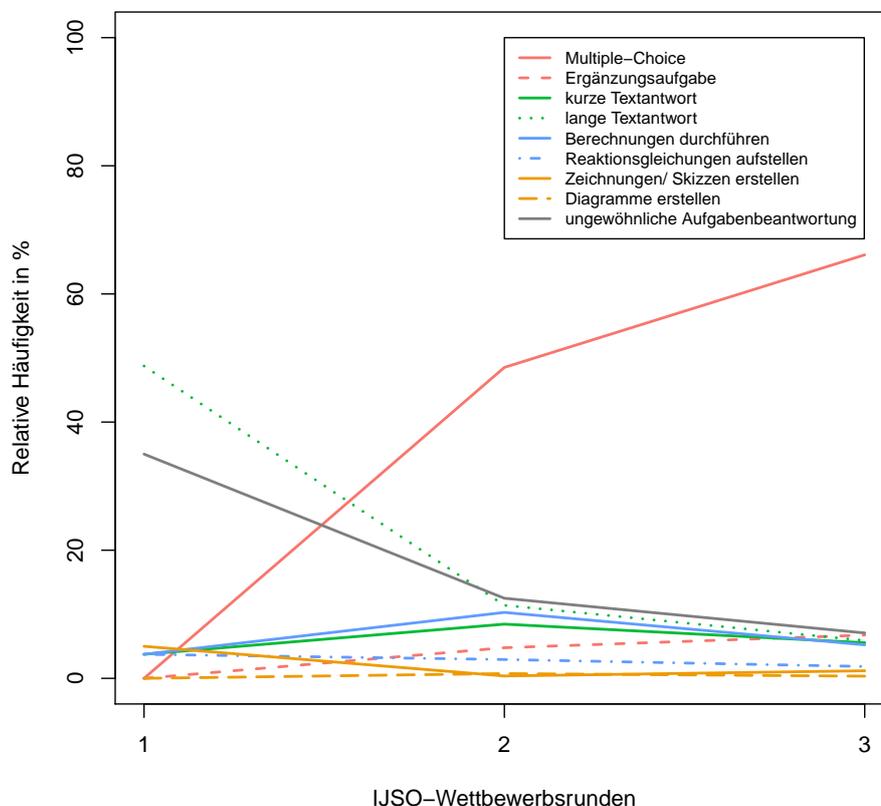


Abbildung 4.3: Verteilung der Aufgabenbeantwortung über die IJSO-Wettbewerbsrunden

Thematisch häufig vertreten waren Aufgaben der Allgemeinen Chemie mit 18.3%, Aufgaben der Mechanik mit 14.6% und Aufgaben aus dem Bereich der Tierphysiologie mit 13.1% (siehe Tabelle 4.5). Betrachtet man die Themengebiete im Detail über die IJSO-Wettbewerbsrunden und die IJSO-Wettbewerbsjahre hinweg, sind keine Auffälligkeiten in der Verteilung festzustellen. Lediglich in den ersten Runden waren die Themen nicht derart breit gestreut wie in den folgenden Wettbewerbsrunden, was dadurch zu erklären ist, dass die Aufgaben der ersten Runden zumeist einem Oberthema unterstehen.

Für die Tiefenmerkmale der klassifizierten Aufgaben wurden, wie für die Oberflächenmerkmale, die relativen Häufigkeiten der einzelnen Stufen der Lösungskomplexität sowie der einzelnen Kompetenzbereiche insgesamt und über die IJSO-Wettbewerbsjahre sowie über die IJSO-Wettbewerbsrunden ermittelt. Die Ergebnisse über die IJSO-Wettbewerbsrunden und IJSO-Wettbewerbsjahre sind in Abbildung 4.4 dargestellt.

Betrachtet man die Verteilung der einzelnen Stufen der Lösungskomplexität über alle klas-

sifizierten Aufgaben hinweg, stellte die zweite Stufe „beschreibende Zusammenhänge“ mit einem Anteil von 70.0% die mit Abstand am häufigsten vertretene Stufe der Lösungskomplexität dar. Es folgte die Stufe „Fakten“ mit 15.7% und die Stufe „erklärende Zusammenhänge“ mit 14.0%. Die höchste Stufe der Lösungskomplexität, „übergeordnetes Konzept“, war mit 0.2%, d. h. insgesamt 3 von 1305 Aufgaben, vertreten. Bei der Verteilung der Kompetenzbereiche zeichnete sich der Kompetenzbereich Fachwissen mit 85.7% als am häufigsten vertretener Kompetenzbereich ab, gefolgt vom Kompetenzbereich Erkenntnisgewinnung mit 10.6%. Der Kompetenzbereich Kommunikation war mit 3.1% vertreten, der Kompetenzbereich Bewertung mit 0.5%, d. h. insgesamt lediglich 4 von 1305 Aufgaben wurden diesem Kompetenzbereich zugeordnet.

In den Analysen über die IJSO-Wettbewerbsjahre hinweg, siehe Abbildung 4.4, bestätigte sich das Ergebnis der Betrachtung aller klassifizierten Aufgaben, dass die zweite Stufe der Lösungskomplexität, „beschreibende Zusammenhänge“, sowie der Kompetenzbereich Fachwissen mit Abstand am stärksten vertreten waren. Die Anteile der Komplexitätsstufen, ausgenommen der Komplexitätsstufe „erklärende Zusammenhänge“, waren relativ konstant (zwischen 12.4 und 18.9% für die Komplexitätsstufe „Fakten“, zwischen 69.1 und 75.4% für die Komplexitätsstufe „beschreibende Zusammenhänge“ und zwischen keiner und einer Aufgabe für die Komplexitätsstufe „übergeordnetes Konzept“). Der Anteil der Komplexitätsstufe „erklärende Zusammenhänge“ variierte zwischen 6.8 und 17.8%. Für das Kriterium Kompetenzbereich ließ sich eine Entwicklung ablesen, da der Anteil des Kompetenzbereichs Fachwissen seit 2011 sank, wobei ein Anteil von 80.7% bislang nicht unterschritten wurde. Im Gegensatz hierzu stieg der Anteil des Kompetenzbereichs Erkenntnisgewinnung seit 2011, ein Anteil von mehr als 14.4% lag bis zum Jahr 2013 allerdings nicht vor.

Bei der Betrachtung der Analysen über die einzelnen IJSO-Wettbewerbsrunden, siehe Abbildung 4.4, fielen zunächst für die Lösungskomplexität als auch für den Kompetenzbereich große Unterschiede zwischen den Wettbewerbsrunden auf. Von der ersten bis zur dritten Wettbewerbsrunde stieg der Anteil an Aufgaben der Lösungskomplexität „Fakten“ und „beschreibende Zusammenhänge“ an (von 2.5% über 17.3% auf 22.4% für die Stufe „Fakten“ und von 35.0% über 62.1% auf 64.6% für die Stufe „beschreibende Zusammenhänge“). Der Anteil der Stufe „erklärende Zusammenhänge“ sank hingegen von 58.5% für die erste Runde über 20.6% und 13.0% für die dritte Runde. Die höchste Stufe der Lösungskomplexität, „übergeordnetes Konzept“, kam lediglich in der ersten Wettbewerbsrunde zu einem Anteil von 3.8% vor. Für den Kompetenzbereich bot sich ein grafisch ähnliches Bild (siehe Abbildung 4.4). Der Anteil des Kompetenzbereichs Fachwissen stieg über die Wettbewerbsrunden von 35.0% über 85.7% auf 89.2% in der dritten Runde an. Die Anteile der Kompetenzbereiche Erkenntnisgewinnung und Kommunikation fielen hingegen ab (von 46.3% über 10.7% auf 7.4% für den Kompetenzbereich Erkenntnisgewinnung und für den Kompetenzbereich Kommunikation von 15.0% für die erste Wettbewerbsrunde über 2.9% auf 3.0%). Der Kompetenzbereich Bewertung sank ebenso von 3.8% über 0.7% auf 0.3% für die dritte Wettbewerbsrunde des deutschen Auswahlverfahrens der IJSO.

#### 4.2.5 Interpretation und Diskussion der deskriptiven Ergebnisse

Die Analyse der klassifizierten Aufgaben der IJSO ergab, dass hauptsächlich Aufgaben im Multiple-Choice-Format vorlagen (56.3%). Dieser große Anteil lässt sich dadurch erklären,

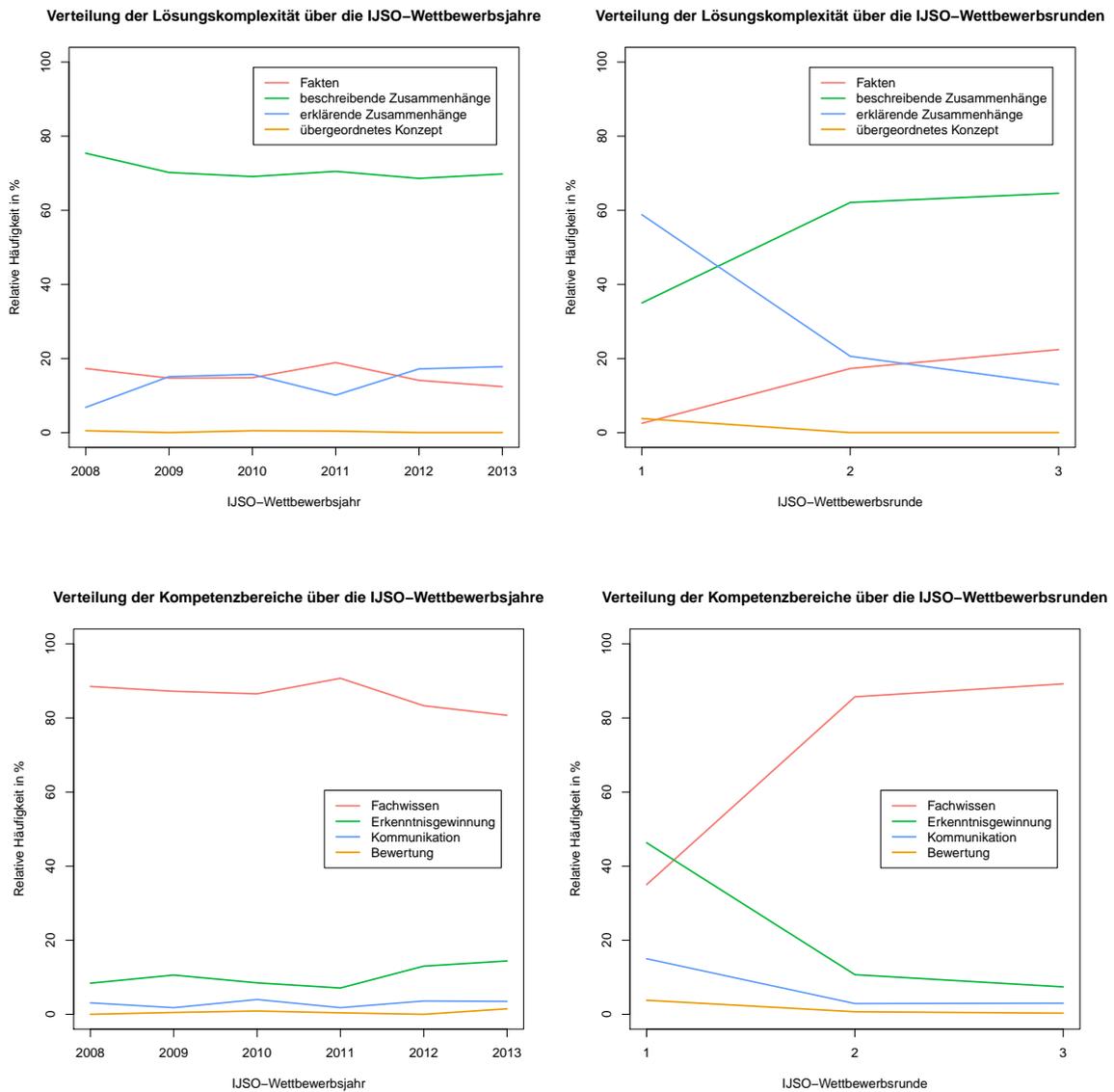


Abbildung 4.4: **Ergebnisse der Analyse von Wettbewerbsaufgaben der IJSO** Abgebildet sind die Verteilungen der Lösungskomplexität sowie der Kompetenzbereiche über die Wettbewerbsrunden sowie über die IJSO-Wettbewerbsjahre.

dass es bis 2015 ab der zweiten Wettbewerbsrunde Multiple-Choice-Aufgaben innerhalb der Klausur bzw. eigenständige Multiple-Choice-Klausuren ab der dritten Wettbewerbsrunde gab. Im freien Aufgabenformat wurden hauptsächlich Berechnungen verlangt (11.2%). Allerdings kommt dieser Wert vor allem dadurch zustande, dass im Jahr 2013 im Vergleich zu den Vorjahren außergewöhnlich viele Aufgaben das Durchführen von Berechnungen erforderten. An zweiter und dritter Stelle der Häufigkeit unter den Möglichkeiten der Aufgabenbeantwortung standen die ungewöhnliche Aufgabenbeantwortung (9.6%) und die lange Textantwort (8.0%). Diese beiden Formate kamen auch über die einzelnen Wettbewerbsrunden betrachtet am häufigsten vor. Die am häufigsten vertretenen Themengebiete waren Allgemeine Chemie, Mechanik und Tierphysiologie, wobei anzumerken ist, dass die Themengebiete noch relativ weit gefasst wurden.

Die Analyse der Tiefenmerkmale ergab, dass die zweite Stufe „beschreibende Zusammenhänge“ des Kriteriums Lösungskomplexität mit Abstand am häufigsten, gefolgt von der ersten Stufe „Fakten“, in den Aufgaben der IJSO vorkam. Der mit Abstand am häufigsten vertretene Kompetenzbereich war der Bereich Fachwissen, gefolgt vom Kompetenzbereich Erkenntnisgewinnung.

Der große Anteil an Aufgaben mit der zweitniedrigsten Stufe der Lösungskomplexität, „beschreibende Zusammenhänge“, lässt zunächst vermuten, dass die Wettbewerbsaufgaben der IJSO als „eher leicht“ einzustufen seien. Bei der Interpretation sind allerdings mehrere Aspekte zu beachten. So berücksichtigt das Schema zur Klassifikation und Analyse von Wettbewerbsaufgaben nicht das Alter oder die Klassenstufe der Teilnehmenden. Dies gewinnt in dem Moment an Bedeutung, in dem die Wettbewerbsaufgaben Inhalte abfragen, die nicht oder noch nicht im Unterricht der Teilnehmenden behandelt wurden. So gibt es auch schon in der ersten Wettbewerbsrunde Aufgaben aus bspw. dem Bereich der organischen Chemie. Inhalte der organischen Chemie werden jedoch selten im Anfangsunterricht vermittelt. Andere Klassifikationsschemata wie z. B. jenes von Bloom et al. (1956) verstehen dies schon als eine erste Transferstufe (Stufe 3 *application*, siehe Tabelle 4.1). Somit wäre zu überlegen, ein weiteres Kriterium in das Schema aufzunehmen, das berücksichtigt, ob das Thema der Aufgabe einen Bezug zum Lehrplan hat. Hier ergäbe sich jedoch ein neues Hindernis: Da die IJSO bundesweit ausgetragen wird und Schülerinnen und Schüler verschiedener Alters- und Klassenstufen an dem Wettbewerb teilnehmen, könnte eine Klassifikation lediglich für ein spezifisches Bundesland sowie eine einzelne Klassenstufe und einzeln für Biologie, Chemie und Physik erfolgen. Eine allgemeingültige Klassifikation wäre dementsprechend kaum möglich. Die Einführung eines derartigen Kriteriums könnte allerdings wettbewerbsintern oder für betreuende Lehrkräfte von Bedeutung sein, die somit ihre Schülerinnen und Schüler gezielter auf die Teilnahme an der IJSO vorbereiten könnten. Für die Analyse der Anforderungen wurde ein solches Kriterium nicht weiter berücksichtigt. Die teilweise für den Unterricht oder die Zielgruppe untypischen Inhalte sollten jedoch in der späteren Erstellung des Testentwurfs berücksichtigt werden.

Die Unterschiede in den Anteilen der vier Stufen der Lösungskomplexität in den einzelnen Wettbewerbsrunden lassen sich durch Betrachtung der Formen, in der die Wettbewerbsrunden durchgeführt werden, erklären. In der ersten Wettbewerbsrunde der IJSO ist zu berücksichtigen, dass die Teilnehmenden sämtliche Quellen zur Verfügung haben, um die Aufgaben zu lösen, da es sich um eine Hausaufgabenrunde handelt. Die Schwierigkeit der Aufgaben ist hier dementsprechend unter dem Aspekt zu betrachten, dass sich das Fachwissen zum Lösen einer

Aufgabe aus beliebigen Quellen angeeignet werden kann. Zudem können Teilnehmende in der ersten Wettbewerbsrunde in Gruppen bis maximal drei Schülerinnen und/ oder Schülern zusammenarbeiten. Ab der zweiten Wettbewerbsrunde werden Klausuren geschrieben, was die Abnahme der Stufe „erklärende Zusammenhänge“ und die Zunahme der Stufen „beschreibende Zusammenhänge“ und „Fakten“ erklärt, denn das Fachwissen zum Lösen einer Aufgabe muss für die zweite und dritte Wettbewerbsrunde bei den Teilnehmenden im Moment der Klausur vorhanden sein und kann sich nicht, wie in der ersten Runde, erst noch angeeignet werden. Der hohe Anteil der Stufe „beschreibende Zusammenhänge“ lässt sich zudem mit Blick auf den internationalen Wettbewerb erklären, dessen Aufgaben zu 92.8% der Stufe „beschreibende Zusammenhänge“ zugeordnet wurden.

Der hohe Anteil des Kompetenzbereichs Fachwissen lässt sich ebenso mit Blick auf die internationale Runde erklären, da der Kompetenzbereich Fachwissen hier zu 91.4% vertreten war. In Hinblick auf die Unterschiede zwischen den einzelnen Runden sind die verschiedenen Funktionen der einzelnen Wettbewerbsrunden zu nennen: In der ersten Wettbewerbsrunde steht die Breitenförderung noch im Fokus, von der zweiten über die dritte Runde (bzw. aktuell die vierte Runde) verändert sich dies zu Enrichment und Spitzenförderung sowie der Auswahl des Nationalteams (vgl. Peters et al., 2013). Dass die Kompetenzbereiche Kommunikation und Bewertung kaum vertreten sind, lässt sich vor allem damit erklären, dass die Bewertung der Antworten auf die Wettbewerbsaufgaben objektiv und eindeutig erfolgen muss. Eine objektive und eindeutige Bewertung von Antworten auf Aufgaben aus den Kompetenzbereichen Kommunikation und Bewertung scheint meist schwieriger zu sein als die Bewertung von Antworten auf Aufgaben aus den anderen Kompetenzbereichen.

Zusammenfassend ist anzumerken, dass die vorgenommene Klassifikation und Analyse von Wettbewerbsaufgaben anhand des adaptierten und weiterentwickelten Schemas lediglich erste Hinweise auf die Anforderungen der IJSO geben konnte, die für den Zweck in dieser Arbeit jedoch als zielführend erachtet wurden. Für eine tiefgründige Analyse der Anforderungen der Wettbewerbsaufgaben sollten einige Kriterien vertieft bzw. erweitert werden. Bei den Oberflächenmerkmalen beträfe dies vor allem das Themengebiet. Um hier differenziertere Aussagen treffen zu können, könnte die Themenliste an Schulbüchern und Lehrplänen der einzelnen Fächer und ggf. der einzelnen Bundesländer ausgerichtet werden. Des Weiteren könnte die Repräsentationsform einer Aufgabe berücksichtigt werden, wie z. B. im allgemeindidaktischen Klassifikationsschema nach Maier et al. (2013). Bezüglich der Tiefenmerkmale könnten die realen Antworten der Teilnehmenden für die Ermittlung der diagnostischen Güte herangezogen und somit weitere Maße für die Schwierigkeit der Wettbewerbsaufgaben ermittelt werden. Auch die Analyse fachlicher und spezifischer Anforderungsmerkmale der einzelnen Aufgaben, die z. B. erforderliche Kenntnisse zum Lösen einer Aufgabe wie Rechenfertigkeiten oder Textverständnis berücksichtigen (vgl. H. E. Fischer & Draxler, 2001), würden differenziertere Aussagen über die Schwierigkeit der Aufgaben und somit die Anforderungen des Wettbewerbs erlauben. Eine derartige Analyse wurde bereits für die IChO von Deussing (2014) für Aufgaben des deutschen Auswahlverfahrens vorgenommen, die dem Basiskonzept „chemische Reaktion“ zuzuordnen sind. Bei einer Erweiterung der Analyse wäre jedoch die jeweilige Zielsetzung zu berücksichtigen, da andere Kriterien von Bedeutung sein könnten, je nachdem, ob man z. B. die Aufgabenentwicklung unterstützen, die Qualität des Wettbewerbs prüfen, Lehrkräfte bei der Unterstützung von Teilnehmenden helfen oder Teilnehmende selbst vorbereiten möchte.

Beschreibt man abschließend die Anforderungen der IJSO unter Berücksichtigung der unterschiedlichen Strukturen der einzelnen Wettbewerbsrunden, ist die IJSO ein Wettbewerb, der im Besonderen Schülerinnen und Schüler anspricht, die sich für alle drei Naturwissenschaften interessieren, was als ein Alleinstellungsmerkmal des Wettbewerbs anzusehen ist. Dies bedeutet auch, dass Teilnehmende ein breites Wissen in Biologie, Chemie und Physik aufweisen sollten, um erfolgreich im Wettbewerb zu sein. Dabei kann sich dieses Wissen in der ersten Wettbewerbsrunde noch eigenständig anhand unterschiedlicher Quellen erarbeitet werden, was an komplexeren Aufgaben zu erkennen ist, bei denen die Erklärung von Vorgängen oder Ursache-Wirkungs-Zusammenhängen (Lösungskomplexität „erklärende Zusammenhänge“) gefragt ist. Im weiteren Verlauf des Wettbewerbs steht vorwiegend das Beschreiben von Vorgängen, das Erkennen von Ursache-Wirkungs-Zusammenhängen sowie das Durchführen von Berechnungen (Lösungskomplexität „beschreibende Zusammenhänge“) im Vordergrund. Da viele Themen in der IJSO vorkommen, die nicht schulrelevant sind oder für die Teilnehmenden oftmals noch nicht schulrelevant waren, sind das Transferieren von bekannten auf neue Inhalte und ein hohes Abstraktionsvermögen wesentliche Anforderungen der in den höheren Runden vorkommenden Wettbewerbsaufgaben. Die unterschiedlichen, zum Teil ungewöhnlichen und aus dem Unterricht teilweise unbekannteren Aufgabenformate, die theoretische sowie experimentelle Inhalte abbilden, stellen zusätzlich eine Herausforderung für Teilnehmende des Wettbewerbs dar.

Vergleicht man die geschilderten Anforderungen der IJSO mit den durch die KMK definierten Zielen naturwissenschaftlicher Grundbildung, wird deutlich, dass die IJSO als eine wertvolle unterrichtsergänzende Maßnahme für naturwissenschaftlich interessierte Schülerinnen und Schüler anzusehen ist. Laut den Bildungsstandards der Fächer Biologie, Chemie und Physik ist das Ziel naturwissenschaftlicher Grundbildung,

Phänomene erfahrbar zu machen, die Sprache und Historie der Naturwissenschaften zu verstehen, ihre Ergebnisse zu kommunizieren sowie sich mit ihren spezifischen Methoden der Erkenntnisgewinnung und deren Grenzen auseinander zu setzen. Dazu gehört das theorie- und hypothesengeleitete naturwissenschaftliche Arbeiten, das eine analytische und rationale Betrachtung der Welt ermöglicht. Darüber hinaus bietet naturwissenschaftliche Grundbildung eine Orientierung für naturwissenschaftlich-technische Berufsfelder und schafft Grundlagen für anschlussfähiges berufsbezogenes Lernen. (Kultusministerkonferenz, 2005b, S. 6)

Der hohe Anteil von Wettbewerbsaufgaben des Kompetenzbereichs Fachwissen führt auf den ersten Blick dazu, dass die IJSO den umfangreichen Zielen naturwissenschaftlicher Grundbildung nicht in Gänze gerecht werden kann. Der interdisziplinäre Charakter, das eigenständige Erarbeiten von Fachinhalten in der ersten Wettbewerbsrunde sowie die Bearbeitung von Wettbewerbsaufgaben der höheren Runde, die sich mit oftmals nicht schulrelevanten Inhalten beschäftigen und somit eine hohe Transferleistung sowie ein hohes Abstraktionsvermögen bei einer Vielfalt an Aufgabenformaten verlangen, ergänzen den naturwissenschaftlichen Unterricht jedoch auf sinnvolle Art und Weise. Den berufsbezogenen Zielen naturwissenschaftlicher Grundbildung kommt die IJSO in dem Sinne nach, dass der Wettbewerb die Möglichkeit bietet, naturwissenschaftliche Interessen vertiefend zu verfolgen und sich auf nationaler und internationaler Ebene mit Gleichgesinnten auszutauschen, was naturwissenschaftlicher Unterricht in der Form zumeist nicht bieten kann. In diesem Sinne sollte der Wettbewerb auch als unterrichtsergänzende Maßnahme für besonders naturwissenschaftlich interessierte Schü-

lerinnen und Schüler gesehen werden, ohne den Anspruch, die Ziele naturwissenschaftlicher Grundbildung im Rahmen der allgemeinbildenden Schulen in vollem Umfang zu verfolgen.

### **4.3 Analyse der Anforderungen im Wettbewerb Jugend forscht bzw. Schüler experimentieren**

Den Wettbewerb Jugend forscht, der sich an Teilnehmende bis 21 Jahren richtet, die sich für MINT interessieren, gibt es bereits seit 1965, seit 1969 gibt es zusätzlich die Jugendsparte Schüler experimentieren, die sich an Schülerinnen und Schüler bis 14 Jahre wendet. Im Jahr 2015 hatten sich mehr als 11500 Schülerinnen und Schüler für den Wettbewerb (Jugend forscht und Schüler experimentieren zusammengefasst) mit mehr als 6000 Projekten angemeldet (Stiftung Jugend forscht e. V., 2015a). Auf die Jugendsparte Schüler experimentieren fielen dabei ca. 6000 Anmeldungen mit mehr als 3100 Projekten. Der Wettbewerb ist ein Ideenwettbewerb, der von den Teilnehmenden Kreativität und Eigenverantwortlichkeit bei der Suche und Bearbeitung eines möglichst interessanten und originellen Themas fordert (Peters et al., 2013).

Im Rahmen des Wettbewerbs wählen die Teilnehmenden, oft auch in Gruppen bis maximal drei Schülerinnen und/ oder Schülern, ein Thema aus den sieben Bereichen Arbeitswelt, Biologie, Chemie, Geo- und Raumwissenschaften, Mathematik/ Informatik, Physik oder Technik. Zu ihrem selbst gewählten Thema finden die Teilnehmenden eine Forschungsfrage und erstellen dazu eine ca. fünfzehnteilige Projektarbeit, die meist von einem Projektbetreuer (Lehrkräfte, Kooperationspartner aus der Wirtschaft oder Wissenschaft) begleitet wird. Für den Regionalwettbewerb, zu dem alle Teilnehmenden eingeladen werden, die ihre Projektarbeit fristgerecht eingereicht haben, gestalten die Teilnehmenden einen Wettbewerbsstand und bereiten einen Kurzvortrag vor einer Jury vor, die sich aus Experten aus Wirtschaft, Wissenschaft und Schule zusammensetzt. Dem Kurzvortrag schließt sich eine Befragung durch die Jury an. Die Bewertung der Teilnehmenden erfolgt mithilfe eines Jury-Leitfadens anhand von Bewertungskriterien, die fortlaufend weiterentwickelt und überarbeitet werden. Die Regionalsieger qualifizieren sich für den Landeswettbewerb, für Schüler experimentieren ist dies die höchste Stufe. Die Landessieger der Sparte Jugend forscht qualifizieren sich für das Bundesfinale, das jährlich Ende Mai stattfindet (Peters et al., 2013).

Im Gegensatz zu der IJSO handelt es sich bei Jugend forscht bzw. Schüler experimentieren um einen projektorientierten Wettbewerb. Da die Teilnehmenden und ihre Projekte durch eine Jury beurteilt werden, stellte sich die Analyse der Anforderungen als weniger formalisiert möglich als bei der IJSO dar. Da die Jury sowie die ihr zur Verfügung gestellten Handreichungen entscheidend für die Beurteilung der Arbeiten der Teilnehmenden zu sein schienen, wurden diese verstärkt in den Fokus der iterativen Analysen genommen, die im Folgenden vorgestellt werden.

#### **4.3.1 Analyse der Jury-Arbeit und wettbewerbsinterner Richtlinien**

Um einen ersten Eindruck vom Ablauf des Wettbewerbs zu erhalten, wurden die Landeswettbewerbe 2013 in Hamburg, Schleswig-Holstein und Niedersachsen besucht, wobei in Niedersachsen nur der Landeswettbewerb Schüler experimentieren, der dort getrennt vom Landeswettbewerb Jugend forscht stattfand, besucht wurde. In Hamburg wurde es ermöglicht, den

Jury-Gesprächen und den Jury-Sitzungen beizuwohnen, wodurch ausführliche Einblicke in die Jury-Arbeit gewonnen werden konnten.

Den Jurorinnen und Juroren wurde seitens der Wettbewerbsleitungen ein bundeseinheitlicher, nicht öffentlicher Leitfaden zur Jury-Arbeit sowie bundeseinheitliche, öffentlich verfügbare Bewertungskriterien (siehe [http://www.jugend-forscht.de/fileadmin/user\\_upload/Downloadcenter/Teilnahme/Bewertungskriterien.pdf](http://www.jugend-forscht.de/fileadmin/user_upload/Downloadcenter/Teilnahme/Bewertungskriterien.pdf), Zugriffsdatum 09.02.2017) vorgegeben. Der Leitfaden zur Jury-Arbeit gab dabei allgemeine Informationen zum Wettbewerb selbst sowie zum Ablauf am Wettbewerbstag. Auch die Bewertungskriterien und deren Gewichtung wurden hier thematisiert. In den Bewertungskriterien wurden Aspekte wie der Eigenanteil, die Originalität des Themas, Schwierigkeitsgrad und Eignung der Fragestellung, Kreativität, Ideenreichtum, Berücksichtigung des aktuellen Stands von Wissenschaft und Technik, inhaltliche und formale Qualität der schriftlichen Projektarbeit, Darstellung der Ergebnisse, Fehleranalyse, Selbstkritik und Ausblick sowie die Qualität des mündlichen Vortrags und die Qualität der Antworten im Jury-Gespräch genannt (Stiftung Jugend forscht e. V., 2015b). Besonders betont wurde dabei der Eigenanteil, die wissenschaftliche Exzellenz sowie die Präsentation der Projektarbeiten (Stiftung Jugend forscht e. V., 2015c). Des Weiteren wurde den Jurorinnen und Juroren ein Bewertungsbogen zur Verfügung gestellt, der die oben genannten Kriterien berücksichtigte.

Auf den besuchten Landeswettbewerben dienten die Bewertungskriterien den Jurorinnen und Juroren vor allem als Empfehlung und nicht als strikte Richtlinie. Es war daher kaum eindeutig zu ermitteln, ob der Schwerpunkt für die Jurorinnen und Juroren auf der Projektarbeit, der Präsentation oder dem Jury-Gespräch lag. Es deutete sich jedoch an, dass die Projektarbeit im Vergleich zur Präsentation am Wettbewerbstag sowie dem Jury-Gespräch eine eher untergeordnete Rolle spielte. Auch wurde der pädagogische Aspekt, unter dem die Arbeiten der Teilnehmenden betrachtet wurden, zum Teil stark gewichtet. So sollten vor allem Teilnehmende der Sparte Schüler experimentieren motiviert werden, ihr Interesse weiter zu verfolgen und im nächsten Jahr erneut am Wettbewerb teilzunehmen. Bei Projekten, bei denen es um die Qualifikation für das Bundesfinale ging, stand die wissenschaftliche Qualität der Arbeit im Fokus der Jury.

Zusammenfassend gesagt konnte anhand der Besuche der verschiedenen Landeswettbewerbe kein eindeutiger Aspekt herausgearbeitet werden, der vertiefend der Analyse der Anforderungen des Wettbewerbs hätte dienen können. Nach Gesprächen mit Jurorinnen und Juroren und Wettbewerbsleitungen wurden Kompetenzen des Bereichs Kommunikation als äußerst wichtig für eine erfolgreiche Teilnahme an Jugend forscht bzw. Schüler experimentieren eingeschätzt. Nach der Durchsicht einiger Projektarbeiten und der Ermittlung prinzipieller Anforderungen, die diese stellten, wurden Kenntnisse und Fähigkeiten des Kompetenzbereichs Erkenntnisgewinnung, und hier im Speziellen das Entwickeln von Fragestellungen, das Durchführen von (eigenen) Experimenten, die Auswertung gewonnener Daten sowie die Fehleranalyse als weitere wichtig zu erachtende Faktoren für eine erfolgreiche Wettbewerbsteilnahme identifiziert. Um diese Hypothesen zu festigen und die Bedeutung einzelner Kriterien zu erfassen, wurde ein Fragebogen in Anlehnung an Meißner (2013), der Leistungsanforderungen an Teilnehmende von Informatikwettbewerben untersuchte, entwickelt, der im Rahmen einer qualitativen Studie mit  $N = 12$  Jurorinnen und Juroren eingesetzt wurde.

Tabelle 4.6: **Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Für wie wichtig erachten Sie die folgenden Aspekte für eine gute Leistung im Wettbewerb Jugend forscht?** Beantwortet auf einer vierstufigen Likert-Skala.

<i>Item</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Kreativität	12	3.92	0.29
Intelligenz	12	3.83	2.48
Fähigkeit, wissenschaftlich Ergebnisse zu präsentieren	12	3.75	0.62
Eigenständigkeit	12	3.58	0.51
Problemlösefähigkeit	12	3.42	0.51
Fachwissen im Bereich des eigenen Projekts	12	3.33	0.49
Beherrschen wissenschaftlicher Arbeitsweisen	12	3.25	0.75
Wissen, das über das eigene Projekt/ die aktuelle Aufgabenstellung hinausgeht und in anderen, benachbarten Bereichen liegt	12	3.17	0.39
Fähigkeit, sich selbst darzustellen	12	3.08	0.79

#### 4.3.2 Qualitative Befragung von Jurorinnen und Juroren und deskriptive Ergebnisse

Der Fragebogen entwickelte sich aus den Erkenntnissen der Besuche des Wettbewerbs Jugend forscht bzw. Schüler experimentieren sowie aus dem Fragebogen von Meißner (2013). Anhand von sechs Fragen wurden zwölf Jurorinnen und Juroren des Landeswettbewerbs Hamburg 2014 befragt, für wie wichtig sie bestimmte Aspekte für eine gute Leistung im Wettbewerb erachten und wie sie bestimmte Aussagen zu den Teilnehmenden beurteilen. Zudem wurden die Jurorinnen und Juroren gebeten, bestimmte Aspekte und Teilleistungen in eine Reihenfolge zu bringen. Der Fragebogen ist in Anhang C dargestellt. Deskriptive Ergebnisse der Befragung werden im Folgenden vorgestellt.

Bei der Beurteilung der Wichtigkeit von Aspekten wie Kreativität oder Problemlösefähigkeit für eine gute Leistung im Wettbewerb auf einer vierstufigen Likert-Skala, siehe Tabelle 4.6, unterschied sich kein Aspekt in seiner Wichtigkeit wesentlich von anderen Aspekten, die Mittelwerte aller Aspekte lagen alle über einem Mittelwert von  $M = 3.00$ . Hervorzuheben ist der Aspekt Kreativität, der mit einem Mittelwert von  $M = 3.92$ ,  $SD = 0.29$  den höchsten Wert aufwies. Auch die Intelligenz wurde mit einem Mittelwert von  $M = 3.83$  als wichtig eingeschätzt, die hohe Standardabweichung von  $SD = 2.48$  deutete hier jedoch auf ein unterschiedliches Meinungsbild unter den Jurorinnen und Juroren hin. Mit dem geringsten Mittelwert und damit als eher wichtigen Aspekt beurteilten die Jurorinnen und Juroren die Fähigkeit, sich selbst darzustellen ( $M = 3.08$ ,  $SD = 0.79$ ).

Wie in Tabelle 4.7 zu sehen ist, wurde der Kompetenzbereich Erkenntnisgewinnung mit einem Mittelwert von  $M = 3.92$ ,  $SD = 0.29$  am höchsten in der Bedeutung für eine gute Leistung im Wettbewerb bewertet, der Kompetenzbereich Fachwissen am zweithöchsten mit  $M = 3.08$ ,  $SD = 0.51$ . Aber auch die beiden anderen Kompetenzbereiche wurden mit einem Mittelwert von  $M > 3.00$  und damit eher wichtig beurteilt.

Für die dritte Frage des Fragebogens wurden Jurorinnen und Juroren gebeten, die Plätze 1

Tabelle 4.7: **Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Für wie wichtig erachten Sie die folgenden Kompetenzbereiche für eine gute Leistung im Wettbewerb Jugend forscht?** Beantwortet auf einer vierstufigen Likert-Skala.

<i>Item</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Erkenntnisgewinnung	12	3.92	0.29
Bewertung	12	3.42	0.51
Kommunikation	12	3.33	0.49
Fachwissen	12	3.08	0.51

(große Bedeutung) bis 7 (kleine Bedeutung) auf verschiedene Aspekte einer guten Leistung im Wettbewerb Jugend forscht zu verteilen. Hierbei ergab sich, wie in Tabelle 4.8 dargestellt ist, ein sehr heterogenes Bild mit ähnlichen Mittelwerten und hohen Standardabweichungen. Für die Bedeutung der Teilleistungen im Wettbewerb in der vierten Frage bot sich ein homogeneres Bild, siehe Tabelle 4.9. Hier wurde der Projektarbeit die größte Bedeutung beigemessen ( $M = 1.67$ ,  $Mdn = 1.0$ ,  $SD = 0.97$ ). Der Standgestaltung hingegen wurde die geringste Bedeutung zugeschrieben ( $M = 3.33$ ,  $Mdn = 4.0$ ,  $SD = 0.89$ ).

Für die Beurteilung von Aussagen bezüglich des Erkenntnisprozesses, den die Teilnehmenden im Rahmen des Wettbewerbs durchlaufen, ergab sich, wie schon beim Ranking verschiedener Aspekte für eine gute Leistung im Wettbewerb (Tabelle 4.8), kein eindeutiges Bild. Auch hier fanden sich, wie in Tabelle 4.10 dargestellt, recht hohe Mittelwerte mit teilweise hohen Standardabweichungen. Wie auch schon für die Beurteilung der Wichtigkeit von Aspekten wie Kreativität oder Problemlösefähigkeit deuteten auch diese Ergebnisse auf ein unterschiedliches Meinungsbild unter den Jurorinnen und Juroren hin.

In einer letzten Frage beurteilten die Jurorinnen und Juroren die Wichtigkeit einzelner Aspekte der Präsentation am Wettbewerbstag. Am wichtigsten wurde hier die fachliche Richtigkeit ( $M = 3.50$ ,  $SD = 0.52$ ) eingeschätzt. Mit einem Mittelwert von  $M = 2.5$ ,  $SD = 0.67$  wurde das Einhalten von Rahmenbedingungen in seiner Wichtigkeit am geringsten bewertet. Die Ergebnisse aller Items der letzten Frage sind in Tabelle 4.11 dargestellt.

### 4.3.3 Interpretation und Diskussion der deskriptiven Ergebnisse

Bei der Interpretation der Ergebnisse ist zu beachten, dass aufgrund der Stichprobengröße von  $N = 12$  und der Tatsache, dass Jurorinnen und Juroren eines einzelnen Landeswettbewerbs befragt wurden, keine allgemeingültigen Aussagen getroffen werden können. Die durchgeführte qualitative Befragung konnte lediglich erste Anhaltspunkte liefern.

Insgesamt konnten die deskriptiven Ergebnisse der Befragung von Jurorinnen und Juroren nur teilweise die Beobachtungen, die bei Besuchen der Landeswettbewerbe gemacht wurden, bestätigen. Vielmehr zeichnete sich ab, dass es sowohl regionale Unterschiede als auch Unterschiede innerhalb eines Wettbewerbs zu geben schien, was durch die zum Teil recht unterschiedlichen Meinungsbilder, die die Befragung von Jurorinnen und Juroren lieferten, bestätigt wurde. Im Folgenden werden daher nur einige Aspekte aufgegriffen, die als geeignet für die

Tabelle 4.8: **Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Verteilen Sie auf die folgenden Aspekte entsprechend ihrer Bedeutung für eine gute Leistung im Wettbewerb Jugend forscht die Plätze 1 bis 7.** 1 = große Bedeutung, 7 = kleine Bedeutung.

<i>Item</i>	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>
Kombiniert, spezialisiert, erweitert und variiert bekannte Methoden, Verfahren und Lösungswege auf ein neues Problem	12	1.92	2.0	1.00
Wählt aus verschiedenen Ideen und Lösungsvorschlägen die passende aus	12	2.58	2.0	1.38
Ist bereit und in der Lage dazu neue Wege als bisher bekannte zu gehen	12	2.83	3.0	1.85
Findet viele Ideen und Lösungsvorschläge	12	3.00	3.0	1.28
Präsentiert Ergebnisse sicher, selbstbewusst und verwendet dabei die Fachsprache	12	3.08	3.0	1.31
Arbeitet formal korrekt und präzise entsprechend wissenschaftlicher Standards	12	3.67	3.0	1.92
Kennt sich gut mit Fakten und Verfahren aus dem Umfeld des Problems aus	12	3.58	3.5	1.78

Tabelle 4.9: **Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Verteilen Sie auf die folgenden Teilleistungen entsprechend ihrer Bedeutung für eine gute Leistung im Wettbewerb Jugend forscht die Plätze 1 bis 4.** 1 = große Bedeutung, 4 = kleine Bedeutung.

<i>Item</i>	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>
Projektarbeit	12	1.67	1.0	0.97
Präsentation der Ergebnisse	12	1.92	2.0	0.90
Jurybefragung	12	1.92	2.0	0.67
Standgestaltung	12	3.33	4.0	0.89

Beschreibung der Anforderungen des Wettbewerbs Jugend forscht bzw. Schüler experimentieren angesehen wurden.

Bei den Kompetenzbereichen deckte sich die Beobachtung, dass der Bereich Erkenntnisgewinnung als wichtig erachtet wurde, mit den Ergebnissen der Befragung von Jurorinnen und Juroren. Die als gering eingeschätzte Bedeutung des Kompetenzbereichs Fachwissen steht jedoch teilweise im Widerspruch zu anderen Items des Fragebogens, die das Fachwissen betrafen. So wurde die Wichtigkeit von „Fachwissen im Bereich des eigenen Projekts“ sowie von „Wissen, das über das eigene Projekt/ die aktuelle Aufgabenstellung hinausgeht und in anderen, benachbarten Bereichen liegt“ mit Mittelwerten von  $M = 3.33$ ,  $SD = 0.49$  bzw.  $M = 3.17$ ,  $SD = 0.39$  bewertet. Für die Präsentation wurde die fachliche Richtigkeit sogar als wichtigster Aspekt bewertet ( $M = 3.50$ ,  $SD = 0.52$ ). Die bei Besuchen der Landeswettbewerbe gemachte Beobachtung, dass der Kompetenzbereich Kommunikation eine große Rolle spielt, konnte ebenso nicht eindeutig bestätigt werden. So bewerteten die Jurorinnen und Juroren den Kompetenzbereich Bewertung in seiner Wichtigkeit minimal höher ( $M = 3.42$ ,  $SD = 0.51$ ) als den

Tabelle 4.10: **Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Bitte beurteilen Sie die folgenden Aussagen. Es ist besser, dass der/ die Teilnehmende. . .** Beantwortet auf einer vierstufigen Likert-Skala.

<i>Item</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Viele unkonventionelle (eigene) Lösungsideen und Methoden entwickelt, die eventuell Schwächen besitzen, als dass er/ sie nur wenige Lösungsideen und Methoden entwickelt, die wissenschaftlich üblich und gut anwendbar sind.	12	3.42	1.93
Einfache Probleme bearbeitet, die wirklich neu sind, als dass er/ sie komplizierte Probleme bearbeitet, die vielleicht gar nicht neu sind.	12	3.33	0.78
Ein Problem aus dem Alltag bearbeitet, selbst wenn es formale Mängel aufweist, als dass er/ sie ein Alltag-fernes Problem wissenschaftlich nahezu perfekt bearbeitet.	12	3.00	2.17
Ein Problem bearbeitet, das er/ sie zwar selbst gefunden hat, das aber zu umfangreich ist, als dass er/ sie ein Problem bearbeitet, das vom Betreuer vorgegeben wurde und den „richtigen“ Umfang besitzt.	12	2.92	0.67

Tabelle 4.11: **Deskriptive Ergebnisse der qualitativen Befragung von Jurorinnen und Juroren: Für wie wichtig erachten Sie die folgenden Aspekte der Präsentation?** Beantwortet auf einer vierstufigen Likert-Skala.

<i>Item</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Fachliche Richtigkeit	12	3.50	0.52
Qualität der Darstellung	12	3.42	0.51
Strukturiertheit	12	3.42	0.51
Klarheit	12	3.33	0.49
Beschränkung auf das Wesentliche	12	2.92	0.51
Verdeutlichung des eigenen Arbeitsanteils der Projektmitglieder	12	2.33	0.65
Einhaltung der Rahmenbedingungen (z. B. zeitliche Vorgaben)	12	2.50	0.67

Kompetenzbereich Kommunikation ( $M = 3.33$ ,  $SD = 0.49$ ). Auch hier finden sich jedoch teilweise Widersprüche innerhalb des Fragebogens. So wurde z. B. die „Fähigkeit, wissenschaftlich Ergebnisse zu präsentieren“, mit einem Mittelwert von  $M = 3.75$ ,  $SD = 0.62$  in seiner Wichtigkeit als vergleichsweise hoch eingeschätzt. Im Vergleich zum Kompetenzbereich Bewertung ist jedoch zu beachten, dass der Fragebogen keine Items enthielt, die nach der Wichtigkeit von bestimmten Aspekten dieses Kompetenzbereichs fragen. Des Weiteren ist anzunehmen, dass die Jurorinnen und Juroren, die unterschiedlichste berufliche Hintergründe aufweisen, zum einen kein einheitliches Verständnis der Kompetenzbereiche aufwiesen und zum anderen aufgrund ihres Hintergrundes verschiedene Aspekte in ihrer Wichtigkeit auch unterschiedlich bewerteten.

Ein Widerspruch ergab sich auch in der Bedeutung der Teilleistungen des Wettbewerbs. So wurde hier die Projektarbeit als wichtigste Teilleistung bewertet, die Standgestaltung als im Vergleich unwichtigste Teilleistung. Dieses Ergebnis deckt sich jedoch nicht mit den Beobachtungen, die auf den Landeswettbewerben gemacht wurden. Hier wurde vermehrt her-

ausgestellt, dass die schriftliche Projektarbeit zwar von Bedeutung sei, die Leistung auf dem Wettbewerbstag selbst jedoch im Vordergrund stehe. Eine mögliche Erklärung ist, dass dieses Item eventuell missverständlich formuliert wurde und der Unterschied zwischen der gemeinten schriftlich verfassten Projektarbeit und der Arbeit, die die Teilnehmenden für ihr Projekt leisten, nicht deutlich wurde.

Die unterschiedliche Gewichtung von Aspekten wie der Kreativität oder der Problemlösefähigkeit (siehe Tabelle 4.6, Tabelle 4.8 sowie Tabelle 4.10) könnte darauf zurückzuführen sein, dass die unterschiedlichen Fachbereiche, die die Teilnehmenden wählen können, unterschiedliche Anforderungen stellen. Dies konnte im Rahmen der durchgeführten qualitativen Befragung jedoch nicht ermittelt werden und bedarf weiterer Untersuchungen. Meißner (2013), der die Leistungsanforderungen von Teilnehmenden an Informatikwettbewerben untersuchte, konnte jedoch bereits zeigen, dass Problemlösen das wichtigste Leistungskonzept für Jurorinnen und Juroren von Informatikwettbewerben (darunter auch Jugend forscht) darstellt, wohingegen in der hier dargestellten Untersuchung vor allem Kreativität als wichtig erachtet wurde. Die Wichtigkeit des Aspekts Problemlösefähigkeit wurde hier mit einem Mittelwert von  $M = 3.42$ ,  $SD = 0.51$  bewertet.

Zusammengefasst ergab die Analyse der Anforderungen des Wettbewerbs Jugend forscht bzw. Schüler experimentieren kein eindeutiges Bild. Dies lag vor allem an der projektorientierten Struktur des Wettbewerbs sowie der Bewertungen durch eine Jury. Letztere konnte nicht eindeutig und nicht objektiv analysiert werden, da viele Aspekte in die Bewertung mit einfließen, die nicht direkt messbar waren, wie z. B. die Eigenständigkeit der Teilnehmenden oder auch die Berücksichtigung pädagogischer Aspekte. Durch Besuche verschiedener Landeswettbewerbe und durch eine qualitative Jurybefragung konnten jedoch einige objektive und messbare Aspekte identifiziert werden, die für gute Leistungen im Wettbewerb angesehen wurden. Beschreibt man abschließend die gewonnenen Erkenntnisse über die Anforderungen von Jugend forscht bzw. Schüler experimentieren, spricht der Wettbewerb im Besonderen Schülerinnen und Schüler an, die ihre Ideen im MINT-Bereich erforschen möchten. Das Aneignen von Wissen und Methoden im eigenen Themenbereich, was auch außerschulische Inhalte betreffen kann, stellt eine Herausforderung für Teilnehmende des Wettbewerbs dar. Teilnehmende sollten Kompetenzen des Bereichs Erkenntnisgewinnung (z. B. das Formulieren von Forschungsfragen oder das Planen von Versuchen) aufweisen, wie auch die Analysen der Anforderungen des Wettbewerbs zeigten. Neben der zu verfassenden Projektarbeit stellen die Präsentation der Ergebnisse und die Jurybefragung wichtige Teilleistungen im Wettbewerb dar, so dass Kompetenzen der Bereiche Fachwissen, Kommunikation und Bewertung für eine erfolgreiche Wettbewerbsteilnahme auch nicht zu vernachlässigen sind. Des Weiteren wurde der Kreativität der Teilnehmenden ein hoher Wert zugeschrieben, was auch schon in Gesprächen mit Wettbewerbsleitenden und Jurorinnen und Juroren sowie bei Meißner (2013) herausgestellt wurde. Im Verlauf der einzelnen Wettbewerbsrunden und im Vergleich der Jugendsparte Schüler experimentieren zu Jugend forscht treten Leistungsaspekte immer mehr in den Vordergrund, während pädagogische Aspekte vor allem am Anfang berücksichtigt werden. Diese Erkenntnisse sind jedoch, wie bereits erwähnt, nur unter Vorbehalt zu betrachten, da die Anforderungen des Wettbewerbs im Detail mit einer größeren Stichprobe und nach Fachbereichen getrennt erfolgen sollte.

## 4.4 Zusammenfassung und Fazit

Die Analyse der Anforderungen der Schülerwettbewerbe IJSO sowie Jugend forscht bzw. Schüler experimentieren diente zur Eingrenzung des in den folgenden Studien zu messenden Konstruktes und der Definition des Messgegenstandes. Die Erkenntnisse dienten somit der Entwicklung des Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten.

Zur Analyse der Anforderungen der IJSO wurden die Aufgaben dieses aufgabenorientierten Wettbewerbs nach einem adaptierten und validierten Schema zur Klassifikation und Analyse von Wettbewerbsaufgaben analysiert. Die Analyse der Anforderungen des projektorientierten Wettbewerbs Jugend forscht bzw. Schüler experimentieren erfolgte über Besuche verschiedener Landeswettbewerbe und Gespräche mit Wettbewerbsleitenden sowie Jurorinnen und Juroren, Beobachtung der Bewertungsprozesse inklusive Analyse der Handreichungen zur Bewertung der Teilnehmenden und einer qualitativen Befragung von Jurorinnen und Juroren. Die wesentlichen und in die folgende Entwicklung des Testinstruments eingeflossenen Ergebnisse der Analyse der Anforderungen beider Schülerwettbewerbe sind im Folgenden zusammengefasst:

- 56.3% der IJSO-Aufgaben von 2008 bis 2014 lagen im Multiple-Choice-Format vor. Im Bereich der freien Formate wurde das Durchführen von Berechnungen am häufigsten (11.2%) in den Aufgaben verlangt.
- Häufig vertreten waren Aufgaben aus dem Bereich der Allgemeinen Chemie (18.3%), der Mechanik (14.6%) und der Tierphysiologie (13.1%).
- Die Aufgaben der IJSO waren zu 70.0% der Lösungskomplexität „beschreibende Zusammenhänge“ zuzuordnen, gefolgt von der Stufe „Fakten“ mit 15.7%.
- Der mit Abstand am häufigsten vertretene Kompetenzbereich in den Aufgaben der IJSO war der Bereich Fachwissen mit 85.7%, gefolgt vom Kompetenzbereich Erkenntnisgewinnung mit 10.6%.
- Teilnehmende des Wettbewerbs Jugend forscht bzw. Schüler experimentieren benötigen für eine gute Leistung im Wettbewerb vorwiegend Kompetenzen aus dem Bereich Erkenntnisgewinnung.
- Die Kreativität und Intelligenz von Teilnehmenden spielt eine große Rolle beim Erreichen guter Leistungen im Wettbewerb Jugend forscht bzw. Schüler experimentieren.

Die Erkenntnisse beeinflussten z. B. die Wahl des Aufgabenformates oder auch die Wahl der Kompetenzbereiche für die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten, wie im folgenden Kapitel 5 erläutert wird.

## Kapitel 5

# Entwicklung eines Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten von Wettbewerbsteilnehmenden

Auf Basis der Erkenntnisse aus der Analyse der Anforderungen der beiden Schülerwettbewerbe IJSO und Jugend forscht bzw. Schüler experimentieren (siehe Kapitel 4) wurde mit der Erstellung eines ersten Testentwurfs begonnen. Die Entwicklung, Prä-Pilotierung sowie die Überarbeitung und Pilotierung des in Studie 1 eingesetzten Testinstruments werden im Folgenden erläutert.

### 5.1 Erstellung eines Testentwurfs

Die folgende Vorgehensweise in der Dokumentation der Erstellung des Testentwurfs orientiert sich im Wesentlichen an Bühner (2011).

#### 5.1.1 Festlegung der Zielgruppe

Im Fokus der geplanten Erhebungen standen naturwissenschaftlich interessierte und talentierte Schülerinnen und Schüler, die sich bspw. unter Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe finden lassen würden. Diese sollten im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern sowie im Vergleich zu Teilnehmenden weiterer Schülerwettbewerbe hinsichtlich ihrer naturwissenschaftlichen Kenntnisse und Fähigkeiten untersucht werden. Um eine Vergleichbarkeit zu vorangegangenen Arbeiten gewährleisten zu können (Dierks et al., 2014; Dierks, Höffler, Blankenburg, Peters & Parchmann, 2016; Wentorf et al., 2015) und um die Kenntnisse aus der Analyse von Wettbewerbsanforderungen (siehe Kapitel 4) nutzen zu können, wurden für repräsentative Teilnehmende eines aufgabenorientierten Wettbewerbs die Teilnehmenden des Bundesfinales der IJSO ausgewählt. Die Gruppe der Teilnehmenden eines projektorientierten Schülerwettbewerbs sollte durch Teilnehmende des Wettbewerbs Jugend forscht bzw. Schüler experimentieren repräsentiert werden. Die Gruppe von Teilnehmenden des Wettbewerbs Jugend forscht bzw. Schüler experimentieren und die

Gruppe der Nichtteilnehmenden sollten so weit wie möglich auf die Gruppe der Teilnehmenden des IJSO-Bundesfinales in Alter und regionaler Herkunft abgestimmt werden.

Bezüglich des Alters der Teilnehmenden des IJSO-Bundesfinales galt, dass die Teilnehmenden im Wettbewerbsjahr unabhängig von der Klassenstufe nicht älter als 15 Jahre sein durften (Peters et al., 2013). Da der größte Anteil an teilnehmenden Schülerinnen und Schülern der IJSO aus der Altersgruppe zwischen 12 und 15 Jahren bestand, wurde diese Altersgruppe in Absprache mit der IJSO-Wettbewerbsleitung für die geplanten Erhebungen ausgewählt.

In einer Analyse statistischer Angaben zur Verteilung der Teilnehmenden über die Bundesländer, die über die Geschäftsführung der IJSO bezogen wurden, fanden sich starke regionale Unterschiede. So stammte ein großer Teil der teilnehmenden Schülerinnen und Schüler aus den östlichen Bundesländern, was mit der in den östlichen Bundesländern traditionell intensiven Förderung von Begabten in den Naturwissenschaften und der daher generell hohen Teilnahmezahl an naturwissenschaftlichen Schülerolympiaden erklärt werden kann (Olbertz, 2007). Aus diesem Grund sollten die Teilnehmenden des Wettbewerbs Jugend forscht bzw. Schüler experimentieren und die Gruppe der nicht an Wettbewerben teilnehmenden Schülerinnen und Schüler in ihrer regionalen Zusammensetzung auf die Stichprobe des IJSO-Bundesfinales 2013 (und evtl. des IJSO-Bundesfinales 2014) ausgerichtet werden.

In Bezug auf die Referenzgruppe im Wettbewerb Jugend forscht bzw. Schüler experimentieren stellte sich ein Vergleich mit den Teilnehmenden des IJSO-Bundesfinales zunächst als herausfordernd dar. Im Gegensatz zu IJSO-Teilnehmenden, die im Wettbewerbsjahr nicht älter als 15 Jahre sein durften, richtete sich der Wettbewerb Schüler experimentieren an Schülerinnen und Schüler bis zum Alter von 14 Jahren, ältere Teilnehmende (bis maximal 21 Jahre, Nagel, 1999) nahmen üblicherweise in der Sparte Jugend forscht teil. Die Analyse ausgewählter statistischer Angaben zur Altersstruktur einzelner vergangener Wettbewerbe, die über die Stiftung Jugend forscht bezogen wurden, zeigte jedoch, dass mehrere Teilnehmende zum Zeitpunkt des Wettbewerbs Schüler experimentieren und somit zum Zeitpunkt der Befragung ein Alter von 15 Jahren erreicht haben würden, so dass diese Gruppe als Referenz gewählt wurde.

Eine weitere Analyse statistischer Angaben zur besuchten Schulform der IJSO-Teilnehmenden, die ebenso über die Geschäftsführung der IJSO bezogen wurden, ergab, dass IJSO-Teilnehmende hauptsächlich Gymnasien und Schulen mit mathematisch-naturwissenschaftlichem Schwerpunkt sowie zum Teil Gemeinschaftsschulen (oder ähnliche Schulformen) besuchten. Daher wurden diese drei Schulformen für die Erhebungen ausgewählt. Die genaue Zusammensetzung der Stichprobe nach Schulform sollte wiederum auf die Stichprobe des IJSO-Bundesfinales 2013 (und evtl. des IJSO-Bundesfinales 2014) abgestimmt werden.

Zusammenfassend gesagt wurden für die geplanten Erhebungen Schülerinnen und Schüler zwischen 12 und 15 Jahren (und somit aus unterschiedlichen Klassenstufen) mit einem unterschiedlichen Maß an Wettbewerbserfahrung und, betreffend die Teilnehmenden mit Wettbewerbserfahrung, auf einem unterschiedlichen Level bezüglich ihres Wettbewerbserfolgs ausgewählt. Zusätzlich würde sich die Stichprobe aus mehreren Bundesländern zusammensetzen, so dass mit unterschiedlichen Lernvoraussetzungen gerechnet werden musste.

### 5.1.2 Rahmenbedingungen

Teilnehmende verschiedener naturwissenschaftlicher Schülerwettbewerbe und nichtteilnehmende Schülerinnen und Schüler sollten quantitativ im Querschnitt anhand eines gemeinsamen Testinstruments befragt werden, weshalb die Erhebungen im besten Fall sowohl zeitnah zu den entsprechenden Wettbewerben als auch gleichzeitig in Schulen durchgeführt werden sollten. In Gesprächen mit Wettbewerbsleitenden der IJSO und Jugend forscht bzw. Schüler experimentieren und Lehrkräften wurde vereinbart, dass die Testzeit für die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten insgesamt 60 Minuten nicht überschreiten durfte. So war laut Wettbewerbsleitenden bzw. Lehrkräften eine Erhebung im Rahmen einer Doppelstunde in Schulen möglich, und die Wettbewerbe würden durch diese Beschränkung wenig in ihrem Ablauf gestört werden. Des Weiteren wurde eine Papier-Bleistift-Testung gewählt.

### 5.1.3 Testziel und Konstruktionsstrategie

Für den zu konstruierenden Test wurden die folgenden Ziele formuliert:

- *Erfassung der naturwissenschaftlichen Kenntnisse und Fähigkeiten.* Der Test sollte in einem einzigen Erhebungsinstrument für die anvisierte Zielgruppe Kenntnisse und Fähigkeiten von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe und nichtteilnehmenden Schülerinnen und Schülern in den drei Naturwissenschaften Biologie, Chemie und Physik und verschiedenen Kompetenzbereichen erfassen.
- *Trennung von Gruppen.* Der Test sollte mögliche Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten zwischen Teilnehmenden verschiedener naturwissenschaftlicher Schülerwettbewerbe sowie zu nichtteilnehmenden Schülerinnen und Schülern aufdecken.

Einhergehend mit den Testzielen bot sich eine Kombination zweier Konstruktionsstrategien für den Testentwurf an, und zwar die rationale und die externale Testkonstruktion. Bei der rationalen Testkonstruktion geht man von einem deduktiven Ansatz aus, d. h. es liegt bereits zu Beginn der Testkonstruktion eine fundierte Theorie vor (Bühner, 2011). Die Strategie der externalen oder auch kriteriumsorientierten Testkonstruktion bietet sich vor allem an, wenn es um die Trennung von Gruppen geht, da für die Aufgabenkonstruktion vor allem Items gesucht werden, die zwischen den Gruppen besonders gut unterscheiden können (Bühner, 2011).

### 5.1.4 Generieren von Indikatoren und Eingrenzen des Konstrukts

Wie bereits in Kapitel 4 erläutert, wurde für die Schülerwettbewerbe, aus denen die Stichprobe rekrutiert werden sollte, eine Analyse der Anforderungen, die an die Wettbewerbsteilnehmenden gestellt werden, durchgeführt. Die wesentlichen Erkenntnisse, die für die Erstellung des Testentwurfs genutzt wurden, sind in Abschnitt 4.4 einzusehen.

### 5.1.5 Definition des Messgegenstands

In der vorliegenden Arbeit werden naturwissenschaftliche Kenntnisse und Fähigkeiten, wie bereits in Abschnitt 2.3.1 dargelegt, über die Kompetenzbereiche Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung der Bildungsstandards der KMK definiert (Klieme

et al., 2007; Kultusministerkonferenz, 2005a, 2005b, 2005c). Die Kompetenzbereiche umfassen eine inhaltliche Dimension sowie eine Handlungsdimension. Letztere wird v. a. durch den Kompetenzbereich Erkenntnisgewinnung repräsentiert und umfasst neben theoretischem auch experimentelles Arbeiten. Aufgrund der praktischen Umsetzbarkeit wurden für die geplanten Erhebungen jedoch zunächst keine experimentellen Aufgaben berücksichtigt. Des Weiteren wurden die Kompetenzbereiche Kommunikation und Bewertung zunächst ausgenommen, da im Besonderen die Kompetenzbereiche Fachwissen und Erkenntnisgewinnung in den Wettbewerben IJSO und Jugend forscht bzw. Schüler experimentieren vertreten waren und daher in diesen Kompetenzbereichen die größten Unterschiede vermutet wurden.

### 5.1.6 Wahl des Itemformats

Der Testentwurf sollte zu gleichen Teilen Aufgaben im gebundenen und freien Antwortformat enthalten, da auch Aufgaben der IJSO zu mehr als der Hälfte aus Aufgaben im gebundenen Antwortformat bestanden. Für das gebundene Format wurden, wie auch in den Aufgaben der IJSO, Multiple-Choice-Aufgaben (genauer: Multiple-Choice-Single-Select-Aufgaben) ausgewählt. Weitere Formate wie Ratingskalen oder Zuordnungsaufgaben sollten nicht weiter berücksichtigt werden, um in dem verhältnismäßig kurzen Test nicht noch mehr Varianz durch eine Vielfalt an Aufgabenformaten zu erzeugen. Ein Vorteil von Multiple-Choice-Aufgaben ist zudem ihre objektive und ökonomische Auswertung (Bühner, 2011). Als nachteilig anzumerken ist das Vorhandensein einer Ratewahrscheinlichkeit, je nach Anzahl und Qualität der Distraktoren. Herausfordernd ist zudem das Entwickeln geeigneter Distraktoren für die Beantwortung der Aufgabe. Im Hinblick auf Aufgaben bspw. aus dem Kompetenzbereich Erkenntnisgewinnung, in denen z. B. auch das Formulieren von Forschungsfragen oder Hypothesen vorgesehen war, war zu beachten, dass Multiple-Choice-Aufgaben eher auf das Wiedererkennen als auf die Reproduktion von Wissen ausgerichtet sein können (Bühner, 2011).

Die Wahl von Aufgaben im freien Antwortformat lag u. a. darin begründet, dass auch diese Art der Aufgabenbeantwortung in den IJSO-Aufgaben verwendet wurde. Des Weiteren wird in diesem Aufgabenformat Wissen frei verbalisiert, weshalb offene Aufgabenformate mehr Informationen über eine Merkmalsausprägung beinhalten können als gebundene Antwortformate. Auch können kreativ ausgelegte Aufgaben mit diesem Antwortformat getestet werden (Bühner, 2011). Als Nachteil ist hierbei allerdings zu nennen, dass auch mehr Text frei formuliert werden muss und die Beantwortung offener Aufgaben mehr Testzeit in Anspruch nimmt als die Beantwortung von Multiple-Choice-Aufgaben. Die Auswertung offener Aufgaben kann sich zudem problematisch gestalten, da eindeutige Auswertungskriterien schwierig festzulegen sind. Somit besteht die Gefahr, dass die Auswertung nur eingeschränkt objektiv erfolgen kann (Bühner, 2011).

## 5.2 Entwicklung und Auswahl der Aufgaben für den ersten Testentwurf

### 5.2.1 Entwicklung von Aufgaben

Der zu erstellende Testentwurf sollte zunächst Aufgaben der Kompetenzbereiche Fachwissen und Erkenntnisgewinnung enthalten. Für die Aufgaben sollten interdisziplinäre Themen ausgewählt werden, um die drei Fächer Biologie, Chemie und Physik gleichermaßen abzudecken und

die Themenvarianz möglichst gering zu halten. Auch sollten die Themen in ihrer Schwierigkeit gut variierbar sein, um in der favorisierten Stichprobe möglichst breit naturwissenschaftliche Kenntnisse und Fähigkeiten erfassen zu können.

Die Themenbereiche, die in den Wettbewerbsaufgaben der IJSO häufig vertreten waren (siehe Abschnitt 5.1.4), konnten für die Aufgaben kaum berücksichtigt werden, da sie sehr fachspezifisch waren und es vermieden werden sollte, Teilnehmende der IJSO zu bevorteilen. Für die Aufgaben des Kompetenzbereichs Fachwissen sollten daher neben möglichen anderen Themen zwei interdisziplinäre Schwerpunktthemen gewählt werden: Ein Thema sollte dabei für die meisten Schülerinnen und Schüler, unabhängig vom Bundesland und Klassenstufe der favorisierten Stichprobe, bereits in der Schule behandelt worden und somit im jeweiligen Lehrplan aufgeführt sein. Um ein solches Thema zu identifizieren, wurden exemplarisch die Lehrpläne aus Niedersachsen, Nordrhein-Westfalen und Sachsen-Anhalt analysiert und nach Gemeinsamkeiten in der Themenwahl in der entsprechenden Alters- bzw. Klassenstufe gesucht. Auf Basis der Inhaltsanalyse der Lehrpläne wurde das Thema „Energie in Natur und Technik“ als ein Schwerpunktthema gewählt, da dieses Thema in der zu testenden Altersgruppe laut der analysierten Lehrpläne bereits in einem gewissen Umfang im Unterricht in Biologie, Chemie oder Physik behandelt worden sein sollte. Zwar bestand die Möglichkeit, dass, je nach Alters- bzw. Klassenstufe, das Thema nicht in jedem Fach oder nicht ausführlich behandelt worden war, allerdings wurde angenommen, dass dieses Thema in jedem Fall einen Anknüpfungspunkt für Schülerinnen und Schüler bieten würde. Als außerschulisches Thema, und somit ein Thema, wie es auch in einem Schülerwettbewerb vorkommen könnte, wurde „Forensische Kriminalistik“ gewählt, da Teilnehmende der Studie möglicherweise aus Interesse Aufgaben aus diesem Bereich bearbeiten würden, auch wenn ihnen die Fachinhalte auf den ersten Blick unbekannt oder zu schwer vorkommen würden.

Für den Kompetenzbereich Erkenntnisgewinnung stand eine Auswahl möglichst interessanter und interdisziplinärer Themen im Vordergrund, da nur wenige Fachinhalte zur Lösung der Aufgaben notwendig sein sollten. Bedeutender stellten sich hier Aufgabenkategorien des Kompetenzbereichs Erkenntnisgewinnung dar: Aus der Frage, welche Anforderungen Teilnehmende von Jugend forscht bzw. Schüler experimentieren für eine erfolgreiche Teilnahme am Wettbewerb erfüllen müssen, und unter Berücksichtigung von Konzepten der *Scientific Literacy* (z. B. Evans, Gräber, Koballa & Nentwig, 2002; Klieme et al., 2007), wurden für den Kompetenzbereich Erkenntnisgewinnung die vier Aufgabenkategorien „Fragestellungen und Hypothesen finden“, „Versuche planen“, „Versuche auswerten“ und „Fehler analysieren“ für den Testentwurf gewählt.

Um möglichst wenige Aufgaben entwickeln zu müssen und die Erstellung des Testentwurfs somit zu vereinfachen, wurden zunächst vorhandene und bereits erprobte Aufgaben gesichtet. Dazu wurden Aufgaben aus den naturwissenschaftlichen PISA- und TIMSS-Erhebungen (Beschreibungen und Beispiele siehe Baumert, Bos & Lehmann, 2000a; Bos, Wendt, Köller & Selter, 2012; Neubrand et al., 2004; Senkbeil, Rost, Carstensen & Walter, 2005), aus dem durch das IQB durchgeführten Ländervergleich in Mathematik und den Naturwissenschaften 2012 (Beschreibungen und Beispiele siehe Pant, Stanat et al., 2013; Walpuski et al., 2008; Wellnitz et al., 2012), Wettbewerbsaufgaben aus verschiedenen Runden des deutschen Auswahlverfahrens der IJSO (Beispiele siehe Peters et al., 2013), Aufgaben aus verschiedenen institutsinternen Projekten, Aufgaben aus unterrichtspraktischen Fachzeitschriften (z. B.

„Naturwissenschaften im Unterricht“ oder „Praxis der Naturwissenschaften“) und Aufgaben aus naturwissenschaftlichen Schulbüchern (z. B. Asselborn, Jäckel & Risch, 2011; Tausch & von Wachtendonk, 2010) gesichtet. Dabei fiel jedoch auf, dass die meisten Aufgaben in ihrer bestehenden Form nicht auf die Zielgruppe passen würden. Dies hatte mehrere Gründe: Die Aufgaben der großen Schulleistungsstudien intendierten Kompetenzstufen zu erfassen, um das vorhandene Spektrum an Stufen abzubilden oder mit gesetzten Zielen, wie bspw. den Anforderungen des mittleren Schulabschlusses, zu vergleichen (vgl. Klieme & Baumert, 2001; Pant, Stanat et al., 2013; PISA-Konsortium Deutschland, 2007). Die Erfassung von definierten Kompetenzstufen oder von Kompetenzentwicklung sollte hier allerdings nicht Teil der Testziele sein. Darüber hinaus wurde angenommen, dass bspw. die Aufgaben aus dem IQB-Ländervergleich in Mathematik und den Naturwissenschaften 2012 eventuell nicht in der Gruppe der Teilnehmenden des IJSO-Bundesfinales differenzieren würden, da die Aufgaben, im Gegensatz zu den Aufgaben der IJSO, den Umgang mit Fachwissen anstelle des Fachwissens testen sollten, weshalb die Aufgaben alle zur Lösung relevanten Fachinformationen enthielten (vgl. Sumfleth et al., 2013). Die IJSO-Wettbewerbsaufgaben betreffend wurde vermutet, dass diese für die Gruppe der nichtteilnehmenden Schülerinnen und Schüler aufgrund des zum Teil hohen Grades an erforderlichen Fachinformationen zu schwer sein würden, so dass in dieser Subgruppe womöglich keine Differenzierung erfolgen würde. Aus diesem Grund wurden die Aufgaben für den Testentwurf zum einen neu entwickelt, zum anderen wurden vorhandene Aufgaben adaptiert bzw. modifiziert, um in der heterogenen Zielgruppe bestmöglich differenzieren zu können.

Die Aufgabenentwicklung erfolgte dabei mit dem in Abschnitt 5.1.3 genannten Testzielen, die naturwissenschaftlichen Kenntnisse und Fähigkeiten möglichst breit zu erfassen und zwischen den Teilgruppen der favorisierten Zielgruppe möglichst scharf zu trennen. Folgende drei Dimensionen wurden berücksichtigt:

- **Fächer:** Biologie, Chemie und Physik
- **Kompetenzbereiche:** Fachwissen und Erkenntnisgewinnung
- **Schwierigkeiten:** einfach – mittel – schwer

Die Einteilung der Aufgabenschwierigkeiten in die Kategorien einfach, mittel und schwer lag darin begründet, dass für die Erstellung des Testentwurfs aus verschiedenen Gründen nicht mit existierenden Schemata für die Konstruktion von Aufgaben gearbeitet wurde. So wurde in Erwägung gezogen, die an die Bildungsstandards gekoppelten Anforderungsbereiche als Schwierigkeitsstufen für die Aufgabenentwicklung zu übernehmen. Sumfleth et al. (2013) merken zu den Anforderungsbereichen allerdings an, dass „in einer Aufgabe jedoch oft mehrere schwierigkeits erzeugende Merkmale gleichzeitig und in unterschiedlicher Ausprägung vorkommen“ und daher „die Beschreibungen der Anforderungsbereiche der Bildungsstandards als Grundlage für die Entwicklung präziser Erhebungsinstrumente zu ungenau [sind]“ (Sumfleth et al., 2013, S. 39). Auch die Adaption von Schemata, die im IQB-Ländervergleich (z. B. Walpuski et al., 2010) oder in PISA (z. B. Neubrand et al., 2004) der Aufgabenentwicklung zugrunde gelegt wurden, wurden für die vorliegende Arbeit als nicht zielführend erachtet, da in diesen Vergleichsstudien die Erfassung von Kompetenzstufen im Vordergrund stand. Modelle zur Erfassung von Kompetenzstufen nutzen dabei mehrere Komplexitätsstufen als Maß für die Schwierigkeit der Aufgaben, wobei sich meist fünf (z. B. Baumert et al., 2000a; Bernholt et al., 2009; Walpuski et al., 2010) bis sechs (Neubrand et al., 2004; Neumann et al., 2007)

Komplexitätsstufen finden lassen. Zur Erfassung von mehr als drei Schwierigkeitsstufen wären allerdings auch deutlich mehr Aufgaben erforderlich und ein längerer Testentwurf nötig gewesen, was mit den gesetzten Rahmenbedingungen nicht zu vereinbaren gewesen wäre. Bei der Verwendung der vier Schwierigkeitsstufen aus der Aufgabenanalyse der IJSO-Aufgaben, siehe Kapitel 4, wäre zu beachten gewesen - und dies ist ebenso auf die Verwendung von Anforderungsbereichen oder Komplexitätsstufen als Merkmale für die Schwierigkeit einer Aufgabe zu übertragen - dass auch schon existierende Aufgaben für den Testentwurf adaptiert oder modifiziert werden sollten. Diese Aufgaben wurden jedoch nach unterschiedlichen Modellen entwickelt, die wiederum unterschiedliche Maße oder Stufen für die Schwierigkeit der Aufgaben vorsahen, so dass es kaum möglich gewesen wäre, andere Modelle einheitlich auf diese bereits vorhandenen Aufgaben anzuwenden. Die Einteilung der Aufgabenschwierigkeiten in die Kategorien einfach, mittel und schwer war zwar, ähnlich wie die Anforderungsbereiche, wenig präzise, jedoch ein globales Maß. Zur Prüfung der Einteilung wurde diese schließlich von mehreren Personen (Lehrkräfte, Wettbewerbsleitende, Fachdidaktikerinnen und Fachdidaktiker) unabhängig voneinander vorgenommen.

Für die Entwicklung neuer Multiple-Choice-Aufgaben wurden Distraktoren über Schülerantworten generiert. Hierzu erhielten drei Klassen der Stufen 7, 9 und 10 niedersächsischer Gymnasien potentielle Multiple-Choice-Aufgaben im offenen Format. Die anschließende Auswertung der Schülerantworten ergab jedoch, dass das offene Format für die meisten Aufgaben sinnvoller erschien, v. a. im Kompetenzbereich Erkenntnisgewinnung: Auch wenn die Teilnehmenden der Studie nur wenig zu den Aufgaben schreiben würden, würde man mehr Informationen über ihre Kenntnisse und Fähigkeiten erhalten, als wenn sie eine Aufgabe einfach nur falsch ankreuzen würden. Daher wurde die gleichmäßige Verteilung von Multiple-Choice-Aufgaben und offenen Aufgaben auf die Kompetenzbereiche verworfen. Stattdessen sollten Multiple-Choice-Aufgaben nur für den Kompetenzbereich Fachwissen eingesetzt werden.

### 5.2.2 Auswahl der Aufgaben für den ersten Testentwurf

Insgesamt wurden 63 Aufgaben erstellt, die potentiell in einen Testentwurf hätten aufgenommen werden können. Um die Anzahl an Aufgaben für den ersten Testentwurf zu reduzieren wurden mehrere Rahmenbedingungen gesetzt, die im Testentwurf in jedem Fall berücksichtigt werden sollten. Aufgrund der Ansicht, dass Multiple-Choice-Aufgaben für den Kompetenzbereich Erkenntnisgewinnung ungeeignet erschienen, wurden für diesen Kompetenzbereich nur offene Aufgaben, gleichermaßen verteilt auf die vier Aufgabenkategorien „Fragestellungen und Hypothesen finden“, „Versuche planen“, „Versuche auswerten“ und „Fehler analysieren“, ausgewählt. Im Kompetenzbereich Fachwissen wurden pro Fach und pro Thema jeweils drei offene und drei Multiple-Choice-Aufgaben verwendet, wobei für das Thema „Energie in Natur und Technik“ eine Multiple-Choice-Aufgabe und zwei offene Aufgaben und für das Thema „Forensische Kriminalistik“ eine offene Aufgabe und zwei Multiple-Choice-Aufgaben verwendet werden sollten. Durch diese Verteilung sollte berücksichtigt werden, dass die Teilnehmenden der Studie offene Aufgaben eines nicht schulrelevanten Themas, die ihnen eventuell schwer vorkommen könnten, unter Umständen erst gar nicht bearbeiten würden. Obendrein sollten alle Schwierigkeitsstufen abgebildet werden, wobei leichte Aufgaben am wenigsten vertreten sein sollten, um auch im Bereich leistungsstarker Schülerinnen und Schüler, hier im Speziellen der Wettbewerbsteilnehmenden, gut differenzieren zu können.

Tabelle 5.1: **Auswahl und Verteilung der für den ersten Testentwurf ausgewählten Aufgaben** Aufgeführt ist immer die Anzahl an offenen Aufgaben + die Anzahl an Multiple-Choice-Aufgaben.

	<i>Thema</i>	<i>Biologie</i>	<i>Chemie</i>	<i>Physik</i>	GESAMT
<i>Fachwissen</i>	Energie in Natur und Technik	2+1	1+1	2+1	5+3
	Forensische Kriminalistik	1+2	1+2	1+2	3+6
	Luftige Angelegenheiten	0+0	1+0	0+0	1+0
<i>Erkenntnis- gewinnung</i>	Forensische Kriminalistik	1+0	0+0	0+0	1+0
	Luftige Angelegenheiten	0+0	2+0	3+0	5+0
	Pflanzenwachstum	2+0	2+0	0+0	4+0
	So schnell wie der Schall	0+0	0+0	2+0	2+0
	Wasser marsch	2+0	1+0	0+0	3+0
GESAMT		8+3	8+3	8+3	24+9

Um die Auswahl der Aufgaben für den ersten Testentwurf zu erleichtern, wurden alle 63 Aufgaben noch einmal kategorisiert. Dazu wurde ein Kategoriensystem aus verschiedenen Aufgabenmerkmalen zusammengestellt, die bspw. in PISA (Prenzel, Häußler, Rost & Senkbeil, 2002) oder allgemein in der Analyse von Lern- und Testaufgaben angewendet wurden (z. B. Kleinknecht et al., 2011; Pickard, 2007). Folgende Aufgabenmerkmale wurden dabei berücksichtigt:

- **Formale Aufgabenmerkmale:** Länge des Aufgabentexts (Anzahl der Wörter), Multiple-Choice-Antwort, freie kurze Antwort, freie lange Antwort, numerischer Output
- **Kognitive Anforderungen beim Lösen der Aufgaben:** Textinformationen verarbeiten, aus Wissensbasis abrufen, logisch verknüpfen, etwas ausrechnen
- **Art des für das Lösen der Aufgabe erforderlichen Wissens:** Faktenwissen, prozedurales Wissen, konzeptuelles Wissen
- **Aufgabenschwierigkeit:** einfach - mittel - schwer
- **Inhalt der Aufgabe:** Vorhandene Informationen logisch verknüpfen und Rückschlüsse ziehen, Fragestellungen und Hypothesen finden, Versuche planen, Versuche auswerten, Fehleranalyse und Statistik, Abfrage von Wissen zu den Themen Analytik - Ballistik - chemische Reaktion - Energie - Genetik - Kernphysik - Kinetik - Optik - Stoffwechsel

Aufgrund der auf 60 Minuten begrenzten Testzeit wurden für den ersten Testentwurf 33 Aufgaben ausgewählt. Die Auswahl erfolgte in Kombination mit den Fächern, Kompetenzbereichen und Schwierigkeiten nach einer möglichst breiten Verteilung über die im Kategoriensystem enthaltenen Aufgabenmerkmale. Eine Übersicht ist in Tabelle 5.1 dargestellt.

Die Anordnung der ausgewählten Aufgaben im ersten Testentwurf erfolgte nach Themen (und nicht nach Fächern) um zu vermeiden, dass Schülerinnen und Schüler bestimmte Aufgaben aufgrund einer Abneigung gegen ein bestimmtes Fach unabhängig ihrer Kompetenzen auslassen würden. Zu jedem Themengebiet wurde ein Einleitungstext verfasst, der die Schülerinnen und Schüler für die folgenden Aufgaben motivieren sollte und zwischen 45 und 80 Wör-

ter umfasste. Es wurde weiterhin, ebenso aus motivationalen Gründen, darauf geachtet, leichte Aufgaben an den Anfang eines Thementeils zu setzen.

### 5.3 Prä-Pilotierungen des ersten Testentwurfs

Für eine erste Prä-Pilotierung wurden 33 Aufgaben ausgewählt, die von 46 niedersächsischen Schülerinnen und Schülern der Klassenstufen 7 und 8 eines Gymnasiums (Alter 12-14,  $M = 13.39$ ,  $SD = 0.65$ , 60.9% Mädchen) in 60 Minuten bearbeitet wurden. Erfragt wurden zudem das Alter, das Geschlecht und die letzten Zeugnisnoten in Biologie, Chemie und Physik der Schülerinnen und Schüler. Zusätzlich zu den Aufgaben erhielten die Schülerinnen und Schüler einen ergänzenden einseitigen Fragebogen, in dem sie gebeten wurden, die Aufgaben als Gesamtheit auf einer vierstufigen Likert-Skala (bspw. „Ich musste bei den Aufgaben viel zu viel lesen.“ auf einer Skala von „Ich stimme überhaupt nicht zu“ bis „Ich stimme voll zu“) einzuschätzen. Zudem wurden die Schülerinnen und Schüler anhand der gleichen Likert-Skala befragt, welche Art von Aufgabenstellungen sie gerne bearbeiten (bspw. „Offene Fragestellungen bereiten mir Probleme.“ oder „Mir gefallen Aufgaben, bei denen ich eigene Ideen einbringen kann.“). Die Ergebnisse hieraus zeigten keine Auffälligkeiten, so dass weder an der Textlänge noch am Aufgabenformat Änderungen vorgenommen wurden.

Die Korrektur der Aufgaben erfolgte anhand einer zuvor erstellten Musterlösung, die je nach Aufgabe unterschiedliche Anforderungen für eine bestimmte Anzahl an Punkten vorsah. Die Auswertung der korrigierten Schülerantworten erfolgte auf einer rein deskriptiven Ebene. Zudem wurde ermittelt, wie oft die einzelnen Aufgaben bearbeitet und wie häufig gar keine Punkte bzw. die maximale Punktzahl erreicht wurden. Bei der Korrektur der Aufgaben fiel des Weiteren auf, dass sich die Antworten der Schülerinnen und Schüler in bestimmte Kategorien einteilen ließen, weshalb für die Auswertung weiterer Schülerantworten ein mehrstufiges Kategoriensystem, beginnend mit den Schülerantworten aus der Prä-Pilotierung, erstellt wurde.

Aufgrund der Ergebnisse der deskriptiven Auswertung der Schülerantworten wurden 14 Aufgaben aus dem Test entfernt. Die beiden Aufgaben mit einem numerischen Output wurden aus dem Test entfernt, da diese Aufgaben von den Schülerinnen und Schülern kaum richtig beantwortet bzw. gar nicht bearbeitet wurden. Um logisch abhängige Aufgaben auszuschließen, wurden weitere fünf Aufgaben aus dem Test entfernt. Somit sollte gewährleistet werden, dass alle Schülerinnen und Schüler die gleiche Chance haben, jede einzelne Aufgabe unabhängig von der Lösung einer anderen Aufgabe zu bearbeiten. Auch wurden drei der neun Multiple-Choice-Aufgaben nicht weiter berücksichtigt. Eine Aufgabe wurde in der weiteren Testentwicklung ausgeschlossen, da sie keine Antwortstrukturen aufwies, die eine Auswertung anhand eines mehrstufigen Kategoriensystems ermöglicht hätte. Weitere drei Aufgaben wurden als zu schwierig eingestuft und waren von der Thematik her nicht oder kaum zu vereinfachen, so dass diese Aufgaben ebenso nicht weiter berücksichtigt wurden.

Die verbliebenen Aufgaben wurden in ihren Formulierungen überarbeitet und ausgeschärft. Eine Aufgabe des Kompetenzbereichs Fachwissen (Biologie) wurde noch einmal grundlegend verändert, da sie in ihrer ursprünglichen Form zu schwierig war, aber von der Thematik und ihrer Struktur her nach wie vor als passend angesehen wurde. Eine weitere Aufgabe des Kompetenzbereichs Fachwissen (Physik) wurde umstrukturiert, da die Antworten der Schülerinnen

Tabelle 5.2: **Verteilung der Aufgaben für die zweite Prä-Pilotierung auf die verschiedenen Kompetenzbereiche und Fächer** Aufgeführt ist immer die Anzahl an offenen Aufgaben + die Anzahl an Multiple-Choice-Aufgaben.

	<i>Thema</i>	<i>Biologie</i>	<i>Chemie</i>	<i>Physik</i>	GESAMT
<i>Fachwissen</i>	Energie in Natur und Technik	2+1	2+1	2+1	6+3
	Forensische Kriminalistik	1+1	1+1	1+1	3+3
<i>Erkenntnis- gewinnung</i>	Düt un dat ut Pütt un Pann	0+0	1+0	0+0	1+0
	Energie in Natur und Technik	0+0	0+0	1+0	1+0
	Forensische Kriminalistik	1+0	0+0	0+0	1+0
	Luftige Angelegenheiten	0+0	0+0	2+0	2+0
	Pflanzenwachstum	1+0	2+0	0+0	3+0
	So schnell wie der Schall	0+0	0+0	1+0	1+0
	Wasser marsch	2+0	1+0	0+0	3+0
<i>Kreativ ausgelegte Aufgaben</i>	Arzneimittelforschung	1+0	0+0	0+0	1+0
	Düt un dat ut Pütt un Pann	0+0	1+0	0+0	1+0
	Energie in Natur und Technik	0+0	0+0	1+0	1+0
GESAMT		8+2	8+2	8+2	24+6

und Schüler in der anfänglichen Form der Aufgabe nicht eindeutig einem mehrstufigen Kategoriensystem zuzuordnen waren.

Dadurch, dass 14 Aufgaben nach der ersten Prä-Pilotierung entfernt wurden, musste eine entsprechende Anzahl an Aufgaben (und Themen) hinzugefügt werden, um eine gleiche Verteilung der Aufgaben auf Fächer und Kompetenzbereiche weiterhin zu gewährleisten. Die Aufgaben wurden teils neu entwickelt, teils aus dem vorhandenen Aufgabenpool in modifizierter Form entnommen. In Anbetracht der Tatsache, dass der erste Testentwurf mit 33 Aufgaben noch sehr lang war, wurde die künftige Testlänge auf 30 Aufgaben reduziert. Um auch Aspekte zu berücksichtigen, die Teilnehmende des Wettbewerbs Jugend forscht bzw. Schüler experimentieren für eine gute Leistung im Wettbewerb benötigen (siehe Abschnitt 5.1.4), wurden drei Aufgaben dem Test hinzugefügt, die offen ausgelegt waren und auf kreative und problem-lösende Aspekte abzielten. So verlangte eine Aufgabe z. B. die Aufzählung von Möglichkeiten, wie Forscherinnen und Forscher Naturstoffe so verändern können, dass sie als Medikament eingesetzt werden können (siehe Anhang D). Dadurch ergab sich die in Tabelle 5.2 abgebildete Verteilung der Aufgaben. Um die überarbeiteten und die neu entwickelten Aufgaben hinsichtlich der Grammatik und des Satzbaus noch einmal eingehend auf Verständlichkeit für Schülerinnen und Schüler zu prüfen, wurden sie einer Deutschlehrerin vorgelegt, bevor eine weitere Prä-Pilotierung stattfand.

Für die überarbeiteten Aufgaben wurde ein Erwartungshorizont formuliert, der für die offenen Aufgaben musterhafte und als komplett richtig zu bewertende Antworten enthielt. Da sich, wie bereits erwähnt, die Antworten der Schülerinnen und Schüler aus der ersten Prä-Pilotierung in bestimmte Kategorien einteilen ließen, wurde für die Auswertung weiterer Schülerantworten ein mehrstufiges Kategoriensystem entwickelt. Es wurde vorgesehen, Multiple-Choice-Aufgaben mit 0 und 1 Punkten sowie offene Aufgaben mit 0, 1, 2 oder 3 Punkten zu bewerten. Die Formulierung der einzelnen Kategorien erfolgte dabei wiederum

Tabelle 5.3: **Beispiel (FB11) aus dem mehrstufigen Kategoriensystem zur Bewertung der Schülerantworten** In der Aufgabe ging es um die Erklärung des Begriffes „autotroph“ anhand von drei Abbildungen, die verschiedene Tiere bei der Nahrungsaufnahme zeigten. Die Schülerantworten wurden nicht auf Rechtschreibung oder Grammatik geprüft.

<i>Kategorie</i>	<i>Schülerantwort</i>	<i>Anforderungen</i>	<i>Punkte</i>
1	Die Pflanze muss keine anderen Lebewesen fressen wie die bsp. Tiere. Sie kann sich selbst ernähren.	Es wird nur erklärt, dass Pflanzen keine anderen Lebewesen fressen, Tiere aber schon. Es wird nicht auf die Energiegewinnung eingegangen.	1
2	Die Pflanzen brauchen nur anorganische Nahrung (Licht; Mineralien; Wasser), die Tiere hingegen ernähren sich von organischer Nahrung (andere Tiere; Pflanzen).	Es wird erklärt, dass Pflanzen im Gegensatz zu Tieren nicht andere Lebewesen zur Nährstoffaufnahme benötigen, da sie ihre Nährstoffe anderweitig beziehen. Es wird nicht auf die Energiegewinnung eingegangen.	2
3	Autotrophe Tiere [Pflanzen] sind selbst ernährend, da sie aus anorganischen Stoffen organische Stoffe herstellen und daraus ihre Energie gewinnen können. Die drei Bilder zeigen heterotrophe Tiere, die nicht wie Autotrophe sich selbst ihre Energie herstellen, sondern die Energie von anderen Lebewesen durch die Verdauung aufnehmen. Dies kann durch Jagen oder Pflanzenessen geschehen.	Es wird erklärt, dass Pflanzen im Gegensatz zu Tieren selbst energiereiche organische Stoffe produzieren können.	3

aus den Schülerantworten heraus. Zu jeder Kategorie wurde jeweils eine reale Schülerantwort als Beispiel sowie die übergeordneten Anforderungen der jeweiligen Kategorie formuliert. Das Kategoriensystem von Aufgabe FB11 ist als Beispiel in Tabelle 5.3 abgebildet. (Die im Folgenden verwendeten Aufgabenkürzel setzen sich aus den Bereichen Fachwissen (F), Erkenntnisgewinnung (E), Kreativität(K), den Fächern Biologie (B), Chemie (C), Physik (P) und einer laufenden Nummer zusammen.)

In einer zweiten Prä-Pilotierung wurde die neue Testversion mit 28 schleswig-holsteinischen Gymnasiasten, die teilweise auch schon an naturwissenschaftlichen Wettbewerben teilgenommen hatten, erprobt (Alter 13-15,  $M = 14.29$ ,  $SD = 0.59$ , 10.7% Mädchen). Hierbei ging es vor allem um die erneute Prüfung der Aufgaben hinsichtlich ihrer Verständlichkeit und um die Ergänzung des Kategoriensystems. Daher wurden die Testaufgaben in drei Teile geteilt, wobei jedes Drittel von jeweils einem Drittel der teilnehmenden Schülerinnen und Schüler bearbeitet wurde. Eine Auswertung der Schülerantworten erfolgte daher auch auf einer rein qualitativen

Ebene, das Kategoriensystem wurde vervollständigt und die Aufgaben nur noch leicht in ihren Formulierungen überarbeitet.

## 5.4 Pilotierung des zweiten Testentwurfs

Für die Pilotierung des zweiten Testentwurfs wurden 30 Aufgaben gewählt, die nach Themen sortiert in drei Teile geteilt wurden. Die Teile wiederum wurden in drei verschiedenen Testentwürfen variiert, um Reihenfolge- sowie Positionseffekte zu vermeiden. In jedem Testentwurf wurden als einfach eingestufte Aufgaben an den Anfang gesetzt, um den teilnehmenden Schülerinnen und Schülern den Einstieg in die Aufgaben zu erleichtern.

Für die Pilotierung wurden 99 Schülerinnen und Schüler von zwei niedersächsischen und einem schleswig-holsteinischem Gymnasien ausgewählt (Alter 12-18,  $M = 13.90$ ,  $SD = 1.11$ , 42.4% Mädchen). Da die anvisierte Stichprobe für die erste Erhebung (Studie 1, siehe Kapitel 6) recht umfangreich ausfallen würde, wurde versucht, bereits in der Pilotierung ein relativ breites Spektrum an Schülerinnen und Schülern zu gewinnen. Um der Gruppe der Wettbewerbsteilnehmenden gerecht zu werden, wurde die Pilotierung auch in einer Jugendforscht-AG einer Schule durchgeführt. Bezogen auf die Gesamtstichprobe der Pilotierung fanden sich somit 51.5% an Schülerinnen und Schülern, die bereits an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten, davon 17.2% an Jugend forscht bzw. Schüler experimentieren. Schülerinnen und Schüler, die an der IJSO teilgenommen hatten, konnten bei der Pilotierung nicht im Besonderen berücksichtigt werden. Der Grund hierfür war vor allem, dass schulische Arbeitsgemeinschaften, die sich ausschließlich mit der IJSO beschäftigen, kaum in Norddeutschland existieren, d. h. die Rekrutierung einer repräsentativen Gruppe im Rahmen einer Pilotierung war kaum möglich.

Die Ergebnisse aus der Pilotierung wurden, anders als es später in der Studie 1 der Fall sein sollte, mit Methoden der Klassischen Testtheorie statt mit Methoden und Modellen der Item Response Theory (IRT) ausgewertet. Dieses Vorgehen lag darin begründet, dass es unterschiedliche Aussagen darüber gibt, ab welcher Stichprobengröße die Prüfung der Modellgeltung bzw. eine Schätzung von Parametern mithilfe der IRT zuverlässig ist. Nach D. H. Rost et al. (2004) ist zur Prüfung der Modellgeltung eine Stichprobengröße von  $N = 100$  schon ausreichend. Liegen starke a priori Hypothesen vor, kann auch schon eine Stichprobengröße von  $N = 40 - 50$  ausreichend sein. Andere Quellen wiederum sprechen, je nach zu schätzendem Modell, von erforderlichen Stichprobengrößen zwischen  $N > 150$  bis  $N < 1000$  (Becker, 2004; Zickar & Broadfoot, 2009). Edelen und Reeve (2007) nennen je nach Komplexität des Modells Stichprobengrößen zwischen  $N = 100$  bis  $N = 500$ , wobei davon ausgegangen wird, dass je komplexer ein Modell ist, desto höher die Stichprobe sein sollte. Des Weiteren wird empfohlen, dass zumindest für alle Kategorien eines Items Antworten vorliegen sollten. Da dies nach der Pilotierung noch nicht der Fall war, wurden die Pilotierungsergebnisse mit klassischen Verfahren ausgewertet. Ein weiterer Grund, der gegen die IRT-Modellierung der Pilotierungsergebnisse sprach, war, dass die Stichprobe noch keine Teilnehmenden des IJSO-Bundesfinales enthielt und daher nicht repräsentativ für die Hauptstichprobe war. Nach Edelen und Reeve (2007) stellt dies eine weitere Voraussetzung für die IRT-Modellierung dar.

Zur Prüfung der Reliabilität und der Objektivität des Kategoriensystems wurde eine dop-

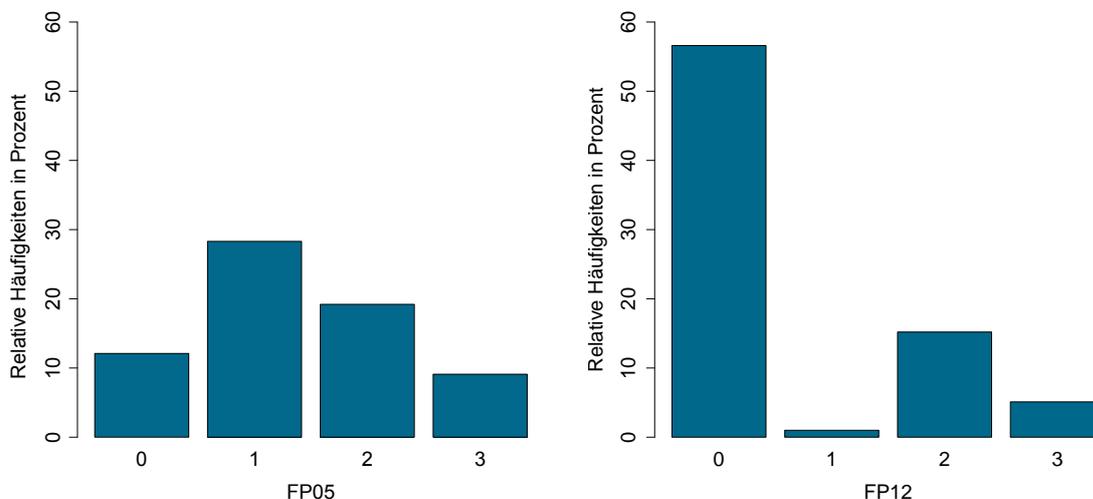


Abbildung 5.1: **Beispiele aus der Pilotierung einer erwartungskonformen (links) und unüblichen Verteilung (rechts) der Häufigkeiten auf die verschiedenen Kategorien bzw. erreichten Punkte**

pelte Bewertung von 10.1% der Tests aus der Pilotierung (10 aus 99) vorgenommen und die Beurteilerreliabilität, wie sie für die Prüfung der Reliabilität und der Objektivität des Schemas zur Klassifikation und Analyse von Wettbewerbsaufgaben verwendet wurde (siehe Abschnitt 4.2.2), ermittelt. Als Maß für die Beurteilerreliabilität wurde hier ebenso die Spearman-Rangkorrelation  $\rho$  gewählt. Es zeigte sich eine signifikante Spearman-Rangkorrelation von  $\rho = .79$ ,  $p < .001$ . Nach Heller und Rosemann (1981, S. 122) ist dieser Wert als ein hoher Zusammenhang zu interpretieren. Im Anschluss wurde eine Schwierigkeits- sowie eine Reliabilitätsanalyse nach der Klassischen Testtheorie durchgeführt. In einer Schwierigkeitsanalyse wird geprüft, wie schwer den Personen die Itembearbeitung fällt und wie gut die Items das zu messende Konstrukt erfassen (Bühner, 2011). Die Reliabilitätsanalyse beinhaltet die Betrachtung der Itemtrennschärfen sowie der Reliabilität und sollte zudem die Güte der Items bzw. des Testentwurfs prüfen.

#### 5.4.1 Schwierigkeitsanalyse

Für die Schwierigkeitsanalyse wurde zunächst die Verteilung der Aufgaben über Histogramme geprüft. 14 der 24 offenen Aufgaben zeigten dabei eine erwartungskonforme Verteilung. Erwartungskonform meinte hier, dass die Verteilung der Häufigkeiten von einem zu drei Punkten bzw. von Kategorie 1 zu Kategorie 3 abnahm, was für eine Zunahme der Schwierigkeit über die einzelnen Kategorien und somit für eine Differenzierung innerhalb der Aufgabe sprach. Beispiele für eine erwartungskonforme und eine unübliche Verteilung der Häufigkeiten sind in Abbildung 5.1 dargestellt. Auffällig ist hierbei, dass unabhängig von der erwartungskonformen Häufigkeitsverteilung über die Kategorien Schülerantworten relativ häufig mit null Punkten bewertet wurden, was auf eine höhere Schwierigkeit der Aufgaben hindeutete.

Bei zehn der 24 offenen Aufgaben ließen sich im Histogramm Muster erkennen, die vor

allem auf keine geordnete Reihenfolge der Kategorien (d.h. eine höhere Kategorie ist bspw. leichter als eine niedrigere Kategorie) oder auch als sehr schwierig zu beantwortende Kategorien hindeuteten. Ein Beispiel ist in Abbildung 5.1 dargestellt. Um die Schwierigkeiten vergleichen zu können, wurde der Schwierigkeitsindex ermittelt, der bei intervallskalierten Items dem Mittelwert aller Antworten über alle Personen entspricht (Bühner, 2011). Hierbei wurden die Schwierigkeitsindizes derart korrigiert, dass unbeantwortete Aufgaben in der Mittelwertberechnung nicht berücksichtigt wurden (Bühner, 2011). Zehn der 24 offenen Aufgaben lagen in ihrem Mittelwert dabei unterhalb der Skalenmitte ( $M < 1.50$ ), aber noch über einem Mittelwert von  $M = 1.00$ . Elf der 24 offenen Aufgaben lagen sogar unter einem Mittelwert von  $M = 1.00$ , nur eine offene Aufgabe hatte einen Schwierigkeitsindex, der über der Skalenmitte lag. Ein ähnliches Bild zeigte sich bei den Multiple-Choice-Aufgaben, bei denen fünf der sechs Aufgaben unter dem Skalenmittelwert ( $M < 0.50$ ) lagen. Die Standardabweichungen waren dabei durchgehend hoch (bei den offenen Aufgaben zwischen  $SD = 0.55$  und  $SD = 1.14$ , bei den Multiple-Choice-Aufgaben zwischen  $SD = 0.41$  und  $SD = 0.50$ ).

#### 5.4.2 Reliabilitätsanalyse

Für die Reliabilitätsanalyse wurde keine Korrektur der Daten vorgenommen, d. h. fehlende Werte gingen als Bewertung mit null Punkten in die Analyse ein. Für die Reliabilität des Gesamttests mit 30 Aufgaben ergab sich ein akzeptabler Wert für Cronbachs Alpha von  $\alpha = .71$ . Die Reliabilitätsanalyse zeigte dabei, dass ein Ausschluss von Items den Wert für Cronbachs Alpha maximal auf  $\alpha = .72$  verbessert hätte. Die Trennschärfen von elf Items lagen über  $r_{i(x-i)} > .30$ . Trennschärfen unter  $r_{i(x-i)} < .30$  werden in der Regel als niedrig eingestuft, wobei es nach Bühner (2011) keine feste Untergrenze gibt, ab der man Aufgaben sinnvoll ausschließt.

Neben der Reliabilität über alle Aufgaben wurden ebenso die Reliabilitäten für einzelne Testteile (Kompetenzbereiche und Fächer) bestimmt. Die Ergebnisse lagen in nicht akzeptablen Bereichen zwischen Werten für Cronbachs Alpha von  $\alpha = .34$  (Chemie-Aufgaben) und  $\alpha = .61$  (Aufgaben Erkenntnisgewinnung Biologie, Chemie und Physik), was im Vergleich zur akzeptablen Reliabilität des Gesamttests einerseits an der höheren Anzahl an Aufgaben im Gesamttest im Vergleich zu den Subtests gelegen haben könnte, andererseits gab dieser Befund erste Hinweise auf das Vorliegen eines eindimensionalen Konstrukts.

#### 5.4.3 Weitere Auswertungen

Neben der Schwierigkeits- und der Reliabilitätsanalyse wurden Korrelationen nach Pearson bzw. bei dichotomen Variablen punktbiserial Korrelationen zwischen der Testleistung (die hier der im Test erreichten Gesamtpunktzahl entsprach) und den erhobenen Begleitvariablen (Alter, Zeugnisnoten in Biologie, Chemie und Physik, Geschlecht, Wettbewerbsteilnahme) ermittelt. Es zeigte sich erwartungskonform eine mittlere (negative) Korrelation zwischen der Testleistung und dem Mittelwert aus den Zeugnisnoten in Biologie, Chemie und Physik ( $r = -.41$ ,  $p < .01$ ), die negative Korrelation kommt hier dadurch zustande, dass geringe Zahlenwerte gute Leistungen beschreiben und umgekehrt) sowie eine mittlere (positive) Korrelation zwischen der Testleistung und dem Alter ( $r = .34$ ,  $p < .001$ ). Zwischen der Testleistung und der Frage, ob die Teilnehmenden der Studie schon an einem naturwissenschaftlichen Wettbewerb teilgenommen hatten, zeigte sich eine niedrige (positive) Korrelation ( $r_{pb} = .27$ ,

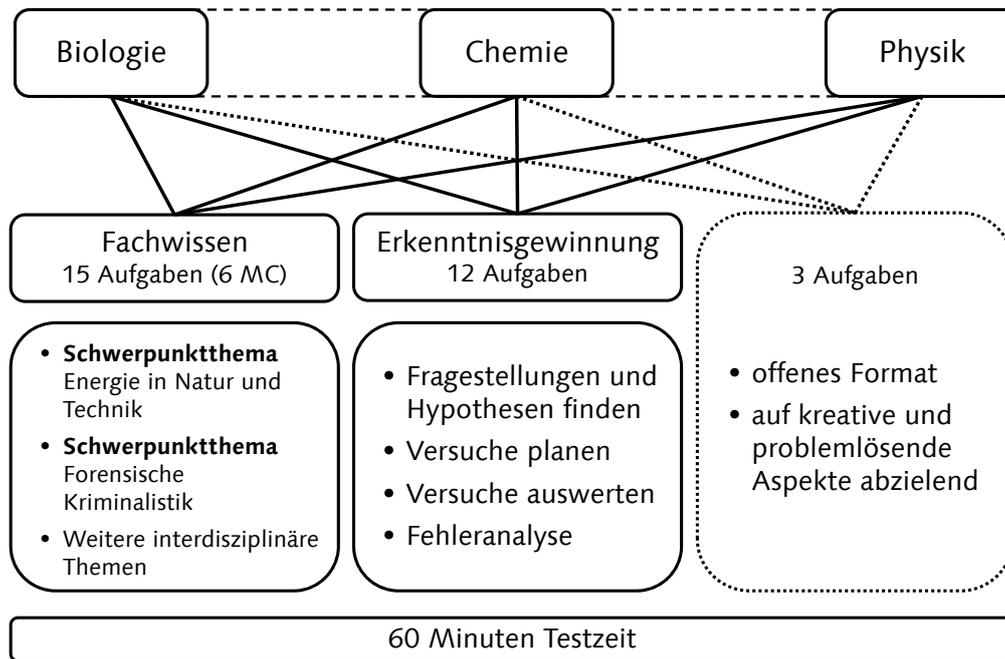


Abbildung 5.2: **Finales Testinstrument nach Abschluss und Auswertung der Pilotierungen**

$p < .01$ ). Zwischen der Testleistung und dem Geschlecht konnte keine signifikante Korrelation ermittelt gefunden.

## 5.5 Zusammenfassung und Fazit

Nach mehreren Schritten der Entwicklung und Erprobung eines Instruments zur Erfassung der naturwissenschaftlichen Kenntnisse und Fähigkeiten von Wettbewerbsteilnehmenden lag ein Testinstrument vor, welches für eine Testzeit von 60 Minuten 30 Aufgaben der Kompetenzbereiche Fachwissen und Erkenntnisgewinnung der Fächer Biologie, Chemie und Physik umfasste. Drei weitere Aufgaben waren offen ausgelegt und zielten auf kreative und problemlösende Aspekte ab. Eine grafische Darstellung des Testinstrumentes ist in Abbildung 5.2 zu sehen. Die Bewertung der Antworten der Schülerinnen und Schüler erfolgte anhand eines an den realen Schülerantworten orientierten mehrstufigen Kategoriensystems, welches eine hohe Beurteilerreliabilität aufwies.

Die Schwierigkeitsanalyse zeigte, dass das Testinstrument eine zu hohe Schwierigkeit für die Schülerinnen und Schüler aus der Pilotierung aufwies. Allerdings musste beachtet werden, dass die Gruppe der Teilnehmenden des IJSO-Bundesfinales die Aufgaben noch nicht bearbeitet hatten. Da leichtere Aufgaben evtl. nicht präzise genug in dieser Gruppe differenziert hätten, wurden die Aufgaben in der Form, wie sie in der Pilotierung eingesetzt wurden, beibehalten.

Die Reliabilitätsanalyse für den gesamten Testentwurf zeigte ein akzeptables Ergebnis für

Cronbachs Alpha von  $\alpha = .71$ . Da durch Ausschluss von Aufgaben die Reliabilität nur minimal hätte verbessert werden können, wurde aus inhaltlichen Gründen keine Aufgabe aus dem Testinstrument entfernt. Trotz zum Teil niedriger Trennschärfen wurden auch hier aus v. a. inhaltlichen Gründen keine Aufgaben für die folgenden Erhebungen ausgeschlossen. Bezüglich der Trennschärfen ist zudem anzumerken, dass niedrige Trennschärfen mit hohen Itemschwierigkeiten zusammenhängen können (Bühner, 2011), was bei den meisten Aufgaben der Fall war.

Des Weiteren fanden sich erwartungskonform signifikante Korrelationen zwischen der Testleistung und dem Alter, den Zeugnisnoten in Biologie, Chemie und Physik sowie zu der Frage, ob die Schülerinnen und Schüler schon an einem naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten, was im Rahmen der Prüfung der Testqualität erste Hinweise auf die Validität des Testinstruments gab.

Die Aufgaben, die Auffälligkeiten in der Schwierigkeitsanalyse und ihren Trennschärfen zeigten, wurden noch einmal hinsichtlich der Arbeitsaufträge überarbeitet, wobei die Arbeitsaufträge meist konkretisiert, präzisiert oder zum Teil durch Hervorhebung wichtiger Wörter oder Satzteile der Aufgabenstellung vereinfacht wurden. Das Testinstrument wurde anschließend im Rahmen von Studie 1 unter der Fragestellung, ob sich naturwissenschaftliche Kenntnisse und Fähigkeiten von erfolgreichen Wettbewerbsteilnehmenden im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern mit einem gemeinsamen Testinstrument erfassen lassen, eingesetzt und validiert.

## Kapitel 6

# Validierung des Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten von Wettbewerbsteilnehmenden (Studie 1)

Im Rahmen von Studie 1 wurde das entwickelte Testinstrument zu naturwissenschaftlichen Kenntnissen und Fähigkeiten von Wettbewerbsteilnehmenden (Kapitel 5) mit Teilnehmenden des Bundesfinales der IJSO 2013, Teilnehmenden zweier Landeswettbewerbe Jugend forscht bzw. Schüler experimentieren 2014 sowie Schülerinnen und Schülern, die mindestens einmal oder gar nicht an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten, erprobt, wobei sich eine Gesamtstichprobe von  $N = 226$  ergab. Zur Prüfung der Testqualität und der Testziele wurden Methoden der probabilistischen Testtheorie angewendet sowie Korrelationen zu Begleitvariablen berechnet. Des Weiteren wurden zur Prüfung der Testziele multiple lineare sowie binäre logistische Regressionen zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten, für die Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb sowie für Erfolg in einem naturwissenschaftlichen Schülerwettbewerb durchgeführt. Zudem wurden mögliche Unterschiede zwischen verschiedenen Subgruppen varianzanalytisch untersucht. Im Zentrum stand die Beantwortung der folgenden Forschungsfrage:

Lassen sich die naturwissenschaftlichen Kenntnisse und Fähigkeiten von erfolgreichen Wettbewerbsteilnehmenden im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern mit einem gemeinsamen Testinstrument erfassen?

### 6.1 Stichprobe

Die Stichprobe von Studie 1 setzte sich aus 226 Schülerinnen und Schülern zusammen, wobei 39 Schülerinnen und Schüler Teilnehmende des IJSO-Bundesfinales 2013 und 30 Schülerinnen und Schüler Teilnehmende von Landeswettbewerben Jugend forscht bzw. Schüler experimentieren 2014 waren (durchschnittliches Alter der Gesamtstichprobe  $M = 14.47$ ,

$SD = 1.07$ , Altersgruppe 11-18, 41.1% Mädchen). Die Gruppe der Teilnehmenden des IJSO-Bundesfinales 2013 wies ein durchschnittliches Alter von  $M = 14.47$ ,  $SD = 0.64$ , (Altersgruppe 13-15) auf, 33.3% der Teilnehmenden des IJSO-Bundesfinales 2013 waren Mädchen. Die Teilnehmenden entstammten dem gesamten Bundesgebiet ausgenommen Berlin, Bremen, Hamburg, Mecklenburg-Vorpommern und dem Saarland. Die Gruppe der Teilnehmenden von Landeswettbewerben Jugend forscht bzw. Schüler experimentieren 2014 wies ein durchschnittliches Alter von  $M = 14.10$ ,  $SD = 1.81$ , (Altersgruppe 11-18) auf, 40.0% der Teilnehmenden waren Mädchen. Die Erhebungen wurden in den Landeswettbewerben Jugend forscht und Schüler experimentieren in Hamburg sowie dem Landeswettbewerb Schüler experimentieren in Niedersachsen durchgeführt, wobei in Niedersachsen nur der Fragebogen eingesetzt wurde (zu den weiteren Erhebungsinstrumenten und dem Ablauf der Erhebungen siehe Abschnitt 6.3). Die Gruppe der Schülerinnen und Schüler aus Klassen bzw. Schulen von IJSO-Teilnehmenden des Bundesfinales 2013 stammte von drei Schulen aus Bayern, Niedersachsen und Sachsen. Es handelte sich hierbei um eine Gemeinschaftsschule, ein staatliches Gymnasium sowie eine MINT-EC-Schule, d. h. ein Gymnasium des nationalen Excellence-Netzwerks von Schulen mit ausgeprägtem MINT-Profil. Die Schülerinnen und Schüler besuchten die Klassenstufen 8, 9 und 10. Die Auswahl dieser Teilstichprobe war über den Kontakt von Lehrkräften, die ihre Schülerinnen oder Schüler bis zum Bundesfinale der IJSO betreut hatten, erfolgt. Die Bereitschaft zur Teilnahme an Studie 1 war jedoch bedauerlicherweise eher gering. Die Gruppe von Schülerinnen und Schülern aus Klassen bzw. Schulen von Teilnehmenden des IJSO-Bundesfinales 2013 wies ein durchschnittliches Alter von  $M = 14.52$ ,  $SD = 0.95$  (Altersgruppe 12-16) auf, 48.7% waren Mädchen.

## 6.2 Erhebungsinstrumente

Neben dem entwickelten Testinstrument zu naturwissenschaftlichen Kenntnissen und Fähigkeiten in drei verschiedenen, sich in der Reihenfolge der Themenblöcke unterscheidenden Versionen (Beispielaufgaben siehe Anhang D), wurden ein umfangreicher Fragebogen sowie Teile des KFT (Heller et al., 2000) eingesetzt. Die Zuordnung der einzelnen Erhebungsinstrumente zueinander erfolgte durch einen Code, den die Teilnehmenden vor Beginn der Bearbeitung eines Erhebungsinstruments selbst reproduzierbar generierten. Im Folgenden werden der Fragebogen sowie der KFT kurz erläutert.

### 6.2.1 Fragebogen

Der Fragebogen zum IKoN bestand aus bereits getesteten und zum Teil standardisierten Skalen. Er wurde im Rahmen des Gesamtprojektes eingesetzt (Dierks et al., 2014, 2016; Wentorf et al., 2015), weshalb im Folgenden nicht näher auf die Hintergründe des Fragebogens eingegangen wird. Die einzelnen erhobenen Informationen und Konstrukte sind in Tabelle 6.1 aufgeführt, der Fragebogen ist in Anhang E einzusehen.

### 6.2.2 Kognitiver Fähigkeitstest

Zur Bewertung allgemeiner kognitiver Teilfähigkeiten wurde der KFT nach Heller et al. (2000) eingesetzt. Aus testökonomischen Gründen wurden zwei Subskalen der Testform A genutzt, der Q-Test 2 (quantitativer Teil - Zahlenreihen) sowie der N-Test 1 (nonverbaler (figuraler)

Tabelle 6.1: In Studie 1 erhobene Informationen und Konstrukte im Fragebogen zum IKoN

<i>Erhobene Information bzw. erhobenes Konstrukt</i>	<i>Beschreibung und Quelle</i>
Demographische Informationen	Geschlecht, Alter, Schulform, Bundesland, letzte Zeugnisnoten und Interesse an den jeweiligen Fächern (eigene Entwicklung)
Allgemeine Fragen zu naturwissenschaftlichen Wettbewerben	Bekanntheit naturwissenschaftlicher Wettbewerbe, interessante Preise für die Teilnahme an einem Wettbewerb, Image Jugend forscht/ IJSO (eigene Entwicklung)
Allgemeine Fragen an Schülerinnen und Schüler, die schon an naturwissenschaftlichen Wettbewerben teilgenommen haben	An welchen Wettbewerben hatten Schülerinnen und Schüler schon teilgenommen, welcher war der erste Wettbewerb, wie oft hatten Schülerinnen und Schüler schon teilgenommen, Beweggründe zur Wettbewerbsteilnahme (eigene Entwicklung)
Akademisches (absolutes schulisches) Selbstkonzept	5 Items, vierstufige Likert-Skala (Schöne et al., 2002)
Naturwissenschaftliches Selbstkonzept	7 Items, vierstufige Likert-Skala (Daniels, 2008; Köller et al., 2000; Schöne et al., 2002)
Zielorientierungen im Naturwissenschaftsunterricht	16 Items, vierstufige Likert-Skala (Spinath et al., 2002; Blankenburg et al., 2015)
Interesse an Tätigkeiten, die häufig in der Schule ausgeübt werden, und Frage, ob die Tätigkeiten schon durchgeführt wurden	30 Items, vierstufige Likert-Skala (Dierks et al., 2014)
Interesse an Tätigkeiten, die im naturwissenschaftlichen Schulunterricht, in naturwissenschaftlichen Fördermaßnahmen oder in einem naturwissenschaftlichen Beruf ausgeübt werden	3 x 28 Items, vierstufige Likert-Skala (Dierks et al., 2014)
Eigenschaften eines Naturwissenschaftlers	8 Items, fünfstufige Antwortskala („Spinne“, Wentorf et al., 2015)
Einschätzung, welche Tätigkeiten ein Wissenschaftler regelmäßig ausübt	30 Items, vierstufige Likert-Skala (Wentorf et al., 2015)
Interesse daran, Tätigkeiten eines Wissenschaftlers später selbst durchzuführen	30 Items, vierstufige Likert-Skala (Wentorf et al., 2015)
Selbstwirksamkeitserwartung für die spätere Ausübung von Tätigkeiten eines Wissenschaftlers	30 Items, vierstufige Likert-Skala (Wentorf et al., 2015)

Teil - Figurenklassifikation). Der Q-Test 2 ist ein Test zur Ermittlung zahlengebundenen, logischen Denkens, der N-Test 1 prüft das logische Denken mit Bezug zu anschauungsgebundenen Aspekten räumlichen Vorstellungsvermögens (Heller et al., 2000). Auf den verbalen Teil wurde hier verzichtet, da diese Tests vor allem das Sprachverständnis bzw. sprachgebundenes logisches Denken erfassen. Die teilnehmenden Schülerinnen und Schüler bearbeiteten zunächst den quantitativen und im Anschluss den nonverbalen Teil. Zum Ende der Erhebungen von Studie 1 wurde auf den quantitativen Teil verzichtet, um den teilnehmenden Schülerinnen und Schülern eine Pause zwischen dem Ausfüllen der verschiedenen Erhebungsinstrumente zu ermöglichen. Die nonverbale (figurale) Skala wurde beibehalten, da mit nonverbalen Testskalen die fluide Intelligenz erfasst wird (vgl. Cattell, 1987), die die Basisprozesse des Denkens umfasst und weniger soziokulturell beeinflusst wird als die kristalline (oder kristallisierte) Intelligenz, die mit den verbalen und quantitativen Testskalen des KFT erfasst wird (Heller et al., 2000; Schmechtig, Adolph & Melzer, 2008).

Da hier lediglich Subskalen des KFT verwendet wurden, wurden keine genormten Intelligenzwerte ermittelt. Für die spätere Auswertung wurden daher nur Summenwerte berücksichtigt, die somit auch nur innerhalb der hier vorgesehenen Stichprobe zu interpretieren waren.

### 6.3 Ablauf der Erhebungen

Die Teilnahme an der Studie war freiwillig. Im Vorfeld der Erhebungen wurden die Genehmigungen zur Durchführung der wissenschaftlichen Erhebungen von allen potentiell teilnehmenden Bundesländern eingeholt. Je nach Bundesland wurden auch die schriftlichen Einverständniserklärungen der teilnehmenden Schulleitungen erfragt. Da alle teilnehmenden Schülerinnen und Schüler minderjährig waren, wurde vor jeder Erhebung das schriftliche Einverständnis der Erziehungsberechtigten eingeholt, in Abhängigkeit vom jeweiligen Bundesland wurde teilweise auch das schriftliche Einverständnis der teilnehmenden Schülerinnen und Schüler erbeten. Der Ablauf der Erhebungen selbst in den einzelnen Wettbewerben und Schulen unterschied sich aus organisatorischen Gründen minimal voneinander.

Teilnehmende des IJSO-Bundesfinales 2013 wurden mit der Einladung zum Bundesfinale über die Wettbewerbsleitung angeschrieben, sie erhielten neben einem Informationsschreiben die Einverständniserklärung für die Erziehungsberechtigten sowie den Fragebogen. Der Fragebogen wurde vor dem Bundesfinale von den Teilnehmenden zu Hause bearbeitet (die Bearbeitung nahm ca. 30 Minuten in Anspruch) und beim Bundesfinale im Oktober 2013 bei der Wettbewerbsleitung abgegeben. Somit sollte sichergestellt werden, dass die Teilnehmenden nicht von Erlebnissen und Eindrücken des Bundesfinales beeinflusst werden würden. Die Bearbeitung der Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten sowie des KFT erfolgte während des Bundesfinales vor Ort im Oktober 2013 innerhalb von 90 Minuten (60 Minuten Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten, 30 Minuten Aufgaben des KFT nach standardisiertem Testverfahren) nach Abschluss aller für den Wettbewerb zu erbringenden Leistungen, jedoch vor Bekanntgabe der Ergebnisse. So sollte zum einen gewährleistet werden, dass der Ablauf des Wettbewerbs nicht gestört wird, zum anderen sollten die Teilnehmenden in ihrer Motivation zur Teilnahme an den Erhebungen nicht durch Ergebnisse des Wettbewerbs beeinflusst werden. Die Erhebungen selbst wurden von der Autorin durchgeführt.

Im Anschluss an das Bundesfinale wurden betreuende Lehrkräfte von Teilnehmenden des IJSO-Bundesfinales 2013 kontaktiert und um Unterstützung für die Durchführung der Erhebungen gebeten. Bedauerlicherweise beschränkten sich die Rückmeldungen lediglich auf drei Schulen. Das anvisierte Ziel, in den drei Schulformen Gemeinschaftsschule, Gymnasium und Gymnasium mit mathematisch-naturwissenschaftlichen Schwerpunkt (bzw. MINT-EC-Schule) zu erheben sowie die Schulen möglichst breit im Bundesgebiet zu verteilen (erhoben wurde für Studie 1 in Bayern, Niedersachsen und Sachsen) wurde dennoch erreicht, auch wenn für Studie 1 kaum von einer repräsentativen Auswahl gesprochen werden konnte. Die Auswahl, welche Klassenstufen (8, 9 oder 10) sowie welche und wie viele Klassen an den Erhebungen teilnahmen, wurde von den Schulen bestimmt, wobei darum gebeten wurde, Klassen von IJSO-Teilnehmenden zu wählen. Auch in den Klassen bzw. Schulen von IJSO-Teilnehmenden des Bundesfinales 2013 erhielten die Schülerinnen und Schüler den Fragebogen vorab und bearbeiteten ihn zu Hause. Die Bearbeitung der Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten sowie des KFT erfolgte im Rahmen einer Doppelstunde unter Anleitung der Autorin sowie einer vorab in standardisierter Weise instruierten studentischen Hilfskraft vor Ort der entsprechenden Schulen. Die Erhebungen wurden zwischen Januar und März 2014 durchgeführt.

Für die Erhebungen in den Landeswettbewerben Jugend forscht bzw. Schüler experimentieren wurde ein Vorgehen ähnlich dem in der IJSO angestrebt, dies konnte aus organisatorischen Gründen allerdings nicht konsequent durchgeführt werden. Aufgrund des engen Zeitplans im Wettbewerb Jugend forscht bzw. Schüler experimentieren, für die im Gegensatz zum sechstägigen IJSO-Bundesfinale ein bis maximal drei Tage angesetzt waren, fanden sich lediglich zwei Landeswettbewerbe, in denen die Erhebungen ermöglicht werden konnten. Im Landeswettbewerb Schüler experimentieren in Niedersachsen konnte so auch nur der Fragebogen eingesetzt werden, den die Teilnehmenden jedoch auch hier vor dem Wettbewerb selbst bearbeiteten. Im Landeswettbewerb Hamburg bearbeiteten die Teilnehmenden den Fragebogen ebenso vor dem Wettbewerb zu Hause, die Bearbeitung der Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten sowie des KFT erfolgte, wie auch beim IJSO-Bundesfinale, während des Landeswettbewerbs vor Ort. Als Zeitpunkt für die Erhebungen wurde analog zum IJSO-Bundesfinale ein Zeitpunkt nach den für den Wettbewerb erbrachten Leistungen und vor Bekanntgabe der Ergebnisse gewählt. Die Erhebungen fanden im März und April 2014 statt und wurden von einer studentischen Hilfskraft durchgeführt, die vorab in standardisierter Weise instruiert worden war.

Ursprünglich sollten in Analogie zur IJSO Teilnehmende aus Klassen bzw. Schulen von Teilnehmenden des Wettbewerbs Jugend forscht bzw. Schüler experimentieren für die Erhebungen gewonnen werden. Da jedoch nur zwei Landeswettbewerbe an den Erhebungen teilnahmen, beschränkte sich die Anzahl potentiell teilnehmender Schulen stark. Des Weiteren hatte die Erfahrung mit Schulen von IJSO-Teilnehmenden gezeigt, dass die (positiven) Rückmeldungen sehr gering ausfielen. Die Anzahl teilnehmender Schülerinnen und Schüler des Wettbewerbs Jugend forscht bzw. Schüler experimentieren war mit 30, von denen einige auch nur den Fragebogen bearbeitet hatten, auch zu gering, um aussagekräftige Analysen im Vergleich zu IJSO-Teilnehmenden durchzuführen. Aus diesem Grund wurde Studie 1 nach den Erhebungen in den Landeswettbewerben Jugend forscht bzw. Schüler experimentieren nicht mehr fortgeführt.

## 6.4 Auswertung

Für die Auswertung wurde das bestehende Kategoriensystem gegenüber dem Kategoriensystem aus der Pilotierung zunächst überarbeitet und ergänzt. Bei einigen Aufgaben wurde das dreistufige Kategoriensystem auf ein zweistufiges Kategoriensystem reduziert, da sich die Antworten der Schülerinnen und Schüler zum Teil nicht eindeutig der einen oder anderen Kategorie zuordnen ließen. Für eine der offenen, auf kreative und problemlösende Aspekte abzielende Aufgabe (KC01) wurden zwei Kategoriensysteme entwickelt, die unterschiedliche Aspekte in der Bewertung berücksichtigten. Für diese Aufgabe gab es daher zwei voneinander unabhängige Bewertungen (KC01i, und KC01o). Schließlich wurden die Antworten aller Schülerinnen und Schüler, die an Studie 1 teilnahmen, nach dem Kategoriensystem bewertet. Mit Methoden der probabilistischen Testtheorie wurden darauf folgend die Testqualität und die Testziele des eingesetzten Instruments geprüft, zudem wurden Korrelations-, Regressions- und Varianzanalysen durchgeführt. Im Folgenden werden die eingesetzten Methoden zunächst erläutert, die Ergebnisse werden in Abschnitt 6.5 präsentiert.

### 6.4.1 Beurteilerreliabilität

Da das bestehende Kategoriensystem im Anschluss an die Pilotierung überarbeitet und ergänzt wurde, wurde zur Prüfung der Reliabilität und der Objektivität des Kategoriensystems erneut die Beurteilerreliabilität bestimmt. Es wurde eine doppelte Bewertung von 10.1% der Tests (22 aus insgesamt 218, ohne Pilotierung) vorgenommen und die Beurteilerreliabilität über die Spearman-Rangkorrelation  $\rho$  (siehe Abschnitte 4.2.2 und 5.4.3) ermittelt. Es ergab sich eine signifikante Spearman-Rangkorrelation mit einem Wert von  $\rho = .98$ ,  $p < .001$ . Nach Heller und Rosemann (1981, S. 122) ist dieser Wert als ein hoher Zusammenhang zu interpretieren.

### 6.4.2 Umgang mit fehlenden Werten in Itemantworten

Es gibt verschiedene Situationen, durch die fehlende Werte in Itemantworten entstehen können (Pohl & Carstensen, 2012; Pohl, Gräfe & Rose, 2014; Schöps & Saß, 2012):

- Den Teilnehmenden einer Studie werden einige Items gar nicht erst vorgelegt, z. B. wenn verschiedene Testhefte existieren.
- Ein Item wird ungültig beantwortet, z. B. wenn in einer einfachen Multiple-Choice-Aufgabe zwei mögliche Antworten angekreuzt werden.
- Items werden ausgelassen.
- Items werden aufgrund von limitierter Testzeit nicht erreicht.

Da in Studie 1 alle Teilnehmenden dieselben Aufgaben erhielten, existierten im Datensatz keine fehlenden Werte aufgrund nicht vorgelegter Items. Ungültige Antworten wurden im Rahmen dieser Studie mit null Punkten bewertet und daher nicht als fehlender Wert gekennzeichnet.

Während fehlende Werte aufgrund von nicht vorgelegten Items meist vollkommen zufällig sind, trifft dies auf ausgelassene und aufgrund der Testzeit nicht erreichte Items in der Regel nicht zu, weshalb diese Arten von fehlenden Werten gewöhnlicherweise nicht einfach

ignoriert werden können. Die Tendenz Items auszulassen kann mit der Itemschwierigkeit und der Personenfähigkeit zusammenhängen (z. B. Mislevy & Wu, 1988), während nicht erreichte Items mit der Bearbeitungsstrategie der Aufgaben, der Motivation und/ oder der Personenfähigkeit zusammenhängen können (z. B. Stocking, Eignor & Cook, 1988). Mehrere Studien untersuchten bislang verschiedene Ansätze, wie mit diesen Arten von fehlenden Werten umgegangen werden kann (z. B. Ayala, Plake & Impara, 2001; Finch, 2008; Lord, 1974; Ludlow & O’Leary, 1999; Pohl et al., 2014). Dabei zeigte sich, dass das Ignorieren fehlender Werte, multiple Imputationen sowie modellbasierte Herangehensweisen zu weitgehend unverfälschten Parameterschätzungen führten. Verfälschte Parameterschätzungen und überschätzte Reliabilitäten wurden hingegen gefunden, sofern fehlende Werte mit falsch beantworteter Antworten gleichgesetzt wurden (Pohl et al., 2014). Pohl et al. (2014) verglichen im Rahmen des Projektes Nationales Bildungspanel (NEPS) die Parameterschätzungen, wenn fehlende Werte ignoriert werden und wenn die Tendenz zu fehlenden Werten im IRT-Modell berücksichtigt wird. Dabei wurden keine nennenswerten Unterschiede in den Parameterschätzungen gefunden, lediglich das Gleichsetzen fehlender mit falsch beantworteter Antworten führte zu nennenswerten Unterschieden in den Schätzungen der Parameter. Da multiple Imputationen im Rahmen von Kompetenztestungen nicht häufig angewendet werden (Pohl et al., 2014), wurden die fehlenden Werte für die Auswertungen ignoriert.

Um einen ersten Eindruck von der Schwierigkeit der Aufgaben sowie vom Umfang des Tests zu erhalten, wurden die fehlenden Werte analysiert. Dazu wurde die Anzahl an Teilnehmenden der Studie ermittelt, die das Ende des Tests nicht erreichten. Maßgeblich war hierfür das Erreichen der letzten Aufgabe im Test. Des Weiteren wurde analysiert, wie viele Items pro Person ausgelassen wurden und wie viele fehlende Werte insgesamt pro Person auftraten. Auf Ebene der Items wurde der Anteil fehlender Werte pro Aufgabe berechnet.

### 6.4.3 Skalierung der Itemantworten

#### Probabilistische versus klassische Testmodelle

Die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten wurden mit Hilfe eines probabilistischen Testmodells für ordinal skalierte Itemantworten ausgewertet. Hieraus generiert wurden die Fähigkeitsparameter für die Personen, mit denen weitere analytische Verfahren durchgeführt wurden. Zur Anwendung kam hier das ordinale Rasch-Modell, genannt Partial-Credit-Modell (PCM). Die Gründe für die Verwendung und somit die Vorteile probabilistischer Testmodelle im Vergleich zu Verfahren der klassischen Testtheorie werden im Folgenden kurz erläutert.

#### Modellierung der Lösungswahrscheinlichkeit mithilfe einer logistischen Funktion

Viele probabilistische Testmodelle basieren auf der logistischen Funktion. Im Gegensatz zu deterministischen Funktionen, die nur die Lösungswahrscheinlichkeiten null und eins für ein Item definieren, berücksichtigt die logistische Funktion, dass eine fähige Person immer auch eine (geringe) Wahrscheinlichkeit aufweist, ein leichtes Item nicht zu lösen, gleichermaßen wie berücksichtigt wird, dass eine weniger fähige Person immer auch eine (geringe) Wahrscheinlichkeit aufweist, ein schweres Item zu lösen. Das Antwortverhalten einer Person ist folglich probabilistisch und liegt im Bereich zwischen null und eins. Sind die Personenfähigkeit und die Itemschwierigkeit bekannt, kann mit Hilfe der so genannten *Item-Characteristic-Curves* (ICC)

für jedes Item die Lösungswahrscheinlichkeit für beliebige Personen ermittelt werden. Mit zunehmender Differenz zwischen Personenfähigkeit und Itemschwierigkeit, d. h. je mehr die Personenfähigkeit die Itemschwierigkeit übersteigt, desto höher ist die Lösungswahrscheinlichkeit für ein Item (Bühner, 2011).

**Betrachtung von Antwortmustern statt von Summenwerten** Modelle der klassischen Testtheorie gehen üblicherweise davon aus, dass alle Items eines Leistungstests gleich und mit dem Faktor eins zur zu messenden Fähigkeits- oder Eigenschaftsausprägung beitragen. Aus den einzelnen Punkten, die bei der Beantwortung der Items erreicht wurden, kann somit ein Summenwert gebildet werden, der aufgrund der gleichen Gewichtung aller Items und der erreichten Punkte eine erschöpfende Statistik für die Fähigkeits- oder Eigenschaftsausprägung darstellt. Der Nachteil hierbei ist jedoch, dass nicht immer davon ausgegangen werden kann, dass alle Items eines Leistungstests gleich sind, weshalb Modelle der klassischen Testtheorie in diesem Fall nicht eindeutig interpretierbar sein und keine erschöpfende Statistik für die Fähigkeits- oder Eigenschaftsausprägung darstellen können. Eine Varianz in der Schwierigkeit der Items ist jedoch gerade bei Leistungstests meist erwünscht, um leistungsschwache Schülerinnen und Schüler nicht zu überfordern und leistungsstarke nicht zu unterfordern. Probabilistische Testmodelle betrachten statt dem Summenwert Antwortmuster. Die Itemantworten werden dabei als manifeste Ausprägungen einer latenten Variable angesehen, die auch die Korrelationen der Itemantworten bedingt (Bühner, 2011).

**Vorhersage des Antwortverhaltens** Durch die Betrachtung von Antwortmustern statt von Summenwerten wie in der klassischen Testtheorie, kann das Antwortverhalten einer Person durch ihre Fähigkeits- oder Eigenschaftsausprägung und weiterer Itemparameter wie z. B. der Itemschwierigkeit vorhergesagt werden. Andersherum betrachtet treffen probabilistische Testmodelle Annahmen darüber, von welchen Modellparametern (Fähigkeits- oder Eigenschaftsausprägung einer Person, Itemschwierigkeit, Ratewahrscheinlichkeit, Itemtrennschärfe, etc.) die Lösungswahrscheinlichkeit eines Items abhängt. Es gilt das Konzept der lokalen stochastischen Unabhängigkeit: Es besagt, dass die Wahrscheinlichkeit der Beantwortung eines Items ausschließlich mit der Fähigkeits- oder Eigenschaftsausprägung einer Person und der Itemschwierigkeit zusammenhängt. Die Wahrscheinlichkeit ein Item zu lösen beeinflusst somit nicht die Wahrscheinlichkeit ein anderes Item zu lösen. Lokale stochastische Unabhängigkeit liegt vor, wenn das Produkt der Lösungswahrscheinlichkeiten jedes einzelnen Items der Wahrscheinlichkeit entspricht, alle Items zu lösen (Masters, 1982; Bühner, 2011).

### Das Partial-Credit-Modell

Da es sich bei den Antworten auf die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten um ordinal skalierte Daten handelt, wurden die Antworten anhand eines entsprechenden Modells ausgewertet. Hierfür wurde in den Auswertungen das einparametrische logistische PCM (auch: das ordinale Rasch-Modell) genutzt. Als Itemparameter werden hier Schwellenparameter berechnet, die den Punkt definieren, an dem es genauso wahrscheinlich ist, Kategorie  $k$  oder Kategorie  $k + 1$  (usw.) zu wählen. Die Schwellenparameter sollten dabei geordnet sein, d. h. von Kategorie zu Kategorie zunehmen, was bedeutet, dass eine höhere Kategorie auch eine höhere Fähigkeits- oder Eigenschaftsausprägung repräsentiert (Bühner, 2011; D. H. Rost et al., 2004). Bei nicht geordneten Schwellenparametern kann jedoch trotzdem von geordneten Kategorien ausgegangen werden, sofern mit zunehmender Personenfähigkeit die

erreichte Kategorie zunimmt (R. J. Adams, Wu & Wilson, 2012; Wetzel & Carstensen, 2014), was z. B. an dem Mittelwert der Personenfähigkeiten aller Probandinnen und Probanden in der jeweiligen Kategorie abgelesen werden kann (R. J. Adams et al., 2012). R. J. Adams et al. (2012) empfehlen zudem, weitere Itemstatistiken wie Itemfit und punktbiseriale Korrelationen zwischen der jeweiligen Kategorie und der Personenfähigkeit zu betrachten, bevor Maßnahmen wie z. B. das Zusammenfassen von Kategorien ergriffen werden.

### Schätzung der Item- und Personenparameter

Um die Item- und Personenparameter schätzen zu können, wurden die Antworten je nach Aufgabe und erreichter Kategorie mit null, ein, zwei oder drei Punkten bewertet. Falsche Multiple-Choice-Aufgaben wurden mit null Punkten, richtige Multiple-Choice-Aufgaben mit einem Punkt bewertet. Die Schwellenparameter und die Itemschwierigkeiten wurden in *Logits* angegeben. Ein Logit-Wert von 0 stellt den Durchschnitt dar. Schwellenparameter und Itemschwierigkeiten  $< 0$  Logits sind als eher leicht und Itemschwierigkeiten  $> 0$  Logits als eher schwierig einzustufen (Bond & Fox, 2007; M. Wu & Adams, 2007). Itemschwierigkeiten oder Schwellenparameter  $> 2$  Logits werden als schwierig angesehen (Schöps & Saß, 2012). Die Personenparameter wurden über die *Weighted-Likelihood-Estimation* (WLE) ermittelt (D. H. Rost et al., 2004). Die Werte sind in Analogie zu den Itemschwierigkeiten und Schwellenparametern zu interpretieren: Je positiver der Personenparameter  $\theta$ , desto fähiger die Person, und je negativer der Personenparameter  $\theta$ , desto weniger fähig ist die Person. Um zu prüfen, wie gut die Items zur Stichprobe passten, wurde eine Wright Map erstellt, bei der die *Thurstone Thresholds*, d. h. die Punkte auf dem latenten Kontinuum, an dem die Lösungswahrscheinlichkeit der jeweiligen Kategorie bei 50.0% liegt, gegen die Fähigkeit der Teilnehmenden der Studie aufgetragen wurden.

#### 6.4.4 Testqualität

##### Itemfit

Zur Bewertung der Qualität der Items wurden aus den Ergebnissen der Skalierung der Mean-Squared-Infit sowie der Infit-*t*-Wert herangezogen. Dabei finden sich in Abhängigkeit verschiedener Faktoren, wie z. B. dem Testtyp oder der Stichprobengröße, unterschiedliche Aussagen zu akzeptablen Grenzen der Fit-Werte. Nach Wilson (2005) können Werte für den Mean-Squared-Infit zwischen 0.75 und 1.33 als akzeptabel angesehen werden. Für den Infit-*t*-Wert geben Bond und Fox (2007) Werte zwischen -2 und 2 als akzeptabel an. Nach Wilson (2005) werden jedoch vor allem solche Aufgaben als problematisch angesehen, die sowohl für den Mean-Squared-Infit als auch für den Infit-*t*-Wert Werte außerhalb der o. a. Grenzen aufweisen.

##### Differential Item Functioning

Zur Prüfung der Testfairness wurden die Items auf *Differential Item Functioning* (DIF) geprüft. Die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten sollten im Idealfall für alle Teilnehmenden der Studie ein und dasselbe Konstrukt messen. Sofern dies nicht gegeben wäre, wäre ein Vergleich der Merkmalsausprägungen verzerrt und die entsprechende Aufgabe bzw. der Test insgesamt nicht valide (Wilbert & Linnemann, 2011). DIF liegt vor,

wenn zwei Personen mit gleicher Personenfähigkeit für ein Item unterschiedliche Lösungswahrscheinlichkeiten aufgrund der Zugehörigkeit zu verschiedenen Subgruppen besitzen (z. B. Bond & Fox, 2007; P. W. Holland & Wainer, 1993). In diesem Fall könnte die Lösungswahrscheinlichkeit einer Aufgabe nicht allein durch die Personenfähigkeit und die Itemschwierigkeit erklärt werden (R. Adams & Wu, 2002). Bei der Prüfung auf DIF wird zunächst ein Modell, das nur den Haupteffekt modelliert, gegen ein Modell getestet, das einen Interaktionseffekt zwischen den entsprechenden Gruppen modelliert (Pohl & Carstensen, 2012). Sofern das Modell, das einen Interaktionseffekt zwischen den entsprechenden Gruppen modelliert, signifikant besser zu den Daten passt, sollten die Items im Einzelnen weiter untersucht werden. Die Modellpassung wird dabei anhand des Akaike Information Criterion (AIC) und des Bayesian Information Criterion (BIC) geprüft. Des Weiteren wird ein Likelihood-Ratio-Test zur Prüfung der Signifikanz durchgeführt. Zur weiteren Analyse möglicher von DIF betroffener Items wird für jedes Item die Differenz der für die unterschiedlichen Gruppen geschätzten Itemschwierigkeiten betrachtet und je nach ihrer Größe in drei Gruppen klassifiziert (nach Paek, 2002, zit. n. Wilson, 2005, S. 167):

- A Nicht nennenswert von DIF betroffene Aufgaben: Die Differenz in den geschätzten Itemschwierigkeiten ist kleiner als 0.426 Logits.
- B Mäßig von DIF betroffene Aufgaben: Die Differenz in den geschätzten Itemschwierigkeiten liegt zwischen 0.426 und 0.638 Logits.
- C Stark von DIF betroffene Aufgaben: Die Differenz in den geschätzten Itemschwierigkeiten ist größer als 0.638 Logits.

Unter Berücksichtigung des Testziels, eventuell vorhandene Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten zwischen Teilnehmenden verschiedener naturwissenschaftlicher Schülerwettbewerbe sowie zu nichtteilnehmenden Schülerinnen und Schülern zu erfassen, wurden die Aufgaben hinsichtlich des Merkmals Wettbewerbserfahrung auf DIF geprüft. Hier liegt die theoretische Annahme zugrunde, dass gerade die Teilnehmenden des IJSO-Bundesfinales durch ihre Erfahrung in einem Aufgabenwettbewerb gegenüber anderen Schülerinnen und Schülern in der Bearbeitung der Testaufgaben möglicherweise bevorteilt wurden. Darüber hinaus wurden die Aufgaben hinsichtlich des Geschlechts auf DIF geprüft, da es geschlechtsspezifische Interessensgebiete von Mädchen und Jungen gibt (z. B. Holstermann & Bögeholz, 2007; Su, Rounds & Armstrong, 2009), die sich je nach Kontext einer Testaufgabe auf die Bearbeitung der Aufgabe auswirken und DIF erzeugen können (Walpuski & Ropohl, 2014). Eine Prüfung auf DIF hinsichtlich verschiedener Schülerwettbewerbe erfolgte aufgrund der geringen Stichprobengröße der Subgruppen nicht.

Sofern substanzieller DIF vorliegt, sollte nach möglichen inhaltlichen Gründen für das Vorliegen von DIF gesucht und das entsprechende Item ggf. überarbeitet werden. Es ist jedoch zu beachten, dass in der Testpraxis bei großen Stichproben jedes Item DIF aufweisen kann, da eine entsprechend große Stichprobe schon relativ kleine Unterschiede zwischen Subgruppen aufdecken kann (M. Wu & Adams, 2007). Die Entscheidung, wie mit Items umgegangen wird, die DIF aufweisen, ist daher unter Beachtung mehrerer Faktoren zu treffen.

### Mehrdimensionalität

Die Pilotierung hatte gezeigt, dass die Reliabilität des gesamten Testinstruments im Vergleich zu der Reliabilität einzelner Subtests deutlich höher lag (siehe Abschnitt 5.4.2). Dies könnte einerseits an der höheren Anzahl an Aufgaben im Gesamttest im Vergleich zu den Subtests gelegen haben, andererseits deutete dies darauf hin, dass naturwissenschaftliche Kenntnisse und Fähigkeiten als ein Konstrukt angesehen werden können und die eingesetzten Testaufgaben somit eine Fähigkeit messen. Im ersten Ansatz wurde daher ein eindimensionales PCM berechnet. Aufgrund der Anlage der Aufgaben und der Aufgabenkonstruktion nach den Fächern Biologie, Chemie und Physik (jeweils zehn Aufgaben) sowie den Kompetenzbereichen Erkenntnisgewinnung (zwölf Aufgaben) und Fachwissen (15 Aufgaben) wurden in Ergänzung zu den eindimensionalen Modellen zur weiteren Prüfung sowohl zweidimensionale als auch dreidimensionale Modelle geschätzt. Dies erfolgte unter der Annahme, dass die Aufgaben möglicherweise verschiedene Fähigkeiten erfassten. Der Vergleich der Modelle untereinander erfolgte anhand des AIC, des BIC, der Reliabilität sowie anhand der Korrelationen der Dimensionen untereinander. Hohe Korrelationen (über .95) der Dimensionen untereinander sprechen dabei eher gegen ein mehrdimensionales Modell (Pohl & Carstensen, 2012), da sich die einzelnen Dimensionen durch hohe Korrelationen schwer voneinander abgrenzen lassen.

### Testziele

Zur Prüfung der Validität des Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten wurden Korrelationen nach Pearson bzw. bei dichotomen Variablen punktbiserial Korrelationen zwischen der Personenfähigkeit der Schülerinnen und Schüler und den im Fragebogen zum IKoN erhobenen Konstrukten (siehe Tabelle 6.1) berechnet. Die im Fragebogen zum IKoN erhobenen Konstrukte wurden ausgewählt, da in einer Vielzahl an Studien bereits gezeigt werden konnte, dass diese Konstrukte als Einflussgrößen (schulischer) Leistung angesehen werden können (vgl. Abschnitt 2.3.2, z. B. Helmke & Schrader, 2010; Köller et al., 2001; Krapp et al., 1993; Marsh & Martin, 2011; Marsh et al., 2005; Wirthwein et al., 2013). Des Weiteren wurden durch multiple lineare und binäre logistische Regressionen Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten sowie für eine Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb und für Erfolg in einem naturwissenschaftlichen Schülerwettbewerb ermittelt. Diese Analysen wurden durchgeführt um zu prüfen, ob das Testinstrument bekannte Prädiktoren für z. B. Erfolg in einem naturwissenschaftlichen Schülerwettbewerb, vgl. Urhahne et al. (2012), aufdecken konnte.

Im Hinblick auf das Testziel, eventuell vorhandene Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten zwischen Teilnehmenden (verschiedener) naturwissenschaftlicher Schülerwettbewerbe sowie zu nichtteilnehmenden Schülerinnen und Schülern zu erfassen, wurden die Daten zu den naturwissenschaftlichen Kenntnissen und Fähigkeiten varianzanalytisch ausgewertet. Hierzu wurde die Stichprobe in mehrere Subgruppen unterteilt, die im weiteren Verlauf miteinander verglichen wurden. Zum einen wurde die Stichprobe nach dem Geschlecht aufgeteilt, zum anderen in Gruppen nach Wettbewerbserfahrung: Es wurden drei Subgruppen gebildet, die Teilnehmenden des IJSO-Bundesfinales ( $n = 39$ ,  $M = 14.59$ ,  $SD = 0.64$ , Altersgruppe 13 bis 15, 33.3% Mädchen), Schülerinnen oder Schüler, die bereits mindestens einmal an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten (im Folgenden: Wettbewerbsteilnehmende,  $n = 70$ ,  $M = 15.10$ ,  $SD = 1.11$ ,

Altersgruppe 13 bis 18, 42.9% Mädchen) und Schülerinnen und Schüler, die noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten (im Folgenden: Nichtteilnehmende,  $n = 86$ ,  $M = 14.21$ ,  $SD = 0.80$ , Altersgruppe 13 bis 16, 54.7% Mädchen). Aufgrund der geringen Anzahl an Schülerinnen und Schülern ( $n = 30$ ), die am Landeswettbewerb Jugend forscht bzw. Schüler experimentieren teilgenommen und die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten bearbeitet hatten ( $n = 15$ ), wurde diese Gruppe in die Subgruppe der Schülerinnen oder Schüler, die bereits mindestens einmal an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten, integriert.

Alle Analysen wurden mit der Statistik-Software R durchgeführt. Die Skalierung der Daten über das PCM und zugehörige Analysen wurden mit dem R-Programmpaket *Test Analysis Modules* (TAM) von Kiefer, Robitzsch und Wu (2015), Version 1.9-0, durchgeführt.

## 6.5 Ergebnisse

Im Folgenden werden die Ergebnisse der Auswertungen der Daten aus Studie 1 vorgestellt. Die Interpretation und Diskussion der Ergebnisse folgt in Abschnitt 6.6.

### 6.5.1 Analyse fehlender Werte

#### Fehlende Werte pro Person

Insgesamt erreichten 62.4% der teilnehmenden Schülerinnen und Schüler nicht die letzte Aufgabe des Tests zu naturwissenschaftlichen Kenntnissen und Fähigkeiten. Die Anzahl ausgelassener Items pro Person sowie die Anzahl fehlender Werte insgesamt pro Person sind in Abbildung 6.1 dargestellt. Zu erkennen ist, dass lediglich 11.0% der Schülerinnen und Schüler kein einziges Item ausließen. Die meisten der Teilnehmenden (31.7%) ließen vier, fünf oder sechs Items von 30 aus. 6.4% ließen mehr als die Hälfte der Testitems aus. Was die Anzahl insgesamt fehlender Werte betrifft, ergab sich ein ähnliches Bild. 10.6% der Schülerinnen und Schüler ließen kein einziges Item unbearbeitet. 29.4% der Teilnehmenden ließen vier, fünf oder sechs Items unbearbeitet. 7.3% bearbeiteten weniger als die Hälfte der Testitems.

#### Fehlende Werte pro Item

Die Anzahl der insgesamt fehlenden Werte pro Aufgabe ist ebenso in Abbildung 6.1 dargestellt. Auf der Ebene der Items gab es drei Aufgaben, die von weniger als der Hälfte der Schülerinnen und Schüler bearbeitet wurden. Insgesamt fällt auf, dass die meisten Aufgaben einen hohen Anteil an fehlenden Werten aufwiesen.

### 6.5.2 Skalierung der Itemantworten

Da im Rahmen des Gesamtprojektes auf lange Sicht alle vier Aspekte des IKoN (siehe Abschnitt 2.1) zur Beschreibung von naturwissenschaftlich interessierten und talentierten Schülerinnen und Schülern herangezogen werden sollten, wurden die naturwissenschaftlichen Kenntnisse und Fähigkeiten nur für diejenigen, die alle Erhebungsinstrumente bearbeitet hatten, unter Verwendung des eindimensionalen PCM modelliert. Darüber hinaus war es für weitere Analysen hinsichtlich verschiedener Testgütekriterien von Bedeutung, Informationen zu allen Aspekten sowie im Besonderen demographische Informationen aus dem Fragebogen vorliegen

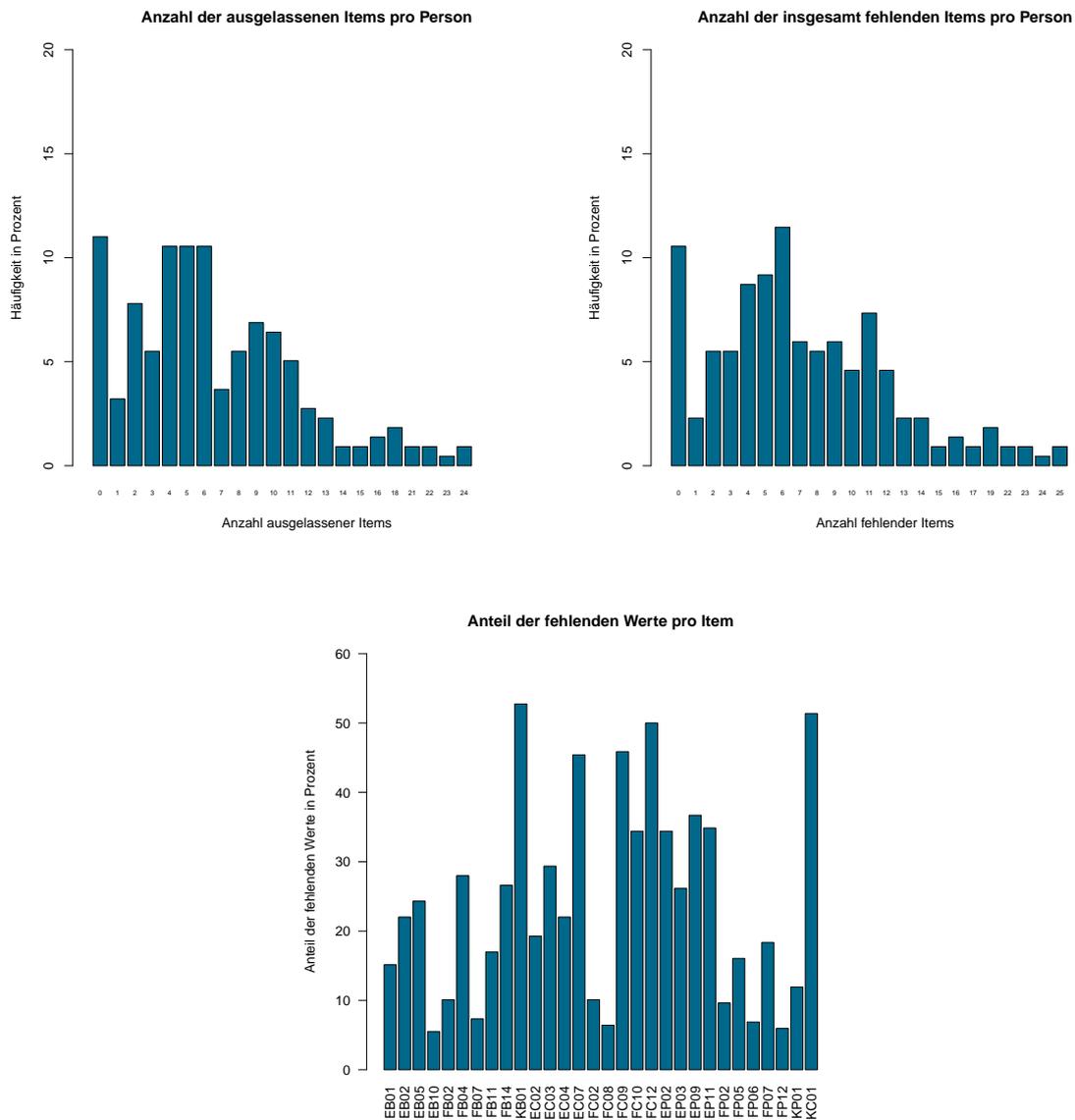


Abbildung 6.1: **Ergebnisse der Analyse fehlender Werte für Studie 1** Abgebildet sind die Anzahl der in Studie 1 von den teilnehmenden Schülerinnen und Schülern ausgelassenen Items pro Person, die Anzahl der insgesamt fehlenden Items pro Person sowie der Anteil der insgesamt fehlenden Werte pro Aufgabe.

zu haben. Ausgeschlossen wurden somit 51 von insgesamt 246 Teilnehmenden (wobei nur 218 Teilnehmende die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten bearbeitet hatten). Die reduzierte Stichprobe bestand aus  $n = 195$  Schülerinnen und Schülern mit einem durchschnittlichen Alter von  $M = 14.61$ ,  $SD = 0.98$ , Altersgruppe 13-18, 46.2% Mädchen.

Nach mehreren Überarbeitungsschritten, in denen eine Aufgabe aufgrund ihrer Fit-Werte aus dem Test gänzlich entfernt und für neun Aufgaben Kategorien zusammengefasst wurden, ergab sich eine zufriedenstellende EAP/PV Reliabilität von .86 und eine WLE-Reliabilität von .84.

### Schätzung der Item- und Personenparameter

**Itemparameter und Itemschwierigkeit** Die Itemparameter wurden als Schwellenparameter über ein eindimensionales PCM geschätzt. Bei 27 von 30 Aufgaben lagen geordnete Schwellenparameter vor. Für die Aufgaben EC03, EP09 und FP12, deren Schwellenparameter nicht geordnet waren, wurden nach R. J. Adams et al. (2012) neben dem Itemfit (siehe Abschnitt 6.5.3) die punktbiserialen Korrelationen zwischen der jeweiligen Kategorie und der Personenfähigkeit und die Mittelwerte der Personenfähigkeiten aller Probandinnen und Probanden in den jeweiligen Kategorien analysiert. Sowohl die punktbiserialen Korrelationen als auch die Mittelwerte der Personenfähigkeiten aller Probandinnen und Probanden in den jeweiligen Kategorien nahmen mit steigender Kategorie zu, weshalb für diese Aufgaben trotz ungeordneter Schwellenparameter von geordneten Kategorien ausgegangen werden konnte. Die Itemschwierigkeiten variierten zwischen -2.112 für das leichteste Item (FP06) und 1.884 (KC01i) für das schwierigste Item. Die Itemparameter, die Itemschwierigkeiten sowie die weiterführenden Itemstatistiken sind in Anhang F einzusehen.

**Personenparameter** Die Schätzung der Personenparameter ( $\theta$ ) erfolgte über die WLE. Sie variierten zwischen Werten von -2.700 Logits bis 2.360 Logits für die fähigste Person.

**Wright Map** Die Passung der Items auf die Stichprobe ist in Abbildung 6.2 dargestellt. Es ist zu erkennen, dass die einzelnen Kategorien das Fähigkeitsspektrum der Schülerinnen und Schüler relativ gut abdecken konnten, wobei die Vielzahl der Aufgaben das obere Fähigkeitsspektrum umfasste. Im Bereich unter -1 Logit waren lediglich wenige Kategorien angesiedelt.

### 6.5.3 Testqualität

#### Itemfit

Unter Verwendung des eindimensionalen PCM wurden die Aufgaben hinsichtlich ihrer Fit-Statistiken geprüft. Nach Wilson (2005) können Werte für den Mean-Squared-Infit zwischen 0.75 und 1.33 als akzeptabel angesehen werden. Alle Aufgaben wiesen einen Mean-Squared-Infit in diesem Wertebereich auf, sie lagen zwischen 0.77 und 1.25. Für den Infit- $t$ -Wert geben Bond und Fox (2007) Werte zwischen -2 und 2 als akzeptabel an, 27 der 30 Aufgaben lagen in diesem Bereich. Nach Wilson (2005) werden jedoch vor allem solche Aufgaben als problematisch angesehen, die sowohl für den Mean-Squared-Infit als auch für den Infit- $t$ -Wert Werte außerhalb der o. a. Grenzen aufweisen. Dies traf auf keine der 30 Aufgaben zu. Die Itemfit-Statistiken sind in Anhang F einzusehen.

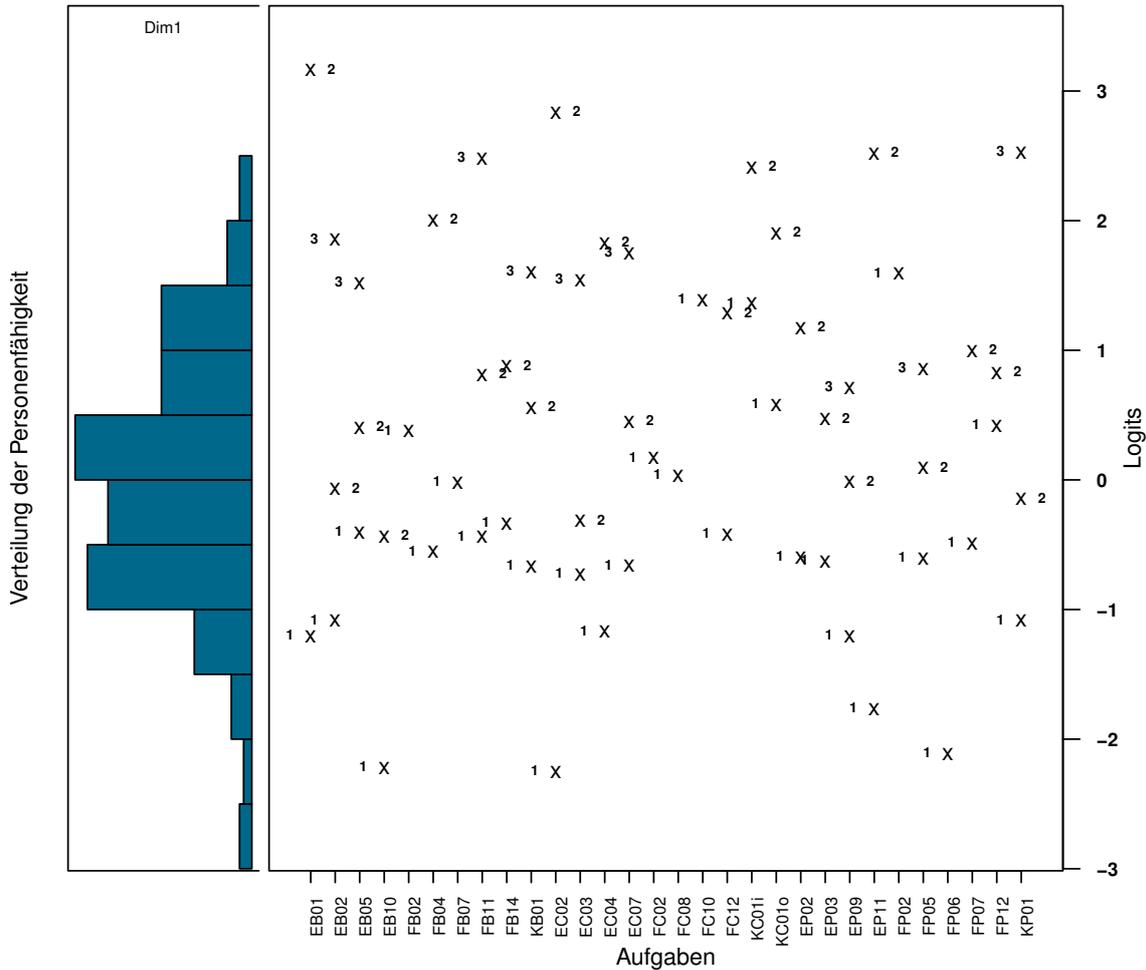


Abbildung 6.2: **Wright Map** Die Zahlen 1, 2, 3 bezeichnen die Kategorien der einzelnen Aufgaben.

### Differential Item Functioning

Es wurden DIF-Analysen für das Merkmal Geschlecht sowie für das Merkmal Wettbewerbs- erfahrung (Teilnehmende des IJSO-Bundesfinales versus nichtteilnehmende Schülerinnen und Schüler) durchgeführt. Die Ergebnisse sind in Tabelle 6.2 dargestellt. In beiden Analysen gab es keinen signifikanten Unterschied in der Passung des einen oder des anderen Modells auf die Daten. Betrachtet man jeweils die Werte für AIC und BIC, fällt auf, dass in beiden Fällen die Werte günstiger für das jeweilige Modell ohne DIF ausfallen.

### Mehrdimensionalität

Zur Prüfung der Dimensionalität des eingesetzten Tests zu naturwissenschaftlichen Kennt- nissen und Fähigkeiten wurde das eindimensionale PCM mit einem zweidimensionalen PCM verglichen, wobei hier die Kompetenzbereiche Erkenntnisgewinnung und Fachwissen in die Analyse eingingen. Zudem wurde das eindimensionale PCM mit einem dreidimensionalen PCM

Tabelle 6.2: Vergleich zwischen Modellen mit und ohne DIF

<i>Merkmal</i>	<i>Modell</i>	<i>Devianz</i>	<i>Anzahl Parameter</i>	<i>AIC</i>	<i>BIC</i>	<i>p<sub>LRT</sub></i>
Geschlecht	Haupteffekt	7537.41	64	7665.41	7874.88	n. s.
	DIF	7497.57	93	7683.57	7987.96	
Wettbewerbs- erfahrung	Haupteffekt	7583.10	64	7711.10	7920.57	n. s.
	DIF	7541.41	93	7727.41	8031.80	

*Anmerkungen.*  $p_{LRT}$  =  $p$ -Wert für den Likelihood-Ratio-Test. n. s. = nicht signifikant.

Tabelle 6.3: Vergleich des eindimensionalen PCM (1-dim) mit einem zwei- (2-dim) und einem dreidimensionalen (3-dim) Modell

<i>Modell</i>	<i>Devianz</i>	<i>Anzahl der Parameter</i>	<i>AIC</i>	<i>BIC</i>	<i>p<sub>LRT</sub></i>
1-dim	7600.41	63	7726.41	7932.61	n. s.
2-dim	7604.77	65	7734.77	7947.52	
3-dim	7595.35	68	7731.35	7953.91	n. s.

*Anmerkungen.*  $p_{LRT}$  =  $p$ -Wert für den Likelihood-Ratio-Test. n. s. = nicht signifikant.

verglichen. In das dreidimensionale Modell gingen die Fächer Biologie, Chemie und Physik als Dimensionen in die Analyse ein. Die Modellvergleiche ergaben in beiden Fällen keine signifikant bessere Passung des einen oder des anderen Modells, allerdings deuteten die jeweils niedrigeren Werte für AIC und BIC darauf hin, dass das eindimensionale PCM eine bessere Passung auf die Daten aufwies. Tabelle 6.3 zeigt die einzelnen Parameter des Modellvergleichs.

Die EAP/PV-Reliabilitäten waren in den einzelnen Dimensionen der mehrdimensionalen Modelle zufriedenstellend, während die WLE-Reliabilitäten sich als nicht zufriedenstellend erwiesen:

- **Erkenntnisgewinnung** EAP/PV-Reliabilität = .82 und WLE-Reliabilität = .56
- **Fachwissen** EAP/PV-Reliabilität = .85 und WLE-Reliabilität = .76
- **Biologie** EAP/PV-Reliabilität = .84 und WLE-Reliabilität = .64
- **Chemie** EAP/PV-Reliabilität = .83 und WLE-Reliabilität = .43
- **Physik** EAP/PV-Reliabilität = .84 und WLE-Reliabilität = .62

In Ergänzung der Reliabilitäten wurden zudem die Korrelationen der Dimensionen untereinander bestimmt. Die Kompetenzbereiche Erkenntnisgewinnung und Fachwissen korrelierten dabei zu  $r = .93$  miteinander. Die Korrelationen der einzelnen Fächer untereinander lagen alle bei  $r = .95$ , was die bessere Passung des eindimensionalen Modells gegenüber den mehrdimensionalen Modellen erklärt.

## Testziele

**Korrelationen** Korrelationen nach Pearson bzw. bei dichotomen Variablen punkt-biserial Korrelationen zwischen der Personenfähigkeit, d. h. den naturwissenschaftlichen Kenntnissen

Tabelle 6.4: Korrelationen zwischen der Personenfähigkeit und erhobenen Begleitvariablen

<i>Prädiktor</i>	<i>M</i>	<i>SD</i>	<i>r</i> bzw. <i>r<sub>pb</sub></i>	<i>p</i>
Nonverbale kognitive Fähigkeiten <sup>a</sup>	21.37	4.12	0.33 ***	< .001
Schulnoten in Biologie, Chemie und Physik <sup>b</sup>	2.13	0.83	-0.53 ***	< .001
Akademisches Selbstkonzept	2.99	0.55	0.44 ***	< .001
Naturwissenschaftliches Selbstkonzept	3.02	0.70	0.59 ***	< .001
Geschlecht (46.2% Mädchen) <sup>c</sup>			-0.27 ***	< .001
Interesse (Schule)	2.36	0.60	0.31 ***	< .001
Interesse (Fördermaßnahme)	2.29	0.69	0.34 ***	< .001
Interesse (Beruf)	2.30	0.66	0.31 ***	< .001
Selbstwirksamkeitserwartung für die spätere Ausübung von Tätigkeiten eines Wissenschaft- lers	2.60	0.52	0.34 ***	< .001
Annäherungsleistungsziele	2.17	0.71	-0.01	.844
Lernziele	2.99	0.65	0.32 ***	< .001
Vermeidungslernziele	1.68	0.62	-0.21 **	.003
Arbeitsvermeidung	2.02	0.73	-0.20 **	.005
Wettbewerbsteilnahme (55.9% ja)			0.47 ***	< .001
Anzahl Wettbewerbsteilnahmen	2.15	2.43	0.57 ***	< .001
Alter	14.61	0.98	0.30 ***	< .001

*Anmerkungen.* <sup>a</sup> Für die Berechnung der Korrelation wurden Teilnehmende des IJSO-Bundesfinales 2013 ausgeschlossen. <sup>b</sup> Gemittelt in einer Variablen. Die negative Korrelation kommt hier dadurch zustande, dass geringe Zahlenwerte gute Leistungen beschreiben und umgekehrt. <sup>c</sup> Jungen wurden mit 0 kodiert, Mädchen mit 1. \*\*  $p < .01$ , \*\*\*  $p < .001$ .

und Fähigkeiten, und den einzelnen im Fragebogen zum IKoN erhobenen Begleitvariablen sowie zugehörige Mittelwerte und Standardabweichungen sind in Tabelle 6.4 aufgeführt. Keine signifikante Korrelation fand sich zu den Annäherungsleistungszielen. Bei der Korrelation zwischen der nonverbalen Skala des KFT ist anzumerken, dass die Gruppe der Teilnehmenden des IJSO-Bundesfinales nicht in die Analyse einging. Dies lag darin begründet, dass bereits bei der Testung auffiel, dass die Schülerinnen und Schüler gegen Ende des KFT unkonzentriert und unmotiviert erschienen, viele der Teilnehmenden beendeten schon nach der Hälfte der Zeit die Bearbeitung. Dies kann einerseits auf die gesamte Testsituation zurückgeführt werden, da die Teilnehmenden bereits die Tage zuvor theoretische und praktische Klausuren im Rahmen des Wettbewerbs bearbeitet hatten. Andererseits waren vielen der Schülerinnen und Schüler nach eigenen Aussagen Aufgaben, die kognitive Teilfähigkeiten testen, gut bekannt, so dass sich hier eventuell eine Testmüdigkeit eingestellt hatte. Einige der Ergebnisse im KFT stützen durchaus diese Vermutungen, da viele der Werte, die die Teilnehmenden in der nonverbalen Skala erreichten, gerade im Vergleich zu ihrer jeweiligen Personenfähigkeit als höchst unplausibel anzusehen waren.

**Multiple lineare und binäre logistische Regressionen** Die zur Personenfähigkeit signifikant korrelierenden Variablen bzw. Konstrukte wurden zur Vorhersage hoher naturwissenschaftlicher Kenntnisse und Fähigkeiten (bzw. einer hohen Personenfähigkeit) in eine multiple

Tabelle 6.5: **Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten**  $R^2 = .528$ ,  $F(14, 120) = 9.56$ ,  $p < .001$ .

<i>Prädiktor</i>	<i>B</i>	<i>SE</i>	$\beta$	<i>p</i>
Nonverbale kognitive Fähigkeiten	0.06	0.01	0.32 ***	< .001
Schulnoten in Biologie, Chemie und Physik	-0.19	0.08	-0.18 *	.026
Akademisches Selbstkonzept	0.07	0.13	0.04	.602
Naturwissenschaftliches Selbstkonzept	0.18	0.11	0.14	.107
Geschlecht	-0.28	0.11	-0.18 **	.009
Interesse (Schule)	0.24	0.18	0.15	.173
Interesse (Fördermaßnahme)	0.06	0.16	0.05	.729
Interesse (Beruf)	-0.28	0.17	-0.23	.094
Selbstwirksamkeitserwartung für die spätere Ausübung von Tätigkeiten eines Wissenschaftlers	0.00	0.15	0.00	.992
Lernziele	0.04	0.10	-0.03	.663
Vermeidungslernziele	-0.14	0.10	-0.11	.155
Arbeitsvermeidung	0.13	0.09	0.11	.159
Anzahl Wettbewerbsteilnahmen	0.08	0.03	0.21 *	.011
Alter	0.22	0.06	0.29 ***	< .001

*Anmerkungen.* \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

lineare Regression aufgenommen. Alle Prädiktoren wurden in einem ersten Schritt für die reduzierte Stichprobe ( $n = 195$ ) simultan getestet. Die Ergebnisse sind in Tabelle 6.5 dargestellt. Es zeigten sich fünf signifikante Prädiktoren: Die nonverbalen kognitiven Fähigkeiten, das Alter der Schülerinnen und Schüler, das Geschlecht, die Anzahl an Wettbewerbsteilnahmen und die Schulnoten in Biologie, Chemie und Physik.

Zur detaillierten Analyse des Einflusses des Geschlechts auf hohe naturwissenschaftliche Kenntnisse und Fähigkeiten wurden nach Geschlecht getrennte multiple lineare Regressionen durchgeführt, in denen zunächst der für die reduzierte Stichprobe ( $n = 195$ ) stärkste Prädiktor in die Regression aufgenommen und dann schrittweise um weitere signifikant korrelierende Variablen bzw. Konstrukte ergänzt wurde. Für die Mädchen ergaben sich als signifikante Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten die nonverbalen kognitiven Fähigkeiten, die Schulnoten in Biologie, Chemie und Physik (gemittelt in einer Variablen) sowie das Alter. Für die Jungen hingegen zeigte sich als stärkster Prädiktor das fachspezifische Selbstkonzept neben dem Alter und den nonverbalen kognitiven Fähigkeiten. Die Ergebnisse der nach Geschlecht getrennten multiplen linearen Regressionen sind in Tabelle 6.6 und Tabelle 6.7 dargestellt.

Um mögliche Unterschiede in den nach Wettbewerbserfahrung eingeteilten Subgruppen zu identifizieren, wurden nach Wettbewerbserfahrung getrennte multiple lineare Regressionen durchgeführt, in denen zunächst der für die reduzierte Stichprobe ( $n = 195$ ) stärkste Prädiktor in die Regression aufgenommen und dann schrittweise um weitere signifikant korrelierende Variablen bzw. Konstrukte ergänzt wurde. Für die Gruppe der Teilnehmenden

des IJSO-Bundesfinales und der Schülerinnen und Schüler, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten, wurde zusätzlich zu den in Tabelle 6.5 aufgeführten möglichen Prädiktoren die Anzahl der Wettbewerbsteilnahmen in die Regression eingeschlossen, da z. B. von Urhahne et al. (2012) gezeigt werden konnte, dass der stärkste Prädiktor für Erfolg in der vierten Runde der IChO die wiederholte Teilnahme am Wettbewerb war. Für Schülerinnen und Schüler, die noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten, konnte dieser Prädiktor nicht untersucht werden, da die Frage diesen Schülerinnen und Schüler nicht vorgelegt worden war.

Für die Teilnehmenden des IJSO-Bundesfinales mit  $n = 39$  ergaben sich das Alter und das Geschlecht als signifikante Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten. (Die nonverbalen kognitiven Fähigkeiten wurden aus o. a. Gründen nicht in die Regression für die Teilnehmenden des IJSO-Bundesfinales aufgenommen.) Gerade die Ergebnisse der Teilnehmenden des IJSO-Bundesfinales sind jedoch nur unter Vorbehalt zu betrachten, da für Regressionsanalysen eine Stichprobengröße von mindestens  $N = 50$  vorliegen sollte (van Voorhis & Morgan, 2007). Bei den Schülerinnen und Schülern, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten (Wettbewerbsteilnehmende), konnten das Alter, die nonverbalen kognitiven Fähigkeiten sowie das naturwissenschaftliche Selbstkonzept als signifikante Prädiktoren identifiziert werden. Für Schülerinnen und Schüler, die noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten (Nichtteilnehmende), ergaben sich ebenso drei signifikante Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten: Die nonverbalen kognitiven Fähigkeiten, die Schulnoten in Biologie, Chemie und Physik (gemittelt in einer Variablen) sowie das Geschlecht. Die Ergebnisse der multiplen linearen Regressionen sind in Tabelle 6.8, Tabelle 6.9 und Tabelle 6.10 aufgeführt.

Um mögliche Prädiktoren für eine Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb und Erfolg in einem naturwissenschaftlichen Schülerwettbewerb, hier die Teilnahme am Bundesfinale der IJSO 2013, zu ermitteln, wurden zunächst punktbiseriale Korrelationen zu im Fragebogen zum IKoN erhobenen Konstrukten und für Korrelationen zwischen dichotomen Variablen der Phi-Koeffizient ermittelt. Die Ergebnisse sind in Tabelle 6.11 aufgeführt.

Die zu einer Wettbewerbsteilnahme bzw. zum Wettbewerbserfolg signifikant korrelierenden Variablen bzw. Konstrukte wurden zur Vorhersage einer Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb schrittweise in jeweils eine binäre logistische Regression aufgenommen. Für die Regression zur Ermittlung von Prädiktoren für die Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb wurde der mögliche Prädiktor „Anzahl der Wettbewerbsteilnahmen“ ausgeschlossen, da alle Schülerinnen und Schüler, die dieses Item beantwortet hatten, auch an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten. Für die Regression zur Ermittlung von Prädiktoren für die Teilnahme am IJSO-Bundesfinale 2013 wurden nur Schülerinnen und Schüler unter 16 Jahren berücksichtigt, da das maximale Alter für eine Teilnahme an der IJSO 15 Jahre beträgt. Aus o. a. Gründen gingen die nonverbalen kognitiven Fähigkeiten nicht in die Regression zur Ermittlung von Prädiktoren für die Teilnahme am IJSO-Bundesfinale 2013 ein.

Die Ergebnisse der binären logistischen Regressionen sind in Tabelle 6.12 und Tabelle 6.13

Tabelle 6.6: **Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für Mädchen ( $n = 90$ )  $R^2 = .336$ ,  $F(3, 71) = 11.99$ ,  $p < .001$ .**

<i>Prädiktor</i>	<i>B</i>	<i>SE</i>	$\beta$	<i>p</i>
Nonverbale kognitive Fähigkeiten	0.08	0.02	0.43 ***	< .001
Schulnoten in Biologie, Chemie und Physik	-0.26	0.08	-0.30 **	.003
Alter	0.25	0.09	0.34 **	.002

*Anmerkungen.* \*\*  $p < .01$ , \*\*\*  $p < .001$

Tabelle 6.7: **Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für Jungen ( $n = 105$ )  $R^2 = .581$ ,  $F(3, 75) = 34.68$ ,  $p < .001$ .**

<i>Prädiktor</i>	<i>B</i>	<i>SE</i>	$\beta$	<i>p</i>
Naturwissenschaftliches Selbstkonzept	0.75	0.11	0.51 ***	< .001
Alter	0.31	0.07	0.35 ***	< .001
Nonverbale kognitive Fähigkeiten	0.08	0.02	0.36 ***	< .001

*Anmerkungen.* \*\*\*  $p < .001$

Tabelle 6.8: **Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für die Teilnehmenden des IJSO-Bundesfinales ( $n = 39$ )  $R^2 = .335$ ,  $F(2, 36) = 9.07$ ,  $p < .001$ .**

<i>Prädiktor</i>	<i>B</i>	<i>SE</i>	$\beta$	<i>p</i>
Alter	0.45	0.12	0.50 ***	< .001
Geschlecht	-0.35	0.16	-0.30 *	.034

*Anmerkungen.* \*  $p < .05$ , \*\*\*  $p < .001$

Tabelle 6.9: **Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für Schülerinnen und Schüler, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten ( $n = 70$ )  $R^2 = .640$ ,  $F(3, 66) = 39.10$ ,  $p < .001$ .**

<i>Prädiktor</i>	<i>B</i>	<i>SE</i>	$\beta$	<i>p</i>
Naturwissenschaftliches Selbstkonzept	0.79	0.11	0.54 ***	< .001
Alter	0.37	0.07	0.42 ***	< .001
Nonverbale kognitive Fähigkeiten	0.12	0.02	0.50 ***	< .001

*Anmerkungen.* \*\*\*  $p < .001$

Tabelle 6.10: **Multiple lineare Regression zur Ermittlung von Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten für Schülerinnen und Schüler, die noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten ( $n = 86$ )  $R^2 = .240$ ,  $F(3, 77) = 8.10$ ,  $p < .001$ .**

<i>Prädiktor</i>	<i>B</i>	<i>SE</i>	$\beta$	<i>p</i>
Nonverbale kognitive Fähigkeiten	0.06	0.02	0.35 **	.001
Schulnoten in Biologie, Chemie und Physik	-0.17	0.09	-0.20 *	.048
Geschlecht	-0.33	0.13	-0.27 *	.011

*Anmerkungen.* \*  $p < .05$ , \*\*  $p < .01$

Tabelle 6.11: Korrelationen zwischen der Teilnahme an einem mathematisch-naturwissenschaftlichen Wettbewerb (Wett-TN) bzw. am Bundesfinale der IJSO 2013 (Bufi-TN) und erhobenen Begleitvariablen

Prädiktor	Variable	<i>M</i>	<i>SD</i>	$r_{pb}$ bzw. $\phi^2$	<i>p</i>
Personenfähigkeit	Wett-TN			0.47 ***	< .001
	Bufi-TN	-0.03	0.81	0.49 ***	< .001
Nonverbale kognitive Fähigkeiten <sup>a</sup> Schulnoten in Biologie, Chemie und Physik <sup>b</sup>	Wett-TN	21.40	3.98	0.04	.633
	Bufi-TN	2.13	0.83	-0.51 ***	< .001
Akademisches Selbstkonzept	Wett-TN			0.35 ***	< .001
	Bufi-TN	2.99	0.55	0.47 ***	< .001
Naturwissenschaftliches Selbstkonzept	Wett-TN			0.48 ***	< .001
	Bufi-TN	3.02	0.70	0.52 ***	< .001
Geschlecht (46.2% Mädchen) <sup>c</sup>	Wett-TN			-0.16 *	.028
	Bufi-TN			-0.13	.073
Interesse (Schule)	Wett-TN			0.30 ***	< .001
	Bufi-TN	2.35	0.61	0.23 **	.001
Interesse (Fördermaßnahme)	Wett-TN			0.41 ***	< .001
	Bufi-TN	2.29	0.69	0.38 ***	< .001
Interesse (Beruf)	Wett-TN			0.39 ***	< .001
	Bufi-TN	2.30	0.66	0.30 ***	< .001
Selbstwirksamkeitserwartung für die spätere Ausübung von Tätigkeiten eines Wissenschaftlers	Wett-TN			0.35 ***	< .001
	Bufi-TN	2.60	0.52	0.30 ***	< .001
Annäherungsleistungsziele	Wett-TN			-0.02	.829
	Bufi-TN	2.17	0.71	-0.05	.480
Lernziele	Wett-TN			0.32 ***	< .001
	Bufi-TN	2.99	0.65	0.29 ***	< .001
Vermeidungslernziele	Wett-TN			-0.14 *	.045
	Bufi-TN	1.68	0.62	-0.16 *	.025
Arbeitsvermeidung	Wett-TN			-0.23 **	.001
	Bufi-TN	2.02	0.73	-0.18 *	.012
Alter	Wett-TN			0.37 ***	< .001
	Bufi-TN	14.61	0.98	0.49 ***	< .001
Anzahl Wettbewerbsteilnahmen <sup>d</sup>	Wett-TN			0.31 **	.002
	Bufi-TN	3.94	1.91	0.31 **	.002

*Anmerkungen.* <sup>a</sup> Für die Berechnung der Korrelation wurden Teilnehmende des IJSO-Bundesfinales 2013 ausgeschlossen, weshalb die Korrelation nur zur Wettbewerbsteilnahme ermittelt werden konnte. <sup>b</sup> Gemittelt in einer Variablen. Die negative Korrelation kommt hier dadurch zustande, dass geringe Zahlenwerte gute Leistungen beschreiben und umgekehrt. <sup>c</sup> Jungen wurden mit 0 kodiert, Mädchen mit 1. <sup>d</sup> Die Korrelation wurde nur zur Teilnahme am IJSO-Bundesfinale ermittelt, da alle Schülerinnen und Schüler, die dieses Item beantwortet hatten, auch an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Tabelle 6.12: **Binäre logistische Regression zur Ermittlung von Prädiktoren für die Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb**  $R^2 = .350$  (Hosmer-Lemeshow),  $.378$  (Cox-Snell),  $.509$  (Nagelkerke).  $\chi^2(3) = 86.36$ ,  $p < .001$ .

<i>Prädiktor</i>	<i>B</i>	<i>SE</i>	<i>Odds ratio</i>
Schulnoten in Biologie, Chemie und Physik	-1.47 ***	0.31	0.23
Alter	1.09 ***	0.24	2.96
Interesse (Fördermaßnahme)	1.07 **	0.34	2.92

*Anmerkungen.* \*\*  $p < .01$ , \*\*\*  $p < .001$

Tabelle 6.13: **Binäre logistische Regression zur Ermittlung von Prädiktoren für Wettbewerbserfolg (Teilnahme am Bundesfinale der IJSO 2013)**  $R^2 = .723$  (Hosmer-Lemeshow),  $.540$  (Cox-Snell),  $.820$  (Nagelkerke).  $\chi^2(3) = 112.47$ ,  $p < .001$ .

<i>Prädiktor</i>	<i>B</i>	<i>SE</i>	<i>Odds ratio</i>
Naturwissenschaftliche Kenntnisse und Fähigkeiten (Personenfähigkeit)	2.52 ***	0.76	12.43
Schulnoten in Biologie, Chemie und Physik	-2.79 **	1.03	0.06
Anzahl der Wettbewerbsteilnahmen	0.59 **	0.21	1.81

*Anmerkungen.* \*\*  $p < .01$ , \*\*\*  $p < .001$

dargestellt. Es zeigten sich für die Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb drei signifikante Prädiktoren, die Schulnoten in Biologie, Chemie und Physik (gemittelt in einer Variablen), das Alter der Schülerinnen und Schüler sowie das Interesse an Tätigkeiten, die in naturwissenschaftlichen Fördermaßnahmen ausgeübt werden. Für Wettbewerbserfolg, d. h. die Teilnahme am Bundesfinale der IJSO 2013, fanden sich die Personenfähigkeit, d. h. die naturwissenschaftlichen Kenntnissen und Fähigkeiten, die Schulnoten in Biologie, Chemie und Physik (gemittelt in einer Variablen) sowie die Anzahl der Wettbewerbsteilnahmen als signifikante Prädiktoren.

**Varianzanalysen** In der varianzanalytischen Auswertung der Daten zeigte sich ein signifikanter Haupteffekt der Wettbewerbserfahrung auf die naturwissenschaftlichen Kenntnisse und Fähigkeiten bzw. die Personenfähigkeit der Schülerinnen und Schüler,  $F(1, 191) = 60.71$ ,  $p < .001$ . So zeigten die Teilnehmenden des IJSO-Bundesfinales ( $M = 0.94$ ,  $SD = 0.56$ ) signifikant ( $p < .001$ ) höhere naturwissenschaftliche Kenntnisse und Fähigkeiten als Wettbewerbsteilnehmende ( $M = 0.08$ ,  $SD = 0.99$ ) und Nichtteilnehmende ( $M = -0.54$ ,  $SD = 0.64$ ,  $p < .001$ ),  $d = 1.00$  bzw.  $d = 2.40$ . Zudem zeigten die Wettbewerbsteilnehmenden signifikant höhere naturwissenschaftliche Kenntnisse und Fähigkeiten als Nichtteilnehmende ( $p < .001$ ,  $d = 0.76$ ). Es konnte kein signifikanter Haupteffekt des Geschlechts auf die naturwissenschaftlichen Kenntnisse und Fähigkeiten gefunden werden. In Abbildung 6.3 sind die Ergebnisse der zweifaktoriellen Varianzanalyse grafisch dargestellt.

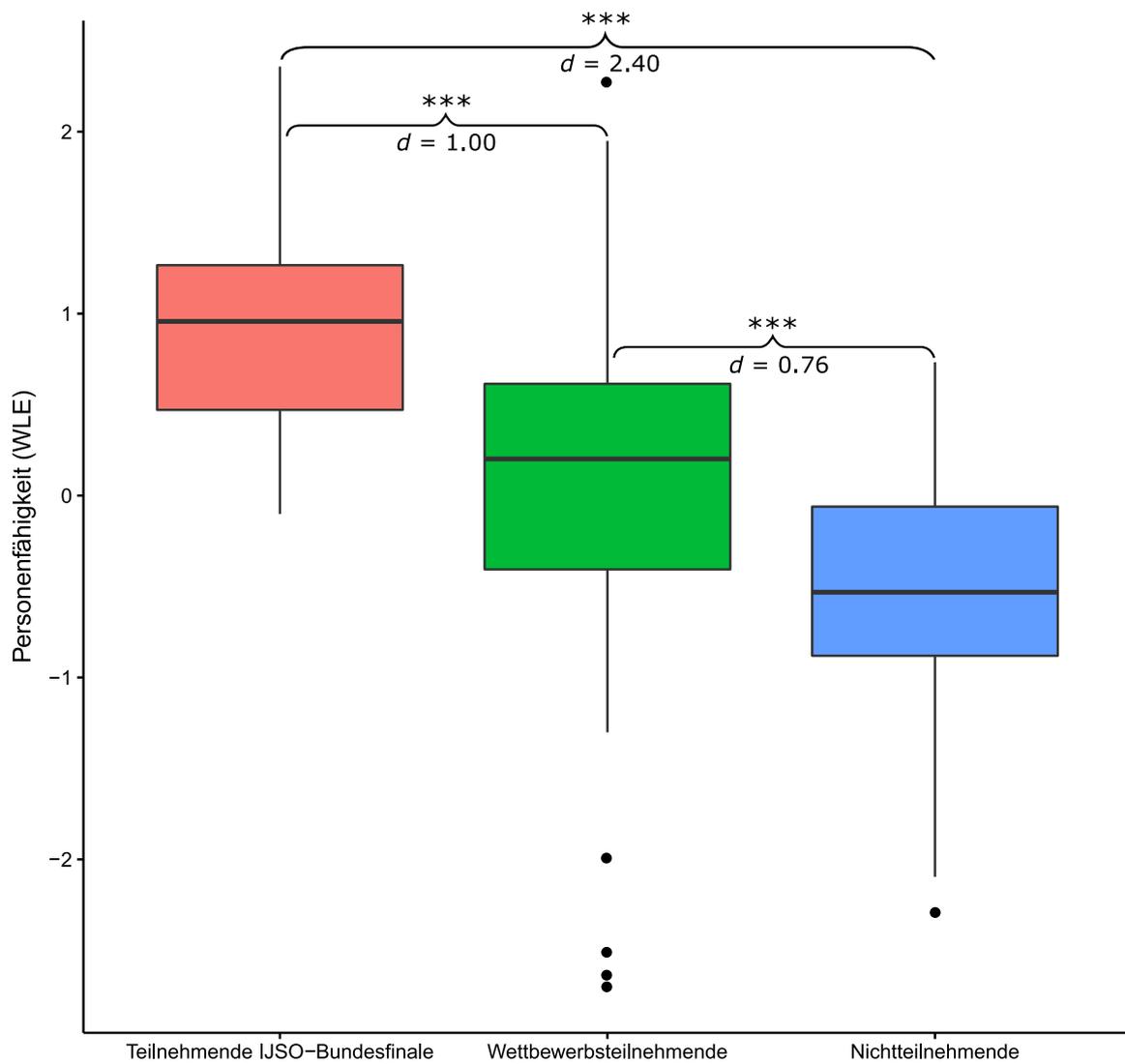


Abbildung 6.3: Ergebnisse der varianzanalytischen Auswertung der Daten aus Studie 1 \*\*\*  $p < .001$

## 6.6 Interpretation und Diskussion der Ergebnisse

Die folgende Interpretation und Diskussion der Ergebnisse fokussiert auf das Testinstrument, d. h. auf die Testqualität und die Testziele. Die Ergebnisse werden in den Forschungsstand eingeordnet. Dies dient vor allem der Prüfung der Testgüte sowie der Forschungsfrage, ob sich naturwissenschaftliche Kenntnisse und Fähigkeiten von erfolgreichen Wettbewerbsteilnehmenden im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern mit einem gemeinsamen Testinstrument erfassen lassen.

### 6.6.1 Analyse fehlender Werte

Insgesamt erreichten knapp zwei Drittel der Schülerinnen und Schüler, die die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten bearbeiteten, nicht das Ende des Tests, was hier über die nicht erfolgte Bearbeitung der letzten Aufgabe erfasst wurde. Trotz einer Testzeit von 60 Minuten und 30 Aufgaben sprach dies für zu viele bzw. zu lange Items. Darüber hinaus ließen die Schülerinnen und Schüler recht viele Items aus, was auf insgesamt schwierige Aufgaben deutete. Während der Testungen konnte zudem beobachtet werden, dass die Konzentration und die Motivation der Schülerinnen und Schüler nach maximal 45 Minuten stark nachließ, was die deskriptiven Befunde unterstützt. Hier lag die Überlegung nahe, Schülerinnen und Schüler mit einer hohen Anzahl ausgelassener und nicht erreichter Aufgaben von der Datenanalyse auszuschließen. Ausgeschlossene Personen hätten hier jedoch einer Zufallsstichprobe entsprechen müssen, um zu gewährleisten, dass die Parameterschätzungen für die übrige Stichprobe trotz Ausschluss von Personen angemessen gewesen wären (Walter, 2005). In Zufallsstichproben sind fehlende Werte wiederum ignorierbar (Walter, 2005), so dass hier schließlich keine Personen ausgeschlossen wurden.

### 6.6.2 Skalierung der Itemantworten

Für die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten, die über das eindimensionale PCM geschätzt wurden, ergab sich eine EAP/PV-Reliabilität von .86 und eine WLE-Reliabilität von .84, was als zufriedenstellend anzusehen ist. Für alle Aufgaben war mit steigender Personenfähigkeit ein Anstieg der erreichten Punkte bzw. Kategorien zu verzeichnen (vgl. R. J. Adams et al., 2012), bei 27 von 30 Aufgaben waren die Schwellenparameter geordnet. Die Aufgaben deckten den Schwierigkeitsbereich zwischen -2.112 und 1.884 ab. Obwohl dieser Bereich recht breit gestreut war, ist im Besonderen unter Beachtung der Anzahl fehlender Werte und der Wright Map anzumerken, dass die Aufgaben vor allem das obere Fähigkeitsspektrum der Schülerinnen und Schüler abdeckten und die Aufgaben somit insgesamt eine hohe Schwierigkeit aufwiesen. Da die Personenfähigkeiten jedoch auch recht breit gestreut waren und eine Differenzierung in einem breiten Spektrum an Personenfähigkeiten erfolgen sollte, wäre für folgende Erhebungen ein Ausschluss der Items mit hoher Schwierigkeit vermutlich kontraproduktiv, eher sollte das untere Fähigkeitsspektrum durch weitere Kategorien bzw. Aufgaben ergänzt werden.

### 6.6.3 Testqualität

#### Itemfit

Aufgrund des schlechten Itemfits (Mean-Squared-Infit 0.73, Infit- $t$  -2.66) wurde Aufgabe FC09 vollständig aus dem Testinstrument entfernt. Die verbleibenden Aufgaben wiesen nach mehreren Überarbeitungsschritten, in denen in neun Aufgaben Kategorien zusammengefasst wurden, alle einen Mean-Squared-Infit in den vorgeschriebenen Grenzen auf. Lediglich drei Aufgaben wiesen einen Infit- $t$ -Wert außerhalb der geforderten Grenzen auf, was hier jedoch in Anlehnung an Wilson (2005) als unproblematisch angesehen wurde.

#### Differential Item Functioning

Die Analysen zum DIF ergaben, dass diejenigen Modelle, die DIF berücksichtigten, nicht signifikant besser auf die Daten passten als die Modelle, die nur einen Haupteffekt modellierten. Somit ist davon auszugehen, dass die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten keine der untersuchten Subgruppen, eingeteilt nach Wettbewerbserfahrung und Geschlecht, systematisch bevorzugten. Dieser Befund spricht zudem für die Konstruktvalidität des Testinstruments (vgl. Wilbert & Linnemann, 2011).

#### Mehrdimensionalität

Bezüglich der möglichen Mehrdimensionalität der Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten erwies sich kein mehrdimensionales Modell als signifikant besser als das eindimensionale Modell. Da zumindest die WLE-Reliabilität für das eindimensionale Modell höher lag und die einzelnen Dimensionen recht hoch miteinander korrelierten (Kompetenzbereiche:  $r = .93$ , Fächer: jeweils  $r = .95$ ), wurde entschieden, das eindimensionale Modell beizubehalten. Das sprach zunächst gegen die intendierte Testkonstruktion, da die Aufgaben im ursprünglichen Sinn die beiden Kompetenzbereiche Erkenntnisgewinnung und Fachwissen sowie die drei Fächer Biologie, Chemie und Physik abdecken sollten. Allerdings wurden im Test pro Kompetenzbereich und pro Fach lediglich vier (Erkenntnisgewinnung) bzw. fünf (Fachwissen) Aufgaben eingesetzt. Womöglich war es mit dieser geringen Anzahl an Items nicht möglich, die Kompetenzbereiche bzw. Fächer klar voneinander abzugrenzen. Zudem wurden die Aufgaben aus dem Bereich Erkenntnisgewinnung zusätzlich zum Fach auch in die Unterkategorien „Fragestellungen und Hypothesen finden“, „Versuche planen“, „Versuche auswerten“ und „Fehler analysieren“ eingeteilt, was eine Ausdifferenzierung mehrerer Dimensionen mit der geringen Anzahl an eingesetzten Testaufgaben nochmals erschwert haben könnte.

#### Testziele - Korrelationen

Betrachtet man die signifikanten Korrelationen zur Personenfähigkeit, lag die Korrelation der nonverbalen kognitiven Fähigkeiten zur Personenfähigkeit mit  $r = .33$  niedriger als die mittlere Korrelation zwischen allgemeiner Intelligenz und Schulleistung ( $r = .50$  bis  $r = .60$ , Helmke & Schrader, 2010). Hier ist jedoch zu beachten, dass nur kognitive Teilfähigkeiten, und hier die nonverbalen Fähigkeiten, mit einer einzigen Skala erhoben wurden, so dass kaum Rückschlüsse auf die allgemeine Intelligenz der Schülerinnen und Schüler gezogen werden können. Des Weiteren kann die Testleistung, bzw. die Personenfähigkeit, nicht mit schulischer Leistung gleichgesetzt werden. Die hier gefundene mittlere Korrelation zwischen der Personenfähigkeit und den nonverbalen kognitiven Fähigkeiten von  $r = .33$  deutet jedoch darauf hin, dass zum

Erreichen einer guten Leistung bei den Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten nicht ausschließlich die Intelligenz von Bedeutung war. Eine weitere Begründung für die im Vergleich niedrigere Korrelation könnte auch sein, dass ein hohes Maß domänenspezifischen Wissens für eine gute Leistung im Test erforderlich war, was nicht durch hohe intellektuelle Fähigkeiten kompensiert werden kann (vgl. Abschnitt 2.3.2). Betrachtet man in Ergänzung die Korrelation zwischen der Personenfähigkeit und den Schulnoten in Biologie, Chemie und Physik zeigt sich erwartungskonform eine mittlere Korrelation von  $r = -.53$  (die negative Korrelation kommt hier dadurch zustande, dass geringe Zahlenwerte für die Schulnoten gute Leistungen beschreiben und umgekehrt). Annähernd vergleichbar sind hier die Ergebnisse von Lau und Roeser (2002), die für die Korrelation zwischen kognitiven Fähigkeiten und Leistungen in den Naturwissenschaften (hier operationalisiert über das Ergebnis eines naturwissenschaftlichen Multiple-Choice-Tests sowie über die Schulnoten der Teilnehmenden in den Naturwissenschaften) einen Wert von  $r = .67$ ,  $p < .01$  fanden.

Vielfach wurden positive Zusammenhänge zwischen dem Selbstkonzept und (schulischer) Leistung nachgewiesen (z. B. Marsh & Martin, 2011; Marsh et al., 2005). Hansford und Hattie (1982) fanden in einer Metaanalyse eine mittlere Korrelation von  $r = .42$  für den Zusammenhang zwischen dem akademischen Selbstkonzept und (schulischer) Leistung. Für den Zusammenhang zwischen dem fachspezifischen Selbstkonzept und der (schulischen) Leistung finden sich tendenziell höhere Korrelation in Größenordnungen von  $r = .60$  (nach Möller & Köller, 1995; 2000, zit. n. Köller, 2004). Die aus Studie 1 generierten Ergebnisse mit einer Korrelation von  $r = .44$  für das akademische Selbstkonzept und  $r = .59$  für das naturwissenschaftliche Selbstkonzept spiegelt diese Werte wider.

Für Korrelationen des fachspezifischen Interesses mit schulischer Leistung fanden Krapp et al. (1993) in einer Metaanalyse eine durchschnittliche Korrelation von  $r = .30$ . Betrachtet man die Interesse-Leistungs-Korrelation in den Naturwissenschaften, fand sich eine durchschnittlich gewichtete Korrelation von  $r_M = .31$ . Helmke und Schrader (2010) berichteten eine mittlere Korrelation zwischen Interesse und Leistung von  $r = .41$ . Auch die Ergebnisse von Studie 1 zeigten Korrelationen in dieser Größenordnung, sie lagen zwischen  $r = .31$  (Interesse an Tätigkeiten, die im naturwissenschaftlichen Schulunterricht bzw. in einem naturwissenschaftlichen Beruf ausgeübt werden) und  $r = .34$  für das Interesse an Tätigkeiten, die in einer naturwissenschaftlichen Fördermaßnahme ausgeübt werden.

Schülerinnen und Schüler mit einer höheren Selbstwirksamkeitserwartung in einem bestimmten Fach zeigen auch bessere Leistungen (z. B. Lopez & Lent, 1992; Pajares, 1997). Für die Selbstwirksamkeitserwartung der Schülerinnen und Schüler für die spätere Ausübung von Tätigkeiten eines Wissenschaftlers fand sich eine signifikante mittlere Korrelation von  $r = .34$ , die etwas niedriger lag als jene, wie sie z. B. von Hackett und Betz (1989) für den Zusammenhang zwischen Selbstwirksamkeitserwartung und Leistung in Mathematik berichtet wurde ( $r = .44$ ).

Wirthwein et al. (2013) sowie Steinmayr et al. (2011) konnten zeigen, dass die Korrelationen zwischen den Zielorientierungen und der Leistung tendenziell eher gering ausfallen. Dabei korrelieren Lernziele und Annäherungsleistungsziele meist positiv mit der Leistung, wohingegen Vermeidungslernziele und Arbeitsvermeidung meist negativ mit der Leistung zusammenhängen (Steinmayr et al., 2011). Wirthwein et al. (2013) fanden in einer Metaanalyse

Korrelationen zwischen  $r = -.12$  und  $r = .13$ , Steinmayr et al. (2011) zeigten Korrelationen zwischen den Zielorientierungen und den Schulnoten zwischen  $r = .24$  (Lernziele) und  $r = -.11$  (Arbeitsvermeidung). Ähnliche Werte fanden sich auch hier, wobei keine signifikante Korrelation zwischen den Annäherungsleistungszielen und der Personenfähigkeit gefunden werden konnte. Tendenziell lagen die Werte hier etwas höher als in der Literatur berichtete Werte für die Korrelation zwischen den Zielorientierungen und der Leistung:  $r = .32$  für die Lernziele,  $r = -.20$  für die Arbeitsvermeidung und  $r = -.21$  für die Vermeidungslernziele.

Campbell (1996) konnte zeigen, dass Mathematik-Olympioniken sehr gute mathematisch-naturwissenschaftlichen Schulnoten aufwiesen und schon viele Wettbewerbe in der Schulzeit gewonnen hatten. In Einklang mit diesen Ergebnissen fand sich bei Urhahne et al. (2012) als einziger signifikanter Prädiktor für eine erfolgreiche Wettbewerbsbeteiligung, hier das Erreichen der vierten Runde der IChO, die vorherige Teilnahme (vgl. Abschnitt 2.2.5). Der Einfluss der Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb auf die naturwissenschaftlichen Kenntnisse und Fähigkeiten wurde hier über die Korrelation zur Personenfähigkeit ermittelt. Erwartungskonform fand sich ein signifikanter mittlerer Zusammenhang von  $r_{pb} = .47$ . Betrachtet man den Zusammenhang zwischen der Anzahl an Wettbewerbsteilnahmen und der Personenfähigkeit zeigte sich sogar eine hohe Korrelation von  $r = .57$ .

Auch für das Alter zeigte sich eine signifikante mittlere Korrelation von  $r = .30$  zur Personenfähigkeit, was im Wesentlichen auf den Lernfortschritt zurückzuführen ist, den ältere Schülerinnen und Schüler gegenüber jüngeren Schülerinnen und Schülern aufweisen.

Fasst man die aus den Korrelationsanalysen erhaltenen Ergebnisse zusammen ist festzustellen, dass die Korrelationen zwischen den aufgeführten Variablen und Konstrukten und der Testleistung weitgehend Erkenntnissen aus der Forschung entsprechen. Somit ist davon auszugehen, dass das eingesetzte Testinstrument zu naturwissenschaftlichen Kenntnissen und Fähigkeiten als extern valide bezeichnet werden kann.

### **Testziele - Multiple lineare und binäre logistische Regressionen**

Die multiplen linearen Regressionen zeigten, dass die nonverbalen kognitiven Fähigkeiten, das Alter der Schülerinnen und Schüler, das Geschlecht zugunsten der Jungen, die Schulnoten in Biologie, Chemie und Physik (gemittelt in einer Variablen) sowie die Anzahl an Wettbewerbsteilnahmen einen Einfluss auf die naturwissenschaftlichen Kenntnisse und Fähigkeiten (d. h. hier die Personenfähigkeit) hatten. Abgesehen vom Geschlecht stehen diese Befunde in Einklang mit den Kenntnissen zu Einflussfaktoren von (schulischer) Leistung (vgl. Abschnitt 2.3.2). Bei Betrachtung der Zusammenhänge zwischen den Prädiktoren untereinander zeigte sich, dass die Schulnoten in Biologie, Chemie und Physik signifikant zu  $r = -.46$ ,  $p < .001$  mit der Anzahl an Wettbewerbsteilnahmen korrelierten (die negative Korrelation kommt hier dadurch zustande, dass geringe Zahlenwerte für die Schulnoten gute Leistungen beschreiben und umgekehrt). Auch bestand ein signifikanter mittlerer Zusammenhang zwischen der Anzahl an Wettbewerbsteilnahmen und dem Alter,  $r = .42$ ,  $p < .001$ . Zur detaillierten Analyse des Einflusses des Geschlechts wurden nach Geschlecht getrennte multiple lineare Regressionen durchgeführt. Hier zeigten sich unterschiedliche Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten, was zum Einfluss des Geschlechts in der gemeinsamen Analyse geführt haben könnte. So zeigten sich für Mädchen die nonverbalen

kognitiven Fähigkeiten, die Schulnoten in Biologie, Chemie und Physik sowie das Alter prädiktiv für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten, folglich abgesehen vom Alter rein kognitive Faktoren. Allerdings konnte hier auch nur eine Varianzaufklärung von 33.6% erreicht werden, so dass weitere Einflussfaktoren existieren, die hier nicht erfasst wurden. Für die Jungen waren das fachspezifische Selbstkonzept, das Alter und die nonverbalen kognitiven Fähigkeiten signifikante Einflussfaktoren, jedoch nicht, wie bei den Mädchen, die Schulnoten. Hinsichtlich des fachspezifischen Selbstkonzepts bestand ein signifikanter Unterschied der Jungen ( $M = 3.16, SD = 0.67$ ) im Vergleich zu den Mädchen ( $M = 2.86, SD = 0.69$ ),  $t(182.78) = -3.03, p < .01, r = 0.23$  (siehe auch Kessels & Hannover, 2004; Köller et al., 2000), was den im Gegensatz zu den Mädchen vorhandenen Einfluss auf die naturwissenschaftlichen Kenntnisse und Fähigkeiten erklärt. Ebenso signifikant,  $t(178.51) = -2.23, p < .05, r = 0.17$ , war der Unterschied in den Schulnoten in Biologie, Chemie und Physik zugunsten der Jungen ( $M = 2.00, SD = 0.75$ , Mädchen:  $M = 2.27, SD = 0.89$ ), wobei der Effekt des fachspezifischen Selbstkonzepts den Effekt der Schulnoten für die Jungen vermutlich überlagerte, während das vergleichsweise niedrigere fachspezifische Selbstkonzept der Mädchen durch den Einfluss der Schulnoten kompensiert wurde. Die Anzahl der Wettbewerbsteilnahmen wurden in den nach Geschlecht getrennten Analysen nicht signifikant prädiktiv.

Die nach Wettbewerbserfahrung getrennten Regressionen zeigten interessante Unterschiede in den prädiktiven Variablen zwischen den einzelnen Subgruppen: So konnten für die Teilnehmenden des IJSO-Bundesfinales als signifikante Prädiktoren für naturwissenschaftliche Kenntnisse und Fähigkeiten das Alter sowie das Geschlecht zugunsten der Jungen identifiziert werden, aufgrund der Stichprobengröße sind diese Ergebnisse allerdings unter Vorbehalt zu interpretieren. Zudem konnte mit diesen beiden Faktoren lediglich eine Varianzaufklärung von 33.5% erreicht werden. Insgesamt entsprechen die Ergebnisse nicht den Forschungsergebnissen von Urhahne et al. (2012), nach denen man einen stärkeren Einfluss der Anzahl vorheriger Wettbewerbsteilnahmen erwartet hätte. Allerdings untersuchten Urhahne et al. (2012), welche Variablen vorhersagen, ob die IChO-Teilnehmenden die vierte Runde erreichen, und nicht, welche Variablen ein gutes Abschneiden in den Klausuren des Wettbewerbs bedingen. Zudem lag der Mittelwert an Wettbewerbsteilnahmen mit  $M = 4.79, SD = 1.62$  recht hoch (18 von 39 Schülerinnen und Schülern hatten schon mindestens sechs Mal an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen), so dass vermutlich aus diesem Grund die Anzahl der Wettbewerbsteilnahmen als Prädiktor nicht signifikant wurde. Eine mögliche Erklärung des Einflusses des Geschlechts für die Teilnehmenden des IJSO-Bundesfinales könnte im Anteil der Jungen im IJSO-Bundesfinale gesehen werden, da zwei Drittel der Teilnehmenden männlich waren. (Das Geschlecht korrelierte zudem signifikant mit der Anzahl an Wettbewerbsteilnahmen ( $r = -.25, p < .001$ ), so dass dieser Effekt den Einfluss der Anzahl an Wettbewerbsteilnahmen möglicherweise überlagerte.) Allerdings besteht auch die Möglichkeit, dass Unterschiede in bspw. dem Wettbewerbsselbstkonzept (vgl. Blankenburg et al., 2015) bestanden, die hier nicht untersucht wurden. Nach Geschlecht getrennte Regressionen konnten hier aufgrund der Stichprobengröße nicht durchgeführt werden. Der Einfluss des Alters ist auf den Lernfortschritt zurückzuführen. Dass Prädiktoren wie das naturwissenschaftliche Selbstkonzept oder die Noten in Biologie, Chemie und Physik in der Gruppe der Teilnehmenden des IJSO-Bundesfinales nicht signifikant wurden, kann möglichen Deckeneffekten zugeschrieben werden. Der Mittelwert für das naturwissenschaftliche Selbstkonzept lag bspw. bei  $M = 3.74, SD = 0.28$  mit einem Minimum von 3 auf einer vierstufigen Likert-Skala. Ähnlich hohe Mittelwerte fanden sich bei den Schulnoten in Biologie, Chemie und Physik: Der

Mittelwert lag hier bei  $M = 1.33$ ,  $SD = 0.44$  bei einem Maximum von 3 bei einer Notenskala von 1 bis 6.

Für die Gruppe der Wettbewerbsteilnehmenden konnten das naturwissenschaftliche Selbstkonzept, die nonverbalen kognitiven Fähigkeiten sowie das Alter als prädiktiv für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten gefunden werden. Das Geschlecht fiel in der Gruppe der Wettbewerbsteilnehmenden als Prädiktor heraus. In der Gruppe der Teilnehmenden des IJSO-Bundesfinales fanden sich jedoch auch nur 33.3% Mädchen, in der Gruppe der Wettbewerbsteilnehmenden waren es 42.9%. Der Einfluss des Alters lässt sich wiederum auf den Lernfortschritt älterer Schülerinnen und Schüler zurückführen, der Einfluss kognitiver Fähigkeiten und des fachspezifischen Selbstkonzepts auf die Leistung ist aus der Literatur bekannt und spricht wiederum für die Validität des eingesetzten Testinstruments.

In der Gruppe der Nichtteilnehmenden wurde das Geschlecht (54.7% Mädchen) in der Regression neben den nonverbalen kognitiven Fähigkeiten und den Schulnoten in Biologie, Chemie und Physik signifikant. Allerdings konnten durch die identifizierten Prädiktoren nur 24.0% der Varianz aufgeklärt werden, so dass auch hier noch weitere Einflussfaktoren existierten, die nicht erfasst wurden. Zur Erklärung des Geschlechterunterschieds wurden  $t$ -Tests für die Begleitvariablen durchgeführt, die allerdings keine signifikanten Unterschiede zeigten. Der Einfluss des Geschlechts auf die Testleistung zugunsten der Jungen ist somit vermutlich auf Faktoren zurückzuführen, die im Rahmen der Studie nicht erhoben wurden, wie z. B. möglicherweise vorhandene Geschlechterstereotypen (vgl. z. B. Nosek et al., 2009). Nach Geschlecht getrennte Regressionsanalysen wurden aufgrund der Stichprobengröße nicht durchgeführt.

Die Ergebnisse der binären logistischen Regressionen zeigten, dass die Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb durch die Schulnoten in Biologie, Chemie und Physik, dem Alter sowie dem Interesse an Tätigkeiten, die in naturwissenschaftlichen Fördermaßnahmen ausgeübt werden, vorhergesagt werden konnte. Dass auch die Schulnoten in Biologie, Chemie und Physik einen Einfluss auf die Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb hatten, könnte dadurch erklärt werden, dass Lehrkräfte vermutlich eher Schülerinnen und Schüler mit entsprechend guten bis sehr guten Noten für eine Teilnahme am Wettbewerb motivieren. Das Alter als Prädiktor erklärt sich dadurch, dass ältere Schülerinnen und Schüler, unabhängig der Anzahl der Teilnahmen, am ehesten an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten als jüngere Schülerinnen und Schüler, was sich auch in den Daten zeigte (83.3% der 16-jährigen hatten bereits an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen, aber nur 42.7% der 14-jährigen). Dies mag hier allerdings auch der Zusammensetzung der Stichprobe geschuldet sein. Das Interesse an Tätigkeiten, die in naturwissenschaftlichen Fördermaßnahmen ausgeübt werden, als Prädiktor für eine Wettbewerbsteilnahme erklärt sich im Wesentlichen dadurch, dass Schülerinnen und Schüler mit einem hohen Interesse in diesem Bereich mit hoher Wahrscheinlichkeit auch an derartigen Fördermaßnahmen teilnehmen, wozu auch mathematisch-naturwissenschaftliche Schülerwettbewerbe zählen.

Wettbewerbserfolg, hier die Teilnahme am IJSO-Bundesfinale 2013, konnte mit der Personenfähigkeit, d. h. den naturwissenschaftlichen Kenntnissen und Fähigkeiten, den Schulnoten in Biologie, Chemie und Physik sowie der Anzahl der Wettbewerbsteilnahmen erklärt werden. Der Einfluss der Personenfähigkeit, d. h. der naturwissenschaftlichen Kenntnisse und

Fähigkeiten, auf den Wettbewerbserfolg steht in Einklang mit den Erkenntnissen aus der (Expertise-) Forschung zum Einfluss (domänenspezifischen) Wissens auf die Schulleistung (siehe Abschnitt 2.3.2). Der Einfluss der Schulnoten ist dadurch zu erklären, dass die Gruppe der Teilnehmenden des IJSO-Bundesfinales vermutlich eine Vielzahl hochleistender und hochbegabter Schülerinnen und Schüler einschließt, deren Leistung sich auch in ihren schulischen Noten widerspiegelt. Allerdings fanden Stang et al. (2014) und Urhahne et al. (2012) keinen signifikanten Einfluss der Schulnoten auf Wettbewerbserfolg. Gerade die Bedeutung der wiederholten Teilnahme an einem Wettbewerb konnte schon von Blankenburg et al. (2015) und Urhahne et al. (2012) gezeigt werden und wird durch die vorliegenden Ergebnisse bestätigt. Die kognitiven Fähigkeiten konnten hier nicht als signifikanter Prädiktor für Wettbewerbserfolg geprüft werden, da die Werte für die IJSO-Bundesfinalisten nicht verwendet werden konnten (siehe Abschnitt 6.5.3). Ohne diesen Faktor wurde jedoch bereits eine Varianzaufklärung von 72.3% (Hosmer-Lemeshow) erreicht.

Fasst man die Ergebnisse zu den multiplen linearen und binären logistischen Regressionen zusammen, wird deutlich, dass die Ergebnisse Erkenntnisse aus der Forschung im Wesentlichen bestätigen, was wiederum für die Güte des eingesetzten Testinstruments spricht.

### Testziele - Varianzanalysen

In der Varianzanalyse zeigten sich signifikante Unterschiede mit großen Effektstärken zwischen den nach Wettbewerbserfahrung unterteilten Subgruppen, was den Erwartungen entsprach. Das Testziel, zwischen Teilnehmenden verschiedener naturwissenschaftlicher Schülerwettbewerbe und nichtteilnehmenden Schülerinnen und Schülern zu unterscheiden, wurde somit erreicht. Zwischen den Geschlechtern zeigten sich keine signifikanten Unterschiede, auch wenn eine geringe signifikante Korrelation zugunsten der Jungen ( $r_{pb} = -.27$ ) vorlag. Hierzu wurden in bisherigen Studien unterschiedliche Ergebnisse gefunden. Urhahne et al. (2012) fanden im Bereich fachspezifischer Tests (hier: Klausuren der dritten Auswahlrunde der IChO) im Vergleich zwischen weiblichen und männlichen Teilnehmenden, dass männliche Teilnehmende signifikant besser abschnitten als weibliche Teilnehmende, während der IQB-Ländervergleich 2012 in Mathematik und den Naturwissenschaften zeigte, dass Mädchen in den naturwissenschaftlichen Kompetenzbereichen höhere Werte als Jungen erzielten (Schroeders, Penk, Jansen & Pant, 2013). Lengfelder und Heller (2002) konnten keinen Unterschied zwischen den Geschlechtern bei deutschen Olympioniken finden. Die jüngste PISA-Erhebung (Reiss, Sälzer, Schiepe-Tiska, Klieme & Köller, 2016) zeigte wiederum eine signifikant bessere naturwissenschaftliche Kompetenz für Jungen als für Mädchen (Schiepe-Tiska, Simm & Schmidtner, 2016).

## 6.7 Zusammenfassung und Fazit

In Studie 1 wurde das in Kapitel 5 entwickelte Testinstrument zu naturwissenschaftlichen Kenntnissen und Fähigkeiten von Wettbewerbsteilnehmenden erprobt und validiert. An den Erhebungen nahmen Teilnehmende des Bundesfinales der IJSO 2013, Teilnehmende zweier Landeswettbewerbe Jugend forscht bzw. Schüler experimentieren 2014 sowie Schülerinnen und Schüler, die mindestens einmal oder gar nicht an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten, teil. Für die Prüfung der Testqualität sowie der Testziele wurden Methoden der probabilistischen Testtheorie angewendet sowie Korrelations-,

Regressions- und Varianzanalysen durchgeführt.

Zusammenfassend ist festzustellen, dass das eingesetzte Testinstrument naturwissenschaftliche Kenntnisse und Fähigkeiten mit einer EAP/PV-Reliabilität von .86 (WLE-Reliabilität .84) reliabel und valide in der untersuchten Stichprobe erfassen konnte. Die Objektivität des entwickelten Kategoriensystem zur Bewertung der Schülerantworten wurde über die Beurteilerreliabilität ermittelt, wobei sich eine signifikante Spearman-Rangkorrelation mit einem Wert von  $\rho = .98$ ,  $p < .001$ . fand, was als ein hoher Zusammenhang zu interpretieren ist. Die Prüfung der Testqualität ergab, dass die eingesetzten Aufgaben akzeptable Fit-Werte aufwiesen. DIF-Analysen führten zu keinem signifikanten Unterschied zwischen einer Modellierung der Daten mit und ohne DIF. Des Weiteren deuteten die Ergebnisse der Skalierung über das PCM darauf hin, dass die naturwissenschaftlichen Kenntnisse und Fähigkeiten als ein eindimensionales Konstrukt angesehen werden konnten. Dies bedeutete jedoch gleichzeitig, dass mithilfe der Aufgaben keine spezifischen Kenntnisse in den Fächern Biologie, Chemie und Physik oder den Kompetenzbereichen Fachwissen und Erkenntnisgewinnung erfasst werden konnten. Die Prüfung der Testziele zeigte erwartete Zusammenhänge zwischen den naturwissenschaftlichen Kenntnissen und Fähigkeiten, d. h. der Personenfähigkeit, und verschiedenen Einflussgrößen (schulischer) Leistung (vgl. Abschnitt 2.3.2). In multiplen linearen Regressionen wurden die nonverbalen kognitiven Fähigkeiten, das Alter der Schülerinnen und Schüler, das Geschlecht, die Anzahl an Wettbewerbsteilnahmen und die Schulnoten in Biologie, Chemie und Physik als signifikante Prädiktoren für hohe naturwissenschaftliche Kenntnisse und Fähigkeiten ermittelt. Unterschiede in den Prädiktoren ergaben sich auch in nach Wettbewerbserfahrung eingeteilten Subgruppen. Logistische Regressionen identifizierten die Schulnoten in Biologie, Chemie und Physik, das Alter der Schülerinnen und Schüler sowie das Interesse an Tätigkeiten, die in naturwissenschaftlichen Fördermaßnahmen ausgeübt werden, als signifikante Prädiktoren für eine Wettbewerbsteilnahme. Für Wettbewerbserfolg, d. h. hier das Erreichen des IJSO-Bundesfinales 2013, wurden in Einklang mit Erkenntnissen aus der Forschung die naturwissenschaftlichen Kenntnisse und Fähigkeiten (d. h. die Personenfähigkeit), die Schulnoten in Biologie, Chemie und Physik sowie die Anzahl der Wettbewerbsteilnahmen als signifikante Prädiktoren ermittelt. Das Testziel, zwischen Teilnehmenden verschiedener naturwissenschaftlicher Schülerwettbewerbe sowie zu nichtteilnehmenden Schülerinnen und Schülern aufzudecken, wurde erreicht. So zeigten sich in zweifaktoriellen Varianzanalysen Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten zwischen Teilnehmenden des IJSO-Bundesfinales 2013, Schülerinnen und Schülern, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten und nichtteilnehmenden Schülerinnen und Schülern. Auch zwischen den beiden letztgenannten Gruppen gab es signifikante Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten. Mögliche Unterschiede zwischen Teilnehmenden spezifischer Schülerwettbewerbe, z. B. der IJSO und Jugend forscht bzw. Schüler experimentieren, konnten aufgrund der geringen Stichprobe nicht geprüft werden.

Die Testziele, in einem einzigen Erhebungsinstrument für die anvisierte Zielgruppe Kenntnisse und Fähigkeiten von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe und nichtteilnehmenden Schülerinnen und Schülern in den drei Naturwissenschaften Biologie, Chemie und Physik und verschiedenen Kompetenzbereichen zu erfassen und Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten zwischen Teilnehmenden verschiedener naturwissenschaftlicher Schülerwettbewerbe sowie zu nichtteilnehmenden Schülerinnen und

Schülern aufzudecken, wurden erreicht. Die Forschungsfrage, ob sich die naturwissenschaftlichen Kenntnisse und Fähigkeiten von erfolgreichen Wettbewerbsteilnehmenden im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern mit einem gemeinsamen Testinstrument erfassen lassen, ist somit für Studie 1 positiv zu beantworten.

Einschränkend zu erwähnen ist, dass die Analyse fehlender Werte und die Wright Map zeigten, dass der Test offensichtlich zu viele und zu schwierige Aufgaben enthielt. Von der organisatorischen Seite betrachtet ergaben sich in Studie 1 zudem Schwierigkeiten in der Rekrutierung der Stichprobe sowie im Ablauf der Erhebungen insgesamt. Um den Bedürfnissen der Wettbewerbsforschung gerecht zu werden, sollte daher in Studie 2, siehe das folgende Kapitel 7, ein reduziertes Testinstrument an einer umfangreichen Stichprobe erprobt werden.

## Kapitel 7

# Erprobung eines reduzierten Testinstruments zu naturwissenschaftlichen Kenntnissen und Fähigkeiten (Studie 2)

Für die Forschung zu Schülerwettbewerben und die Charakterisierung ihrer Teilnehmenden sind eine Reihe affektiver und kognitiver Merkmale von Interesse, weshalb meist eine Vielfalt verschiedener Erhebungsinstrumente eingesetzt wird. Da Teilnehmende derartiger Studien meist weit verstreut und somit außerhalb des Wettbewerbs schwer zu erfassen sind, müssen querschnittliche Erhebungen, die zeitnah zum oder mit dem Wettbewerb erfolgen, den Rahmenbedingungen der jeweiligen Wettbewerbe angepasst werden. Hieraus resultiert, dass die Testzeit gerade in der Wettbewerbsforschung meist äußerst knapp bemessen ist, weshalb das in Studie 1 erprobte Testinstrument reduziert und in Studie 2 mit einer umfangreichen Stichprobe, die ein breites Spektrum an Wettbewerbsteilnehmenden und Schülerinnen und Schülern enthalten sollte, erprobt wurde. Die Prüfung der Testqualität erfolgte, wie auch schon in Studie 1, anhand von Methoden der probabilistischen Testtheorie (siehe Kapitel 6).

Im Folgenden werden zunächst die Folgerungen geschildert, die aus den Erhebungen für Studie 1 abgeleitet wurden. Anschließend wird das reduzierte Testinstrument vorgestellt und die Durchführung, die Auswertung sowie die Ergebnisse der Studie 2 berichtet. Im Zentrum stand die Beantwortung der folgenden Forschungsfrage:

Gelingt die Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten mit einem an die Bedürfnisse der Wettbewerbsforschung angepassten Testinstrument?

### 7.1 Folgerungen aus Studie 1

Im Laufe der Erhebungen von Studie 1 traten Hindernisse in der Planung und Durchführung der Erhebungen auf, die zuvor nicht in dem Maße abzusehen waren. Im Folgenden werden die Schwierigkeiten, die während Studie 1 auftraten, noch einmal kurz erläutert und die Lösungs-

ansätze für die Erhebungen in Studie 2 vorgestellt.

### 7.1.1 Folgerungen für die Rekrutierung der Stichprobe

Die Auswahl der Stichprobe für Studie 1 erfolgte, abgesehen von den Teilnehmenden der IJSO-Bundesfinale, die direkt über die IJSO-Wettbewerbsleitung kontaktiert wurden, über den Kontakt zu Lehrkräften, die Schülerinnen oder Schüler bis zum IJSO-Bundesfinale 2013 betreut hatten. Eine mögliche Teilnahme der Schülerinnen und Schüler richtete sich folglich nach der Bereitschaft der jeweiligen Lehrkräfte und ihrer Schulen, die jedoch sehr gering ausfiel. Die Rekrutierung von Teilnehmenden an Jugend forscht bzw. Schüler experimentieren gestaltete sich ebenso schwierig: Aufgrund des eng getakteten Zeitplans der Landeswettbewerbe Jugend forscht bzw. Schüler experimentieren konnten die Erhebungen nicht überall in den Wettbewerbsablauf eingebettet werden, so dass keine ausreichend große Stichprobe gewonnen werden konnte. Da Studie 1 nach den Erhebungen in den Landeswettbewerben Jugend forscht bzw. Schüler experimentieren 2013 in Hamburg und Niedersachsen aufgrund der geringen Anzahl an Teilnehmenden beendet wurden, konnten auch keine Klassen bzw. Schulen von Teilnehmenden mehr für die Erhebungen gewonnen werden.

Für Studie 2 wurden Schulen angeschrieben, die in den Jahren 2012, 2013 und 2014 in Folge mit Schülerinnen und Schülern an der IJSO teilgenommen hatten und von denen in jedem Jahr Teilnehmende mindestens die zweite Runde erreicht hatten. Nach Aussortieren der Schulen, die schon einmal an den Erhebungen teilgenommen hatten oder eine Erhebung im Rahmen von Studie 1 abgelehnt hatten, verblieben 33 zu kontaktierende Schulen. Für die Rekrutierung von Teilnehmenden an Jugend forscht bzw. Schüler experimentieren wurde auf die Landeswettbewerbe weitestgehend verzichtet. Stattdessen wurden Schulen angeschrieben, die hauptsächlich 2014, teilweise auch 2013 oder 2012, den „Jugend forscht Schulpreis“ erhalten hatten. Dieser Preis wird Schulen verliehen, die „hervorragende Leistungen unter Beweis stellen konnten“ (Stiftung Jugend forscht e. V., 27.06.2014). Bewertet wurde bei der Verleihung des Preises „neben der Anzahl vor allem die Qualität der eingereichten Forschungsprojekte“ sowie „die besondere Förderkultur von Schulen im naturwissenschaftlich-technischen Bereich und der Stellenwert, den der Wettbewerb Jugend forscht dort einnimmt“ (Stiftung Jugend forscht e. V., 27.06.2014). Hierdurch sollten möglichst viele Schülerinnen und Schüler für die Erhebungen gewonnen werden, die bereits an dem Wettbewerb, bestenfalls auch erfolgreich, teilgenommen hatten. Nach Aussortieren von nichtgymnasialen Schulen sowie Schulen, die schon im Rahmen von Studie 1 angeschrieben wurden, verblieben 17 zu kontaktierende Schulen.

Insgesamt wurden 50 Schulen angeschrieben, wobei der Kontakt im Gegensatz zu Studie 1 direkt über die Schulleitungen erfolgte. Dies hatte den Vorteil, dass die Schulleitungen bei einer positiven Rückmeldung in jedem Fall die Erhebungen unterstützten. Des Weiteren wurden nur neunte und zehnte Klassen bzw. Jahrgänge angefragt, da sich in Studie 1 gezeigt hatte, dass die Teilnehmenden der IJSO-Bundesfinale ein durchschnittliches Alter von  $M = 14.59$  Jahren,  $SD = 0.64$  aufwiesen und somit tendenziell etwas älter waren als erwartet. Auch wurden nur Gymnasien, teilweise MINT-EC-Schulen, angeschrieben, da sich gezeigt hatte, dass die Teilnehmenden der IJSO-Bundesfinale in der Regel diese Schularten besuchten. Zudem wurde angestrebt, eine möglichst breite Auswahl an Schulen, verteilt über das gesamte Bundesgebiet, zu erreichen. Die Teilnehmenden der IJSO-Bundesfinale 2014 sollten in gleicher Weise wie

schon die Teilnehmenden der IJSO-Bundesfinale 2013 kontaktiert werden.

### **7.1.2 Folgerungen für den Ablauf der Erhebungen und die eingesetzten Erhebungsinstrumente**

In Studie 1 fand die Bearbeitung des Fragebogens sowie der übrigen Erhebungsinstrumente getrennt voneinander statt. Dadurch konnten einige Erhebungsinstrumente nicht einander zugeordnet werden. Zudem stellte sich im Laufe von Studie 1 heraus, dass der organisatorische Aufwand (Organisation von potentiell zu testenden Klassen, Verteilen und Einsammeln von Einverständniserklärungen und Fragebögen, Organisieren eines Raumes für die Erhebungen, etc.) für die Lehrkräfte recht hoch war und dies die Lehrkräfte womöglich davon abhielt, an den Erhebungen teilzunehmen. Auch für die Schülerinnen und Schüler, die an den Erhebungen teilnahmen, schien der Umfang der Erhebungsinstrumente zu groß zu sein. In den neunzigminütigen Erhebungen in Studie 1 (Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten, KFT) fiel auf, dass die Konzentration und die Motivation der Schülerinnen und Schüler nach spätestens 45 Minuten der anvisierten 60 Minuten Testzeit für die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten stark nachließ. Die Bearbeitung des KFT hingegen schien die Schülerinnen und Schüler zu motivieren, hier lag die Vermutung nahe, dass den Schülerinnen und Schülern diese Art von Aufgaben weniger bekannt war und sie daher nicht so „schulisch“ erschienen wie die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten. Aus den genannten Gründen wurde für Studie 2 die Testzeit für alle Erhebungsinstrumente zusammengefasst auf 90 Minuten begrenzt, zudem wurden alle Erhebungsinstrumente gekürzt und die Erhebungen der einzelnen Instrumente nicht mehr voneinander getrennt. Für die einzelnen Erhebungsinstrumente wurde eine Zeit von maximal 20 Minuten veranschlagt.

In Ergänzung zu den Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten wurde zudem ein Kreativitätstest entwickelt. Dies hatte mehrere Gründe: Zum einen wird in der Forschung vielfach die Meinung vertreten, dass Kreativität ein unverzichtbarer Bestandteil der Leistungsdiagnostik sein sollte (z. B. König, 1986; Krampen, 1993; McComas, 1998; Osborne, Collins, Ratcliffe, Millar & Duschl, 2003; Subotnik et al., 2011). Zum anderen sollte mit kreativen Aufgaben eine zusätzliche Differenzierung zwischen Teilnehmenden verschiedener Schülerwettbewerbe erfolgen, da die Analyse der Anforderungen des Wettbewerbs Jugend forscht bzw. Schüler experimentieren gezeigt hatte, dass die Kreativität von Teilnehmenden eine große Rolle beim Erreichen guter Leistungen spielte. Hinzu kam, dass während der Erhebungen beobachtet wurde, dass Aufgaben, deren Form den Schülerinnen und Schülern aus ihrem schulischen Unterricht kaum oder wenig bekannt waren (Aufgaben des KFT, offen und kreativ ausgelegte Aufgaben), die Schülerinnen und Schüler anscheinend motivierten, möglichst gute Ergebnisse zu erzielen.

Mit einer Beschränkung auf vier Erhebungsinstrumente (Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten, Fragebogen, Kreativitätstest, KFT) à 20 Minuten sollten möglichst viele Schulen zu einer Teilnahme an den Erhebungen bewegt sowie die Schülerinnen und Schüler für die Bearbeitung der Aufgaben motiviert werden, um sie dann möglichst breit anhand verschiedener Konstrukte beschreiben zu können.

### 7.1.3 Kürzung des Testinstruments

Da der Testentwurf auf ein Drittel der ursprünglich vorgesehenen Testzeit gekürzt werden sollte, mussten dementsprechend auch die Testaufgaben auf zehn Aufgaben reduziert werden. Da abgesehen von drei Aufgaben, deren Infit- $t$ -Werte außerhalb der geforderten Grenzen lagen, alle Fit-Werte der Aufgaben im Rahmen vorgegebener Grenzen lagen, wurde bei der Auswahl der Aufgaben zunächst inhaltlich vorgegangen. Dabei wurde auch angestrebt, die Vielzahl an Themen zu reduzieren, um zu vermeiden, dass Schülerinnen und Schülern sich in viele verschiedene Themenstellungen eindenken müssen. Trotz der Eindimensionalität des zu messenden Konstrukts wurde zudem angestrebt, eine möglichst homogene Verteilung der Aufgaben über die Kompetenzbereiche und die Fächer zu erreichen, was eine ausschließlich auf die Fit-Werte oder Schwellenparameter der Aufgaben beschränkte Auswahl nicht ermöglicht hätte.

- In einem ersten Schritt wurden alle sechs Multiple-Choice-Aufgaben aus dem Test entfernt, da bei zehn verbleibenden Aufgaben das Itemformat nicht mehr variiert werden und durch die offenen Aufgaben mehr Informationen über die Schülerinnen und Schüler erhalten werden sollte.
- Im Anschluss wurden zwei der eher kreativ ausgelegten Aufgaben entfernt, da die kreativen Kompetenzen der Schülerinnen und Schüler in dem Kreativitätstest erfasst werden sollten. Eine dritte Aufgabe wurde beibehalten, aber dem Kompetenzbereich Fachwissen zugeordnet, da die Auswertungen von Studie 1 ergeben hatten, dass die Aufgabe eher fachlich als kreativ zu bewerten war. Der Fit-Wert und die Schwellenparameter dieser Aufgabe waren unauffällig.
- Sodann wurden alle Aufgaben entfernt, die ungeordnete Schwellenparameter aufwiesen (EC03, EP09 und FP12). Zwar stellte sich dieser Befund in Studie 1 als unproblematisch dar (vgl. Abschnitt 6.5.2), jedoch sollte dieser Befund in Studie 2 nicht gänzlich ignoriert werden.
- Für den Kompetenzbereich Erkenntnisgewinnung sollte jeweils eine Aufgabe pro Fach bestehen bleiben. Dabei sollten die Bereiche „Fragestellungen und Hypothesen finden“, „Versuche auswerten“ und „Fehler analysieren“ abgedeckt werden. Die Unterkategorie „Versuche planen“ wurde mit der verbliebenen kreativ ausgelegten Physik-Aufgabe abgedeckt, die somit sowohl den Kompetenzbereich Fachwissen als auch den Kompetenzbereich Erkenntnisgewinnung abdeckte. Pro Unterkategorie gab es drei potentiell auszuwählende Aufgaben (eine Aufgabe pro Fach). Das Themengebiet „Luftige Angelegenheiten“ wurde mit zwei Physik-Aufgaben komplett ausgeschlossen. Eine Aufgabe wies zudem einen Infit- $t$ -Wert außerhalb der geforderten Grenzen auf, die andere (EP09) besaß keine geordneten Schwellenparameter. Eine weitere Biologie-Aufgabe wurde ausgeschlossen, da für die Lösung dieser Aufgabe nicht biologisches Wissen erforderlich war. Für die Unterkategorie „Versuche auswerten“ blieb somit nur noch eine Physik-Aufgabe (EP11), deren Fit-Werte auch innerhalb der geforderten Grenzen lagen und deren Schwellenparameter geordnet waren. Für die Unterkategorie „Fehler analysieren“ wurde eine Biologie-Aufgabe gewählt, da eine Chemie-Aufgabe (EC03) aufgrund der nicht geordneten Schwellenparameter bereits ausgeschlossen worden war. Als einzige Chemie-Aufgabe aus dem Bereich Erkenntnisgewinnung blieb somit Aufgabe EC02, deren Fit-Werte und Schwellenparameter in Studie 1 unauffällig waren.

- Für den Kompetenzbereich Fachwissen sollten zwei Aufgaben pro Fach bestehen bleiben, ebenso sollten die Kontexte „Energie in Natur und Technik“ und „Forensische Kriminalistik“ erhalten bleiben und die Aufgaben gleichmäßig auf die beiden Kontexte verteilt werden. Für den Kontext „Energie in Natur und Technik“ standen nach Ausschluss der Multiple-Choice-Items noch sechs bzw. sieben (einschließlich der ursprünglich eher kreativ ausgelegten Physik-Aufgabe KP01) zur Verfügung. Für die Biologie wurde die Aufgabe FB11 ausgeschlossen. Zwar lag für Aufgabe FB14 der Infit- $t$ -Wert außerhalb der geforderten Grenzen, jedoch wurde die Aufgabe aus inhaltlichen Gründen beibehalten. Für die Chemie verblieb Aufgabe FC10, für die Physik Aufgabe FP07, beide mit unauffälligen Fit-Werten und Schwellenparametern. Da der Kontext „Forensische Kriminalistik“ nur drei offene Aufgaben des Kompetenzbereichs Fachwissen enthielt (FB04, FC12, FP05), blieben alle diese Aufgaben bestehen. Auch die Fit-Werte und Schwellenparameter für diese Aufgaben lagen in den geforderten Grenzen bzw. waren geordnet.

Insgesamt blieben drei Aufgaben des Kompetenzbereichs Erkenntnisgewinnung und sieben Aufgaben des Kompetenzbereichs Fachwissen erhalten. Die Physik war dabei mit drei Aufgaben des Kompetenzbereichs Fachwissen vertreten, da die ursprünglich kreativ ausgelegte Aufgabe KP01 dem Fachwissen, Kontext „Energie in Natur und Technik“, zugeordnet wurde. Allerdings nahm diese Aufgabe eine gewisse Sonderstellung zwischen Erkenntnisgewinnung und Fachwissen ein, da die Schülerinnen und Schüler hier einen Versuch planen sollten. Um den gesamten Testentwurf zu kürzen und die Lesezeit für die Schülerinnen und Schüler zu reduzieren, wurden zudem alle Einleitungstexte gestrichen. Lediglich zum Thema „Forensische Kriminalistik“ blieb ein kurzer einleitender Satz zur Erklärung, womit sich die Forensische Kriminalistik beschäftigt, bestehen.

Die zehn verbliebenen Aufgaben wurden noch einmal über das eindimensionale PCM skaliert, um die Reliabilität sowie die Fit-Werte der Aufgaben in dieser Zusammensetzung zu prüfen. Dabei ergab sich eine EAP/PV-Reliabilität von .70 und eine WLE-Reliabilität von .64. Der Abfall der Reliabilität ist hier jedoch vor allem durch die verringerte Anzahl an Items zu erklären. Die Fit-Werte aller Items lagen zwischen 0.85 und 1.10 (Mean-Squared-Infit) sowie zwischen -1.62 und 0.99 (Infit- $t$ -Wert) und somit innerhalb der geforderten Grenzen nach Wilson (2005). Alle Schwellenparameter waren geordnet.

Für den endgültigen Testentwurf wurden die Aufgaben noch einmal neu angeordnet, das Layout optimiert sowie zusammengelegte Kategorien (Aufgaben FC10, FB14, EP11, EB10, FC12, EC02) zunächst wieder getrennt, um eine bestmögliche Differenzierung zu erreichen. Auch wurde im Gegensatz zu Studie 1 darauf verzichtet, mehrere Testentwürfe mit einer unterschiedlichen Reihenfolge in den Aufgaben zu erstellen. Dies wurde einerseits aufgrund der geringen Anzahl an Aufgaben und andererseits aus praktischen Gründen (Ablauf der Erhebungen) entschieden.

## 7.2 Stichprobe

Für Studie 2 wurden die Fälle aus Studie 1 und den nachfolgenden Erhebungen zusammengefasst, da alle Erhebungsinstrumente in Studie 2 lediglich gekürzt und nur geringfügig verändert wurden. Des Weiteren konnten so die Gruppen von Teilnehmenden der IJSO-Bundesfinale sowie von Teilnehmenden an Jugend forscht bzw. Schüler experimentieren vergrößert werden.

Zudem konnte eine weitere Gruppe mit Schülerinnen und Schülern gebildet werden, die zwar an der IJSO, aber nicht an einem IJSO-Bundesfinale und nicht auch an Jugend forscht bzw. Schüler experimentieren teilgenommen hatte. Daten aus der Pilotierung wurden, wie auch schon in Studie 1, ausgeschlossen. Insgesamt wurden Daten von 1021 Schülerinnen und Schülern erhoben. Die Teilnehmenden entstammten dem gesamten Bundesgebiet ausgenommen Bremen, Hamburg, Mecklenburg-Vorpommern und dem Saarland. Inklusive der Schulen aus Studie 1 wurde in Bayern, Brandenburg, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz und Sachsen erhoben, wobei es sich um eine Gemeinschaftsschule, vier reguläre Gymnasien und vier Gymnasien des nationalen Excellence-Netzwerks von Schulen mit ausgeprägtem MINT-Profil (MINT-EC-Schulen) handelte. Zwei der teilnehmenden Schulen waren mit dem „Jugend forscht Schulpreis“ ausgezeichnet worden. Wie auch schon in Studie 1 wurden auch hier nur vollständige Datensätze, d. h. Fälle, für die ein gültiger Fragebogen sowie ein gültiger Test vorlagen, für sämtliche Analysen verwendet, wodurch sich eine Stichprobengröße von  $n = 955$  ergab. Das durchschnittliche Alter der Gesamtstichprobe betrug  $M = 14.75$ ,  $SD = 0.83$ , Altersgruppe 12 bis 18. Der Anteil der Mädchen lag bei 43.4%. Die Jungen wiesen ein durchschnittliches Alter von  $M = 14.82$ ,  $SD = 0.81$ , Altersgruppe 13 bis 18 auf, die Mädchen von  $M = 14.67$ ,  $SD = 0.85$ , Altersgruppe 12 bis 18. 47.4% der Schülerinnen und Schüler hatten schon mindestens einmal an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen.

### 7.2.1 Teilnehmende des Bundesfinales der Internationalen JuniorScience-Olympiade 2013 und 2014

Die Gruppe der Teilnehmenden der IJSO-Bundesfinale aus den Jahren 2013 und 2014 ( $n = 62$ ) wies ein durchschnittliches Alter von  $M = 14.60$ ,  $SD = 0.66$  (Altersgruppe 12 bis 15) auf, der Anteil an Mädchen betrug 35.5%. Die Teilnehmenden der IJSO-Bundesfinale von 2013 und 2014 entstammten dem gesamten Bundesgebiet ausgenommen Bremen, Hamburg, Mecklenburg-Vorpommern und dem Saarland. Sechs Teilnehmende der IJSO-Bundesfinale 2013 oder 2014 hatten auch an Jugend forscht bzw. Schüler experimentieren teilgenommen.

### 7.2.2 Wettbewerbsteilnehmende

Die Gruppe der Schülerinnen und Schüler, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb, aber nicht am Bundesfinale der IJSO 2013 oder 2014 teilgenommen hatten, bestand aus  $n = 391$  Personen, wovon 42.7% Mädchen waren. Sie waren im Durchschnitt 14.75 Jahre alt,  $SD = 0.86$ , Altersgruppe 13 bis 18 Jahre. Sieben Schülerinnen und Schüler hatten sowohl an der IJSO als auch an Jugend forscht bzw. Schüler experimentieren teilgenommen.

### Teilnehmende an Jugend forscht bzw. Schüler experimentieren

In der Stichprobe gab es 47 Schülerinnen und Schüler, die schon einmal an Jugend forscht bzw. Schüler experimentieren und nicht auch an der IJSO teilgenommen hatten. 40.4% von ihnen waren Mädchen. Die Gruppe wies ein durchschnittliches Alter von  $M = 14.87$ ,  $SD = 1.06$ , Altersgruppe 13 bis 18 auf.

### Teilnehmende an der Internationalen JuniorScienceOlympiade (ohne Teilnehmende des Bundesfinales)

In der Stichprobe gab es 66 Schülerinnen und Schüler, die an der IJSO, aber nicht am IJSO-Bundesfinale 2013 oder 2014 sowie nicht an Jugend forscht bzw. Schüler experimentieren teilgenommen hatten. 51,5% von ihnen waren Mädchen. Die Gruppe wies ein durchschnittliches Alter von  $M = 14.48$ ,  $SD = 0.66$ , Altersgruppe 13 bis 16 auf.

#### 7.2.3 Nichtteilnehmende

Die Gruppe an Schülerinnen und Schülern, die noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatte ( $n = 502$ ), wies ein durchschnittliches Alter von  $M = 14.77$ ,  $SD = 0.83$  (Altersgruppe 13-17) auf, der Anteil an Mädchen betrug 44,8%.

## 7.3 Erhebungsinstrumente

Neben den gekürzten Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten (für Beschreibungen der Aufgaben siehe Abschnitt 8.3) wurden, wie bereits in Studie 1, ein Fragebogen, ein Kreativitätstest (neu) sowie eine Subskala des KFT (Heller et al., 2000) eingesetzt. Die Zuordnung der einzelnen Erhebungsinstrumente zueinander erfolgte, wie auch schon in Studie 1, durch einen Code, den die Teilnehmenden vor Beginn der Bearbeitung eines Erhebungsinstruments selbst reproduzierbar generierten. Da in Studie 1 dennoch nicht alle Instrumente eindeutig einander zuzuordnen waren, erhielten die Schülerinnen und Schüler dieses Mal alle Instrumente gesammelt in einem Umschlag. Am Ende der Erhebung wurden alle Instrumente im verschlossenen Umschlag abgegeben. Dies hatte neben der eindeutigen Zuordnung der einzelnen Instrumente zueinander datenschutzrechtliche Vorteile, da somit gewährleistet werden konnte, dass abgesehen von der Testleitung niemand Einblicke in die ausgefüllten Erhebungsinstrumente erhielt. Da die Instrumente vor Studie 2 gekürzt bzw. neu entwickelt wurden (Kreativitätstest), werden die Instrumente im Folgenden kurz erläutert.

### 7.3.1 Fragebogen

Der Fragebogen wurde gegenüber Studie 1 substantiell auf Basis der Reliabilitäten einzelner Skalen und Forschungsfragen des IKoN-Gesamtprojektes gekürzt und überarbeitet. Für die Bearbeitungszeit des Fragebogens wurden 20 Minuten angesetzt. Die einzelnen in Studie 2 erhobenen Informationen und Konstrukte sind in Tabelle 7.1 aufgeführt, der Fragebogen ist in Anhang G einzusehen.

### 7.3.2 Kreativitätstest

Der Kreativitätstest stellte eine Adaption des *Scientific Creativity Test for Secondary School Students* von Hu und Adey (2002) dar, der vor Beginn von Studie 2 pilotiert wurde. Es wurden vier der ursprünglich sieben Aufgaben eingesetzt, die folgende Bereiche abdeckten:

- Flüssigkeit, Flexibilität und Originalität für die Verwendung eines Objekts für naturwissenschaftliche Zwecke
- Grad der Sensibilität für naturwissenschaftliche Fragestellungen

Tabelle 7.1: **In Studie 2 erhobene Informationen und Konstrukte im Fragebogen zum IKoN**

<i>Erhobene Information bzw. erhobenes Konstrukt</i>	<i>Beschreibung und Quelle</i>
Demographische Informationen	Geschlecht, Alter, Schulform, Bundesland, letzte Zeugnisnoten und Interesse an den jeweiligen Fächern (eigene Entwicklung)
Fragen zum Image von Jugend forscht und der IJSO	2 x 5 Items (eigene Entwicklung)
Allgemeine Fragen an Schülerinnen und Schüler, die schon an naturwissenschaftlichen Wettbewerben teilgenommen haben	An welchen Wettbewerben hatten Schülerinnen und Schüler schon teilgenommen, welcher war der erste Wettbewerb, wie oft hatten Schülerinnen und Schüler schon teilgenommen, Beweggründe zur Wettbewerbsteilnahme (eigene Entwicklung)
Akademisches (absolutes schulisches) Selbstkonzept	5 Items, vierstufige Likert-Skala (Schöne et al., 2002)
Naturwissenschaftliches Selbstkonzept	7 Items, vierstufige Likert-Skala (Daniels, 2008; Köller et al., 2000; Schöne et al., 2002)
Zielorientierungen im Naturwissenschaftsunterricht	16 Items, vierstufige Likert-Skala (Spinath et al., 2002; Blankenburg et al., 2015)
Interesse an Tätigkeiten, die im naturwissenschaftlichen Schulunterricht, in naturwissenschaftlichen Fördermaßnahmen oder in einem naturwissenschaftlichen Beruf ausgeübt werden	3 x 27 Items, vierstufige Likert-Skala (Dierks et al., 2014)
Eigenschaften eines Naturwissenschaftlers	8 Items, fünfstufige Likert-Skala (weiterentwickelt nach Wentorf et al., 2015)
Einschätzung, welche Tätigkeiten ein Wissenschaftler regelmäßig ausübt	29 Items, vierstufige Likert-Skala (Wentorf et al., 2015)

- Naturwissenschaftliche Problemlösefähigkeit
- Kreative experimentelle Fähigkeiten

Für den Papier-Bleistift-Test wurde eine Bearbeitungszeit von 20 Minuten veranschlagt, wobei die einzelnen Aufgaben auf fünf Minuten zeitlich begrenzt wurden. Auf den Kreativitätstest wird an dieser Stelle nicht näher eingegangen, da er für die folgenden Auswertungen nicht relevant war.

### 7.3.3 Kognitiver Fähigkeitstest

Wie auch schon in Studie 1 wurde zur Bewertung allgemeiner kognitiver Teilfähigkeiten eine Subskala des KFT nach Heller et al. (2000) eingesetzt. Wie schon gegen Ende von Studie 1 wurde auch in Studie 2 nur die nonverbale (figurale) Skala (N-Test 1, Testform A) eingesetzt. Für

die Durchführung des standardisierten Testverfahrens (Einführung und Bearbeitung) wurden 20 Minuten veranschlagt.

## 7.4 Ablauf der Erhebungen

Die Teilnahme an der Studie war freiwillig. Für die gekürzten Erhebungsinstrumente wurden im Vorfeld der Erhebungen von allen teilnehmenden Bundesländern die Nachgenehmigungen zur Durchführung der wissenschaftlichen Erhebungen eingeholt. Das Einholen von Einverständniserklärungen aller Beteiligten (je nach Bundesland Schülerinnen und Schüler, Erziehungsberechtigte, Schulleitungen) erfolgte wie auch schon in Studie 1 erläutert.

Die Rekrutierung der Teilnehmenden der IJSO-Bundesfinale 2014 erfolgte wie auch schon in Studie 1 über die Wettbewerbsleitung der IJSO. Die Erhebungen selbst fanden vor Ort während des IJSO-Bundesfinales im Oktober 2014 statt. Aus organisatorischen Gründen konnte die Bearbeitung der Erhebungsinstrumente nicht zusammen und nicht nach erbrachter Wettbewerbsleistung erfolgen. So bearbeiteten die teilnehmenden Schülerinnen und Schüler den Fragebogen am Anreisetag und die übrigen Erhebungsinstrumente am Vormittag vor der experimentellen Klausur des IJSO-Bundesfinales. Die Erhebungen selbst wurden von der Autorin durchgeführt.

Die 50 an den Erhebungen potentiell teilnehmenden Schulen wurden zwischen Juli und Oktober 2014 angeschrieben, um den Schulen einen möglichst großen Zeitrahmen für die Organisation der Erhebungen zur Verfügung stellen zu können. 18 Schulen bekundeten zunächst ihr Interesse an den Erhebungen, an sechs Gymnasien fanden die Erhebungen letztendlich statt. Da an jeder Schule mindestens ein Jahrgang komplett zur Verfügung gestellt werden konnte, nahmen erfreulicherweise weitaus mehr Schülerinnen und Schüler als in Studie 1 an den Erhebungen teil. Die Erhebungen an fünf der sechs Schulen fanden im Block im Rahmen einer Doppelstunde (90 Minuten) unter Anleitung der Autorin sowie von vorab in standardisierter Weise instruierten studentischen Hilfskräften vor Ort statt. Eine Schule führte die Erhebungen selbst nach einem vorgegebenen Testverfahren durch. Alle Erhebungen an Schulen wurden zwischen November 2014 und Januar 2015 durchgeführt.

## 7.5 Auswertung

Die Auswertungen von Studie 2 erfolgten im Wesentlichen analog zu Studie 1, daher werden sie im Folgenden nur kurz beschrieben. Für eine ausführliche Erläuterung der einzelnen Auswertungsschritte sei an dieser Stelle auf Abschnitt 6.4 verwiesen.

### 7.5.1 Beurteilerreliabilität

Während der Korrekturen der Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten mussten einige Kategorien des Bewertungssystems der Aufgaben erneut überarbeitet, d. h. vor allem ergänzt werden, da in der größeren Stichprobe Antworten auftraten, die in der Form in Studie 1 nicht aufgetreten waren. Zudem wurden zur besseren Differenzierung der Schülerantworten Null-Punkte-Antworten (Kategorie 0) dem Kategoriensystem hinzugefügt.

Daraufhin wurden die Tests aus Studie 1 noch einmal durchgesehen und vereinzelt Bewertungen dem neuen Kategoriensystem angepasst.

Aus diesem Grund wurde zur Prüfung der Reliabilität und der Objektivität des Kategoriensystems erneut eine doppelte Bewertung von 16.9% der Tests (167 aus insgesamt 991, ohne Pilotierung) vorgenommen und die Beurteilerreliabilität über die Spearman-Rangkorrelation  $\rho$  (siehe Abschnitte 4.2.2 und 5.4.3) ermittelt. Es ergab sich eine signifikante Spearman-Rangkorrelation mit einem Wert von  $\rho = .83$ ,  $p < .001$ . Nach Heller und Rosemann (1981, S. 122) ist dieser Wert als ein hoher Zusammenhang zu interpretieren.

### 7.5.2 Umgang mit fehlenden Werten in Itemantworten

Wie auch schon in Studie 1 (siehe Abschnitt 6.4.2) wurden die fehlenden Werte für die Auswertung von Studie 2 ignoriert, um verfälschte Parameterschätzungen und überschätzte Reliabilitäten zu vermeiden (vgl. Pohl et al., 2014). Eine Analyse der fehlenden Werte mit Fokus auf die Testlänge und die Aufgabenschwierigkeit wurde für Studie 2 nicht noch einmal vorgenommen.

### 7.5.3 Ausschluss von Fällen

Während der Erhebungen selbst sowie der Auswertung der Antworten auf die Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten fiel auf, dass einige Schülerinnen und Schüler die Aufgaben aller Vermutung nach nicht sorgfältig bearbeitet hatten. Daher wurde beschlossen, Fälle von den Analysen auszuschließen, die mehr als sechs fehlende Werte bei den zehn Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten aufwiesen. Somit betraf dies Schülerinnen und Schüler, die weniger als 40.0% der Aufgaben bearbeitet hatten. Dies betraf 13 von 955 Schülerinnen und Schülern, d. h. knapp 1.4%. Dies liegt für den fallweisen Ausschluss von Personen unter einer akzeptablen Quote von 5.0% (Graham et al., 2003, zit. n. Lüdtke, Robitzsch, Trautwein & Köller, 2007).

Für die Bearbeitung des KFT konnte in Studie 2 zum Teil Ähnliches beobachtet werden, zudem wiesen einige Schülerinnen und Schüler unplausible Rohwerte im Bereich von unter zehn Punkten (von maximal 25 Punkten) auf. Nach der klassenweisen Analyse der Verteilung der erreichten Rohwerte wurden diese mit den Mittelwerten und Standardabweichungen der gymnasialen Normstichprobe (Heller et al., 2000) verglichen. Von allen Analysen, in die die Ergebnisse der nonverbalen Skala des KFT eingeschlossen wurden, wurden daher die Fälle ausgeschlossen, die mehr als eine Standardabweichung unter dem Mittelwert der gymnasialen Normstichprobe lagen. Von allen vertretenen Klassenstufen wurde sich an dem insgesamt geringsten Mittelwert minus einer Standardabweichung orientiert. Tendenziell sinken die Werte für Mittelwert und Standardabweichung mit Höhe der Klassenstufe, so dass hier die Klassenstufe 10 als Referenz gewählt wurde. (Die Klassenstufen 11 und 12 waren mit insgesamt 16 Schülerinnen und Schülern vertreten, so dass diese Subgruppen als nicht repräsentativ angesehen wurden. Des Weiteren wurde durch das höhere Kriterium niemand aus dieser Subgruppe benachteiligt, da die Schülerinnen und Schüler über dem Cut-Off-Wert bzw. weit darunter lagen.) Für die Klasse 10 des Gymnasiums wurde ein Mittelwert von  $M = 19.10$ ,  $SD = 3.73$  (Heller et al., 2000, S. 82) angegeben, so dass das Ausschlusskriterium zugunsten der Teilnehmenden auf 15 Punkte festgelegt wurde. Somit wurden alle Fälle ausgeschlossen, in denen Schülerinnen und Schüler in der nonverbalen Skala des KFT weniger als 15 Punkte erreicht

hatten. Dies betraf 99 von 955 Schülerinnen und Schülern.

#### 7.5.4 Skalierung der Itemantworten

Wie bereits in Studie 1 erfolgt (siehe Abschnitt 6.4.3), wurden die Daten auch in Studie 2 über ein eindimensionales PCM skaliert. Des Weiteren wurden die Item- und Personenparameter, letztere über die WLE, geschätzt und eine Wright Map zur Einschätzung der Passung der Items zur Stichprobe erstellt.

#### 7.5.5 Testqualität

Zur Prüfung der Testqualität wurden analog zu Studie 1 die Itemfits betrachtet und die Daten auf DIF geprüft. Die Prüfung auf Mehrdimensionalität entfiel, da die Aufgaben aufgrund der geringen Anzahl nicht mehr in einzelne Dimensionen getrennt werden konnten. Die Hintergründe sind in Abschnitt 6.4.4 erläutert.

Alle Analysen wurden mit der Statistik-Software R durchgeführt. Die Skalierung der Daten über das PCM und zugehörige Analysen wurden mit dem R-Programmpaket TAM von Kiefer et al. (2015), Version 1.13-0, durchgeführt.

## 7.6 Ergebnisse

### 7.6.1 Skalierung der Itemantworten

Für die Skalierung der Itemantworten wurden Fälle ausgeschlossen, die mehr als sechs fehlende Werte bei den zehn Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten aufwiesen, was 13 von 955 (knapp 1.4%) der Schülerinnen und Schülern betraf (siehe dazu Punkt 7.5). In die IRT-Skalierung ging somit eine Stichprobe von  $n = 942$  Schülerinnen und Schülern ein. Die Skalierung erfolgte, wie auch schon in Studie 1, über ein eindimensionales PCM. Hierbei ergab sich eine als nicht zufriedenstellend anzusehende EAP/PV-Reliabilität von .54 sowie eine WLE-Reliabilität von .51.

Aufgrund dieses Ergebnisses wurden weiterführende Itemstatistiken (siehe Anhang H) betrachtet, es ergaben sich jedoch keine Auffälligkeiten. Eine Überarbeitung des Modells, wie sie auch schon durch das Zusammenlegen von Kategorien in Studie 1 erfolgte, erbrachte lediglich eine Verbesserung der EAP/PV-Reliabilität auf .57 (WLE-Reliabilität .54). Die Stichprobe wurde daher in mehrere Subgruppen (nach Geschlecht, Wettbewerbserfahrung, Alter, Klassen, Bundesland, u. a.) eingeteilt und die Subgruppen einzeln über ein PCM skaliert. Hierdurch wurde geprüft, ob es womöglich einzelne Subgruppen gab, für die das PCM nicht galt, weshalb derartige Subgruppen als Konsequenz in der Gesamtskalierung hätten ausgeschlossen werden müssen. Hierbei ergaben sich jedoch keine Auffälligkeiten.

Um zu prüfen, ob das Ausschließen bestimmter Itemgruppen einen Einfluss auf die Passung des PCM auf die Daten hatte, wurden die Aufgaben nach Kompetenzbereichen sowie nach Fächern aufgeteilt und erneut skaliert. Auch hier zeigten sich keine Auffälligkeiten und somit auch keine Verbesserung der Reliabilität.

### Schätzung der Item- und Personenparameter

Die im Folgenden berichteten Schätzungen der Item- und Personenparameter sind mit Vorbehalt zu interpretieren, da die geringe EAP/PV-Reliabilität von .54 gegen eine Passung des PCM auf die Daten sprach. Um mögliche Ursachen hierfür zu finden, wurden die Item- und Personenparameter dennoch analysiert.

**Itemparameter und Itemschwierigkeit** Die Itemparameter wurden wie auch schon für Studie 1 als Schwellenparameter über ein eindimensionales PCM geschätzt. Bei allen Aufgaben lagen geordnete Schwellenparameter vor, d. h. die Schwierigkeit der Items nahm mit steigender Kategorie zu. Die Itemschwierigkeiten variierten zwischen -0.107 für das leichteste Item (EB10) und 1.446 (FC12) für das schwierigste Item. Die Itemparameter und die Itemschwierigkeiten sind in Anhang H einzusehen.

**Personenparameter** Die Schätzung der Personenparameter ( $\theta$ ) erfolgte über die WLE. Die Personenparameter variierten zwischen Werten von -2.758 Logits bis 2.560 Logits für die fähigste Person.

### Wright Map

Die Passung der Items auf die Stichprobe ist in Abbildung 7.1 dargestellt. Es ist zu erkennen, dass das untere Fähigkeitsspektrum der Schülerinnen und Schüler kaum abgedeckt werden konnte, im Bereich unter -1 Logit waren lediglich drei Kategorien angesiedelt. Im oberen Fähigkeitsspektrum lagen hingegen relativ viele Kategorien, obwohl wenige Schülerinnen und Schüler eine Personenfähigkeit in dem Bereich aufwiesen.

## 7.6.2 Testqualität

### Itemfit

Unter Verwendung des eindimensionalen PCM wurden die Aufgaben auch für Studie 2 hinsichtlich ihrer Fit-Statistiken geprüft. Nach Wilson (2005) können Werte für den Mean-Squared-Infit zwischen 0.75 und 1.33 als akzeptabel angesehen werden, für den Infit- $t$ -Wert geben Bond und Fox (2007) Werte zwischen -2 und 2 an. Die Werte des Mean-Squared-Infit der Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten lagen zwischen 0.96 und 1.09, die Infit- $t$ -Werte zwischen -0.94 und 1.62 und somit innerhalb der geforderten Grenzen. Die Fit-Werte der Aufgaben sind in Anhang H einzusehen.

### Differential Item Functioning

Es wurden DIF-Analysen für die Merkmale Alter, Klasse, Geschlecht, Teilnahme am IJSO-Bundesfinale und Wettbewerbsgruppe (Teilnehmende der IJSO-Bundesfinale sowie Schülerinnen und Schüler, die mindestens einmal bzw. noch nie an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten) durchgeführt. Die Ergebnisse sind in Tabelle 7.2 dargestellt.

Für das Merkmal Alter fand sich ein signifikanter Unterschied zwischen dem Modell das nur den Haupteffekt modelliert und dem Model welches DIF berücksichtigt. Die jeweils niedrigeren Werte für den AIC bzw. für den BIC sprachen dabei für eine bessere Passung des

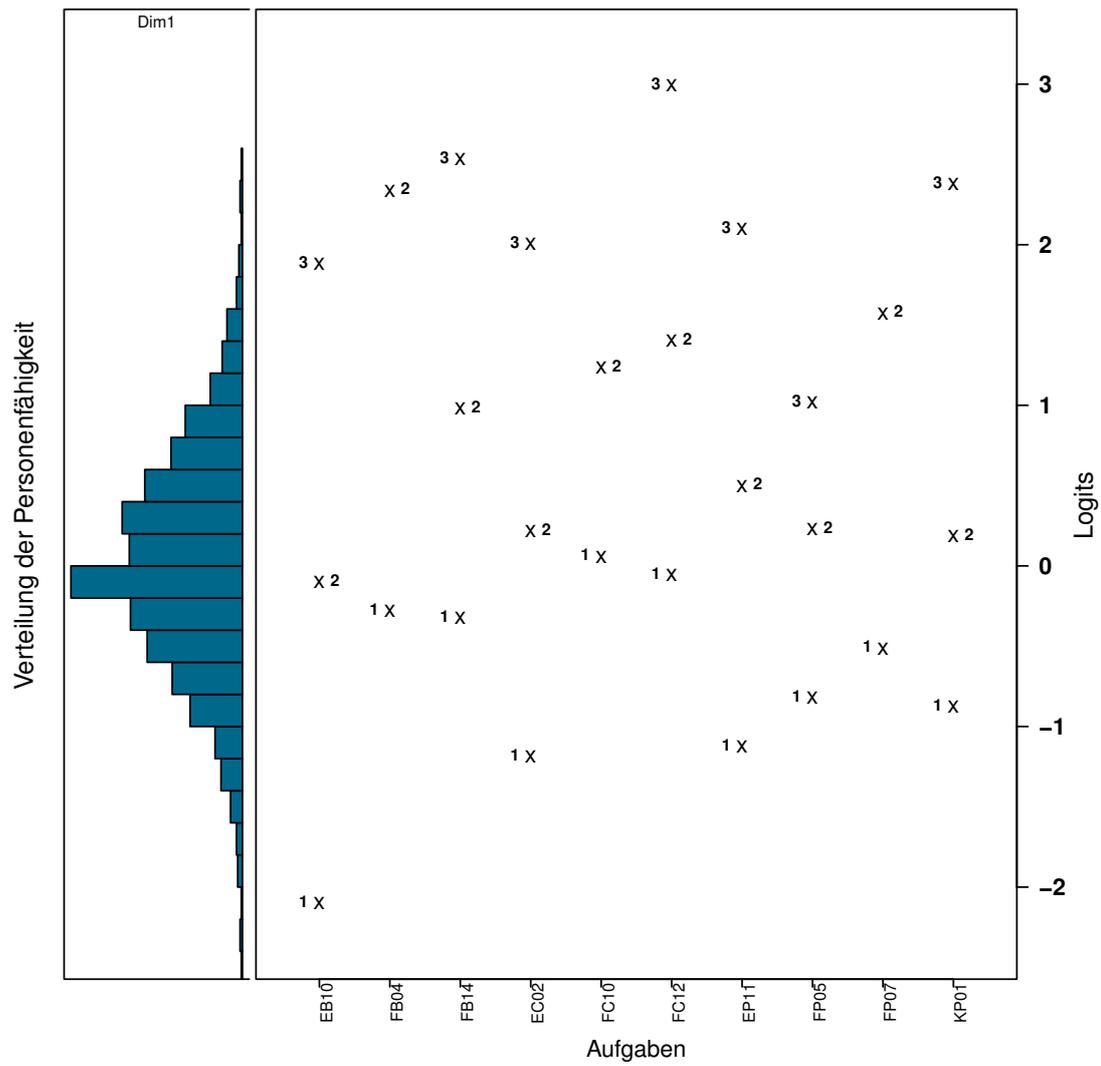


Abbildung 7.1: **Wright Map** Die Zahlen 1, 2, 3 bezeichnen die Kategorien der einzelnen Aufgaben.

Tabelle 7.2: Vergleich zwischen Modellen mit und ohne DIF

<i>Merkmal</i>	<i>Modell</i>	<i>Devianz</i>	<i>Anzahl der Parameter</i>	<i>AIC</i>	<i>BIC</i>	<i>pLRT</i>
Alter	Haupteffekt	16883.23	35	16953.23	17122.91	< .05
	DIF	16798.51	98	16994.51	17469.61	
Klassen	Haupteffekt	16835.84	32	16899.84	17054.97	n. s.
	DIF	16855.96	68	16991.96	17321.62	
Geschlecht	Haupteffekt	16878.65	30	16938.65	17084.09	< .001
	DIF	16798.84	48	16894.84	17127.54	
Teilnahme am IJSO-Bundesfinale	Haupteffekt	16792.00	29	16850.00	16990.59	< .001
	DIF	16757.46	38	16833.46	17017.68	
Wettbewerbs Erfahrung	Haupteffekt	16765.30	30	16825.30	16970.74	< .001
	DIF	16702.23	48	16798.23	17030.94	

*Anmerkungen.*  $p_{LRT}$  =  $p$ -Wert für den Likelihood-Ratio-Test. n. s. = nicht signifikant.

Modells ohne DIF.

Für das Merkmal Klassen fand sich kein signifikanter Unterschied in der Passung des Modells ohne bzw. mit DIF auf die Daten, die niedrigeren Werte für den AIC bzw. für den BIC sprachen jedoch auch hier für die bessere Passung des Modells, das nur den Haupteffekt modelliert.

Für die Merkmale Geschlecht, Teilnahme am IJSO-Bundesfinale und Wettbewerbserfahrung fand sich ein signifikanter Unterschied zwischen den jeweiligen Modellen ohne DIF und dem Modell, das DIF berücksichtigt. Betrachtet man bei den drei Merkmalen die Werte für AIC und BIC fällt auf, dass der Wert des AIC für das jeweilige Modell mit DIF niedriger ausfällt als der Wert für das Modell, welches nur den Haupteffekt modelliert. Für den Wert des BIC verhält es sich jedoch andersherum: Der Wert für das jeweilige Modell ohne DIF fällt niedriger aus als der Wert für das Modell, das DIF berücksichtigt. Nach D. H. Rost et al. (2004, S. 344) ist bei kleinen Itemanzahlen mit großen Patternhäufigkeiten der AIC und bei großen Itemanzahlen mit geringen Patternhäufigkeiten der BIC vorzuziehen, weshalb hier der AIC in der Interpretation des Modellvergleichs stärker gewichtet wurde. Zudem berücksichtigt der BIC im Vergleich zum AIC ergänzend die Anzahl der zu schätzenden Modellparameter sowie die Stichprobengröße, so dass der Wert für den BIC umso größer wird, je größer die Anzahl der zu schätzenden Parameter und je größer die Stichprobe ist (Eid, Gollwitzer & Schmitt, 2013), was in diesem Fall wiederum für die Bevorzugung des AIC sprach. Somit passten für die Merkmale Geschlecht, Teilnahme am IJSO-Bundesfinale und Wettbewerbserfahrung jeweils die Modelle besser auf die Daten, die DIF berücksichtigen. Aus diesem Grund wurde zur detaillierten Analyse möglicher von DIF betroffener Aufgaben für jedes Item die Differenz der für die unterschiedlichen Gruppen geschätzten Itemschwierigkeiten betrachtet und je nach ihrer Größe in drei Gruppen klassifiziert (siehe dazu Abschnitt 6.4.4). Die Ergebnisse sind in Tabelle 7.3 zusammengefasst.

Aufgrund des hohen Anteils an Kategorie-B- und Kategorie-C-Aufgaben für die Merkmale Geschlecht, Teilnahme am IJSO-Bundesfinale und Wettbewerbserfahrung wurden die Aufga-

Tabelle 7.3: **Analyse möglicher von DIF betroffener Aufgaben** Angegeben ist die Anzahl der von DIF betroffenen Aufgaben von insgesamt zehn Aufgaben.

<i>Kategorie</i>	<i>Geschlecht</i>	<i>Teilnahme am IJSO-Bundesfinale</i>	<i>Wettbewerbs- erfahrung</i>
<b>A</b> Nicht nennenswert von DIF betroffene Aufgaben	7	1	0
<b>B</b> Mäßig von DIF betroffene Aufgabe	3	1	1
<b>C</b> Stark von DIF betroffene Aufgaben	0	8	9

ben noch einmal differenzierter analysiert, d. h. es wurde geprüft, für welche Aufgaben und für welche Subgruppen DIF auftrat.

Für das Merkmal Geschlecht waren die drei Physik-Aufgaben des Kompetenzbereichs Fachwissen mäßig von DIF betroffen. Der Vorteil lag hier zugunsten der Jungen, für die die Itemschwierigkeit der von DIF betroffenen Aufgaben niedriger lag als die Itemschwierigkeit der Mädchen.

Für das Merkmal Teilnahme am IJSO-Bundesfinale waren bis auf zwei Items (EC02, EP11) alle Aufgaben stark von DIF betroffen. Bei vier dieser Aufgaben lag die Differenz in der Itemschwierigkeit bei mehr als 1 Logit, eine Aufgabe wies sogar eine Differenz in den Itemschwierigkeiten von 1.538 Logits auf (FP07).

Für das Merkmal Wettbewerbserfahrung fanden sich Unterschiede im Ausmaß des DIF zwischen den verschiedenen Subgruppen. Die Ergebnisse spiegeln prinzipiell die Ergebnisse der DIF-Analyse für das Merkmal Teilnahme am IJSO-Bundesfinale wider:

- Bei sechs der zehn Aufgaben (KP01, FB04, FB14, FC12, FP05, FP07) lag DIF der Kategorie C zwischen den IJSO-Bundesfinalisten und den Schülerinnen und Schülern, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten (Wettbewerbsteilnehmende), sowie zwischen den Teilnehmenden der IJSO-Bundesfinale und Schülerinnen und Schülern, die noch nie an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten (Nichtteilnehmende), vor. Bei all diesen Aufgaben lag mindestens zwischen zwei Subgruppen eine Differenz in den Itemschwierigkeiten größer als 1 Logit vor.
- Bei drei der zehn Aufgaben (EB10, EP11, FC10) lag DIF der Kategorie C zwischen den Teilnehmenden der IJSO-Bundesfinale und den Nichtteilnehmenden vor. Zwischen Teilnehmenden der IJSO-Bundesfinale und Wettbewerbsteilnehmenden lag bei diesen Aufgaben DIF der Kategorie A (EP11) bzw. Kategorie B (EB10, FC10) vor.
- Zwischen den Wettbewerbsteilnehmenden und den Nichtteilnehmenden lag bei Aufgabe FB14 DIF der Kategorie B vor, ansonsten ausschließlich der Kategorie A.
- Die Aufgabe EC02 wies DIF der Kategorie B zwischen den Teilnehmenden der IJSO-Bundesfinale und den Wettbewerbsteilnehmenden bzw. den Nichtteilnehmenden auf,

zwischen den Wettbewerbsteilnehmenden und den Nichtteilnehmenden DIF der Kategorie A.

Die aus den DIF-Analysen errechneten Itemparameter, die der detaillierten DIF-Analyse der Aufgaben zugrunde lagen, sind in Anhang I aufgeführt.

### Testziele

Aufgrund der Ergebnisse aus der Skalierung der Itemantworten und der Prüfung auf DIF musste davon ausgegangen werden, dass die Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten mit dem in Studie 2 eingesetzten, an die Bedürfnisse der Wettbewerbsforschung angepasstem Testinstrument nicht gelang. Aus diesem Grund wurde von weiteren statistischen Analysen abgesehen, da kaum aussagekräftige Ergebnisse hätten generiert werden können.

## 7.7 Interpretation und Diskussion der Ergebnisse

Die Auswertung der Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten, und somit auch die Schätzung der Item- und Personenparameter, erfolgte, wie auch schon in Studie 1, über das eindimensionale PCM. Die Skalierung der Daten ergab jedoch eine nicht zufriedenstellende EAP/PV-Reliabilität von .54 und eine WLE-Reliabilität von .51 (bzw. nach Überarbeitung des Modells eine EAP/PV-Reliabilität von .57 und eine WLE-Reliabilität von .54), was auf eine nicht optimale Passung des PCM auf die Daten und somit auf ein nicht valides Testinstrument hindeutete. Mögliche Gründe werden im Folgenden erläutert.

Betrachtet man die Verteilung der Schwellenparameter fällt auf, dass lediglich zehn von 27 Kategorien in den Bereich zwischen -2 und 0 Logits fielen. In Studie 1 traf dies auf 28 von 74 Kategorien zu, was jeweils einem Anteil von ca. einem Drittel der Kategorien entsprach. Allerdings wurden in Studie 1 drei Mal so viele Aufgaben eingesetzt, so dass das untere Fähigkeitsspektrum der Schülerinnen und Schüler im Vergleich zu Studie 2 trotz recht schwieriger Items verhältnismäßig gut erfasst werden konnte. Bei Betrachtung der Itemfits ließen sich keine Auffälligkeiten erkennen, die Fit-Werte der einzelnen Aufgaben lagen alle im Bereich der geforderten Grenzen, so dass hier das Ausschließen einzelner Aufgaben keine Verbesserung der Reliabilität bewirkt hätte. Auch in den weiterführenden Itemstatistiken, siehe Anhang H, fanden sich keine Hinweise auf die nicht optimale Modellpassung. Aus diesem Grund wurden für alle Aufgaben die *Expected Score Curves* analysiert. Lediglich die Aufgaben EB10 und EP11 zeigten Auffälligkeiten, die darauf hindeuteten, dass in bestimmten Bereichen mit steigender Personenfähigkeit kein höherer Punktwert erreicht wurde, siehe Abbildung 7.2. Ein Ausschluss dieser Items erbrachte jedoch keine Verbesserung der Reliabilität. Insgesamt konnte die Betrachtung der *Expected Score Curves* demnach keine Hinweise auf die nicht optimale Passung des PCM auf die Daten geben.

Betrachtet man die Itemschwierigkeiten auf Ebene der Aufgaben, ist festzustellen, dass die leichteste Aufgabe (EB10) mit einer Itemschwierigkeit von -0.107 Logits immer noch verhältnismäßig schwierig war. Die Wright Map, siehe Abbildung 7.1, zeigte, dass nur wenige Kategorien das untere und relativ viele Kategorien das obere Fähigkeitsspektrum der Schülerinnen und Schüler abdeckten, obwohl nur wenige Schülerinnen und Schüler eine Personenfähigkeit in den Randbereichen aufwiesen. So liegt die Vermutung nahe, dass die verbliebenen Aufgaben

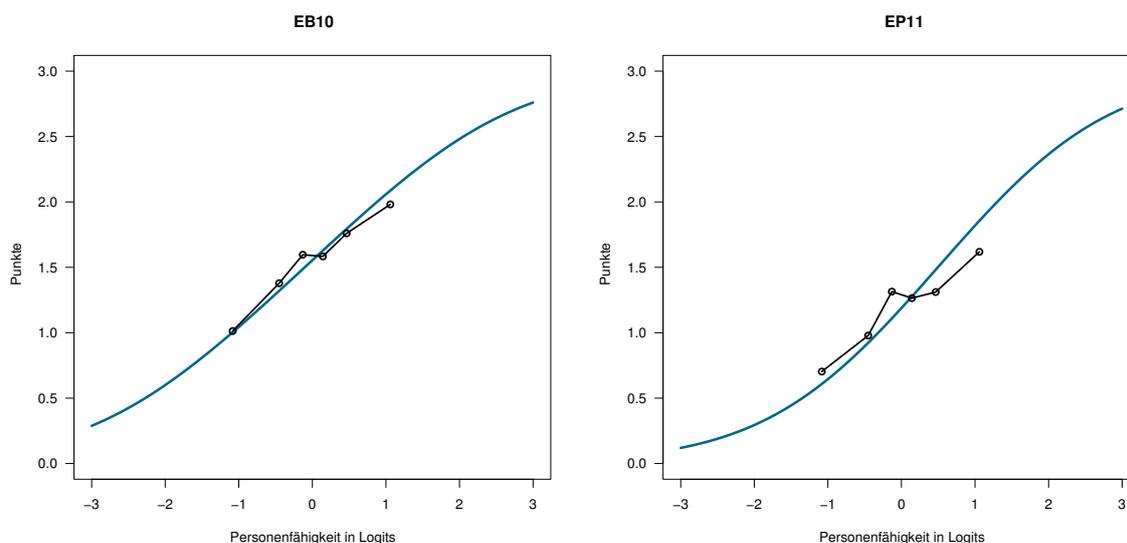


Abbildung 7.2: *Expected Score Curves* für Aufgaben EB10 und EP11

bzw. Kategorien im Mittelfeld der Personenfähigkeiten nicht ausreichend differenzieren konnten. (Zum Vergleich: Die Itemschwierigkeiten in Studie 1 variierten zwischen -2.112 und 1.884.)

Die Prüfung auf DIF lieferte weitere Hinweise auf mögliche Ursachen der nicht optimalen Passung des PCM auf die Daten, da die Modelle, die DIF für die Merkmale Geschlecht, Teilnahme am Bundesfinale und Wettbewerbserfahrung simulierten, signifikant besser auf die Daten passten. Aus diesem Grund wurden die Aufgaben für die einzelnen DIF-Analysen auf inhaltlicher Ebene betrachtet:

- Für das Merkmal Geschlecht gab es einen Vorteil zugunsten der Jungen bei der Bearbeitung der drei Physik-Aufgaben des Kompetenzbereichs Fachwissen. Dies lässt sich einerseits auf das höhere Fach-, Berufs- und Freizeitinteresse in Physik von Jungen im Vergleich zu Mädchen (z. B. Hoffmann, Häußler & Lehrke, 1998) zurückführen, was wiederum einen Einfluss auf die Leistung gehabt haben könnte (vgl. Abschnitt 2.2.3). Andererseits lieferte auch die inhaltliche Betrachtung der einzelnen Aufgaben eine Erklärung für das Vorliegen von DIF zugunsten der Jungen, da die Aufgaben mit den Themen Optik, Kinetik und Technik vermutlich vor allem Jungen ansprachen (vgl. Holstermann & Bögeholz, 2007).
- Für die Merkmale Teilnahme am Bundesfinale und Wettbewerbserfahrung lieferte die inhaltliche Betrachtung der Aufgaben kaum Hinweise auf die Ursachen des Vorliegens von DIF, da hier bei fast allen Aufgaben DIF der Kategorie C vorlag. Nur mäßig von DIF betroffen (Kategorie B) war sowohl für das Merkmal Teilnahme am Bundesfinale als auch für das Merkmal Wettbewerbserfahrung die Chemie-Aufgabe EC02 des Kompetenzbereichs Erkenntnisgewinnung, bei der es um die Formulierung einer Forschungsfrage ging. Für das Merkmal Teilnahme am Bundesfinale war die Physik-Aufgabe EP11 des Kompetenzbereichs Erkenntnisgewinnung nicht nennenswert von DIF betroffen (Kategorie A). In dieser Aufgabe ging es um die Auswertung eines Experiments zur „goldenen

Regel der Mechanik“ und der Formulierung einer allgemeingültigen Regel aus den Ergebnissen des Experiments. Für das Merkmal Wettbewerbserfahrung war diese Aufgabe nur zwischen Teilnehmenden der IJSO-Bundesfinale und Wettbewerbsteilnehmenden sowie zwischen Wettbewerbsteilnehmenden und Nichtteilnehmenden nicht nennenswert von DIF betroffen, zwischen Teilnehmenden der IJSO-Bundesfinale und Nichtteilnehmenden lag hier DIF der Kategorie C vor.

Zur Erklärung des Vorliegens von substanziellen DIF für die Merkmale Teilnahme am Bundesfinale und Wettbewerbserfahrung mussten somit andere Faktoren in Betracht gezogen werden. Anzuführen sind hier für das Merkmal Teilnahme am Bundesfinale vor allem die Erfahrung der Teilnehmenden der IJSO-Bundesfinale in der Bearbeitung von Testaufgaben, im Besonderen des Kompetenzbereichs Fachwissen. Damit einher geht möglicherweise auch ein höheres Fachwissen, wie die Ergebnisse aus Studie 1 bereits andeuteten. Auch für die Erklärung des Vorliegens von DIF für das Merkmal Wettbewerbserfahrung können diese Gründe angeführt werden, da zwischen den Wettbewerbsteilnehmenden und den Nichtteilnehmenden bis auf eine Ausnahme lediglich DIF der Kategorie A vorlag. Die Gründe für die Bevorteilung der Teilnehmenden der IJSO-Bundesfinale sind somit vermutlich in spezifischen Charakteristika, wie z. B. einem höheren Fachwissen oder der Vertrautheit mit der Testsituation, zu sehen.

Unabhängig der spezifischen Gründe für das Vorliegen von DIF bedeutete dies jedoch, dass die zehn eingesetzten Aufgaben nicht fair die naturwissenschaftlichen Kenntnisse und Fähigkeiten von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern messen konnten. Mögliche Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten zwischen den entsprechenden Gruppen hätten somit nicht durch Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten bzw. in der Personenfähigkeit erklärt werden können. Dies führt wiederum zu der Annahme, dass die Aufgaben zumindest nicht ausschließlich die naturwissenschaftlichen Kenntnisse und Fähigkeiten erfassten, sondern ein weiteres Merkmal, welches die Unterschiede zwischen den entsprechenden Gruppen zu erklären vermochte (vgl. Wilbert & Linnemann, 2011). Dies wird in der Regel als Verstoß gegen die Konstruktvalidität eines Erhebungsinstruments gewertet (z. B. Jurecka, 2010; Wilbert & Linnemann, 2011). Allerdings existieren auch andere Interpretationen des Vorliegens von DIF, in denen davon ausgegangen wird, dass DIF die Stärken und Schwächen in der Performanz von Subgruppen repräsentieren kann, die z. B. durch unterschiedliche Lerngelegenheiten entstanden sein können (vgl. Scheuneman & Gerritz, 1990; Jurecka, 2010; Klieme & Baumert, 2001). In diesem Sinne kann DIF auch als diagnostisches Instrument angesehen werden, das zur Analyse differenzieller Stärken und Schwächen verwendet werden kann (Jurecka, 2010; Klieme & Baumert, 2001), was hier auf die Analyse spezifischer Charakteristika übertragen werden könnte.

Eine weitere Ursache für die geringe Reliabilität kann in der Auswahl der Aufgaben für das reduzierte Testinstrument gesehen werden. So wurden die Aufgaben hauptsächlich nach inhaltlichen Kriterien gewählt (siehe Abschnitt 7.1.3), was zu einer Heterogenität des Testinstruments führte. Für die externe Validität der Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten war dies zwar von Vorteil, jedoch führt dies in der Regel zu einer nicht optimalen Modellpassung und somit zu einer geringen Reliabilität, was als Reliabilitäts-Validitäts-Dilemma bezeichnet wird (vgl. Lienert & Raatz, 1998; D. H. Rost et al., 2004). Die Auswahl von Aufgaben mit idealen Itemfits und mittleren Schwierigkeiten hätte somit

vermutlich zu einer akzeptablen Reliabilität geführt, allerdings stellt sich die Frage, ob ein derart zusammengestelltes Testinstrument Unterschiede zwischen den relevanten Subgruppen hätte aufdecken können. Dies wäre in einer weiterführenden Studie zu prüfen.

## 7.8 Zusammenfassung und Fazit

In Studie 2 wurde ein Testinstrument eingesetzt, das gegenüber Studie 1, siehe Kapitel 6, auf ein Drittel reduziert und somit den Bedürfnissen der Wettbewerbsforschung angepasst wurde. An den Erhebungen nahmen Teilnehmende des IJSO-Bundesfinales 2014, Teilnehmende von Jugend forscht bzw. Schüler experimentieren, die aus Schulen stammten, die mit dem „Jugend forscht Schulpreis“ ausgezeichnet wurden, sowie Schülerinnen und Schüler teil, die zumindest einmal oder auch noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten. Für die Auswertungen wurden die Teilnehmenden aus Studie 1 und Studie 2 zusammengefasst.

Das gekürzte Testinstrument zur Erfassung der naturwissenschaftlichen Kenntnisse und Fähigkeiten von Schülerinnen und Schülern musste aufgrund der nicht optimalen Passung des PCM auf die Daten, was an der niedrigen EAP/PV-Reliabilität von .54 zu erkennen war, als nicht reliabel angesehen werden. Mehrere Aspekte können zur Erklärung der niedrigen Reliabilität herangezogen werden. Die verbliebenen Aufgaben wiesen eine vergleichsweise hohe Schwierigkeit auf, was in Studie 1 noch durch die Vielzahl der Aufgaben kompensiert werden konnte. Die Wright Map lieferte Hinweise darauf, dass nicht genug Aufgaben bzw. Kategorien vorhanden waren, um im breiten Mittelfeld der Personenfähigkeiten differenzieren zu können. Zudem lag in Studie 2 für die Merkmale Geschlecht, Teilnahme am IJSO-Bundesfinale und Wettbewerbserfahrung DIF vor, im Fall der Merkmale Teilnahme am IJSO-Bundesfinale und Wettbewerbserfahrung in starker Ausprägung und immer zugunsten der Teilnehmenden der IJSO-Bundesfinale. Testfairness war somit im gekürzten Instrument nicht gegeben. Das Vorliegen von DIF kann somit einerseits als Verstoß gegen die Konstruktvalidität des Testinstruments gewertet werden, andererseits kann DIF auch als diagnostisches Instrument zur Analyse differenzieller Stärken und Schwächen einzelner Gruppen verwendet werden (Jurecka, 2010; Klieme & Baumert, 2001), was für die Merkmale Teilnahme am IJSO-Bundesfinale und Wettbewerbserfahrung den zu Beginn formulierten Testzielen entsprach. Ein weiterer Grund für die niedrige Reliabilität kann darin gesehen werden, dass die Auswahl der Aufgaben für das reduzierte Erhebungsinstrument im Wesentlichen nach inhaltlichen Kriterien erfolgte, wodurch ein recht heterogen zusammengestelltes Testinstrument entstand. Da jedoch vor allem homogene Items zu einer guten Modellpassung und hohen Reliabilitäten führen, ist hier ein weiterer Grund für die nicht optimale Modellpassung der Daten auf das PCM zu sehen. Einen möglichen Lösungsansatz hierfür stellen formative Messmodelle dar, die eine andere Herangehensweise als hier genutzte reflektive Messmodelle verwenden (siehe Abschnitt 9.2).

Die Forschungsfrage, ob die Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten mit einem an die Bedürfnisse der Wettbewerbsforschung angepasstem Testinstrument gelingt, ist somit für Studie 2 negativ zu beantworten.

Sämtliche Qualitäts- und Gütekriterien der Aufgaben erwiesen sich jedoch als zufriedenstellend. Das Vorliegen von DIF gab zudem Hinweise auf existierende spezifische Charakteristika

der untersuchten Subgruppen, v. a. für die Teilnehmenden der IJSO-Bundesfinale. So wurden in der folgenden Studie 3, siehe Kapitel 8, qualitative Methoden angewendet, um mögliche Unterschiede in den Antworten von Teilnehmenden unterschiedlicher naturwissenschaftlicher Schülerwettbewerbe, auch im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern, zu analysieren.

## Kapitel 8

# Varianzanalytischer Vergleich und qualitative Analyse der Verteilungen der Antwortkategorien für einzelne Aufgaben und Subgruppen (Studie 3)

Nach der Betrachtung quantitativer Merkmale der Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten in Studie 1 und Studie 2 wurden für Studie 3 qualitative Merkmale der Aufgaben und ihrer Antwortkategorien aus dem Testinstrument aus Studie 2 näher betrachtet und für verschiedene Subgruppen analysiert. Von Interesse waren hier mögliche Muster im Antwortverhalten bzw. der Verteilung der Häufigkeiten über die Antwortkategorien der unterschiedlichen Subgruppen, aus denen mögliche Charakteristika und Trends der einzelnen Gruppen abgeleitet werden sollten. Das Vorliegen von DIF für das in Studie 2 eingesetzte reduzierte Testinstrument, im Besonderen zugunsten der Teilnehmenden der IJSO-Bundesfinale, gab zudem Hinweise auf existierende spezifische Charakteristika einzelner Gruppen.

Für die qualitative Analyse, die vorwiegend explorativen Charakter hatte, wurden verschiedene Schritte unternommen. Zunächst wurden die Mittelwerte für die einzelnen Aufgaben verschiedener Gruppen varianzanalytisch ausgewertet um zu prüfen, ob es generell Unterschiede zwischen einzelnen Subgruppen gab. Anschließend wurde die Verteilung der Häufigkeiten über die Kategorien innerhalb der verschiedenen Gruppen untersucht und mit den anderen Gruppen verglichen. Dies diente dazu, möglicherweise spezifische Muster oder auch Auffälligkeiten in den Verteilungen der einzelnen Gruppen zu identifizieren. Anschließend wurde mithilfe des Kategoriensystems nach inhaltlichen Aspekten zur Erklärung der spezifischen Muster bzw. Auffälligkeiten gesucht. In einem letzten Schritt wurden Eigenschaften und Trends der einzelnen untersuchten Subgruppen aus den Ergebnissen abgeleitet. Im Zentrum stand die Beantwortung der folgenden Forschungsfrage:

Zeigen sich qualitative Unterschiede in den Antworten von Teilnehmenden unterschiedlicher naturwissenschaftlicher Schülerwettbewerbe, auch im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern?

## 8.1 Varianzanalytischer Vergleich der Mittelwerte

Die varianzanalytischen Vergleiche der in den jeweiligen Aufgaben erreichten Mittelwerte ist in Tabelle 8.1 dargestellt. Verglichen wurden die drei Gruppen der Teilnehmenden der IJSO-Bundesfinale mit Schülerinnen und Schülern, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten (Wettbewerbsteilnehmende), sowie mit Schülerinnen und Schülern, die noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten (Nichtteilnehmende). Dies erfolgte aufgrund der in Studie 1, siehe Kapitel 6, gefundenen Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten zwischen diesen Gruppen. Des Weiteren wurden Teilnehmende der IJSO mit Teilnehmenden der Wettbewerbe Jugend forscht bzw. Schüler experimentieren (im Folgenden: Jufo-Teilnehmende) verglichen, was in Einklang mit den ursprünglich intendierten Testzielen stand, auch Teilnehmende verschiedener naturwissenschaftlicher Schülerwettbewerbe miteinander zu vergleichen. Die Teilnehmenden der IJSO-Bundesfinale wurden hierbei ausgeklammert, da diese Gruppe den Vergleich zwischen Teilnehmenden verschiedener Schülerwettbewerbe verzerrt hätte, weil keine Teilnehmenden einer entsprechenden Wettbewerbsstufe in Jugend forscht bzw. Schüler experimentieren befragt wurden.

Der varianzanalytische Vergleich der in den Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten erreichten Mittelwerte zeigte, wie auch schon anhand der DIF-Analysen zu vermuten war (siehe Abschnitt 7.6.2), dass die Teilnehmenden der IJSO-Bundesfinale bei neun von zehn Aufgaben signifikant besser abschnitten als Wettbewerbsteilnehmende und Nichtteilnehmende. Darüber hinaus schnitten die Wettbewerbsteilnehmenden bei sieben von zehn Aufgaben signifikant besser ab als die Nichtteilnehmenden. Die Unterschiede in den Mittelwerten zwischen IJSO-Teilnehmenden sowie Jufo-Teilnehmenden waren in nur drei Fällen signifikant. So schnitten die IJSO-Teilnehmenden in zwei Aufgaben signifikant besser ab, wohingegen die Jufo-Teilnehmenden in einer Aufgabe signifikant besser abschnitten.

## 8.2 Vergleich der nach Gruppen getrennten Verteilung der Kategorien

Im Folgenden werden die nach Gruppen getrennten Verteilungen der Kategorien von 0 bis 2 bzw. 0 bis 3 für die einzelnen Aufgaben miteinander verglichen. Die Verteilungen wurden nach Wettbewerbserfahrung (Teilnehmende der IJSO-Bundesfinale, Wettbewerbsteilnehmende, Nichtteilnehmende) sowie nach Wettbewerb (IJSO ohne Teilnehmende eines Bundesfinales, Jugend forscht bzw. Schüler experimentieren) verglichen. Berücksichtigt wurden hier die absoluten Häufigkeiten, d. h. die entsprechenden Prozentangaben beziehen sich auf die gesamte Stichprobe und nicht nur auf gültige Werte, weshalb die Summe der Häufigkeiten nicht immer 100.0% ergeben muss. Die Verteilungen sind in den Abbildungen 8.1 sowie 8.1 zu sehen.

### 8.2.1 Vergleich der Wettbewerbsgruppen

Aufgrund der Ergebnisse des varianzanalytischen Vergleichs der in den jeweiligen Aufgaben erreichten Mittelwerte hätte man im Vergleich der Wettbewerbsgruppen idealtypisch erwartet, dass die Häufigkeit der jeweiligen Kategorie für die Teilnehmenden der IJSO-Bundesfinale im Verhältnis zu den beiden anderen Gruppen mit steigender Kategorie zunahm. Ähnliches

Tabelle 8.1: **Varianzanalytischer Vergleich der in den jeweiligen Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten erreichten Mittelwerte**

<i>Aufgabe (Punkte)</i>	<i>Gruppe</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Sig.</i>	<i>d</i>
EB10 (3)	Teilnehmende IJSO-Bundesfinale	61	1.87	0.64	.005 **	0.43
	Wettbewerbsteilnehmende	383	1.55	0.72		
	Teilnehmende IJSO-Bundesfinale	61	1.87	0.64	.002 **	0.45
	Nichtteilnehmende	488	1.52	0.80		
	Wettbewerbsteilnehmende	383	1.55	0.72	n. s.	
	Nichtteilnehmende	488	1.52	0.80		
EC02 (3)	IJSO-Teilnehmende	59	1.53	0.73	n. s.	
	Jufo-Teilnehmende	47	1.66	0.70		
	Teilnehmende IJSO-Bundesfinale	57	1.61	0.77	.022 *	0.39
	Wettbewerbsteilnehmende	357	1.29	0.82		
	Teilnehmende IJSO-Bundesfinale	57	1.61	0.77	.026 *	0.33
	Nichtteilnehmende	431	1.32	0.88		
Wettbewerbsteilnehmende	357	1.29	0.82	n. s.		
Nichtteilnehmende	431	1.32	0.88			
EP11 (3)	IJSO-Teilnehmende	53	1.23	0.82	n. s.	
	Jufo-Teilnehmende	45	1.20	0.81		
	Teilnehmende IJSO-Bundesfinale	36	1.39	0.84	n. s.	
	Wettbewerbsteilnehmende	191	1.35	0.83		
	Teilnehmende IJSO-Bundesfinale	36	1.39	0.84	n. s.	
	Nichtteilnehmende	253	1.08	0.81		
Wettbewerbsteilnehmende	191	1.35	0.83	.002 **	0.33	
Nichtteilnehmende	253	1.08	0.81			
FB04 (2)	IJSO-Teilnehmende	23	1.70	0.63	.000 ***	1.49
	Jufo-Teilnehmende	25	0.72	0.68		
	Teilnehmende IJSO-Bundesfinale	50	1.04	0.64	.000 ***	0.62
	Wettbewerbsteilnehmende	350	0.66	0.61		
	Teilnehmende IJSO-Bundesfinale	50	1.04	0.64	.000 ***	0.81
	Nichtteilnehmende	430	0.57	0.57		
Wettbewerbsteilnehmende	350	0.66	0.61	.039 *	0.15	
Nichtteilnehmende	430	0.57	0.57			
FB14 (3)	IJSO-Teilnehmende	50	0.58	0.54	n. s.	
	Jufo-Teilnehmende	44	0.64	0.61		
	Teilnehmende IJSO-Bundesfinale	45	1.71	0.87	.000 ***	0.96
	Wettbewerbsteilnehmende	321	0.94	0.79		
	Teilnehmende IJSO-Bundesfinale	45	1.71	0.87	.000 ***	1.40
	Nichtteilnehmende	360	0.65	0.74		
Wettbewerbsteilnehmende	321	0.94	0.79	.000 ***	0.38	
Nichtteilnehmende	360	0.65	0.74			
FB14 (3)	IJSO-Teilnehmende	56	1.05	0.75	.048 *	0.52
	Jufo-Teilnehmende	38	0.66	0.75		

Anmerkungen. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ , n. s. = nicht signifikant.

<i>Aufgabe (Punkte)</i>	<i>Gruppe</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Sig.</i>	<i>d</i>
FC10 (2)	Teilnehmende IJSO-Bundesfinale	36	1.00	0.99	.026 *	0.37
	Wettbewerbsteilnehmende	289	0.71	0.75		
	Teilnehmende IJSO-Bundesfinale	36	1.00	0.99	.002 **	0.61
	Nichtteilnehmende	355	0.56	0.69		
	Wettbewerbsteilnehmende	289	0.71	0.75	.020 *	0.21
	Nichtteilnehmende	355	0.56	0.69		
	IJSO-Teilnehmende	42	0.55	0.67	n. s.	
	Jufo-Teilnehmende	36	0.86	0.83		
FC12 (3)	Teilnehmende IJSO-Bundesfinale	30	1.33	0.80	.000 ***	0.82
	Wettbewerbsteilnehmende	283	0.73	0.72		
	Teilnehmende IJSO-Bundesfinale	30	1.33	0.80	.000 ***	1.13
	Nichtteilnehmende	324	0.54	0.69		
	Wettbewerbsteilnehmende	283	0.73	0.72	.001 **	0.27
	Nichtteilnehmende	324	0.54	0.69		
	IJSO-Teilnehmende	37	0.54	0.73	n. s.	
	Jufo-Teilnehmende	36	0.75	0.73		
FP05 (3)	Teilnehmende IJSO-Bundesfinale	56	2.25	0.96	.000 ***	0.81
	Wettbewerbsteilnehmende	378	1.42	1.04		
	Teilnehmende IJSO-Bundesfinale	56	2.25	0.96	.000 ***	1.09
	Nichtteilnehmende	473	1.18	0.98		
	Wettbewerbsteilnehmende	378	1.42	1.04	.001 **	0.24
	Nichtteilnehmende	473	1.18	0.98		
	IJSO-Teilnehmende	59	1.31	0.99	n. s.	
	Jufo-Teilnehmende	44	1.48	1.05		
FP07 (2)	Teilnehmende IJSO-Bundesfinale	49	1.37	0.70	.000 ***	0.93
	Wettbewerbsteilnehmende	338	0.74	0.67		
	Teilnehmende IJSO-Bundesfinale	49	1.37	0.70	.000 ***	0.99
	Nichtteilnehmende	436	0.73	0.64		
	Wettbewerbsteilnehmende	338	0.74	0.67	n. s.	
	Nichtteilnehmende	436	0.73	0.64		
	IJSO-Teilnehmende	50	0.82	0.63	n. s.	
	Jufo-Teilnehmende	42	0.62	0.70		
KP01 (3)	Teilnehmende IJSO-Bundesfinale	44	1.80	0.88	.001 **	0.55
	Wettbewerbsteilnehmende	344	1.30	0.91		
	Teilnehmende IJSO-Bundesfinale	44	1.80	0.88	.000 ***	0.81
	Nichtteilnehmende	404	1.14	0.81		
	Wettbewerbsteilnehmende	344	1.30	0.91	.009 **	0.19
	Nichtteilnehmende	404	1.14	0.81		
	IJSO-Teilnehmende	52	1.02	0.87	.019 *	0.58
	Jufo-Teilnehmende	43	1.53	0.88		

*Anmerkungen.* \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ , n. s. = nicht signifikant.

war bei den meisten Aufgaben auch für das Verhältnis von Wettbewerbsteilnehmenden zu Nichtteilnehmenden zu erwarten. Grundsätzlich zeigten die Verteilungen über die Kategorien, siehe Abbildung 8.1, das erwartete Verhalten. Um eine erste Übersicht über die verschiedenen Verteilungen der Kategorien zu erhalten, wurden diese in sechs Muster eingeteilt:

**Muster 1** Die Häufigkeit sinkt von Kategorie 0 zu Kategorie 2 bzw. Kategorie 3 kontinuierlich. Die meisten derer, die Aufgaben mit diesem Muster bearbeiteten, beantworteten die Aufgabe falsch.

**Muster 2** Die Häufigkeit steigt von Kategorie 0 zu Kategorie 2 bzw. Kategorie 3 kontinuierlich. Die meisten derer, die Aufgaben mit diesem Muster bearbeiteten, beantworteten die Aufgabe komplett richtig.

**Muster 3** Die Häufigkeit konzentriert sich auf Kategorie 0 und auf die höchste zu erreichende Kategorie, d. h. die Häufigkeit sinkt von Kategorie 0 zu Kategorie 1 bzw. Kategorie 2 und steigt danach wieder an. Die Aufgabe wird kaum oder sehr gut gelöst, d. h. es gibt nur wenige zum Teil richtige Antworten.

**Muster 4** Die Häufigkeit steigt von Kategorie 0 zu Kategorie 1 oder Kategorie 2 an, konzentriert sich auf Kategorie 1 bzw. Kategorie 2 und fällt danach ab. Die Aufgabe wird hauptsächlich zum Teil richtig gelöst, die meisten erreichen Kategorie 1 oder Kategorie 2. Nur wenige erreichen die höchste Kategorie.

**Muster 5** (Nur bei Aufgaben mit den Kategorien 0, 1, 2, 3) Das Muster ähnelt im Wesentlichen Muster 4, jedoch zeigt sich eine Art „Plateau“ zwischen Kategorie 1 und Kategorie 2. Auch diese Aufgabe wird hauptsächlich zum Teil richtig gelöst, Kategorie 1 und Kategorie 2 werden mit einem Unterschied von maximal 8.0% etwa gleich häufig erreicht.

**Muster 6** Das Muster ähnelt im Wesentlichen Muster 1, jedoch erreichen annähernd gleich viele Schülerinnen und Schüler mit einem Unterschied von maximal 2.0% Kategorie 0 und Kategorie 1.

Die Zuordnungen sind durch Zahlen in Abbildung 8.1 dargestellt. Es lässt sich erkennen, dass in einigen Aufgaben die Verteilungen der Kategorien für alle drei oder zumindest zwei Gruppen gleich waren, während in anderen Aufgaben unterschiedliche Muster für die einzelnen Wettbewerbsgruppen vorhanden waren oder auch Besonderheiten in der Verteilung auftraten:

- Das gleiche Muster für alle drei Wettbewerbsgruppen fand sich in Aufgabe EC02 (Muster 5).
- Das gleiche Muster für Teilnehmende der IJSO-Bundesfinale und Wettbewerbsteilnehmende fand sich in Aufgabe FB14 (Muster 4).
- Das gleiche Muster für Wettbewerbsteilnehmende und Nichtteilnehmende fand sich in den Aufgaben EB10 (Muster 5), FC10 (Muster 1), FP07 (Muster 4) und KP01 (Muster 5).
- Die Aufgabe FP05 und FC12 zeigten unterschiedliche Muster für alle drei Wettbewerbsgruppen.

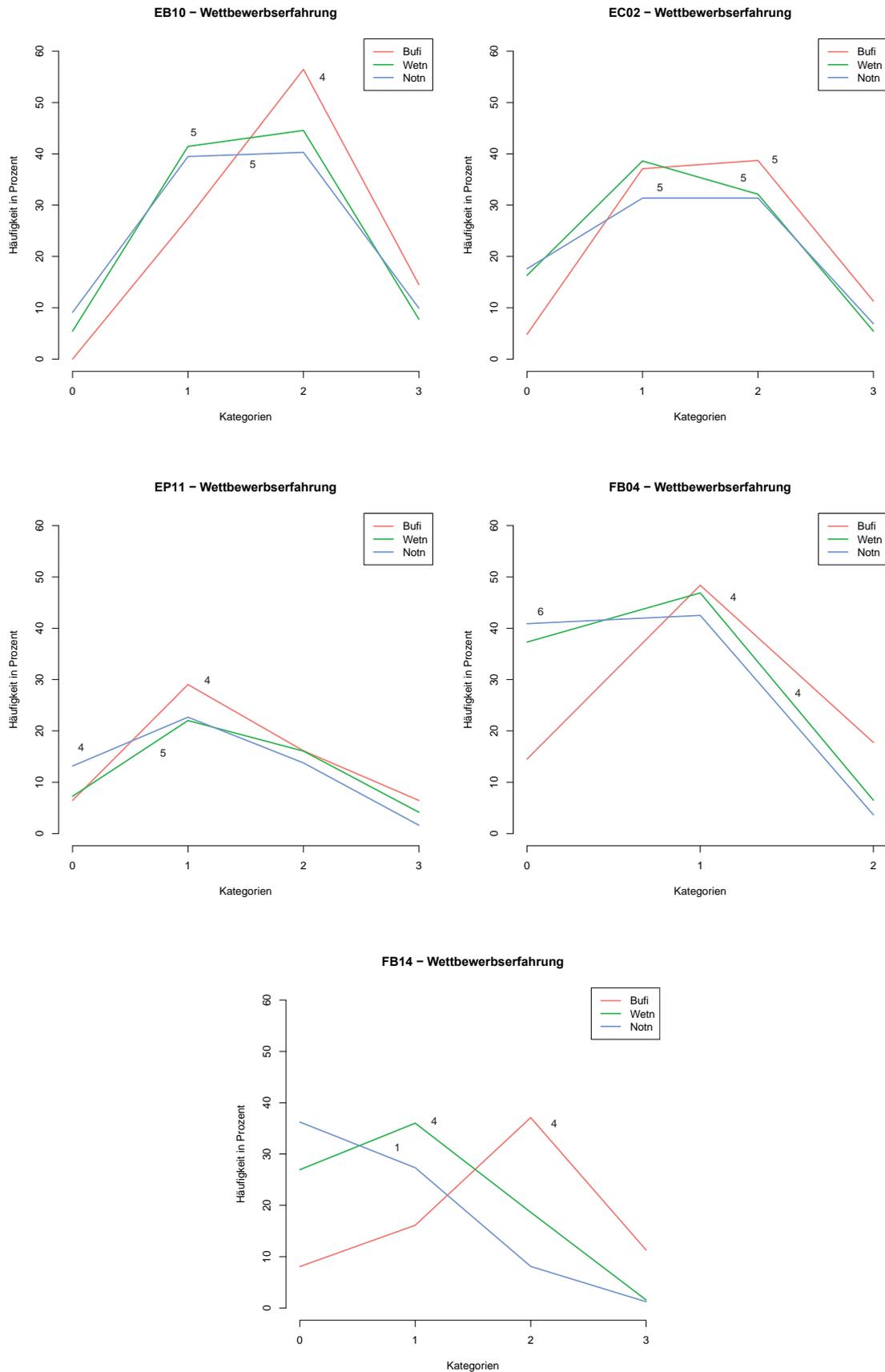


Abbildung 8.1: Vergleich der nach Wettbewerbsgruppe getrennten Verteilung der Kategorien für einzelne Aufgaben BuFi = Teilnehmende der IJSO-Bundesfinale, Wetn = Wettbewerbsteilnehmende, Notn = Nichtteilnehmende.

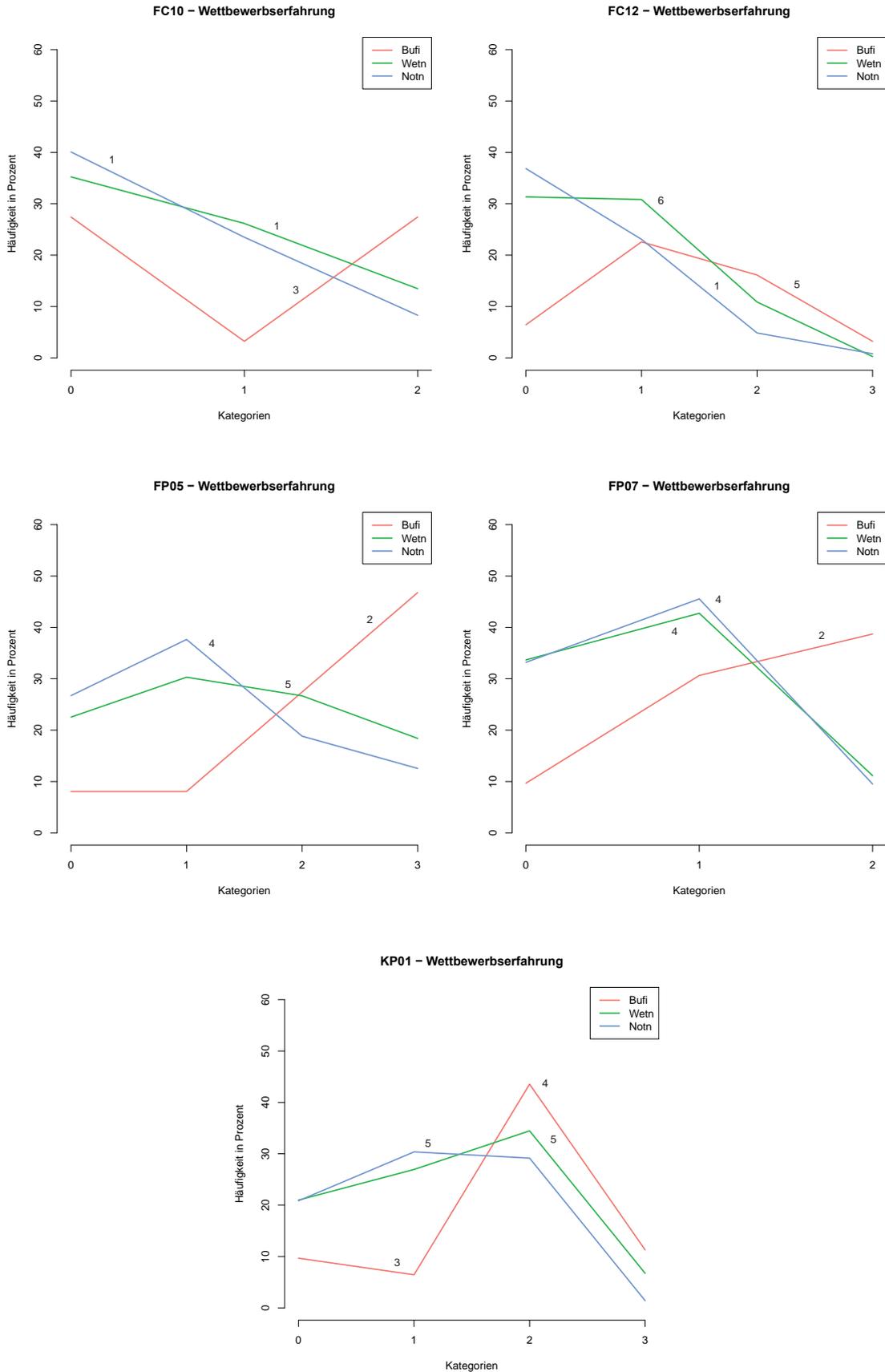


Abbildung 8.0: Vergleich der nach Wettbewerbsgruppe getrennten Verteilung der Kategorien Bufi = Teilnehmende der IJSO-Bundesfinale, Wetn = Wettbewerbsteilnehmende, Notn = Nichtteilnehmende.

- In den Aufgaben EP11 und FB04 traten annähernd gleiche Muster auf, die als Varianten voneinander zu interpretieren waren und sich daher nur minimal unterschieden.

Das am häufigsten aufgetretene Muster war Muster 4, gefolgt von Muster 5, welches im Grunde genommen eine Variation von Muster 4 darstellte. Muster 1 und 6 traten nur für Wettbewerbsteilnehmende und Nichtteilnehmende auf, während Muster 2 und 3 nur für Teilnehmende der IJSO-Bundesfinale zu finden waren.

Von Interesse zur Identifikation möglicher spezifischer Unterschiede zwischen den Wettbewerbsgruppen waren vor allem jene Aufgaben, die für bestimmte Gruppen Abweichungen zeigten. Somit waren die Aufgaben EC02, EP11 und FB04 von geringerem Interesse. Aufgabe EC02 zeigte für alle Gruppen das selbe Muster. Zwar fanden sich für Aufgabe EC02 signifikante Mittelwertsunterschiede zwischen Teilnehmenden der IJSO-Bundesfinale und Wettbewerbsteilnehmenden sowie Nichtteilnehmenden, jedoch kann dies vermutlich auf generell höhere Fähigkeiten zurückgeführt werden und nicht auf spezifische Charakteristika der einzelnen Gruppen. Aufgabe EP11 zeigte zwar für die Wettbewerbsteilnehmenden ein anderes Muster (Muster 5 statt Muster 4 für die anderen beiden Wettbewerbsgruppen), dies war jedoch lediglich eine Variation von Muster 4. Ein signifikanter Mittelwertsunterschied fand sich bei dieser Aufgabe auch nur zwischen Wettbewerbsteilnehmenden und Nichtteilnehmenden. Auffällig waren hier die Anteile derer, die die Aufgabe überhaupt bearbeiteten: So nahmen nur 54.5% der Teilnehmenden der IJSO-Bundesfinale die Aufgabe in Angriff, im Vergleich zu 48.9% der Wettbewerbsteilnehmenden und 50.4% der Nichtteilnehmenden. Bei Aufgabe FB04 bestand der wesentliche Unterschied darin, dass die Teilnehmenden der IJSO-Bundesfinale deutlich seltener nur Kategorie 0 erreichten (14.5%) als die beiden anderen Gruppen (37.3% für Wettbewerbsteilnehmende und 40.9% für Nichtteilnehmende). Die Wettbewerbsteilnehmenden und die Nichtteilnehmenden erreichten fast genauso häufig Kategorie 0 wie Kategorie 1. Kategorie 1 erreichten alle Gruppen zu annähernd gleichen Teilen, Kategorie 2 wurde nur von verhältnismäßig wenigen Schülerinnen und Schülern beantwortet.

Aufgabe FB14 wies das gleiche Muster für Teilnehmende der IJSO-Bundesfinale und Wettbewerbsteilnehmende auf (Muster 4), allerdings zeigte sich eine Verschiebung der Schwerpunkte in den Häufigkeiten. So erreichten Nichtteilnehmende am häufigsten Kategorie 0, Wettbewerbsteilnehmende am häufigsten Kategorie 1 und Teilnehmende der IJSO-Bundesfinale hauptsächlich Kategorie 2. Kategorie 3 erreichten nur 11.3% der Teilnehmenden der IJSO-Bundesfinale, in den anderen Gruppen erreichten diese Kategorie weniger als 2.0% der Schülerinnen und Schüler. Hier stellt sich die Frage, aus welchem Grund diese Aufgabe den Nichtteilnehmenden im Gegensatz zu den anderen Gruppen offensichtlich so schwer fiel.

In Aufgabe EB10, in der die Verläufe für Wettbewerbsteilnehmende und Nichtteilnehmende gleich waren und Muster 5 entsprachen, erhielt niemand der Teilnehmenden der IJSO-Bundesfinale null Punkte, zudem erreichten sie deutlich häufiger (56.5%) die zweite Kategorie als Wettbewerbsteilnehmende (44.6%) oder Nichtteilnehmende (40.3%). Hier stellt sich die Frage, aus welchem Grund die Teilnehmenden der IJSO-Bundesfinale nicht auch häufiger Kategorie 3 erreichten, da die Aufgabe offensichtlich eher leicht für die Teilnehmenden der IJSO-Bundesfinale war. In Aufgabe FC10 erhielten Teilnehmende der IJSO-Bundesfinale relativ häufig null Punkte (27.4%), erreichten dann allerdings kaum Kategorie 1 (3.2%) sondern hauptsächlich zu gleichen Anteilen wie Kategorie 0 auch Kategorie 2. Insgesamt bearbeiteten jedoch auch

nur 54.5% der Teilnehmenden der IJSO-Bundesfinale diese Aufgabe (zum Vergleich: 73.9% Wettbewerbsteilnehmende und 70.7% Nichtteilnehmende). In Aufgabe FP07 erreichten die Wettbewerbsteilnehmenden und Nichtteilnehmenden am häufigsten Kategorie 1. Teilnehmende der IJSO-Bundesfinale erreichten hier am häufigsten die maximale Punktzahl bzw. die höchste Kategorie. Hier stellt sich die Frage, aus welchem Grund die Aufgabe anhand der Muster für zwei Gruppen eher schwierig und für eine Gruppe eher leicht einzuschätzen war. In Aufgabe KP01 zeigte sich ein ähnliches Bild. Wettbewerbsteilnehmende und Nichtteilnehmende wiesen beide das gleiche Muster in den Verteilungen der Häufigkeiten auf und erzielten hauptsächlich zum Teil richtige Antworten. Nur wenige der Teilnehmenden der IJSO-Bundesfinale erreichten Kategorie 1 (sogar weniger als null Punkte erreicht wurden), dafür erreichten die meisten Kategorie 2 und nur wenige Kategorie 3 (allerdings mehr als Kategorie 1), weshalb hier auch zwei Muster (3 und 4) für die Verteilungen angegeben wurden. Auch hier stellt sich die Frage nach den Gründen für diesen Verlauf bei den Teilnehmenden der IJSO-Bundesfinale.

Aufgabe FC12 zeigte für alle drei Gruppen unterschiedliche Muster (Muster 1 für die Nichtteilnehmenden, Muster 6 für die Wettbewerbsteilnehmenden und Muster 5 für Teilnehmende der IJSO-Bundesfinale). Hier war auffällig, dass lediglich 45.5% der Teilnehmenden der IJSO-Bundesfinale diese Aufgabe im Vergleich zu 72.4% der Wettbewerbsteilnehmenden und 64.5% der Nichtteilnehmenden bearbeiteten. Insgesamt erreichten Teilnehmende der IJSO-Bundesfinale kaum Kategorie 0 und am häufigsten Kategorie 1, Wettbewerbsteilnehmende erreichten annähernd zu gleichen Teilen Kategorie 0 (31.3%) und Kategorie 1 (30.8%) und Nichtteilnehmende erreichten am häufigsten Kategorie 0. Kaum jemand erreichte Kategorie 3. Aufgabe FP05 zeigte ebenso wie Aufgabe FC12 für alle drei Gruppen unterschiedliche Muster. Für Nichtteilnehmende zeigte sich Muster 4, für Wettbewerbsteilnehmende Muster 5. Für Teilnehmende der IJSO-Bundesfinale zeigte sich Muster 2, sie erreichten also am häufigsten die höchste Kategorie. Die Nichtteilnehmenden erreichten am häufigsten Kategorie 1, während Wettbewerbsteilnehmende annähernd gleich häufig Kategorie 1 und 2 erreichten. Auch hier stellt sich die Frage, welche spezifischen Anforderungen diese Aufgabe verlangt und wie die Unterschiede erklärt werden könnten.

Zusammenfassend lässt sich feststellen, dass acht von zehn Aufgaben Muster bzw. Auffälligkeiten in den Verteilungen der Häufigkeiten über die Kategorien aufwiesen, die potentiell Unterschiede zwischen den verschiedenen Subgruppen, im Besonderen zwischen Teilnehmenden der IJSO-Bundesfinale und Wettbewerbsteilnehmenden sowie Nichtteilnehmenden, aufdecken könnten. Einige Muster traten gruppenspezifisch auf. Um für mögliche Unterschiede sowie für die aufgetretenen Fragen Erklärungsansätze finden zu können, wurden die Ergebnisse anhand des Kategoriensystems in einem nächsten Schritt detailliert analysiert, siehe Abschnitt 8.3.

### **8.2.2 Vergleich von Teilnehmenden der Internationalen JuniorScience-Olympiade und Jugend forscht bzw. Schüler experimentieren**

Im varianzanalytischen Vergleich von IJSO- und Jufo-Teilnehmenden zeigten sich nur drei signifikante Mittelwertsunterschiede, dabei einer zugunsten der Jufo-Teilnehmenden, so dass man in den Verteilungen der Häufigkeiten über die Kategorien kaum unterschiedliche Muster erwartet hätte. Zudem konnte keine spezifische Annahme über die Verteilungen getroffen werden, da man keine allgemeinen Fähigkeitsunterschiede, sondern vor allem Unterschiede in

bestimmten Fähigkeitsbereichen (vgl. Kapitel 4) oder einzelnen Aufgaben wie bspw. dem Formulieren von Forschungsfragen (Aufgabe EC02) vermutet hätte. Bei detaillierter Betrachtung der Verteilungen der Häufigkeiten über die Kategorien zeigten sich entgegen der Erwartungen spezifische Muster, wie sie auch bei den Wettbewerbsgruppen zu finden waren. Die Muster (für die Beschreibung siehe Abschnitt 8.2.1) sind in Abbildung 8.1 durch Zahlen gekennzeichnet.

Wie auch schon bei den Wettbewerbsgruppen fanden sich auch hier in einigen Aufgaben das gleiche und in anderen unterschiedliche Muster:

- Das gleiche Muster für beide Subgruppen fand sich in den Aufgaben EB10 (Muster 5), FB04 (Muster 4), FC10 (Muster 1) und FP05 (Muster 4).
- Die Aufgaben EC02, EP11, FB14, FP07 und KP01 zeigten unterschiedliche Muster für IJSO- und Jufo-Teilnehmende.
- In Aufgabe FC12 traten annähernd gleiche Muster auf, die als Varianten voneinander zu interpretieren sind.

Am häufigsten trat Muster 4, gefolgt von Muster 1 und Muster 5, auf. Muster 6 (welches eine Variante von Muster 1 darstellte) trat einmalig für die Jufo-Teilnehmenden in Aufgabe FC12 und Muster 3 einmalig für IJSO-Teilnehmende in Kombination mit Muster 5 in Aufgabe KP01 auf. Aufgrund des einmaligen Auftretens der genannten Muster in den Gruppen konnte jedoch nicht von gruppenspezifischen Mustern gesprochen werden.

Auch hier waren vor allem Aufgaben von Interesse, die bestimmte Abweichungen zeigten, die möglicherweise zur Identifikation spezifischer Unterschiede zwischen den beiden Gruppen dienen konnten. Da die Aufgaben EB10 und FB04 gleiche Muster und auch keine signifikanten Mittelwertsunterschiede zeigten, waren diese Aufgaben von geringerem Interesse. In Aufgabe FC10 trat ebenso das gleiche Muster auf. Hier erreichten IJSO- und Jufo-Teilnehmende am häufigsten Kategorie 0 und annähernd gleich häufig Kategorie 1, die höchste Kategorie 2 erreichten dann allerdings deutlich mehr Jufo- als IJSO-Teilnehmende. Jufo-Teilnehmende erzielten dabei annähernd gleich häufig Kategorie 1 (23.4%) und Kategorie 2 (21.3%). Da dieser Unterschied trotz des gleichen Musters für die Verteilungen der Häufigkeiten über die Kategorien auffällig war, wurde mithilfe des Kategoriensystems nach Erklärungsansätzen gesucht. Aufgabe FP05 zeigte gleiche Muster für IJSO-Teilnehmende und Jufo-Teilnehmende, jedoch erreichten IJSO-Teilnehmende am häufigsten Kategorie 1 und Jufo-Teilnehmende am häufigsten Kategorie 2. Auch hier ist zu prüfen, wie diese Verschiebung zustande gekommen sein könnte.

Die Aufgaben EC02, EP11, FB14, FP07 und KP01 zeigten unterschiedliche Muster für die Verteilung der Häufigkeiten über die Kategorien. In Aufgabe EC02 erreichten Jufo-Teilnehmende am häufigsten Kategorie 1 (mit einem Unterschied von 14.5% zu IJSO-Teilnehmenden) und IJSO-Teilnehmende am häufigsten Kategorie 1 und 2. Für die Aufgaben EP11 und FB14 lagen signifikante Mittelwertsunterschiede zugunsten der IJSO-Teilnehmenden vor. In diesen Aufgaben erreichten nur wenige Schülerinnen und Schüler die höchste Kategorie, in den Aufgabe EP11 und FB14 erreichte sogar niemand der Jufo-Teilnehmenden Kategorie 3. Am häufigsten erreichten Jufo-Teilnehmende Kategorie 1 (EP11) bzw. Kategorie 0 (FB14), IJSO-Teilnehmende erreichten am häufigsten Kategorie 1 und 2 (EP11) bzw. Kategorie 1 (FB14).

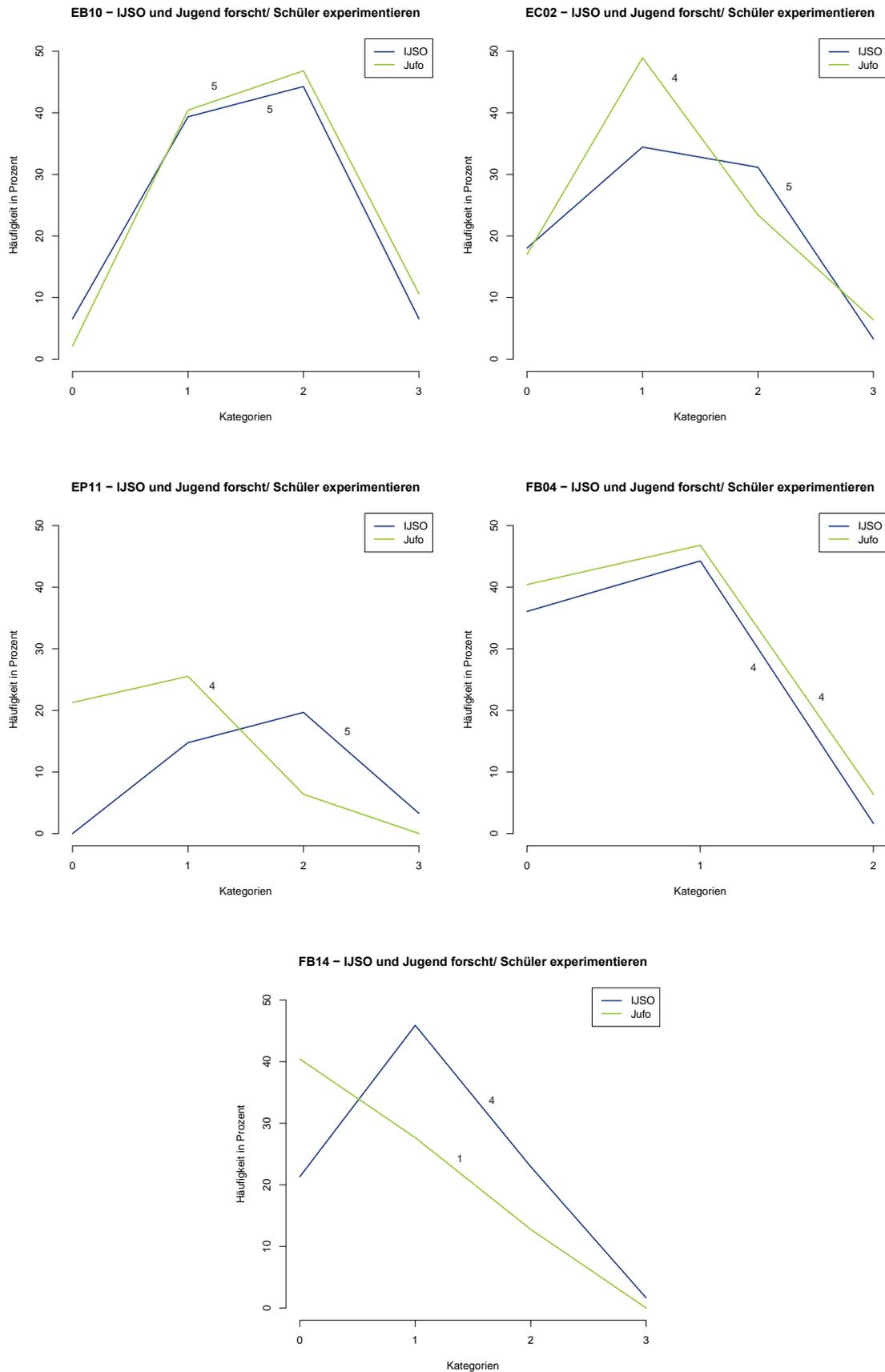


Abbildung 8.1: Vergleich der nach Teilnehmenden der IJSO bzw. Jufo-Teilnehmenden getrennten Verteilung der Kategorien für einzelne Aufgaben

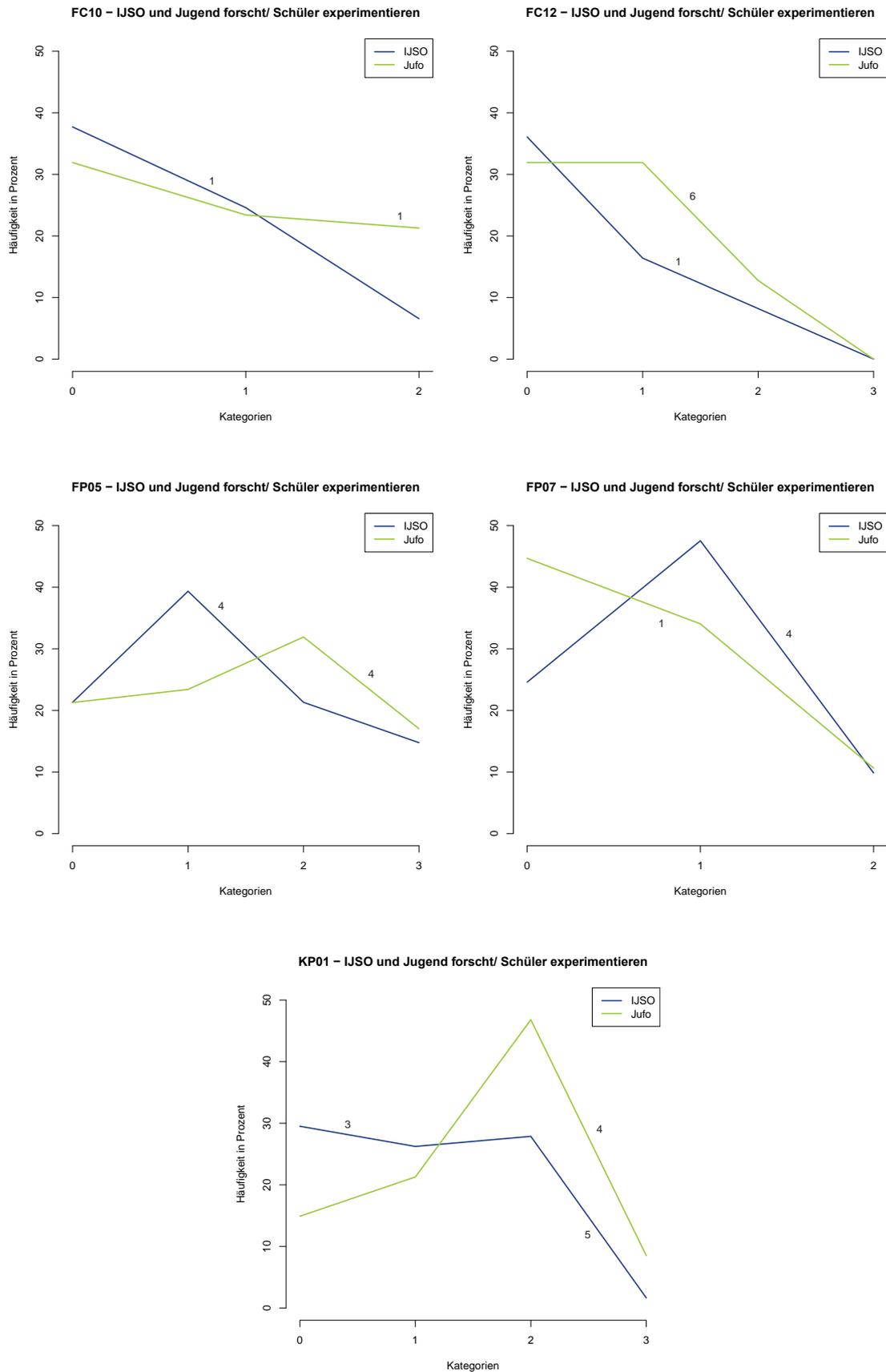


Abbildung 8.0: Vergleich der nach Teilnehmenden der IJSO bzw. Jufo-Teilnehmenden getrennten Verteilung der Kategorien für einzelne Aufgaben

In Aufgabe EP11 erreichten sogar alle IJSO-Teilnehmenden mindestens Kategorie 1. Auffällig war hier jedoch der geringe Anteil an Schülerinnen und Schülern, die diese Aufgabe überhaupt bearbeiteten. So bearbeiteten lediglich 34.9% der IJSO-Teilnehmenden im Vergleich zu 53.2% der Jufo-Teilnehmenden diese Aufgabe. In Aufgabe FP07 erreichten IJSO-Teilnehmende meist Kategorie 1, während Jufo-Teilnehmende meist lediglich Kategorie 0 erzielten. In Aufgabe KP01 wiesen die IJSO-Teilnehmenden einen interessanten Verlauf auf, der denen der Teilnehmenden der IJSO-Bundesfinale ähnelte. So erreichten mehr Teilnehmende Kategorie 0 als Kategorie 1, Kategorie 1 wurde dann allerdings annähernd gleich häufig erreicht wie Kategorie 2. Nur wenige Schülerinnen und Schüler erreichten Kategorie 3. 46.8% der Jufo-Teilnehmenden hingegen erreichten Kategorie 2 (im Gegensatz zu 27.9% der IJSO-Teilnehmenden), der Mittelwertunterschied war signifikant. Bei all diesen Aufgaben ist zu prüfen, wodurch Unterschiede im Abschneiden der beiden untersuchten Subgruppen entstanden sein könnten und ob dies auf mögliche spezifische Unterschiede zwischen IJSO- und Jufo-Teilnehmenden zurückzuführen ist.

In Aufgabe FC12 erreichten Jufo-Teilnehmende (Muster 6) gleich häufig wie Kategorie 0 auch Kategorie 1. IJSO-Teilnehmende (Muster 1) erreichten hier meist lediglich Kategorie 0. Jufo-Teilnehmende erreichten dabei auch deutlich häufiger Kategorie 1 als IJSO-Teilnehmende (31.9% ggü. 16.4%). Niemand der Teilnehmenden erreichte Kategorie 3, zudem bearbeiteten mehr Jufo-Teilnehmende (76.6%) als IJSO-Teilnehmende (56.1%) die Aufgabe.

Zusammenfassend ist festzustellen, dass acht von zehn Aufgaben Unterschiede in den Verteilungen der Häufigkeiten über die Kategorien zeigten, die ursächlich auf spezifische Charakteristika der einzelnen Gruppen zurückgeführt werden könnten. Die gefundenen Muster waren nicht gruppenspezifisch. Die im folgenden Abschnitt ausgeführte inhaltliche Betrachtung der einzelnen Kategorien der Aufgaben diente zur Identifikation möglicher Erklärungsansätze für die geschilderten Unterschiede in den Verteilungen der Häufigkeiten über die Kategorien.

### 8.3 Interpretation der Ergebnisse des Vergleichs der Verteilungen der Antwortkategorien anhand des Kategoriensystems

Im Folgenden werden die einzelnen Aufgaben beschrieben, ihre Kategorien aufgelistet und die Ergebnisse des Vergleichs der Verteilungen der Häufigkeiten über die Kategorien zusammengefasst und interpretiert.

#### 8.3.1 Aufgabe EB10

In dieser Aufgabe musste begründet entschieden werden, welcher von zwei Forschergruppen man einen Auftrag zur Untersuchung von Fischkrankheiten bzw. Fischsterben erteilen würde.

**Kategorie 0** Es wird die falsche Forschergruppe und eine falsche Begründung genannt.

**Kategorie 1** Nennen der richtigen Forschergruppe. Begründung zu ungenau oder zu allgemein (es wird mehr untersucht, man kommt schneller zu einer Lösung, u.ä.). **Oder:** Begründung

gibt quasi den Aufgabentext wieder.

**Kategorie 2** Nennen der richtigen Forschergruppe. Begründung ist eher allgemein formuliert (lebende und tote Fische sollten verglichen werden, Fischkrankheiten treten oft aufgrund falscher Lebensbedingungen auf, man könnte die Bedingungen für die lebenden Fische verbessern, etc.) und bezieht nicht konkret die für die Fische erforderlichen Lebensbedingungen im Fluss mit ein. **Oder:** Antwort ist fachsprachlich ungenau oder nicht ausreichend begründet.

**Kategorie 3** Nennen der richtigen Forschergruppe. Begründete Aussage mit Bezugnahme auf die für die Fische erforderlichen Lebensbedingungen im Fluss (können auch andere als die in der Aufgabe genannten sein), Nennen von Beispielen (Fluss könnte zu warm sein, Wassertemperatur zu hoch, etc.).

### Zusammenfassung und Interpretation der Ergebnisse

- Niemand der Teilnehmenden der IJSO-Bundesfinale erreichte Kategorie 0. Am häufigsten, und häufiger als die anderen beiden Wettbewerbsgruppen - 56.6% ggü. 44.6% für Wettbewerbsteilnehmende und 40.3% für Nichtteilnehmende - erreichten Teilnehmende der IJSO-Bundesfinale Kategorie 2. Die Aufgabe war vermutlich besonders für Teilnehmende der IJSO-Bundesfinale leicht zum Teil richtig zu beantworten. Da Kategorie 2 das Begründen der Antwort verlangt, könnte man daher vermuten, dass Teilnehmende der IJSO-Bundesfinale das Begründen von Antworten leichter als Wettbewerbsteilnehmenden und Nichtteilnehmenden fiel.
- Nur wenige Schülerinnen und Schüler aller drei Gruppen nach Wettbewerbserfahrung erreichten Kategorie 3, auch nur wenige der Teilnehmenden der IJSO-Bundesfinale. Dies könnte hier daran gelegen haben, dass das Nennen von Beispielen nicht explizit im Aufgabentext gefordert wurde.

### 8.3.2 Aufgabe EC02

Hier war die Formulierung einer möglichst präzisen Forschungsfrage anhand zweier gegensätzlicher Aussagen zum Thema Düngemittel gefragt.

**Kategorie 0** Es wird keine (auch keine indirekte) Frage formuliert. **Oder:** Es wird gefragt, ob Pflanzen Mineralstoffe überhaupt benötigen.

**Kategorie 1** Frage ist eher allgemein oder zu unpräzise formuliert. Kein (vergleichender) Bezug zwischen Mineraldünger und Stallmist. **Oder:** Frage, ob Mineraldünger/ Stallmist überhaupt düngt. **Oder:** Frage nach den Inhaltsstoffen von Mineraldünger/ Stallmist. **Oder:** Frage gibt den Aufgabentext wieder.

**Kategorie 2** Frage ist unpräzise formuliert: Es wird nicht genannt, was konkret untersucht werden soll. Ein (vergleichender) Bezug zwischen Mineraldünger und Stallmist ist in jedem Fall vorhanden. **Oder:** Frage zielt darauf ab, welcher Dünger besser ist (auch bezogen auf die Inhaltsstoffe). **Oder:** Es wird sich nur auf das Pflanzenwachstum bezogen, aber nicht darauf eingegangen, dass Dünger dem Boden Mineralstoffe zurückgeben sollte.

**Kategorie 3** Frage ist präzise ausformuliert. Es wird berücksichtigt, dass der Dünger dem Boden in jedem Fall Mineralstoffe in ausreichender Form zurückgeben sollte.

#### Zusammenfassung und Interpretation der Ergebnisse

- Jufo-Teilnehmende erreichten zu 48.9% Kategorie 1.
- IJSO-Teilnehmende erreichten annähernd gleich häufig Kategorie 1 (34.4%) und 2 (31.1%). Ursächlich für die Unterschiede zwischen den beiden Gruppen könnte sein, dass zwar vielen Jufo-Teilnehmenden das Formulieren einer Frage relativ leicht fiel, aber es am Grad der Präzisierung scheiterte, in dem wiederum IJSO-Teilnehmende einen Vorteil gehabt haben könnten.
- Die nach Wettbewerbserfahrung eingeteilten Gruppen wiesen alle das gleiche Muster (Muster 5) auf, Kategorie 3 wurde kaum beantwortet, was vermutlich daran lag, dass für Kategorie 3 die Forschungsfrage sehr präzise formuliert werden musste.

#### 8.3.3 Aufgabe EP11

In dieser Aufgabe ging es um die Auswertung eines Experiments und Ableitung einer allgemeingültigen Regel zur Beziehung zwischen Zugkraft und Neigungswinkel einer schiefen Ebene.

**Kategorie 0** Der geschilderte Zusammenhang ist falsch.

**Kategorie 1** Formulierung einer Regel bezogen auf das Experiment (Rampe statt Weg oder Strecke). **Oder:** Die Daten werden ausgewertet.

**Kategorie 2** Formulierung einer allgemeingültigen Regel (Weg oder Strecke statt Rampe). **Oder:** Interpretation der Ergebnisse und Formulierung einer Regel bezogen auf das Experiment (Rampe statt Weg oder Strecke).

**Kategorie 3** Interpretation der Ergebnisse und Formulierung einer allgemeingültigen Regel (Weg oder Strecke statt Rampe).

#### Zusammenfassung und Interpretation der Ergebnisse

- Niemand der IJSO-Teilnehmenden erreichte Kategorie 0. Am häufigsten erreichten sie Kategorie 2. Allerdings bearbeiteten auch nur 34.9% der IJSO-Teilnehmenden diese Aufgabe, so dass davon auszugehen ist, dass Schülerinnen und Schüler diese Aufgabe nur bearbeiteten, wenn sie auch die Antwort wussten. Hierdurch könnte auch der signifikante Mittelwertsunterschied zugunsten der IJSO-Teilnehmenden entstanden sein, da 53.2% der Jufo-Teilnehmenden diese Aufgabe bearbeiteten, d. h. mehr Schülerinnen und Schüler versuchten sich an dieser Aufgabe, die dann aber hauptsächlich einen Punkt erhielten (niemand erreichte Kategorie 3).
- Die nach Wettbewerbserfahrung eingeteilten Gruppen wiesen alle ein annähernd gleiches Muster auf. Die Aufgabe wurde jedoch nicht sehr häufig bearbeitet (zu 54.5% von Teilnehmenden der IJSO-Bundesfinale, zu 48.9% von Wettbewerbsteilnehmenden und zu 50.4% von Nichtteilnehmenden), nur wenige erreichten die höchste Kategorie. Hier liegt

die Vermutung nahe, dass die im Aufgabentext geforderte allgemeingültige Regel nicht wie gefordert umgesetzt wurde, was hier bedeutet, dass vermutlich der Aufgabentext anders hätte formuliert werden müssen. Alternativ wäre hier auch eine Änderung der Kategorien zu überdenken. Dies könnte auch erklären, weshalb dies die einzige Aufgabe war, in der Teilnehmende der IJSO-Bundesfinale nicht signifikant besser abschnitten.

### 8.3.4 Aufgabe FB04

In dieser Aufgabe ging es um das Formulieren begründeter Vermutungen zum Vorteil einer Haaranalyse gegenüber einer Blutanalyse (siehe auch Anhang D in der in Studie 1 eingesetzten Version).

**Kategorie 0** Als Begründung wird genannt, dass Haare leichter zugänglich sind, Blutanalysen schmerzhaft sind, Haare sich länger halten u. ä.

**Kategorie 1** Die Vermutung ist nicht oder nicht richtig begründet. **Oder:** Begründung ist prinzipiell richtig, aber Antwort ist nicht über den Stoffwechsel begründet. **Oder:** Die Antwort ist fachsprachlich ungenau (Blut bzw. Haare werden schneller gereinigt, Haare sind nur indirekt am Stoffwechsel beteiligt, etc.).

**Kategorie 2** Vermutung wird korrekt darüber begründet, dass die Haare bzw. Haarspitzen nicht mehr am Stoffwechsel teilnehmen.

#### Zusammenfassung und Interpretation der Ergebnisse

- Alle Gruppen ausgenommen der Teilnehmenden der IJSO-Bundesfinale wiesen hier annähernd gleiche Muster in den Verteilungen der Häufigkeiten über die Kategorien auf. Die meisten erreichten hier nur Kategorie 0 oder Kategorie 1, was ein Hinweis auf eine hohe Schwierigkeit der Aufgabe war.
- Teilnehmende der IJSO-Bundesfinale erhielten weniger häufig null Punkte als alle anderen Gruppen und erreichten häufiger die höchste Kategorie. Die meisten beantworteten die Aufgabe jedoch nur zum Teil richtig. Dies könnte daran gelegen haben, dass für die höchste Stufe neben dem Fachwissen auch eine präzise Formulierung der Lösung erforderlich war.

### 8.3.5 Aufgabe FB14

Diese Aufgabe erfragte, wie Nährstoffe aus dem Verdauungstrakt in den Blutkreislauf gelangen.

**Kategorie 0** Die Erklärung ist falsch oder zu allgemein formuliert.

**Kategorie 1** Allgemeine Erklärung ohne Berücksichtigung, dass die Nahrung erst in ihre Bestandteile zerlegt werden muss. Dabei muss deutlich werden, dass die Moleküle gespalten werden, und nicht, dass die Nahrung durch Kauen zerkleinert wird. **Oder:** Es wird zwar gesagt, dass die Nahrung aufgespalten werden muss, aber der Ort der Aufnahme fehlt ganz oder ist falsch.

**Kategorie 2** Erklärung mit Berücksichtigung, dass die Nahrung aufgespalten werden muss und im (Dünn)Darm in den Blutkreislauf gelangt.

**Kategorie 3** Erklärung, dass Nährstoffe an verschiedenen Orten (Mund, Magen, Darm) durch Enzyme zersetzt und im Dünndarm aufgenommen werden. Nennen von mindestens einem Beispiel, z. B. die Stärkeverdauung. Fehlt der Ort der Nahrungsaufnahme oder ein Beispiel, wird ein Punkt abgezogen.

### Zusammenfassung und Interpretation der Ergebnisse

- Nichtteilnehmende und Jufo-Teilnehmende erlangten in dieser Aufgabe am häufigsten Kategorie 0. Die Aufgabe verlangte insgesamt ein hohes Maß an Fachwissen im Themengebiet der Aufgabe sowie eine differenzierte Schilderung des Verdauungsvorgangs unter Verwendung mehrerer Fachbegriffe, was in den beiden Gruppen evtl. derart nicht vorhanden war.
- Wettbewerbsteilnehmende und IJSO-Teilnehmende erreichten am häufigsten Kategorie 1, was einer relativ simplen Beschreibung des Verdauungsvorgangs entsprach.
- Kategorie 2 wurde am häufigsten von Teilnehmenden der IJSO-Bundesfinale erreicht. Diese Gruppe wies vermutlich den Erwartungen entsprechend am ehesten Fachwissen in diesem Bereich auf, was auch das signifikant bessere Abschneiden von IJSO-Teilnehmenden ggü. Jufo-Teilnehmenden erklären könnte.
- Niemand der Jufo-Teilnehmenden, weniger als 2.0% der Wettbewerbsteilnehmenden und der Nichtteilnehmenden sowie lediglich 11.3% der Teilnehmenden der IJSO-Bundesfinale erreichten hier die höchste Kategorie. In der höchsten Kategorie war jedoch auch eine recht differenzierte Beschreibung des Verdauungsvorgangs unter Verwendung von Fachbegriffen gefragt, zudem wurde in der Aufgabe nicht explizit nach dem Ort der Nahrungsaufnahme oder einem Beispiel gefragt, so dass dies dazu geführt haben könnte, dass nur wenige der Teilnehmenden der IJSO-Bundesfinale drei Punkte in dieser Aufgabe erzielten. Hier wäre interessant zu erfahren, ob Schülerinnen und Schüler, die hier Kategorie 3 erreichten, womöglich Schülerinnen und Schüler des IJSO-Nationalteams oder Landessieger im Wettbewerb Jugend forscht oder Schüler experimentieren waren. Anhand der vorliegenden Daten war dies allerdings nicht möglich, so dass dies ein Aspekt wäre, der in weiteren Studien zu prüfen wäre.

### 8.3.6 Aufgabe FC10

Anhand zweier Aussagen zum Thema Aktivierungsenergie sollte in dieser Aufgabe begründet entschieden werden, welche der genannten Aussagen korrekt war.

**Kategorie 0** Es wird die falsche Person und eine falsche Begründung genannt.

**Kategorie 1** Nennen der richtigen Person. Begründung fehlt oder ist falsch. **Oder:** Begründung ist richtig, aber die falsche Person wird genannt. **Oder:** Begründung ist zwar prinzipiell richtig, aber zu ungenau formuliert (es wird z.B. nicht gesagt, welchen Vorteil es bringt, dass mehr Luft zwischen den Holzspänen ist).

**Kategorie 2** Begründung darüber, dass die Aktivierungsenergie für den gleichen Stoff auch gleich ist. **Oder:** Begründung darüber, dass die Holzspäne eine größere Oberfläche haben, weshalb sie besser brennen. **Oder:** Begründung darüber, dass zwischen den Holzspänen mehr Luft bzw. Sauerstoff ist. Da Luft bzw. Sauerstoff brandfördernd sind, brennen die Holzspäne leichter.

### Zusammenfassung und Interpretation der Ergebnisse

- Teilnehmende der IJSO-Bundesfinale erreichten genau so häufig Kategorie 0 wie Kategorie 2. Kategorie 1 wurde von Teilnehmenden der IJSO-Bundesfinale nur zu 3.2% erreicht. Die Aufgabe wurde von Teilnehmenden der IJSO-Bundesfinale dementsprechend entweder falsch oder komplett richtig gelöst, wobei anzumerken ist, dass auch nur 54.5% der Teilnehmenden die Aufgabe bearbeiteten. Wettbewerbsteilnehmende und Nichtteilnehmende bearbeiteten die Aufgabe häufiger (zu 73.9 bzw. 70.7%), erreichten jedoch in den meisten Fällen keinen Punkt. Da die höchste Kategorie neben der richtigen Antwort auch eine richtige Begründung einforderte, ist auch hier zu vermuten, dass Teilnehmende der IJSO-Bundesfinale das Begründen von Antworten leichter fiel.
- Jufo-Teilnehmende erreichten deutlich häufiger Kategorie 2 als IJSO-Teilnehmende. Dies hätte man aufgrund des Ergebnisses für Teilnehmende der IJSO-Bundesfinale andersherum erwartet, so dass das Begründen von Antworten eine Fähigkeit sein könnte, die sich erst im fortgeschrittenen Wettbewerb ausprägt. Für das bessere Abschneiden könnte hier auch schlicht der Aufgabenkontext verantwortlich gewesen sein.

### 8.3.7 Aufgabe FC12

Hier waren begründete Vermutungen zur Funktionsweise der Sicherung von Fingerabdrücken auf Metallen gefragt.

**Kategorie 0** Die Vermutung wird nicht begründet oder ist zu ungenau. **Oder:** Es wird gesagt, dass die Essigsäure den Dreck wegätzt oder die Oberfläche reinigt.

**Kategorie 1** Unpräzise oder fachsprachlich bzw. fachlich nicht richtige Begründung bzw. Vermutung. Beispiel: Die Essigsäure reagiert mit dem Metall oder dem Fingerabdruck. Keine Angabe, dass der Fettfilm des Fingerabdrucks das Metall schützt. **Oder:** Die Essigsäure setzt sich am Fingerabdruck/ am Metall ab, so dass der Fingerabdruck sichtbar wird.

**Kategorie 2** Begründung über eine chemische Reaktion: Die Essigsäure reagiert mit dem Metall. Der Fingerabdruck wird sichtbar, da der Fettfilm bzw. der Fingerabdruck das Metall schützt.

**Kategorie 3** Begründung über eine Oxidation des Metalls (Fachbegriff Oxidation oder Korrosion muss genannt werden). Der Fingerabdruck wird sichtbar, da der Fettfilm des Fingerabdrucks das Metall schützt. Wird nicht gesagt, warum der Fingerabdruck sichtbar wird, wird ein Punkt abgezogen.

### Zusammenfassung und Interpretation der Ergebnisse

- IJSO-Teilnehmende und Nichtteilnehmende erreichten am häufigsten Kategorie 0, was für eine sehr hohe Schwierigkeit der Aufgabe sprach.
- Wettbewerbsteilnehmende und Jufo-Teilnehmende erreichten annähernd gleich häufig Kategorie 0 wie Kategorie 1 (allerdings häufiger Kategorie 0), d. h. Schülerinnen und Schüler dieser Gruppen formulierten zumindest einen zum Teil richtigen Lösungsansatz.
- Teilnehmende der IJSO-Bundesfinale erreichten annähernd gleich häufig Kategorie 1 wie Kategorie 2 (zu 22.6 und 16.1%). Der niedrige Anteil von 45.5% der Teilnehmenden der IJSO-Bundesfinale, die diese Aufgabe überhaupt bearbeiteten, im Vergleich zu 72.4% der Wettbewerbsteilnehmenden und 64.5% der Nichtteilnehmenden, spricht in Kombination mit dem niedrigen Anteil an Antworten der Kategorie 0 (6.5%) dafür, dass die Aufgabe von dieser Gruppe nur bearbeitet wurde, wenn zumindest eine zum Teil richtige Idee der Antwort vorhanden war.
- Weniger als 1.0% der Wettbewerbsteilnehmenden und der Nichtteilnehmenden sowie niemand der IJSO- und Jufo-Teilnehmenden erreichten die höchste Kategorie. Nur 3.2% der Teilnehmenden der IJSO-Bundesfinale erreichten drei Punkte. Dies mag hier daran gelegen haben, dass die Aufgabe insgesamt eine zu hohe Schwierigkeit aufwies oder auch zu wenig Zeit war, das Problem zu durchdenken, da hier zwei Aspekte zur richtigen Beantwortung der Aufgabe wichtig waren (zum einen die Reaktion der Säure mit dem Metall, zum anderen der Schutz des Metalls durch den Fettfilm der Haut). Dass gerade Kategorie 3 so selten erreicht wurde, mag daran gelegen haben, dass der Fachbegriff Oxidation bzw. Korrosion genannt werden musste. D. h., dass viele Schülerinnen und Schüler den Sachverhalt zwar durchschaut hatten, aber alleine aufgrund des fehlenden Fachbegriffs lediglich Kategorie 2 erreichten.

#### 8.3.8 Aufgabe FP05

In dieser Aufgabe zum Thema Lichtausbreitung musste begründet werden welche der beiden genannten Aussagen korrekt war.

**Kategorie 0** Es wird der falsche Kriminologe und eine falsche Begründung genannt.

**Kategorie 1** Begründung fehlt oder ist falsch bzw. unverständlich formuliert.

**Kategorie 2** Begründung ist prinzipiell richtig, aber ungenau bzw. fachsprachlich nicht richtig formuliert. **Oder:** Es wird so argumentiert, dass im Dunkeln nur ein Punkt zu sehen ist.

**Kategorie 3** Begründung ist zutreffend und richtig formuliert. Es wird ausgesagt, dass Licht auf etwas treffen muss, um sichtbar zu werden. Dabei ist es unerheblich, ob Licht gebrochen, reflektiert, gestreut o. ä. wird.

### Zusammenfassung und Interpretation der Ergebnisse

- Teilnehmende der IJSO-Bundesfinale erreichten gleich häufig Kategorie 0 und Kategorie 1 (8.1%).

- Nichtteilnehmende und IJSO-Teilnehmende erreichten am häufigsten Kategorie 1, die Begründung der Antwort stellte hier vermutlich eine Schwierigkeit dar.
- Wettbewerbsteilnehmende erreichten annähernd gleich häufig Kategorie 1 (30.3%) und Kategorie 2 (26.7%).
- Jufo-Teilnehmende erreichten am häufigsten die zweite Kategorie.
- 46.8% der Teilnehmenden der IJSO-Bundesfinale erreichten Kategorie 3, die Aufgabe fiel den Teilnehmenden der IJSO-Bundesfinale eher leicht. Neben dem genauen Formulieren der Begründung, was den Teilnehmenden der IJSO-Bundesfinale meist gut zu gelingen schien, könnte hier auch der Kontext bzw. das Fach Physik eine Rolle gespielt haben. Diese Vermutung wird dadurch gestützt, dass Muster 2 nur bei Teilnehmenden der IJSO-Bundesfinale im Bereich Fachwissen Physik auftrat (siehe Aufgabe FP07).

### 8.3.9 Aufgabe FP07

Hier ging es um die Bestimmung und Begründung des Zeitpunkts der maximalen kinetischen Energie eines geworfenen Balls.

**Kategorie 0** Der Zeitpunkt des Abwurfs und die Begründung sind falsch.

**Kategorie 1** Der Zeitpunkt ist ungenau (am Anfang, kurz nach dem Abwurf, kurz nach dem Loslassen, etc.), aber die Begründung ist richtig. **Oder:** Der Zeitpunkt ist richtig, aber die Begründung fehlt.

**Kategorie 2** Der Zeitpunkt wird detailliert beschrieben und richtig begründet.

### Zusammenfassung und Interpretation der Ergebnisse

- Jufo-Teilnehmende erreichten zumeist die niedrigste Kategorie.
- Wettbewerbsteilnehmende, Nichtteilnehmende und IJSO-Teilnehmende erreichten am häufigsten Kategorie 1. IJSO-Teilnehmende erzielten mit 47.5% deutlich häufiger Kategorie 1 als Jufo-Teilnehmende (34.0%). Da bspw. in Aufgabe FC10 den Jufo-Teilnehmenden das Begründen ihrer Antwort besser gelang, kann kein eindeutiger Trend für die eine oder andere Wettbewerbsgruppe gefolgert werden.
- Teilnehmende der IJSO-Bundesfinale erreichten am häufigsten die höchste Kategorie. Hier war wieder für das Erreichen der höchsten Kategorie das Begründen der Antwort erforderlich, was den Teilnehmenden der IJSO-Bundesfinale offensichtlich erneut besser gelang. Zudem mag neben dem Begründen der Antwort der physikalische Kontext bzw. das Fach (siehe Aufgabe FP05, in der die Gruppe ebenso am häufigsten auch die höchste Kategorie erreichte) dazu geführt haben, dass diese Gruppe so gut abschnitt.

### 8.3.10 Aufgabe KP01

In dieser Aufgabe war die Planung eines Versuchs gefragt, in dem mithilfe von Sonnenlicht Wasser erwärmt werden sollte.

**Kategorie 0** Versuchsplanung ist weder in einem Text noch in einer Zeichnung verständlich dargestellt oder funktioniert in der beschriebenen Weise nicht. **Oder:** Wasser wird mit einer solarbetriebenen Herdplatte erhitzt. **Oder:** Wasser wird einfach nur in die Sonne gestellt. **Oder:** Der Behälter wird nur isoliert.

**Kategorie 1** Der Text oder die Zeichnung sind nicht klar: Die Lichtbündelung wird nicht deutlich, Text und Zeichnung widersprechen sich, o.ä. **Oder:** Es wird nur die Lichtabsorption durch schwarze Materialien erhöht. **Oder:** Sonnenlicht wird nur über einen einzelnen Spiegel o. ä. auf das Wasser reflektiert (Gefäß kann dabei auch isoliert sein).

**Kategorie 2** Licht wird in irgendeiner Art und Weise gebündelt (Lupe, Sammellinse, Trichter aus Alufolie, mehrere Spiegel, Parabolspiegel, ...). Bündelung muss dabei in irgendeiner Art und Weise deutlich werden (entweder im Text (wird gebündelt, verstärkt, es wird vom Brennpunkt gesprochen, usw.) oder in der Zeichnung). **Oder:** Licht wird in eine isolierte Kiste, eine Glaskuppel o. ä. reflektiert (ähnlich einem Solarofen). **Oder:** Es wird die Lichtabsorption erhöht und isoliert. **Oder:** Es wird die Lichtabsorption erhöht und Sonnenlicht über einen einzelnen Spiegel o. ä. auf das Wasser reflektiert.

**Kategorie 3** Licht wird gebündelt und die Lichtabsorption durch schwarze Materialien erhöht/ Isolation wird bedacht/ Erhöhung des Dampfdrucks und ähnliches. Wenn die Bündelung dabei nicht deutlich wird, wird ein Punkt abgezogen.

### Zusammenfassung und Interpretation der Ergebnisse

- Teilnehmende der IJSO-Bundesfinale und IJSO-Teilnehmende zeigten hier einen atypischen Verlauf: Beide Gruppen erreichten weniger häufig Kategorie 1 als Kategorie 0, aber häufiger Kategorie 2 als Kategorie 1. Der Unterschied bei IJSO-Teilnehmenden zwischen Kategorie 1 und 2 war jedoch minimal, während der Unterschied bei Teilnehmenden der IJSO-Bundesfinale 37.0% betrug. Die Aufgabe war somit für Teilnehmende der IJSO-Bundesfinale offenbar leichter als für IJSO-Teilnehmende. Letztere erreichten im Vergleich aller Gruppen rein deskriptiv (d. h. nicht signifikant) sogar den geringsten Mittelwert.
- Nichtteilnehmende erreichten annähernd gleich häufig Kategorie 1 und 2, ähnlich den IJSO-Teilnehmenden.
- Wettbewerbsteilnehmende und Jufo-Teilnehmende erreichten am häufigsten Kategorie 2, wobei Jufo-Teilnehmende deutlich häufiger als IJSO-Teilnehmende Kategorie 2 erzielten (27.9% versus 46.8%). Die Aufgabe forderte neben Fachwissen auch Kreativität, was hier den signifikanten Unterschied für die Jufo-Teilnehmenden gegenüber den IJSO-Teilnehmenden ausgemacht haben könnte. Allerdings ist dies eine Aufgabe, wie sie auch eher im Wettbewerb Jugend forscht bzw. Schüler experimentieren bearbeitet wird, so dass Teilnehmende des Wettbewerbs hier evtl. einen Vorteil hatten. Teilnehmende der IJSO-Bundesfinale wiesen möglicherweise neben dem nötigen Fachwissen zur Beantwortung der Aufgabe ebenso wie die Jufo-Teilnehmenden das nötige Maß an Kreativität auf, so dass dies auch ein Merkmal sein könnte, was für das Erreichen höherer IJSO-Wettbewerbsrunden erforderlich ist.

- Kategorie 3 wurde vergleichsweise wenig erreicht, wobei hier auch mehrere Aspekte kombiniert werden mussten. Hier erfolgte womöglich eine weitere Differenzierung im Leistungsspektrum, auch innerhalb der Gruppe der Teilnehmenden der IJSO-Bundesfinale. Hier wäre, wie auch schon in Aufgabe FB14, interessant zu erfahren, ob Schülerinnen und Schüler, die hier Kategorie 3 erreichten, womöglich Schülerinnen und Schüler des IJSO-Nationalteams oder Landessieger im Wettbewerb Jugend forscht oder Schüler experimentieren waren.

## 8.4 Spezifische Charakteristika und erkennbare Trends einzelner Gruppen

Aus den vorangegangenen Analysen lassen sich die folgenden Charakteristika und Trends ableiten:

- Teilnehmende der IJSO-Bundesfinale schnitten in neun von zehn Aufgaben signifikant besser als Wettbewerbsteilnehmende und Nichtteilnehmende ab. Teilnehmende der IJSO-Bundesfinale erreichten zudem im Vergleich zu den anderen Gruppen in jeder Aufgabe am seltensten null Punkte und am häufigsten die maximale Punktzahl. Muster 1 und 6, bei denen am häufigsten Kategorie 0 bzw. in etwa gleich häufig Kategorie 0 und Kategorie 1 erreicht wurden, traten bei Teilnehmenden der IJSO-Bundesfinale nicht auf. Wettbewerbsexpertise beeinflusste offenbar die Leistung in den jeweiligen Aufgaben positiv.
- Wettbewerbsteilnehmende schnitten in sieben von zehn Aufgaben signifikant besser als Nichtteilnehmende ab. In acht von zehn Aufgaben erreichten die Wettbewerbsteilnehmenden seltener null Punkte als Nichtteilnehmende, bei sechs von zehn Aufgaben erreichten sie häufiger die maximale Punktzahl als die Nichtteilnehmenden. Auch dies ist ein Indiz für den Einfluss von Wettbewerbsexpertise auf die Leistung in den Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten.
- Jufo-Teilnehmende und IJSO-Teilnehmende (ohne Teilnehmende der IJSO-Bundesfinale) schnitten in sieben von zehn Aufgaben gleich ab.
- Teilnehmende der IJSO-Bundesfinale erlangten in Aufgaben, in denen das fachliche und fachsprachliche Begründen von Antworten gefordert war (EB10, FC10, FC12, FP07), häufiger höhere Kategorien, so dass hier eine Stärke vorliegen könnte. Es lässt sich vermuten, dass diese Stärke generell durch höhere Fähigkeiten oder durch die Übung durch die Wettbewerbsteilnahme entstanden sein könnte. Die Fähigkeit, unbekannte Phänomene und mehrere kausale Verknüpfungen beinhaltende Prozesse erklären zu können, zeichnete auch Schülerinnen und Schüler aus, die in PISA 2015 die oberen Kompetenzstufen V und VI erreichten (Schiepe-Tiska, Rönnebeck et al., 2016).
- In mehreren Aufgaben (EP11, FC10) war der Anteil an Teilnehmenden der IJSO-Bundesfinale und weiteren IJSO-Teilnehmenden, welche die Aufgabe bearbeiteten, relativ gering, gleichzeitig schnitten diese dann jedoch verhältnismäßig gut ab. Dies ist Hinweis darauf, dass Teilnehmende der IJSO-Bundesfinale und IJSO-Teilnehmende die Aufgaben sowie ihre eigene Fähigkeit, diese zu lösen, gut einschätzen können.

- Aufgabe FB14, die ein fundiertes Fachwissen erforderte, wurde erwartungskonform am besten von Teilnehmenden der IJSO-Bundesfinale und IJSO-Teilnehmenden gelöst. Dies gibt einen Hinweis auf den Umfang vorliegender Kenntnisse bei Teilnehmenden der IJSO.
- In Physik-Aufgaben des Kompetenzbereichs Fachwissen erreichte fast die Hälfte der Teilnehmenden der IJSO-Bundesfinale, die diese Aufgabe bearbeiteten, die volle Punktzahl. Muster 2 trat auch nur für diese Gruppe und nur in diesen beiden Aufgaben auf. Teilnehmende der IJSO-Bundesfinale scheinen somit auch fundierte physikalische Kenntnisse zu besitzen.
- In Aufgabe KP01 erreichten Teilnehmende der IJSO-Bundesfinale und Jufo-Teilnehmende die höchsten Mittelwerte, einen signifikanten Mittelwertsunterschied zwischen diesen beiden Gruppen gab es jedoch nicht. Der Unterschied zu den Wettbewerbsteilnehmenden und den Nichtteilnehmenden bzw. zu den IJSO-Teilnehmenden war jedoch signifikant. Beide Gruppen wiesen bei dieser Aufgabe neben dem nötigen Fachwissen auch die nötige Kreativität zur Lösung der Aufgabe auf.
- Für IJSO- und Jufo-Teilnehmende waren über die Kreativität hinaus kaum generelle Charakteristika oder Trends abzulesen. In einigen Aufgaben zeigten die untersuchten Gruppen Ähnlichkeiten zu Wettbewerbsteilnehmenden, in anderen zu Nichtteilnehmenden, allerdings lag hier keine Systematik vor. So gelang es IJSO-Teilnehmenden bspw. in den Aufgaben EC02 und FP07 eher als den Jufo-Teilnehmenden, die Antwort richtig zu formulieren bzw. zu begründen, während dies den Jufo-Teilnehmenden bei den Aufgaben FC10 und FP05 im Vergleich besser gelang. Signifikante Unterschiede zugunsten der IJSO-Teilnehmenden zeigten sich in den Aufgaben EP11 und FB14, zugunsten der Jufo-Teilnehmenden in Aufgabe KP01.

Betrachtete man die Mittelwerte der nach übergeordneten Merkmalen zusammengefassten Aufgaben, um hieraus weitere Spezifika oder Trends ablesen zu können, zeigte sich das in Abbildung 8.1 dargestellte Bild für die verschiedenen Wettbewerbsgruppen. Stärken zeigten sich rein deskriptiv demnach für die Teilnehmenden der IJSO-Bundesfinale in Aufgaben aus dem Bereich Fachwissen Physik (Aufgaben FP05, FP07), im Versuche planen (Aufgabe KP01, wobei diese ebenso als Physik-Fachwissensaufgabe gewertet werden kann) und in der Fehleranalyse/ Statistik (Aufgabe EB10). Die Wettbewerbsteilnehmenden zeigten Stärken im Versuche planen, Fehleranalyse/ Statistik, Fragestellungen finden (Aufgabe EC02) und Versuche auswerten (EP11), was eher dem Bereich Erkenntnisgewinnung zuzuordnen ist. Das gleiche Bild zeigte sich für die Nichtteilnehmenden, so dass zusammengefasst in den Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten von einer Stärke der Teilnehmenden der IJSO-Bundesfinale vor allem im Bereich des Fachwissens (Physik) ausgegangen werden kann (was den Erwartungen entsprach, vgl. Kapitel 4), während die Wettbewerbsteilnehmenden und Nichtteilnehmenden vor allem eine Stärke im Bereich der Erkenntnisgewinnung zu haben schienen.

Für die IJSO- und die Jufo-Teilnehmenden zeigte sich das in Abbildung 8.2 zu sehende Bild. Die IJSO-Teilnehmenden zeigten den Erwartungen entsprechend (vgl. Kapitel 4) vor allem Stärken im Bereich Versuche auswerten (Aufgabe EP11), wohingegen die Jufo-Teilnehmenden vor allem im Versuche planen (Aufgabe KP01) Stärken zeigten. Im Bereich Fehleranalyse/ Statistik (Aufgabe EB10) schnitten beide Gruppen annähernd gleich gut ab.

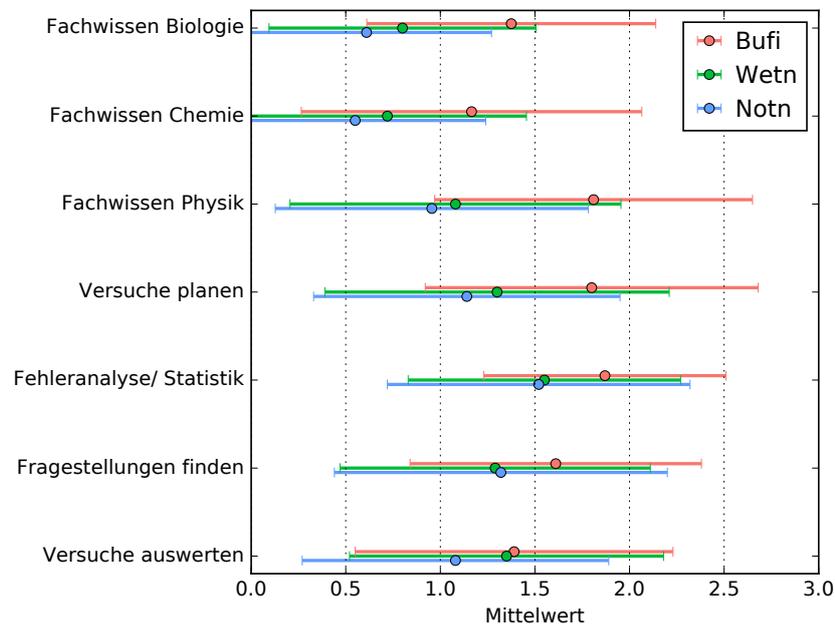


Abbildung 8.1: **Übersicht über das Abschneiden der Wettbewerbsgruppen in den verschiedenen Aufgabenkategorien** Bufi = Teilnehmende der IJSO-Bundesfinale, Wetn = Wettbewerbsteilnehmende, Notn = Nichtteilnehmende. Dargestellt ist neben dem Mittelwert die mittlere Standardabweichung.

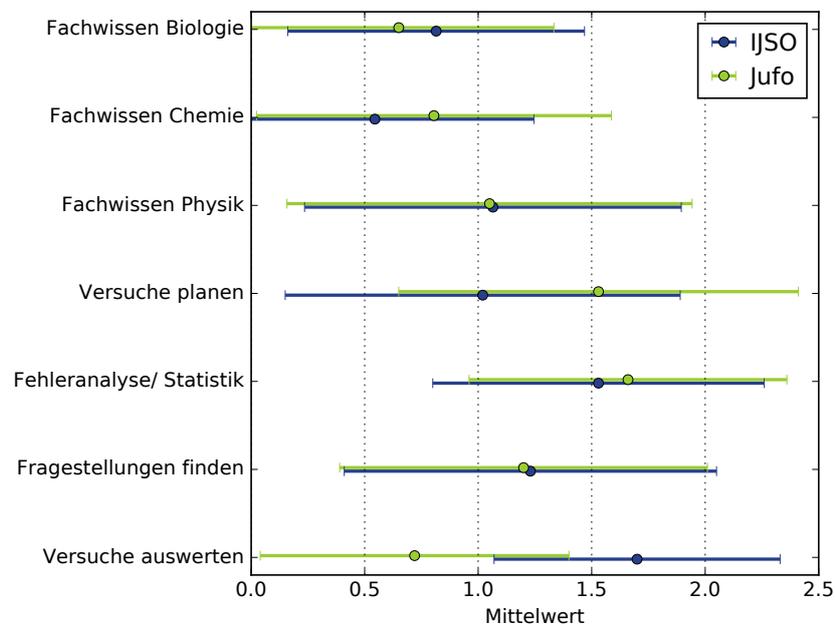


Abbildung 8.2: **Übersicht über das Abschneiden der IJSO- und Jufo-Teilnehmenden in den verschiedenen Aufgabenkategorien** IJSO = Teilnehmende der IJSO, Jufo = Teilnehmende an Jugend forscht bzw. Schüler experimentieren. Dargestellt ist neben dem Mittelwert die mittlere Standardabweichung.

## 8.5 Diskussion

Bei den Ergebnissen der vorangegangenen Analysen sind einige Anmerkungen aufzuführen. So sollten durch die angewendeten Methoden spezifische Charakteristika und Trends der einzelnen Subgruppen identifiziert werden. Dabei sollte beachtet werden, dass jede Aufgabe für sich betrachtet wurde und Verallgemeinerungen kaum vorgenommen werden können. Die Ergebnisse liefern erste Hinweise, die in weiterführenden Studien untersucht werden sollten. Im Besonderen für den Vergleich von IJSO- und Jufo-Teilnehmenden ist zu beachten, dass die untersuchten Gruppen mit  $n = 66$  und  $n = 47$  recht klein waren.

Bei der inhaltlichen Betrachtung der Aufgaben und ihrer Kategorien fiel auf, dass zum Teil nicht eindeutig war, ob die in den Aufgaben erzielten Ergebnisse auf spezifische Charakteristika der Gruppen oder der Aufgaben zurückzuführen waren. So wurde in vielen Aufgaben die höchste Stufe kaum erreicht. Zum Teil lag dies daran, dass die höchste Kategorie einen bestimmten Grad der Präzisierung in der Antwort (z. B. Aufgabe EC02) oder den Gebrauch von Fachsprache (z. B. Aufgabe FC12) erforderte. Bei derartigen Aufgaben wäre zu prüfen, ob das Erreichen der höchsten Kategorie auch mit insgesamt höheren naturwissenschaftlichen Kenntnissen und Fähigkeiten einhergeht (bzw. ob in ähnlichen Aufgaben ebenso die höchste Kategorie erreicht wurde) oder ob dies mit anderen Merkmalen, wie z. B. der sprachlichen Fähigkeit, zusammenhängt.

Eine weitere Herausforderung bei der Interpretation der Ergebnisse ist die Frage der Kausalität. So kann nicht mit Sicherheit gesagt werden, ob bestimmte Charakteristika auf die Teilnahme eines Wettbewerbs und die damit verbundene Übung bzw. Routine oder auf generelle (höhere) Fähigkeiten von bestimmten Gruppen zurückgeführt werden können. Dies wäre bspw. in Interventionsstudien zu prüfen.

Einige Aufgaben vermochten, unabhängig von den erwähnten Einschränkungen, bestimmte Unterschiede zwischen Gruppen aufzudecken. Dies betraf vor allem Aufgaben, in denen sich für jede Gruppe bzw. für mindestens zwei Gruppen Unterschiede im Muster der Verteilungen der Häufigkeiten über die Kategorien fanden (Aufgaben FB14, FC10, FC12, FP05, FP07, KP01 für die Wettbewerbsgruppen und Aufgaben EC02, EP11, FB14, FC12, FP05, FP07, KP01 für IJSO- und Jufo-Teilnehmende). Dabei war jedoch, wie bereits erwähnt, nicht immer eindeutig festzustellen, ob die Unterschiede auf spezifische Charakteristika der Gruppen, der Aufgaben oder auf generell höhere Fähigkeiten bestimmter Gruppen zurückzuführen waren. Dies betraf vor allem Aufgaben, in denen keine signifikanten Mittelwertsunterschiede für bestimmte Gruppen vorlagen (vgl. Tabelle 8.1). Aufgaben, die zwischen allen untersuchten Gruppen signifikant und mit hohen Effektstärken trennen konnten (zwischen den Wettbewerbsgruppen sowie zwischen IJSO-/ Jufo-Teilnehmenden), waren die Aufgaben FB14 und KP01. Diese Aufgaben zeigten auch den Erwartungen entsprechende Ergebnisse und Charakteristika für die einzelnen Gruppen.

Im Vergleich zu Studie 2 ließen sich insgesamt interessante Aspekte identifizieren, die aufgrund der nicht optimalen Modellpassung (d. h. hier aufgrund der geringen Reliabilität, aber nicht aufgrund schlechter Itemparameter oder Fit-Werte), nicht mithilfe probabilistischer Modelle gefunden werden konnten. Dennoch sind die Ergebnisse aus Studie 2 nicht zu ignorieren, da die geringe Reliabilität Hinweise darauf lieferte, dass die Aufgaben für die Stichprobe

keine einheitliche Skala naturwissenschaftlicher Kenntnisse und Fähigkeiten abbildete (vgl. Abschnitt 7.7). Die im vorliegenden Kapitel beschriebene qualitative Analyse lieferte zwar Hinweise auf den Einfluss von Wettbewerbsexpertise auf die in den Aufgaben erzielte Leistung sowie auf Unterschiede zwischen den einzelnen Gruppen, jedoch können von Ergebnissen einzelner Aufgaben keine Schlüsse auf naturwissenschaftliche Kenntnisse und Fähigkeiten im Allgemeinen gezogen werden. Betrachtet man bspw. Aufgabe FB14, bei der sowohl die Unterscheidung zwischen Gruppen als auch die Identifikation spezifischer Charakteristika gelang, sagt die Aufgabe jedoch ausschließlich etwas über das vorhandene biologische Fachwissen im Themenbereich der Aufgabe aus.

## 8.6 Zusammenfassung und Fazit

In Studie 3 wurden die in Studie 2 eingesetzten Aufgaben zur Erfassung der naturwissenschaftlichen Kenntnisse und Fähigkeiten von Wettbewerbsteilnehmenden im Detail qualitativ analysiert, um mögliche Charakteristika und Trends für einzelne Gruppen zu identifizieren. Zunächst wurden Varianzanalysen durchgeführt, bei denen Teilnehmende der IJSO-Bundesfinale bis auf eine Ausnahme besser als Wettbewerbsteilnehmende und Nichtteilnehmende abschnitten. Wettbewerbsteilnehmende schnitten in sieben von zehn Aufgaben besser als Nichtteilnehmende ab. IJSO-Teilnehmende waren in zwei Aufgaben, Jufo-Teilnehmende in einer Aufgabe besser als die jeweils andere Gruppe.

Aus dem Vergleich der Verteilungen der Häufigkeiten über die Kategorien resultierten sechs Muster, zu denen die gruppenspezifischen Verteilungen zugeordnet werden konnten. Hier fanden sich erste Hinweise auf Merkmale der Aufgaben hinsichtlich ihrer Schwierigkeit für bestimmte Gruppen. Anschließend wurden die Muster sowie im Vergleich der Verteilungen identifizierte Auffälligkeiten mithilfe der Antwortkategorien der einzelnen Aufgaben interpretiert und Charakteristika und Trends für die Gruppen, wenn möglich, abgeleitet. Es zeigten sich vor allem Charakteristika der Teilnehmenden der IJSO-Bundesfinale im Vergleich zu den anderen beiden Gruppen. Aus dem Vergleich von IJSO- und Jufo-Teilnehmenden ließen sich keine spezifischen Merkmale ableiten.

Zwei der zehn Aufgaben vermochten deutlich zwischen allen anvisierten Gruppen den Erwartungen entsprechend zu trennen und spezifische Charakteristika zu identifizieren, in den anderen Aufgaben waren die Unterschiede auch z. B. auf Merkmale der Aufgaben bzw. Antwortkategorien zurückzuführen.

Die Forschungsfrage, ob sich qualitative Unterschiede in den Antworten von Teilnehmenden unterschiedlicher naturwissenschaftlicher Schülerwettbewerbe, auch im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern zeigten, ist somit positiv zu beantworten.

Aus den gewonnenen Erkenntnissen konnten vor allem erste Hinweise auf spezifische Charakteristika einzelner Gruppen gefunden werden, die sich bereits in der Analyse von Wettbewerbsanforderungen andeuteten (siehe Kapitel 4) und anhand weiterer Studien untersucht werden sollten.

# Kapitel 9

## Diskussion und Ausblick

In diesem Kapitel werden die Studien und Ergebnisse der vorliegenden Arbeit unter Berücksichtigung der Forschungsfragen zusammengefasst und diskutiert. Im Anschluss werden theoretische und praktische Implikationen der Arbeit erläutert. Abschließend werden die Limitationen der Arbeit erörtert und ein Ausblick auf relevante weiterführende Forschung gegeben.

### 9.1 Zusammenfassung und Diskussion

Die vorliegende Arbeit hatte zum Ziel, die Anforderungen naturwissenschaftlicher Schülerwettbewerbe sowie deren Teilnehmende anhand spezifischer Leistungsmerkmale zu charakterisieren. Zunächst wurde hierzu der Forschungsstand zu naturwissenschaftlichen Schülerwettbewerben zusammengefasst und Forschungsdesiderata aufgezeigt (siehe Kapitel 2). Ein Forschungsdesiderat ergab sich im Bereich von quantitativen Quer- und Längsschnittstudien, die spezifische Leistungsmerkmale von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe erfassen. So ist zwar der Einfluss fachspezifischer kognitiver Fähigkeiten, wie z. B. des domänenspezifischen Wissens oder der Intelligenz, auf die Leistung im Allgemeinen in einer Vielzahl an Studien erforscht worden (siehe Abschnitt 2.3.2), jedoch liegen zu den fachspezifischen kognitiven Fähigkeiten von Schülerinnen und Schülern, die an naturwissenschaftlichen Schülerwettbewerben teilnehmen, nur wenige Erkenntnisse vor (vgl. Abschnitt 2.2.3). Fundierte Kenntnisse sind allerdings notwendig, um naturwissenschaftlich interessierte und talentierte Schülerinnen und Schüler zu fördern und für Berufe im naturwissenschaftlich-technischen Bereich zu begeistern. Dazu gehören neben fundierten Kenntnissen über die Teilnehmenden auch Kenntnisse über die Anforderungen der einzelnen Fördermaßnahmen, hier die Wettbewerbe, zu denen bislang gleichermaßen kaum Erkenntnisse vorliegen.

Studien im Bereich der Wettbewerbsforschung stellen sich als herausfordernd dar, da die Testzeit durch bspw. die Anbindung an die Wettbewerbe und die Vielzahl an notwendigen Erhebungsinstrumenten meist knapp bemessen ist. Zudem existieren meist große Differenzen in den zu messenden Merkmalsausprägungen, insbesondere im Bereich der Leistung (z. B. zwischen erfolgreichen Wettbewerbsteilnehmenden und wettbewerbsunerfahrenen Schülerinnen und Schülern). Aus testökonomischen Gründen und um einen direkten Vergleich verschiedener Gruppen zu ermöglichen ist es daher von besonderem Interesse, die jeweiligen Merkmalsausprägungen anhand eines Instruments für verschiedene Gruppen der Zielpopulation zu erfassen.

Aus den genannten Gründen wurden die folgenden Forschungsfragen für die vorliegende Arbeit formuliert, die in drei Studien beantwortet wurden:

1. Lassen sich die naturwissenschaftlichen Kenntnisse und Fähigkeiten von erfolgreichen Wettbewerbsteilnehmenden im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern mit einem gemeinsamen Testinstrument erfassen? (Studie 1, Kapitel 6)
2. Gelingt die Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten mit einem an die Bedürfnisse der Wettbewerbsforschung angepassten Testinstrument? (Studie 2, Kapitel 7)
3. Zeigen sich qualitative Unterschiede in den Antworten von Teilnehmenden unterschiedlicher naturwissenschaftlicher Schülerwettbewerbe, auch im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern? (Studie 3, Kapitel 8)

Zur Vorbereitung eines ersten Testentwurfs wurde zunächst eine Charakterisierung der Anforderungen zweier exemplarisch gewählter Schülerwettbewerbe vorgenommen (Kapitel 4), anhand derer spezifische Merkmale der Wettbewerbe identifiziert und im Rahmen naturwissenschaftlicher Kenntnisse und Fähigkeiten mittels eines Testinstruments erfasst werden sollten. Ausgewählt wurden zum einen ein aufgabenorientierter Schülerwettbewerb, die IJSO, zum anderen der projektorientierte Schülerwettbewerb Jugend forscht und seine Jugendsparte Schüler experimentieren. Die Charakterisierung der Anforderungen der IJSO erfolgte über eine Analyse der Wettbewerbsaufgaben anhand eines aus existierenden Modellen entwickelten Schemas zur Klassifikation und Analyse von Wettbewerbsaufgaben (siehe Abbildung 4.1). Die Charakterisierung der Anforderungen von Jugend forscht bzw. Schüler experimentieren erfolgte anhand mehrerer Aspekte: die Analyse der Jury-Arbeit und wettbewerbsinterner Richtlinien sowie einer qualitativen Befragung von Jurorinnen und Juroren. Die Ergebnisse zeigten, dass in der IJSO hauptsächlich Aufgaben im Multiple-Choice-Format verwendet wurden, die häufig der Allgemeinen Chemie, der Mechanik und der Tierphysiologie entstammten. Zu 70.0 % waren IJSO-Aufgaben der zweitniedrigsten Lösungskomplexität „beschreibende Zusammenhänge“ zuzuordnen, der mit Abstand am häufigsten vertretene Kompetenzbereich war Fachwissen, gefolgt von Erkenntnisgewinnung. Für eine gute Leistung im Wettbewerb Jugend forscht bzw. Schüler experimentieren benötigten Teilnehmende nach Ansicht der Jurorinnen und Juroren Kompetenzen aus dem Bereich Erkenntnisgewinnung sowie Kreativität und Intelligenz. Die Ergebnisse dienten zwar vorrangig der Auswahl des Messgegenstandes für das Testinstrument, jedoch konnten auch Anforderungsprofile der beiden Wettbewerbe erstellt werden, die eine hohe praktische Relevanz aufweisen (siehe Abschnitt 9.2). So kann eine zielgerichtete nachhaltige Förderung von Schülerinnen und Schülern anhand naturwissenschaftlicher Schülerwettbewerbe nur gelingen, wenn neben Kenntnissen zu den Teilnehmenden auch Kenntnisse zu den Wettbewerben vorliegen. Die gewonnen Erkenntnisse wären für theoretische und praktische Zwecke jedoch noch weiter auszudifferenzieren, wie in Abschnitt 4.2.5 erläutert, und hinsichtlich organisatorischer und weiterer Aspekte zu einem ausführlichen Profil zu ergänzen. Zusätzlich ist zu beachten, dass die Wettbewerbe einer stetigen Entwicklung unterliegen, und somit davon ausgegangen werden kann, dass die Anforderungen Änderungen unterliegen können.

Von den Ergebnissen der Charakterisierung der Anforderungen ausgehend wurde in Kapitel 5 die Entwicklung und Pilotierung eines Erhebungsinstrument dargestellt, welches die naturwissenschaftlichen Kenntnisse und Fähigkeiten von Teilnehmenden der IJSO im Vergleich zu Teilnehmenden anderer naturwissenschaftlicher Schülerwettbewerbe sowie zu Schülerinnen und Schülern, die noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten, in Biologie, Chemie und Physik sowie den Kompetenzbereichen Fachwissen und Erkenntnisgewinnung anhand eines Testinstruments erfassen sollte. Das Testinstrument enthielt Aufgaben im offenen sowie im Multiple-Choice-Format und wurde anhand eines Kategoriensystems bewertet, das aus den Antworten der Schülerinnen und Schüler aus der Pilotierung entwickelt wurde.

In Studie 1 (Kapitel 6) wurde das Testinstrument zu naturwissenschaftlichen Kenntnissen und Fähigkeiten validiert. Die Studie diente somit der Beantwortung der ersten Forschungsfrage. Mit  $n = 195$  Schülerinnen und Schülern zeigte sich, dass das Erhebungsinstrument, welches über ein eindimensionales PCM skaliert wurde, mit einer EAP/PV-Reliabilität von .86 (WLE-Reliabilität .84) naturwissenschaftliche Kenntnisse und Fähigkeiten erfassen konnte. Die Itemparameter und Itemfit-Werte entsprachen den jeweiligen Vorgaben. Die Validität des Testinstruments wurde anhand von Korrelations- und Regressionsanalysen bestätigt. Zudem wurde die Mehrdimensionalität des Instruments sowie das Vorliegen von DIF geprüft. Varianzanalytische Unterschiede in den naturwissenschaftlichen Kenntnissen und Fähigkeiten zeigten sich zwischen Teilnehmenden des IJSO-Bundesfinales ( $M = 0.94$ ,  $SD = 0.56$ ), Schülerinnen und Schülern, die mindestens einmal ( $M = 0.08$ ,  $SD = 0.99$ ,  $p < .001$ ,  $d = 1.00$ ) oder noch nie ( $M = -0.54$ ,  $SD = 0.64$ ,  $p < .001$ ,  $d = 2.40$ ) an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten. Zudem zeigten Schülerinnen und Schüler, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten, signifikant höhere naturwissenschaftliche Kenntnisse und Fähigkeiten als Schülerinnen und Schüler, die noch nie teilgenommen hatten ( $p < .001$ ,  $d = 0.76$ ). Die Forschungsfrage, ob sich die naturwissenschaftlichen Kenntnisse und Fähigkeiten von erfolgreichen Wettbewerbsteilnehmenden im Vergleich zu nicht an Wettbewerben teilnehmenden Schülerinnen und Schülern mit einem gemeinsamen Testinstrument erfassen ließen, konnte somit positiv beantwortet werden. Des Weiteren zeigten die Varianzanalysen, siehe Abbildung 6.3, dass es deutliche Überlappungsbereiche in den naturwissenschaftlichen Kenntnissen und Fähigkeiten zwischen den einzelnen Gruppen gab. Hier stellt sich die Frage, aus welchem Grund Schülerinnen und Schüler, die offenbar die nötigen Kenntnisse und Fähigkeiten aufwiesen, nicht oder nicht erfolgreich an einem mathematisch-naturwissenschaftlichem Wettbewerb teilnahmen. Weitere interessante Ergebnisse gingen aus den bereits erwähnten Regressionsanalysen hervor, die zeigten, dass für eine Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb die Schulnoten in Biologie, Chemie und Physik (gemittelt in einer Variablen), das Alter sowie das Interesse an Tätigkeiten, die in einer naturwissenschaftlichen Fördermaßnahme ausgeübt werden, prädiktiv waren. Für Wettbewerbserfolg, der hier über die Teilnahme am IJSO-Bundesfinale 2013 operationalisiert wurde, waren die naturwissenschaftlichen Kenntnisse und Fähigkeiten (d. h. die Personenfähigkeit), die Schulnoten in Biologie, Chemie und Physik (gemittelt in einer Variablen) sowie die Anzahl der Wettbewerbsteilnahmen prädiktiv. Gerade der Einfluss wiederholter Wettbewerbsteilnahmen auf die Teilnahmebereitschaft bzw. Wettbewerbserfolg konnte schon von Blankenburg et al. (2015) und Urhahne et al. (2012) gezeigt werden. In Verbindung mit den Ergebnissen der Varianzanalyse verdeutlichen die Ergebnisse umso mehr die Notwendigkeit der Untersuchung weiterer möglicher Einflussfak-

toren auf eine Wettbewerbsteilnahme oder einen Wettbewerbserfolg, bspw. aus dem Bereich von *Scientific Thinking and Reasoning* (Dunbar & Klahr, 2012). Zudem zeigte sich hier, dass nicht nur die Schulleistungen oder die kognitiven Fähigkeiten über eine Wettbewerbsteilnahme oder Wettbewerbserfolg entscheiden. Was die Motivation und Förderung möglichst junger Schülerinnen und Schüler zur Wettbewerbsteilnahme betrifft, bieten insbesondere das Interesse an Tätigkeiten, die in einer naturwissenschaftlichen Fördermaßnahme ausgeübt werden, sowie die wiederholte Wettbewerbsteilnahme und die naturwissenschaftlichen Kenntnisse und Fähigkeiten als Prädiktoren einer Wettbewerbsteilnahme bzw. eines Wettbewerbserfolgs einen Ansatzpunkt.

In Studie 2 (Kapitel 7) wurde ein gegenüber Studie 1 reduziertes Erhebungsinstrument mit  $n = 955$  Schülerinnen und Schülern erprobt, was der Beantwortung der zweiten Forschungsfrage diente. Die Kürzung des Testinstruments resultierte aus der knapp bemessenen Testzeit und der Herausforderung, Schülerinnen und Schüler für eine Testbearbeitung zu motivieren. Das reduzierte Testinstrument wies eine EAP/PV-Reliabilität von .54 (WLE-Reliabilität .51) auf und genügte somit nicht den in Studie 1 noch erfüllten Gütekriterien, weshalb die zweite Forschungsfrage, ob die Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten mit einem an die Bedürfnisse der Wettbewerbsforschung angepasstem Testinstrument gelang, negativ beantwortet werden musste. Gründe hierfür waren die geringe Anzahl an Aufgaben sowie das aufgrund inhaltlicher Aspekte heterogen zusammengesetzte Erhebungsinstrument, was sich in der geringen Reliabilität zeigte, sowie das Vorliegen von DIF, im Besonderen zugunsten der Teilnehmenden der IJSO-Bundesfinale. Gerade der letzte Punkt bietet allerdings Diskussionsbedarf, da DIF auch als diagnostisches Instrument zur Analyse von differentiellen Stärken und Schwächen verwendet werden kann (Jurecka, 2010), was in der vorliegenden Arbeit von besonderem Interesse war. So wurde bspw. für die erste TIMS-Studie untersucht, ob Unterschiede in der mathematischen Kompetenz auf z. B. länderspezifische Unterrichtstraditionen zurückgeführt werden können, was bestätigt werden konnte (Klieme & Baumert, 2001). Im Fall der Teilnehmenden der IJSO-Bundesfinale könnte das Vorliegen von DIF für alle Aufgaben des reduzierten Testinstruments somit möglicherweise auf differenzielle Lerngelegenheiten zurückzuführen sein, die sich für Teilnehmende der IJSO-Bundesfinale mehrheitlich geboten haben könnten. Dazu könnte auch die Wettbewerbsteilnahme bzw. der Wettbewerbserfolg zählen, was zu untersuchen wäre (siehe Abschnitt 9.4).

In der in Studie 3 (Kapitel 8) durchgeführten Varianzanalyse auf Basis der in den Aufgaben erreichten Mittelwerte zeigte sich dementsprechend auch, dass in neun von zehn Aufgaben Teilnehmende der IJSO-Bundesfinale 2013 und 2014 besser als andere Wettbewerbsteilnehmende und Nichtteilnehmende abschnitten. Wettbewerbsteilnehmende erreichten in sieben von zehn Aufgaben signifikant höhere Mittelwerte als Nichtteilnehmende. Die intendierte Trennung zwischen Gruppen vermochten die Aufgaben demzufolge trotz nicht optimaler Passung der Daten auf das eindimensionale PCM zu leisten. Die Ergebnisse zeigten zudem, dass Wettbewerbsexpertise einen Einfluss auf die Leistung in den Aufgaben zu naturwissenschaftlichen Kenntnissen und Fähigkeiten hatte. Neben varianzanalytischen Unterschieden zeigten sich des Weiteren in den Aufgaben unterschiedliche Verläufe in den Verteilungen der Antwortkategorien, die, wie das Vorliegen von DIF in Studie 2 andeutete, Hinweise auf spezifische Charakteristika der Gruppen lieferten. So zeigten Teilnehmende der IJSO-Bundesfinale Stärken im Begründen von Antworten, im Fachwissen (insbesondere in Physik-Aufgaben) sowie im Planen von Versuchen, wofür auch Kreativität erforderlich war. Zudem zeigten sich auch Unterschiede

zwischen IJSO- und Jufo-Teilnehmenden, bspw. im Fachwissen Biologie zugunsten von IJSO-Teilnehmenden und im Planen von Versuchen zugunsten von Jufo-Teilnehmenden. Hier bieten sich Ansatzpunkte für die Vorbereitung auf die entsprechenden Wettbewerbe oder auch die Auswahl eines Wettbewerbs (siehe Abschnitt 9.2). Die in Studie 3 behandelte Forschungsfrage, ob sich qualitative Unterschiede in den Antworten von Teilnehmenden unterschiedlicher naturwissenschaftlicher Schülerwettbewerbe, auch im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern, zeigten, konnte somit positiv beantwortet werden.

Die wesentlichen Erkenntnisse der vorliegenden Arbeit werden abschließend noch einmal zusammengefasst:

- Zunächst wurde ein Schema zur Klassifikation und Analyse von Wettbewerbsaufgaben adaptiert und validiert, was zur Analyse weiterer aufgabenorientierter Wettbewerbe genutzt werden kann.
- Anhand des Klassifikationsschemas sowie weiterer qualitativer Aspekte wurden Anforderungsprofile zweier naturwissenschaftlichen Schülerwettbewerbe, der IJSO und Jugend forscht (Schüler experimentieren), erstellt.
- Ausgehend von den Erkenntnissen aus der Analyse der Wettbewerbsanforderungen wurde ein Testinstrument zur Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten in den Fächern Biologie, Chemie und Physik sowie den Kompetenzbereichen Fachwissen und Erkenntnisgewinnung entwickelt und validiert (EAP/PV-Reliabilität .86, WLE-Reliabilität .84). Das Testinstrument kann somit im Rahmen des IKoN zur Charakterisierung von Wettbewerbsteilnehmenden im Vergleich zu wettbewerbsunerfahrenen Schülerinnen und Schülern eingesetzt werden.
- Als signifikante Prädiktoren einer Wettbewerbsteilnahme wurden die Schulnoten in Biologie, Chemie und Physik (gemittelt in einer Variablen), das Alter sowie das Interesse an Tätigkeiten, die in einer naturwissenschaftlichen Fördermaßnahme ausgeübt werden, identifiziert.
- Als signifikante Prädiktoren für den Wettbewerbserfolg, hier der Teilnahme am IJSO-Bundesfinale 2013, fanden sich die naturwissenschaftlichen Kenntnisse und Fähigkeiten (d. h. die Personenfähigkeit), die Schulnoten in Biologie, Chemie und Physik (gemittelt in einer Variablen) sowie die Anzahl der Wettbewerbsteilnahmen.
- Teilnehmende des IJSO-Bundesfinales zeigten signifikant höhere naturwissenschaftliche Kenntnisse und Fähigkeiten als Schülerinnen und Schüler, die mindestens einmal an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten, und wettbewerbsunerfahrene Schülerinnen und Schüler. Zudem zeigte sich ein signifikanter Unterschied zwischen Schülerinnen und Schülern, die mindestens einmal oder noch nie an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb teilgenommen hatten.
- Das Vorliegen von substantiellem DIF im reduzierten Testinstrument gab Hinweise auf die Existenz spezifischer Charakteristika, die zu den unterschiedlichen Parameterschätzungen für die einzelnen Subgruppen geführt haben könnten.

- In einer varianzanalytischen und qualitativen Untersuchung der im reduzierten Testinstrument eingesetzten Aufgaben konnten erste Hinweise auf spezifische Charakteristika einzelner Subgruppen, insbesondere der Teilnehmenden der IJSO-Bundesfinale, identifiziert werden. Insbesondere zwei Aufgaben zeigten signifikante Unterschiede zwischen allen untersuchten Subgruppen mit hohen Effektstärken, aus denen Aussagen zu spezifischen Charakteristika abgeleitet werden konnten.

## 9.2 Theoretische und praktische Implikationen

Aus den Ergebnissen der Arbeit ergeben sich Implikationen für die naturwissenschaftlichen Schülerwettbewerbe als Fördermaßnahmen und deren Teilnehmenden sowie für die Wettbewerbsforschung, im Besonderen für die Erfassung domänenspezifischen Wissens, die im Folgenden erläutert werden.

Zunächst können im Rahmen des IKoN mithilfe des entwickelten Testinstruments naturwissenschaftliche Kenntnisse und Fähigkeiten in Kombination mit affektiven Merkmalen erfasst werden, wodurch eine umfassende Charakterisierung naturwissenschaftlich interessierter und talentierter Schülerinnen und Schüler ermöglicht wird. Die Erkenntnisse hieraus können zudem zur Identifizierung naturwissenschaftlich interessierter und talentierter Schülerinnen und Schüler genutzt werden.

Die Kenntnis über die Anforderungen der Wettbewerbe im Allgemeinen ermöglicht die spezifische Förderung von naturwissenschaftlich interessierten und talentierten Schülerinnen und Schülern, da Lehrkräfte (und auch Schülerinnen und Schüler selbst) so gezielt einen auf Bedürfnisse und Interessen angepassten Wettbewerb auswählen können. Des Weiteren kann die Kenntnis über die Anforderungen von Schülerwettbewerben Lehrkräften sowie Schülerinnen und Schülern der gezielten Vorbereitung auf den Wettbewerb dienen, was Erfolgchancen erhöhen und die Motivation steigern kann. Auch die Entwicklung von auf den entsprechenden Wettbewerb zugeschnittenen Arbeitsgemeinschaften, Fortbildungen oder Fördermaßnahmen kann auf Basis der identifizierten Anforderungen erfolgen. Berücksichtigt man hier zusätzlich den in dieser Arbeit bestätigten Einfluss wiederholter Wettbewerbsteilnahmen auf die naturwissenschaftlichen Kenntnisse und Fähigkeiten und vor allem auf den Wettbewerbserfolg (Urhahne et al., 2012) wird deutlich, dass Schülerinnen und Schüler schon früh zu einer Wettbewerbsteilnahme ermutigt werden sollten. Mit Misserfolgen sollte offensiv umgegangen werden, damit Schülerinnen und Schüler nicht enttäuscht bzw. entmutigt werden. An dieser Stelle tritt die Wichtigkeit der unterstützenden Betreuung durch die Lehrkräfte sowie die Förderung junger Teilnehmender seitens der Wettbewerbe im Rahmen von (ergänzenden) Juniorprogrammen (wie bspw. in der IJSO, vgl. Blankenburg et al., 2015; Blankenburg, Höffler, Peters & Parchmann, 2016) oder Juniorsparten, in denen auch pädagogische Aspekte berücksichtigt werden (bspw. Schüler experimentieren), in den Vordergrund. Verstärkt wird dies zudem dadurch, dass sich als signifikanter Prädiktor für die Teilnahme an einem mathematisch-naturwissenschaftlichen Wettbewerb das Interesse an Tätigkeiten, die in naturwissenschaftlichen Fördermaßnahmen durchgeführt werden, zeigte. In Kombination mit den Erkenntnissen zu den Anforderungen der entsprechenden Wettbewerbe könnten gezielte Qualifizierungsangebote für Betreuende, wie sie bspw. für den Wettbewerb Jugend forscht bzw. Schüler experimentieren in regelmäßigen Abständen angeboten werden, dazu dienen, Schülerinnen und

Schüler optimal auf eine Wettbewerbsteilnahme vorzubereiten.

Die Erkenntnis, dass Wettbewerbserfolg u. a. durch die naturwissenschaftlichen Kenntnisse und Fähigkeiten der Teilnehmenden und der Anzahl ihrer Wettbewerbsteilnahmen vorhergesagt werden konnte, zeigt, wie auch schon von bspw. Urhahne et al. (2012) erwähnt, dass bereits junge Schülerinnen und Schüler zur Teilnahme an einem naturwissenschaftlichen Schülerwettbewerb ermutigt werden sollten. Frühe Wettbewerbsteilnahmen können somit als Training angesehen werden, in denen Schülerinnen und Schüler neben dem allgemeinen Ablauf eines Wettbewerbs auch an domänenspezifischem Wissen gewinnen können, was sie sich durch die Vorbereitung auf den Wettbewerb und die Beschäftigung mit den Aufgaben oder Projekten aneignen können. Erfolg im Wettbewerbskontext ist also „trainierbar“, was sich für Schülerinnen und Schüler motivationssteigernd auswirken kann. Einschränkend zu erwähnen ist an dieser Stelle jedoch die Frage nach der Kausalität, da die Wirkungen naturwissenschaftlicher Schülerwettbewerbe bislang nicht untersucht wurden (siehe Abschnitt 9.4) und an dieser Stelle nicht ausgesagt werden kann, ob eine Zunahme des Wissens auf die Wettbewerbsteilnahme oder andere Faktoren zurückgeführt werden kann. Betrachtet man zu den in Studie 1 erhaltenen Ergebnissen die Ergebnisse der qualitativen Analyse aus Studie 3, zeigten sich Hinweise auf spezifische Charakteristika einzelner Wettbewerbsgruppen. Die Kenntnis dieser kann in der Vorbereitung auf einen Wettbewerb nützlich sein, gerade in Kombination mit den hier eingesetzten Aufgaben. So könnten Lehrkräfte und Schülerinnen und Schüler anhand der Aufgaben, in denen eindeutige Unterschiede zwischen Gruppen gefunden wurden, einen ersten Eindruck gewinnen, ob die Teilnahme am entsprechenden Wettbewerb für bestimmte Schülerinnen und Schüler in Frage käme. Die Aufgaben könnten somit als eine Art Diagnoseinstrument eingesetzt werden und bei der Auswahl eines Wettbewerbs unterstützen. Eine andere Möglichkeit des Einsatzes der Aufgaben wären Maßnahmen wie simulierte Wettbewerbssituationen, die nachweislich die Teilnahmebereitschaft an einem mathematisch-naturwissenschaftlichen Wettbewerb steigern können (Blankenburg, Höffler & Parchmann, 2016), und für die die entwickelten Aufgaben genutzt werden könnten.

Das in Studie 2 eingesetzte reduzierte Testinstrument stellte einen Versuch dar, ein testökonomisches und an die Bedürfnisse der Wettbewerbsforschung angepasstes Instrument zu entwickeln, was teilnehmende Jugendliche auch nicht überforderte. Die Kürzung des Testinstruments auf zehn Aufgaben erbrachte jedoch nicht den gewünschten Erfolg. Die Problematik der begrenzten Testzeit, gerade in Veranstaltungen wie den Schülerwettbewerben, stellt jedoch nach wie vor eine Herausforderung in der Wettbewerbsforschung dar. Zudem zeigten die in Studie 1 generierten Ergebnisse, dass auf die Erfassung kognitiver Fähigkeiten und domänenspezifischen Wissens nicht verzichtet werden sollte. Im Folgenden werden nun zwei weitere Ansätze vorgestellt, die vorbehaltlich empirischer Prüfung eine Möglichkeit der Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten darstellen könnten.

- Beim **computerbasierten adaptiven Testen** werden Probandinnen und Probanden nur solche Items vorgelegt, die aufgrund des individuellen Antwortverhaltens möglichst umfangreiche Informationen über das zu messende Konstrukt liefern (z. B. Frey, 2012). So wird nach jeder Antwort die Fähigkeit der Probandin bzw. des Probanden neu geschätzt und das nächste Item auf Grundlage der ermittelten Personenfähigkeit ausgewählt (van der Linden, 2005). Im Vergleich zu konventionellen Testverfahren müssen beim computerbasierten adaptiven Testen bei gleicher Messpräzision nur etwa die Hälfte

te an Aufgaben den Probandinnen und Probanden vorgelegt werden (Frey & Ehmke, 2007; Segall, 2005). Vorteile der Methode sind des Weiteren, dass auch bei Probandinnen und Probanden mit extrem hoher oder extrem niedriger Merkmalsausprägung eine hohe Messpräzision erreicht werden kann, was in der Wettbewerbsforschung von besonderem Interesse ist. Nachteilig sind neben der Größe des erforderlichen Itempools auch die praktische Umsetzung, da die nötige Hardware zur Verfügung stehen muss (Asseburg, 2011).

- **Formative Messmodelle** nutzen eine andere Herangehensweise als reflektive Messmodelle, die auch in der vorliegenden Arbeit genutzt wurden. So wird in formativen Messmodellen ein theoretisches Konstrukt als Folge der Messindikatoren verstanden, während reflektive Messmodelle davon ausgehen, dass die latente Variable die Ausprägungen der Messvariablen verursacht (Bühner, 2011; Weiber & Mühlhaus, 2014a). Die Auswahl formativer Indikatoren hängt ausschließlich von der Definition und den Bestandteilen des zu messenden Konstrukts ab, woraus resultiert, dass formative Indikatoren im Gegensatz zu reflektiven Indikatoren nicht alle dasselbe messen müssen (Bühner, 2011). Auch müssen sie im Gegensatz zu reflektiven Indikatoren untereinander nicht korreliert und somit nicht eindimensional sein (Weiber & Mühlhaus, 2014b). Dies hat den Vorteil, dass ein Konstrukt durch eine Vielfalt an unterschiedlichen Indikatoren erfasst werden kann, wobei gleichzeitig der Nachteil resultiert, dass die Operationalisierung des Konstrukts und die Auswahl der Messindikatoren sehr sorgfältig erfolgen muss. Da formative Messindikatoren i. d. R. nicht auch als reflektive Messvariablen verwendet werden können (Weiber & Mühlhaus, 2014b), müsste eine formative Erfassung naturwissenschaftlicher Kenntnisse und Fähigkeiten dementsprechend von Grund auf entwickelt werden. Mögliche formative Messindikatoren für naturwissenschaftliche Kenntnisse und Fähigkeiten könnten affektive Faktoren wie das naturwissenschaftliche Interesse, aber auch Fähigkeiten im Sinne von *Scientific Literacy* sein, wie sie auch in der vorliegenden Arbeit berücksichtigt wurden.

### 9.3 Limitationen der Arbeit

Die aus Studie 1 gewonnenen Erkenntnisse basierten auf einer Stichprobe von  $n = 195$ , was einer verhältnismäßig kleinen Stichprobe entsprach. Im Besonderen ist die Gruppe der Teilnehmenden des IJSO-Bundesfinales 2013 zu nennen, die mit  $n = 39$  Schülerinnen und Schülern recht klein ausfiel. Erkenntnisse für diese besondere Gruppe sollten entsprechend anhand größerer Stichproben repliziert werden. Bei maximal 39 Teilnehmenden des Bundesfinales pro Jahr kann dies jedoch nur durch Studien realisiert werden, in denen die Teilnehmenden über mehrere Jahre hinweg befragt werden, ähnlich den Studien von Heller (2008). Gleiches gilt für Teilnehmende des Wettbewerbs Jugend forscht bzw. Schüler experimentieren. Dass die Stichprobengröße gerade im oberen Leistungssegment generell eine Herausforderung in der Wettbewerbsforschung darstellt, zeigen jedoch auch die Studien von Stang et al. (2014) und Urhahne et al. (2012).

Eine weitere Herausforderung stellte die Motivation zur Testbearbeitung dar. In den durchgeführten Studien erhielten die Teilnehmenden aus datenschutzrechtlichen Gründen keine Rückmeldungen über ihr Abschneiden in den Aufgaben. Preise wurden verlost, so dass nicht jeder etwas erhielt. Auch den Lehrkräften konnten, ebenso aus datenschutzrechtlichen Gründen,

keine Rückmeldungen auf Klassen- oder Individualebene gegeben werden. Bei der Testbearbeitung entstand so zum Teil der Eindruck, dass nicht alle Schülerinnen und Schüler gewissenhaft an der Studie teilnahmen. Hier wäre über alternative Rückmeldungen oder Preise nachzudenken, bspw. über anonymisierte Rückmeldungen oder Geldpreise für die Klassenkasse, die allen Teilnehmenden zu Gute kommen würden.

Studie 3 konnte zwar erste Hinweise auf spezifische Charakteristika einzelner Gruppen liefern, jedoch sollten diese noch quantitativ überprüft werden. Da der Einsatz eines reduzierten Testinstruments in Studie 2 nicht erfolgreich gelang, wäre für folgende Studien über alternative methodische Ansätze der Leistungsmessung im Bereich der Wettbewerbsforschung, wie sie in Abschnitt 9.2 genannt wurden, nachzudenken.

Auf Ebene der Erhebungsinstrumente ist anzumerken, dass für Teilnehmende des IJSO-Bundesfinales 2013 bei vielen Variablen und Konstrukten Deckeneffekte auftraten, so dass hier nur das Alter und Geschlecht der Teilnehmenden prädiktiv für das Abschneiden in den Aufgaben waren. Dies gibt Hinweise darauf, dass für diese Leistungsgruppe eigene Instrumente zur Erfassung affektiver Faktoren wie dem Selbstkonzept oder dem Interesse entwickelt werden sollten, sofern man in leistungsstarken Gruppen wie dieser eine Differenzierung erreichen möchte.

Insgesamt sollte die Charakterisierung naturwissenschaftlich interessierter und talentierter Schülerinnen und Schüler anhand mehrerer affektiver und kognitiver Merkmale erfolgen, so wie im Modell IKoN (siehe Abschnitt 2.1) vorgesehen. Die vorliegende Arbeit lieferte Erkenntnisse zu einem Teilbereich des Modells, für eine umfassende Charakterisierung von Wettbewerbsteilnehmenden sind weitere Analysen in Kombination der Teilbereiche des IKoN notwendig.

## 9.4 Ausblick auf weiterführende Forschung

Neben der Erkenntnis, dass das in Studie 1 eingesetzte Erhebungsinstrument die naturwissenschaftlichen Kenntnisse und Fähigkeiten von erfolgreichen Wettbewerbsteilnehmenden im Vergleich zu nichtteilnehmenden Schülerinnen und Schülern erfassen konnte, zeigten die Analysen, die vorrangig der Validierung des Instruments dienten, aufschlussreiche Ergebnisse. In Längsschnittanalysen sollte daher untersucht werden, welche Faktoren beeinflussen, ob und aus welchen Gründen Schülerinnen und Schüler wiederholt an einem Wettbewerb teilnehmen oder aus welchen Gründen eine erneute Teilnahme abgelehnt wird. Einen ersten Ansatz hierzu bietet das am Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN) verortete Projekt WinnerS (Wirkungen naturwissenschaftlicher Schülerwettbewerbe), das untersuchen möchte, welche Faktoren Erfolg bzw. Misserfolg in naturwissenschaftlichen Schülerwettbewerben bestimmen und welchen Einfluss Erfolg und Misserfolg auf die kognitive und affektive Entwicklung von Teilnehmenden naturwissenschaftlicher Wettbewerbe und deren Berufswahl haben (IPN, 2017, Zugriffsdatum 13.02.2017). In Einklang mit diesen Erkenntnissen stehen die Ergebnisse der in Studie 1 durchgeführten Varianzanalysen. So fanden sich nicht nur Unterschiede zwischen Teilnehmenden des IJSO-Bundesfinales und Schülerinnen und Schülern, die noch nie an einem mathematisch-naturwissenschaftlichen Wettbewerb teilgenommen hatten, sondern auch Unterschiede zu Schülerinnen und Schülern, die mindestens einmal an einem entsprechenden Schülerwettbewerb teilgenommen hatten. Zwar kann

im Rahmen der vorliegenden Arbeit nicht beantwortet werden, ob höhere naturwissenschaftliche Kenntnisse und Fähigkeiten sich durch eine Wettbewerbsteilnahme manifestieren oder ob Schülerinnen und Schüler mit höheren naturwissenschaftlichen Kenntnissen und Fähigkeiten bevorzugt an Schülerwettbewerben teilnehmen (was zu untersuchen wäre), jedoch lieferten die Ergebnisse dieser Arbeit erste Hinweise, dass bereits eine einmalige Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb einen Einfluss auf naturwissenschaftliche Kenntnisse und Fähigkeiten haben könnte bzw. dass vorhandenes Wissen die Bereitschaft zur Teilnahme an einem Wettbewerb beeinflussen kann.

Die hier durchgeführten Regressionsanalysen lieferten u. a. Hinweise auf Einflussfaktoren einer Teilnahme an einem mathematisch-naturwissenschaftlichen Schülerwettbewerb. Um die Anzahl Teilnehmender an entsprechenden Wettbewerben erhöhen zu können, wären hier weitere Erkenntnisse unter Einbezug zusätzlicher Faktoren wünschenswert. Hier sei auf die Arbeiten von Blankenburg et al. (2015), Stang et al. (2014) oder Urhahne et al. (2012) verwiesen, die Merkmale wie das Wettbewerbsselbstkonzept, die elterliche Unterstützung oder die erlebte Langeweile im Unterricht untersuchten. In Kombination mit den in dieser Arbeit untersuchten Faktoren könnte so eine differenzierte Analyse möglicher Einflussgrößen einer Wettbewerbsteilnahme erfolgen. Erstrebenswert wären an dieser Stelle auch Analysen von Subgruppen, z. B. getrennt nach Geschlecht oder Alter. So könnten auf lange Sicht detaillierte Fördermaßnahmen zur Erhöhung der Teilnahmebereitschaft für einzelne Subgruppen konzipiert werden.

Betrachtet man das Vorliegen von DIF zugunsten der Teilnehmenden der IJSO-Bundesfinale in Studie 2, wäre hier in Anlehnung an Klieme und Baumert (2001) zu untersuchen, worauf das Vorliegen von DIF zurückgeführt werden kann. Hierzu könnten zunächst qualitative und quantitative Studien mit Lehrkräften, Wettbewerbsleitungen oder auch Teilnehmenden selbst durchgeführt werden, die Hinweise auf besondere Lerngelegenheiten (bspw. besondere Förderprogramme, weitere extracurriculare Aktivitäten, *summer schools*, u. a.) oder Anforderungen geben könnten, die erfolgreiche Wettbewerbsteilnehmende im Vergleich zu anderen Schülerinnen und Schülern nutzen bzw. aufweisen. (Erste Hinweise auf mögliche Faktoren, wie z. B. die wiederholte Teilnahme an naturwissenschaftlichen Schülerwettbewerben, wurden bereits in dieser Arbeit identifiziert.) Anschließend könnte man in Kombination aus Fragebogenstudien und Leistungsmessungen durch Korrelationsanalysen prüfen, ob bestimmte Lerngelegenheiten, Anforderung oder weitere Faktoren mit den DIF-Parametern zusammenhängen. Hieraus könnten dann systematisch weitere Charakteristika von Teilnehmenden naturwissenschaftlicher Schülerwettbewerbe generiert werden.

Durch die gezielte und nachhaltige Förderung von naturwissenschaftlich interessierten Schülerinnen und Schülern, z. B. anhand von Schülerwettbewerben, könnte mit den geschilderten Maßnahmen das Gelingen, was Stefan Schmitt in DIE ZEIT Nr. 29/2015 vorschlägt:

„Wollen wir mehr *21st century scientists*, müssen wir vor allem die kindliche Begeisterung erhalten.“

## Anhang A

# Sonderfälle bei der Klassifikation von Aufgaben der Internationalen JuniorScienceOlympiade

### A.1 Sonderfälle bei der Lösungskomplexität

- War in einer Aufgabe eine Begründung, Erklärung oder Erläuterung der Antwort gefragt, erfolgte immer eine Zuordnung zur Lösungskomplexität „erklärende Zusammenhänge“. Dies erfolgte auch, wenn eine Begründung, Erklärung oder Erläuterung in der Lösung gefordert, aber in der Aufgabe nicht verlangt wurde (und umgekehrt).
- Berechnungen wurden den „beschreibenden Zusammenhängen“ zugeordnet. Wurde anhand dieser Rechnung jedoch etwas begründet, erklärt, erläutert oder auch (mathematisch) bewiesen, wurde diese Aufgabe der Lösungskomplexität „erklärende Zusammenhänge“ zugeordnet.
- Sofern Teilnehmende eine (experimentelle) Beobachtung beschreiben oder protokollieren sollten, wurde solch eine theoretische Aufgabe der Lösungskomplexität „beschreibende Zusammenhänge“ zugeordnet. Handelte es sich dabei um eine Multiple-Choice-Aufgabe, wurde jedoch die Stufe „Fakten“ zugeordnet, da hierbei keine aktive Beschreibung von Vorgängen stattfand.
- Das Ergänzen von Diagrammen, Abbildungen, Zeichnungen etc. wurde der Lösungskomplexität „beschreibende Zusammenhänge“ zugeordnet. Sollte dies begründet erfolgen, wurde die Aufgabe der Lösungskomplexität „erklärende Zusammenhänge“ zugeordnet.
- Sollten Diagramme vollständig erstellt werden, wurde dies der Lösungskomplexität „erklärende Zusammenhänge“ zugeordnet. Dies lag darin begründet, dass sich die Konstruktion von Diagrammen in mehrere Anforderungen unterteilen lässt (vgl. Lachmayer, Nerdel & Prechtel, 2007). Daher wurde auch die dritte Komplexitätsstufe „erklärende Zusammenhänge“ zugeordnet, wenn in theoretischen Aufgaben Informationen aus Diagrammen, Zeichnungen oder Abbildungen entnommen werden sollten, um die Aufgabe zu lösen. War dies bei Multiple-Choice-Aufgaben der Fall, wurde dies jedoch nur der Lösungskomplexität „beschreibende Zusammenhänge“ zugeordnet. Sollten die Informationen nur genannt werden, wurde die Lösungskomplexität „Fakten“ zugeordnet.

- Ebenso der Lösungskomplexität „erklärende Zusammenhänge“ zugeordnet wurden Aufgaben, in denen die Teilnehmenden Informationen recherchieren mussten, um mit diesen die Aufgabe zu lösen. Der gleichen Komplexitätsstufe zugeordnet wurden zudem Aufgaben, in denen Teilnehmende Nahrungsnetze oder Stammbäume entwickeln sollten.
- Lückentexte wurden der Stufe „beschreibende Zusammenhänge“ zugeordnet, da Teilnehmende hierbei die Begriffe in einen (gegebenen) Zusammenhang bringen und zuordnen sollten (anders als bei der isolierten Beschriftung von Zeichnungen u. ä., siehe unten).
- Aufgaben, die das Aufstellen, Ergänzen oder Ausgleichen von Reaktionsgleichungen erforderten, wurden „beschreibenden Zusammenhängen“ zugeordnet. Dies beinhaltete auch die Ermittlung von Oxidationszahlen im Fach Chemie. Sollte eine Wortgleichung aufgestellt oder eine Reaktionsgleichung nur ergänzt und nicht ausgeglichen werden, wurde die Aufgabe der Lösungskomplexität „Fakten“ zugeordnet, da die Zusammenhänge der Reaktion nicht mehr verstanden werden mussten, sondern meist war es lediglich erforderlich, einen Stoff in die Summenformel zu übersetzen.
- Sollten Sachverhalte unbegründet mit einer Abbildung, Zeichnung oder Skizze veranschaulicht, Schlussfolgerungen gezogen sowie Eigenschaften u. ä. unbegründet abgeleitet werden, wurden diese Aufgaben ebenso „beschreibenden Zusammenhängen“ zugeordnet. Dies galt auch für die Formulierung unbegründeter Vermutungen.
- Aufgaben, in denen die Teilnehmenden großtechnische Verfahren in einer Zeichnung oder Skizze darstellen sollten, wurden der Lösungskomplexität „beschreibende Zusammenhänge“ zugeordnet. Sollten Versuchsanordnungen mit einer gegebenen Skizze geplant, gegebene Versuchsanordnungen mit einer Zeichnung oder Skizze veranschaulicht, Versuchsskizzen mit gegebenen Geräten entwickelt bzw. geplant oder aber gegebene Versuchsskizzen verschriftlicht werden, wurde ebenso diese Komplexitätsstufe zugeordnet.
- Das Planen von Versuchsanordnungen ohne eine gegebene Skizze bzw. die Planung einer Versuchsskizze ohne gegebene Geräte wurde aufgrund fehlender Hilfen der Lösungskomplexität „erklärende Zusammenhänge“ zugeordnet.
- Aufgaben, in denen Zeichnungen und Skizzen (auch mit gegebenen Begriffen) beschriftet werden sollten, wurden der Lösungskomplexität „Fakten“ zugeordnet. Sollten Zeichnungen und Skizzen komplett erstellt oder begründet ergänzt werden, wurde dies der Lösungskomplexität „erklärende Zusammenhänge“ zugeordnet. Bei unbegründeten Ergänzungen wurde die Lösungskomplexität „beschreibende Zusammenhänge“ zugeordnet.

## A.2 Sonderfälle bei den Kompetenzbereichen

- In Anlehnung an die Bildungsstandards Chemie wurde das Zuordnen von Abbildungen, Diagrammen u. ä. zu Begriffen, Sachverhalten etc. dem Kompetenzbereich Erkenntnisgewinnung zugeordnet, da laut diesem Standard Schülerinnen und Schüler „in erhobenen oder recherchierten Daten, Trends, Strukturen und Beziehungen [finden], [diese] erklären und geeignete Schlussfolgerungen [ziehen]“ (E6; Kultusministerkonferenz, 2005a, S. 12).
- Berechnungen sowie (mathematische) Beweise wurden generell dem Kompetenzbereich Fachwissen zugeordnet. Dies erfolgte gemäß den Standards in Physik, die verlangen,

dass Schülerinnen und Schüler „ihre Kenntnisse über physikalische Grundprinzipien, Größenordnungen, Messvorschriften, Naturkonstanten sowie einfache physikalische Gesetze“ wiedergeben (F2; Kultusministerkonferenz, 2005b, S.11) sowie „diese Kenntnisse zur Lösung von Aufgaben und Problemen“ nutzen (F3; Kultusministerkonferenz, 2005b, S.11).

- Verlangte eine Aufgabe das Beschreiben oder Protokollieren von (theoretischen) Experimenten, wurde diese generell dem Kompetenzbereich Erkenntnisgewinnung zugeordnet. In den Bildungsstandards finden sich hierzu Zuordnungen zu mehreren Kompetenzbereichen, so dass eine gesonderte Vorschrift nötig war. So heißt es zwar in den Standards des Kompetenzbereichs Erkenntnisgewinnung, dass Schülerinnen und Schüler „qualitative und einfache quantitative experimentelle und andere Untersuchungen durch[führen] und protokollieren“ (Chemie, E3; Kultusministerkonferenz, 2005a, S. 12) und „einfache Experimente [planen], durch[führen] und dokumentieren“ (Physik, E8; Kultusministerkonferenz, 2005b, S. 11), aber auch im Kompetenzbereich Kommunikation finden sich ähnliche Standards. So sollen Schülerinnen und Schüler „Ergebnisse und Methoden biologischer Untersuchung dar[stellen] und [damit] argumentieren“ (Biologie, K6; Kultusministerkonferenz, 2005c, S. 15), „den Verlauf und die Ergebnisse von Untersuchungen und Diskussionen in angemessener Form [protokollieren]“ (Chemie, K6; Kultusministerkonferenz, 2005a, S. 13) sowie „die Ergebnisse ihrer Arbeit [dokumentieren]“ (Physik, K5; Kultusministerkonferenz, 2005b, S. 12). Um den Bezug zum Experimentellen bzw. zum Prozess der Erkenntnisgewinnung aufrecht zu erhalten, wurde bei derartigen Aufgaben die Zuordnung zum Kompetenzbereich Erkenntnisgewinnung vorgenommen.
- Sollten Teilnehmende in einer Aufgabe begründet Stellung nehmen oder einen Sachverhalt in fachlicher Hinsicht beurteilen, wurde eine solche Aufgabe dem Kompetenzbereich Kommunikation zugeordnet. In den Bildungsstandards findet man hierzu bspw. die Chemiestandards K3 und K8, die vorsehen, dass Schülerinnen und Schüler „fachlich korrekt und folgerichtig [argumentieren]“ sowie „Darstellungen in Medien hinsichtlich ihrer fachlichen Richtigkeit [prüfen]“ (Kultusministerkonferenz, 2005a, S. 12-13). Hier ist zu beachten, dass die Beurteilung oder Bewertung von Sachverhalten in gesellschaftlicher Hinsicht jedoch dem Kompetenzbereich Bewertung zugeordnet wird (vgl. Biologie B3; Kultusministerkonferenz, 2005c, S. 15).
- Differenzierte Zuordnungen waren bei Aufgaben erforderlich, die das Erstellen oder Ergänzen von Zeichnungen, Diagrammen, Abbildungen oder Versuchsskizzen verlangten. War das Erstellen oder Ergänzen von Diagrammen, Abbildungen oder Zeichnungen gefordert, wurden solche Aufgaben dem Kompetenzbereich Kommunikation zugeordnet, so wie es auch die Standards für die einzelnen Fächer vorsehen. Gemäß diesen Standards sollen Schülerinnen und Schüler „Daten messbarer Größen zu Systemen, Struktur und Funktion sowie Entwicklung angemessen mit sprachlichen, mathematischen oder bildlichen Gestaltungsmitteln [veranschaulichen]“ (Biologie K3; Kultusministerkonferenz, 2005c, S. 11), „chemische Sachverhalte unter Verwendung der Fachsprache und/oder mit Hilfe von Modellen und Darstellungen [beschreiben, veranschaulichen oder erklären]“ (Chemie K4; Kultusministerkonferenz, 2005a, S. 12) sowie „sich über physikalische Erkenntnisse und deren Anwendungen unter angemessener Verwendung der Fachsprache und fachtypischer Darstellungen aus[tauschen]“ (Physik K1; Kultusministerkonferenz, 2005b, S. 12). Der Kompetenzbereich Kommunikation wurde auch zugeordnet, wenn neben der graphischen

Darstellung eine Erklärung verlangt wurde, da Schülerinnen und Schüler „Originale oder naturgetreue Abbildungen mit Zeichnungen oder idealtypischen Bildern [beschreiben und erklären]“ (Biologie K2; Kultusministerkonferenz, 2005c, S. 15). Sofern ein schon vorhandenes Diagramm, eine Abbildung oder Zeichnung lediglich beschriftet oder in einem vorhandenen Diagramm, einer Abbildung oder Zeichnung etwas markiert oder gekennzeichnet werden sollte, wurde diese Aufgabe jedoch dem Kompetenzbereich Fachwissen zugeordnet, was auch der allgemeinen Definition des Kompetenzbereichs in den Bildungsstandards der jeweiligen Fächer entspricht. Sollte eine Darstellung im Rahmen der Planung eines Experiments erstellt werden, war dies dem Kompetenzbereich Erkenntnisgewinnung zuzuordnen. Dies entspricht den Bildungsstandards, da das Planen von Experimenten auch hier dem Kompetenzbereich Erkenntnisgewinnung zugeordnet ist (Biologie E6; Kultusministerkonferenz, 2005c, S. 14; Chemie E2; Kultusministerkonferenz, 2005a, S. 12; Physik E8; Kultusministerkonferenz, 2005b, S. 11). Ebenso dem Kompetenzbereich Kommunikation zugeordnet wurden Aufgaben, in denen Sachverhalte mit einer Abbildung, Zeichnung oder Skizze veranschaulicht werden sollten (Biologie K2; Kultusministerkonferenz, 2005c, S. 15) sowie Aufgaben, in denen (großtechnische) Verfahren in einer Zeichnung oder Skizze dargestellt werden sollten (Biologie K2 und K3; Kultusministerkonferenz, 2005c, S. 15; Chemie K4; Kultusministerkonferenz, 2005a, S. 12). Sollte eine schon gegebene Versuchsskizze verschriftlicht werden, wurde gemäß der Bildungsstandards Biologie E6, Chemie E2 und Physik E8 (Kultusministerkonferenz, 2005c, S. 14, Kultusministerkonferenz, 2005a, S. 12, Kultusministerkonferenz, 2005b, S. 11) der Kompetenzbereich Erkenntnisgewinnung zugeordnet.

- Experimente und unmittelbar zum Experiment zugehörige (Teil-)Aufgaben wurden dem Kompetenzbereich Erkenntnisgewinnung zugeordnet, was im Allgemeinen auch den Definitionen dieses Kompetenzbereichs in den einzelnen Fächern entspricht.
- War es erforderlich, dass Teilnehmende aus gegebenen Diagrammen, Abbildungen oder Tabellen Informationen entnehmen und im weiteren Lösungsprozess der Aufgabe diese Informationen anwenden sollten, ordnete man diese Aufgaben dem Bereich Erkenntnisgewinnung zu. Diese Vorgehensweise folgte den Standards für die Fächer Chemie (E5 und E6; Kultusministerkonferenz, 2005a, S. 12) und Physik (E2; Kultusministerkonferenz, 2005b, S. 11). Sofern eine reine Informationsentnahme ohne die Anwendung der entnommenen Informationen verlangt wurde, ordnete man diese Aufgabe dem Kompetenzbereich Fachwissen zu, da die Informationsentnahme dann nicht mehr als Teil eines Erkenntnisprozesses angesehen wurde.
- Auch dem Kompetenzbereich Fachwissen zugeordnet wurden Lückentexte, was in Anlehnung an die allgemeine Definition des Kompetenzbereichs in den Bildungsstandards der jeweiligen Fächer vorgenommen wurde.
- Laut Bildungsstandards sollen Schülerinnen und Schüler „zu einem chemischen Sachverhalt in unterschiedlichen Quellen“ (Chemie K1; Kultusministerkonferenz, 2005a, S. 12) bzw. „in unterschiedlichen Quellen“ recherchieren (Physik K3, Kultusministerkonferenz, 2005b, S. 11). Daher wurden auch Aufgaben, in denen Teilnehmende Informationen recherchieren sollten um mit diesen Informationen Aufgaben zu lösen, dem Kompetenzbereich Kommunikation zugeordnet.

- In einigen Aufgaben sollten Teilnehmende Nahrungsnetze aus gegebenen Informationen entwickeln. Die Bildungsstandards Biologie sehen im Kompetenzbereich Kommunikation vor, dass Schülerinnen und Schüler „Daten messbarer Größen zu Systemen, Struktur und Funktion sowie Entwicklung angemessen mit sprachlichen, mathematischen oder bildlichen Gestaltungsmitteln“ veranschaulichen sowie „Informationen zu biologischen Fragestellungen aus verschiedenen Quellen zielgerichtet aus[werten] und diese auch mit Hilfe verschiedener Techniken und Methoden adressaten- und situationsgerecht [verarbeiten]“ (Biologie K3 und K4; Kultusministerkonferenz, 2005c, S. 15). Aus diesem Grund wurden derartige Aufgaben auch dem Kompetenzbereich Kommunikation zugeordnet.
- Alle Aufgaben, die das Aufstellen, Ergänzen oder Ausgleichen von Wort- und Reaktionsgleichungen erforderten, wurden dem Kompetenzbereich Fachwissen zugeordnet. Auch zählte hierzu das Ergänzen von Oxidationszahlen. Dazu heißt es in den Bildungsstandards Chemie, dass Schülerinnen und Schüler „Reaktionsschemata/ Reaktionsgleichungen durch Anwendung der Kenntnisse über die Erhaltung der Atome und die Bildung konstanter Atomzahlenverhältnisse in Verbindungen [erstellen]“ (Chemie F3.4; Kultusministerkonferenz, 2005a, S. 11).
- Da Schülerinnen und Schüler gemäß den Bildungsstandard Chemie E6 „in erhobenen oder recherchierten Daten, Trends, Strukturen und Beziehungen [finden], [diese] erklären und geeignete Schlussfolgerungen [ziehen]“ (Kultusministerkonferenz, 2005a, S. 12), wurden Aufgaben, in denen Schlussfolgerungen gezogen werden sollten, auch dem Kompetenzbereich Erkenntnisgewinnung zugeordnet.
- Der Kompetenzbereich Erkenntnisgewinnung wurde auch zugeordnet, wenn in einer Aufgabe ein Stammbaum entwickelt werden sollte, da die Bildungsstandards Biologie in E3 vorsehen, dass Schülerinnen und Schüler „die stammesgeschichtliche Verwandtschaft bzw. ökologisch bedingte Ähnlichkeit bei Organismen durch kriteriengeleitetes Vergleichen [analysieren]“ (Kultusministerkonferenz, 2005c, S. 14).
- Verlangte eine Aufgabe die Formulierung von Vermutungen bzw. Hypothesen, wurde sie dem Kompetenzbereich Erkenntnisgewinnung zugeordnet, so wie es die Bildungsstandards Physik vorsehen: Schülerinnen und Schüler „stellen an einfachen Beispielen Hypothesen auf“ (E6; Kultusministerkonferenz, 2005b, S. 11).
- Zu planende Versuchsanordnungen wurden dem Kompetenzbereich Erkenntnisgewinnung zugeordnet, so wie es die Bildungsstandards Biologie E6, Chemie E2 und Physik E8 (Kultusministerkonferenz, 2005a, S. 12, Kultusministerkonferenz, 2005b, S. 11, Kultusministerkonferenz, 2005c, S. 14,) vorsehen. Sollte jedoch eine gegebene Versuchsanordnung mit einer Zeichnung oder Skizze veranschaulicht werden, wurde dies dem Kompetenzbereich Kommunikation zugeordnet, da Schülerinnen und Schüler laut der Bildungsstandards „Ergebnisse und Methoden biologischer Untersuchung dar[stellen] und [damit] argumentieren“ (Biologie K6; Kultusministerkonferenz, 2005c, S. 15) sowie den „Verlauf und die Ergebnisse von Untersuchungen und Diskussionen in angemessener Form [protokollieren]“ (Chemie K7; Kultusministerkonferenz, 2005a, S. 13).

## A.3 Sonderfälle auf der Ebene einzelner Aufgaben

Aufgabe	Kompetenzbereich	Lösungskomplexität	Beispiel	Bildungsstandards
<b>A</b>				
Etwas in einem Diagramm, einer Abbildung oder Zeichnung markieren oder kennzeichnen	<i>Siehe Diagramm.</i>			
Abbildungen, Diagramme u.ä. <b>unbegründet</b> zuordnen	2	2	2012_2rd_TH_CHE_02a 2012_2rd_MC_PHY_07	Chemie E6
Abbildungen, Diagramme u.ä. <b>begründet</b> zuordnen	2	3	Kein Beispiel vorhanden.	
<b>B</b>				
Begründung der Antwort ist in der <b>Lösung</b> gefordert, aber in der Aufgabe <b>nicht</b> verlangt	Je nach Aufgabe.	3	2012_1rd_01c	–
Berechnungen durchführen	1	2	2011_3rd_TH_CHE_03c 2011_3rd_TH_PHY_02c	Physik F2, F3
Beweise (mathematisch) und Berechnungen, die einen Sachverhalt erklären sollen	1	3	2009_2rd_TH_PHY_b	
Beschreiben und Protokollieren von (experimentellen) Beobachtungen (nur TH)	2	2 (TH)	2012_1rd_03a 2012_2rd_TH_CHE_03b	Chemie E3 Physik E8
		1 (MC)	2014_2rd_MC_CHE_02	
Beurteilen, zu etwas (fachlich) begründet Stellung nehmen	3	3	2009_3rd_TH_CHE_04c	Chemie K3, K4, K8
<b>C</b>				
<b>D</b>				
Diagramme ergänzen	3	2	2008_3rd_TH_PHY_01b	Biologie K2, K3 Chemie K4
Etwas in einem Diagramm, einer Abbildung oder Zeichnung markieren oder kennzeichnen	1	Je nach Aufgabe.	2012_3rd_TH_CHE_04b	Fachwissen
Diagramme vollständig erstellen	3	3	2010_3rd_TH_BIO_01d	Biologie K2, K3 Chemie K4
<b>E</b>				
Experimente und unmittelbar zum Experiment zugehörige (Teil-)Aufgaben	2	Je nach Aufgabe.	2014_1rd_02b 2014_1rd_02c	Biologie E6 Chemie E2, E3 Physik E7, E8
<b>F</b>				
<b>G</b>				
<b>H</b>				
Hypothesen formulieren	<i>Siehe Vermutungen.</i>			

Aufgabe	Kompetenzbereich	Lösungskomplexität	Beispiel	Bemerkung
<b>I</b>				
Informationen aus Diagrammen, Zeichnungen, Abbildungen, Tabellen entnehmen, um mit diesen Informationen die Aufgabe zu lösen	2	2 (MC)	2010_2rd_MC_PHY_12	Physik E2
		3 (TH)	2013_2rd_TH_PHY_01a	
Informationen aus Diagrammen, Zeichnungen, Abbildungen, Tabellen entnehmen und aufführen	1	1	2014_2rd_TH_PHY_02c	Physik E2
Informationen recherchieren, um mit diesen Informationen die Aufgabe zu lösen	3	3	2012_1rd_04b 2014_1rd_02e	Chemie K1 Physik K3
<b>J</b>				
<b>K</b>				
<b>L</b>				
Lückentexte	1	2	2009_3rd_TH_BIO_01a	–
<b>M</b>				
<b>N</b>				
Nahrungsnetze entwickeln	3	3	2009_2rd_TH_BIO_d	Biologie K3, K4
<b>O</b>				
Oxidationszahlen in einer Reaktionsgleichung ermitteln	1	2	2013_2rd_TH_CHE_03c	Chemie F3.4
<b>P</b>				
Protokollieren von (experimentellen) Beobachtungen	<i>Siehe Beschreiben.</i>			
<b>Q</b>				
<b>R</b>				
Reaktionsgleichung vollständig aufstellen	1	2	2009_3rd_TH_CHE_01a	Chemie F3.4
Reaktionsgleichung ergänzen <b>und</b> ausgleichen	1	2	2009_2rd_TH_CHE_a	
Reaktionsgleichung ergänzen <b>ohne</b> ausgleichen	1	1	Kein Beispiel vorhanden.	
Reaktionsgleichung nur ausgleichen	1	2	2013_3rd_MC_CHE_25	
Recherche von Informationen um mit diesen Informationen die Aufgabe zu lösen	<i>Siehe Informationen recherchieren.</i>			
<b>S</b>				
Sachverhalt mit einer Abbildung, Zeichnung oder Skizze <b>unbegründet</b> veranschaulichen	3	2	2014_2rd_TH_BIO_02c	Biologie K2
Schlussfolgerungen ziehen, Eigenschaften u.ä. <b>begründet</b> ableiten	2	3	2011_1rd_02b 2013_1rd_01d	Chemie E6
Schlussfolgerungen ziehen, Eigenschaften u.ä. <b>unbe-</b>	2	2	2010_3rd_TH_CHE_03b	

<b>gründet</b> ableiten				
Skizzen erstellen	<i>Siehe Zeichnungen.</i>			
Stammbäume entwickeln	2	3	2012_3rd_TH_BIO_02	Biologie E3
<b>T</b>				
<b>U</b>				
<b>V</b>				
Verfahren (großtechnisch) in einer Zeichnung oder Skizze darstellen	3	2	2008_3rd_TH_CHE_04a 2013_3rd_TH_CHE_01d	Biologie K2, K3 Chemie K4
Vermutungen <b>unbegründet</b> formulieren	2	2	2010_3rd_TH_BIO_03c	Physik E6
Vermutungen <b>begründet</b> formulieren	2	3	2013_2rd_TH_BIO_04c	
Versuchsordnung <b>ohne gegebene Skizze</b> planen	2	3	2013_2rd_TH_CHE_01b	Biologie E6 Chemie E2 Physik E8
Versuchsordnung <b>mit einer gegebenen</b> Skizze planen	2	2	Kein Beispiel vorhanden.	
Versuchsordnung (gegeben) mit einer Zeichnung oder Skizze veranschaulichen	3	2	2013_3rd_TH_BIO_03b	Biologie K6 Chemie K7
Versuchsskizze <b>ohne gegebene Geräte</b> entwickeln bzw. planen	2	3	Kein Beispiel vorhanden.	Biologie E6 Chemie E2 Physik E8
Versuchsskizze <b>mit gegebenen Geräten</b> entwickeln bzw. planen	2	2	2012_3rd_TH_CHE_01a	
Versuchsskizze (gegeben) verschriftlichen	2	2	2012_3rd_TH_CHE_01b 2013_3rd_TH_CHE_03a	
<b>W</b>				
Wortgleichung aufstellen	1	1	2014_2rd_TH_CHE_03a	Chemie F3.4
<b>X</b>				
<b>Y</b>				
<b>Z</b>				
Zeichnungen u.ä. beschriften (auch mit vorgegebenen Begriffen)	1	1	2009_3rd_TH_BIO_01b	Biologie K2, K3 Chemie K4
Zeichnungen und Skizzen erstellen	3	3	2010_3rd_TH_PHY_01a	Physik K8
Zeichnungen und Skizzen <b>begründet</b> ergänzen	3	3	2010_3rd_TH_PHY_03b	Biologie K2, K3 Chemie K4
Zeichnungen und Skizzen <b>unbegründet</b> ergänzen	3	2	2012_2rd_TH_BIO_03	
Etwas in einem Diagramm, einer Abbildung oder Zeichnung markieren oder kennzeichnen	<i>Siehe Diagramm.</i>			

## Anhang B

# Musterlösungen zu den Beispielen von klassifizierten Aufgaben

Die Musterlösungen sind online verfügbar unter  
<http://wettbewerbe.ipn.uni-kiel.de/ijso> (Zugriffsdatum 09.12.2016).

**Experiment 4 – „vom Holz zum Papier“**

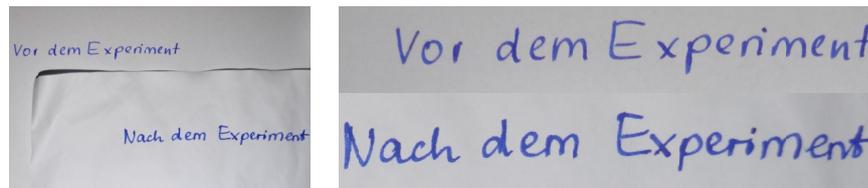
Miss genau die Länge und Breite eines DIN A4-Blattes. Weiche das Blatt für einige Minuten in Wasser ein und bestimme die Größe des noch nassen Blattes. Wiederhole deine Messung, nachdem das Papier getrocknet ist. Schreibe mit Tinte einige Worte auf das so behandelte Blatt sowie auf ein neues DIN A4-Blatt.

**Aufgabe 4:**

4a) Notiere deine Beobachtungen und erkläre die Veränderungen.

[4,0 Punkte]

Beobachtungen: [2,0 P]



DIN A4-Blatt	trocken	nass	getrocknet
Breite in cm	21,0	21,7 (+ 3,3%)	20,8 (-1,0 %)
Länge in cm	29,7	29,5 (-0,7%)	29,4 (-1,0%)

Wässerungsdauer: 2 Minuten; Wassertemperatur 30 °C; Raumtemperatur 21 °C

Werte in Klammern beziehen sich auf ursprüngliche Papiermaße vor dem Einweichen.

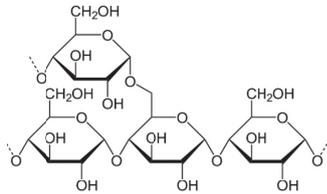
- Das trockene Papierblatt fühlt sich vor dem Einweichen in Wasser an der **Oberfläche glatt** an. Der **Schriftzug** ist nach dem Trocknen der Tinte **scharf**. Das eingeweichte und anschließend wieder getrocknete Papier ist **deutlich gewellt** und die Oberfläche fühlt sich **vergleichsweise rau** an. Die Tinte wird vom Papier aufgesaugt, die **Schriftzüge** wirken **verschwommen** und in den Konturen ausgefranst, ähnlich wie wenn man mit Tinte auf saugfähigem Papier (z. B. Löschpapier, Papiertaschentuch, Küchenrolle) schreibt.
- Die **Länge des Papiers** bleibt nahezu **unverändert** (Abweichung max. 1%), wenn das Blatt nass gemacht und anschließend wieder getrocknet wird.
- Die **Breite des Papiers nimmt deutlich zu** (+ 3,3%), wenn das Blatt nass gemacht wird. **Nach dem Trocknen schrumpft die Breite** wieder und erreicht in etwa wieder den Ausgangszustand.

Erklärungen: [2,0 P]

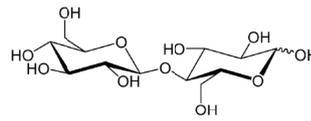
Das DIN A4- Blatt enthält Leime, die das Papier glatter, weniger saugfähig und stärker wasserabweisend machen. Das Blatt ist mit Tinte gut beschreibbar. Wässert man das Papier, wird der Leim teilweise gelöst. In wieder getrocknetem Zustand ist das Papier weniger wasserabweisend und somit saugfähiger. Deshalb verläuft die Tinte etwas beim Beschreiben.

Das Blatt Papier zeigt bezüglich der Quellung bzw. Schrumpfung eine Richtungsabhängigkeit. Es verhält sich anisotrop. Hieraus kann man schließen, dass die im Papier enthaltenen Fasern nicht statistisch beliebig verteilt sind, sondern in einer bestimmten Richtung vorzugsweise ausgerichtet (orientiert) sind. Die Fasern dehnen sich beim Einweichen in Wasser in radialer Richtung stark aus (Quellen), beim Trocknen schrumpfen sie entsprechend und nehmen ihren ursprünglichen Durchmesser wieder ein. In der Längsrichtung hingegen ist ein Dehnen bzw. Schrumpfen vernachlässigbar gering. Da sich das Blatt beim Wässern ausschließlich in der Breite ausdehnt, müssen die Fasern im Papier vorzugsweise mit ihrer Längsachse parallel zur Längsseite im Papier angeordnet sein.

- 4b) Cellulose besteht wie Stärke aus dem Grundbaustein Glucose. Informiere dich über die Struktur von Cellulose und Stärke.



Verzweigter Aufbau eines Stärkemakulärs



Cellobiose: Glucose-Dimer in Sesselkonfiguration

[http://upload.wikimedia.org/wikipedia/commons/7/7e/Amylopektin\\_Haworth.svg](http://upload.wikimedia.org/wikipedia/commons/7/7e/Amylopektin_Haworth.svg)

[http://upload.wikimedia.org/wikipedia/commons/thumb/1/1c/Cellobiose\\_skeletal.png/800px-Cellobiose\\_skeletal.png](http://upload.wikimedia.org/wikipedia/commons/thumb/1/1c/Cellobiose_skeletal.png/800px-Cellobiose_skeletal.png)

Begründe, weshalb sich Stärke für die Papierherstellung nicht eignet.

[2,0 Punkte]

Stärke ist ein **Mehrfachzucker** (Polysaccharid) mit der Formel  $(C_6H_{10}O_5)_n$ , der aus  **$\alpha$ -D-Glucose-Einheiten** besteht, die über **glykosidische Bindungen** miteinander verknüpft sind. Sie besteht zum überwiegenden Anteil (etwa 80%) aus Amylopektin, das keine linearen Ketten, sondern **stark verzweigte Strukturen** bildet. Deshalb bilden Stärkekörner keine Fasern, sondern sehen eher kugelig, oval, linsenförmig aus.

Für die Papierherstellung wird aber faseriges Material (z. B. Holzfasern oder Textilfasern) benötigt.

Cellulose ist wie Stärke ein **Mehrfachzucker** mit der Formel  $(C_6H_{10}O_5)_n + 2 H + O$ , der aus mehreren hundert bis zehntausend **Cellobiose-Einheiten ( $\beta$ -D-Glucose-Dimere)** besteht. Das Makromolekül bildet **Ketten** und ist **unverzweigt**. Die Cellulosemoleküle lagern sich zu höheren Strukturen zusammen, die als **reißfeste Fasern in Pflanzen** häufig festigende Funktionen haben.

- 4c) Verschiedene Prozesse tragen zum „Altern“ von Papier bei. Bibliotheken schützen ihre wertvollen Bücher durch verschiedene Maßnahmen: Sie werden in dunklen und klimatisierten Räumen aufbewahrt und dürfen nur mit Handschuhen berührt werden. Nach dem verheerenden Brand in der Herzogin-Anna-Amalia-Bibliothek in Weimar wurden die nach den Löscharbeiten geretteten Bücher gefriergetrocknet. Nenne vier Faktoren, vor denen diese Maßnahmen schützen sollen, und erläutere die zugehörigen Alterungsprozesse.

[4,0 Punkte]

Die genannten Maßnahmen sollen das Papier vor Einwirkungen durch Licht, Wärme, Luftfeuchtigkeit und Säure schützen.

Erläuterungen:

- 1) Abdunklung – Schutz vor Lichteinwirkung:

Die Einwirkung von **Licht** führt zum Vergilben. Dafür sind vor allem im Papier enthaltene Reste an Lignin verantwortlich. Lignin ist sehr fest bis spröde, hell- bis dunkelbraun gefärbt und absorbiert UV-Anteile im Licht fast vollständig. Lignin denaturiert, wenn es über lange Zeit Licht mit UV-Anteilen ausgesetzt wird.

[1,0 P]

- 2) Klimatisierung – Schutz vor Temperaturschwankungen und zu großer Feuchtigkeit bzw. Trockenheit:

**Wärme** beschleunigt fast alle biologischen und chemischen Prozesse der Papieralterung. [1,0 P]

**Feuchtigkeit** ermöglicht die Ansiedlung von Schädlingen (z. B. Schimmelpilze). Neben der normalen Luftfeuchtigkeit spielen Wasserschäden z. B. auch in Folge der Löscharbeiten bei einem Brand eine Rolle. Feuchtigkeitsschäden können am schonendsten durch Gefrier-trocknen behoben werden. Zu große **Trockenheit** ist ebenfalls nicht förderlich, weil das Papier dann spröde wird und reißen kann.

[1,0 P]

## 3) Das Tragen von Handschuhen – Schutz vor Fingerabdrücken

Fingerabdrücke enthalten **Säureanteile**, die das Blatt beschädigen können. Bei diesem sogenannten „Säurefraß“ bauen Säuren Cellulose ab, indem sie deren Makromolekül verkürzen. Das führt zu einer **Degradation der Cellulosefasern** und einer damit verbundenen **Senkung der Reißfestigkeit**. Die mechanische Festigkeit des Papiers sinkt und führt so zu einer irreversiblen Schädigung des Papiers.

[1,0 P]

*Ergänzung: Auch Restanteile verschiedener saurer Substanzen, die beim Herstellungsprozess verwendet wurden und im Papier verbleiben (z.B. Aluminiumsulfate) oder die Beschriftung mit sauren Tinten, wie sie früher verbreitet Anwendung fanden, verursachen den Säurefraß (siehe auch Aufgabenteil 4d).*

## 4d) Das Papier mit einem Selbstporträt Dürers aus dem 15. Jahrhundert ist einwandfrei erhalten. Jedoch „sterben“ massenweise Bücher und Handschriften aus der zweiten Hälfte des 19. Jahrhunderts.

Finde eine Begründung

[0,5 Punkte]

Die von Dürer verwendeten Papiere wurden noch auf der Basis von Textilabfällen (Lumpen, Hadern) hergestellt.

[0,5 P]

und erläutere die dafür verantwortlichen chemischen Prozesse.

[3,0 Punkte]

Diese Hadern wurden in Wasser mit Soda oder Pottasche eingeweicht und langsam in ihre Fasern zerlegt. Die Restanteile der alkalischen Lösungen reichten aus, um die Säureanteile aus dem Zerfall der in pflanzlichen Fasern enthaltenen Carboxylgruppen zu neutralisieren. Durch das Tauchen in tierische Leime (z. B. Knochenleim, Gelatine) wurden die Papiere mit einem Schutzfilm überzogen und beschreibbar gemacht.

[1,5 P]

**Mitte des 19. Jahrhunderts** erfand Friedrich Gottlob Keller ein Verfahren zur Herstellung von **Papier aus Holzschliff**, welches die Papierherstellung aus Lumpen (Hadern) weitgehend ablöste. Für den Einsatz größerer Mengen von Holz bei der Papierherstellung, musste die Cellulose aus dem zersetzten Holz aufgeschlossen werden, **Restanteile verschiedener saurer Substanzen aus dem chemischen Aufschlussverfahren verbleiben im Papier**. Zusätzlich wurde auch häufig eine **Leimung mit Harzsäuren und Alaun** verwendet. Die über das Alaun eingebrachten Aluminiumsulfate bilden in feuchter Luft durch Hydrolyse Schwefelsäure. Diese ist verantwortlich für braune Stockflecken und einen leicht säuerlichen Geruch, den man von alten Büchern kennt.

[1,5 P]

*Ergänzung: Aber auch Handschriften auf Hadernpapier (Rohmaterial auf der Basis von Stofflumpen), das im Vergleich zum Holzschliffpapier weniger problembehaftet ist, blieben vom Säurefraß nicht vollständig verschont, denn Tinte wurde früher auf Gallapfelbasis mit Eisensulfat hergestellt.*

Formuliere dazu eine chemische Reaktionsgleichung.

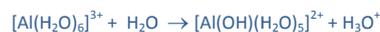
[1,5 Punkte]

Das Lösen von Alaun in Wasser ist im Detail kompliziert und nur in Teilschritten zu beschreiben:

1. Das Aluminium-Ion hydratisiert in wässriger Lösung und bildet einen Hexaaquakomplex, der aber nur bei  $\text{pH} < 3$  stabil ist.



2. Der Hexaaquakomplex des Aluminium-Ions ist eine Kationensäure. Im Bereich  $\text{pH} = 4-7$  findet eine mehrschrittige Hydrolyse statt, wie z. B.:

*Hinweis zur Bewertung:*

*Eine genaue quantitative Beschreibung in dieser Weise wird von den Teilnehmenden nicht erwartet. Volle Punktzahl wird gegeben, falls eine der folgenden Gleichungen stöchiometrisch korrekt formuliert werden.*



oder



[1,5 P]

## Musterlösung

11th International Junior Science Olympiad, Mendoza IJSO 

IJSO 2014 – Aufgaben erste Runde

## Aufgabe 1

(Insgesamt 8,0 Punkte)

1a) Nenne drei tierische und drei pflanzliche Fasern. Gib in einer Tabelle jeweils an, von welchen Tieren, Pflanzen und Pflanzenteilen (z.B. Wurzel, Stängel, Blatt oder Frucht) sie gewonnen werden.

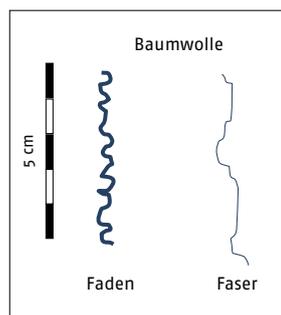
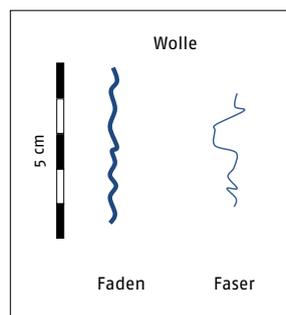
1,5 Punkte

Mögliche Antworten:

Fasertyp	Faser	Tier- oder Pflanzennamen	Pflanzenteil	
tierisch	max. [0,6 P]	Alpaka [0,1]	Alpaka (Kamel) [0,1]	
		Angora [0,1]	Kaninchen [0,1]	
		Kaschmir [0,1]	Kaschmirziege [0,1]	
		Weitere Beispiele:		
		Seide	Raupe o. Seidenspinner	
		Mohair	Ziege	
		Wolle (Schurwolle)	Schaf	
pflanzlich	max. [0,9 P]	Baumwolle [0,1]	Baumwollpflanze oder Gossypium [0,1] Frucht – Samenhaare [0,1]	
		Hanf [0,1]	Hanfpflanze [0,1] Stängel – Bastfaser [0,1]	
		Jute [0,1]	Jute [0,1] Stängel – Bastfaser [0,1]	
		Weitere Beispiele:		
		Kokos	Kokospalme	Fruchthülle
		Leinen/ Flachs	Flachs oder Gemeiner Lein	Stängel- Bastfaser
		Sisal	Agave	Blatt

1b) Führe das Experiment in Aufgabe 1 durch und fertige jeweils eine Skizze der Fäden und Fasern von Baumwoll- und Schurwollzwirn an. Nutze dazu Lupe oder Binokular.

2,0 Punkte



Je korrekter Skizze für Faden bzw. Faser [0,5 P]; insgesamt max. [2,0 P].

- [0,25] Abzug bei fehlendem Maßstab
- [0,25] Abzug bei fehlender Beschriftung

- 1c) Bestimme in beiden Zwirnen die Anzahl der Fäden. Miss die mittlere Faserlänge und vergleiche Oberflächenbeschaffenheit, Dehnbarkeit und Elastizität beider Faserarten. Wo notwendig, nutze Lupe oder Binokular. Fertige eine Vergleichstabelle mit deinen Untersuchungsergebnissen an.

4,5 Punkte

Untersuchte Eigenschaft		Wolle	Baumwolle
<b>Je nach Zwirn verschieden. Bitte vergleichen Sie mit der eingereichten Schülerprobe.</b>			
Anzahl der Fäden	[0,8]	In unserer Probe (Merinowolle): 4 verdrehte Fäden	6 verdrehte Fäden
Faserlänge	[0,8]	<i>viele mehrere Zentimeter lange Fasern</i>	<i>Gemenge aus vielen kürzeren und einigen längeren Fasern</i>
Messergebnisse	[0,5]	3,2 cm / 7,2 cm / 8,3 cm / 9,7 cm / 10,5 cm	2,6 cm / 2,3 cm / 2,9 cm / 2,8 cm
Oberflächenbeschaffenheit	[0,8]	Oberfläche rau, glänzt	Oberfläche glatt, glänzt
Dehnbarkeit	[0,8]	einzelne Faser etwas dehnbar	einzelne Faser kaum dehnbar
Elastizität	[0,8]	elastisch (springt in ursprüngliche Form zurück)	wenig elastisch

**Anmerkung:** Bei gekämmter Baumwolle sind die Fasern länger und elastischer, sehr ähnlich denen der Wolle.

Sind die Beobachtungen und Messergebnisse für Baumwolle und Wolle zu den genannten Eigenschaften nachvollziehbar und angemessen beschrieben, gibt es volle Punktzahl.

- [0,5] Abzug, falls keine Einzelmessungen der Faserlänge aufgeführt sind.

## Anhang C

# Fragebogen für Jury-Mitglieder des Wettbewerbs Jugend forscht

**Fragebogen für Juroren im Wettbewerb  
Jugend forscht / Schüler experimentieren**



In dieser Umfrage möchten wir gerne wissen, für wie wichtig Sie als Jurorin bzw. Juror bestimmte Aspekte für eine gute Leistung im Wettbewerb Jugend forscht / Schüler experimentieren erachten.

**1. Für wie wichtig erachten Sie die folgenden Aspekte für eine gute Leistung im Wettbewerb Jugend forscht?**

	unwichtig	eher unwichtig	eher wichtig	wichtig
Kreativität				
Problemlösefähigkeit				
Intelligenz				
Fachwissen im Bereich des eigenen Projekts				
Wissen, das über das eigene Projekt/ die aktuelle Aufgabenstellung hinausgeht und in anderen, benachbarten Bereichen liegt				
Beherrschen wissenschaftlicher Arbeitsweisen				
Fähigkeit, wissenschaftlich Ergebnisse zu präsentieren				
Fähigkeit, sich selbst darzustellen				
Eigenständigkeit				

**2. Für wie wichtig erachten Sie die folgenden Kompetenzbereiche für eine gute Leistung im Wettbewerb Jugend forscht?**

	unwichtig	eher unwichtig	eher wichtig	wichtig
<b>Fachwissen</b> (Lebewesen, naturwissenschaftliche Phänomene, Begriffe, Prinzipien, Gesetzmäßigkeiten und Fakten kennen)				
<b>Erkenntnisgewinnung</b> (Beobachten, vergleichen, experimentelle und andere Untersuchungsmethoden sowie Modelle nutzen)				
<b>Kommunikation</b> (Informationen sach- und fachbezogen erschließen und austauschen)				
<b>Bewertung</b> (naturwissenschaftliche Sachverhalte in verschiedenen Kontexten erkennen und bewerten)				

**3. Verteilen Sie auf die folgenden Aspekte entsprechend ihrer Bedeutung für eine gute Leistung im Wettbewerb Jugend forscht die Plätze 1 bis 7 (1 = große Bedeutung, 7 = kleine Bedeutung). Sie dürfen Plätze doppelt vergeben!**

Findet viele Ideen und Lösungsvorschläge	
Arbeitet formal korrekt und präzise entsprechend wissenschaftlicher Standards	
Wählt aus verschiedenen Ideen und Lösungsvorschlägen die passende aus	
Kennt sich gut mit Fakten und Verfahren aus dem Umfeld des Problems aus	
Kombiniert, spezialisiert, erweitert und variiert bekannte Methoden, Verfahren und Lösungswege auf ein neues Problem	
Präsentiert Ergebnisse sicher, selbstbewusst und verwendet dabei die Fachsprache	
Ist bereit und in der Lage dazu neue Wege als bisher bekannte zu gehen	

**Fragebogen für Juroren im Wettbewerb  
Jugend forscht / Schüler experimentieren**



4. Verteilen Sie auf die folgenden Teilleistungen entsprechend ihrer Bedeutung für eine gute Leistung im Wettbewerb Jugend forscht die Plätze 1 bis 4 (1 = große Bedeutung, 4 = kleine Bedeutung). Benutzen Sie jeden Platz nur ein Mal!

Projektarbeit	
Standgestaltung	
Präsentation der Ergebnisse	
Jurybefragung	

5. Bitte beurteilen Sie die folgenden Aussagen. Es ist besser, dass der / die Teilnehmende...

	Ich stimme überhaupt nicht zu	Ich stimme eher nicht zu	Ich stimme eher zu	Ich stimme voll zu
Einfache Probleme bearbeitet, die wirklich neu sind, als dass er / sie komplizierte Probleme bearbeitet, die vielleicht gar nicht neu sind.				
Viele unkonventionelle (eigene) Lösungsideen und Methoden entwickelt, die eventuell Schwächen besitzen, als dass er / sie nur wenige Lösungsideen und Methoden entwickelt, die wissenschaftlich üblich und gut anwendbar sind.				
Ein Problem bearbeitet, das er / sie zwar selbst gefunden hat, das aber zu umfangreich ist, als dass er / sie ein Problem bearbeitet, das vom Betreuer vorgegeben wurde und den „richtigen“ Umfang besitzt.				
Ein Problem aus dem Alltag bearbeitet, selbst wenn es formale Mängel aufweist, als dass er / sie ein Alltag-fernes Problem wissenschaftlich nahezu perfekt bearbeitet.				

6. Für wie wichtig erachten Sie die folgenden Aspekte der Präsentation?

	unwichtig	eher unwichtig	eher wichtig	wichtig
Qualität der Darstellung				
Beschränkung auf das Wesentliche				
Strukturiertheit				
Klarheit				
Einhaltung der Rahmenbedingungen (z.B. zeitliche Vorgaben)				
Verdeutlichung des eigenen Arbeitsanteils der Projektmitglieder				
Fachliche Richtigkeit				

**Vielen Dank für Ihre Teilnahme!**

Christine Köhr | koehr@ipn.uni-kiel.de



# Anhang D

## Beispielaufgaben aus Studie 1

### Naturwissenschaftliche Kenntnisse und Fähigkeiten



FB04

Mithilfe von Haaranalysen kann untersucht werden, ob jemand zu einem bestimmten Zeitpunkt unter Einfluss von Drogen oder Medikamenten stand. Dabei sind Haare biologisch gesehen tot, nur die Haarfollikel, in denen die Haarwurzeln gebildet werden, sind am Stoffwechsel beteiligt. Im Gegensatz dazu ist Blut direkt am Stoffwechsel beteiligt. **Nenne begründete Vermutungen, welchen Vorteil Haaranalysen im Gegensatz zu Blutanalysen haben könnten.**

### Musterlösung

In Haaren sind Drogen oder Medikamente länger nachweisbar, denn Haare bzw. Haarspitzen nehmen nicht mehr am Stoffwechsel teil. In ihnen enthaltene Substanzen werden nicht abgebaut, so dass diese Substanzen dort länger nachweisbar sind.

### Kategoriensystem

Kategorie	Schülerantwort	Anforderungen	Punkte
1	Eventuell gelangen bestimmte Stoffe der Drogen in die Haarwurzel und wandern mit dem Haar.	Vermutung wird nicht oder nicht richtig begründet.	1
2	Drogen im Blut werden abgebaut. Da die Haare nicht am Stoffwechsel beteiligt sind werden Stoffe im Haar auch nicht wieder abgebaut. So kann der Stoff an sich oder Folgen der z.B. Droge im Haar nachgewiesen werden.	Vermutung wird korrekt darüber begründet, dass die Haare bzw. Haarspitzen nicht mehr am Stoffwechsel teilnehmen.	2

**Naturwissenschaftliche Kenntnisse und Fähigkeiten**

EC07

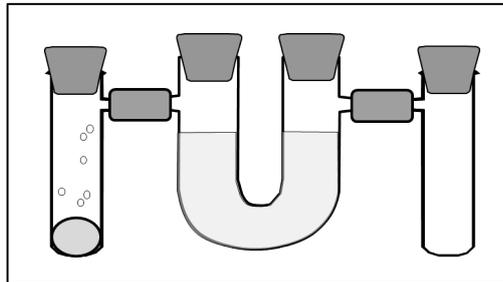
Beim Pizzabacken kommt Nick auf die Idee, die Hefe etwas genauer zu untersuchen. Vor allem möchte er gerne wissen, wie viel Kohlenstoffdioxid beim Gären der Hefe entsteht und welchen Einfluss die Temperatur dabei hat. **Plane ein Experiment, mit dem Nick seine Forschungsfrage beantworten kann.**

**Musterlösung und beispielhafter Versuchsaufbau**1. *Untersuchung, wie viel Kohlenstoffdioxid beim Gären entsteht*

Definierte Menge an Hefeteig herstellen, in ein Reagenzglas geben und verschließen. U-Rohr mit Wasser füllen und an das Hefeteig-Reagenzglas schließen. Das andere Ende mit einem weiteren Reagenzglas verschließen, um überlaufendes Wasser aufzufangen. Wasserstand markieren und Gärung über bestimmten Zeitraum beobachten. Notieren, wie viel Wasser verdrängt wurde und auf Gasvolumen umrechnen. (Eigentlich müsste hier zusätzlich noch bedacht werden, dass sich Kohlenstoffdioxid zu einem gewissen Anteil im Wasser löst.)

2. *Untersuchung des Einflusses der Temperatur*

Wiederholung des Versuches bei verschiedenen Temperaturen.

**Kategoriensystem**

Kategorie	Schülerantwort	Anforderungen	Punkte
1	Man könnte die Hefe luftdicht abisolieren ein Rohr an dieser luftdichten Schüssel befestigen und das Kohlenstoffdioxid in ein Gefäß weiterleiten. // Man baut einen Kohlenstoffdioxidmesser und misst die Temperatur und stellt die Abhängigkeit zueinander in einem Diagramm dar.	Es wird nur ein Versuchsaufbau genannt, aber keine Messmethode. // Es wird nur eine Messmethode genannt, aber kein Versuchsaufbau.	1
2	Nick kann den Teig in verschiedene Behälter packen, die er dann unterschiedlich warm hat. Nach z.B. einer Stunde kann er mit Hilfe meines CO <sub>2</sub> -Messgeräts sehen, wie die Temperatur damit zusammen hängt. // Er könnte die Hefe mit verschiedenen Temperaturen gären lassen. Anschließend fängt er die entstehenden Gase	Es wird ein Versuchsaufbau mit Messung des Volumens des entstehenden Kohlenstoffdioxids ODER der Temperatur beschrieben. // Fachsprachlich ungenaue Formulierung, kein Ansatz, wie das Volumen oder die Temperatur genau gemes-	2

## Naturwissenschaftliche Kenntnisse und Fähigkeiten



	auf. Den Behälter in dem die Gase aufgefangen werden misst man vor und nach dem Versuch. Diesen Versuch führt man mit verschiedenen Temperaturen durch. Dadurch erkennt man, ob die Temperatur Einfluss auf die Entstehung von Kohlenstoffdioxid hat und wie viel jeweils entsteht.	sen werden.	
3	Die Hefe muss in einem geschlossenen Gas sich befinden, wo oben eine kleine Öffnung zum Inneren einer Seite führt. Bei einer bestimmten Temperatur entsteht eine bestimmte Menge an CO <sub>2</sub> , was den Stopper der Spritze hochschiebt. Die Menge kann abgelesen werden. Dies kann man bei mehreren Temperaturen durchführen.	Es wird ein vollständiger Versuchsaufbau mit Messung des Volumens des entstehenden Kohlenstoffdioxids UND der Temperatur beschrieben.	3

KB01

Naturstoffe sind eine wichtige Quelle für die Entwicklung von Arzneimitteln. Oft besitzen Naturstoffe jedoch sehr komplexe Strukturen, deren einzelne Strukturelemente nicht immer wichtig für die therapeutische Wirksamkeit eines Medikaments sind. **Nenne Möglichkeiten, wie Forscherinnen und Forscher Naturstoffe so verändern können, dass sie als Medikament eingesetzt werden können.**

### Musterlösung

Verschiedene Möglichkeiten – Eliminierung einzelner Strukturelemente, Änderung Hydrophilie / Hydrophobie, etc.

### Kategoriensystem

Kategorie	Schülerantwort	Anforderungen	Punkte
1	<ul style="list-style-type: none"> <li>– In andere Aggregatzustände</li> <li>– Mit anderen Stoffen verbinden</li> </ul>	Stoff selbst wird nicht verändert, sondern nur Darreichungsform, Herstellungsbedingungen, o.ä. // Ein ganz neuer Stoff wird entwickelt. // Antwort zu ungenau.	1
2	Der Mediziner könnte die wichtigen Abschnitte des Wirkstoffs extrahieren und diese mit neuen Abschnitten kombinieren, um z.B. die Verträglichkeit zu verbessern.	Modifikation zu anderen Stoffen durch Entfernen von Elementen und ähnliches.	2
3	Sie können die Oberflächenstruktur minimal ändern um den Platz für die ähnlich aussehenden Moleküle zu blockieren (z.B. bei Krebs). Sie können bestimmte Moleküle in andere einschleusen, die die Medikamente dann zu dem gewünschten Platz bringen.	Modifikation zu anderen Stoffen durch Entfernen von Elementen und ähnliches, dabei Nennen konkreter Beispiele.	3



# Anhang E

## Fragebogen aus Studie 1

### Fragebogen zum individuellen Konzept über die Naturwissenschaften



Liebe Teilnehmerin, lieber Teilnehmer,

im Rahmen eines Projekts zur Charakterisierung naturwissenschaftlicher Talente führt die Abteilung Didaktik der Chemie des Leibniz-Instituts für die Pädagogik der Naturwissenschaften und Mathematik (IPN) Erhebungen durch. Dabei möchten wir gerne wissen, welche naturwissenschaftlichen Überzeugungen, Interessen sowie Kenntnisse und Fähigkeiten du hast.

Um zu erfahren, welche Überzeugungen, Interessen sowie Erfahrungen mit naturwissenschaftlichen Wettbewerben du hast, bitten wir dich, diesen Fragebogen auszufüllen. Einige Fragen scheinen sich zu ähneln. Wundere dich darüber nicht sondern versuche, genau zu lesen. Gerade die kleinen Unterschiede sind für uns sehr wichtig! Bei einigen Aufgaben sind auch Mehrfachnennungen möglich.

Die Teilnahme ist für dich freiwillig. Du hast keine Nachteile, wenn du nicht mitmachen möchtest, und kannst die Teilnahme jederzeit ohne Angabe von Gründen abbrechen. Deine Mitarbeit würde uns jedoch sehr bei unserer Arbeit helfen, den Unterricht in den naturwissenschaftlichen Fächern und außerschulische Angebote (wie zum Beispiel Wettbewerbe wie die *Internationale JuniorScienceOlympiade* oder *Jugend forscht*) weiter zu verbessern.

**Vielen Dank für deine Hilfe!**

Christine Köhr

**Deine Antworten bleiben anonym. Schreibe deshalb bitte nicht deinen Namen auf den Fragebogen, sondern trage hier bitte deinen Code ein.**

Kreuz bitte beim vorletzten Buchstaben des Vornamens deiner Mutter.																														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte beim letzten Buchstaben des Vornamens deiner Mutter.																														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte beim letzten Buchstaben deines Vornamens.																														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte bei deinem Geburtstag.																														
01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31

Beispiel: Vorname der Mutter CHRISTIANE, dein Vorname MIRIAM, dein Geburtstag 02. Mai.

Kreuz bitte beim vorletzten Buchstaben des Vornamens deiner Mutter.																														
A	B	C	D	E	F	G	H	I	J	K	L	M	<input checked="" type="checkbox"/>	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte beim letzten Buchstaben des Vornamens deiner Mutter.																														
A	B	C	D	<input checked="" type="checkbox"/>	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte beim letzten Buchstaben deines Vornamens.																														
A	B	C	D	E	F	G	H	I	J	K	L	<input checked="" type="checkbox"/>	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte bei deinem Geburtstag.																														
01	<input checked="" type="checkbox"/>	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Allgemeine Fragen**



1. Du bist ein  Mädchen und  13  14  15  16 Jahre alt.  
 Junge
2. Du besuchst zurzeit ein / eine  Gymnasium  Gesamtschule  Spezialschule  Privatschule  sonstige Schule.
3. In welchem Bundesland gehst du zur Schule?
- |  |   |  |   |
|--|---|--|---|
| <input type="checkbox"/> Baden-Württemberg | <input type="checkbox"/> Bremen                 | <input type="checkbox"/> Niedersachsen       | <input type="checkbox"/> Sachsen            |
| <input type="checkbox"/> Bayern            | <input type="checkbox"/> Hamburg                | <input type="checkbox"/> Nordrhein-Westfalen | <input type="checkbox"/> Sachsen-Anhalt     |
| <input type="checkbox"/> Berlin            | <input type="checkbox"/> Hessen                 | <input type="checkbox"/> Rheinland-Pfalz     | <input type="checkbox"/> Schleswig-Holstein |
| <input type="checkbox"/> Brandenburg       | <input type="checkbox"/> Mecklenburg-Vorpommern | <input type="checkbox"/> Saarland            | <input type="checkbox"/> Thüringen          |

4. Trage deine letzten Zeugnisnoten in die Tabelle ein und kreuze an, wie sehr dich die jeweiligen Schulfächer interessieren.

Schulfach	Note	Das interessiert mich gar nicht; das tue ich nicht gerne	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr; das tue ich sehr gerne
Deutsch	1 2 3 4 5 6				
Englisch	1 2 3 4 5 6				
Mathematik	1 2 3 4 5 6				
Biologie	1 2 3 4 5 6				
Chemie	1 2 3 4 5 6				
Physik	1 2 3 4 5 6				
Musik	1 2 3 4 5 6				
Kunst	1 2 3 4 5 6				
Wirtschaft / Politik o. Recht	1 2 3 4 5 6				
Geschichte	1 2 3 4 5 6				
Erkunde	1 2 3 4 5 6				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Wettbewerben**



Im Folgenden möchten wir gerne wissen, welche Erfahrungen du mit naturwissenschaftlichen Wettbewerben hast.

1. Welche der folgenden naturwissenschaftlichen Wettbewerbe kennst du?
- |  |   |
|--|---|
| <input type="checkbox"/> BundesUmweltWettbewerb                  | <input type="checkbox"/> Bundesweiter Wettbewerb Physik für die Sekundarstufe 1                         |
| <input type="checkbox"/> Internationale JuniorScienceOlympiade   | <input type="checkbox"/> Internationale Chemie-Olympiade  |
| <input type="checkbox"/> Jugend forscht/ Schüler experimentieren | <input type="checkbox"/> Internationale Biologie-Olympiade  |
| <input type="checkbox"/> Bundeswettbewerb Mathematik             | <input type="checkbox"/> Bundeswettbewerb Informatik  |
| <input type="checkbox"/> Mathematik-Olympiaden                   | <input type="checkbox"/> Europäische ScienceOlympiade   |
| <input type="checkbox"/> Bundeswettbewerb Naturwissenschaften    | <input type="checkbox"/> Wettbewerb der Siemens Stiftung in Mathematik, Naturwissenschaften und Technik |
| <input type="checkbox"/> Internationale Physik-Olympiade         |   |
2. Kreuze an, welche der folgenden Dinge du gerne für die Teilnahme an einem Wettbewerb gewinnen würdest.

	Auf keinen Fall	Eher nein	Eher ja	Ja, sehr gern
Auszeichnungen (Urkunden, Medaillen) .....				
Geldpreise .....				
Büchergutscheine .....				
Naturwissenschaftliche Sachpreise (Mikroskope, Experimentierkästen, usw.) .....				
Praktika .....				
Exkursionen (Forschungseinrichtungen, Schülerlabore, usw.) .....				
Teilnahme an Veranstaltungen (z. B. Ausstellungen, Messen, weiterführende Wettbewerbe) .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Wettbewerben**



3. Im Folgenden findest du einige Aussagen zu *Jugend forscht* aufgeführt. Kreuze an, wie sehr du den Aussagen zustimmst.

	ich stimme überhaupt nicht zu	ich stimme eher nicht zu	ich stimme eher zu	ich stimme voll zu
1. Wenn ich bei Jugend forscht teilnehme, steigt mein Ansehen in der Klasse. ....				
2. Bei Jugend forscht kann ich meine Fähigkeiten und Talente beweisen. ....				
3. Ich hätte schon eine Idee für ein Jugend forscht Projekt, aber ich finde keinen Betreuer. ....				
4. Bei Jugend forscht finde ich heraus, ob Naturwissenschaften und Technik auch im Beruf und Studium Spaß bringen können.				
5. Jugend forscht ist nur etwas für Hochbegabte. ....				
6. Bei Jugend forscht werde ich in den naturwissenschaftlichen Fächern besonders gefördert. ....				

4. Im Folgenden findest du einige Aussagen zur *Internationalen JuniorScienceOlympiade (IJSO)* aufgeführt. Kreuze an, wie sehr du den Aussagen zustimmst.

	ich stimme überhaupt nicht zu	ich stimme eher nicht zu	ich stimme eher zu	ich stimme voll zu
1. Wenn ich an der IJSO teilnehme, steigt mein Ansehen in der Klasse. ....				
2. Bei der IJSO kann ich meine Fähigkeiten und Talente beweisen. ....				
3. Ich würde gerne an der IJSO teilnehmen, aber ich finde keinen Betreuer. ....				
4. Bei der IJSO finde ich heraus, ob Naturwissenschaften und Technik auch im Beruf und Studium Spaß bringen können.				
5. Die IJSO ist nur etwas für Hochbegabte. ....				
6. Wenn ich an der IJSO teilnehme, werde ich in den naturwissenschaftlichen Fächern besonders gefördert. ....				

5. Hast du schon an einem naturwissenschaftlichen Wettbewerb teilgenommen?

Ja  Nein

Wenn du schon an einem naturwissenschaftlichen Wettbewerb teilgenommen hast, beantworte bitte auch die folgenden Fragen. Wenn nicht, kannst du die Fragen 6 bis 9 überspringen und auf Seite 5 weitermachen.

3

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Wettbewerben**



6. An welchen der folgenden naturwissenschaftlichen Wettbewerbe hast du schon teilgenommen?

- |  |   |   |
|--|---|---|
| <input type="checkbox"/> BundesUmweltWettbewerb                  | <input type="checkbox"/> Internationale Physik-Olympiade                        | <input type="checkbox"/> Europäische ScienceOlympiade   |
| <input type="checkbox"/> Internationale JuniorScienceOlympiade   | <input type="checkbox"/> Bundesweiter Wettbewerb Physik für die Sekundarstufe 1 | <input type="checkbox"/> Wettbewerb der Siemens Stiftung in Mathematik, Naturwissenschaften und Technik |
| <input type="checkbox"/> Jugend forscht/ Schüler experimentieren | <input type="checkbox"/> Internationale Chemie-Olympiade                        |   |
| <input type="checkbox"/> Bundeswettbewerb Mathematik             | <input type="checkbox"/> Internationale Biologie-Olympiade                      |   |
| <input type="checkbox"/> Mathematik-Olympiaden                   | <input type="checkbox"/> Bundeswettbewerb Informatik                            |   |
| <input type="checkbox"/> Bundeswettbewerb Naturwissenschaften    |   |   |

7. Was war dein erster naturwissenschaftlicher Wettbewerb, an dem du teilgenommen hast?

- |  |   |   |
|--|---|---|
| <input type="checkbox"/> BundesUmweltWettbewerb                  | <input type="checkbox"/> Internationale Physik-Olympiade                        | <input type="checkbox"/> Europäische ScienceOlympiade   |
| <input type="checkbox"/> Internationale JuniorScienceOlympiade   | <input type="checkbox"/> Bundesweiter Wettbewerb Physik für die Sekundarstufe 1 | <input type="checkbox"/> Wettbewerb der Siemens Stiftung in Mathematik, Naturwissenschaften und Technik |
| <input type="checkbox"/> Jugend forscht/ Schüler experimentieren | <input type="checkbox"/> Internationale Chemie-Olympiade                        |   |
| <input type="checkbox"/> Bundeswettbewerb Mathematik             | <input type="checkbox"/> Internationale Biologie-Olympiade                      |   |
| <input type="checkbox"/> Mathematik-Olympiaden                   | <input type="checkbox"/> Bundeswettbewerb Informatik                            |   |
| <input type="checkbox"/> Bundeswettbewerb Naturwissenschaften    |   |   |

8. Wie oft hast du schon an einem naturwissenschaftlichen Wettbewerb teilgenommen?

- |                                |                                |                                |
|--------------------------------|--------------------------------|--------------------------------|
| <input type="checkbox"/> 1 Mal | <input type="checkbox"/> 3 Mal | <input type="checkbox"/> 5 Mal |
| <input type="checkbox"/> 2 Mal | <input type="checkbox"/> 4 Mal | <input type="checkbox"/> 6 Mal |

9. Wer oder was hat dich dazu bewegt, an einem naturwissenschaftlichen Wettbewerb teilzunehmen (mehrere Nennungen sind möglich)?

- |   |   |   |
|---|---|---|
| <input type="checkbox"/> Arbeitsgemeinschaft (AG) an der Schule | <input type="checkbox"/> Meine Eltern oder Großeltern         | <input type="checkbox"/> Die Aussicht auf Erfolg                      |
| <input type="checkbox"/> Hochbegabtenförderprogramm             | <input type="checkbox"/> Meine Mitschüler/ innen oder Freunde | <input type="checkbox"/> Ein Bericht in der Zeitung                   |
| <input type="checkbox"/> Mein Lehrer/ meine Lehrerin            | <input type="checkbox"/> Die zu gewinnenden Preise            | <input type="checkbox"/> Die Möglichkeit meine Ideen zu verwirklichen |

4

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zum Selbstkonzept**



Der nächste Fragebogenteil besteht aus einer Reihe unterschiedlicher Aussagen. Kreuze bitte an, inwieweit du den Aussagen zustimmen kannst.

	Ich stimme überhaupt nicht zu	Ich stimme eher nicht zu	Ich stimme eher zu	Ich stimme voll zu
1. Ich bin für die Schule sehr begabt. ....				
2. Neues zu lernen fällt mir schwer. ....				
3. Ich bin sehr intelligent. ....				
4. Ich kann in der Schule viel. ....				
5. In der Schule fallen mir viele Aufgaben schwer. ....				

	Ich stimme überhaupt nicht zu	Ich stimme eher nicht zu	Ich stimme eher zu	Ich stimme voll zu
1. Naturwissenschaften würde ich viel lieber machen, wenn das Fach/ die Fächer nicht so schwer wäre/n. ....				
2. Naturwissenschaften liegen mir nicht besonders. ....				
3. Naturwissenschaften fallen mir schwerer als vielen meiner Mitschüler/innen. ....				
4. Bei manchen Themen im Naturwissenschaftsunterricht weiß ich gleich: „Das verstehe ich nie.“ ....				
5. Für Naturwissenschaften habe ich einfach keine Begabung. ....				
6. Ich denke, ich bin für den Naturwissenschaftsunterricht begabter als viele meiner Mitschüler/innen. ....				
7. Mit den Aufgaben im Naturwissenschaftsunterricht komme ich besser zurecht als viele meiner Mitschüler/innen. ....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zum naturwissenschaftlichen Unterricht**



Auch dieser Fragebogenteil besteht aus einer Reihe unterschiedlicher Aussagen. Kreuze auch hier bitte an, inwieweit du den Aussagen zustimmen kannst.

**Im Naturwissenschaftsunterricht geht es mir darum, ...**

	Ich stimme überhaupt nicht zu	Ich stimme eher nicht zu	Ich stimme eher zu	Ich stimme voll zu
1. etwas Interessantes zu lernen. ....				
2. komplizierte Inhalte zu verstehen. ....				
3. die Dinge, die wir im Unterricht besprechen, wirklich zu verstehen. ....				
4. so viel wie möglich zu lernen. ....				
5. zu zeigen, dass ich bei einer Sache gut bin. ....				
6. bessere Noten zu bekommen als andere. ....				
7. Klassenarbeiten besser zu schaffen als andere. ....				
8. dass andere denken, dass ich klug bin. ....				
9. zu verheimlichen, wenn ich weniger weiß als andere. ....				
10. nicht zu zeigen, falls ich weniger schlau bin als andere. ....				
11. mich nicht zu blamieren (z.B. durch falsche Ergebnisse oder dumme Fragen). ....				
12. dass niemand merkt, wenn ich etwas nicht verstehe. ....				
13. mich nicht zu sehr anstrengen zu müssen. ....				
14. mit wenig Arbeit durch die Schule zu kommen. ....				
15. keine schwierigen Fragen oder Aufgaben lösen zu müssen. ....				
16. zu Hause keine Arbeiten erledigen zu müssen. ....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



Im folgenden Fragebogenteil findest du verschiedene **Tätigkeiten, die häufig in der Schule ausgeübt** werden und die du dort vielleicht auch schon ausgeübt hast. Kreuze bitte dein Interesse an diesen Tätigkeiten an und ob du diese Tätigkeiten schon einmal ausgeübt hast. **Interessieren bedeutet dabei: etwas gerne tun, etwas wegen der Sache selbst tun.**

	Das interessiert mich <b>gerade</b> ; das tue ich <b>nicht gerne</b>	Das interessiert mich <b>wenig</b>	Das interessiert mich <b>ziemlich</b>	Das interessiert mich <b>sehr</b> ; das tue ich <b>sehr gerne</b>	Schon gemacht?	
					ja	nein
1. Untersuchungen eines Gewässers durchführen.....						
2. Unterschiedliche Quellen nach wichtigen Informationen durchsuchen.....						
3. Ohne Anleitung etwas Neues kreieren (z.B. Experimente, Zeichnungen).....						
4. Selbst andere Schülerinnen und Schüler unterrichten.....						
5. Ergebnisse eines Schulprojekts öffentlich vorstellen.....						
6. Die Klassenkasse verwalten.....						
7. Zusammen mit anderen an einer Aufgabe arbeiten.....						
8. Kristalle nach einer Vorschrift züchten.....						
9. Eigene Experimente planen.....						
10. Artikel / Reportagen für die Schülerzeitung schreiben.....						
11. Eine AG betreuen.....						
12. Andere beim Experimentieren anleiten und beaufsichtigen.....						
13. In der Schülerbücherei mitarbeiten.....						

7

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



	Das interessiert mich <b>gerade</b> ; das tue ich <b>nicht gerne</b>	Das interessiert mich <b>wenig</b>	Das interessiert mich <b>ziemlich</b>	Das interessiert mich <b>sehr</b> ; das tue ich <b>sehr gerne</b>	Schon gemacht?	
					ja	nein
14. Selbstständig Kontakt zu Experten aufnehmen.....						
15. Produkte im Chemieunterricht herstellen (z.B. Kosmetika, Farbstoffe).....						
16. In Fachbüchern / Zeitschriften Informationen suchen.....						
17. In einer Musik-AG aktiv sein.....						
18. Ein Computerprogramm schreiben.....						
19. Nachhilfeunterricht geben.....						
20. Die Organisation eines Projektes übernehmen.....						
21. Das Klassenbuch führen.....						
22. Aktiv an Diskussionen im Unterricht teilnehmen.....						
23. An einer Werken-AG teilnehmen.....						
24. Knifflige „Aufgaben der Woche“ lösen.....						
25. Die Schule künstlerisch gestalten / verschönern.....						
26. Im Schulsanitätsdienst tätig sein.....						
27. Sich an einer Schülerfirma beteiligen.....						
28. Die Sammlung mit betreuen.....						
29. In einer Gruppe ein Problem lösen.....						
30. Einen Roboter programmieren.....						

8

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



Der folgende Fragebogenteil besteht aus einer Liste verschiedener **Tätigkeiten, die im naturwissenschaftlichen Schulunterricht, in naturwissenschaftlichen Fördermaßnahmen oder auch in einem naturwissenschaftlichen Beruf ausgeübt** werden können. Gib für jede einzelne davon an, wie sehr dich deren Ausübung interessiert bzw. interessieren würde. **Interessieren bedeutet dabei: etwas gerne tun, etwas wegen der Sache selbst tun.**

Erläuterung:

<i>In der Schule</i>	Wie sehr bist du an der Ausübung dieser Tätigkeit in der Schule interessiert?
<i>In einer Fördermaßnahme</i>	Wie sehr bist du an der Ausübung dieser Tätigkeit in einer Fördermaßnahme interessiert?  (Fördermaßnahmen sind z.B. AGs, Wettbewerbe, Enrichment-Kurse, etc.)  Naturwissenschaftliche Fördermaßnahmen haben das Ziel, interessierte oder auch talentierte Jugendliche besonders in den Naturwissenschaften zu fördern und zu fordern.
<i>Im Beruf</i>	Wie sehr bist du an der Ausübung dieser Tätigkeit in deinem späteren Beruf interessiert?

**Blättere jetzt bitte um und beginne mit der Bearbeitung!**

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



Das interessiert mich gar nicht; das tue ich nicht gerne	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr; das tue ich sehr gerne
--	-----------------------------	--------------------------------	--

1. Mit Laborgeräten oder Experimentier-Sets arbeiten				
<i>In der Schule</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>In einer Fördermaßnahme</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Im Beruf</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Naturwissenschaftliche Fachliteratur lesen und interpretieren				
<i>In der Schule</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>In einer Fördermaßnahme</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Im Beruf</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Naturwissenschaftliche Inhalte nach künstlerischen Gesichtspunkten darstellen				
<i>In der Schule</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>In einer Fördermaßnahme</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Im Beruf</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Andere Menschen in den Naturwissenschaften unterrichten				
<i>In der Schule</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>In einer Fördermaßnahme</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Im Beruf</i> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



	Das interessiert mich gar nicht; das tue ich nicht gerne	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr; das tue ich sehr gerne
5. Eine naturwissenschaftliche Arbeitsgruppe leiten				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
6. Chemikalienbestände sortieren und verwalten				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
7. Sich mit anderen Menschen über naturwissenschaftliche Inhalte austauschen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
8. Versuchsaufbauten zeichnen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



	Das interessiert mich gar nicht; das tue ich nicht gerne	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr; das tue ich sehr gerne
9. Das Verhalten / die Eigenschaften von Stoffen untersuchen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
10. Romane / Geschichten mit naturwissenschaftlichen Inhalten lesen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
11. Andere Menschen über naturwissenschaftliche Inhalte aufklären				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
12. Andere beim Experimentieren beaufsichtigen oder kontrollieren				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



	Das interessiert mich gar nicht; das tue ich nicht gerne	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr; das tue ich sehr gerne
13. Messwerte in ein Computerprogramm eingeben				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
14. Einen naturwissenschaftlichen Vortrag halten				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
15. Geräte für naturwissenschaftliche Messungen reinigen / reparieren				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
16. Ein Modell entwickeln				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



	Das interessiert mich gar nicht; das tue ich nicht gerne	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr; das tue ich sehr gerne
17. Eine naturwissenschaftliche Fachsprache lernen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
18. Mithilfe der Naturwissenschaften etwas „Gutes“ für Menschen tun				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
19. Für das Ansehen der Naturwissenschaften in der Öffentlichkeit auftreten				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
20. Insekten, Kristalle o.ä. sammeln und ordnen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



	Das interessiert mich gar nicht; das tue ich nicht gerne	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr; das tue ich sehr gerne
21. Mit anderen eine naturwissenschaftliche Fragestellung diskutieren				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
22. Experimente nach einer Anleitung durchführen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
23. Die Ursachen eines Phänomens erforschen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
24. Naturwissenschaftliche Präparate zeichnen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu naturwissenschaftlichen Interessen**



	Das interessiert mich gar nicht; das tue ich nicht gerne	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr; das tue ich sehr gerne
25. Die Versorgung mit Medikamenten durch die Naturwissenschaften verbessern				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
26. Mit Sponsoren / Geldgebern verhandeln				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
27. Die Einhaltung von Sicherheitsvorschriften überwachen				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
28. Einen naturwissenschaftlichen Vortrag hören				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu Eigenschaften und Tätigkeiten eines Naturwissenschaftlers**

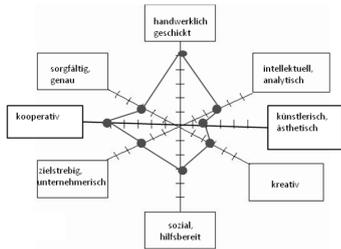


Wie wichtig sind die folgenden Eigenschaften für einen Naturwissenschaftler?

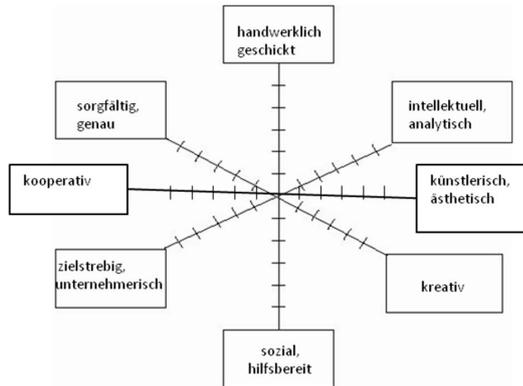
Kreuze auf der jeweiligen Achse an, ob du sie für sehr wichtig (Kreuz außen) oder weniger wichtig (Kreuz innen) hältst. Verbinde anschließend die markierten Punkte zu einem Netz (siehe Beispiel: Eigenschaften eines Fußballers).

**Beispiel: Eigenschaften eines Fußballers**

Eine Testperson wurde befragt, welche Eigenschaften einen guten Fußballer auszeichnen. Sie hielt das handwerkliche Geschick und die Kooperation für besonders wichtig; am unwichtigsten war der Testperson das Künstlerische bei der Beschreibung eines Fußballers (siehe unten).



**Eigenschaften eines Naturwissenschaftlers**



**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu Eigenschaften und Tätigkeiten eines Naturwissenschaftlers**



Der letzte Fragebogenteil besteht aus einer Reihe unterschiedlicher Aussagen. Kreuze an, inwieweit du den Aussagen zustimmen kannst.

Ein Wissenschaftler übt folgende Tätigkeiten regelmäßig aus...

	Stimmt gar nicht	Stimmt eher nicht	Stimmt eher	Stimmt genau
1. Beobachtungen in der Natur durchführen.....				
2. Neue Messgeräte entwickeln.....				
3. Studenten betreuen.....				
4. Lehrveranstaltungen vorbereiten und durchführen.....				
5. Zeitschriftenartikel verfassen.....				
6. Abrechnungen für verwendete Forschungsgelder fertigstellen.....				
7. Austausch mit Wissenschaftlern anderer Universitäten.....				
8. Stoffe (zum Beispiel Medikamente) im Labor herstellen.....				
9. Eigene Theorien entwickeln.....				
10. Ergebnisse aus Experimenten auswerten und interpretieren.....				
11. Interviews geben.....				
12. An eigenen Erfindungen arbeiten.....				
13. Fachliteratur lesen.....				
14. Untersuchungen durchführen (zum Beispiel Lebensmittel oder Gewässer) .....				
15. Finanzen planen und verwalten.....				
16. Eigene Messverfahren entwickeln.....				
17. An Büchern schreiben.....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu Eigenschaften und Tätigkeiten eines Naturwissenschaftlers**



	Stimmt gar nicht	Stimmt eher nicht	Stimmt eher	Stimmt genau
18. Modelle entwickeln.....				
19. Ein Team aufbauen und managen.....				
20. Verwaltungsaufgaben erledigen.....				
21. Fächerübergreifende Projekte durchführen.....				
22. Doktoranden betreuen.....				
23. Zeichnungen, Fotografien oder Protokolle anfertigen.....				
24. Eigene Versuchsapparaturen konstruieren.....				
25. Ein Projekt organisieren und leiten.....				
26. Treffen mit Kollegen anderer Fachbereiche durchführen.....				
27. Abschlussarbeiten von Studenten begleiten.....				
28. Messungen durchführen.....				
29. Eine Arbeitsgruppe leiten.....				
30. Ideen für neue Forschungsansätze entwickeln.....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften**  
**Fragen zu Eigenschaften und Tätigkeiten eines Naturwissenschaftlers**



Würdest du die folgende Tätigkeit später selbst gern durchführen?

	Auf keinen Fall	Eher nein	Eher ja	Ja, sehr gern
1. Beobachtungen in der Natur durchführen.....				
2. Neue Messgeräte entwickeln.....				
3. Studenten betreuen.....				
4. Lehrveranstaltungen vorbereiten und durchführen.....				
5. Zeitschriftenartikel verfassen.....				
6. Abrechnungen für verwendete Forschungsgelder fertigstellen.....				
7. Austausch mit Wissenschaftlern anderer Universitäten.....				
8. Stoffe (zum Beispiel Medikamente) im Labor herstellen.....				
9. Eigene Theorien entwickeln.....				
10. Ergebnisse aus Experimenten auswerten und interpretieren.....				
11. Interviews geben.....				
12. An eigenen Erfindungen arbeiten.....				
13. Fachliteratur lesen.....				
14. Untersuchungen durchführen (zum Beispiel Lebensmittel oder Gewässer).....				
15. Finanzen planen und verwalten.....				
16. Eigene Messverfahren entwickeln.....				
17. An Büchern schreiben.....				
18. Modelle entwickeln.....				
19. Ein Team aufbauen und managen.....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu Eigenschaften und Tätigkeiten eines Naturwissenschaftlers**



	Auf keinen Fall	Eher nein	Eher ja	Sehr gern
20. Verwaltungsaufgaben erledigen.....				
21. Fächerübergreifende Projekte durchführen.....				
22. Doktoranden betreuen.....				
23. Zeichnungen, Fotografien oder Protokolle anfertigen.....				
24. Eigene Versuchsapparaturen konstruieren.....				
25. Ein Projekt organisieren und leiten.....				
26. Treffen mit Kollegen anderer Fachbereiche durchführen.....				
27. Abschlussarbeiten von Studenten begleiten.....				
28. Messungen durchführen.....				
29. Eine Arbeitsgruppe leiten.....				
30. Ideen für neue Forschungsansätze entwickeln.....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu Eigenschaften und Tätigkeiten eines Naturwissenschaftlers**



Schätze deine Fähigkeit für die spätere Ausübung dieser Tätigkeit ein.

	Sehr schlecht	Eher schlecht	Eher gut	Sehr gut
1. Beobachtungen in der Natur durchführen.....				
2. Neue Messgeräte entwickeln.....				
3. Studenten betreuen.....				
4. Lehrveranstaltungen vorbereiten und durchführen.....				
5. Zeitschriftenartikel verfassen.....				
6. Abrechnungen für verwendete Forschungsgelder fertigstellen.....				
7. Austausch mit Wissenschaftlern anderer Universitäten.....				
8. Stoffe (zum Beispiel Medikamente) im Labor herstellen.....				
9. Eigene Theorien entwickeln.....				
10. Ergebnisse aus Experimenten auswerten und interpretieren.....				
11. Interviews geben.....				
12. An eigenen Erfindungen arbeiten.....				
13. Fachliteratur lesen.....				
14. Untersuchungen durchführen (zum Beispiel Lebensmittel oder Gewässer).....				
15. Finanzen planen und verwalten.....				
16. Eigene Messverfahren entwickeln.....				
17. An Büchern schreiben.....				
18. Modelle entwickeln.....				
19. Ein Team aufbauen und managen.....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu Eigenschaften und Tätigkeiten eines Naturwissenschaftlers**



	Sehr schlecht	Eher schlecht	Eher gut	Sehr gut
20. Verwaltungsaufgaben erledigen.....				
21. Fächerübergreifende Projekte durchführen.....				
22. Doktoranden betreuen.....				
23. Zeichnungen, Fotografien oder Protokolle anfertigen.....				
24. Eigene Versuchsapparaturen konstruieren.....				
25. Ein Projekt organisieren und leiten.....				
26. Treffen mit Kollegen anderer Fachbereiche durchführen.....				
27. Abschlussarbeiten von Studenten begleiten.....				
28. Messungen durchführen.....				
29. Eine Arbeitsgruppe leiten.....				
30. Ideen für neue Forschungsansätze entwickeln.....				

**Geschafft!**  
**Vielen Dank für deine Hilfe!!!**



## Anhang F

# Itemparameter, Itemfit und Itemstatistiken (Studie 1)

Tabelle F.1: Schwellenparameter in Logits

<i>Item</i>	<i>Kategorie 1</i>	<i>Kategorie 2</i>	<i>Kategorie 3</i>
EB01	-1.19	3.15	
EB02	-0.68	-0.36	1.74
EB05	0.09	0.22	1.23
EB10	-2.04	-0.62	
FB02	0.38		
FB04	-0.48	1.92	
FB07	-0.02		
FB11	-0.14	0.68	2.31
FB14	0.01	0.52	
KB01	-0.39	0.63	1.24
EC02	-2.25	2.83	
EC03	0.30	-1.21	1.46
EC04	-1.12	1.78	
EC07	-0.33	0.37	1.49
FC02	0.17		
FC08	0.03		
FC10	1.39		
FC12	-0.22	1.08	
KC01i	1.79	1.98	
KC01o	0.89	1.59	
EP02	-0.41	0.98	
EP03	-0.22	0.06	
EP09	-0.95	0.27	0.14
EP11	-1.76	2.50	
FP02	1.59		
FP05	-0.06	0.03	0.38
FP06	-2.11		
FP07	-0.23	0.74	
FP12	1.51	-0.27	
KP01	-0.61	-0.57	2.48

Tabelle F.2: Mean-Squared-Infit und Infit-*t*

<i>Item</i>	<i>Mean-Squared-Infit</i>	<i>Infit-t</i>
EB01	0.99	-0.01
EB02	1.25	2.31
EB05	1.16	1.46
EB10	1.10	0.93
FB02	0.96	-0.59
FB04	0.83	-1.68
FB07	0.93	-1.33
FB11	0.95	-0.48
FB14	0.77	-2.68
KB01	1.00	0.06
EC02	1.05	0.37
EC03	0.96	-0.33
EC04	0.88	-1.14
EC07	1.00	0.00
FC02	1.02	0.40
FC08	0.92	-1.54
FC10	0.98	-0.18
FC12	0.97	-0.18
KC01i	1.00	0.09
KC01o	1.13	0.91
EP02	1.23	2.18
EP03	0.87	-1.51
EP09	1.09	0.83
EP11	1.12	0.80
FP02	1.01	0.14
FP05	0.99	-0.10
FP06	0.96	-0.20
FP07	0.95	-0.52
FP12	1.04	0.45
KP01	1.00	0.00

Tabelle F.3: Itemstatistiken

<i>Item</i>	<i>N</i>	<i>xi.item</i>	<i>Kategorie</i>	<i>rpb.WLE</i>	<i>M.WLE</i>	<i>SD.WLE</i>
EB01	167	0.978	0	-0.38	-0.58	0.83
			1	0.26	0.18	0.86
			2	0.20	0.86	0.83
EB02	151	0.235	0	-0.35	-0.57	0.85
			1	-0.07	-0.06	0.79
			2	0.14	0.20	0.87
			3	0.32	0.82	0.76
EB05	147	0.512	0	-0.47	-0.61	0.86
			1	-0.04	-0.12	0.71
			2	0.33	0.50	0.73
			3	0.32	0.72	0.70
EB10	186	-1.331	0	-0.25	-0.80	0.79
			1	-0.22	-0.28	1.00
			2	0.35	0.28	0.80
FB02	177	0.375	0	-0.44	-0.38	0.83
			1	0.44	0.43	0.82
FB04	141	0.723	0	-0.54	-0.64	0.77
			1	0.23	0.23	0.73
			2	0.45	1.16	0.60
FB07	182	-0.023	0	-0.48	-0.45	0.85
			1	0.48	0.44	0.81
FB11	162	0.952	0	-0.52	-0.63	0.86
			1	0.05	0.12	0.76
			2	0.35	0.65	0.71
			3	0.34	1.42	0.51
FB14	144	0.269	0	-0.64	-0.70	0.72
			1	0.11	0.21	0.68
			2	0.57	0.92	0.64
KB01	92	0.495	0	-0.31	-0.36	0.62
			1	-0.21	-0.22	0.64
			2	0.19	0.28	0.61
			3	0.53	1.17	0.69
EC02	158	0.291	0	-0.36	-0.87	1.17
			1	0.26	0.22	0.86
			2	0.04	0.24	0.86
EC03	141	0.183	0	-0.51	-0.69	0.79
			1	-0.18	-0.27	0.73
			2	0.17	0.36	0.68
			3	0.48	1.11	0.69
EC04	155	0.327	0	-0.51	-0.69	0.77
			1	0.15	0.24	0.71
			2	0.39	0.91	0.63
EC07	109	0.512	0	-0.43	-0.56	0.82
			1	-0.17	-0.19	0.87
			2	0.33	0.56	0.55
			3	0.41	1.11	0.69

<i>Item</i>	<i>N</i>	<i>xi.item</i>	<i>Kategorie</i>	<i>rbp.WLE</i>	<i>M.WLE</i>	<i>SD.WLE</i>
FC02	175	0.167	0	-0.38	-0.36	0.85
			1	0.38	0.33	0.86
FC08	184	0.034	0	-0.48	-0.48	0.88
			1	0.48	0.44	0.79
FC10	127	1.387	0	-0.39	-0.17	0.92
			1	0.39	0.69	0.63
FC12	97	0.430	0	-0.48	-0.63	0.87
			1	0.12	0.08	0.76
			2	0.44	0.80	0.82
KC01i	95	1.884	0	-0.30	-0.10	0.78
			1	0.16	0.33	0.77
			2	0.30	1.19	0.28
KC01o	95	1.239	0	-0.32	-0.17	0.84
			1	0.18	0.26	0.62
			2	0.26	0.67	0.70
EP02	126	0.286	0	-0.25	-0.21	0.83
			1	0.02	0.12	0.81
			2	0.25	0.47	0.72
EP03	145	-0.080	0	-0.63	-0.79	0.63
			1	0.14	0.17	0.62
			2	0.48	0.55	0.71
EP09	121	-0.181	0	-0.40	-0.71	0.93
			1	-0.19	-0.21	0.55
			2	-0.01	0.03	0.80
			3	0.53	0.78	0.69
EP11	128	0.372	0	-0.08	-0.03	0.73
			1	-0.10	0.09	0.80
			2	0.24	0.73	0.72
FP02	176	1.590	0	-0.29	-0.17	0.91
			1	0.29	0.54	0.95
FP05	165	0.115	0	-0.52	-0.71	0.81
			1	-0.15	-0.25	0.75
			2	0.17	0.32	0.54
			3	0.52	0.85	0.67
FP06	183	-2.112	0	-0.36	-0.91	0.79
			1	0.36	0.11	0.91
FP07	162	0.253	0	-0.41	-0.44	0.72
			1	-0.08	-0.05	0.79
			2	0.53	0.84	0.67
FP12	183	0.619	0	-0.51	-0.41	0.81
			1	0.10	0.20	0.89
			2	0.49	0.76	0.74
KP01	172	0.434	0	-0.39	-0.65	0.96
			1	-0.28	-0.41	0.64
			2	0.40	0.40	0.66
			3	0.30	0.99	0.72

*Anmerkungen.* *xi.item* = mittlere Itemschwierigkeit, *rbp.WLE* = punktbiseriale Korrelation der Kategorie zur WLE, *M.WLE* = Mittelwert der WLE aller Probandinnen und Probanden in dieser Kategorie, *SD.WLE* = Standardabweichungen der WLE aller Probandinnen und Probanden in dieser Kategorie.

# Anhang G

## Fragebogen aus Studie 2

### Fragebogen zum individuellen Konzept über die Naturwissenschaften



Liebe Teilnehmerin, lieber Teilnehmer,

ich komme vom IPN in Kiel und führe mit euch Erhebungen im Rahmen des Projektes „Individuelles Konzept über die Naturwissenschaften“ (IKoN) durch. In diesem Projekt möchte ich gerne mehr über Schülerinnen und Schüler erfahren, die sich für Naturwissenschaften interessieren. Ich finde besonders spannend, was dich genau an den Naturwissenschaften interessiert und wie du dir Naturwissenschaftlerinnen und Naturwissenschaftler vorstellst. Auch möchte ich gerne wissen, was du schon alles im Bereich der Naturwissenschaften weißt und was du für Ideen im Zusammenhang mit Naturwissenschaften hast.

Lies dir alle Fragen bitte ganz genau durch. Einige Fragen scheinen sich zu ähneln, aber gerade die kleinen Unterschiede sind für mich sehr wichtig! Es kann auch vorkommen, dass du mehrere Antworten ankreuzen kannst.

Die Teilnahme ist für dich freiwillig. Du hast keine Nachteile, wenn du nicht mitmachen möchtest, und kannst die Teilnahme jederzeit ohne Angabe von Gründen abbrechen. Deine Mitarbeit würde mir jedoch sehr dabei helfen, den Unterricht in den naturwissenschaftlichen Fächern und außerschulische Angebote (wie zum Beispiel naturwissenschaftliche Schülerwettbewerbe) weiter zu verbessern.

Vielen Dank für deine Hilfe!

*Christiane*

Deine Antworten bleiben anonym. Schreibe deshalb bitte *nicht* deinen Namen auf den Fragebogen, sondern trage hier deinen Code ein.

Kreuz bitte beim <b>ersten</b> Buchstaben des Vornamens deiner Mutter.																														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte beim <b>zweiten</b> Buchstaben des Vornamens deiner Mutter.																														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte beim <b>ersten</b> Buchstaben deines Vornamens.																														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte bei deinem Geburtstag.																														
01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31

Beispiel: Vorname der Mutter CHRISTIANE, dein Vorname MIRIAM, dein Geburtstag 02. Mai.

Kreuz bitte beim <b>ersten</b> Buchstaben des Vornamens deiner Mutter.																														
A	B	<input checked="" type="checkbox"/>	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte beim <b>zweiten</b> Buchstaben des Vornamens deiner Mutter.																														
A	B	C	D	E	F	G	<input checked="" type="checkbox"/>	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte beim <b>ersten</b> Buchstaben deines Vornamens.																														
A	B	C	D	E	F	G	H	I	J	K	L	<input checked="" type="checkbox"/>	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
Kreuz bitte bei deinem Geburtstag.																														
01	<input checked="" type="checkbox"/>	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Allgemeine Fragen**



1. Du bist ein  Mädchen und  12  13  14  15  16  17  18 Jahre alt.  
 Junge

2. Du besuchst zurzeit ein / eine  Gymnasium  Gesamtschule  Privatschule  sonstige Schule.

3. In welchem Bundesland gehst du zur Schule?

- |  |   |  |   |
|--|---|--|---|
| <input type="checkbox"/> Baden-Württemberg | <input type="checkbox"/> Bremen                 | <input type="checkbox"/> Niedersachsen       | <input type="checkbox"/> Sachsen            |
| <input type="checkbox"/> Bayern            | <input type="checkbox"/> Hamburg                | <input type="checkbox"/> Nordrhein-Westfalen | <input type="checkbox"/> Sachsen-Anhalt     |
| <input type="checkbox"/> Berlin            | <input type="checkbox"/> Hessen                 | <input type="checkbox"/> Rheinland-Pfalz     | <input type="checkbox"/> Schleswig-Holstein |
| <input type="checkbox"/> Brandenburg       | <input type="checkbox"/> Mecklenburg-Vorpommern | <input type="checkbox"/> Saarland            | <input type="checkbox"/> Thüringen          |

4. Trage deine letzten Zeugnisnoten in die Tabelle ein und kreuze an, wie sehr dich die jeweiligen Schulfächer interessieren.

Schulfach	Note	Das interessiert mich gar nicht, das tue ich nicht gerne	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr, das tue ich sehr gerne
Deutsch	1 2 3 4 5 6				
Englisch	1 2 3 4 5 6				
Mathematik	1 2 3 4 5 6				
Biologie	1 2 3 4 5 6				
Chemie	1 2 3 4 5 6				
Physik	1 2 3 4 5 6				
Musik	1 2 3 4 5 6				
Kunst	1 2 3 4 5 6				
Wirtschaft / Politik o. Recht	1 2 3 4 5 6				
Geschichte	1 2 3 4 5 6				
Erdkunde	1 2 3 4 5 6				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Wettbewerben**



5. Im Folgenden findest du einige Aussagen zu naturwissenschaftlichen Wettbewerben aufgeführt. Kreuze an, wie sehr du den Aussagen für jeden einzelnen Wettbewerb zustimmst.

	Ich stimme überhaupt nicht zu	Ich stimme eher nicht zu	Ich stimme eher zu	Ich stimme voll zu
<b>1. Wenn ich an dem Wettbewerb teilnehme, steigt mein Ansehen in der Klasse.</b>				
Jugend forscht .....				
Internationale JuniorScienceOlympiade (IJSO) .....				
<b>2. Bei dem Wettbewerb kann ich meine Fähigkeiten und Talente beweisen.</b>				
Jugend forscht .....				
Internationale JuniorScienceOlympiade (IJSO) .....				
<b>3. Bei dem Wettbewerb finde ich heraus, ob Naturwissenschaften und Technik auch im Beruf und Studium Spaß bringen können.</b>				
Jugend forscht .....				
Internationale JuniorScienceOlympiade (IJSO) .....				
<b>4. Der Wettbewerb ist nur etwas für Hochbegabte.</b>				
Jugend forscht .....				
Internationale JuniorScienceOlympiade (IJSO) .....				
<b>5. Wenn ich an dem Wettbewerb teilnehme, werde ich in den naturwissenschaftlichen Fächern besonders gefördert.</b>				
Jugend forscht .....				
Internationale JuniorScienceOlympiade (IJSO) .....				

6. Hast du schon an einem naturwissenschaftlichen Wettbewerb teilgenommen?

- Ja  
 Nein

Wenn du schon an einem naturwissenschaftlichen Wettbewerb teilgenommen hast, beantworte auch die folgenden Fragen. Wenn nicht, kannst du die Fragen 7 bis 10 überspringen und auf Seite 5 weitermachen.

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Wettbewerben**



- 7. An welchen der folgenden naturwissenschaftlichen Wettbewerbe hast du schon teilgenommen?**
- BundesUmweltWettbewerb
  - Internationale JuniorScienceOlympiade
  - Jugend forscht/ Schüler experimentieren
  - Bundeswettbewerb Mathematik
  - Mathematik-Olympiade
  - Internationale Physik-Olympiade
  - Bundesweiter Wettbewerb Physik für die Sekundarstufe 1
  - Internationale Chemie-Olympiade
  - Internationale Biologie-Olympiade
  - Bundeswettbewerb Informatik
  - Andere
- 8. Welche war die höchste Stufe, die du jemals bei der Teilnahme an einem naturwissenschaftlichen Wettbewerb erreicht hast?**
- Erste Runde
  - Zweite Runde
  - Dritte Runde
  - Vierte Runde
  - Nationales Finale
  - Internationales Finale
- 9. Wie oft hast du schon an einem naturwissenschaftlichen Wettbewerb teilgenommen?**
- 1 Mal
  - 2 Mal
  - 3 Mal
  - 4 Mal
  - 5 Mal
  - mindestens 6 Mal
- 10. Wer oder was hat dich dazu bewegt, an einem naturwissenschaftlichen Wettbewerb teilzunehmen (mehrere Nennungen sind möglich)?**
- Arbeitsgemeinschaft (AG) an der Schule
  - Hochbegabtenförderprogramm
  - Mein Lehrerin / mein Lehrer
  - Meine Eltern oder Großeltern
  - Meine Mitschülerinnen / Mitschüler oder Freunde
  - Die zu gewinnenden Preise
  - Die Aussicht auf Erfolg
  - Ein Bericht in der Zeitung
  - Die Möglichkeit meine Ideen zu verwirklichen

4

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zum naturwissenschaftlichen Unterricht**



Auf den Seiten 5 und 6 geht es um eine Reihe unterschiedlicher Aussagen. Kreuze an, inwieweit du den Aussagen zustimmen kannst.

	Ich stimme überhaupt nicht zu	Ich stimme eher nicht zu	Ich stimme eher zu	Ich stimme voll zu
1. Ich bin für die Schule sehr begabt. ....				
2. Neues zu lernen fällt mir schwer. ....				
3. Ich bin sehr intelligent. ....				
4. Ich kann in der Schule viel. ....				
5. In der Schule fallen mir viele Aufgaben schwer. ....				

	Ich stimme überhaupt nicht zu	Ich stimme eher nicht zu	Ich stimme eher zu	Ich stimme voll zu
1. Naturwissenschaften würde ich viel lieber machen, wenn das Fach/ die Fächer nicht so schwer wäre/n. ....				
2. Naturwissenschaften liegen mir nicht besonders. ....				
3. Naturwissenschaften fallen mir schwerer als vielen meiner Mitschüler/innen. ....				
4. Bei manchen Themen im Naturwissenschaftsunterricht weiß ich gleich: „Das verstehe ich nie.“ ....				
5. Für Naturwissenschaften habe ich einfach keine Begabung. ....				
6. Ich denke, ich bin für den Naturwissenschaftsunterricht begabter als viele meiner Mitschüler/innen. ....				
7. Mit den Aufgaben im Naturwissenschaftsunterricht komme ich besser zurecht als viele meiner Mitschüler/innen. ....				

5

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zum naturwissenschaftlichen Unterricht**



Im Naturwissenschaftsunterricht geht es mir darum, ...

		Ich stimme überhaupt nicht zu	Ich stimme eher nicht zu	Ich stimme eher zu	Ich stimme voll zu
1. etwas Interessantes zu lernen. ....					
2. komplizierte Inhalte zu verstehen. ....					
3. die Dinge, die wir im Unterricht besprechen, wirklich zu verstehen. ....					
4. so viel wie möglich zu lernen. ....					
5. zu zeigen, dass ich bei einer Sache gut bin. ....					
6. bessere Noten zu bekommen als andere. ....					
7. Klassenarbeiten besser zu schaffen als andere. ....					
8. dass andere denken, dass ich klug bin. ....					
9. zu verheimlichen, wenn ich weniger weiß als andere. ....					
10. nicht zu zeigen, falls ich weniger schlau bin als andere. ....					
11. mich nicht zu blamieren (z.B. durch falsche Ergebnisse oder dumme Fragen). ....					
12. dass niemand merkt, wenn ich etwas nicht verstehe. ....					
13. mich nicht zu sehr anstrengen zu müssen. ....					
14. mit wenig Arbeit durch die Schule zu kommen. ....					
15. keine schwierigen Fragen oder Aufgaben lösen zu müssen. ....					
16. zu Hause keine Arbeiten erledigen zu müssen. ....					

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Interessen**



Die Seiten 7 bis 12 bestehen aus einer Liste verschiedener **Tätigkeiten**, die im **naturwissenschaftlichen Schulunterricht**, in **naturwissenschaftlichen Fördermaßnahmen** oder auch in einem **naturwissenschaftlichen Beruf** ausgeübt werden können. Gib für jede einzelne davon an, wie sehr dich deren Ausübung interessiert bzw. interessieren würde. **Interessieren bedeutet dabei: etwas gerne tun, etwas wegen der Sache selbst tun.**

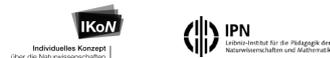
Erläuterung:

In der Schule	Wie sehr bist du an der Ausübung dieser Tätigkeit in der Schule interessiert?
In einer Fördermaßnahme	Wie sehr bist du an der Ausübung dieser Tätigkeit in einer naturwissenschaftlichen Fördermaßnahme interessiert? (Fördermaßnahmen sind z.B. AGs, Wettbewerbe, Enrichment-Kurse, etc.) Naturwissenschaftliche Fördermaßnahmen haben das Ziel, interessierte oder auch talentierte Jugendliche besonders in den Naturwissenschaften zu fördern und zu fordern.
Im Beruf	Wie sehr bist du an der Ausübung dieser Tätigkeit in deinem späteren Beruf interessiert?

Das interessiert mich gar nicht, das interessiert mich überhaupt nicht <b>gerne</b>	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr, das tue ich <b>sehr gerne</b>
---	-----------------------------	--------------------------------	---

<b>1. Mit Laborgeräten oder Experimentier-Sets arbeiten</b>				
In der Schule.....				
In einer Fördermaßnahme.....				
Im Beruf.....				
<b>2. Naturwissenschaftliche Fachliteratur lesen und interpretieren</b>				
In der Schule.....				
In einer Fördermaßnahme.....				
Im Beruf.....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Interessen**



Das interessiert mich <b>gar nicht</b> , überhaupt <b>nicht</b> gerne	Das interessiert mich <b>wenig</b>	Das interessiert mich <b>ziemlich</b>	Das interessiert mich <b>sehr</b> , das tue ich <b>sehr gerne</b>
---	------------------------------------	---------------------------------------	---

<b>3. Naturwissenschaftliche Inhalte nach künstlerischen Gesichtspunkten darstellen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>4. Andere Menschen in den Naturwissenschaften unterrichten</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>5. Eine naturwissenschaftliche Arbeitsgruppe leiten</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>6. Chemikalienbestände sortieren und verwalten</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>7. Sich mit anderen Menschen über naturwissenschaftliche Inhalte austauschen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Interessen**



Das interessiert mich <b>gar nicht</b> , überhaupt <b>nicht</b> gerne	Das interessiert mich <b>wenig</b>	Das interessiert mich <b>ziemlich</b>	Das interessiert mich <b>sehr</b> , das tue ich <b>sehr gerne</b>
---	------------------------------------	---------------------------------------	---

<b>8. Versuchsaufbauten zeichnen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>9. Das Verhalten / die Eigenschaften von Stoffen untersuchen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>10. Romane / Geschichten mit naturwissenschaftlichen Inhalten lesen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>11. Andere Menschen über naturwissenschaftliche Inhalte aufklären</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>12. Andere beim Experimentieren beaufsichtigen oder kontrollieren</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Interessen**



Das interessiert mich <b>gar nicht</b> ; das <b>will ich nicht</b> gerne	Das interessiert mich <b>wenig</b>	Das interessiert mich <b>ziemlich</b>	Das interessiert mich <b>sehr</b> ; das tue ich <b>sehr gerne</b>
--	------------------------------------	---------------------------------------	---

<b>13. Messwerte in ein Computerprogramm eingeben</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>14. Einen naturwissenschaftlichen Vortrag halten</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>15. Geräte für naturwissenschaftliche Messungen reinigen / reparieren</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>16. Ein Modell entwickeln</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>17. Eine naturwissenschaftliche Fachsprache lernen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Interessen**



Das interessiert mich <b>gar nicht</b> ; das <b>will ich nicht</b> gerne	Das interessiert mich <b>wenig</b>	Das interessiert mich <b>ziemlich</b>	Das interessiert mich <b>sehr</b> ; das tue ich <b>sehr gerne</b>
--	------------------------------------	---------------------------------------	---

<b>18. Mithilfe der Naturwissenschaften etwas „Gutes“ für Menschen tun</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>19. Für das Ansehen der Naturwissenschaften in der Öffentlichkeit auftreten</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>20. Insekten, Kristalle o.ä. sammeln und ordnen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>21. Mit anderen eine naturwissenschaftliche Fragestellung diskutieren</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>22. Experimente nach einer Anleitung durchführen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu naturwissenschaftlichen Interessen**

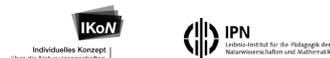


Das interessiert mich gar nicht über das Maß hinaus	Das interessiert mich wenig	Das interessiert mich ziemlich	Das interessiert mich sehr, das tue ich sehr gerne
---	-----------------------------	--------------------------------	--

<b>23. Die Ursachen eines Phänomens erforschen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>24. Naturwissenschaftliche Präparate zeichnen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>25. Die Versorgung mit Medikamenten durch die Naturwissenschaften verbessern</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>27. Die Einhaltung von Sicherheitsvorschriften überwachen</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				
<b>28. Einen naturwissenschaftlichen Vortrag hören</b>				
<i>In der Schule</i> .....				
<i>In einer Fördermaßnahme</i> .....				
<i>Im Beruf</i> .....				

12

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu Eigenschaften und Tätigkeiten eines Naturwissenschaftlers**



Nun geht es um Eigenschaften und Tätigkeiten von Naturwissenschaftlerinnen und Naturwissenschaftlern. Kreuze an, wie wichtig du die folgenden Eigenschaften für Naturwissenschaftlerinnen / Naturwissenschaftler hältst.

	überhaupt nicht wichtig	nicht wichtig	weder noch	wichtig	sehr wichtig
1. handwerklich geschickt.....					
2. intellektuell, analytisch.....					
3. künstlerisch, ästhetisch.....					
4. kreativ.....					
5. sozial, hilfsbereit.....					
6. zielstrebig, unternehmerisch.....					
7. kooperativ.....					
8. sorgfältig, genau.....					

Auf den letzten beiden Seiten geht es um eine Reihe unterschiedlicher Aussagen. Kreuze an, inwieweit du den Aussagen zustimmen kannst.

Ein Wissenschaftler übt folgende Tätigkeiten regelmäßig aus...

	Stimmt gar nicht	Stimmt eher nicht	Stimmt eher	Stimmt genau
1. Beobachtungen in der Natur durchführen.....				
2. Neue Messgeräte entwickeln.....				
3. Studenten betreuen.....				
4. Lehrveranstaltungen vorbereiten und durchführen.....				
5. Zeitschriftenartikel verfassen.....				
6. Abrechnungen für verwendete Forschungsgelder fertigstellen.....				
7. Austausch mit Wissenschaftlern anderer Universitäten.....				
8. Stoffe (zum Beispiel Medikamente) im Labor herstellen.....				
9. Eigene Theorien entwickeln.....				
10. Ergebnisse aus Experimenten auswerten und interpretieren.....				

13

**Fragebogen zum individuellen Konzept über die Naturwissenschaften  
Fragen zu Eigenschaften und Tätigkeiten eines Naturwissenschaftlers**



	Stimmt gar nicht	Stimmt eher nicht	Stimmt eher	Stimmt genau
11. Interviews geben.....				
12. An eigenen Erfindungen arbeiten.....				
14. Untersuchungen durchführen (zum Beispiel Lebensmittel oder Gewässer) .....				
15. Finanzen planen und verwalten.....				
16. Eigene Messverfahren entwickeln.....				
17. An Büchern schreiben.....				
18. Modelle entwickeln.....				
19. Ein Team aufbauen und managen.....				
20. Verwaltungsaufgaben erledigen.....				
21. Fächerübergreifende Projekte durchführen.....				
22. Doktoranden betreuen.....				
23. Zeichnungen, Fotografien oder Protokolle anfertigen.....				
24. Eigene Versuchsaapparaturen konstruieren.....				
25. Ein Projekt organisieren und leiten.....				
26. Treffen mit Kollegen anderer Fachbereiche durchführen.....				
27. Abschlussarbeiten von Studenten begleiten.....				
28. Messungen durchführen.....				
29. Eine Arbeitsgruppe leiten.....				
30. Ideen für neue Forschungsansätze entwickeln.....				

*Geschafft!  
Vielen Dank.*

## Anhang H

# Itemparameter, Itemfit und Itemstatistiken (Studie 2)

Tabelle H.1: Schwellenparameter in Logits

<i>Item</i>	<i>Kategorie 1</i>	<i>Kategorie 2</i>	<i>Kategorie 3</i>
EB10	-1.97	-0.10	1.75
FB04	-0.20	2.26	
FB14	-0.05	0.90	2.34
EC02	-0.94	0.12	1.86
FC10	0.43	0.87	
FC12	0.17	1.36	2.80
EP11	-0.94	0.51	1.91
FP05	-0.50	0.40	0.51
FP07	-0.38	1.441	
KP01	-0.477	-0.119	2.293

Tabelle H.2: Mean-Squared-Infit und Infit-*t*

<i>Item</i>	<i>Mean-Squared-Infit</i>	<i>Infit-t</i>
EB10	1.04	0.84
FB04	0.96	-0.94
FB14	0.96	-0.75
EC02	1.06	1.43
FC10	1.02	0.41
FC12	0.95	-0.95
EP11	1.10	1.62
FP05	0.97	-0.78
FP07	0.98	-0.62
KP01	0.99	-0.27

Tabelle H.3: Itemstatistiken

<i>Item</i>	<i>N</i>	<i>xi.item</i>	<i>Kategorie</i>	<i>rpb.WLE</i>	<i>M.WLE</i>	<i>SD.WLE</i>
EB10	932	-0.107	0	-0.32	-0.82	0.79
			1	-0.19	-0.16	0.63
			2	0.22	0.18	0.63
			3	0.23	0.51	0.67
FB04	830	1.029	0	-0.42	-0.34	0.64
			1	0.26	0.19	0.61
			2	0.30	0.82	0.73
			3	0.15	0.64	0.73
FB14	726	1.066	0	-0.47	-0.37	0.59
			1	0.10	0.12	0.57
			2	0.41	0.61	0.54
			3	0.15	0.64	0.73
EC02	845	0.347	0	-0.38	-0.56	0.68
			1	-0.02	0.00	0.64
			2	0.22	0.22	0.61
			3	0.20	0.51	0.61
FC10	680	0.648	0	-0.32	-0.23	0.65
			1	0.03	0.00	0.58
			2	0.40	0.59	0.57
			3	0.18	1.19	0.70
FC12	637	1.446	0	-0.45	-0.35	0.62
			1	0.19	0.16	0.58
			2	0.36	0.67	0.62
			3	0.18	1.19	0.70
EP11	480	0.490	0	-0.20	-0.24	0.71
			1	-0.13	-0.07	0.60
			2	0.20	0.23	0.57
			3	0.24	0.66	0.65
FP05	907	0.136	0	-0.42	-0.52	0.63
			1	-0.21	-0.22	0.52
			2	0.24	0.30	0.49
			3	0.47	0.71	0.58
FP07	823	0.530	0	-0.38	-0.36	0.63
			1	0.11	0.06	0.61
			2	0.37	0.63	0.58
			3	0.27	0.79	0.62
KP01	792	0.566	0	-0.39	-0.49	0.61
			1	-0.14	-0.15	0.56
			2	0.36	0.29	0.56
			3	0.27	0.79	0.62

*Anmerkungen.* *xi.item* = mittlere Itemschwierigkeit, *rpb.WLE* = punktbiseriale Korrelation der Kategorie zur WLE, *M.WLE* = Mittelwert der WLE aller Probandinnen und Probanden in dieser Kategorie, *SD.WLE* = Standardabweichungen der WLE aller Probandinnen und Probanden in dieser Kategorie.

# Anhang I

## Itemparameter der DIF-Analysen (Studie 2)

Tabelle I.1: Itemparameter der DIF-Analysen für das Merkmal Geschlecht

<i>Item</i>	<i>Gruppe</i>	<i>N</i>	<i>M</i>	<i>Itemschwierigkeit in Logits</i>
EB10	männlich	531	1.48	0.032
EB10	weiblich	398	1.65	-0.307
KP01	männlich	452	1.38	0.368
KP01	weiblich	337	1.06	0.876
EC02	männlich	475	1.34	0.337
EC02	weiblich	367	1.31	0.373
EP11	männlich	285	1.18	0.525
EP11	weiblich	192	1.24	0.422
FB04	männlich	469	0.67	0.944
FB04	weiblich	360	0.60	1.129
FB14	männlich	412	0.90	0.955
FB14	weiblich	312	0.77	1.199
FC10	männlich	388	0.72	0.496
FC10	weiblich	290	0.55	0.861
FC12	männlich	360	0.72	1.332
FC12	weiblich	276	0.59	1.592
FP05	männlich	523	1.53	-0.080
FP05	weiblich	381	1.09	0.437
FP07	männlich	474	0.88	0.289
FP07	weiblich	347	0.63	0.894

Tabelle I.2: Itemparameter der DIF-Analysen für das Merkmal Teilnahme am IJSO-Bundesfinale (ja/ nein)

<i>Item</i>	<i>Gruppe</i>	<i>N</i>	<i>M</i>	<i>Itemschwierigkeit in Logits</i>
EB10	nein	871	1.53	-0.062
EB10	ja	61	1.87	-0.721
KP01	nein	748	1.21	0.622
KP01	ja	44	1.80	-0.306
EC02	nein	788	1.30	0.373
EC02	ja	57	1.61	-0.104
EP11	nein	444	1.19	0.505
EP11	ja	36	1.39	0.106
FB04	nein	780	0.61	1.118
FB04	ja	50	1.04	-0.107
FB14	nein	681	0.78	1.219
FB14	ja	45	1.71	-0.288
FC10	nein	644	0.63	0.686
FC10	ja	36	1.00	-0.021
FC12	nein	607	0.63	1.560
FC12	ja	30	1.33	0.261
FP05	nein	851	1.29	0.203
FP05	ja	56	2.25	-0.925
FP07	nein	774	0.74	0.633
FP07	ja	49	1.37	-0.905

Tabelle I.3: **Itemparameter der DIF-Analysen für das Merkmal Wettbewerbserfahrung** Bufi = Teilnehmende des IJSO-Bundesfinales 2013 oder 2014, Wetn = Wettbewerbsteilnehmende, Notn = Nicht-Teilnehmende.

<i>Item</i>	<i>Gruppe</i>	<i>N</i>	<i>M</i>	<i>Itemschwierigkeit in Logits</i>
EB10	Bufi	61	1.87	-0.687
EB10	Wetn	383	1.55	-0.106
EB10	Notn	488	1.52	-0.031
KP01	Bufi	44	1.80	-0.312
KP01	Wetn	344	1.30	0.490
KP01	Notn	404	1.14	0.727
EC02	Bufi	57	1.61	-0.079
EC02	Wetn	357	1.29	0.388
EC02	Notn	431	1.32	0.352
EP11	Bufi	36	1.39	0.039
EP11	Wetn	191	1.35	0.292
EP11	Notn	253	1.08	0.688
FB04	Bufi	50	1.04	-0.211
FB04	Wetn	350	0.66	0.990
FB04	Notn	430	0.57	1.244
FB14	Bufi	45	1.71	-0.265
FB14	Wetn	321	0.94	0.950
FB14	Notn	360	0.65	1.497
FC10	Bufi	36	1.00	-0.058
FC10	Wetn	289	0.71	0.523
FC10	Notn	355	0.56	0.827
FC12	Bufi	30	1.33	0.283
FC12	Wetn	283	0.73	1.355
FC12	Notn	324	0.54	1.766
FP05	Bufi	56	2.25	-0.955
FP05	Wetn	378	1.42	0.056
FP05	Notn	473	1.18	0.324
FP07	Bufi	49	1.37	-0.840
FP07	Wetn	338	0.74	0.612
FP07	Notn	436	0.73	0.638



# Literaturverzeichnis

- Abernathy, T. V. & Vineyard, R. N. (2001). Academic competitions in science: What are the rewards for students? *The Clearing House*, 74 (5), 269–276. doi: 10.1080/00098650109599206
- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, 102 (1), 3–27. doi: 10.1037/0033-2909.102.1.3
- Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, 77 (5), 598–614. doi: 10.1037/0021-9010.77.5.598
- Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, 22 (2), 227–257. doi: 10.1016/S0160-2896(96)90016-1
- Ackerman, P. L. & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, 121 (2), 219–245. doi: 10.1037/0033-2909.121.2.219
- Adams, R. & Wu, M. (2002). *PISA 2000 technical report*. Paris: OECD.
- Adams, R. J., Wu, M. L. & Wilson, M. (2012). The Rasch Rating Model and the disordered threshold controversy. *Educational and Psychological Measurement*, 72 (4), 547–573. doi: 10.1177/0013164411432166
- Adamson, L. B., Foster, M. A., Roark, M. L. & Reed, D. B. (1998). Doing a science project: Gender differences during childhood. *Journal of Research in Science Teaching*, 35 (8), 845–857.
- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Arbeitsgemeinschaft bundesweiter Schülerwettbewerbe (Hrsg.). (2016). *Wettbewerbe*. Zugriff am 19.08.2016 auf <http://www.bundeswettbewerb.de/wettbewerbe.html>
- Arnold, K., Dietrich, V., Eberle, A., Grimmer, A., Labahn, B., Malz, R. & Schäfer, S. (2010). *Chemie Oberstufe: Gesamtband : Allgemeine Chemie, Physikalische Chemie, Organische Chemie*. Berlin: Cornelsen.
- Asseburg, R. (2011). *Motivation zur Testbearbeitung in adaptiven und nicht-adaptiven Leistungstests* (Dissertation). Christian-Albrechts-Universität Kiel, Kiel.
- Asselborn, W., Jäckel, M. & Risch, K. T. (Hrsg.). (2011). *Chemie heute: Kontextorientierter Ansatz: Sekundarstufe I*. Braunschweig: Schroedel.
- Atkins, P. W. & de Paula, J. (2013). *Physikalische Chemie* (5. Aufl.). Weinheim: VCH.
- Ayala, R. J., Plake, B. S. & Impara, J. C. (2001). The impact of omitted responses on the accuracy of ability estimation in Item Response Theory. *Journal of Educational Measurement*, 38 (3), 213–234. doi: 10.1111/j.1745-3984.2001.tb01124.x

- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.). (2000a). *Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (Bd. 2). Opladen: Leske & Budrich.
- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.). (2000b). *TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Opladen: Leske & Budrich.
- Baumert, J. & Köller, O. (1998). Interest research in secondary Level I: An overview. In L. Hoffmann, A. Krapp, K. A. Renninger & J. Baumert (Hrsg.), *Interest and learning* (Bd. 164). Kiel: IPN.
- Becker, J. (2004). *Computergestütztes Adaptives Testen (CAT) von Angst entwickelt auf der Grundlage der Item Response Theorie (IRT)* (Dissertation). Freie Universität Berlin, Berlin.
- Bell, B. S. & Kozlowski, S. W. J. (2002). Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *Journal of Applied Psychology*, 87 (3), 497–505. doi: 10.1037/0021-9010.87.3.497
- Berger, J.-L. & Karabenick, S. A. (2011). Motivation and students' use of learning strategies: Evidence of unidirectional effects in mathematics classrooms. *Learning and Instruction*, 21 (3), 416–428. doi: 10.1016/j.learninstruc.2010.06.002
- Bernholt, S., Parchmann, I. & Commons, M. L. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 219–245.
- Beutel, W. & Tetzlaff, S. (2007). Schülerwettbewerbe und Schulentwicklung: Zur pädagogischen Wirksamkeit eines Förderinstruments. *Die Deutsche Schule*, 99 (H. 2), 242–244.
- Biggs, J. B. & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. Academic Press.
- Bitter, T., Sommer, B. & Tompert, A. (2010). *Elemente Chemie 2* (1. Aufl.). Stuttgart: Klett.
- Blankenburg, J. S., Höffler, T. N. & Parchmann, I. (2015). Naturwissenschaftliche Wettbewerbe – Was kann junge Schülerinnen und Schüler zur Teilnahme motivieren? *Zeitschrift für Didaktik der Naturwissenschaften*, 1–13. Zugriff auf <http://dx.doi.org/10.1007/s40573-015-0031-y> doi: 10.1007/s40573-015-0031-y
- Blankenburg, J. S., Höffler, T. N. & Parchmann, I. (2016). Fostering today what is needed tomorrow: Investigating students' interest in science. *Science Education*, 100 (2), 364–391. doi: 10.1002/sce.21204
- Blankenburg, J. S., Höffler, T. N., Peters, H. & Parchmann, I. (2016). The effectiveness of a project day to introduce sixth grade students to science competitions. *Research in Science & Technological Education*, 1–17. doi: 10.1080/02635143.2016.1222361
- Blömeke, S., Risse, J., Müller, C., Eichler, D. & Schulz, W. (2006). Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht: Ein allgemeines Modell und seine exemplarische Umsetzung im Unterrichtsfach Mathematik. *Unterrichtswissenschaft*, 34 (4), 330–357. Zugriff auf <http://www.pedocs.de/volltexte/2013/5522>
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives*. New York: McKay.
- Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed. Aufl.). Mahwah: Lawrence Erlbaum Associates Publishers.
- Bos, W., Wendt, H., Köller, O. & Selter, C. (Hrsg.). (2012). *TIMSS 2011: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.

- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. (1997). *Gutachten zur Vorbereitung des Programms "Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts"* (Nr. 60). Bonn.
- Bundesministerium für Bildung und Forschung (Hrsg.). (2012). *Perspektive MINT: Wegweiser für MINT-Förderung und Karrieren in Mathematik, Informatik, Naturwissenschaften und Technik*. Berlin.
- Bundesministerium für Bildung und Forschung (Hrsg.). (2015). *Entdecken. Mitmachen. Bewegen. Schüler- und Jugendwettbewerbe des Bundesministeriums für Bildung und Forschung*. Zugriff am 18.05.2015 auf [http://www.bmbf.de/pub/Entdecken\\_Mitmachen\\_Bewegen.pdf](http://www.bmbf.de/pub/Entdecken_Mitmachen_Bewegen.pdf)
- Campbell, J. R. (1996). Early identification of mathematics talent has long-term positive consequences for career contributions. *International Journal of Educational Research*, 25 (6), 497–522. doi: 10.1016/S0883-0355(97)86728-6
- Campbell, J. R., Tirri, K., Ruohotie, P. & Walberg, H. J. (Hrsg.). (2004). *Cross-cultural research: Basic issues, dilemmas, and strategies*. Hämeenlinna: Research Centre for Vocational Education, University of Tampere.
- Campbell, J. R., Wagner, H. & Walberg, H. J. (2000). Academic competitions and programs designed to challenge the exceptionally talented. In K. A. Heller, F. J. Mönks, R. F. Subotnik & R. J. Sternberg (Hrsg.), *The international handbook of giftedness and talent* (S. 523–535). Amsterdam: Elsevier.
- Campbell, J. R. & Walberg, H. J. (2010). Olympiad studies: Competitions provide alternatives to developing talents that serve national interests. *Roeper Review*, 33 (1), 8–17. doi: 10.1080/02783193.2011.530202
- Campbell, J. R. & Wu, W.-T. (1996). Development of exceptional academic talent: International research studies. *International Journal of Educational Research*, 25 (6), 479–484. Zugriff auf <http://www.sciencedirect.com/science/article/pii/S0883035597867262> doi: 10.1016/S0883-0355(97)86726-2
- Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action* (Bd. 35). Amsterdam: North Holland.
- Ceci, S. J. & Liker, J. K. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology: General*, 115 (3), 255–266. doi: 10.1037/0096-3445.115.3.255
- Chambers, D. W. (1983). Stereotypic images of the scientist: The draw-a-scientist test. *Science Education*, 67 (2), 255–265. doi: 10.1002/sce.3730670213
- Czerniak, C. M. (1996). Predictors of success in a district science fair competition: An exploratory study. *School Science and Mathematics*, 96 (1), 21–27. doi: 10.1111/j.1949-8594.1996.tb10208.x
- Czerniak, C. M. & Lumpe, A. T. (1996). Predictors of science fair participation using the theory of planned behavior. *School Science and Mathematics*, 96 (7), 355–361. doi: 10.1111/j.1949-8594.1996.tb15853.x
- Dai, D. Y., Moon, S. M. & Feldhusen, J. F. (1998). Achievement motivation and gifted students: A social cognitive perspective. *Educational Psychologist*, 33 (2-3), 45–63. doi: 10.1080/00461520.1998.9653290
- Daniels, Z. (2008). *Entwicklung schulischer Interessen im Jugendalter*. Münster: Waxmann.

- Deussing, C. (2014). *Inhalt und Anforderungen von Aufgaben der Chemie Olympiade* (unveröffentlichte Masterarbeit). Christian-Albrechts-Universität Kiel, Kiel.
- Deutscher Bildungsserver (Hrsg.). (2016). *Wettbewerbe*. Zugriff am 12.08.2016 auf <http://www.bildungsserver.de/wettbew.html>
- Dierks, P. O., Höffler, T. N., Blankenburg, J. S., Peters, H. & Parchmann, I. (2016). Interest in science: a RIASEC-based analysis of students' interests. *International Journal of Science Education*, *38* (2), 238–258. doi: 10.1080/09500693.2016.1138337
- Dierks, P. O., Höffler, T. N. & Parchmann, I. (2014). Profiling interest of students in science: Learning in school and beyond. *Research in Science & Technological Education*, *32* (2), 1–18. doi: 10.1080/02635143.2014.895712
- Dionne, L., Reis, G., Trudel, L., Guillet, G., Kleine, L. & Hancianu, C. (2012). Students' sources of motivation for participating in science fairs: An exploratory study within the Canada-wide science fair 2008. *International Journal of Science and Mathematics Education*, *10* (3), 669–693. doi: 10.1007/s10763-011-9318-8
- Dochy, F., de Rijdt, C. & Dyck, W. (2002). Cognitive prerequisites and learning: How far have we progressed since Bloom? Implications for educational practice and teaching. *Active Learning in Higher Education*, *3* (3), 265–284. doi: 10.1177/1469787402003003006
- Dörner, D. & Kreuzig, H. W. (1983). Problemlösefähigkeit und Intelligenz. *Psychologische Rundschau*, *34* (4), 185–192.
- Dunbar, K. N. & Klahr, D. (2012). Scientific thinking and reasoning. In K. J. Holyoak & R. G. Morrison (Hrsg.), *The Oxford Handbook of thinking and reasoning* (S. 701–718). Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199734689.013.0035
- Eccles, J. S. (1983). Expectancies, values, and academic choice: Origins and changes. In J. T. Spence (Hrsg.), *Achievement and achievement motivation* (S. 87–134). San Francisco: W.H. Freeman.
- Edelen, M. O. & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, *16*, 5–18. doi: 10.1007/s11136-007-9198-0
- Eid, M., Gollwitzer, M. & Schmitt, M. (2013). *Statistik und Forschungsmethoden*. Weinheim: Beltz.
- Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. J. Elliot & C. S. Dweck (Hrsg.), *Handbook of competence and motivation* (S. 52–72). New York: Guilford Press.
- Elliot, A. J. & Harackiewicz, J. M. (1994). Goal setting, achievement orientation, and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, *66* (5), 968–980. doi: 10.1037/0022-3514.66.5.968
- Elliott, E. S. & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54* (1), 5–12. doi: 10.1037/0022-3514.54.1.5
- Europäische Kommission. (2004). *Europe needs more scientists: Report by the high level group on increasing human resources for science and technology in Europe 2004*. Luxemburg.
- Europäische Kommission. (2007). *Naturwissenschaftliche Erziehung JETZT: Eine erneuerte Pädagogik für die Zukunft Europas*. Luxemburg: Amt für aml. Veröff. d. Europ. Gemeinschaften.
- Europäische Kommission. (2010). *Europa 2020: Eine Strategie für intelligentes, nachhaltiges und integratives Wachstum*. Brüssel. Zugriff am 15.01.2017 auf <http://ec.europa.eu/eu2020/pdf/COMPLET%20%20DE%20SG-2010-80021-06-00-DE-TRA-00.pdf>

- Evans, R., Gräber, W., Koballa, T. & Nentwig, P. (2002). *Scientific Literacy: Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung*. Opladen: Leske & Budrich.
- Fausser, P. & Messner, R. (Hrsg.). (2007). *Fordern und fördern: Was Schülerwettbewerbe leisten*. Hamburg: Edition Körber-Stiftung.
- Feng, A. X., Campbell, J. R. & Verna, M. A. (2001). The talent development of American Physics Olympians. *Gifted and Talented International*, 16 (2), 108–114. doi: 10.1080/15332276.2001.11672972
- Finch, H. (2008). Estimation of Item Response Theory parameters in the presence of missing data. *Journal of Educational Measurement*, 45 (3), 225–245. doi: 10.1111/j.1745-3984.2008.00062.x
- Finnerty, V. (2013). *Can participation in a school science fair improve middle school students' attitudes toward science and interest in science careers?* (Unveröffentlichte Dissertation). University of Massachusetts, Lowell, Massachusetts.
- Fischer, C. (2006). *Lernstrategien in der Begabtenförderung: Eine empirische Untersuchung zu Strategien Selbstgesteuerten Lernens in der individuellen Begabungsförderung* (Dissertation). Universität Münster, Münster.
- Fischer, C., Mönks, F. J. & Grindel, E. (Hrsg.). (2004). *Curriculum und Didaktik der Begabtenförderung: Begabungen fördern, Lernen individualisieren*. Münster: LIT Verlag.
- Fischer, H. E. (1994). *Physiklernen: Eine Herausforderung für Unterrichtsforschung: Arbeiten zur Lernprozessforschung im Physikunterricht*. Frankfurt a. M.: Peter Lang.
- Fischer, H. E. & Draxler, D. (2001). Aufgaben und naturwissenschaftlicher Unterricht. *Der mathematische und naturwissenschaftliche Unterricht*, 54 (7), 388–393.
- Fisher, S. L. & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology*, 51 (2), 397–420. doi: 10.1111/j.1744-6570.1998.tb00731.x
- Forrester, J. H. (2010). *Competitive science events: Gender, interest, science self-efficacy, and academic major choice* (Dissertation, North Carolina State University, Raleigh, North Carolina). Zugriff auf <http://repository.lib.ncsu.edu/ir/bitstream/1840.16/6073/1/etd.pdf>
- Frey, A. (2012). Adaptives Testen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 275–293). Springer. Zugriff auf [http://dx.doi.org/10.1007/978-3-642-20072-4\\_11](http://dx.doi.org/10.1007/978-3-642-20072-4_11) doi: 10.1007/978-3-642-20072-4\_11
- Frey, A. & Ehmke, T. (2007). Hypothetischer Einsatz adaptiven Testens bei der Messung von Bildungsstandards in Mathematik. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (Bd. 8, S. 169–184). Wiesbaden: Verlag für Sozialwissenschaften.
- Friege, G. (2002). *Wissen und Problemlösen: Eine empirische Untersuchung des wissenszentrierten Problemlösens im Gebiet der Elektrizitätslehre auf der Grundlage des Experten-Novizen-Vergleichs*. Berlin: Logos.
- Friege, G. (2003). Wissen und Problemlösen: Empirische Untersuchungen des wissenszentrierten Problemlösens auf der Grundlage Experten-Novizen-Vergleich. In A. Pitton (Hrsg.), *Außerschulisches Lernen in Physik und Chemie* (S. 50–59). Münster: LIT Verlag.
- Friege, G. & Lind, G. (2006). Types and qualities of knowledge and their relations to problem solving in physics. *International Journal of Science and Mathematics Education*, 4 (3), 437–465. Zugriff auf <http://dx.doi.org/10.1007/s10763-005-9013-8> doi: 10.1007/s10763-005-9013-8
- Gagné, F. (2005). From gifts to talents: The DMGT as a developmental model. In R. J. Sternberg & J. E. Davidson (Hrsg.), *Conceptions of giftedness* (S. 98–119). Cambridge: Cam-

- bridge University Press.
- Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart and Winston.
- Gagné, R. M. & Briggs, L. J. (Hrsg.). (1974). *Principles of instructional design*. New York: Holt, Rinehart and Winston. doi: 10.1002/pfi.4140440211
- Gau, B. & Gau, R. (1990). On alterations in the structure and requirement level of theoretical problems set in International Physics Olympiads. In W. Gorzkowski (Hrsg.), *International Physics Olympiads* (S. 53–71). Singapur: World Scientific. doi: 10.1142/9789814368285\_0004
- Goetz, T., Pekrun, R., Zirngibl, A., Jullien, S., Kleine, M., Vom Hofe, R. & Blum, W. (2004). Leistung und emotionales Erleben im Fach Mathematik. *Zeitschrift für pädagogische Psychologie*, 18 (3/4), 201–212. doi: 10.1024/1010-0652.18.34.201
- Goldstein, D. & Wagner, H. (1993). After school programs, competitions school olympics, and summer programs. In K. A. Heller, F. J. Mönks & A. H. Passow (Hrsg.), *International handbook of research and development of giftedness and talent* (S. 593–604). Oxford: Pergamon Press.
- Gräber, W. (1992). Untersuchungen zum Schülerinteresse an Chemie und Chemieunterricht. *Praxis der Naturwissenschaften - Chemie in der Schule* (39), 270–273.
- Greene, B. A. & Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology*, 21 (2), 181–192. doi: 10.1006/ceps.1996.0015
- Greenfield, T. A. (1995). An exploration of gender participation patterns in science competitions. *Journal of Research in Science Teaching*, 32 (7), 735–748. doi: 10.1002/tea.3660320707
- Gruber, H. & Stamouli, E. (2015). Intelligenz und Vorwissen. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 25–44). Berlin: Springer.
- Gustafsson, J.-E. & Undheim, J. O. (1996). Individual differences in cognitive functions.
- Hackett, G. & Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education*, 20 (3), 261. doi: 10.2307/749515
- Hagemeister, V. (1999). Was wurde bei TIMSS erhoben? Über die empirische Basis einer aufregenden Studie. *Die Deutsche Schule*, 91 (2), 160–177. Zugriff auf [http://www.digizeitschriften.de/dms/img/?PPN=PPN509092632\\_0091&DMDID=dmdlog46](http://www.digizeitschriften.de/dms/img/?PPN=PPN509092632_0091&DMDID=dmdlog46)
- Hammann, M. (2006). Kompetenzförderung und Aufgabenentwicklung. *Der mathematische und naturwissenschaftliche Unterricht*, 59 (2), 85–95.
- Hannover, B. & Kessels, U. (2002). Challenge the science-stereotype: Der Einfluss von Technik-Freizeitkursen auf das Naturwissenschaften-Stereotyp von Schülerinnen und Schülern. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (Bd. 45, S. 341–358). Weinheim: Beltz. Zugriff auf <http://www.pedocs.de/volltexte/2011/3955>; <http://nbn-resolving.de/urn:nbn:de:0111-opus-39556>
- Hansford, B. C. & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52 (1), 123–142. doi: 10.3102/00346543052001123
- Harder, B. (2012). *Modelle zur Erklärung von Leistungsexzellenz im theoretischen und empirischen Vergleich* (Bd. 13). Berlin: Lit.
- Härtig, H., Heitmann, P. & Retelsdorf, J. (2015). Analyse der Aufgaben zur Evaluation der Bildungsstandards in Physik - Differenzierung von schriftsprachlichen Fähigkeiten

- und Fachlichkeit. *Zeitschrift für Erziehungswissenschaft*, 18 (4), 763–779. doi: 10.1007/s11618-015-0646-2
- Häußler, P. & Lind, G. (1998). *Erläuterungen zu Modul 1 mit Beispielen für den Physikunterricht: Weiterentwicklung der Aufgabenkultur im mathematisch-naturwissenschaftlichen Unterricht*. Zugriff am 15.04.15 auf [http://sinus-transfer.uni-bayreuth.de/module/modul\\_1weiterentwicklung\\_der\\_aufgabenkultur.html](http://sinus-transfer.uni-bayreuth.de/module/modul_1weiterentwicklung_der_aufgabenkultur.html)
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Heilmann, K. (1999). *Begabung - Leistung - Karriere: Die Preisträger im Bundeswettbewerb Mathematik 1971-1995*. Göttingen: Hogrefe.
- Heller, K. A. (2008). *Von der Aktivierung der Begabungsreserven zur Hochbegabtenförderung: Forschungsergebnisse aus vier Dekaden* (Bd. 2). Berlin: LIT Verlag.
- Heller, K. A., Mönks, F. J., Subotnik, R. F. & Sternberg, R. J. (Hrsg.). (2000). *The international handbook of giftedness and talent* (2. Aufl.). Amsterdam: Elsevier.
- Heller, K. A., Perleth, C. & Tock, K. L. (2005). The Munich model of giftedness designed to identify and promote gifted students. In R. J. Sternberg & J. E. Davidson (Hrsg.), *Conceptions of giftedness* (S. 147–170). Cambridge: Cambridge University Press.
- Heller, K. A. & Rosemann, B. (1981). *Planung und Auswertung empirischer Untersuchungen: Eine Einführung in die Wissenschaftsmethodik und Forschungsstatistik für Pädagogen, Psychologen und Soziologen* (2. Aufl.). Stuttgart: Klett-Cotta.
- Helmke, A. & Schrader, F.-W. (2010). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (S. 90–102). Weinheim: Beltz.
- Hetze, P. (2011). *Nachhaltige Hochschulstrategien für mehr MINT-Absolventen* (2. Aufl.). Essen: Stifterverband für die Wissenschaft.
- Höffler, T. N., Bonin, V. & Parchmann, I. (2016). Science vs. sports: motivation and self-concepts of participants in different school competitions. *International Journal of Science and Mathematics Education*. doi: 10.1007/s10763-016-9717-y
- Hoffmann, L., Häußler, P. & Lehrke, M. (1998). *Die IPN-Interessenstudie Physik* (Bd. 158). Kiel: IPN.
- Hoffmann, L., Krapp, A., Renninger, K. A. & Baumert, J. (Hrsg.). (1998). *Interest and learning: Proceedings of the Secon Conference on interest and gender* (Bd. 164). Kiel: IPN.
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3. Aufl.). Odessa: Psychological Assessment Resources.
- Holland, J. L. & Gottfredson, G. D. (1992). Studies of the hexagonal model: An evaluation of the perils of stalking the perfect hexagon. *Journal of Vocational Behavior*, 40 (2), 158–170.
- Holland, P. W. & Wainer, H. (1993). *Differential item functioning*. Hillsdale: Lawrence Erlbaum Associates.
- Holleman, A. F., Wiberg, E. & Wiberg, N. (2007). *Lehrbuch der anorganischen Chemie* (102. Aufl.). Berlin: De Gruyter.
- Holstermann, N. & Bögeholz, S. (2007). Interesse von Jungen und Mädchen an naturwissenschaftlichen Themen am Ende der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 71–86. Zugriff auf [http://archiv.ipn.uni-kiel.de/zfdn/pdf/006\\_Holster\\_13.pdf](http://archiv.ipn.uni-kiel.de/zfdn/pdf/006_Holster_13.pdf)
- Höttecke, D. (2004). Schülervorstellungen über die "Natur der Naturwissenschaften". In D. Höttecke, C. Höhle & E. Kircher (Hrsg.), *Lehren und Lernen über die Natur der Naturwissenschaften* (S. 264–277). Baltmannsweiler: Schneider Verlag Hohengehren.

- Höttecke, D., Höfle, C. & Kircher, E. (Hrsg.). (2004). *Lehren und Lernen über die Natur der Naturwissenschaften*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Hu, W. & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education*, 24 (4), 389–403. doi: 10.1080/09500690110098912
- Internationale BiologieOlympiade (Hrsg.). (2016). *Berichte*. Zugriff am 23.11.2016 auf <http://wettbewerbe.ipn.uni-kiel.de/ibo/>
- Internationale ChemieOlympiade (Hrsg.). (2016). *Mannschaften*. Zugriff am 23.11.2016 auf <http://wettbewerbe.ipn.uni-kiel.de/icho/>
- Internationale JuniorScienceOlympiade (Hrsg.). (2016). *Der Wettbewerb*. Zugriff am 09.12.2016 auf <http://ijso.info/>
- Internationale PhysikOlympiade (Hrsg.). (2016). *Teilnehmer*. Zugriff am 23.11.2016 auf <http://wettbewerbe.ipn.uni-kiel.de/iphoto/teilnehmer.html>
- IPN (Hrsg.). (2017). *Wirkungen naturwissenschaftlicher Schülerwettbewerbe: WinnerS*. Zugriff am 13.02.2017 auf <http://www.ipn.uni-kiel.de/de/forschung/projekte/winners>
- Jatzwauk, P. (2007). *Aufgaben im Biologieunterricht: Eine Analyse der Merkmale und des didaktisch-methodischen Einsatzes von Aufgaben im Biologieunterricht*. Berlin: Logos.
- Jones, G. (1991). Gender differences in science competitions. *Science Education*, 75 (2), 159–167. doi: 10.1002/sce.3730750202
- Jurecka, A. (2010). *Zum Zusammenhang von Differentiellen Item Funktionen und Testkultur* (Unveröffentlichte Dissertation). Johann Wolfgang Goethe-Universität, Frankfurt a. M..
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*. Berlin: Logos.
- Kauertz, A., Fischer, H. E., Lau, A. & Neumann, K. (2008). Kompetenzmessung durch Leistungstests: Hilfe oder Druckmittel? *Der mathematische und naturwissenschaftliche Unterricht*, 61 (2), 75–79.
- Kessels, U. & Hannover, B. (2004). Empfundene “Selbstnähe“ als Mediator zwischen Fähigkeitsselbstkonzept und Leistungskurswahlintentionen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36 (3), 130–138. doi: 10.1026/0049-8637.36.3.130
- Kiefer, T., Robitzsch, A. & Wu, M. (2015). *Package ‘TAM’*. Zugriff am 11.09.2015 auf <http://cran.r-project.org/web/packages/TAM/TAM.pdf>
- Kleinknecht, M., Bohl, T., Maier, U. & Metz, K. (Hrsg.). (2013). *Lern- und Leistungsaufgaben im Unterricht: Fächerübergreifende Kriterien zur Auswahl und Analyse*. Bad Heilbrunn: Klinkhardt.
- Kleinknecht, M., Maier, U., Metz, K. & Bohl, T. (2011). Analyse des kognitiven Aufgabenpotenzials: Entwicklung und Erprobung eines allgemeindidaktischen Auswertungsmanuals. *Unterrichtswissenschaft*, 39 (4), 328–344.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., ... Vollmer, H. J. (2007). *Zur Entwicklung nationaler Bildungsstandards: Eine Expertise*. Bonn: BMBF.
- Klieme, E. & Baumert, J. (2001). Identifying national cultures of mathematics education: Analysis of cognitive demands and differential item functioning in TIMSS. *European Journal of Psychology of Education*, 16 (3), 385–402. doi: 10.1007/BF03173189
- Klieme, E., Lucius, E. R. & Zabel, E. (Hrsg.). (1999). *Report of the 9th International Biology Olympiad in Kiel, Germany* (Bd. 9). Kiel: IPN.
- Köller, O. (1998). *Zielorientierungen und schulisches Lernen*. Münster: Waxmann.
- Köller, O. (2004). *Konsequenzen von Leistungsgruppierungen* (Bd. 37). Münster: Waxmann.

- Köller, O., Baumert, J. & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, 32 (5), 448. doi: 10.2307/749801
- Köller, O., Daniels, Z., Schnabel, K. U. & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse. *Zeitschrift für pädagogische Psychologie*, 14 (1), 26–37. doi: 10.1024//1010-0652.14.1.26
- Köller, O., Trautwein, U., Lüdtke, O. & Baumert, J. (2006). Zum Zusammenspiel von schulischer Leistung, Selbstkonzept und Interesse in der gymnasialen Oberstufe. *Zeitschrift für pädagogische Psychologie* (20), 27–39.
- König, F. (1986). Kreativitätsdiagnostik als essentieller Bestandteil der Intelligenzdiagnostik. *Diagnostica*, 32 (4), 345–357.
- Krampen, G. (1993). Diagnostik der Kreativität. In G. Trost, K. Ingenkamp & R. S. Jäger (Hrsg.), *Tests und Trends* (S. 11–39). Weinheim: Beltz.
- Krapp, A. (2002). Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, 12 (4), 383–409. doi: 10.1016/S0959-4752(01)00011-1
- Krapp, A., Schiefele, U. & Schreyer, I. (1993). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25 (2), 120–148.
- Kröger, J. (2011). *Validierung eines Kategoriensystems für Aufgaben aus Physikolympiaden* (unveröffentlichte Staatsexamensarbeit). Christian-Albrechts-Universität Kiel, Kiel.
- Kuchling, H. (2011). *Taschenbuch der Physik* (20. Aufl.). München: Hanser.
- Kultusministerkonferenz. (2004a). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss: Beschluss vom 04.12.2003*. München: Luchterhand.
- Kultusministerkonferenz. (2004b). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss: Beschluss vom 04.12.2003*. München: Luchterhand.
- Kultusministerkonferenz. (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss: Beschluss vom 16.12.2004*. München: Luchterhand.
- Kultusministerkonferenz. (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss: Beschluss vom 16.12.2004*. München: Luchterhand.
- Kultusministerkonferenz. (2005c). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss: Beschluss vom 16.12.2004*. München: Luchterhand.
- Kultusministerkonferenz. (2015). *Förderstrategie für leistungsstarke Schülerinnen und Schüler: Beschluss der Kultusministerkonferenz vom 11.06.2015*.
- Kultusministerkonferenz. (2016). *Gemeinsame Initiative von Bund und Ländern zur Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler: Beschluss der Kultusministerkonferenz vom 10.11.2016*.
- Lachmayer, S., Nerdel, C. & Prechtel, H. (2007). Modellierung kognitiver Fähigkeiten beim Umgang mit Diagrammen im naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 145–160; 200 KB. Zugriff auf [http://www.ipn.uni-kiel.de/zfdn/pdf/008\\_Lachmayer\\_13.pdf](http://www.ipn.uni-kiel.de/zfdn/pdf/008_Lachmayer_13.pdf)
- Larochelle, M. & Désautels, J. (1991). ‘Of course, it’s just obvious’: adolescents’ ideas of scientific knowledge. *International Journal of Science Education*, 13 (4), 373–389. doi: 10.1080/0950069910130402
- Lau, S. & Roeser, R. W. (2002). Cognitive abilities and motivational processes in High School students’ situational engagement and achievement in Science. *Educational Assessment*,

- 8 (2), 139–162. doi: 10.1207/S15326977EA0802\_04
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29 (4), 331–359. doi: 10.1002/tea.3660290404
- Lederman, N. G. (2007). Nature of science: Past, present, and future. *Handbook of research on science education*, 2, 831–879.
- Lengfelder, A. & Heller, K. A. (2002). German Olympiad studies: Findings from a retrospective evaluation and from in-depth interviews: Where have all the gifted females gone. *Journal of Research in Education*, 12 (1), 86–92.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R* (2. Aufl.). Göttingen: Hogrefe.
- Lopez, F. G. & Lent, R. W. (1992). Sources of mathematics self-efficacy in High School students. *The Career Development Quarterly*, 41 (1), 3–12. doi: 10.1002/j.2161-0045.1992.tb00350.x
- Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39 (2), 247–264. doi: 10.1007/BF02291471
- Lück, G. (2004). Naturwissenschaften im frühen Kindesalter. In W. E. Fthenakis (Hrsg.), *Frühpädagogik international* (S. 331–343). Wiesbaden: Verlag für Sozialwissenschaften. doi: 10.1007/978-3-322-95041-3\_22
- Ludlow, L. H. & O'Leary, M. O. (1999). Scoring omitted and not-reached items: Practical data analysis implications. *Educational and Psychological Measurement*, 59 (4), 615–630. doi: 10.1177/0013164499594004
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58 (2), 103–117. doi: 10.1026/0033-3042.58.2.103
- Maier, U., Bohl, T., Kleinknecht, M. & Metz, K. (2013). Allgemeindidaktische Kategorien für die Analyse von Aufgaben. In M. Kleinknecht, T. Bohl, U. Maier & K. Metz (Hrsg.), *Lern- und Leistungsaufgaben im Unterricht* (S. 9–45). Bad Heilbrunn: Klinkhardt.
- Markl, J. & Sadava, D. (2011). *Purves Biologie* (9. Aufl.). Heidelberg: Spektrum der Wissenschaft.
- Marsh, H. W. & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British journal of educational psychology*, 81 (Pt 1), 59–77. doi: 10.1348/000709910X503501
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child development*, 76 (2), 397–416. doi: 10.1111/j.1467-8624.2005.00853.x
- Mason, C. L., Kahle, J. B. & Gardner, A. L. (1991). Draw-a-Scientist test: Future implications. *School Science and Mathematics*, 91 (5), 193–198. doi: 10.1111/j.1949-8594.1991.tb12078.x
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47 (2), 149–174. doi: 10.1007/BF02296272
- McComas, W. F. (Hrsg.). (1998). *The nature of science in science education: Rationales and strategies*. Dordrecht: Kluwer Academic Publishers.
- Meißner, G. (2011). Aspekte der Bewertung von Jugend-forscht-Arbeiten in einem Landeswettbewerb. In M. Weigand, M. Thomas & F. Otto (Hrsg.), *Informatik mit Kopf, Herz und Hand* (S. 9–18). ZfL-Verlag.

- Meißner, G. (2013). *Leistungsanforderungen an Teilnehmerinnen und Teilnehmer von Informatikwettbewerben: Eine Untersuchung auf Grundlage des Kompetenz-Paradigmas* (Dissertation). Friedrich-Schiller-Universität Jena, Jena.
- Messner, R. (2007). Schülerwettbewerbe leisten methodische Pionierarbeit. In P. Fauser & R. Messner (Hrsg.), *Fordern und fördern* (S. 15–30). Hamburg: Edition Körber-Stiftung.
- Meyling, H. (1997). How to change students' conceptions of the epistemology of science. *Science & Education*, 6 (4), 397–416. doi: 10.1023/A:1017908916810
- Mislevy, R. J. & Wu, P.-K. (1988). Inferring examinee ability when some item responses are missing. *ETS Research Report Series, 1988* (2), i-75. doi: 10.1002/j.2330-8516.1988.tb00304.x
- Möller, J. & Köller, O. (2004). Die Genese akademischer Selbstkonzepte. *Psychologische Rundschau*, 55 (1), 19–27. doi: 10.1026/0033-3042.55.1.19
- Möller, K. (2014). Vom naturwissenschaftlichen Sachunterricht zum Fachunterricht – Der Übergang von der Grundschule in die weiterführende Schule. *Zeitschrift für Didaktik der Naturwissenschaften*, 20 (1), 33–43. doi: 10.1007/s40573-014-0010-8
- Moseley, D. et al. (Hrsg.). (2005). *Frameworks for thinking: A handbook for teaching and learning*. Cambridge: Cambridge University Press.
- Mund, W. (2007). *Jugend forscht und Jugend musiziert: Kognitive Fähigkeiten und Persönlichkeitsmerkmale erfolgreicher Teilnehmer* (Dissertation, Philipps-Universität Marburg, Marburg). Zugriff auf <http://archiv.ub.uni-marburg.de/diss/z2008/0750/pdf/dwm.pdf>
- Nagel, R. (1999). Die Schülerwettbewerbe "Jugend forscht" und "Schüler experimentieren". *Biologie in der Schule*, 48 (2), 110–114.
- Nerdel, C., Neumann, K., Stäudel, L. & Rehm, M. (2013). Fachdidaktische Analyse von Aufgaben in den Naturwissenschaften. In M. Kleinknecht, T. Bohl, U. Maier & K. Metz (Hrsg.), *Lern- und Leistungsaufgaben im Unterricht* (S. 95–100). Bad Heilbrunn: Klinkhardt.
- Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., ... Neubrand, J. (2004). Der Prozess der Itementwicklung bei der nationalen Ergänzungsuntersuchung von PISA 2000: Vom theoretischen Rahmen zu den konkreten Aufgaben. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland* (S. 31–49). Wiesbaden: Verlag für Sozialwissenschaften.
- Neumann, K., Kauertz, A., Lau, A., Notarp, H. & Fischer, H. E. (2007). Die Modellierung physikalischer Kompetenz und ihrer Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 103–123.
- Nokelainen, P., Tirri, K. & Campbell, J. R. (2004). Cross-cultural predictors of mathematical talent and academic productivity. *High Ability Studies*, 15 (2), 229–242. doi: 10.1080/1359813042000314790
- Nokelainen, P., Tirri, K., Campbell, J. R. & Walberg, H. (2007). Factors that contribute to or hinder academic productivity: Comparing two groups of most and least successful Olympians. *Educational Research and Evaluation*, 13 (6), 483–500. doi: 10.1080/13803610701785931
- Nokelainen, P., Tirri, K. & Merenti-Valimäki, H.-L. (2007). Investigating the influence of attribution styles on the development of mathematical talent. *Gifted Child Quarterly*, 51 (1), 64–81. doi: 10.1177/0016986206296659
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., ... Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex

- differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (26), 10593–10597. doi: 10.1073/pnas.0809921106
- Olbertz, J.-H. (2007). Hochbegabtenförderung zwischen Wissenschaft und Bildungspolitik: Ansätze in den neuen Bundesländern. In K. A. Heller & A. Ziegler (Hrsg.), *Begabt sein in Deutschland* (Bd. 1, S. 376–394). Berlin: LIT Verlag.
- Olszewski, P., Kulieke, M. & Buescher, T. (1987). The influence of the family environment on the development of talent: A literature review. *Journal for the Education of the Gifted*, 11 (1), 6–28. doi: 10.1177/016235328701100102
- Osborne, J., Collins, S., Ratcliffe, M., Millar, R. & Duschl, R. (2003). What "ideas-about-science" should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, 40 (7), 692–720. doi: 10.1002/tea.10105
- Osborne, J. & Dillon, J. (2008). *Science education in Europe: Critical reflections*. London: King's College London.
- Pajares, F. (1996). Self-efficacy beliefs and mathematical problem-solving of gifted students. *Contemporary Educational Psychology*, 21 (4), 325–344. doi: 10.1006/ceps.1996.0025
- Pajares, F. (1997). Current directions in self-efficacy research. In M. L. Maehr & P. R. Pintrich (Hrsg.), *Advances in motivation and achievement* (Bd. 10, S. 1–49). Greenwich: Jai Press.
- Pant, H. A., Böhme, K. & Köller, O. (2013). Das Kompetenzkonzept der Bildungsstandards und die Entwicklung von Kompetenzstufenmodellen. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012* (S. 53–61). Münster: Waxmann.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Paul, J. & Groß, J. (2016). Lehr- und Lernforschung in der Biologiedidaktik: Nature of Science (NoS) und Forschendes Lernen im Wettbewerb Jugend forscht. In U. Gebhard & M. Hamann (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (Bd. 7, S. 57–73).
- Paulick, I., Watermann, R. & Nückles, M. (2013). Achievement goals and school achievement: The transition to different school tracks in secondary school. *Contemporary Educational Psychology*, 38 (1), 75–86.
- Peters, H. (2014). Holz gefällt?! Wettbewerbsaufgabe zur IJSO 2012. *Unterricht Chemie*, 25 (144), 44–46.
- Peters, H., Giese, D. & Müller-Balhorn, S. (2013). Ein Bild von Wissenschaft und Forschung vermitteln: Jugend forscht und die Internationale JuniorScienceOlympiade. *Unterricht Chemie*, 24 (136), 18–25.
- Petersen, S. (2010). Oberflächen- und Tiefenmerkmale von Aufgaben in der Internationalen PhysikOlympiade: Projekt zur systematischen Analyse und Entwicklung von Aufgaben in dem Wettbewerb. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung*. Zugriff am 13.04.2015 auf <http://www.phydid.de/index.php/phydid-b/article/view/195/168>
- Pickard, M. J. (2007). The new Bloom's taxonomy: An overview for family and consumer sciences. *Journal of Family and Consumer Sciences Education*, 25 (1), 45–55.
- PISA-Konsortium Deutschland (Hrsg.). (2007). *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.

- Pohl, S. & Carstensen, C. H. (2012). *NEPS technical report - Scaling the data of the competence tests (NEPS working paper no. 14)*. Bamberg: Otto-Friedrich-Universität Bamberg, Nationales Bildungspanel.
- Pohl, S., Gräfe, L. & Rose, N. (2014). Dealing with omitted and not-reached items in competence tests: Evaluating approaches accounting for missing responses in Item Response Theory models. *Educational and Psychological Measurement*, *74* (3), 423–452. doi: 10.1177/0013164413504926
- Prenzel, M., Häußler, P., Rost, J. & Senkbeil, M. (2002). Der PISA-Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten vorhersagen? *Unterrichtswissenschaft*, *30* (1), 120–135.
- Prenzel, M., Heidemeier, H., Ramm, G., Hohensee, F. & Ehmke, T. (2004). Soziale Herkunft und mathematische Kompetenz. In M. Prenzel et al. (Hrsg.), *PISA 2003* (S. 273–282). Münster: Waxmann.
- Racherbäumer, K. & Boltz, S. (2012). Wettbewerbe, Preise, Qualitätssiegel und Co. Ein Überblick über Typen, Varianten, Funktionen und Wirkungen der Auszeichnung von Schülern und Schülerinnen und von Schulen. *Die Deutsche Schule*, *104* (2), 119–137.
- Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E. & Köller, O. (Hrsg.). (2016). *PISA 2015: Eine Studie zwischen Kontinuität und Innovation*. Münster: Waxmann.
- Renzulli, J. S. (2005). The three ring conception of giftedness: A developmental model for creative productivity. In R. J. Sternberg & J. E. Davidson (Hrsg.), *Conceptions of giftedness* (S. 246–279). Cambridge: Cambridge University Press.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? *Psychologische Rundschau*, *57* (2), 69–86.
- Rist, R. S. (1989). Schema creation in programming. *Cognitive Science*, *13* (3), 389–414.
- Ropohl, M., Walpuski, M. & Sumfleth, E. (2015). Welches Aufgabenformat ist das richtige? – Empirischer Vergleich zweier Aufgabenformate zur standardbasierten Kompetenzmessung. *Zeitschrift für Didaktik der Naturwissenschaften*, *21* (1), 1–15. doi: 10.1007/s40573-014-0020-6
- Rost, D. H. (Hrsg.). (2009). *Hochbegabte und hochleistende Jugendliche: Befunde aus dem Marburger Hochbegabtenprojekt* (2. Aufl.). Münster: Waxmann.
- Rost, D. H., Dickhäuser, O., Sparfeldt, J. R. & Schilling, S. R. (2004). Fachspezifische Selbstkonzepte und Schulleistungen im dimensional Vergleich. *Zeitschrift für pädagogische Psychologie*, *18* (1), 43–52. doi: 10.1024/1010-0652.18.1.43
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion* (2. Aufl.). Bern: Huber.
- Sahin, A., Gulacar, O. & Stuessy, C. (2015). High school students' perceptions of the effects of International Science Olympiad on their STEM career aspirations and twenty-first century skill development. *Research in Science Education*, *45* (6), 785–805. doi: 10.1007/s11165-014-9439-5
- Sandmann, A. (2007). Theorien und Methoden der Expertiseforschung in biologiedidaktischen Studien. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung* (S. 231–242). Berlin and New York: Springer. doi: 10.1007/978-3-540-68166-3\_21
- Schabram, K. (2007). *Lernaufgaben im Unterricht. Instruktionspsychologische Analysen am Beispiel der Physik* (Dissertation). Universität Duisburg-Essen, Duisburg-Essen.
- Scheuneman, J. D. & Gerritz, K. (1990). Using Differential Item Functioning procedures to explore sources of item difficulty and group performance characteristics. *Journal of Educational Measurement*, *27* (2), 109–131. doi: 10.1111/j.1745-3984.1990.tb00737.x

- Schiefele, U. (1996). *Motivation und Lernen mit Texten*. Göttingen: Hogrefe.
- Schiefele, U., Krapp, A., Wild, K.-P. & Winteler, A. (1993). Der "Fragebogen zum Studieninteresse" (FSI). *Diagnostica*, 39 (4), 335–351.
- Schiepe-Tiska, A., Rönnebeck, S., Schöps, K., Neumann, K., Schmidtner, S., Parchmann, I. & Prenzel, M. (2016). Naturwissenschaftliche Kompetenz in PISA 2015: Ergebnisse des internationalen Vergleichs mit einem modifizierten Testansatz. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015*. Münster: Waxmann.
- Schiepe-Tiska, A., Simm, I. & Schmidtner, S. (2016). Motivationale Orientierungen, Selbstbilder und Berufserwartungen in den Naturwissenschaften in PISA 2015. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015* (S. 99–132). Münster: Waxmann.
- Schlüter, K. (2009). Eine Studie zu den Merkmalen der Aufgabenschwierigkeit am Beispiel eines Informatik-Schülerwettbewerbs. Erster Teil: Aufgabenklassifizierung. In B. Koerber (Hrsg.), *Zukunft braucht Herkunft - 25 Jahre »INFOS – Informatik und Schule«* (S. 181–192). Bonn: Köllen.
- Schlüter, K. (2010). Eine Studie zu den Merkmalen der Aufgabenschwierigkeit am Beispiel eines Informatik-Schülerwettbewerbs. Zweiter Teil: Empirische Aufgabenanalyse. In I. Diethelm, C. Dörge, C. Hildebrandt & C. Schulte (Hrsg.), *Didaktik der Informatik - Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik* (S. 69–80). Bonn: Köllen.
- Schmechtig, N., Adolph, A. & Melzer, W. (2008). *Wissenschaftliche Begleitung der Schulversuche "Schulen mit besonderem pädagogische Profil/ Gemeinschaftsschulen". Untersuchung im Auftrag des Sächsischen Staatsministeriums für Kultus und Sport. Fortführung des Zwischenarbeitsstandes - Ergebnisse der Untersuchungen in Kohorte I und IV sowie Kohorte II (Längsschnitt) Untersuchungswelle 2010/2011*. Zugriff am 16.06.2015 auf [http://www.schule.sachsen.de/download/download\\_smk/gs\\_zwischenstand4.pdf](http://www.schule.sachsen.de/download/download_smk/gs_zwischenstand4.pdf)
- Schmitt, M. & Altstötter-Gleich, C. (2010). *Differentielle Psychologie und Persönlichkeitspsychologie kompakt* (1. Aufl.). Weinheim: Beltz.
- Schneider, W. & Bjorklund, D. F. (1992). Expertise, aptitude, and strategic remembering. *Child development*, 63 (2), 461–473.
- Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (Hrsg.). (2002). *SESSKO: Skalen zur Erfassung des schulischen Selbstkonzepts*. Göttingen: Hogrefe.
- Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2004). Zielorientierung und Bezugsnormorientierung: Zum Zusammenhang zweier Konzepte. *Zeitschrift für pädagogische Psychologie*, 18 (2), 93–99.
- Schöps, K. & Saß, S. (2012). *NEPS technical report for science - Scaling results of starting cohort 4 in ninth grade (NEPS working paper no. 23)*. Bamberg: Otto-Friedrich-Universität Bamberg, Nationales Bildungspanel.
- Schrader, F.-W. (2006). Kognitive Voraussetzungen. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 569–574). Bad Heilbrunn: Klinkhardt.
- Schroeders, U., Penk, C., Jansen, M. & Pant, H. A. (2013). Geschlechtsbezogene Disparitäten. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012* (S. 249–274). Münster: Waxmann.
- Schwarzer, R. & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. *Zeitschrift für Pädagogik Beiheft*, 44, 28–53.
- Schwarzlow, K. (2014). *Vergleichende Analyse von Aufgabenmerkmalen der ScienceOlympiaden im Fach Chemie* (unveröffentlichte Masterarbeit). Christian-Albrechts-Universität

- Kiel, Kiel.
- Segall, D. O. (2005). Computerized adaptive testing. In K. Kempf-Leonard (Hrsg.), *Encyclopedia of social measurement* (S. 429–438). New York: Academic Press.
- Senkbeil, M., Rost, J., Carstensen, C. H. & Walter, O. (2005). Der nationale Naturwissenschaftstest PISA 2003: Entwicklung und empirische Überprüfung eines zweidimensionalen Facetten-Designs. *Empirische Pädagogik*, 19 (2), 166–189.
- Solomon, J., Scott, L. & Duveen, J. (1996). Large-scale exploration of pupils' understanding of the nature of science. *Science Education*, 80 (5), 493–508. doi: 10.1002/(SICI)1098-237X(199609)80:5<493::AID-SCE1>3.0.CO;2-6
- Sonnert, G., Sadler, P. & Michaels, M. (2013). Gender aspects of participation, support, and success in a state science fair. *School Science and Mathematics*, 113 (3), 135–143. doi: 10.1111/ssm.12007
- Sparfeldt, J. R., Buch, S. R., Wirthwein, L. & Rost, D. H. (2007). Zielorientierungen: Zur Relevanz der Schulfächer. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39 (4), 165–176. doi: 10.1026/0049-8637.39.4.165
- Spinath, B., Stiensmeier-Pelster, J., Schöne, C. & Dickhäuser, O. (Hrsg.). (2002). *SELLMO: Skalen zur Erfassung der Lern- und Leistungsmotivation*. Göttingen: Hogrefe.
- Stang, J., Urhahne, D., Nick, S. & Parchmann, I. (2014). Wer kommt weiter? Vorhersage der Qualifikation zur Internationalen Biologie- und Chemie-Olympiade auf Grundlage des Leistungsmotivations-Modells von Eccles. *Zeitschrift für pädagogische Psychologie*, 28 (3), 105–114. doi: 10.1024/1010-0652/a000127
- Steinmayr, R., Bipp, T. & Spinath, B. (2011). Goal orientations predict academic performance beyond intelligence and personality. *Learning and Individual Differences*, 21 (2), 196–200. doi: 10.1016/j.lindif.2010.11.026
- Stern, E. (1997). Erwerb mathematischer Kompetenzen: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 157–170). Weinheim: Psychologie Verlags Union.
- Stiftung Jugend forscht e. V. (2015a). *Bewertungskriterien der Jury in Kurzform*. Zugriff am 18.05.2015 auf [http://www.jugend-forscht.de/fileadmin/user\\_upload/Downloadcenter/Teilnahme/Bewertungskriterien\\_Kurzform.pdf](http://www.jugend-forscht.de/fileadmin/user_upload/Downloadcenter/Teilnahme/Bewertungskriterien_Kurzform.pdf)
- Stiftung Jugend forscht e. V. (2015b). *Bewertungskriterien für Arbeiten im Wettbewerb Jugend forscht und Schüler experimentieren*. Zugriff am 18.05.2015 auf [http://www.jugend-forscht.de/fileadmin/user\\_upload/Downloadcenter/Teilnahme/Bewertungskriterien.pdf](http://www.jugend-forscht.de/fileadmin/user_upload/Downloadcenter/Teilnahme/Bewertungskriterien.pdf)
- Stiftung Jugend forscht e. V. (2015c). *Statistiken 2015*. Zugriff am 18.05.2015 auf [http://www.jugend-forscht.de/fileadmin/user\\_upload/Downloadcenter/Infomaterial/Statistik\\_Anmeldezahlen\\_2015\\_nach\\_Bundeslaendern.pdf](http://www.jugend-forscht.de/fileadmin/user_upload/Downloadcenter/Infomaterial/Statistik_Anmeldezahlen_2015_nach_Bundeslaendern.pdf)
- Stiftung Jugend forscht e. V. (27.06.2014). *Ehrung der Gewinner des Jugend forscht Schulpreises 2014: Bundesweit 81 Schulen erhalten für ihr besonderes Engagement die von CTS Gruppen- und Studienreisen GmbH geförderte Auszeichnung*. Hamburg. Zugriff am 03.08.2015 auf [http://www.jugend-forscht.de/uploads/media/Jugend\\_forscht\\_Schulpreis\\_2014.pdf](http://www.jugend-forscht.de/uploads/media/Jugend_forscht_Schulpreis_2014.pdf)
- Stocking, M. L., Eignor, D. R. & Cook, L. L. (1988). Factors affecting the sample invariant properties of linear and curvilinear observed- and true-score equating procedures. *ETS Research Report Series*, 1988 (2), i-71. doi: 10.1002/j.2330-8516.1988.tb00297.x
- Strunck, S. (2011). *Schulentwicklung durch Wettbewerbe: Prozesse und Wirkungen der Teilnahme an Schulwettbewerben*. Wiesbaden: Verlag für Sozialwissenschaften.

- Su, R., Rounds, J. & Armstrong, P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological Bulletin*, *135* (6), 859–884. doi: 10.1037/a0017364
- Subotnik, R. F., Duschl, R. A. & Selmon, E. H. (1993). Retention and attrition of science talent: a longitudinal study of Westinghouse Science Talent Search winners. *International Journal of Science Education*, *15* (1), 61–72. doi: 10.1080/0950069930150105
- Subotnik, R. F., Miserandino, A. D. & Olszewski-Kubilius, P. (1996). Implications of the olympiad studies for the development of mathematical talent in schools. *International Journal of Educational Research*, *25* (6), 563–573. doi: 10.1016/S0883-0355(97)86733-X
- Subotnik, R. F., Olszewski-Kubilius, P. & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12* (1), 3–54. doi: 10.1177/1529100611418056
- Subotnik, R. F., Orland, M., Rayhack, K., Schuck, J., Edmiston, A., Earle, J., ... Fuchs, B. (2009). Identifying and developing talent in science, technology, engineering, and mathematics (STEM): An agenda for research, policy, and practice. In L. V. Shavinina (Hrsg.), *International handbook on giftedness* (S. 1313–1326). Springer. doi: 10.1007/978-1-4020-6162-2\_69
- Sumfleth, E., Klebba, N., Kauertz, A., Mayer, J., Fischer, H. E. & Walpuski, M. (2013). Beschreibung der untersuchten naturwissenschaftlichen Kompetenzen. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012* (S. 38–52). Münster: Waxmann.
- Sumfleth, E., Klebba, N., Kauertz, A., Mayer, J., Fischer, H. E., Walpuski, M. & Wellnitz, N. (2013). Das Kompetenzstrukturmodell in den naturwissenschaftlichen Fächern. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012* (S. 38–42). Münster: Waxmann.
- Tacke, H. (2007). Gute Schülerwettbewerbe küren nicht nur Sieger. In P. Fauser & R. Messner (Hrsg.), *Fordern und fördern* (S. 111–122). Hamburg: Edition Körber-Stiftung.
- Tausch, M. & von Wachtendonk, M. (2010). *Chemie 2000+: Niedersachsen 1*. Bamberg: C. C. Buchner.
- Tirri, K. (2002). Developing females' talent: Case studies of Finnish Olympians. *Journal of Research in Education*, *12* (1), 80–85.
- Tirri, K. (2010). Motivation and giftedness. *High Ability Studies*, *21* (2), 77–80. doi: 10.1080/13598139.2010.528923
- Tirri, K. & Nokelainen, P. (2010). The influence of self-perception of abilities and attribution styles on academic choices: Implications for gifted education. *Roeper Review*, *33* (1), 26–32. doi: 10.1080/02783193.2011.530204
- Urhahne, D., Ho, L. H., Parchmann, I. & Nick, S. (2012). Attempting to predict success in the qualifying round of the International Chemistry Olympiad. *High Ability Studies*, *23* (2), 167–182. doi: 10.1080/13598139.2012.738324
- Urhahne, D. & Hopf, M. (2004). Epistemologische Überzeugungen in den Naturwissenschaften und ihre Zusammenhänge mit Motivation, Selbstkonzept und Lernstrategien. *Zeitschrift für Didaktik der Naturwissenschaften*, *10* (1), 71–87.
- Urhahne, D., Kremer, K. & Mayer, J. (2008). Welches Verstandnis haben Jugendliche von der Natur der Naturwissenschaften? Entwicklung und erste Schritte zur Validierung eines Fragebogens. *Unterrichtswissenschaft*, *36* (1), 71.
- van der Linden, W. J. (2005). A comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement*, *42* (3), 283–302. doi:

- 10.1111/j.1745-3984.2005.00015.x
- van Voorhis, C. R. W. & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, 3 (2), 43–50. Zugriff auf <http://www.tqmp.org/Content/vol103-2/p043/p043.pdf>
- Verna, M. A. & Feng, A. X. (2002). American Chemistry Olympians achieve the highest level of equity. *Journal of Research in Education*, 12 (1), 101–105.
- Vock, M., Köller, O. & Nagy, G. (2013). Vocational interests of intellectually gifted and highly achieving young adults. *The British journal of educational psychology*, 83 (Pt 2), 305–328. doi: 10.1111/j.2044-8279.2011.02063.x
- Vollhardt, K. P. C. & Schore, N. E. (2012). *Organische Chemie: Set aus Lehrbuch und Arbeitsbuch* (5. Aufl.). Weinheim: Wiley-VCH.
- Wagner, H. & Neber, H. (2007a). Nationale und internationale Leistungswettbewerbe im Kontext. In K. A. Heller & A. Ziegler (Hrsg.), *Begabt sein in Deutschland* (Bd. 1, S. 209–232). Berlin: LIT Verlag.
- Wagner, H. & Neber, H. (2007b). Schülerwettbewerbe fördern Begabungen. In P. Fauser & R. Messner (Hrsg.), *Fordern und fördern* (S. 73–84). Hamburg: Edition Körber-Stiftung.
- Walpuski, M., Kampa, N. N., Kauertz, A. & Wellnitz, N. (2008). Evaluation der Bildungsstandards in den Naturwissenschaften. *Der mathematische und naturwissenschaftliche Unterricht*, 61 (1), 323–326.
- Walpuski, M., Kauertz, A., Kampa, N. N., Fischer, H. E., Mayer, J., Sumfleth, E. & Wellnitz, N. (2010). ESNaS - Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In A. Gehrman, U. Hericks & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle* (S. 171–184). Bad Heilbrunn: Klinkhardt.
- Walpuski, M. & Ropohl, M. J. (2014). Statistische Verfahren für die Analyse des Einflusses von Aufgabenmerkmalen auf die Schwierigkeit. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 385–398). Berlin: Springer Spektrum.
- Walter, O. (2005). *Kompetenzmessung in den PISA-Studien: Simulationen zur Schätzung von Verteilungsparametern und Reliabilitäten*. Lengerich: Pabst.
- Weiber, R. & Mühlhaus, D. (2014a). Formative Messmodelle. In R. Weiber & D. Mühlhaus (Hrsg.), *Strukturgleichungsmodellierung* (S. 255–273). Berlin: Springer. doi: 10.1007/978-3-642-35012-2\_12
- Weiber, R. & Mühlhaus, D. (Hrsg.). (2014b). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl.). Berlin: Springer. doi: 10.1007/978-3-642-35012-2
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer.
- Weinert, F. E. (1996). Wissen und Denken. Über die unterschätzte Bedeutung des Gedächtnisses für das menschliche Denken. In Bayerische Akademie der Wissenschaften (Hrsg.), *Jahrbuch der Bayerischen Akademie der Wissenschaften 1996* (S. 85–101).
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim: Beltz.
- Weinert, F. E. & Helmke, A. (1995). Interclassroom differences in instructional quality and interindividual differences in cognitive development. *Educational Psychologist*, 30 (1), 15–20. doi: 10.1207/s15326985ep3001\_2
- Welch, A. & Huffman, D. (2011). The effect of robotics competitions on high school students' attitudes toward science. *School Science and Mathematics*, 111 (8), 416–424. doi: 10

- .1111/j.1949-8594.2011.00107.x
- Wellnitz, N., Fischer, H. E., Kauertz, A., Mayer, J., Neumann, I., Pant, H. A., ... Walpuski, M. (2012). Evaluation der Bildungsstandards – eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 261–291.
- Wentorf, W., Höffler, T. N. & Parchmann, I. (2015). Schülerkonzepte über das Tätigkeitsspektrum von Naturwissenschaftlerinnen und Naturwissenschaftlern: Vorstellungen, korrespondierende Interessen und Selbstwirksamkeitserwartungen. *Zeitschrift für Didaktik der Naturwissenschaften*, 21 (1), 207–222. doi: 10.1007/s40573-015-0035-7
- Wetzel, E. & Carstensen, C. H. (2014). Reversed thresholds in Partial Credit Models. *Assessment*, 21 (6), 765–774. doi: 10.1177/1073191114530775
- Wilbert, J. & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufsdagnostik. *Empirische Sonderpädagogik*, 3 (3), 225–242.
- Williams, G. & Clarke, D. M. (1997). Mathematical task complexity and task selection. In D. M. Clarke et al. (Hrsg.), *Mathematics. Imagine the possibilities* (S. 406–415). Brunswick, Victoria: Mathematical Association of Victoria.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah: Lawrence Erlbaum Associates.
- Wirthwein, L., Sparfeldt, J. R., Pinquart, M., Wegerer, J. & Steinmayr, R. (2013). Achievement goals and academic achievement: A closer look at moderating factors. *Educational Research Review*, 10, 66–89.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität : Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Wodzinski, R. (2007). Mädchen im Physikunterricht. In E. Kircher, R. Girwidz & P. Häußler (Hrsg.), *Physikdidaktik* (S. 559–580). Berlin: Springer.
- Wu, M. & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Melbourne: Educational Measurement Solutions.
- Wu, W.-T. & Chen, J.-D. (2001). A follow-up study of Taiwan Physics and Chemistry Olympians: The role of environmental influences in talent development. *Gifted and Talented International*, 16 (1), 16–26. doi: 10.1080/15332276.2001.11672949
- Zickar, M. J. & Broadfoot, A. A. (2009). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Hrsg.), *Statistical and methodological myths and urban legends* (S. 37–59). New York: Routledge.
- Zilker, I., Kauertz, A. & Fischer, H. E. (2009). Kompetenzdiagnose im Bereich Nature of Science und Scientific Inquiry. In D. Höttecke (Hrsg.), *Chemie- und Physikdidaktik für die Lehramtsausbildung* (S. 392–394). Münster: LIT Verlag.
- Zimmermann, P. & Spangler, G. (2001). Jenseits des Klassenzimmers: Der Einfluss der Familie auf Intelligenz, Motivation, Emotion und Leistung im Kontext der Schule. *Zeitschrift für Pädagogik*, 47 (4), 461–479.

## **Eidesstattliche Erklärung**

Hiermit erkläre ich, dass die vorliegende Arbeit - abgesehen von der Beratung durch meine Betreuerinnen und Betreuer - nach Inhalt und Form meine eigene ist. Es wurden keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Die Dissertation ist unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden. Die Arbeit hat weder im Ganzen noch in Teilen an anderer Stelle im Rahmen eines Prüfungsverfahrens vorgelegen und wurde weder im Ganzen noch in Teilen veröffentlicht oder zur Veröffentlichung eingereicht.

Kiel, den 28. Februar 2017

---

Christine Köhler