

**Sophie Bettina Kirschner**

**MODELLIERUNG UND ANALYSE  
DES PROFESSIONSWISSENS VON  
PHYSIKLEHRKRÄFTEN**



---

# MODELLIERUNG UND ANALYSE DES PROFESSIONSWISSENS VON PHYSIKLEHRKRÄFTEN

---

Dissertation

zur

Erlangung des Doktorgrades

Dr. phil. nat.

vorgelegt von

**Sophie Bettina Kirschner**

Geboren in Frankfurt am Main, wohnhaft in Karben

Fakultät für Physik

der Universität Duisburg-Essen

Juli 2013

Erstgutachter: Prof. Dr. Hans E. Fischer

Zweitgutachter: Prof. Dr. Andreas Borowski

Vorsitzende: Prof. Dr. Heike Theyßen

Prüferin: Prof. Dr. Claudia von Aufschnaiter

Tag der Disputation 25. November 2013



## ZUSAMMENFASSUNG

Das Professionswissen von Lehrkräften gilt als Grundlage für erfolgreichen Unterricht, obwohl weder Konsens über das Konstrukt besteht, noch der Zusammenhang zu Lehrerhandeln und Schülerleistung und –motivation geklärt ist.

Erstes Ziel dieser Arbeit ist die Modellierung des Professionswissens, die Umsetzung des Modells in ein Testinstrument und die empirische Absicherung des Modells. Zweites Ziel ist die Untersuchung des Zusammenhangs des Professionswissens von Physiklehrkräften und ihrem demographischen Hintergrund, sowie der Vergleich des Professionswissens von Lehramtsstudierenden der Physik, Physiklehrkräften im Vorbereitungsdienst und Physiklehrkräften.

Das Modell des Professionswissens beinhaltet folgende aus der Literatur abgeleitete Aspekte: (a) Die Dimensionen CK (Fachwissen), PCK (fachdidaktisches Wissen) und PK (pädagogisches Wissen), wobei PCK zwischen CK und PK steht; (b) Die PCK-Facetten Wissen über Schülervorstellungen, Konzepte und Experimente; (c) Die PK-Facetten Wissen über Klassenführung, Unterrichtsmethoden, Lernprozesse und Leistungsbeurteilung; (d) Die CK-Wissensstufen Schulwissen und vertieftes Schulwissen; (e) Den Schwerpunkt Mechanik, der die Tests fach- und themenabhängig macht; (f) Die Wissensbereiche deklaratives, prozedurales und konditionales Wissen in allen drei Dimensionen.

Die Inhaltsvalidität wurde durch den Abgleich mit Curricula und Fachliteratur, Expertenbefragungen und durch die Testentwicklung anhand des Modells gewährleistet. Bevor weitere Analysen durchgeführt wurden, wurde sichergestellt, dass sich die Items objektiv kodieren lassen ( $\text{Gamma} > .7$ ,  $\text{ICC}_{2,2}$  (unjustiert)  $> .7$ ). Zunächst wurde in einer mehrdimensionalen Rasch-Analyse untersucht, ob die angenommene Aufteilung in CK, PCK und PK empirisch nachgewiesen werden kann. Der PK-Test besteht aus einem deklarativen und einen konditional-prozeduralen Teil,  $\text{PK}_D$  und  $\text{PK}_{KP}$ , allerdings ist der  $\text{PK}_{KP}$ -Test nicht Rasch-konform und wird daher nicht weiter untersucht. Das Modell mit den drei Dimensionen CK, PCK und  $\text{PK}_D$  passt signifikant ( $p < .001$ ) besser zu den Daten, als die möglichen ein- oder zweidimensionalen Modelle. Daher können die drei Dimensionen für sich untersucht werden. Die Reliabilität ist zufriedenstellend, die EAP/PV-Reliabilität liegt in der Gruppe der Physiklehrkräfte der Hautstudie, die alle drei Tests ausgefüllt haben, bei .86 für den CK-Test, bei .77 für den PCK-Test und bei .71 für den  $\text{PK}_D$ -Test. Nach dem Vergleich des Wissens von Lehrkräften anderer Fächer ( $N=31$ ), von Physiklehrkräften verschiedener Schulformen (216 Gymnasiallehrkräfte vs. 62 Lehrkräfte anderer Schulformen) und von Diplom-Physikerinnen und -Physikern ( $N=22$ ) kann die Konstruktvalidität überwiegend bestätigt werden. Die Kriteriumsvalidität wird nach einem Vergleich der Korrelationen zwischen CK, PCK und PK als vorwiegend erfüllt angesehen. Da alle drei Dimensionen reliabel sind und als valide betrachtet werden können, wird mit den Personenfähigkeiten weiter gearbeitet. Allerdings muss bedacht werden, dass CK, PCK und PK das darstellen, was die modellbasierten Tests gemessen haben und insbesondere PCK nur auf die oben genannten drei Facetten fokussiert.

Um die fachspezifischen Bereiche des Professionswissens näher zu beleuchten, wird der Zusammenhang von CK und PCK zu den demographischen Variablen der Physiklehrkräfte (N=279) untersucht. Werden nur die drei Dimensionen in einer Regression betrachtet, klärt CK mehr als doppelt so viel PCK auf wie PK ( $R^2=.27$ ,  $F=34$ ,  $Beta_{CK}=0.46$ ,  $Beta_{PK}=0.19$ ). In Regressionen, deren Ergebnisse auf andere Populationen übertragbar sein sollten, haben die Variablen Geschlecht, fachfremder Unterricht und Unterricht am Gymnasium einen signifikanten Einfluss auf CK und die Variablen CK, Alter und Unterricht am Gymnasium einen signifikanten Einfluss auf PCK. Es ist zu bedenken, dass aufgrund der Stichprobengröße nur eine Auswahl der unabhängigen Variablen in die Regression einbezogen werden konnte und eine andere Auswahl vermutlich zu anderen Ergebnissen geführt hätte.

Des Weiteren wird analysiert, wie sich CK und PCK und ihr Zusammenhang zwischen Lehramtsstudierenden, Lehrkräften im Vorbereitungsdienst und Lehrkräften unterscheiden. Es ist festzustellen, dass nur in der Teilstichprobe, die ein gymnasiales Lehramt studiert oder studiert hat, signifikante Unterschiede zwischen den Gruppen auftreten: Studierende des Physik-Lehramts (N=43) schneiden in CK und PCK signifikant schlechter ab als Physiklehrkräfte im Vorbereitungsdienst (N=37) und Physiklehrkräfte (N=71). Zwischen den entsprechenden Gruppen des nicht-gymnasialen Lehramts (N=71) sind keine signifikanten Unterschiede zu finden. Die Korrelation zwischen CK und PCK unterscheidet sich nicht signifikant zwischen Studierenden, Lehrkräften im Vorbereitungsdienst und Lehrkräften.

Die Dimensionalität des Professionswissens wurde empirisch bestätigt, indem ein modellbasiertes Testinstrument entwickelt wurde, das für weitere Studien zur Verfügung steht. Mit Hilfe des Instruments konnte untersucht werden, welche Lehrervariablen besonders eng mit CK und PCK zusammenhängen und die Entwicklung von CK und PCK konnte quasi-längsschnittlich betrachtet werden.

## SUMMARY OF

### “MODELLING AND ANALYSING PHYSICS TEACHERS’ PROFESSIONAL KNOWLEDGE”

Professional knowledge is said to be a basis for effective teaching, however there is neither consensus about the construct in general nor scientific agreement on the relationship between teachers’ actions and students’ outcomes.

The first aim of this thesis is three-tiered: to model professional knowledge, to transform the model into a test and to support the model empirically. The second goal is to analyse the relationship between physics teachers’ professional knowledge and their demographic variables and to compare pre-service and in-service physics teachers’ professional knowledge.

The model of professional knowledge includes the following literature-based aspects: (a) the dimensions of content knowledge (CK), pedagogical content knowledge (PCK) and pedagogical knowledge (PK) with PCK connecting to CK and PK; (b) the PCK facets knowledge about students’ preconceptions, experiments and concepts; (c) the PK facets knowledge about classroom management, teaching methods, individual learning processes and assessment of performance; (d) the CK knowledge steps school subject knowledge and advanced school knowledge; (e) a focus on mechanics; (f) knowledge areas – declarative, procedural and conditional knowledge - in all three dimensions.

Content validity was ensured by aligning content with curricula and literature, by consulting with experts and developing a model-based test. Before further analysis was conducted, the objective coding of the items was investigated ( $\text{Gamma} > .7$ ,  $\text{ICC}_{2,2}$  (not adjusted)  $> .7$ ). The distinction in CK, PCK and PK was empirically analysed as a start. The PK test consists of a declarative and a procedural-conditional part,  $\text{PK}_D$  and  $\text{PK}_{KP}$ . Because the  $\text{PK}_{KP}$  test is not Rasch-conformed, it won’t be analysed in this thesis. The three dimensional model fits the data significantly better ( $p < .001$ ) in a multidimensional Rasch analysis than possible one or two dimensional models. Therefore the dimensions can be analysed separately. The reliability of all dimensions is satisfying. The EAP/PV reliability for physics teachers who finished all three tests in the main study is .86 for the CK test, .77 for the PCK test and .71 for the  $\text{PK}_D$  test. Teachers of other subjects than physics ( $N=31$ ), teachers of different school types ( $N=216$  at a gymnasium,  $N=62$  at other school types) and physicists ( $N=22$ ) were tested to confirm the construct validity. To check criteria validity, the correlations between CK, PCK and PK were compared. Since all three dimensions are reliable and valid, the calculated personal abilities for CK, PCK and PK can be dealt with. However, keep in mind that CK, PCK and PK only represent the model-based tests, not the whole construct.

To explore CK and PCK in more detail, the relationship between teachers’ demographic variables and their CK and PCK is examined ( $N=279$ ). If only CK, PCK and PK are considered in one regression, CK explains twice as much of PCK than PK does ( $R^2 = .27$ ,  $F=34$ ,  $\text{Beta}_{CK} = 0.46$ ,  $\text{Beta}_{PK} = 0.19$ ). In two regressions, which were considered to be generalizable to other populations, the variables *gender*, *out-of-field teaching* and *teaching at a gymnasium* have a significant impact on CK. For PCK the high-impact variables are *CK*, *age* and *teaching at a gymnasium*. Keep in mind that only a selection of independent

variables could be considered in the regression because of the sample size. A different selection of variables may have led to different results.

Furthermore, the knowledge of pre-service teachers at university (N=43) and at school (N=37) and of in-service teachers (N=71) is compared. This affects differences in CK and PCK and the connection between CK and PCK. Significant differences were only found in the group who went to university classes preparing for teaching at a gymnasium: Pre-service physics teachers at university (N=43) performed significantly worse than the more experienced pre-service physics teachers at school (N=37) and in-service physics teachers (N=71). No significant differences were found between the corresponding groups at other school types than the gymnasium.

The correlation between CK and PCK is not significantly different for pre-service teachers at university, at school and for in-service teachers.

The dimensions of professional knowledge were confirmed empirically by developing a model-based test instrument which can be adopted in further studies. This study analysed which teacher variables have a strong relationship with CK and PCK and observed changes in CK and PCK at the different steps of teacher education.

## **INHALT**

1	Einleitung .....	1
2	Professionswissen in den Naturwissenschaften.....	5
2.1	Verständnis des Professionswissens .....	6
2.2	Modelle des Professionswissens .....	8
2.3	Wissen, Handeln und Unterrichtsqualität.....	11
2.4	Quantitative Studien zum Professionswissen .....	12
2.5	Zusammenhang zwischen den Dimensionen des Professionswissens .....	14
2.6	Erwerb des Professionswissens.....	16
3	Modell zur Definition des Professionswissens .....	23
3.1	Zusammenhang der Dimensionen des Professionswissens CK, PCK, PK: Annahme für diese Arbeit .....	24
3.2	Aspekte zur Operationalisierung des CK, PCK und PK .....	25
3.2.1	Wissensbereiche.....	25
3.2.2	Wissensstufen.....	27
3.2.3	Themenabhängigkeit .....	29
3.2.4	Facetten des PCK .....	30
3.2.5	Facetten des PK .....	35
3.3	Gesamtes Modell.....	36
4	Ziele der Untersuchung und Forschungsfragen.....	37
4.1	Ziele .....	38
4.2	Forschungsfragen und Hypothesen .....	38
5	Anlage der Untersuchung .....	41
5.1	Beschreibung der Testinstrumente.....	42
5.1.1	CK-Test und PCK-Test .....	42
5.1.1.1	Testentwicklung.....	42
5.1.1.2	Testformat .....	44
5.1.1.3	Weitere Charakteristika der Tests .....	46
5.1.2	PK-Test .....	47
5.1.2.1	Testentwicklung.....	47
5.1.2.2	Testformat .....	47
5.1.2.3	Weitere Charakteristika der Tests .....	48
5.2	Testdurchführung.....	50
5.3	Stichprobenbeschreibung .....	51

6	Auswertung .....	56
6.1	Kodierung und Interraterreliabilität.....	57
6.2	Rasch-Analyse.....	58
6.2.1	Berechnung der Personenfähigkeit.....	63
6.3	Klassische Analyse .....	64
7	Ergebnisse .....	67
7.1	Ergebnisse zur Objektivität .....	68
7.2	Ergebnisse zur Modellierung.....	68
7.2.1	Erste Dimensionsmodellierung .....	69
7.2.2	Einfluss DIF zwischen den Studien .....	69
7.2.3	Itemfit und Sicherstellung der Dimensionsmodellierung .....	71
7.2.4	Reliabilität im besten Modell .....	73
7.3	Ergebnisse zur Validität .....	75
7.4	Zusammenfassung Testgüte.....	80
7.5	Ergebnisse zu den Personenmerkmalen .....	80
7.5.1	Zusammenhang von CK, PCK und PK.....	81
7.5.2	Korrelationen zwischen den unabh. Variablen und CK sowie PCK .....	84
7.5.3	Regression.....	86
7.5.4	Ausbildungsschritte .....	90
7.5.4.1	Mittelwertunterschiede zwischen den Ausbildungsstufen.....	90
7.5.4.2	Zusammenhang zwischen CK und PCK .....	94
8	Diskussion und Ausblick.....	97
8.1	Rekapitulation der Ziele der Studie.....	98
8.2	Messung des Professionswissens.....	98
8.2.1	Messung des PCK.....	100
8.2.2	Messung des CK .....	101
8.2.3	Messung des PK .....	103
8.2.4	Modell.....	103
8.3	Was lässt sich aus dem Zusammenhang zwischen CK, PCK und PK .....	104
	schließen? .....	104
8.4	Wie hängen Personenmerkmale mit Professionswissen zusammen? .....	106
8.5	Unterschiede zwischen den Ausbildungsstufen .....	113
8.6	Ausblick.....	115
	Verzeichnisse .....	119
	Anhang .. .....	135

*1. KAPITEL*  
*EINLEITUNG*

**Was müssen Lehrkräfte wissen?** Als sicher gilt, dass Lehrkräfte Wissen benötigen, um guten Unterricht zu ermöglichen. Dies scheint zunächst eine oberflächliche Aussage zu sein, bei Betrachtung der Art des im Unterricht wirksamen und von der Lehrperson benötigten Wissens wird deutlich, dass dies differenziert analysiert werden muss. Dieser Studie liegt die Annahme zu Grunde, dass Professionswissen, das sich aus den Dimensionen Fachwissen, fachdidaktisches Wissen und pädagogisches Wissen zusammensetzt, für guten Unterricht relevant ist. Es besteht bisher weder in der Literatur noch in der Forschungsgesellschaft Konsens darüber, was sich hinter dem Begriff Professionswissen verbirgt. Uneinigkeit besteht auch darüber, ob Professionswissen verbalisierbar sein muss oder implizit vorliegen kann und wie es mit dem Handeln der Lehrkraft in Beziehung steht. Insbesondere fehlen für das Fach Physik umfangreich validierte deutschsprachige Testinstrumente.

In dieser Arbeit wird das verbalisierbare Professionswissen von Physiklehrkräften modelliert und das Modell mit einem Test operationalisiert, um Aussagen über den Zusammenhang zu Lehrervariablen machen zu können und um in einem Quasi-Längsschnitt Unterschiede im Professionswissen zu beobachten. Um dies zu beschreiben, gliedert sich die vorliegende Arbeit in acht Kapitel, über die an dieser Stelle ein Überblick gegeben wird.

Kapitel 2 hat den aktuellen Stand der Forschung bezüglich des Professionswissens zum Inhalt. Der Forschungsbedarf wird skizziert und die Dimensionen des Professionswissens werden beschrieben und miteinander in Beziehung gesetzt. Nach einem Exkurs zu dem Zusammenhang von Professionswissen, Lehrerhandeln und Unterrichtsqualität, werden mehrere quantitative Studien zum Professionswissen (angehender) Lehrkräfte der Mathematik und Naturwissenschaften vorgestellt. Ausgewählte Ergebnisse dieser Studien werden thematisch zusammengefasst. Es wird dargestellt, wie die Dimensionen des Professionswissens in den Studien korrelieren und welche Personenmerkmale mit Fachwissen und fachdidaktischem Wissen zusammenhängen. In der Darstellung wird ein Schwerpunkt auf das fachbezogene Professionswissen gelegt. Dieser Abschnitt bildet die Grundlage für die Hypothesenbildung der vorliegenden Studie zur Analyse des Professionswissens.

In Kapitel 3 wird zunächst darauf eingegangen, welche Aspekte in verschiedenen Studien und theoretischen Arbeiten zur Operationalisierung von Professionswissen dienen. Daraus wird ein Modell des Professionswissens abgeleitet, welches Professionswissen als Grundlage für die weitere Arbeit definiert.

In Kapitel 4 werden die Ziele der Arbeit, die Analyse und Modellierung des Professionswissens von Physiklehrkräften, konkreter definiert. Dazu werden auf Grundlage der vorangegangenen Kapitel Forschungsfragen gestellt und Hypothesen formuliert.

Die Entwicklung der Tests auf Basis des Modells und deren Einsatz wird in Kapitel 5 dargelegt. Daran schließt sich eine Beschreibung der Stichprobe an.

Kapitel 6 ist der Auswertung der Daten gewidmet. Es wird die Frage behandelt, wie sich aus Antworten von (angehenden) Physiklehrkräften auf Testitems ein Bild über ihr Professionswissen formt.

Die Ergebnisse der Studie werden in Kapitel 7 vorgestellt. Das Kapitel beinhaltet den Zusammenhang der Dimensionen, die Testgüte, den Zusammenhang zwischen Personenmerkmalen und dem fachspezifischen Professionswissen auf Basis von Korrelationsanalysen und Regressionen, sowie den quasi-längsschnittlichen Vergleich des Wissens von Physiklehrkräften der verschiedenen Ausbildungsstufen.

In Kapitel 8 werden die Ergebnisse der vorgestellten Studie, sowie das Modell, auf dem sie gründen, diskutiert und kritisch reflektiert. Abschließend wird ein Ausblick auf mögliche Reanalysen, eine Erweiterung des Testinstruments und eine bereits begonnene Folgestudie gegeben.



# *2. KAPITEL*

## *PROFESSIONSWISSEN IN DEN NATURWISSENSCHAFTEN*

Obwohl Professionswissen als wichtige Grundlage für guten Unterricht gilt, besteht Uneinigkeit über das Konzept und die Modellierung in Dimensionen. In diesem Kapitel wird die Relevanz des Konzepts Professionswissen geklärt (2.1) und dieses anhand mehrerer Modelle vorgestellt (2.2). Der Zusammenhang zwischen Wissen und Handeln wird skizziert (2.3). Anhand einiger quantitativer Studien (2.4) werden der Zusammenhang zwischen den Dimensionen des Professionswissens (2.5) sowie der Erwerb des Professionswissens (2.6) beleuchtet.

## 2.1 VERSTÄNDNIS DES PROFESSIONSWISSENS

Professionswissen ist das Wissen, das Menschen brauchen, um ihren Beruf auszuüben, ihrer Profession nachzugehen. Im Falle von Lehrkräften eines Faches handelt es sich also um das Wissen, das sie benötigen, um ihr Fach erfolgreich zu unterrichten. Der Begriff des Professionswissens wird im Folgenden als Kurzfassung für das Professionswissen von Lehrkräften verwendet. Üblicherweise werden für das Professionswissen drei Bereiche genannt (Baumert & Kunter, 2013):

Fachwissen (Content Knowledge, CK) (oft auch Subject Matter Knowledge genannt, SMK, Abell, 2007),

Generelles Pädagogisches Wissen (General Pedagogical Knowledge, PK),

Fachdidaktisches Wissen (Pedagogical Content Knowledge, PCK),

in Nordamerika oft ergänzt um das Wissen über (Schul-) Kontexte (z. B. Gess-Newsome, 1999).

Das Fachwissen, das eine Lehrkraft braucht, um Unterricht zu planen, durchzuführen und nachzubereiten, wird nachfolgend mit CK abgekürzt. PK, das pädagogische Wissen, ist fachunabhängig und befähigt die Lehrkraft, ein Klima herzustellen, in dem Unterricht stattfinden kann (Voss, Kunter, & Baumert, 2011). Der Begriff PCK für das fachdidaktische Wissen wurde erstmals von Shulman (1986) als Kategorie von Fachwissen eingeführt und später als „special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding“ (Shulman, 1987, S. 8) ausgeschärft und wird seither als das Wissen, das einzig und allein die Lehrkräfte eines bestimmten Faches haben und brauchen, verwendet.

Shulman stellt das Wissen über Fachcurricula, über die Psychologie Lernender, über Organisation bzw. Kontexte und über Erziehungsphilosophie, Bildungstheorie und –historie (übersetzt in Anlehnung an Baumert & Kunter, 2006), neben PCK, CK und PK, ohne weitere Aspekte explizit auszuschließen. In dieser Kategorisierung des Wissens von Lehrkräften hebt er PCK als Kern des Professionswissens von Lehrkräften aufgrund seiner Einzigartigkeit für Fachlehrkräfte besonders hervor (Shulman, 1987). Kind ergänzt: „The attraction of PCK lies in its ability to tell us something of the unique professional experience that constitutes teaching.“ (Kind, 2009, S. 198).



**Abbildung 1:** Die Aufteilung des Professionswissens in die drei Bereiche CK, PCK und PK

Schon die englischsprachige Wortwahl des Pedagogical Content Knowledge legt die enge Verbindung der drei Bereiche CK, PCK und PK nahe (siehe Abbildung 1). Weitere Kategorien sind von diesem Dreiklang ausgeschlossen und werden nicht flächendeckend genannt (Baumert & Kunter, 2006). Das Wissen über Organisation bzw. Kontexte bleibt allerdings in verhältnismäßig vielen Arbeiten als Kategorie neben CK, PCK und PK bestehen (vgl. Gess-Newsome, 1999; Kunter et al., 2011; Magnusson, Krajcik, & Borko, 1999). Einzelne Kategorien, wie Wissen über Curricula, über die Psychologie Lernender, über Erziehungsphilosophie, Bildungstheorie und -historie, können in Operationalisierungen von PCK und PK aufgehen (siehe 3.2.4 und 3.2.5, Facetten des PCK und PK).

Die Einführung des Konzepts PCK führte international zu einer Veränderung des Forschungsschwerpunktes in der Lehrerforschung weg vom Persönlichkeits-Produkt-Paradigma (z. B. Anderson, Evertson, & Brophy, 1979), hin zur Erforschung des Professionswissens. Es wird heute davon ausgegangen, dass weder allein die Persönlichkeit der Lehrkraft noch allein ihr Fachwissen entscheidend für erfolgreiches Unterrichten ist (Abell, 2007), sondern dass das Professionswissen der Lehrkraft eine Grundlage dafür bildet (Fischer, Borowski, & Tepner, 2012), auch wenn der Effekt auf Schülerleistung und -motivation weiter untersucht werden muss (Abell, 2007).

Obwohl dem Professionswissen eine solche Wichtigkeit zugeordnet wird, ist weiterhin klarer Forschungsbedarf zu erkennen: Erstens besteht bisher keine Einigkeit über das Konzept (Abell, 2007; Gess-Newsome, 2011), zweitens werden weiterhin objektive, reliable und valide Testinstrumente benötigt (Baumert & Kunter, 2013; Gess-Newsome, 2011). Da davon ausgegangen wird, dass die Grundlage des Professionswissens bereits

in der Ausbildung gelegt wird (z. B. Gesetz über die Ausbildung für Lehrämter an öffentlichen Schulen, 2012; Riese, 2009; Terhart, 2002), könnte detailliertes Wissen über Professionswissen die Lehrerausbildung direkt beeinflussen und verbessern (Abell, 2007).

### **2.2 MODELLE DES PROFESSIONSWISSENS**

Wie bereits erwähnt, herrscht keine Einigkeit über das Konzept Professionswissen. Im deutschsprachigen Raum ist eine Trennung in die drei Bereiche CK, PCK und PK üblich (siehe Abbildung 1, vgl. das Kompetenzmodell nach Kunter et al., 2011 (Abbildung 3); Baumert & Kunter, 2006; Blömeke et al., 2010; Borowski et al., 2010a; Riese, 2009; Terhart, 2002), die der Aufteilung in der Lehramtsausbildung entspricht (vgl. Fischler, 2008).

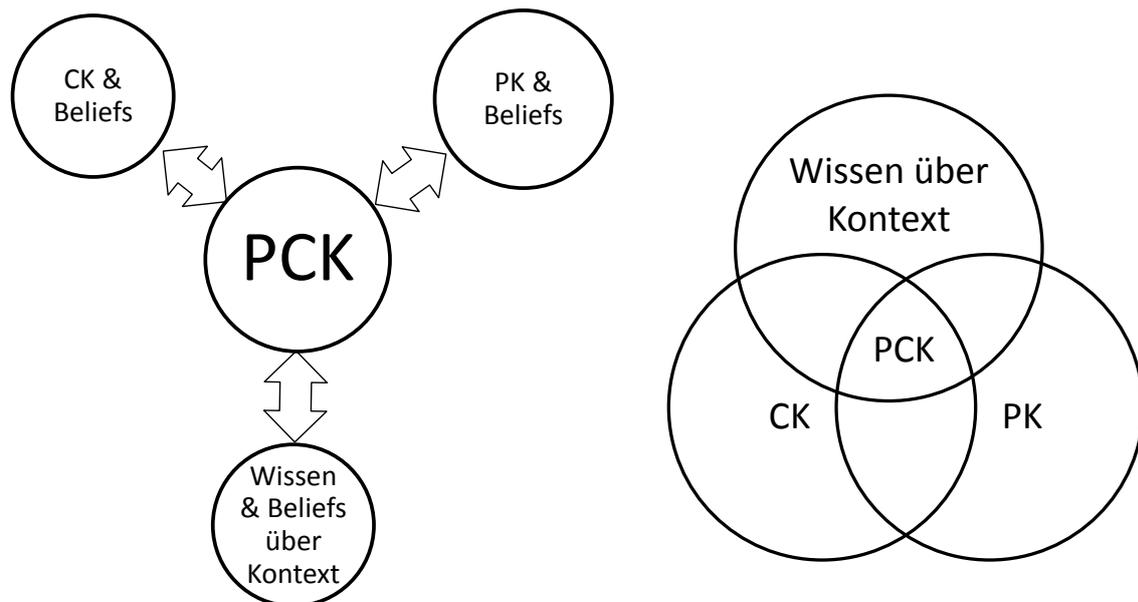
Nach Gess-Newsome sind PK und CK international in den meisten Definitionen in PCK eingebettet (Gess-Newsome, 2011). Viele Autorinnen und Autoren beschäftigen sich nur mit dem Bereich PCK (z. B. Park & Chen, 2012; Schmelzing, 2010; van Driel, Verloop, & Vos, 1998), ohne CK und PK als weitere Dimensionen zu untersuchen.

Zwei Modelle, die CK und PK explizit einbeziehen, sind das Modell nach Magnusson (Magnusson et al., 1999) und das von Gess-Newsome genutzte Modell (Gess-Newsome, 1999). Das Modell nach Magnusson (siehe Abbildung 2) sieht PCK als eine Transformation aus CK, PK und Wissen über Kontext. Die Transformation schafft etwas Neues – PCK. Die anderen Wissensbereiche bleiben zwar erhalten, gelten aber erst nach ihrer Transformation zu PCK als brauchbar für Unterricht (Gess-Newsome, 1999).

Im Gegensatz dazu beschreibt das von Gess-Newsome genutzte Modell (ebenfalls Abbildung 2) PCK als eine Integration der anderen Wissensbereiche. Die einzelnen Wissensbereiche werden als Entscheidungsgrundlage genutzt und nur im Moment des Handelns von der Lehrkraft zu PCK zusammengefasst (Gess-Newsome, 1999).

Allerdings deuten die neusten Ergebnisse an, dass erstens das gemessene PCK ein eigenes, messbares Konstrukt ist, das sich in zwei Faktoren, PCK-CK und PCK-PK, einteilen lässt, die disjunkt neben dem generellen PK und dem akademischen CK stehen und zweitens PK einen Einfluss auf Unterrichtshandeln hat (Gess-Newsome et al., In review). Diese Ergebnisse widersprechen sowohl dem transformativen Modell, da ein anderer Bereich als PCK einen Einfluss auf Unterricht hat, als auch dem integrativen Mo-

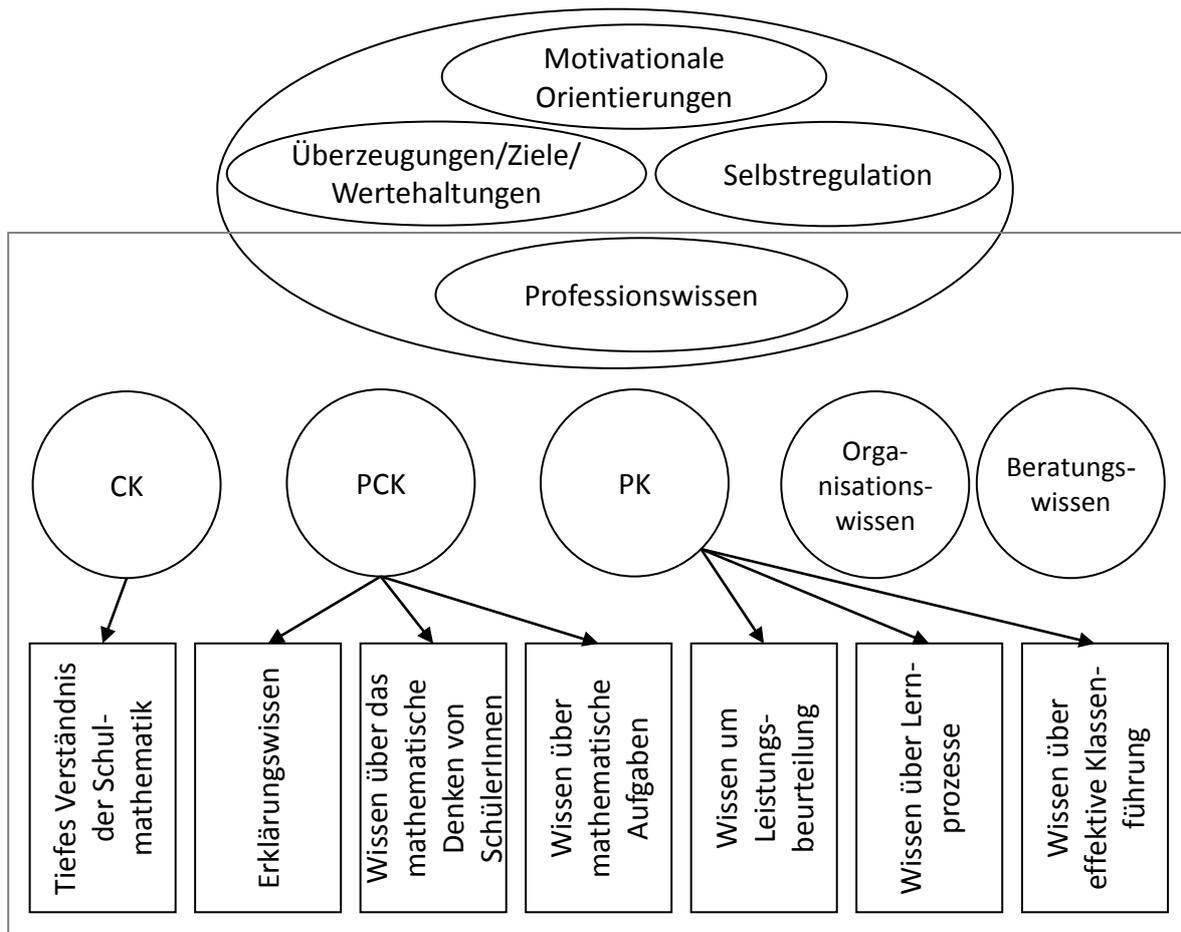
dell, da die Bereiche nebeneinander stehen und sich nicht durchdringen. Beide Ergebnisse finden sich nicht nur in der beschriebenen Studie, sondern z. B. auch in der COACTIV-Studie (Baumert et al., 2010, 2.4). Damit erscheinen weder eine transformative noch eine integrative Modellierung von PCK zielführend. Als erfolgversprechender zeichnet sich die Annahme disjunkter, für Unterricht relevanter Dimensionen ab.



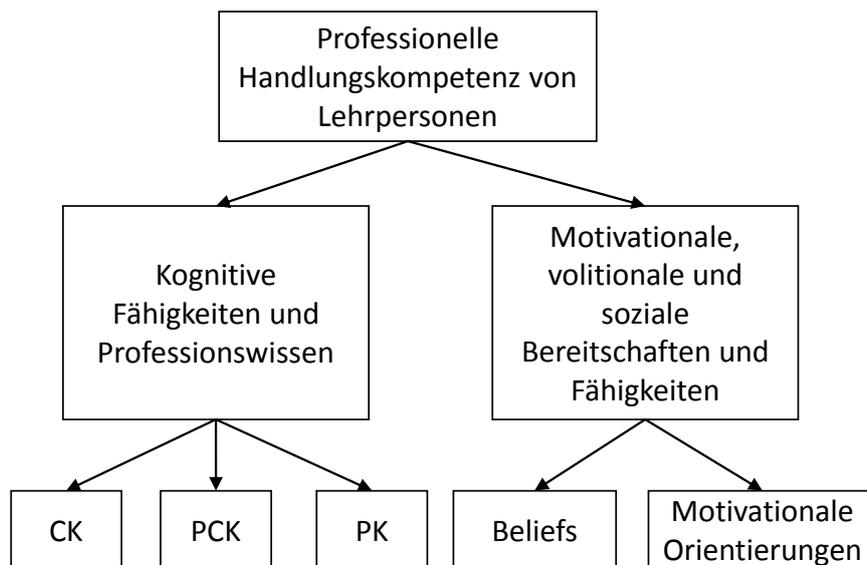
**Abbildung 2:** Modelle von PCK

Links: Das transformative Modell zur Beziehung zwischen den Bereichen des Professionswissens nach Magnusson et al. (1999, S. 98); Rechts: Das integrative Modell von PCK nach Gess-Newsome, (Zusammenführung von Gess-Newsome (1999) und Gess-Newsome et al. (In review)).

Doch nicht nur CK, PCK und PK machen eine Lehrkraft aus: Insbesondere im Kompetenzmodell der COACTIV-Studie ist zu erkennen, dass das Professionswissen nicht allein steht (Abbildung 3): Überzeugungen, Werthaltungen und Ziele, motivationale Orientierungen und Selbstregulation gehören zur professionellen Handlungskompetenz (Kunter et al., 2011). In älteren Modellen wie dem Magnusson-Modell, sind Wissen und Beliefs (bzw. Überzeugungen, Orientierungen, Werthaltungen) nicht eindeutig voneinander zu trennen. In Modellen, die als Grundlage zur Testentwicklung dienen (besonders prägnant in Abbildung 4 zu erkennen) ist das anders (vgl. Baumert & Kunter, 2006).



**Abbildung 3:** Kompetenzmodell von COACTIV mit Schwerpunkt auf dem Professionswissen (Kunter et al., 2011, S. 32)

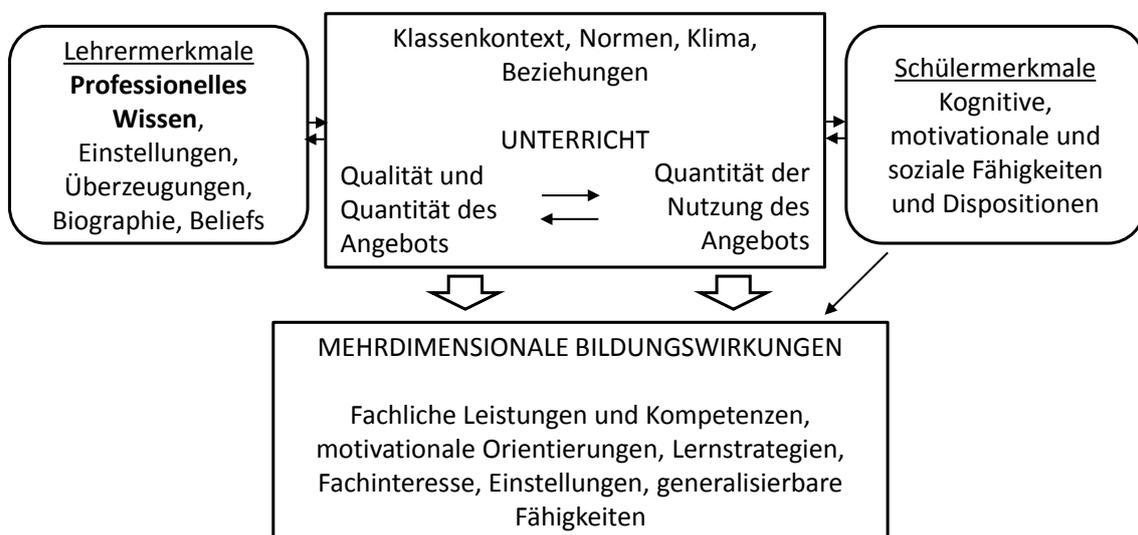


**Abbildung 4:** Kompetenzstrukturmodell (Riese, 2009, S. 26, in Anlehnung an Baumert & Kunter, 2006; Blömeke, Felbrich, & Müller, 2008)

### 2.3 WISSEN, HANDELN UND UNTERRICHTSQUALITÄT

Obwohl Professionswissen als wichtige Grundlage für lernförderlichen Unterricht gilt, ist nicht zweifelsfrei nachgewiesen, dass ein höheres Professionswissen einen positiven Einfluss auf Unterrichtsqualität hat (Terhart, 2002). Kolbe (2004) beschreibt, dass Wissen und Handeln im Lehrerberuf eng und komplex zusammenhängen. Bisher gibt es nur wenige empirische Untersuchungen zum Zusammenhang zwischen Wissen, Handeln und Schülerleistung (Fischer et al., 2012). Nach Fischler (2008) fehlen selbst belastbare Daten über einen Zusammenhang zwischen Wissen und Können. Einerseits fehlen valide Instrumente, um physikspezifisches Professionswissen zu verschiedenen Themen zu messen. Andererseits müsste der Zusammenhang zu Lehrerhandeln und Schülerwissen, bzw. –motivation mit ebendiesen Instrumenten quantitativ untersucht werden.

Eine Einordnung des Professionswissens in den Rahmen des Schulalltags bietet das Angebots-Nutzungs-Modell von Reusser und Pauli, das einen Einfluss des Professionswissens auf den Unterricht und damit unter anderem auch auf die Schülerleistung und –motivation impliziert (siehe Abbildung 5).



**Abbildung 5:** Ausschnitt aus dem Angebots-Nutzungs-Modell (nach Reusser & Pauli, 2003, S. 3)

Quantitative Studien zum Professionswissen in der Mathematik und den Naturwissenschaften

## **2.4 QUANTITATIVE STUDIEN ZUM PROFESSIONSWISSEN IN DER MATHEMATIK UND DEN NATURWISSENSCHAFTEN**

Mittlerweile wurden mehrere größere, quantitative Studien zum Professionswissen von Lehrkräften der Mathematik und Naturwissenschaften in Deutschland durchgeführt. Eine Auswahl soll hier kurz vorgestellt und durch Stichprobengröße und Hauptfragestellungen charakterisiert werden. Für die Studien im Bereich der Physik wird ein Überblick über die Validierung der Instrumente ergänzt. Die für den weiteren Verlauf dieser Arbeit relevanten Ergebnisse werden in den darauf folgenden Abschnitten dargestellt.

Vorgestellt werden die COACTIV-Studie (Cognitive Activation in the Classroom / Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz, Baumert et al., 2010), die MT21-Studie (Mathematics Teaching in the 21st Century, Blömeke, Kaiser, & Lehmann, 2008), die Studie von Riese und Reinhold (Riese, 2009), die QuiP-Studie (Quality of Instruction in Physics – comparing instruction and performance in Finland, Germany and Switzerland, Neumann, Fischer, Labudde, & Viiri, 2010), und das ProwiN-Projekt (Professionswissen in den Naturwissenschaften, Borowski et al., 2010a).

In der **COACTIV-Studie** wurde mit einer Stichprobe von 194 nach Bundesland und Schulzweig für Deutschland repräsentativen Klassen mit 4353 Schülerinnen und Schülern und ihren 181 Mathematiklehrkräften folgendes gefunden: CK und PCK sind verschiedene Wissensbereiche und PCK hat einen Einfluss auf das Lernen der Schülerinnen und Schüler, der durch die Instruktionsqualität mediiert wird (Baumert et al., 2010). Die entsprechenden schriftlichen Tests, ergänzt um einen umfangreich validierten PK-Test, wurden auch im Rahmen der COACTIV-R Studie mit 746 Mathematiklehrkräften im Vorbereitungsdienst genutzt (vgl. Voss et al., 2011).

In der **MT21-Studie** wurde der Effekt der Lehrerausbildung in Bezug auf das Professionswissen empirisch erfasst. Dazu wurden CK, PCK und PK von 849 angehenden Mathematiklehrkräften für die Sekundarstufe I, also Lehramtsstudierenden und Lehrkräften im Vorbereitungsdienst, schriftlich erhoben (Blömeke et al., 2008).

**Riese und Reinhold** entwickelten ein Kompetenzstrukturmodell und ein Testinstrument zum Professionswissen. In einer Stichprobe von 301 Studierenden des Physiklehramts für die Sekundarstufen I und II wurde das Professionswissen im Inhaltsbereich

Mechanik und dessen Entwicklung quasi-längsschnittlich gemessen, sowie der Zusammenhang zwischen den Dimensionen CK, PCK und PK untersucht (Riese, 2009).

In dem von **Olszewski** bearbeiteten Teilprojekt der **QuiP-Studie** standen unter anderem die Entwicklung eines Testinstruments zu PCK im Inhaltsbereich Elektrizitätslehre und die Validierung ihres Modells von PCK im Vordergrund (Olszewski, 2010).

Die Stichprobe bestand aus 103 Klassen aus Gymnasien, Real- und Gesamtschulen aus Deutschland, Finnland und dem deutschsprachigen Teil der Schweiz. Für die Auswertung des PCK-Tests wurden die Antworten von 41 deutschen, 26 schweizer und 25 finnischen Physiklehrkräfte herangezogen (Olszewski, 2010). Informationen über den in der QuiP-Studie genutzten CK-Test sind bisher nicht veröffentlicht.

Die Kriteriums- und Konstruktvalidität der PCK-Tests von Olszewski und Riese wurde in einer gemeinsamen Studie untersucht. Erwartungsgemäß schnitten die 27 Studierenden in beiden PCK-Tests schlechter ab als die 42 Lehrkräfte im Vorbereitungsdienst und diese wiederum schlechter als die 8 untersuchten Lehrkräfte (davon 6 Fachseminarleiter); für die gesamte Gruppe war die Korrelation zwischen den beiden PCK-Tests größer als die zu den jeweiligen CK-Tests (kein Signifikanztest angegeben) (Borowski, Olszewski, & Fischer, 2010b; Borowski & Riese, 2010; Olszewski, 2010; Riese, 2009).

Zur weiteren Validierung des PCK-Tests von Riese wurden Experten interviewt (Riese, 2009). Es findet sich kein Beleg, dass der von Riese entwickelte und genutzte CK-Test validiert wurde. Olszewski testete für eine kriteriale Validierung ihres PCK-Tests ergänzend Lehrkräfte anderer Fächer und Physikerinnen und Physiker, die im Mittel jeweils schlechter als Physiklehrkräfte abschnitten (Olszewski, 2010).

Bisher wurden im deutschsprachigen Raum bereits zwei PCK-Tests im Fach Physik eingesetzt und auf unterschiedliche, allerdings bisher nicht hinreichende Weise, validiert. Die Validierung der entsprechenden CK-Tests steht noch aus. Der Bedarf nach einem deutschsprachigen, objektiven, reliablen und validen Testinstrument für das Professionswissen, also CK, PCK und PK, bleibt für das Fach Physik bestehen.

Im **ProwiN-Projekt**, in das diese Arbeit eingebunden ist, wird der Zusammenhang zwischen CK, PCK und PK in den drei Fächern Biologie, Chemie und Physik untersucht und fächerübergreifend verglichen (Borowski et al., 2010a). In der darauf folgenden Pro-wiN-Videostudie soll der Zusammenhang zwischen Professionswissen und Unterrichtsqualität untersucht werden (Cauet, Borowski, & Fischer, 2012; Fischer et al., 2011). Die Stichprobe der Fachlehrkräfte besteht für die Hauptfragestellung aus 158 Biologielehrkräften (Jüttner, Boon, Park, & Neuhaus, 2013), 166 Chemielehrkräften (Dollny, 2011)

und 186 Physiklehrkräften, ergänzt um Studierende, Lehrkräfte im Vorbereitungsdienst, Lehrkräfte anderer Fächer, Wissenschaftlerinnen und Wissenschaftler zur Validierung und Untersuchung ergänzender Fragestellungen (Dollny, 2011; Kirschner, Borowski, & Fischer, 2013; Mutke & Tepner, 2013; Pollender & Tepner, 2012)).

In den folgenden Abschnitten zum Zusammenhang zwischen CK, PCK und PK (2.5) sowie dem Erwerb des Professionswissens (2.6) werden Ergebnisse der beschriebenen Studien thematisch geordnet vorgestellt.

### **2.5 ZUSAMMENHANG ZWISCHEN DEN DIMENSIONEN DES PROFESSIONSWISSENS**

In diesem Abschnitt werden Ergebnisse zu dem Zusammenhang zwischen CK, PCK und PK vorgestellt. Die eingangs vorgestellten Modelle des Professionswissens machen deutlich, dass die Bereiche eng zusammengehören (siehe auch Abbildung 1). Sie sind aber auch durch eine Abgrenzung voneinander definiert: CK grenzt sich von PCK ab, da es keinen pädagogischen Anteil hat, PK grenzt sich von PCK dadurch ab, dass es keinen Fachanteil hat. PCK hat immer einen fachlichen und einen pädagogischen Anteil. Es scheint auch vorgegeben, dass CK und PK jeweils enger mit PCK verknüpft sind als miteinander. Die drei Bereiche sind im Rahmen des Professionswissens kaum unabhängig voneinander denkbar, auch wenn nicht in jeder Studie CK, PCK und PK gemeinsam untersucht werden.

In Tabelle 1 sind Korrelationen zwischen CK, PCK und PK dargestellt, die in den im vorangegangenen Abschnitt vorgestellten Studien gefunden wurden. Es ist zu sehen, dass die Korrelation zwischen CK und PCK, die in sämtlichen Studien untersucht wurde, sehr unterschiedlich ausfällt. Insbesondere der Schulzweig bzw. dessen Kontrolle scheint eine Rolle zu spielen. Die Korrelationen sind allerdings nur eingeschränkt vergleichbar, da für Ihre Berechnung unterschiedliche Verfahren, Operationalisierungen und Stichproben gewählt wurden.

Für den Mathematikunterricht gilt CK als notwendige, aber nicht hinreichende Bedingung für PCK, da keine oder nur sehr wenige Versuchspersonen mit niedrigem CK und gleichzeitig hohem PCK gefunden wurden, wohl aber Versuchspersonen mit hohem CK und niedrigem PCK (Baumert et al., 2010; Krauss et al., 2008). Für Studierende des Physiklehramts konnte bestätigt werden, dass CK keine hinreichende Bedingung für PCK ist (Riese, 2009).

**Tabelle 1:** Korrelationen zwischen den Dimensionen des Professionswissens

Korrelation	Krauss, Neubrand, Blum, & Baumert, 2008	Voss et al., 2011 <sup>1</sup>	Riese, 2009 <sup>2</sup>	Dollny, 2011	Mutke & Tepner, 15.4.13	Jüttner et al., 2013
CK zu PCK	r=.60, p<.05	r <sub>latent</sub> =.91, p<.05	r=.49, p < .001 r <sub>latent</sub> = .68, p < .001	<sup>3</sup> r <sub>s</sub> =.176, p=.176 <sup>4</sup> r <sub>s</sub> =.498, p<.006	r=.535, p < .001	r=.22, p<.01
PK zu PCK		r <sub>latent</sub> =.42, p<.05	r=.45, p < .001 r <sub>latent</sub> =.70, p < .001		nicht signifi- kant	
CK zu PK		r <sub>latent</sub> =.24, p<.05	r=.29, p < .001 r <sub>latent</sub> =.38, p < .001		nicht signifi- kant	

<sup>1</sup> Unter Kontrolle des Schulzweigs

<sup>2</sup> Die Korrelationen für Studierende des Physiklehramts über die verschiedenen Studiengänge und den Studienfortschritt sind vergleichbar hoch

<sup>3</sup> Nur Gymnasiallehrkräfte

<sup>4</sup> Keine Gymnasiallehrkräfte

PCK klärt genauso viel Varianz an CK auf, wie umgekehrt. Es wird allerdings darüber diskutiert, ob es sich um einen originären oder nur um einen statistischen Zusammenhang handelt. Es scheint eindeutig zu sein, dass ein gewisses CK benötigt wird (und während der Schulzeit als Schülerin oder Schüler auch erworben wird), bevor PCK aufgebaut werden kann. Allerdings sollte danach die Beschäftigung mit dem Stoff im Schulkontext (Erwerb von PCK) auch wieder zu einem höheren CK führen (van Driel, 2013). Im weiteren Verlauf wird CK als Bedingung für PCK betrachtet und der Einfluss von PCK auf CK nicht weiter untersucht.

Oft wird nicht weiter differenziert, ob das Professionswissen von Studierenden, Lehrkräften im Vorbereitungsdienst oder Lehrkräften untersucht wurde. Riese beispielsweise nutzt ein Modell für das Wissen von Lehrkräften und Lehramtsstudierenden (Riese, 2009). Dies entspringt der Vermutung, dass das Professionswissen aller Gruppen prinzipiell ähnlich strukturiert ist. Diese einfache Annahme stammt daher, dass

keine Large-Scale-Studien veröffentlicht wurden, in denen mindestens zwei Dimensionen des Professionswissens aller drei Gruppen nicht nur mit denselben Instrumenten erfasst (wie von Kleickmann et al. (2012) durchgeführt), sondern auch korreliert wurden. Es gibt also keinen Anhaltspunkt dafür, dass die Zusammenhänge zwischen den Dimensionen unterschiedlich sind. Allerdings handelt es sich hierbei um einen Schluss aus fehlenden Daten, der Zusammenhang muss für alle drei Gruppen empirisch untersucht werden.

### 2.6 ERWERB DES PROFESSIONSWISSENS

In diesem Abschnitt wird dargelegt, in welchen Situationen Professionswissen erworben werden kann und welche Voraussetzungen den erfolgreichen Erwerb begünstigen könnten. Es werden insbesondere Lerngelegenheiten analysiert und Ergebnisse der bereits vorgestellten Studien dargestellt.

Es ist nicht abschließend geklärt, wie und wo Professionswissen erworben wird (Baumert & Kunter, 2013; Riese, 2009). Da die Aus- und Weiterbildung von Lehrkräften in vier Teile geteilt ist - 1) der Schulbesuch bis zur allgemeinen Hochschulreife, 2) das Studium, 3) der Vorbereitungsdienst, 4) Schulpraxis und Weiterbildung - sollten alle Phasen Lerngelegenheiten zum Erwerb des Professionswissens bieten. Strukturierte Lerngelegenheiten sollten, neben Fortbildungen, vor allem die Hochschulausbildung, der Vorbereitungsdienst und für den Bereich des Schulwissens das Abitur offerieren.

Als ein Ziel der ersten Phase der Lehrerausbildung gilt der Erwerb des theoretischen Professionswissens (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2010). Riese beschreibt, dass Lehramtsstudierenden „die erforderlichen Routinen und Handlungsschemata noch nicht zur Verfügung stehen können, die für ein Handeln im (Zeit-)Druck der Unterrichtspraxis erforderlich sind.“ (2009, S. 76). Lehrkräfte im Vorbereitungsdienst haben die Gelegenheit, abstraktes Wissen mit eigenen Erfahrungen und angeleiteter Reflexion zu verbinden (vgl. van Driel, Jong, & Verloop, 2002).

Wird davon ausgegangen, dass allein die Anzahl der **Lerngelegenheiten** zählt, müssten Lehrkräfte mit mehr Berufserfahrung immer besser abschneiden. Allerdings nehmen Brunner et al. (2006) an, dass strukturierte Übungsaktivitäten benötigt werden. Lerngelegenheiten allein reichen nicht aus, sie müssen auch genutzt werden. Für Lehrkräfte

bedeutet dies insbesondere, dass sie ihre Tätigkeiten und Handlungen reflektieren müssen (Baumert & Kunter, 2013; Cochran, DeRuiter, & King, 1993; vgl. auch Kolbe, 2004). Lehrkräften selbst bieten sich in jeder Unterrichtsstunde Lerngelegenheiten, sie haben nach dem Abschluss ihrer Ausbildung täglich die Möglichkeit zur Reflexion .

Der Begriff Time on Task beschreibt nicht die Gelegenheit zu lernen, sondern die aktive Lernzeit. Es wird angenommen, dass sich ihr Einfluss auf Lernfortschritte von Schülerinnen und Schülern auf die Fortschritte (angehender) Lehrkräfte übertragen lässt (Hattie, 2009; Riese, 2009). Allerdings lässt sich die aktive Lernzeit weniger gut mit Hilfe eines retrospektiven Fragebogens erfassen als die Anzahl der Lerngelegenheiten, weshalb diese im weiteren Verlauf als Argumentationshilfe dienen.

Lehrkräfte in Deutschland müssen sich zwar fortbilden, allerdings muss der Schwerpunkt dabei nicht auf fachdidaktischen oder fachlichen Inhalten liegen (Brunner et al., 2006; vgl. beispielsweise Bayrisches Staatsministerium für Unterricht und Kultus für Bayern oder Bezirksregierung Düsseldorf für NRW), was dazu führen könnte, dass durch fehlenden externen Anreiz wenig reflektiert wird und dass die Lehrkräfte neues fachdidaktisches Wissen nicht erreicht.

Auch ohne weitere Reflexion müsste Routine zu einem höheren Fachwissen führen, da viele Inhalte jedes Jahr unterrichtet und daher memoriert werden. Zusätzlich werden weitere Inhalte für die Unterrichtsvorbereitung neu erarbeitet, was ebenfalls zu einem Anstieg des Fachwissens führen könnte.

Es kann deshalb angenommen werden, über die verschiedenen **Ausbildungsstufen** ein Wachstum in CK zu messen. Empirische Ergebnisse zu CK und PCK in den Ausbildungsstufen sind in Tabelle 2 dargestellt. Während der formalen Ausbildung von Lehrkräften der Mathematik und der Naturwissenschaften scheinen CK und PCK stetig zu steigen. Riese und Olszewski halten diese Gruppenunterschiede im Bereich PCK für ein Zeichen von Validität (Olszewski, 2010; Riese, 2009).

**Tabelle 2:** Wachstum des Professionswissens in den Dimensionen in und zwischen den Ausbildungsstufen

Studium	<u>CK und PCK</u> Lehramtsstudierende schneiden im Grundstudium signifikant schlechter ab als im Hauptstudium (Mathematik: Blömeke et al., 2008; Physik: Riese & Reinhold, 2012)
Studium und Vorbereitungs- dienst	<u>CK und PCK</u> Lehramtsstudierende schneiden signifikant schlechter ab als Lehrkräfte im Vorbereitungsdienst (Mathematik: Blömeke et al., 2008; Physik: Borowski, Kirschner, Liedtke, & Fischer, 2011; Borowski & Riese, 2010; Olszewski, 2010; Riese, 2009)
Vorbereitungs- dienst	<u>CK</u> Signifikanter Lernzuwachs <u>PCK und PK</u> Kein signifikanter Lernzuwachs (Chemie: Mutke & Tepner, 15.4.13)
Vorbereitungs- dienst und Schule	<u>CK und PCK</u> Lehrkräfte im Vorbereitungsdienst schneiden im Mittel signifikant schlechter ab als Lehrkräfte (Physik: Borowski et al., 2011; Borowski & Riese, 2010; Olszewski, 2010; Riese, 2009)
Studium und Schule	<u>CK</u> Lehramtsstudierende schneiden signifikant schlechter ab als Lehrkräfte <u>PCK</u> Kein signifikanter Unterschied zwischen Lehramtsstudierenden und Lehrkräften (Chemie: Dollny, 2011)

Der **Schulzweig**, an dem die Lehrkräfte unterrichten, bzw. auf den die universitäre Ausbildung der Lehramtsstudierenden ausgerichtet ist, scheint mit einem großen Unterschied in Bezug auf CK und PCK zusammenzuhängen. Ergebnisse aus mehreren Studien sind in Tabelle 3 dargestellt.

Der Unterricht an einem Gymnasium, bzw. das Studium eines gymnasialen Lehramts, hat im Allgemeinen einen positiven Einfluss auf CK und PCK, sofern CK nicht kontrolliert wird. Zumindest für das Studium lässt sich dies durch die Anzahl der Lerngelegenheiten und die geschätzte Time on Task erklären, da die gymnasialen Studiengänge auf die

Fachinhalte fokussieren (Brunner et al., 2006) und für die untersuchten Stichproben länger dauern.

Die Schulformunterschiede in CK und PCK sind so prägnant, dass sie für weitere Studien angenommen werden können. Träte der Unterschied nicht auf, stellte dies die Validität in Frage. Die Ausbildungsstufen müssen aufgrund dieser Ergebnisse differenziert nach der Schulform betrachtet werden. Selbst auf den Zusammenhang zwischen den Dimensionen kann die Schulform einen Einfluss haben (siehe Tabelle 1).

**Tabelle 3:** Schulformunterschiede in den Ausbildungsstufen

Dimension	Studium	Vorbereitungsdienst	Schule
<u>CK und PCK</u>	Fortgeschrittene Studierende des gymnasialen Lehramts schneiden im Mittel signifikant besser ab als die des nicht-gymnasialen Lehramts (Physik: Riese, 2009)	<u>CK</u> Lehrkräfte im Vorbereitungsdienst mit einem absolvierten gymnasialen Studiengang schneiden signifikant besser ab <u>PCK</u> Kein signifikanter Unterschied (Chemie: Mutke & Tepner, 2013)	Gymnasiallehrkräfte schneiden besser ab als Lehrkräfte anderer Schulformen (Chemie: Dollny, 2011; Mathematik: Krauss et al., 2008)
<u>PCK unter Kontrolle von CK</u>	Kein signifikanter Unterschied (Physik: Riese, 2009)		Lehrkräfte anderer Schulformen scheiden besser ab als Gymnasiallehrkräfte (Mathematik: Brunner et al., 2006) Kein signifikanter Unterschied (Chemie: Dollny, 2011)
<u>PK</u>	Kein signifikanter Unterschied (Physik: Riese, 2009)		

Ein Diplomstudiengang der Physik bietet viele Lerngelegenheiten bezüglich CK und wenige bezüglich PCK und der durchgängige Besuch eines Physikkurses in der eigenen Oberstufenzeit lässt einen positiven, bisher aber nicht empirisch bestätigten, Einfluss auf CK und PCK erwarten (Riese, 2009). Physiklehrkräfte, die ein Studium der Physik,

## Erwerb des Professionswissens

ob im Lehramts- oder Diplom/Masterstudium, absolvierten, sollten in CK- und PCK-Tests zur Physik besser abschneiden, als Physiklehrkräfte, die nicht Physik studiert haben, da sie Physik nicht fachfremd unterrichten; ein Studium der Mathematik sollte keinen direkten Einfluss auf CK und PCK haben (vgl. Riese, 2009, kein Einfluss der letzten Mathematiknote).

Lehrkräfte, die ein einzelnes Thema in der Physik selten unterrichten, sollten weniger Wissen in diesem Thema haben, da sie sich weniger häufig mit dem Thema beschäftigen, also weniger Lerngelegenheiten haben. Wird ein bestimmtes Thema, wie z. B. Mechanik, ausschließlich in der Mittel- und Oberstufe unterrichtet, sollte es einen positiven Einfluss haben, wenn in der Mittel- oder Oberstufe unterrichtet wird. Der positive Einfluss des Unterrichts in der Oberstufe sollte verstärkt werden, da diese Lehrkräfte am Gymnasium unterrichten und in der Regel ein Lehramtsstudium für die Sekundarstufe II bzw. ein Diplomstudium absolviert haben.

Der positive Einfluss des Oberstufenkurses, des Studiums der Physik, sowie des Unterrichts auf CK und PCK ist bisher nicht empirisch abgesichert und muss daher weiter untersucht werden.

**Fortbildungen** sind eine Möglichkeit, nach dem Ende der Ausbildung Wissen an Lehrkräfte zu vermitteln. Ob, bzw. wie der Besuch von Fortbildungen mit CK und PCK zusammenhängt, ist allerdings nicht geklärt. Brunner et al. (2006) finden einen tendenziellen, aber keinen statistisch signifikanten Zusammenhang. Allerdings schlüsseln sie nicht nach der Art der Fortbildung auf.

Für Chemielehrkräfte lässt sich kein eindeutiger Effekt von Fortbildungen auf CK zeigen. Regelmäßige chemiespezifische Fortbildungen und Fachgruppenarbeit scheinen das Abschneiden tendenziell zu verbessern; Die Gruppe, die sich mindestens alle sechs Monate individuell fortbildet, schneidet im PCK-Test besser ab als jene, die sich jeden Monat individuell fortbildet (Dollny, 2011). Trotzdem kann davon ausgegangen werden, dass häufige Fortbildungen, individuelle Weiterbildung und Fachgruppenarbeit tendenziell einen positiven Einfluss auf CK und PCK haben, da sie Lerngelegenheiten darstellen, auch wenn diese individuell nicht konsequent genutzt werden. Der Besuch von Fortbildungen zu allgemein pädagogischen Themen könnte einen Einfluss auf PCK haben, da PK und PCK vernetzt sind. Dieser müsste allerdings kleiner sein, als der Einfluss vorwiegend fachdidaktischer oder fachwissenschaftlicher Fortbildungen. Der Besuch von Fortbildungen zum unterrichteten Schulfach sollte tendenziell ebenfalls einen

positiven Einfluss auf CK haben, da auch sie Lerngelegenheiten bieten. Es bleibt darauf hinzuweisen, dass Fortbildungen oft keine messbare Wirkung haben (Lipowsky, 2004) und daher damit gerechnet werden muss, dass andere Variablen - wie der belegte Studiengang - eine weitaus größere Wirkung zeigen.

Die Jahre im Schuldienst, die Anzahl an Jahren, in denen bereits Physik unterrichtet wurde und das Alter der Physiklehrkräfte können als äquivalent betrachtet werden, da sie meist eng zusammenhängen (Brunner et al., 2006). Sie können als **Berufserfahrung** zusammengefasst werden, sofern die Angabe sich auf vollständig ausgebildete Lehrkräfte im Schuldienst bezieht. In der COACTIV-Studie schneiden Lehrkräfte mit längerer Berufserfahrung im Mittel schlechter in CK und PCK ab, dies wird allerdings auf die Gruppe der Lehrkräfte zurückgeführt, die in der ehemaligen Deutschen Demokratischen Republik ausgebildet wurden (Brunner et al., 2006). Es wird das Fazit gezogen, dass eine längere Berufserfahrung allgemein nicht zu mehr CK oder PCK führt (Brunner et al., 2006).

Im Bereich der Chemie schneiden Gymnasiallehrkräfte mit höherer Berufserfahrung schlechter in PCK ab, was auf eine inzwischen verbesserte Berufsausbildung zurückgeführt wird (Dollny, 2011). Physiklehrkräfte mit mehr als fünf Jahren und Biologielehrkräfte mit mehr als 19 Jahren Berufserfahrung schneiden in PCK schlechter ab, als jene mit weniger Berufserfahrung (Jüttner, 2013; Olszewski, 2010). Weder für Biologie-, noch für Physiklehrkräfte wurde ein Einfluss der Berufserfahrung auf CK gefunden (Borowski et al., 2011; Jüttner, 2013).

Die Berufserfahrung scheint auf CK keinen Einfluss zu haben, möglicherweise aber, aufgrund einer inzwischen verbesserten Ausbildung an den Universitäten, auf PCK.

Die Anzahl der Wochenstunden, in denen eine Lehrkraft Physik unterrichtet, müsste einen positiven Einfluss auf CK und PCK haben, da sich mehr Lerngelegenheiten bieten. Allerdings müssen diese nicht zwangsläufig genutzt werden (Brunner et al., 2006), so dass keine belastbare Aussage über die Time on Task gemacht werden kann. Dies könnte den Einfluss der Wochenstunden relativieren.

Schülerinnen und Schüler in Deutschland schneiden in den verschiedenen **Bundesländern** sehr unterschiedlich in den Naturwissenschaften ab (Prenzel & Deutsches PISA-Konsortium, 2008). Wird davon ausgegangen, dass das Professionswissen der Lehr-

## Erwerb des Professionswissens

kräfte mit dem Wissen der Schülerinnen und Schüler zusammenhängt, muss in Betracht gezogen werden, dass die Unterschiede zwischen Schülerinnen und Schülern der Bundesländer auch mit CK, PCK und PK der Lehrkräfte zusammenhängen. Aufgrund der PISA-Ergebnisse (Prenzel & Deutsches PISA-Konsortium, 2008), sollten Lehrkräfte der Naturwissenschaften in Bayern deutlich besser abschneiden als in Nordrhein-Westfalen, es sei denn, das Wissen der Lehrkräfte korreliert nicht mit dem Wissen der Schülerinnen und Schüler.

Im Rahmen der ProWiN-Studie wurde dieser Bundeslandunterschied bisher nur für Chemielehrer des Gymnasiums untersucht und für CK, nicht aber für PCK, gefunden. Für Chemielehrerinnen wurde der Unterschied aufgrund einer zu kleinen Stichprobe nicht untersucht (Dollny, 2011).

Es kann davon ausgegangen werden, dass in den Fächern Physik und Chemie **Lehrerinnen** im Mittelschlechter in CK abschneiden als Lehrer, das Geschlecht aber keinen Einfluss auf PCK hat (Dollny, 2011; Riese, 2009)). Dies entspräche dem Stereotyp, dass Frauen weniger Fachwissen in Physik hätten, und könnte genau durch dieses Stereotyp in Testergebnissen manifestiert werden („stereotype threat“) (Schmader, 2002).

Riese (2009) fasst in Regressionsanalysen den Einfluss **mehrerer Variablen** auf CK und PCK zusammen. Damit kann er 38% der Varianz des CK und 17% der Varianz des PCK von Physiklehramtsstudierenden aufklären. Allerdings ist keine Validierung der Regressionen veröffentlicht, so dass nicht von einer Generalisierbarkeit ausgegangen werden kann.

Wird nicht nach dem Studiengang unterschieden, haben die bereits in Physik belegten Semesterwochenstunden als Maß für die Lerngelegenheiten, die Abiturnote als Maß für die kognitiven Fähigkeiten und das Geschlecht einen signifikanten Einfluss auf CK, wobei mehr Semesterwochenstunden, eine bessere Abiturnote sowie das Attribut männlich mit besseren Ergebnissen zusammenhängen. Der Besuch eines Physikkurses in der eigenen Oberstufenzeit und die letzte schulische Physiknote haben keinen signifikanten Einfluss. Das wird dahingehend interpretiert, dass die Unterschiede im Vorwissen im Laufe des Studiums weniger bedeutsam werden (Riese, 2009). Einen signifikanten Einfluss auf PCK haben die bereits in Physik und Physikdidaktik belegten Semesterwochenstunden und die Abiturnote (Riese, 2009).

# *3. KAPITEL*

## *MODELL ZUR DEFINITION DES PROFESSIONSWISSENS*

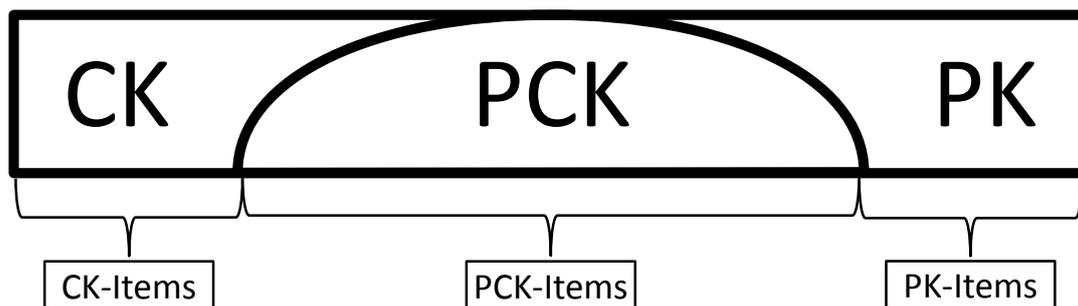
Wie im vorangegangenen Kapitel diskutiert, unterscheiden sich die Modelle des Professionswissens von verschiedenen Autorinnen und Autoren erheblich. In diesem Kapitel wird zuerst das in dieser Arbeit zugrunde gelegte Verständnis des Zusammenhangs von CK, PCK und PK vorgestellt (3.1). Im Verlauf des Kapitels werden weitere Aspekte erörtert, die der Operationalisierung von CK, PCK und PK dienen (3.2). Daraus wird das in dieser Arbeit zur Aufgabenkonstruktion genutzte Modell entwickelt, welches das Professionswissen für die vorgestellte Studie definiert (3.3).

Zusammenhang der Dimensionen des Professionswissens CK, PCK, PK: Annahme für diese Arbeit

### 3.1 ZUSAMMENHANG DER DIMENSIONEN DES PROFESSIONSWISSENS CK, PCK, PK: ANNAHME FÜR DIESE ARBEIT

Professionswissen wird in dieser Studie als theoretisch-formales Wissen unabhängig von anderen Aspekten der professionellen Handlungskompetenz untersucht (vgl. Baumert & Kunter, 2013). Wie im vorangegangenen Kapitel beschrieben, werden die drei Bereiche CK, PCK und PK im deutschsprachigen Raum als disjunkt, aber miteinander zusammenhängend angenommen. Dies wird graphisch in Abbildung 6 durch drei Bereiche verdeutlicht: PCK steht in der Mitte zwischen CK und PK und grenzt an beide. CK und PK stellen entgegengesetzte Enden dar. Lehrkräfte sollten in allen drei Bereichen über viel Wissen verfügen.

CK ist auch Teil des Professionswissens von Fachwissenschaftlerinnen und -wissenschaftlern (hier Physikerinnen und Physikern), PK ist Teil des Professionswissens von Pädagoginnen und Pädagogen, auch jenen, die nicht schulfachgebunden tätig sind und PCK ist das Wissen, das eine Fachlehrkraft besitzt.



**Abbildung 6:** Die Aufteilung des Professionswissens in drei Bereiche CK, PCK und PK und die Zuordnung der Items zu den Professionswissensbereichen

Wie die Unterscheidung von CK, PCK und PK im Testinstrument umgesetzt wird, wird im Folgenden erläutert. Der Schulkontext wird in Items des CK-Tests nicht thematisiert. Damit sollten sie auch von Physikerinnen und Physikern sowie Schülerinnen und Schüler der Oberstufe in Teilen zu lösen sein. Allerdings könnten Physiklehrkräfte besser abschneiden, da sie bei einigen Aufgaben eine größere Routine haben. Der PK-Test hat keinen naturwissenschaftlichen oder physikalischen Fachanteil (wie beispielsweise Wissen über Experimente), beschäftigt sich aber mit dem Kontext Schule. Er sollte von Lehrkräften aller Fächer ähnlich zu lösen sein.

Der PCK-Test gliedert sich wiederum in drei sich überschneidende Bereiche, PCK-Items mit einem vergleichsweise höheren Fachanteil, PCK-Items mit einem vergleichsweise höheren pädagogischen Anteil und PCK-Items, in denen sich der fachwissenschaftliche und der pädagogische Anteil die Waage halten. In allen PCK-Aufgaben wird der Schulkontext thematisiert. Die Aufgaben sollten von Physiklehrkräften besser gelöst werden als von Physikerinnen und Physikern sowie von vergleichbar qualifizierten Lehrkräften anderer Fächer.

### **3.2 ASPEKTE ZUR OPERATIONALISIERUNG DES CK, PCK UND PK**

In Kapitel 2 wurden CK, PCK und PK beschrieben und in Abschnitt 3.1 wurde die Annahme über den Zusammenhang der Dimensionen für diese Arbeit dargestellt. Dies ist nicht ausreichend um Professionswissen für ein Testinstrument zu operationalisieren, da die Begriffe CK, PCK und PK Interpretationsspielraum bieten. Daher werden im folgenden Abschnitt Wissensbereiche und -stufen definiert, die Themenabhängigkeit von CK und PCK diskutiert und themenunabhängige Facetten von PCK und PK dargestellt. Es wird jeweils darauf Bezug genommen, wie diese Aspekte in der Modellierung umgesetzt werden.

#### **3.2.1 WISSENSBEREICHE**

PCK gilt als Zusammenspiel dessen, was eine Lehrkraft weiß, wie sie handelt und den Gründen für diese Handlungen (Baxter & Lederman, 1999, S. 159). Es ist davon auszugehen, dass nicht nur Wissen über Fakten für das Unterrichten nötig ist. Daher soll im gesamten Testinstrument die Vielfalt verschiedener Wissensbereiche repräsentiert werden.

So unterscheidet Kunter zwischen deklarativem und prozeduralem Wissen (Kunter & Pohlmann, 2009), das ProwiN-Projekt zwischen den Wissensbereichen deklarativ, prozedural und konditional (vgl. Tepner et al., 2012, in Anlehnung an Paris, Lipson, & Wixson, 1983). Beschreibungen und Beispiele für die Bereiche PCK und CK finden sich in Tabelle 4. Beispielitems aus dem Bereich PK finden sich in Abschnitt 5.1.2.2. Die drei Wissensbereiche werden fächerübergreifend im ProwiN-Projekt verwendet: Alle drei Wissensbereiche können in allen drei Dimensionen des Professionswissens, CK, PCK und PK, in die systematische Itemkonstruktion einbezogen werden.

**Tabelle 4:** Kurzbeschreibung und Beispiele für die drei Wissensbereiche aus den Bereichen PCK und CK. Die Fragen entsprechen nicht den im Testinstrument verwendeten Items.

Bezeichnung	Beschreibung	Beispiele
Deklaratives Wissen	Wissen über Tatsachen, Inhalte	<b>Nennen</b> Sie... ... möglichst viele Schülervorstellungen zum Thema Kraft. ... das Hebelgesetz.
Prozedurales Wissen	Wissen über Handlungen	<b>Wie</b> kann der Versuch optimiert werden, ... ...damit die Schülerinnen und Schüler den Zusammenhang erkennen können? ...damit ein Wissenschaftler ein vertrauenswürdiges Ergebnis erhält?
Konditionales Wissen	Wissen über Begründungen und Beurteilungen	Warum ... ...ist es wichtig, im Unterricht an Schülervorstellungen anzuknüpfen? ... dauert der Sommer auf der Nordhalbkugel länger als auf der Südhalbkugel?

### WISSENSSTUFEN

Das Fachwissen wird auf verschiedenen Stufen unterschieden: Schulwissen, vertieftes Schulwissen, universitäres Wissen und fachliches Alltagswissen (Baumert et al., 2010; Riese, 2009). Es handelt sich bei Schulwissen um Wissen, das auch Schülerinnen und Schüler haben sollten, das also von Lehrenden unmittelbar zum Unterrichten benötigt wird. Der Begriff universitäres Wissen bezieht sich auf Wissen, das im fachlichen universitären Kontext gelehrt wird und sich nicht unmittelbar im Schulunterricht wiederfindet. Allerdings ist es möglich, dass es zu einem tieferen Verständnis des Faches führt, welches für guten Unterricht nötig sein könnte. Bei vertieftem Schulwissen handelt es sich nach Woitkowski, Riese, & Reinhold (2011) um eine explizite Verknüpfung von Schulwissen und universitärem Wissen.

Zu Beginn der Studie wurden Items zum vertieften Schulwissen wie folgt charakterisiert:

- Unbekannt: Keine Standardaufgaben für die Schule; Aufgaben, die an der Universität nicht explizit gelehrt werden
- Vollständiger Verzicht auf Oberstufen- und universitäre Mathematik (insbesondere Analysis)
- Benötigtes deklaratives Wissen geht im Wesentlichen nicht über die Sekundarstufe II hinaus
- Wissen muss flexibel eingesetzt werden
- Erste Lösungsansätze können in die Irre führen

Vertieftes Schulwissen kann nicht durch deklaratives Wissen abgedeckt werden.

Da ein Charakteristikum die Bekanntheit ist und diese nicht für aller Versuchspersonen gleich ist, können Schulwissen und vertieftes Schulwissen nicht eindeutig voneinander getrennt werden. Allerdings muss es für Unterricht nicht relevant sein, auf welcher Basis die Frage eines Schülers von der Lehrkraft korrekt beantwortet wird.

Der in dieser Studie entwickelte und genutzte CK-Test deckt Schulwissen und vertieftes Schulwissen ab und streift das universitäre Wissen. Ein Beispiel für ein CK-Item ist in Abbildung 7 zu sehen.

Keplersche Gesetze

Die Planeten bewegen sich auf Ellipsen, in deren Brennpunkt die Sonne steht.  
Der Fahrstrahl zwischen Sonne und Planeten überstreicht in gleicher Zeit gleiche Flächen.

Das Verhältnis aus den 3. Potenzen der großen Halbachsen und den Quadraten der Umlaufzeiten ist für alle Planeten konstant.

Jahreszeiten

Der Sommer beginnt mit dem längsten Tag im Jahr und endet mit der Tagundnachtgleiche.

Während auf der Nordhalbkugel der Erde Sommer ist, ist auf der Südhalbkugel Winter und umgekehrt.

Dauert der Sommer auf den beiden Halbkugeln der Erde unterschiedlich lang oder ist er auf beiden Halbkugeln gleich lang?

- Er dauert gleich lang.
- Er dauert unterschiedlich lang.

Bitte begründen Sie Ihre Antwort physikalisch.



—

**Abbildung 7:** Beispielitem aus dem CK-Test (CK 1180)

Kodierung:

Zur vollen Punktzahl tragen das Ankreuzen der korrekten Antwort (Der Sommer ist auf der Nordhalbkugel länger als auf der Südhalbkugel) und ein Verweis auf das zweite Keplersche Gesetz bei. Eine Argumentation über Daten und Kalender gilt nicht als physikalische Begründung.

### 3.2.2 THEMENABHÄNGIGKEIT

Allgemein wird angenommen, dass CK und PCK fachabhängig, wenn nicht sogar themenspezifisch sind (Park & Chen, 2012). Die Themenabhängigkeit springt im Falle des Fachwissens sofort ins Auge: Wer viel über Atomphysik weiß, muss nicht auch alles über Mechanik wissen. Trotzdem ist eine gewisse Vorhersagekraft anzunehmen, wenn von einem allgemeinen physikalischen Verständnis ausgegangen wird. Im Bereich des PCK ist die Lage schwieriger, allerdings wird aus Gründen der Testökonomie oft nur ein Thema innerhalb des Faches getestet oder auch qualitativ untersucht, obwohl es möglich ist, Fragen zum Fach allgemein zu konstruieren. Die neben den Fächern Biologie, Chemie und Physik international angeführte Disziplin „Naturwissenschaften“ (Veal & MaKinster, 1999), spielt in Deutschland aufgrund der Unterrichtskultur bisher keine Rolle.

Da Professionswissen als fachabhängig gilt, wurden PCK- und CK-Tests für Biologie, Chemie und Physik im ProwiN-Projekt getrennt entwickelt (Tepner et al., 2012).

Es gibt keine quantitativen Studien, die systematisch vergleichen, wie eng das PCK zu verschiedenen Themen innerhalb des Faches Physik zusammenhängt. Es konnte allerdings gezeigt werden, dass innerhalb derselben Stichprobe die PCK-Testergebnisse der Skalen von Olszewski und Riese stärker miteinander korrelieren, als mit den Ergebnissen der jeweiligen Fachtests, obwohl unterschiedliche Inhaltsbereiche (Mechanik und Elektrizitätslehre) getestet und verschiedene fachdidaktische Schwerpunkte gesetzt wurden (Borowski et al., 2010b).

Das Thema Mechanik ist sowohl in den Sekundarstufen als auch im Studium hochrelevant (Deutsche Physikalische Gesellschaft, 2004; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005). Es wird davon ausgegangen, dass bei Physiklehrkräften von einem gut entwickelten Verständnis der Mechanik in begrenztem Umfang auf ein gutes Verständnis in anderen Bereichen der Physik geschlossen werden kann.

Da Mechanik gleichermaßen wichtig für Lernende als auch für Lehrende zu sein scheint, liegt der Schwerpunkt der Physik-Tests im Bereich der Mechanik. Ein CK-Item und 2 PCK-Items liegen im Bereich der E-Lehre. Um die Beschränkung auf ein einzelnes Thema zu reduzieren, sind im Modell PCK-Items vorgesehen, die sich themenunabhängig auf den Physikunterricht beziehen. Ein Beispiel ist die Frage „Was spricht für die Verwendung von Einheiten bei Rechnungen im Physikunterricht?“. Diese kann ebenso mit dem Gedanken an eine Schulstunde zur Mechanik beantwortet werden, wie mit

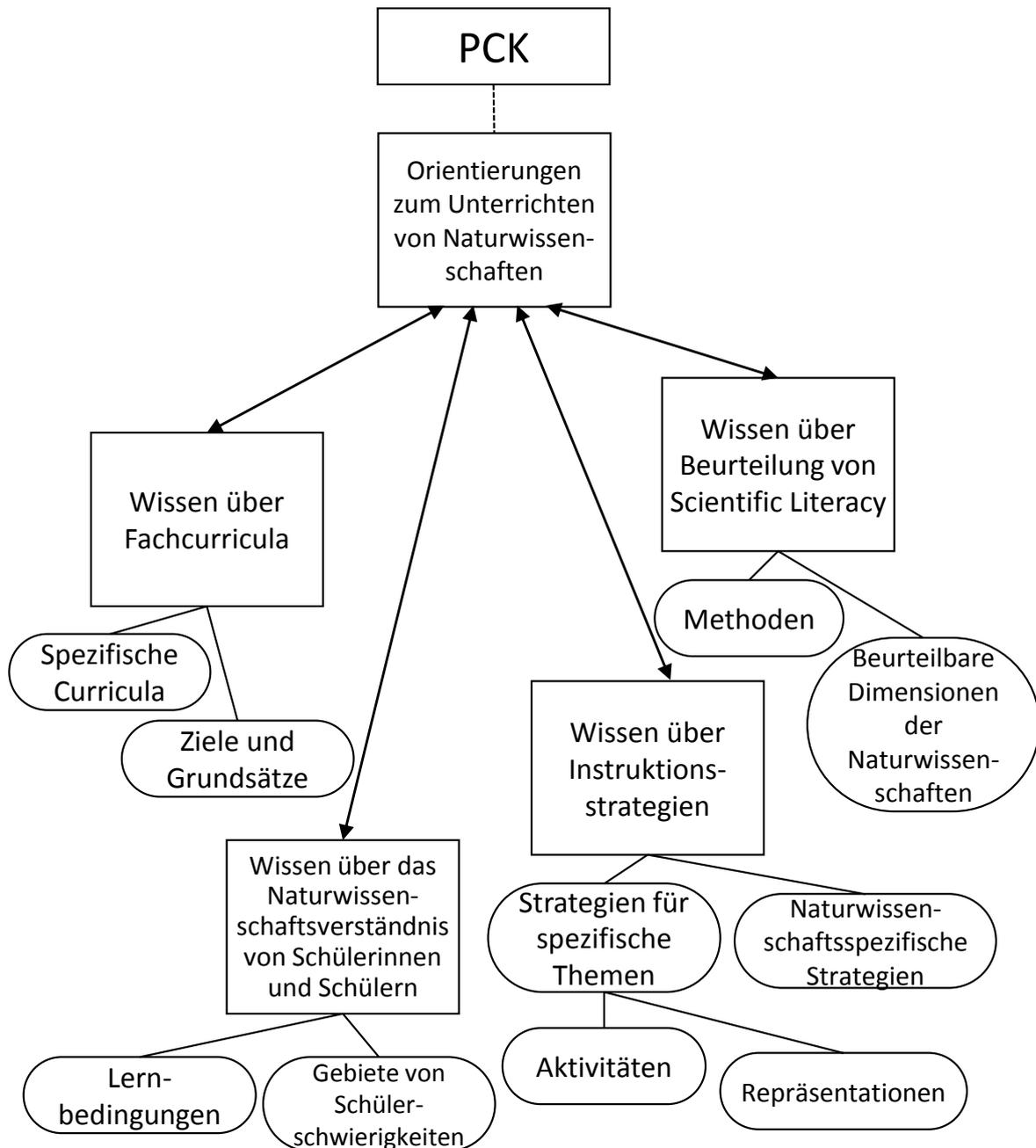
dem an eine Stunde zur E-Lehre, aber nicht aus der Perspektive einer Deutschstunde. Alle PK-Items sind fachunabhängig und sollten von Lehrkräften aller Fächer zu lösen sein.

### **3.2.3 FACETTEN DES PCK**

Es ist Konsens, dass sich das fachdidaktische Wissen aus Wissen über einzelne, nicht unbedingt trennbare, Facetten zusammensetzt (Kind, 2009; Park & Chen, 2012; Park et al., 2012; Riese, 2009; Shulman, 1986). Das international sehr verbreitete Magnusson-Modell bietet eine große Anzahl an Facetten, die als wichtig angesehen werden (siehe Abbildung 8) (z. B. Park & Suh, 2012; Park et al., 2012). Allerdings unterscheiden sich die tatsächlich in Messinstrumenten genutzten Facetten erheblich, da unterschiedliche Schwerpunkte gesetzt werden und aus testökonomischen Gründen die Anzahl der jeweils untersuchten Facetten begrenzt bleibt. Einen Überblick über Facetten, die an unterschiedlicher Stelle genutzt wurden, bietet Tabelle 5.

Sowohl bei Riese, als auch in der MT21-Studie ist eine Facette zu Reaktionen auf (kritische) Unterrichtssituationen zu finden, allerdings scheint sie eine Metaebene zu repräsentieren und sich dadurch von den anderen dargestellten Facetten abzugrenzen (Blömeke et al., 2008; Riese, 2009). Riese selbst definiert sie als Nutzung von PCK (Riese, 2009). Es ist allerdings nicht empirisch überprüft worden, ob das Wissen, das zur Beantwortung von Items zu Reaktionen auf kritische Unterrichtssituationen benötigt wird, auch tatsächlich in kritischen Unterrichtssituationen genutzt wird bzw. hilfreich ist.

Es wird diskutiert, wie das Wissen über die einzelnen Facetten miteinander verknüpft sein muss, um gut zu unterrichten. Allerdings gibt es hierzu noch keine quantitativen Studien. Einen Ansatz bieten Park & Chen (2012), die alle Facetten des Magnusson-Modells miteinander in Beziehung setzen (siehe Abbildung 9) und PCK als die Integration von zwei oder mehr Facetten definieren (S. 928). Sie kommen nach einer Studie mit vier Lehrkräften zu dem Schluss, dass die Facetten „Wissen über Schülerverständnis“ und „Wissen über Instruktionsstrategien“ und ihre Verbindung mit stabilerem und kohärenterem PCK einhergehen und daher für die Entwicklung von PCK besonders relevant sind. Es fällt auf, dass es sich hierbei um die beiden Facetten handelt, die in fast allen Operationalisierungen genannt werden (siehe Tabelle 5).



**Abbildung 8:** Das Magnusson-Modell - Facetten von PCK nach Magnusson et al. (1999, S. 99)

Pfeile zwischen zwei Boxen bedeuten, dass sich diese gegenseitig formen.

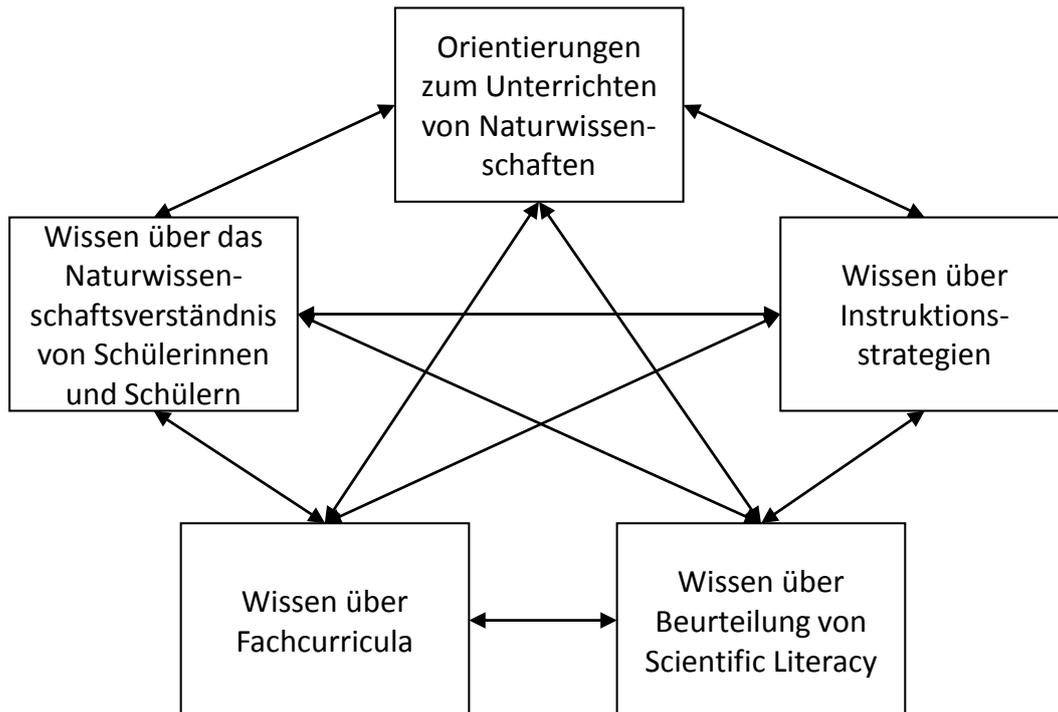
Die gestrichelte Linie bedeutet, dass PCK die Orientierungen beinhaltet.

Die durchgezogenen Linien ordnen den Facetten Teilfacetten zu.

**Tabelle 5:** Facetten von PCK nach Park & Oliver (2008), angepasst und ergänzt um weitere Studien

Studie	Wissen über...							
	Unterrichtsziele	Schülerverständnis	Curriculum	Instruktionsstrategien und Repräsentationen	Leistungsüberprüfung	Fachwissen	Kontext	Pädagogik
Shulman (1987)	-	0	-	0		-	-	-
Tamir (1988)		0	0	0	0	-		-
Grossman (1990)	0	0	0	0		-		
Marks (1990)		0		0		0		
Smith & Neale (1989)	0	0		0		-		
Cochran et al. (1993)		0				0	0	0
Geddis (1993)		0	0	0				
Fernandez-Balboa & Stiehl (1995)	0	0		0		0	0	
Magnusson et al. (1999)	0	0	0	0	0			
Hashweh (1987)	0	0	0	0	0	0	0	0
Loughran, Berry, & Mulhall (2006)	0	0		0		0	0	0
Gess-Newsome et al., In review		0		0		0	0	-
Rollnick, Bennett, Rhemtula, Dharsey, & Ndlovu (2008)		0	0	0	0	0	0	0
COACTIV (Kunter et al., 2011)		0		0	-	-	-	-
MT21 (Blömeke et al., 2008)		0	0	0		-		-
Riese (2009)	0	0	0	0		-		-
QuiP (Olszewski, 2010)		0	0	0		-		-
ProwiN (Tepner et al., 2012)		0		0		-		-

0 Diese Facette gehört explizit zu PCK  
 - Diese Facette gehört explizit nicht zu PCK



**Abbildung 9:** Der Zusammenhang zwischen den Facetten von PCK dargestellt als PCK-Map nach Park & Chen (2012, S. 930)

Wie bereits beschrieben, werden in der Literatur sehr unterschiedliche und breite Konstrukte des PCK vertreten, dabei bilden die Facetten, die PCK zugeordnet werden, eine wesentliche Differenz zwischen verschiedenen Ansätzen.

In dem Modell des PCK, das der Testkonstruktion dieser Studie zu Grunde liegt, werden drei Facetten abgedeckt, das **Wissen über Konzepte**, das **Wissen über Experimente** und das **Wissen über Schülervorstellungen**. Sowohl das Wissen über Konzepte als auch das Wissen über Experimente kann dem in Tabelle 5 genannten „Wissen über Instruktionsstrategien und Repräsentationen“ zugeordnet werden.

Items zur Facette „Konzepte“ liegen am Übergang von CK zum PCK. Die fachlichen Konzepte müssen von der Lehrkraft verstanden werden, dann aber auch aus fachdidaktischer Sicht so bearbeitet werden können, dass die Lernprozesse der Schülerinnen und Schülern unterstützt werden können. Sie lassen sich im Bereich der „Repräsentationen“ als Teil von „Wissen über Instruktionsstrategien für spezifische Themen“ in das Magnusson-Modell einordnen (siehe Abbildung 8).

## Aspekte zur Operationalisierung des CK, PCK und PK

Experimente nehmen im Physikunterricht an deutschen Schulen einen großen Teil der Zeit ein (Tesch, 2005) und sind damit elementarer Teil des Unterrichts. Zudem grenzen sie den naturwissenschaftlichen Unterricht von anderen Fächern, insbesondere von der Mathematik ab. Im Magnusson-Modell werden sie ebenfalls in den Bereich des „Wissens über Instruktionsstrategien für spezifische Themen“ als „Aktivitäten“ eingeordnet.

Das Wissen über Schülervorstellungen, die sich von Fach zu Fach bzw. themenabhängig unterscheiden, gilt als elementar für jede Lehrkraft (z. B. Wodzinski, 1996). Es ist im Magnusson-Modell als „Gebiete von Schülerschwierigkeiten“ dem „Wissen über das Naturwissenschaftsverständnis von Schülerinnen und Schülern“ zugeordnet. Wie im letzten Abschnitt beschrieben, ist das Wissen um Schülervorstellungen damit in nahezu jeder Konzeption des PCK als Wissen über Schülerverständnis zu finden (vgl. Tabelle 5).

### 3.2.4 FACETTEN DES PK

Im Bereich PK ist der Konsens über wichtige Facetten größer als im Bereich PCK. Schon Grossman (1990) zählt als generelles pädagogisches Wissen auf: Wissen und Vorstellungen (Beliefs) über Lernen und Lernende, Wissen über Grundprinzipien der Instruktion, Wissen und Fertigkeiten bezüglich Klassenführung und Wissen und Vorstellungen über den Zweck von Unterricht.

Kunter et al. (2011) führen „weitgehend konsensfähige Vorschläge“ an (S. 39), die allerdings nicht alle als handlungsrelevant gelten. Sie sind in Tabelle 6 dargestellt. In Abbildung 3 ist das Modell von Kunter et al. (2011) zusammenfassend dargestellt. Es zeigt die Trennung der Dimensionen CK, PCK und PK sowie die Facetten im Bereich PCK und PK.

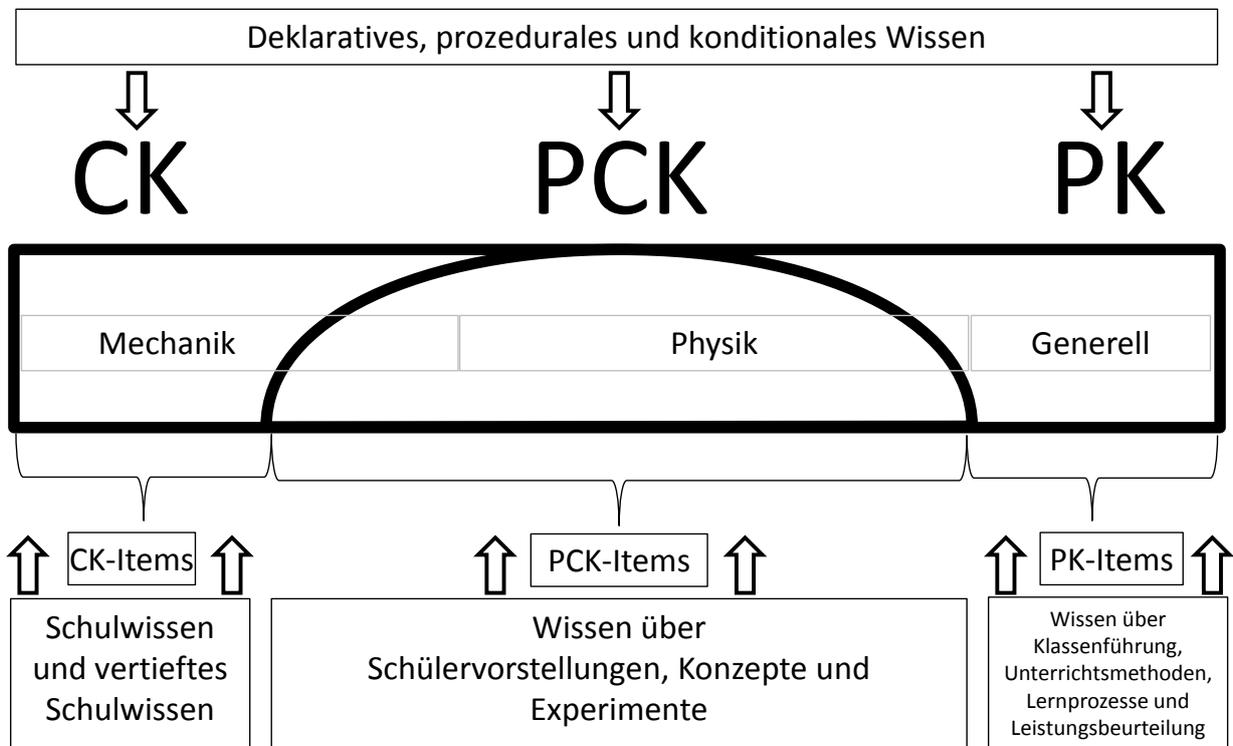
Im ProWiN-Projekt werden im Bereich PK die folgenden Facetten abgedeckt: Wissen über Klassenführung, Unterrichtsmethoden, Lernprozesse, Leistungsbeurteilung (Tepner et al., 2012; vgl. auch Kunter et al., 2011). Ein Beispielitem für den Bereich PK findet sich in Abschnitt 5.2.2.

**Tabelle 6:** Facetten generischen pädagogischen Wissens nach Kunter et al., 2011, S. 39

1. Konzeptuelles bildungswissenschaftliches Grundlagenwissen
- Erziehungsphilosophische, bildungstheoretische und historische Grundlagen von Schule und Unterricht
- Theorie der Institution
- Psychologie der menschlichen Entwicklung, des Lernens und der Motivation
2. Allgemeindidaktisches Konzeptions- und Planungswissen
- Metatheoretische Modelle der Unterrichtsplanung
- Fachübergreifende Prinzipien der Unterrichtsplanung
- Unterrichtsmethoden im weiten Sinne
3. Wissen über Unterrichtsführung und Orchestrierung von Lerngelegenheiten
- Inszenierungsmuster von Unterricht
- Variation von Sozialformen und Methoden
- Effektive Klassenführung (classroom management), Sicherung einer konstruktiv-unterstützenden Lernumgebung
4. Wissen über fachübergreifende Prinzipien des Diagnostizierens, Prüfens und Bewertens
- Lernen und Leisten: Grundlagen der Diagnostik
- Prozessdiagnostik
- Rückmeldungen
- Summatives Prüfen und Bewerten
5. Methodische Grundlagen empirischer Sozialforschung

### 3.3 GESAMTES MODELL

Die Tests zum Professionswissen von Physiklehrkräften werden nach dem in Abbildung 10 gezeigten Modell entwickelt. Es beinhaltet die Dimensionen CK, PCK und PK und die zugehörigen Items. In allen drei Dimensionen werden die drei Wissensbereiche deklaratives, prozedurales und konditionales Wissen abgedeckt. Die Tests sind fach- und themenabhängig mit Schwerpunkt auf der Mechanik. Die im PCK-Test betrachteten Facetten sind Wissen über Schülervorstellungen, Konzepte und Experimente. Die im PK-Test betrachteten Facetten sind Wissen über Klassenführung, Unterrichtsmethoden, Lernprozesse und Leistungsbeurteilung. Die CK-Items liegen im Bereich des Schulwissens und vertieften Schulwissens.



**Abbildung 10:** Das Modell zur Testkonstruktion

# *4. KAPITEL*

## *ZIELE DER UNTERSUCHUNG UND FORSCHUNGSFRAGEN*

Nachdem die Relevanz des Professionswissens erläutert, der Forschungsbedarf skizziert und das in dieser Arbeit genutzte Modell dargestellt wurde, werden in diesem Abschnitt die Ziele der Untersuchung formuliert (4.1), Forschungsfragen gestellt und Hypothesen dazu aufgestellt (4.2).

### **4.1 ZIELE**

Das erste Ziel der Arbeit ist die Modellierung des Professionswissens, die Umsetzung des Modells in ein Testinstrument und die empirische Absicherung des Modells. Das zweite Ziel ist die Untersuchung des Zusammenhangs zwischen dem Professionswissen von Physiklehrkräften und ihrem demographischen Hintergrund, sowie der Vergleich des Professionswissens von Lehramtsstudierenden, Lehrkräften im Vorbereitungsdienst und Lehrkräften.

Die Inhaltsvalidität wird bei der Testentwicklung gewährleistet. Die Auswertobjektivität wird für jedes Items sichergestellt. Zunächst muss untersucht werden, ob die angenommene Aufteilung in CK, PCK und PK empirisch nachgewiesen werden kann. Ist das der Fall, können die drei Dimensionen für sich untersucht werden. Nachdem gezeigt wird, dass die Tests reliabel und valide sind, werden die fachspezifischen Bereiche des Professionswissens näher beleuchtet. Es wird untersucht, welche demographischen Variablen mit CK und PCK von Physiklehrkräften zusammenhängen. Des Weiteren wird analysiert, wie sich CK und PCK und ihr Zusammenhang zwischen Lehramtsstudierenden, Lehrkräften im Vorbereitungsdienst und Lehrkräften unterscheiden.

### **4.2 FORSCHUNGSFRAGEN UND HYPOTHESEN**

Um die oben genannten Ziele zu erreichen, werden die folgenden Forschungsfragen gestellt und die anschließenden Hypothesen getestet.

FF1: Lassen sich die theoretisch angenommenen drei Dimensionen des Professionswissens, CK, PCK und PK, empirisch voneinander trennen?

H1: In einer mehrdimensionalen Rasch-Analyse erreicht (1) ein Modell mit den drei Dimensionen CK, PCK und PK eine signifikant bessere Passung (Final Deviance), als (2) ein eindimensionales Modell, oder als die zweidimensionalen Modelle mit (3) den Dimensionen fachspezifisches Professionswissen (CK und PCK) und pädagogisches Wissen (PK), bzw. (4) den Dimensionen Wissen über das Fach (CK) und Wissen über Unterricht (PCK und PK).

Es wird untersucht, ob die einzelnen Dimensionen reliabel sind und Kriteriums- und Konstruktvalidität gewährleistet sind.

FF2: Kann der erwartete Einfluss der demographischen Variablen zur Ausbildung, zur Lehrerfahrung und aktuellen Situation sowie Alter und Geschlecht der Physiklehrkräfte auf die fachspezifischen Bereiche des Professionswissens, CK und PCK, bestätigt werden? Wie groß ist der Einfluss von CK und PK auf PCK?

H2: Die Korrelationen zwischen den unabhängigen Variablen und CK bzw. PCK haben die in Tabelle 7 dargestellten Richtungen.

**Tabelle 7:** H2 - Erwartete Korrelationen der unabhängigen Variablen mit CK und PCK

Variable	Korrelation mit CK	Korrelation mit PCK
CK	-	positiv
PK	kein Einfluss	positiv
Durchgängiger Besuch eines Physikkurses in der eigenen Oberstufenzeit	positiv	positiv
Belegter Studiengang gymnasiales Lehramt	positiv	positiv
Belegter Studiengang nicht-gymnasiales Lehramt	negativ	negativ
Belegter Studiengang Diplom	positiv	negativ
Studium der Physik/kein fachfremder Unterricht	positiv	positiv
Studium der Mathematik	kein Einfluss	kein Einfluss
Jahre im Schuldienst	kein Einfluss	negativ
Anzahl an Jahren, in denen bereits Physik unterrichtet wurde	kein Einfluss	negativ
Alter	kein Einfluss	negativ
Wochenstunden Physik, aktuell	positiv	positiv
Unterrichtet in der Unterstufe	kein Einfluss	kein Einfluss
Unterrichtet in der Mittelstufe	positiv	positiv
Unterrichtet in der Oberstufe	positiv	positiv
Unterrichtet am Gymnasium	positiv	positiv
Häufigkeit Fortbildungsbesuch Physik	positiv	positiv
Häufigkeit individueller Weiterbildung	positiv	positiv
Häufigkeit Fachgruppenarbeit	positiv	positiv
Häufigkeit pädagogischer Fortbildungen	kein Einfluss	positiv
Bundesland:		
Bayern	positiv	positiv
NRW	negativ	negativ
Geschlecht:		
Lehrer	positiv	kein Einfluss
Lehrerin	negativ	kein Einfluss

## Ziele der Untersuchung und Forschungsfragen

Der Einfluss der Variablen zur aktuellen Situation der Physiklehrkräfte soll auf CK und PCK in Kombination untersucht werden. Dazu wird eine lineare Regression mit den folgenden unabhängigen Variablen durchgeführt:

- Unterrichtet am Gymnasium (stellvertretend für die Schulform)
- fachfremder Unterricht (kein Studium der Physik)
- Häufigkeit Fortbildungsbesuch Physik
- Häufigkeit pädagogischer Fortbildungen
- Häufigkeit individueller Weiterbildung
- Häufigkeit Fachgruppenarbeit
- Anzahl der Wochenstunden, in denen Physik unterrichtet wird
- Alter (stellvertretend für die Berufserfahrung) und Geschlecht
- CK und PK für die Varianzaufklärung von PCK.

Die unabhängigen Variablen in einer Regression sollen nicht hoch miteinander korrelieren, daher wurde jeweils eine Variable stellvertretend für die Schulform und die Berufserfahrung ausgewählt.

FF3: Wie unterscheidet sich das fachspezifische Professionswissen zwischen Lehramtsstudierenden der Physik, Lehrkräften der Physik im Vorbereitungsdienst und Lehrkräften der Physik?

H3: CK und PCK wachsen im Laufe der Ausbildung in beiden Schulformen. Lehrkräfte schneiden besser ab als Lehrkräfte im Vorbereitungsdienst. Diese wiederum schneiden besser ab als Lehramtsstudierende.

FF4: Wie hängen CK und PCK bei Lehramtsstudierenden der Physik, Lehrkräften der Physik im Vorbereitungsdienst und Lehrkräften der Physik zusammen?

H4: Die Korrelationen unterscheiden sich nicht zwischen den Gruppen.

# *5. KAPITEL*

## *ANLAGE DER*

### *UNTERSUCHUNG*

Das folgende Kapitel beschreibt die Entwicklung und Charakterisierung der Testinstrumente (5.1), die Testdurchführung in der ersten und zweiten Studie (5.2) und die befragte Stichprobe (5.3).

### **5.1 BESCHREIBUNG DER TESTINSTRUMENTE**

In dem ersten Abschnitt werden der CK- und der PCK-Test vorgestellt (Eigenentwicklung; Kirschner, Borowski, & Fischer, 2011a; Kirschner, Borowski, & Fischer, 2011b). Der Abschnitt beschreibt die Testentwicklung, das gewählte Testformat und weitere Charakteristika der Tests wie beispielsweise die Anzahl der Items. Der zweite Abschnitt beschreibt die Entwicklung des PK-Tests durch Hubertina Thillmann und charakterisiert besonders die Itemformate näher.

#### **5.1.1 CK-TEST UND PCK-TEST**

Ein Ziel der Studie ist es, ein modellbasiertes, reliables und valides schriftliches Testinstrument zu entwickeln, mit dem CK und PCK der Physiklehrkräfte gemessen werden können.

##### **5.1.1.1 TESTENTWICKLUNG**

Im ersten Schritt der Testentwicklung wurden Items nach dem in Kapitel 3 beschriebenen Modell (Abbildung 10) konstruiert, wobei die zwei Bereiche CK und PCK getrennt entwickelt wurden. Es war nicht beabsichtigt, die einzelnen Facetten, Wissensstufen oder Wissensbereiche eindeutig voneinander trennbar zu konstruieren.

Alle Items sowie das Auswertungsmanual wurden zu mehreren Zeitpunkten mit erfahrenen Fachdidaktikerinnen und Fachdidaktikern sowie Lehrkräften diskutiert und weiterentwickelt, um die Inhaltsvalidität zu gewährleisten. Um eine Anschlussfähigkeit an bereits durchgeführte Studien zu dem Professionswissen deutscher Physiklehrkräfte zu erreichen, wurden einzelne Items aus der Studie von Olszewski (2010) vollständig übernommen. Einige Aufgabenstellungen und Kodierungsgrundlagen sind angelehnt an Riese (2009). Als weitere Grundlagen zur Entwicklung von Testaufgaben wurden auch Schulbücher und der „Force Concept Inventory“ (FCI; Hestenes, Wells, & Swackhamer, 1992) genutzt.

Im Anschluss an die Itemkonstruktion wurden die Items in mehreren Iterationen mit Lehramtsstudierenden der Physik (N=12), Physiklehrkräften im Vorbereitungsdienst (N=29) und Physiklehrkräften der Hauptschule (N=2) und des Gymnasiums (N=3) in Präpilotierungen erprobt. In allen Schritten wurden Kommentare der Versuchspersonen abgefragt, so dass Verständnisprobleme sowie Ungenauigkeiten in den Aufgabenstellungen reduziert werden konnten. Nach der Präpilotierung wurde die Anzahl der

Items in den Testheften an die in der ersten Studie (Pilotstudie) zur Verfügung stehende Zeit angepasst.

Die erste Studie diente vorrangig der Itemauswahl und ihrer Validierung. Ihre Durchführung sowie die detailliertere Beschreibung der 186 Versuchspersonen sind in den Abschnitten 5.2 und 5.3 zu finden.

Die Items wurden nach Kriterien der Rasch-Analyse und der klassischen Testtheorie ausgewählt (Kirschner, Wlotzka, Borowski, & Fischer, 2011). In der Rasch-Analyse mit dem Programm Winsteps (Linacre, 2010) sollte die mean-square-fit-Statistik (MnSq) im Infit für jedes Item nahe eins liegen und die Items mussten in der z-standardisierten Statistik (ZStd) einen Wert kleiner als 1.7 erreichen (vgl. Bond & Fox, 2007). Mit Hilfe einer Item-Map (Bond & Fox, 2007) wurden Aufgabenschwierigkeiten und Personenfähigkeiten analysiert und die zu den Personenfähigkeiten passenden Items ausgewählt. Konkret wurde von mehreren Items gleicher Schwierigkeit nur eines übernommen und im PCK-Test wurden eher schwierige Items verworfen. Aus dem Bereich der klassischen Testtheorie wurden die Trennschärfe und die Interraterübereinstimmung für die Auswahl herangezogen. Die Trennschärfe musste über 0.2 liegen. Goodman und Kruskal Gamma wurde als Maß für die Interraterübereinstimmung bei ordinalskalierten Skalen genutzt, der Cut-Off-Wert lag bei 0.8 (vgl. Wirtz & Caspar, 2002). Dies sollte psychometrisch zufriedenstellende Items gewährleisten. Die in der zweiten Studie angelegten Gütekriterien werden in Kapitel 6 näher erläutert.

Die ausgewählten Items sollten das Modell des Professionswissens (Kapitel 3) möglichst breit abdecken, allerdings hatten die obengenannten psychometrischen Qualitätskriterien Vorrang. Standen unterschiedliche Items zur Auswahl, von denen nur eines in die zweite Studie übernommen werden sollte, wurde das, welches eine Lücke im Modell schloss, ausgewählt, so dass alle Facetten und Wissensbereiche abgebildet werden konnten.

Zwei Items, die aus der ersten in die zweite Studie übernommen wurden, erreichten das Kriterium für die Passung an das Rasch-Modell nicht, eines verfehlte das Kriterium der Interraterübereinstimmung. Diese drei Items wurden deutlich überarbeitet. Einige Items mussten durch kleine Änderungen angepasst werden.

## Beschreibung der Testinstrumente

Die Items der zweiten Studie sind in Kirschner et al., 2011a und Kirschner et al., 2011b zu finden und bei der Autorin und den Autoren erhältlich. Zwei Beispielitems des PCK-Tests sind in Abbildung 11 zu finden. Beispiel 1 ist keinem einzelnen Thema innerhalb der Physik zugeordnet, Beispiel 2 dem Thema Geschwindigkeit. Dies entspricht dem Schwerpunkt des Modells, der Mechanik.

### 5.1.1.2 TESTFORMAT

Die Items im PCK-Test sind überwiegend als halboffene Fragen (Bortz & Döring, 2006) gestaltet, da in offenen Antwortformaten mehrere Möglichkeiten als korrekt bewertet werden können. Damit wird der Tatsache Rechnung getragen, dass es in der Fachdidaktik zu vielen Fragen bisher keine belastbaren empirischen Befunde und keine Einigkeit über eindeutig richtige Antworten gibt. Außerdem kann ein Aha-Effekt bei vorgegebenen Antwortmöglichkeiten vermieden werden (Baxter & Lederman, 1999; Schmelzing, 2010). In den Beispielitems (Abbildung 11) wird deutlich, dass korrekte Antworten aus verschiedenen Perspektiven gegeben werden können und mehrere verschiedene Aspekte genannt werden müssen, um die volle Punktzahl zu erreichen.

Im CK-Test wurden auch geschlossene Items (Bortz & Döring, 2006) verwendet, um die Testökonomie zu erhöhen. Für einige CK-Items wurde eine Kombination der Formate gewählt, es muss eine korrekte Antwort ausgesucht und diese dann begründet werden. Ein Beispiel ist in Abschnitt 3.2.2 zu finden (Abbildung 7).

**1. Was spricht für die Verwendung von Einheiten bei Rechnungen im Physikunterricht?**

**Aufgabe:**

*Bitte finden Sie möglichst viele Begründungen.*



---

---

**2. Sie führen das Thema Geschwindigkeit ein.**

**Ein Schüler erzählt: „Auf der Rückfahrt aus dem Urlaub sind wir eine Geschwindigkeit von 130 Stundenkilometern gefahren.“**

**Aufgabe:**

*Soll der Begriff „Stundenkilometer“ im Unterricht verwendet werden?*

- Ja
- Nein
- Kommt darauf an

*Bitte begründen Sie Ihre Antwort.*



---

---

**Abbildung 11:** Beispielitems aus dem PCK-Test (PCK S230, PCK S180)

Kodierung:

1. Korrekte Antworten beziehen sich auf die Wissenschaftspropädeutik, die Vermeidung und das Finden von Fehlern und das vertiefte Verständnis von Zusammenhängen. Inkorrekte Antworten beziehen sich beispielsweise auf die reine Übung.
2. Korrekte Antworten beziehen sich auf die fachliche Richtigkeit, den Lernprozess, das Handeln von Lehrkräften oder Schülerinnen und Schülern oder die Verwendung von Alltags- und Fachsprache.

### 5.1.1.3 WEITERE CHARAKTERISTIKA DER TESTS

Die **Anzahl** der Items der ersten und zweiten Studie ist in Tabelle 8 dargestellt. Weder die Wissensbereiche, noch die im PCK-Test betrachteten Facetten oder die im CK-Test betrachteten Wissensstufen werden aufgelöst.

6 PCK-Items und 11 CK-Items konnten unverändert aus der ersten in die zweite Studie übernommen werden. In einer Rasch-Analyse konnten die Antworten der Versuchspersonen der ersten Studie zu diesen unveränderten Items gemeinsam mit den Daten der zweiten Studie verarbeitet werden. Zu diesem Zweck wurden die Antworten aus der ersten Studie nach Änderungen im Manual erneut von zwei Raterinnen kodiert.

**Tabelle 8:** Anzahl der Items

	Itemanzahl		
	Erste Studie	Zweite Studie	Davon unverändert
CK	22	15	11
PCK	20	17	6

Der gesamte Test kann als **Speedtest** interpretiert werden, da eine feste Bearbeitungszeit vorgegeben wurde und nicht alle Versuchspersonen ihn in dieser Zeit vollständig bearbeiten konnten. Dies unterscheidet ihn von einem Powertest, in dem allen Versuchspersonen ausreichend Zeit für die Bearbeitung aller Items gegeben wird (genutzt z. B. in der COACTIV-Studie, Krauss et al., 2008). Der PCK-Test enthält Items, bei denen der Zeitfaktor explizit eine Rolle spielt: Für die ersten drei Items stand jeweils genau eine Minute zur Bearbeitung zur Verfügung, um die zeitlichen Rahmenbedingungen für jedes Item konstant zu halten. In der Auswertung wurden alle Items gleich behandelt.

Im Zuge des ProwiN-Projektes wurden Aufgaben gesucht, die von Lehrkräften der Biologie, Chemie und Physik lösbar sein sollten. Es zeigte sich, dass allein Aufgaben zum Experimentieren fächerübergreifend konsensfähig waren. So wurden in der zweiten Studie sechs **Items zum naturwissenschaftlichen Arbeiten** in biologischen, physikalischen und chemischen Kontexten für die Lehrkräfte aller drei Fächer eingesetzt, die bisher nicht analysiert wurden. Das Item aus der Physik gehört zu den hier analysierten PCK-Items (PCK 0040).

### **5.1.2 PK-TEST**

Neben dem CK- und PCK-Test wird ein PK-Test eingesetzt. Dieser Abschnitt beschreibt die Testentwicklung, das Testformat und weitere Charakteristika des PK-Tests.

#### **5.1.2.1 TESTENTWICKLUNG**

Der PK-Test wurde am Lehrstuhl der Lehr-Lernforschung an der Ruhr-Universität-Bochum von Hubertina Thillmann und Joachim Wirth entwickelt. In einem ersten Schritt wurden Lehrkräfte interviewt und Situationsskizzen gesammelt (Fischer et al., 2009). In einem weiteren Schritt wurden Testaufgaben aus der COACTIV-R-Studie (Voss et al., 2011) unter Beibehaltung der Inhalte (Facetten) von Hubertina Tillmann adaptiert und ergänzt (Fischer et al., 2009). Die Wissensbereiche können in einen deklarativen sowie einen prozedural-konditionalen Teil aufgelöst werden. Die Facetten werden nicht aufgelöst.

Die Pilotstudie mit 72 Versuchspersonen, davon 21 Lehrkräfte, 20 Personen aus der universitären Lehrerausbildung und 31 Studierende (weder im Fach Erziehungswissenschaften noch in der Lehramtsausbildung), zeigte zufriedenstellende Reliabilitäten (Fischer et al., 2010).

Die Pilotierung der PK-Items wurde unabhängig von der Pilotierung der CK- und PCK-Items durchgeführt.

#### **5.1.2.2 TESTFORMAT**

In dem CK- und dem PCK-Test wurde das Aufgabenformat unabhängig von Facetten, Wissensstufen und Wissensbereichen konstruiert. Im PK-Test wurden die Items zum deklarativen Wissen grundlegend anders konstruiert als die Items zum prozeduralen und konditionalen Wissen: Sie werden deshalb getrennt voneinander analysiert.

Das deklarative Wissen wurde mit Multiple-Choice-Items im Single Select Verfahren abgefragt (Fischer et al., 2010). Die Antworten wurden dichotomisiert und jeweils mit null Punkten oder einem Punkt bewertet. Ein Beispiel ist in Abbildung 12 dargestellt.

Im Test zum prozedural-konditionalen Wissen wurden Situationen dargestellt und Handlungsalternativen bewertet. Als Grundlage für das gewählte Antwortformat wurde Schlagmüller & Schneider (2007) genutzt. Dabei sollten einige Handlungsalternativen von den Lehrkräften besser bewertet werden als andere. Die Bewertung der Lehrerurteile wurde aus einem Expertenrating mit acht Personen generiert (Thillmann,

## Beschreibung der Testinstrumente

personal communication, 2011). Die Paarvergleiche zwischen den Handlungsalternativen wurden mit null Punkten, einem halben Punkt und einem Punkt bewertet. Ein Punkt wurde für die richtige Rangfolge vergeben, ein halber Punkt bei gleichem Rang und kein Punkt bei falscher Rangfolge. Im Beispiel (Abbildung 12) bedeutet das: Wird a) besser als b) bewertet, werden keine Punkte vergeben. Wird a) schlechter als b) bewertet, wird ein Punkt vergeben.

Werden a) und b) gleich bewertet, wird ein halber Punkt vergeben. Selbstverständlich werden nicht nur a) und b) verglichen, sondern alle möglichen Paarvergleiche durchgeführt. Jeder Paarvergleich wird in der folgenden Rasch-Analyse als ein Item betrachtet.

### **5.1.2.3 WEITERE CHARAKTERISTIKA DER TESTS**

Der PK-Test der Hauptstudie (zweite Studie) besteht aus 9 Aufgaben zum deklarativen Wissen mit je 4 bis 6 Items und 14 Items zum prozedural-konditionalen Wissen mit 125 ausgewerteten Paarvergleichen.

Es gibt zwei Itemtypen, das deklarative pädagogische Wissen  $PK_D$  und das konditional-prozedurale pädagogische Wissen  $PK_{KP}$ . Diese beiden PK-Bereiche werden getrennt ausgewertet.

1. Welche Maßnahmen sind geeignet, das selbstregulierte Lernen zu fördern?

		Sehr	eher	eher nicht	gar nicht
a.	Lerntagebuch führen lassen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b.	Concept-Map erstellen lassen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c.	Gespräch mit den Eltern führen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d.	Gespräch mit dem Schüler führen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e.	Lernprozess gut vorstrukturieren	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Stellen Sie sich vor, eine Lehrperson lässt in einer Klasse die erste Klausur schreiben. Bewerten Sie die Vorgehensweisen danach, wie geeignet sie sind, um die Akzeptanz der Notengebung zu erhöhen.

		Noten					
		1	2	3	4	5	6
a.	Sie baut eine vertrauensvolle Beziehung zu den SuS auf.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b.	Sie macht die Bewertungskriterien vorher explizit und lässt deren Umsetzung einüben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c.	Sie demonstriert immer wieder Beispiele für sehr gute Ausarbeitungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d.	Sie demonstriert ab und zu Beispiele für schlechte Ausarbeitungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e.	Sie macht die Bewertungskriterien vorher explizit und lässt diese auswendig lernen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Abbildung 12:** Beispiel für 1. den deklarativen (PKD 4) und 2. den konditional-prozeduralen Teil des PK-Tests (PK 18)

Kodierung:

1. Korrekte Antworten: a, d, e sehr oder eher; b, c eher nicht oder gar nicht

2. Rangfolge: b, a, c, d, e

## 5.2 TESTDURCHFÜHRUNG

Alle Testungen wurden unter Aufsicht durchgeführt. Hilfsmittel wie Taschenrechner oder Formelsammlungen waren nicht zugelassen. Bevor die Versuchspersonen die Wissenstests ausfüllten, wurden sie gebeten, Angaben zu Ihrer Person zu machen.

In Tabelle 9 sind die Bearbeitungszeiten für die einzelnen Tests dargestellt.

**Tabelle 9:** Bearbeitungszeiten

	Bearbeitungszeit in Minuten	
	Erste Studie	Zweite Studie
CK	60	40
PCK	63	50
PK		30

Die Teilnahme aller Versuchspersonen war freiwillig. Daher muss von einer Positivauslese ausgegangen werden. Alle Lehrkräfte der ersten Studie und alle Versuchspersonen der zweiten Studie, außer den Lehrkräften in Bayern und den Lehrkräften im Vorbereitungsdienst, die im Seminar getestet wurden, erhielten eine Aufwandsentschädigung. Dies könnte die Teilnahmemotivation beeinflusst haben.

Die Versuchspersonen der ersten Studie füllten den CK- und PCK-Test aus, während die Versuchspersonen der zweiten Studie den CK-, PCK- und PK-Test bearbeiteten. Die Teilstichprobe der Studierenden hat auch in der zweiten Studie nur den CK- und PCK-Test bearbeitet.

Die erste Studie wurde im Jahr 2010 in Berlin, Bremen, Hessen, Niedersachsen, Nordrhein-Westfalen und Thüringen durchgeführt, die zweite Studie im Jahr 2011 in Nordrhein-Westfalen und Bayern. Ein Großteil der Testungen der Physiklehrkräfte (Haupterhebung NRW N=80, Haupterhebung Bayern N=64, Nacherhebung NRW N=19) in der zweiten Studie wurde vom DPC (Data Processing and Research Center, Hamburg) organisiert und durchgeführt. Ein Teil der Daten wurde im Rahmen von Examensarbeiten an der Universität Duisburg-Essen erhoben.

### 5.3 STICHPROBENBESCHREIBUNG

Insgesamt wurden 518 Versuchspersonen befragt, davon 279 Physiklehrkräfte. Das Professionswissen von Physiklehrkräften steht im Fokus der Untersuchung, daher sind diese stark repräsentiert. Um die Entwicklung des Professionswissens erhellen zu können, wurden ergänzend angehende Physiklehrkräfte untersucht, die noch studieren bzw. sich im Vorbereitungsdienst befinden. Zu Validierungszwecken wurden Lehrkräfte anderer Naturwissenschaften und der Mathematik, die an Gymnasien unterrichten, sowie Diplom-Physikerinnen und -Physiker und Studierende der Physik in die Stichprobe aufgenommen. Das Wissen der Physikstudierenden wird hier nicht analysiert. Die Teilstichproben sind nicht repräsentativ.

In Tabelle 10 ist die Anzahl der Versuchspersonen nach ihrem **Ausbildungsstand** und ihrem **Studienabschluss** aufgegliedert. Für alle Gruppen sind **Alter** und **Geschlechtsverteilung** aufgeführt. Wie zu erwarten, sind die Studierenden jünger als die Lehrkräfte im Vorbereitungsdienst und diese wiederum jünger als die Lehrkräfte. Die Mehrheit der Physiklehrkräfte ist männlich. Dies entspricht der Lage in Deutschland. In NRW sind z. B. nur 26% der Physiklehrkräfte an Gymnasien und Gesamtschulen sowie 37% der Physiklehrkräfte an Haupt- und Realschulen weiblich (IT.NRW: Information und Technik Nordrhein-Westfalen - Geschäftsbereich Statistik, 2013). Die Mehrheit der angehenden Physiklehrkräfte, also der Physiklehrkräfte im Vorbereitungsdienst und der Lehramtsstudierenden der Physik, in dieser Stichprobe ist ebenfalls männlich.

Es folgt eine deskriptive Beschreibung der demographischen Variablen. Sie beginnt mit einem Überblick, an welcher Schulform Physiklehrkräfte und Lehrkräfte im Vorbereitungsdienst zum Zeitpunkt der Teilnahme unterrichteten und auf welche Schulform ihr Studium ausgerichtet war bzw. das der Studierenden ausgerichtet ist. Für die Gruppe der Lehrkräfte werden anschließend Daten über ihre Ausbildung und Fortbildungsaktivität dargestellt.

Tabelle 11 gibt einen Überblick, in welcher **Schulform** unterrichtet wird. Die Stichprobe besteht hauptsächlich aus Gymnasiallehrkräften. 80% der untersuchten Lehrkräfte unterrichten an der Schulform, für die sie ausgebildet wurden (vgl. Tabelle 12).

**Tabelle 10:** Anzahl, Alter und Geschlecht der Versuchspersonen. Fehlende Prozentangaben zum Geschlecht beruhen auf fehlenden Angaben.

	Anzahl			Geschlecht		Alter	
	1. Studie	2. Studie	Gesamt	♀	♂	M	SD
	N	N	N	%	%		
Physiklehrkräfte	93	186	279	28	71	44	10
Lehrkräfte im Vorbereitungsdienst	21	41	62	35	65	31	7
Lehramtsstudierende Physik	26	77	103	39	60	25	5
Lehrkräfte anderer Fächer	10	21	31	45	55	39	10
Dipl. Phys.	15	7	22	23	77	29	3
Studierende Physik	21	-	21	48	52	22	2

M Mittelwert

SD Standardabweichung

Dipl. Phys. Diplom-Physikerinnen und –Physiker

**Tabelle 11:** Anzahl der Versuchspersonen, aufgeteilt nach der Schulform, an der sie zur Zeit der Untersuchung unterrichteten und der studierten Schulform

	Aktuelle Schulform			Belegter Studiengang		
	GY	HR	GS	GY	NGY	Diplom
Physiklehrkräfte	216	47		166	42	38
LiV	50	2	9	37	12	13
Lehramtsstudierende Physik				49	49	

GY Gymnasium / Studium eines gymnasialen Lehramtes

HR Haupt- und Realschule

GS Gesamtschule

NGY Studium eines nicht-gymnasialen Lehramtes, also Haupt-, Real- und Gesamtschule

LiV Physiklehrkräfte im Vorbereitungsdienst

**Tabelle 12:** Zusammenhang zwischen studierter und aktueller Schulform für alle Physiklehrkräfte, die beide Informationen angegeben haben

	Unterrichtet zur Zeit am Gymnasium	Unterrichtet zur Zeit an einer anderen Schulform
Studium des gymnasialen Lehramtes	159	7
Diplomstudiengang	28	10
Studium des nicht-gymnasialen Lehramtes	3	38

50% der Physiklehrkräfte, die kein Lehramt studiert, sondern einen **Diplom-Studiengang** absolvierten, haben ein Physik-Diplom (19 Personen). Sechs Personen (16%) erwarben ein Diplom im Bereich Ingenieurwissenschaft und Technik. Jeweils drei Personen haben ein Diplom der Biologie, Chemie oder Mathematik. Lehrkräfte mit dem Abschluss Diplom-Lehrkraft fallen nicht in diese Kategorie.

150 Physiklehrkräfte haben in ihrer eigenen Zeit als Schülerin oder Schüler in der **Oberstufe Physik** durchgängig belegt. 11 haben keine Naturwissenschaft belegt, 25 Biologie oder Chemie, aber keine Physik.

225 der Physiklehrkräfte haben das Fach **Physik studiert**. 53 haben das Fach Physik nicht studiert, sie unterrichten fachfremd. 171 haben das Fach Mathematik studiert, 107 nicht.

Tabelle 13 zeigt das **Fortbildungsverhalten** der Physiklehrkräfte der zweiten Studie. Es wurden mehr allgemein pädagogische Fortbildungen besucht als Fortbildungen zum Fach Physik.

106 Physiklehrkräfte unterrichten auch in der **Oberstufe**, 53 höchstens in der **Mittelstufe** und für 13 Physiklehrkräfte stellt die **Unterstufe** die höchste unterrichtete Schulstufe dar.

Die Lehrkräfte unterrichteten zum Zeitpunkt der Erhebung in sieben verschiedenen **Bundesländern**. Die größten Stichproben liefern Nordrhein-Westfalen (N=144) und Bayern (N=64), in denen die zweite Studie durchgeführt wurde. Es folgen Hessen (N=29), Bremen (N=15), Thüringen (N=15), Niedersachsen (N=7) und Berlin (N=5), die an der ersten Studie beteiligt waren.

**Tabelle 13:** Besuchte Fortbildungen und Anzahl der besuchten Fortbildungen der Physiklehrkräfte

		Anzahl
Allgemein pädagogische Fortbildungen in den letzten zwei Jahren	Nie	39
	Einmal	46
	Zwei- bis dreimal	74
	Mehr als dreimal	27
Fortbildungen im Fach Physik in den letzten zwei Jahren	Nie	62
	Einmal	44
	Zwei- bis dreimal	48
	Mehr als dreimal	28
Individuelle Fortbildung	Nie	20
	Mindestens einmal im Jahr	32
	Mindestens alle 6 Mo- nate	69
	Monatlich	63
Fachgruppenarbeit in Physik an der Schule	Nie	33
	Seltener als einmal im Jahr	47
	Seltener als alle 6 Mo- nate	73
	Monatlich	29

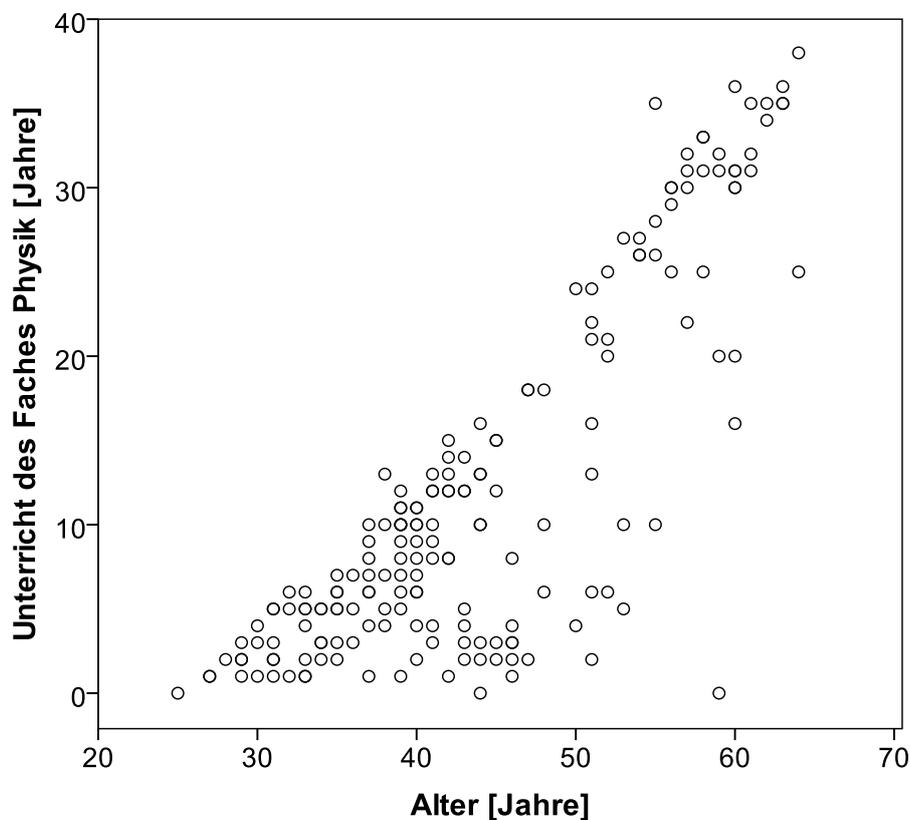
Die Physiklehrkräfte unterrichten im Schnitt seit 12.2 Jahren Physik (Standardabweichung 10.7 Jahre, Maximalwert 38 Jahre).

Die Variablen zur **Berufserfahrung** (Alter, Jahre im Schuldienst und Jahre, in denen bereits Physik unterrichtet wurde) korrelieren erwartungsgemäß hoch miteinander:

- Die Korrelation des Alters mit der Anzahl an Jahren, in denen bereits Physik unterrichtet wurde, liegt bei  $r=.829$ ,  $p<.001$
- Die Korrelation des Alters mit den Jahren im Schuldienst liegt bei  $r=.873$ ,  $p<.001$
- Die Korrelation der Jahre im Schuldienst mit der Anzahl an Jahren, in denen bereits Physik unterrichtet wurde, liegt  $r=.949$ ,  $p<.001$

In Abbildung 13 ist zu erkennen, dass einige Lehrkräfte bei gleichem Alter weniger Berufserfahrung haben als ihre Kolleginnen und Kollegen. Dies könnte auf einen späteren Einstieg, z. B. durch einen Quereinstieg, oder Auszeiten, z. B. Erziehungszeiten, zurückzuführen sein.

Die Physiklehrkräfte unterrichten im Mittel 8.1 **Stunden** Physik in der Woche (Standardabweichung 5 Stunden, Maximalwert 26 Stunden).



**Abbildung 13:** Zusammenhang zwischen Alter und Lehrerfahrung der Physiklehrkräfte

# *6. KAPITEL*

## *AUSWERTUNG DER*

### *UNTERSUCHUNG*

Im vorhergehenden Abschnitt wurde die Datenerhebung dargelegt. Nun soll erläutert werden, wie aus den Rohdaten, das heißt kurzen Texten, Rechnungen und Multiple-Choice-Antworten der Versuchspersonen, Werte gewonnen und diese zu Personenfähigkeiten aggregiert werden.

## 6.1 KODIERUNG UND INTERRATERRELIABILITÄT

Alle Antworten werden in IBM SPSS Statistics Version 20.0 mit 0, 1 oder 2 Punkten kodiert. Den Antworten auf halboffene Fragen wird die Punktzahl nach einem Manual zugeordnet (Kirschner et al., 2011a; Kirschner et al., 2011b). Wird eine Aufgabe nicht bearbeitet, wird sie als fehlend kodiert (Missing).

Bei allen geschlossenen Aufgaben wird davon ausgegangen, dass sie korrekt kodiert werden.

Zur Sicherstellung der Übereinstimmung zwischen den Raterinnen werden in einer Teilstichprobe Goodmans und Kruskal Gamma und die Intraklassenkorrelation (ICC) als Qualitätskriterien untersucht. Da es sich um ordinalskalierte Daten handelt, kann Goodmans und Kruskal Gamma als Maß für die Interraterübereinstimmung betrachtet werden (Wirtz & Caspar, 2002). „Ist die Anzahl der von beiden Ratern gemeinsam tendenziell höher bzw. niedriger eingeschätzten Objekte ähnlich der bei zufälliger Zuteilung erwarteten Anzahl, liegt Gamma nahe 0. (...) Ein Wert nahe 1 zeigt eine sehr hohe Korrelation des Ratings an.“ (Wirtz & Caspar, 2002, S. 145). Werden die Daten als intervallskaliert angenommen, kann für diesen Fall die ICC als am besten geeignete Reliabilitätsbestimmung angewendet werden (Wirtz & Caspar, 2002). Dabei handelt es sich um „Maße für den Anteil der Varianz aller Ratingwerte, der durch die wahren Werte der Personen erklärt werden kann“ (Wirtz & Caspar, 2002).

Es kann zwischen der justierten und unjustierten ICC unterschieden werden, wobei bei der Berechnung der unjustierten ICC Unterschiede im Mittelwert zwischen den Raterinnen und Ratern mit einfließen und bei der Berechnung der justierten ICC nicht. Für die Klassifikation der ICC werden zwei Ziffern angegeben. Die erste bezieht sich darauf, ob alle Kodierungen zur Berechnung der Interraterübereinstimmung von allen Raterinnen und Ratern durchgeführt wurden und diese zufällig aus einer Menge möglicher Raterinnen und Rater ausgewählt wurden. Die zweite Ziffer bezieht sich darauf, wie viele Ratings verschiedener Raterinnen und Rater jeweils einbezogen werden können. In dieser Arbeit wird die ICC<sub>2,2</sub> (unjustiert) genutzt, für die zwei Raterinnen ausgewählt wurden, die beide alle Aufgaben kodieren. Als Grenze gilt  $ICC > .7$ , die die Betrachtung von Gruppenunterschieden zulässt (Wirtz & Caspar, 2002). Die Ergebnisse sind in Abschnitt 7.1 dargestellt.

## 6.2 RASCH-ANALYSE

Zur Auswertung der Daten werden auf Basis der Rohwerte mit Hilfe einer Rasch-Analyse Personenfähigkeiten und Itemschwierigkeiten geschätzt. Beide sind metrisch skaliert; es ist möglich, festzustellen, dass eine Person x-mal fähiger ist als eine andere, oder ein Item x-mal schwieriger zu lösen ist als ein anderes (vgl. z. B. Carstensen, 2000). Das Rasch-Modell ist ein Spezialfall der Item-Response-Theorie (Bond & Fox, 2007). Die Personenfähigkeit wird mit der Itemschwierigkeit in Beziehung gesetzt, um für jede Person und jedes Item eine Lösungswahrscheinlichkeit schätzen zu können (Bortz & Döring, 2006). Dabei sind die Eigenschaften von Personen und Items voneinander trennbar (Carstensen, 2000); Im Rasch-Modell ist die Personenfähigkeit unabhängig von der Auswahl der Items aus einer Itemmenge bzw. die Itemschwierigkeit unabhängig von der untersuchten Stichprobe (Spezifische Objektivität, Bond & Fox, 2007).

Da in dieser Studie die meisten Items nicht dichotom sind, wird ein Partial-Credit-Modell genutzt (Wu, Adams, & Wilson, 1998), wobei es sich um eine Verallgemeinerung des Rasch-Modells auf ordinale Daten handelt (Carstensen, 2000). In diesem Fall müssen beim Lösen eines Items von der Testperson Schwellen überschritten werden, um mehr Punkte zu erreichen. Zur Berechnung der Lösungswahrscheinlichkeit wird die Personenfähigkeit mit der Schwierigkeit, eine bestimmte Schwelle eines bestimmten Items zu überschreiten, in Beziehung gesetzt (Carstensen, 2000).

Neben der metrischen Skala und der spezifischen Objektivität ist ein weiterer Vorteil der Rasch-Analyse, dass den einzelnen Punktzahlen, die eine Person in einer Aufgabe erreicht, weitere Informationen zugeordnet werden können. Zum Beispiel kann angegeben werden, welcher Dimension ein Item zugeordnet ist (z. B. Carstensen, 2000), welches Geschlecht die Person hat (z. B. Hartig, Jude, & Wagner, 2008), oder welches Testheft sie bekommen hat (z. B. Hartig et al., 2008). Werden offene Aufgaben analysiert, kann angegeben werden, wer die Aufgabe kodiert hat (Bond & Fox, 2007; Linacre, 2011a). Werden mehrere Studien zusammengefasst, kann angegeben werden, an welcher Studie die Person teilgenommen hat. All diese Informationen können in die Rasch-Analyse einbezogen werden.

Die Rasch-Analyse für diese Arbeit wird in ConQuest (Wu, Adams, & Haldane, 2007) durchgeführt.

Mit Hilfe der **Fit-Statistik** wird untersucht, ob empirisch gewonnene Daten und geschätzte Itemschwierigkeiten und Personenfähigkeiten gut genug zueinander passen (Linacre, 2011b). Dazu wird der Infit bzw. der weighted Fit analysiert. Dieser reagiert weniger sensibel auf extrem abweichende Rohwerte als der klassische Outfit bzw. unweighted Fit (Bond & Fox, 2007; Linacre, 2011b; Wu et al., 1998). Es werden keine Personen aus der Analyse ausgeschlossen, da erstens die Stichprobe aus Erwachsenen besteht, die freiwillig teilnahmen, und daher davon ausgegangen wird, dass sie die Tests ernst nehmen, und zweitens aufgrund des hohen Anteils an offenen Aufgaben angenommen werden kann, dass das Antwortverhalten nicht zufällig ist.

Allerdings werden durch die Betrachtung des Infits Personen bei der Beurteilung der Rasch-Konformität der Items stärker gewichtet, deren Fähigkeiten nah bei der Itemschwierigkeit liegen (Bond & Fox, 2007). Zur Beurteilung der Itempassung wird die mean-square-fit-Statistik (MnSq) für die Gruppe der Physiklehrkräfte betrachtet. Liegt der MnSq unter zwei, stört das Item die Messung nicht (Linacre, 2011b). Items außerhalb der Grenze  $MnSq > 2.0$  werden aus der weiteren Analyse entfernt.

Um zu untersuchen, ob die angenommene Aufteilung in CK, PCK und PK empirisch nachweisbar ist, wird eine **mehrdimensionale Rasch-Analyse** durchgeführt (Wu et al., 1998). Dabei werden die Items auf theoretischer Grundlage bestimmten Dimensionen zugeordnet. Durch dieses konfirmatorische Verfahren werden die Hypothesen zur Struktur überprüft (Carstensen, 2000). In dieser Arbeit wird untersucht, ob sich das Professionswissen in mehrere Dimensionen gliedert und ob somit ein Modell, in dem die Items der Tests (CK, PCK, PK<sub>D</sub>, PK<sub>KP</sub>) jeweils einer eigenen Dimension zugeordnet werden, besser passt, als Modelle, in denen jeweils zwei, drei oder alle vier Itemgruppen zusammengefasst werden.

Als Kriterium für das beste Modell wird die Final Deviance herangezogen. Dies ist eine Kenngröße, die angibt, wie gut die Modellierung zu den Daten passt (Wu et al., 1998). Es gibt keinen cut-off-Wert für die Final Deviance, da sie nicht absolut bewertet wird. Sie soll im Vergleich zwischen jeweils zwei Modellen mit der gleichen Datenmatrix, also den gleichen Items und Versuchspersonen, möglichst klein sein. Dies wird mit Chi-Quadrat-Tests (Bortz, 2005) untersucht (Wu et al., 1998).

## Rasch-Analyse

Um ein komplexeres Modell zu bevorzugen, muss es signifikant besser passen als das einfachere Modell mit weniger Dimensionen. Passen zwei Modelle mit der gleichen Anzahl an Dimensionen gleich gut, wird das gewählt, in welchem weniger Items als störend eingestuft werden müssen.

Als weitere Kriterien werden die Informationskriterien BIC (Bayes Information Criterion) und CAIC (Akaike Information Criterion) genutzt. Diese beziehen neben der Deviance die Anzahl der Freiheitsgrade und die Stichprobengröße ein. Sie sollen ebenso wie die Final Deviance möglichst klein sein und dürfen nicht absolut bewertet werden (Bühner, 2006; van Vorst, 2013).

In dieser Arbeit steht die Dimensionalität des Professionswissens von Physiklehrkräften im Vordergrund. Zur Untersuchung werden die Antworten der Physiklehrkräfte **iterativ** analysiert. Nach der Dimensionsanalyse (7.2.1) wird in dem Modell mit der besten Passung der MnSq der Items in den einzelnen Dimensionen betrachtet. Nach dem Entfernen der störenden Items wird die Dimensionsanalyse mit den verbliebenen Items erneut ausgeführt, um sicherzustellen, dass nicht ein anderes Modell besser zu dieser Itemauswahl passt (7.2.3).

Da zu den offenen Antworten jeweils zwei Ratings vorliegen (siehe 7.2) und diese Information nicht verloren gehen soll, werden jeweils beide Ratings in die Rasch-Analyse in Form einer **Ratermodellierung** einbezogen. Dabei wird angegeben, von welcher der beiden Raterinnen die angegebene Punktzahl stammt und es wird modelliert, ob eine von ihnen bei der Punktevergabe anspruchsvoller ist als die andere. Zu den geschlossenen Antworten existiert nur eine Kodierung, es spielt keine Rolle, welche der Raterinnen sie kodiert hat.

Einige Items konnten unverändert aus der ersten in die zweite Studie übernommen werden (vgl. Abschnitt 5.2.1). Daher ist es möglich, die Daten **beider Studien** in einer Auswertung zu verknüpfen. Sie werden auf dieselbe Weise ausgewertet, als wären zwei unterschiedliche, über Ankeritems verbundene, Testhefte genutzt worden. Dies ist Grundlage für die weitere Arbeit. Es ist allerdings nicht auszuschließen, dass einzelne Items, z. B. aufgrund ihrer Position im Testheft, in den beiden Studien nicht vollständig gleich bearbeitet wurden. Daher wird in der Rasch-Analyse angegeben, welcher Studie jede Person zugeordnet ist.

Damit kann ein DIF (**Differential Item Function**) (Bond & Fox, 2007) modelliert werden, der Unterschiede zwischen Gruppen einbezieht (Wu et al., 1998). Es wird untersucht, ob es zielführend ist, den Unterschied zwischen den zwei Studien im Rasch-Modell zu modellieren. Dazu wird in den Dimensionen CK und PCK jeweils die Final Deviance mit und ohne Modellierung des DIF miteinander verglichen. Wenn die Modellierung des DIF die Final Deviance signifikant verbessert, wird sie genutzt. Der PK-Test ist von der Untersuchung des DIF nicht betroffen, da er nur während der zweiten Studie eingesetzt wurde.

Für die Datenanalyse werden die Daten in Form einer **Datenmatrix** benötigt. Diese Form wird im Folgenden beschrieben, ein Beispiel ist in Abbildung 14 dargestellt. Jeder Person ist eine ID zugeordnet. Zu jeder offenen Antwort werden jeweils die Kodierungen von Raterin A und B angegeben. Die Multiple-Choice-Items wurden nur einmal kodiert, die Kodierungen werden allerdings zweimal, jeweils einmal bei jeder Raterin, eingefügt, so dass zwei komplette Datensätze entstehen. Mit diesem Vorgehen, durch das Informationen doppelt angegeben werden, wird die Reliabilität künstlich erhöht (Linacre, 2011a), daher muss die so berechnete Reliabilität als obere Grenze verstanden werden.

ID	Studie	Raterin A		Raterin B	
01	1	1221200201	11001	2221200201	11001
02	1	1200012012	10011	1201012012	10011
03	2	1200210210	01111	1200210211	01111
04	2	0012021010	00001	1012021020	00001
05	2	2012010021	01011	2012010021	01011

**Abbildung 14:** Design der Datenmatrix. Für jede Raterin sind links die Kodierungen der offenen Aufgaben zu sehen und rechts die der geschlossenen Aufgaben, die für beide Raterinnen identisch sind.

Als Maß für die **Reliabilität** wird die EAP/PV-Reliabilität angegeben (Walter, 2005), deren Höhe generell mit Cronbachs Alpha vergleichbar ist (Rost, 2004).

Zur Berechnung einer zuverlässigen Reliabilität werden die Dimensionen, die sich in der Analyse ergeben, getrennt untersucht (7.2.4).

Da der PCK-Test nur aus offenen Aufgaben besteht, die zweimal kodiert wurden, werden die Informationen der zwei Raterinnen genutzt, ohne dass Informationen doppelt eingegeben würden.

Der PK-Test hingegen besteht nur aus geschlossenen Items, die nur einmal kodiert wurden, und wurde nur in der zweiten Studie eingesetzt. Da weder der Ratereffekt noch der DIF zwischen den Studien modelliert wird, kann für den PK-Test zusätzlich das klassische Maß Cronbachs Alpha angegeben werden.

Eine obere Grenze für die Reliabilität des CK-Tests ergibt die oben beschriebene Rechnung. Um eine untere Grenze für die CK-Reliabilität zu finden, wird die Reliabilität der offenen und geschlossenen Items getrennt berechnet. Aufgrund der geringen Itemanzahl wird die Reliabilität der Teilskalen kleiner sein als die Reliabilität des gesamten CK-Tests.

Um die Reliabilität des gesamten CK-Tests zu berechnen, ohne Informationen mehrfach in die Rechnung einzubeziehen, wird ein Dummy-Rater eingeführt. Dieser hat nur die geschlossenen Aufgaben kodiert, so dass diese nicht mehr den Raterinnen zugeordnet werden müssen. Ein Beispiel ist in Abbildung 15 dargestellt.

Die Nutzung eines Dummy-Raters kann allerdings einen Einfluss auf die Ratermodellierung und somit auf die geschätzten Personenfähigkeiten haben. Sind zum Beispiel alle geschlossenen Items sehr einfach, werden die beiden Raterinnen im Vergleich zum Dummy-Rater als sehr streng angenommen. Die (hypothetische) Änderung der Itemschwierigkeit kann auch einen Einfluss auf die Reliabilität haben, wenn sie dazu führt, dass die Itemschwierigkeit besser oder schlechter zu den Personenfähigkeiten passt: Die Reliabilität steigt, wenn Itemschwierigkeiten und Personenfähigkeiten besser zueinander passen (Linacre, 2011a). Daher hat die so berechnete Reliabilität eine Unschärfe, die in Kauf genommen wird, um eine Reliabilität für den gesamten CK-Test angeben zu können. Ihre Verlässlichkeit muss in Zusammenhang mit der vorher bestimmten oberen und unteren Grenze diskutiert werden.

ID	Studie	Raterin A		Raterin B		Dummy-Rater	
01	1	1221200201	99999	2221200201	99999	9999999999	11001
02	1	1200012012	99999	1201012012	99999	9999999999	10011
03	2	1200210210	99999	1200210211	99999	9999999999	01111
04	2	0012021010	99999	1012021020	99999	9999999999	00001
05	2	2012010021	99999	2012010021	99999	9999999999	01011

**Abbildung 15:** Design der Datenmatrix zur Berechnung der Reliabilität. Links sind die Kodierungen der offenen Aufgaben zu sehen, rechts die der geschlossenen Aufgaben. 9 stellt eine fehlende Information dar (Missing).

### 6.2.1 BERECHNUNG DER PERSONENFÄHIGKEIT

Aus dem endgültigen Itempool ergeben sich mehrere Dimensionen des Professionswissens (7.2.3). Die Berechnung der Itemschwierigkeiten wird getrennt in den einzelnen Dimensionen durchgeführt.

Aufgrund der spezifischen Objektivität sollte die Auswahl der Stichprobe keinen Einfluss auf die Itemschwierigkeit haben. Da allerdings von Versuchspersonen, die zur Validierung herangezogen wurden, keine gute Modellpassung erwartet wird, werden die Itemschwierigkeiten für die Gruppe der Physiklehrkräfte aus beiden Studien bestimmt und fixiert (anchor item response parameter). In die Berechnung fließen neben Item und Schwellen (zwischen 0 und 1 bzw. 1 und 2 Punkten) auch Rater und Studie mit ein (model rater + item + item\*step + item-studie + item\*studie).

Um Personenparameter für die komplette Stichprobe (auch Lehrkräfte anderer Fächer, Studierende, Lehrkräfte im Vorbereitungsdienst, Physikerinnen und Physiker) zu erhalten, werden die Personenparameter für die komplette Stichprobe auf Basis der fixierten Itemschwierigkeiten geschätzt. Es werden Weighted Likelihood Estimates (WLEs) genutzt, die eine Individualdiagnostik ermöglichen und keine Hintergrundmodelle (wie Geschlecht, Schulzweig oder Personenfähigkeiten in anderen Dimensionen) einbeziehen (Hartig et al., 2008). Die Berechnung der WLEs ist in Walter (2005) dargestellt. Die WLEs werden für jede Dimension getrennt geschätzt.

Alle weiteren Analysen werden auf Basis der den Versuchspersonen zugeordneten WLEs durchgeführt. Zur besseren Vergleichbarkeit werden die WLEs standardisiert. Dem Vorgehen in der PISA-Studie entsprechend (OECD, 2006), liegt der Mittelwert der Personenfähigkeit bei 500 und die Standardabweichung bei 100 (dazu wird zu den z-standardisierten WLEs (Mittelwert 0, Standardabweichung 1) erst 5 addiert und der

Wert dann mit 100 multipliziert). Bei einer Gesamtstichprobe von 518 Personen ergibt sich bei einer Standardabweichung von 100 ein Standardfehler von 4.4. Dies lässt eine Angabe der Ergebnisse auf drei Stellen ohne Nachkommastellen zu (z. B. 512). In kleineren Substichproben steigt der Standardfehler. Aus Gründen der Vergleichbarkeit wird auf eine unterschiedliche Rundung für verschiedene Gruppen verzichtet. Die standardisierten WLEs sind dimensionslos.

Die erhaltenen Werte (standardisierte WLEs) werden Personenfähigkeit genannt und sind ein Maß für das CK, PCK und PK der einzelnen Versuchspersonen.

Die Personenfähigkeit wird dazu genutzt, die Validität zu bestimmen (7.3), Zusammenhänge zwischen den Dimensionen CK, PCK und PK (7.5.1), sowie CK, PCK und unabhängigen Variablen zu berechnen (7.5.2), Regressionen durchzuführen (7.5.3) sowie verschiedene Ausbildungsschritte zu vergleichen (7.5.4). Um dies zu erreichen, werden klassische Verfahren genutzt, die im nächsten Abschnitt beschrieben werden.

### 6.3 KLASSISCHE ANALYSE

Es wird ein kurzer Überblick gegeben, welche Verfahren zur Bestimmung von Gruppenunterschieden und Korrelationen genutzt werden, sowie die Herangehensweise an die Regression beschrieben. Alle Rechnungen werden mit IBM SPSS Statistics Version 20.0 durchgeführt.

**Mittelwertunterschiede** zwischen zwei Gruppen werden mit t-Tests analysiert (Field, 2009). Bei Gruppengrößen von weniger als 30 Versuchspersonen muss mit nonparametrischen Verfahren gearbeitet werden, sofern die Annahme der **Normalverteilung** verletzt ist (Bortz & Döring, 2006). Die Untersuchung der Normalverteilung für kleine Gruppen geschieht mit Hilfe des Kolmogorov-Smirnov-Tests, der die untersuchte Verteilung mit einer Normalverteilung vergleicht (Field, 2009). Ergänzend werden auch Schiefe und Kurtosis getrennt untersucht, indem die zugehörigen z-Werte (Wert/Standardabweichung) berechnet werden. In kleinen Stichproben gilt  $z > 1.96$  als signifikante Abweichung von der Normalverteilung mit  $p < .05$  (Field, 2009). Eine Normalverteilung wird nicht angenommen, wenn entweder der Kolmogorov-Smirnov-Test Signifikanz zeigt, oder die Verteilung eine signifikante Schiefe und Kurtosis aufweist. Kann keine Normalverteilung angenommen werden, werden Mann-Whitney-Tests zur Berechnung von Unterschieden zwischen zwei Gruppen herangezogen (Field, 2009).

Werden mindestens drei Gruppen miteinander verglichen, wird unter der Annahme einer Normalverteilung eine ANOVA mit Post-Hoc-Test berechnet. Sind die Gruppen, die verglichen werden, unterschiedlich groß, wird eine GT2-Analyse nach Hochberg als Post-Hoc-Test durchgeführt (Field, 2009). Kann keine Normalverteilung angenommen werden, wird anstatt einer ANOVA ein Kruskal-Wallis-Test berechnet (Field, 2009). Finden sich dort Unterschiede, werden die einzelnen Gruppen mit Mann-Whitney-Tests verglichen. Dabei muss das Signifikanzkriterium angepasst werden, um Fehler 1. Art zu minimieren. Deshalb wird gefordert, dass die ausgegebene Signifikanz geteilt durch die Anzahl der durchgeführten Tests kleiner als .05 ist (Field, 2009).

Zur Untersuchung von Zusammenhängen zwischen intervallskalierten Daten (wie CK, PCK, PK, oder Alter) oder Ordinaldaten (z. B. Häufigkeiten von Fortbildungsbesuchen) wird Pearsons **Korrelationskoeffizient**  $r_P$  berechnet und angegeben. Werden von drei Variablen jeweils zwei miteinander korreliert, kann die Korrelation zwischen zwei Variablen mit der Korrelation zwischen einer der beiden und der dritten Variable in Verhältnis gesetzt werden um anzugeben, ob sie sich signifikant unterscheiden (Bortz, 2005).

Für Zusammenhänge zwischen echten Nominaldaten (wie Geschlecht) und intervallskalierten Daten wird eine Punkt-Biserial-Korrelation mit dem Korrelationskoeffizient  $r_{PB}$  angegeben (Field, 2009). Im Folgenden wird in beiden Fällen von dem Korrelationskoeffizienten  $r$  gesprochen, der immer zwischen Null und Eins liegt.

Um den Einfluss mehrerer unabhängiger Variablen gleichzeitig zu untersuchen, werden schrittweise **Regressionen** mit den Daten aller untersuchten Physiklehrkräfte durchgeführt. Dabei werden unabhängige Variablen aus der Regression entfernt, die keinen signifikanten Einfluss auf das Modell haben. Diese Form der Regression wird backwards genannt (Field, 2009).

Zur Beurteilung von Regressionen werden die Variablen  $R^2$ , F, Beta (unstandardisiert) und ein standardisierter Beta-Koeffizient angegeben. Diese werden auf Grundlage von Field, 2009 kurz erläutert:

$R^2$  steht für die Aufklärung der abhängigen Variable durch die Regression und liegt zwischen Null und Eins, wobei Eins für eine perfekte Aufklärung steht. Wird nur eine unabhängige Variable betrachtet, ist R identisch mit dem Korrelationskoeffizienten r.

F ist ein Maß für die Modellgüte, das mit dem Fehler des Modells ins Verhältnis gesetzt wurde und daher größer als Eins sein sollte, wobei größere Werte einer höheren Modellgüte entsprechen. Um von einer guten Modellpassung zu sprechen, muss F signifikant, also nur mit geringer Wahrscheinlichkeit zufällig entstanden, sein. Dies ist in allen später berichteten Regressionen der Fall.

Beta beschreibt den Einfluss einer unabhängigen Variable auf die abhängige Variable unter Kontrolle aller anderen unabhängigen Variablen in der Regression. Ändert sich die unabhängige Variable um eine Einheit (z. B. einen Punkt oder von ja zu nein), ändert sich die abhängige Variable um eine Einheit, multipliziert mit dem Faktor Beta (unstandardisiert). Einfacher zu vergleichen als Beta ist der standardisierte Beta-Koeffizient. Wenn sich die unabhängige Variable um eine Standardabweichung ändert, ändert sich die abhängige Variable mit dem standardisierten Beta-Koeffizienten multipliziert. Um Beta einen Einfluss zuzuschreiben, muss ihm ein  $p < .05$  zugeordnet werden. Ein Beispiel für die Interpretation von Beta wird in 7.5.1 gegeben.

Regressionen, bei denen die Reihenfolge der unabhängigen Variablen nicht theoretisch vorgegeben ist und keine empirischen Ergebnisse zum Vergleich vorliegen, wird unterstellt, nicht generalisierbar zu sein, da sie sich nur auf eine Stichprobe beziehen (Tabachnick & Fidell, 2001). Daher werden die dargestellten Regressionen kreuzvalidiert. Dies geschieht, indem die Regressionen mit den theoretisch identifizierten Variablen nur mit einem Teil der Stichprobe durchgeführt werden (backwards) und die erhaltenen Beta-Koeffizienten zur Vorhersage des Wissens der restlichen Stichprobe mit einer linearen Funktion genutzt werden. Das Quadrat der Korrelation zwischen den vorhergesagten Wissen und dem gemessenen Wissen ( $R^2_{\text{Vorhersage}}$ ) sollte nicht wesentlich kleiner sein als die Aufklärung der vorgehenden Regression für die erste Teilstichprobe ( $R^2_{\text{Regression}}$ ) (Field, 2009; Tabachnick & Fidell, 2001; weiterführend Bortz, 2005).

# *7. KAPITEL*

## *ERGEBNISSE*

Dieses Kapitel hat die Ergebnisse zu den in Kapitel 4 aufgestellten Forschungsfragen zum Inhalt. Zuerst werden die Ergebnisse zur Objektivität der Auswertung der offenen Aufgaben dargestellt (7.1), anschließend wird die Modellpassung beleuchtet (7.2). Daraufhin werden Hypothesen zu der Validität der einzelnen Dimensionen getestet (7.3), bevor die Personenmerkmale genauer betrachtet werden (7.5).

## 7.1 ERGEBNISSE ZUR OBJEKTIVITÄT

Im folgenden Abschnitt wird auf die Interraterreliabilität, also die Übereinstimmung zwischen Raterinnen und Ratern, eingegangen. Es wurde eine zufällige Stichprobe von 10% der Physiklehrkräfte mehrfach kodiert, um die Interraterübereinstimmung für die Kodierung sicherzustellen.

Als Kriterium wurde für Goodmans und Kruskal Gamma eine Grenze von .7 gewählt, die überschritten werden musste. Zusätzlich wurde die Intraklassenkorrelation (ICC) untersucht. Im PCK-Teil wurde die ICC<sub>2,2</sub> (unjustiert) genutzt. Für alle später kodierten Items lag die ICC<sub>2,2</sub> (unjustiert) über .7. Daraus ergibt sich, dass alle Antworten von allen Versuchspersonen zu allen offenen PCK-Items doppelt kodiert wurden (siehe Abschnitt 6.1). Der Vollständigkeit halber wurden auch die Antworten zu allen offenen CK-Items doppelt kodiert. Hier hätte allerdings auch eine Einfachkodierung ausgereicht, da die ICC<sub>2,1</sub> (unjustiert) für alle Items, die später ausgewertet wurden, über .7 liegt. Die ICC<sub>2,2</sub> (unjustiert) lag für diese CK-Items über .88.

Zwei PCK-Items und ein CK-Items konnten nicht analysiert werden, da keine zufriedenstellende Interraterreliabilität erreicht werden konnte (siehe Anhang, Tabelle 37).

## 7.2 ERGEBNISSE ZUR MODELLIERUNG

In diesem Abschnitt wird untersucht, welches Modell in einer mehrdimensionalen Rasch-Analyse die beste Passung erreicht:

- Ein dreidimensionales Modell mit den Dimensionen CK, PCK und PK
- Ein eindimensionales Modell: Professionswissen
- Ein zweidimensionales Modell mit den Dimensionen fachspezifisches Professionswissen (CK und PCK) und pädagogisches Wissen (PK)
- Ein zweidimensionales Modell mit den Dimensionen Wissen über das Fach (CK) und Wissen über Unterricht (PCK und PK)

Die Hypothese war, dass es sich bei CK, PCK und PK um drei getrennte Dimensionen handelt. Der Ablauf der Analyse wurde in Abschnitt 6.2 dargestellt. Damit wird die Forschungsfrage „Lassen sich die theoretisch angenommenen drei Dimensionen des Professionswissens, CK, PCK und PK, empirisch voneinander trennen?“ beantwortet. Ergänzend wird betrachtet, ob die Dimensionen, die sich ergeben, reliabel sind.

### 7.2.1 ERSTE DIMENSIONSMODELLIERUNG

Aufgrund der Konzeption des PK-Tests werden vier mögliche Dimensionen betrachtet: CK, PCK, PK<sub>KP</sub> und PK<sub>D</sub> (siehe 5.1.2.3). Es werden sechs verschiedene Modelle gegeneinander getestet, die in Tabelle 14 dargestellt sind. Alle Rechnungen werden mit den gleichen Randbedingungen in ConQuest durchgeführt (Abbruch nach 1000 Iterationen, method=montecarlo, nodes=1000, converge=.001). Die Ergebnisse sind in Tabelle 15 dargestellt.

Der Unterschied der Final Deviance zwischen dem vierdimensionalen Modell und allen anderen ist signifikant ( $p < .001$ ), wie mit dem Chi-Quadrat-Test belegt wurde. Dies gilt sowohl für die Physiklehrkräfte der zweiten Studie, die alle drei Testteile bearbeitet haben, als auch für die Stichprobe aller Physiklehrkräfte. Da der Modelfit (Final Deviance) eindeutig ist, muss die Anzahl der Items, die das Modell stören, nicht als Auswahlkriterium für die Wahl des Modells herangezogen werden. BIC und CAIC zeigen dasselbe Ergebnis. Allerdings kann noch keine Reliabilität angegeben werden, da die Passung der Items in die jeweiligen Modelle noch nicht gewährleistet ist.

### 7.2.2 EINFLUSS DIF ZWISCHEN DEN STUDIEN

Um zu untersuchen, ob es sinnvoll ist, den Unterschied zwischen der ersten und der zweiten Studie im Rasch-Modell zu modellieren, werden in den Dimensionen CK und PCK jeweils die Final Deviance mit und ohne Schätzung des DIF miteinander verglichen. Da das Modell jeweils mit der Modellierung des DIF signifikant besser passt, wird diese in allen weiteren Rechnungen einbezogen, in denen beide Studien zusammengefasst werden.

## Ergebnisse zur Modellierung

**Tabelle 14:** Überblick über die sechs Modelle, die gegeneinander getestet werden

Modell	1	2	3	4	5	6
Anzahl Dimensionen	1	2	2	3	3	4
Dimension 1	CK & PCK & PK <sub>KP</sub> & PK <sub>D</sub>	CK & PCK	CK	CK	CK & PCK	CK
Dimension 2	-	PK <sub>KP</sub> & PK <sub>D</sub>	PCK & PK <sub>KP</sub> & PK <sub>D</sub>	PCK	PK <sub>KP</sub>	PCK
Dimension 3	-	-	-	PK <sub>KP</sub> & PK <sub>D</sub>	PK <sub>D</sub>	PK <sub>KP</sub>
Dimension 4	-	-	-	-	-	PK <sub>D</sub>

**Tabelle 15:** Freiheitsgrade, Final Deviance und die Anzahl der Items, die nicht Raschkonform sind, für die sechs möglichen Modelle. Oben alle Physiklehrkräfte, unten die Physiklehrkräfte der 2. Studie

Modell		1	2	3	4	5	6
f		435	437	437	440	440	444
Physik- lehrkräfte	dev	135313	134261	134315	133967	133823	<b>133536</b>
	dev/ f	311	307	307	304	304	301
	#	100	96	92	94	83	91
	BIC	137441	136400	136455	136122	135976	<b>135706</b>
	CAIC	136812	135768	135822	135485	135340	<b>135065</b>
Physik- lehrkräfte, 2. Studie	dev	130197	129307	129368	129108	128861	<b>128686</b>
	dev/ f	299	296	296	293	293	290
	#	90	90	95	94	89	97
	BIC	132173	131290	131352	131108	130858	<b>130701</b>
	CAIC	131621	130735	130797	130549	130300	<b>130138</b>

Die Final Deviance der Modelle mit verschiedenen Stichproben kann nicht verglichen werden.

f Freiheitsgrade

dev Final Deviance

# Anzahl der Items mit Infit (weighted fit) MnSq>2.0

### 7.2.3 ITEMFIT UND SICHERSTELLUNG DER DIMENSIONSMODELLIERUNG

Da das Modell, in dem jeder Teil des Professionswissens eine eigene Dimension darstellt, am vielversprechendsten erscheint, werden in jeder der vier Dimensionen (CK, PCK, PK<sub>KP</sub>, PK<sub>D</sub>) störende Items entfernt ( $MnSq < 2.0$ ), wie in Kapitel 6.2 beschrieben wurde. Dies führt dazu, dass 3 von 14 CK-Items, 1 von 15 PCK-Items und 28 von 45 Items im PK<sub>D</sub>-Test entfernt werden müssen. Welche Items entfernt wurden, ist im Anhang aufgeführt (Tabelle 40). Es stellt sich heraus, dass der PK<sub>KP</sub>-Test mit der untersuchten Stichprobe nicht Rasch-modellierbar ist, da bei der Entfernung der Items mit einer schlechten Passung keine genügend große Anzahl übrig bleibt, die für eine weitere Auswertung herangezogen werden kann. Nur vier der Paarvergleiche ergeben eine Skala mit einem  $MnSq < 2.0$ . Daher wird im weiteren Verlauf für den Bereich PK nur noch der Testteil PK<sub>D</sub> betrachtet. Dies führt zu einer Reduzierung der Modelle, die verglichen werden können; diese sind in Tabelle 16 dargestellt.

**Tabelle 16:** Modelle, die ohne PK<sub>KP</sub> verglichen werden

Modell	1	2	3	4
Anzahl Dimensionen	1	2	2	3
Dimension 1	CK & PCK & PK <sub>D</sub>	CK & PCK	CK	CK
Dimension 2	-	PK <sub>D</sub>	PCK & PK <sub>D</sub>	PCK
Dimension 3	-	-	-	PK <sub>D</sub>

Auf Basis der Items, die gut zu dem dreidimensionalen Modell passen, wird die endgültige Modellpassung untersucht. Sie ist in Tabelle 17 dargestellt. In der ersten Studie war der PK-Test nicht integriert. Daher werden alle Ergebnisse zusätzlich ausschließlich für die zweite Studie dargestellt. In diesem Fall muss der DIF zwischen den Studien nicht modelliert werden. Im weiteren Verlauf werden die beiden Studien zusammengefasst. Das Ergebnis, dass das Modell mit den drei Dimensionen CK, PCK und PK<sub>D</sub> signifikant ( $p < .001$ ), am besten passt, ist unabhängig von der Beachtung des DIF zwischen den Studien, weshalb die Ergebnisse, wie im vorhergehenden Abschnitt besprochen, mit der Modellierung des DIF dargestellt sind.

In der dreidimensionalen Analyse mit allen Physiklehrkräften haben nur zwei der beibehaltenen Items einen  $MnSq$  größer als 2.0. Er liegt für keines der beiden Items über 2.1. Es kann also davon ausgegangen werden, dass sie die Messung kaum stören.

In Tabelle 17 wird zudem die EAP/PV-Reliabilität angegeben, um eine obere Grenze zu definieren (siehe 6.3).

## Ergebnisse zur Modellierung

**Tabelle 17:** Freiheitsgrade, Final Deviance und Reliabilität der vier möglichen Modelle (Siehe Tabelle 16) unter Ausschluss nicht passender Items. Oben: Alle Physiklehrkräfte mit Modellierung des DIF zwischen den Studien. Unten: Physiklehrkräfte der 2. Studie

Modell		1	2	3	4
Physiklehrkräfte, 1. & 2. Studie	f	126	128	128	131
	dev	25388	24950	25438	<b>24629</b>
	dev/ f	201	195	199	188
	BIC	26004	25576	26064	<b>25270</b>
	CAIC	25822	25391	25879	<b>25080</b>
	Rel Dimension 1	.87	.87	.81	.86
	Rel Dimension 2		.57	.73	.80
	Rel Dimension 3				.58
Physiklehrkräfte, 2. Studie	f	67	69	69	72
	dev	21433	21002	21441	<b>20786</b>
	dev/ f	320	304	311	289
	BIC	21737	21315	21754	<b>21113</b>
	CAIC	21652	21228	21667	<b>21021</b>
	Rel Dimension 1	.89	.88	.80	.86
	Rel Dimension 2		.79	.79	.82
	Rel Dimension 3				.80

f Freiheitsgrade  
 dev Final Deviance  
 Rel obere Grenze für die EAP/PV-Reliabilität

### 7.2.4 RELIABILITÄT IM BESTEN MODELL

In diesem Abschnitt wird näher auf die Reliabilität in dem verwendeten Model eingegangen. Die EAP/PV-Reliabilität ist generell mit Cronbachs Alpha vergleichbar (Rost, 2004). Sie wurde in den einzelnen Dimensionen CK, PCK und PK<sub>D</sub> berechnet. Wie im vorangegangenen Abschnitt beschrieben, lässt sich der PK<sub>KP</sub>-Test nicht Rasch-skalieren und wird daher nicht weiter analysiert.

Die EAP/PV-Reliabilität wurde mit drei verschiedenen Personengruppen berechnet (siehe Tabelle 18). Die kleinste Gruppe (G<sub>1</sub>, N=186) beinhaltet alle Physiklehrkräfte der zweiten Studie, also ausschließlich Physiklehrkräfte, die CK-, PCK- und PK-Test vorgelegt bekamen. Die zweitgrößte Gruppe (G<sub>2</sub>, N=279) umfasst zusätzlich die Physiklehrkräfte der ersten Studie, in der nur CK und PCK getestet wurde. Die dritte Gruppe enthält alle Versuchspersonen (G<sub>3</sub>, N=518). Alle Versuchspersonen aus G<sub>3</sub> haben den CK- und PCK-Test bearbeitet, aber nur zum Teil den PK-Test. Die Reliabilität für CK- und PCK-Test ändert sich kaum, unabhängig davon, welche Gruppe betrachtet wird. Das lässt darauf schließen, dass die Tests auch für Studierende und Lehrkräfte im Vorbereitungsdienst reliabel sind.

Die Reliabilität des PCK-Tests (siehe Tabelle 18) ist gut (Rost, 2004).

**Tabelle 18:** Reliabilität des CK-, PCK- und PK<sub>D</sub>-Tests in verschiedenen Personengruppen

		EAP/PV-Reliabilität		
		G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>
CK, gesamt		.86	.85	.84
CK, offene Items		.82	.83	.82
CK, geschlossene Items		.56	.53	.52
PCK		.77	.78	.78
PK <sub>D</sub>		.71	.47	.38
G <sub>1</sub>	Physiklehrkräfte der zweiten Studie			
G <sub>2</sub>	Alle Physiklehrkräfte			
G <sub>3</sub>	Gesamte Stichprobe			

## Ergebnisse zur Modellierung

Der Wert für die Berechnung der Reliabilität des gesamten CK-Tests mit Hilfe eines Dummy-Raters liegt nicht außerhalb der berechneten oberen und unteren Grenze und wird damit als Wert für die Reliabilität des CK-Tests angesehen. Dieser Wert ist ebenfalls gut. Als untere Grenzen fungieren die Reliabilität der offenen CK-Items, die ebenfalls gut ist, und die der geschlossenen CK-Items. Die geschlossenen Items sollten aufgrund ihrer geringen Reliabilität (siehe Tabelle 18) nicht als einzelne Skala ausgewertet werden. Für alle weiteren Rechnungen werden nur die Ergebnisse des gesamten CK-Tests genutzt.

Die großen Unterschiede in der Reliabilität des PK-Tests (siehe Tabelle 18) lassen sich darauf zurückführen, dass viele Versuchspersonen in den beiden größeren Gruppen  $G_2$  und  $G_3$  den PK-Test nicht bearbeiten konnten, da dieser nur in der zweiten Studie Verwendung fand. Als Reliabilität für den PK-Test wird daher der in Gruppe  $G_1$  berechnete Wert .71 betrachtet (Linacre, 2011b). Da der  $PK_D$ -Test nur aus Multiple-Choice Items besteht, also kein Ratereffekt zu beachten ist, kann für die Physiklehrkräfte der zweiten Studie auch das klassische Reliabilitätsmaß, Cronbachs Alpha, angegeben werden. Es liegt bei .69, was akzeptabel ist. Werden fehlende Werte mit Null Punkten kodiert, liegt Cronbachs Alpha ( $PK_D$ ) bei .75.

Es lässt sich zusammenfassen, dass sich die drei Dimensionen CK, PCK und  $PK_D$  reliabel modellieren und empirisch voneinander trennen lassen. Allerdings bleibt zu bedenken, dass relativ viele Items aus der Analyse entfernt werden mussten. Nichtsdestotrotz kann im folgenden Abschnitt die Validität der verbleibenden Skalen untersucht werden, da die Reliabilität der Skalen hinreichend ist. Das Vorgehen zur Berechnung der Personenfähigkeiten, die für die weitere Auswertung genutzt werden, wurde genauso umgesetzt wie in Abschnitt 6.2.1 erläutert.

### 7.3 ERGEBNISSE ZUR VALIDITÄT

Da die Reliabilität der einzelnen Dimensionen CK, PCK und PK<sub>D</sub> akzeptabel bis gut ist, können aus den Personenfähigkeiten Schlüsse über die Validität der Tests gezogen werden. Die Validität soll auf verschiedenen Ebenen sichergestellt werden. Dafür werden Inhalts-, Konstrukt-, Kriteriums- und externe Validität beleuchtet.

Bisher ist die Validität des gesamten Konstrukts Professionswissen nicht geklärt. Für eine Validierung und zur Klärung der Relevanz des Konstrukts für Unterricht muss erforscht werden, welche Einflüsse CK, PCK und PK auf das Handeln der Lehrkräfte sowie auf Schülerleistung und –motivation haben. Ist der Zusammenhang gering, stellt dies die Validität des Konstrukts in Frage.

Durch den Abgleich mit Curricula und Fachliteratur, Expertenbefragungen und der Testentwicklung anhand des Modells wird die **Inhaltsvalidität** gesichert. Diese steht dafür, dass die Tests den Inhalt messen, den sie messen sollen.

Ergänzend belegen Kriteriums- und Konstruktvalidität die Validität indirekt (Bühner, 2006). Die Kriteriumsvalidität wird dadurch überprüft, dass der Zusammenhang mit anderen Kriterien, wie anderen Ausbildungsgängen oder dem unterrichteten Schulfach, analysiert wird. Im Rahmen der Prüfung der Konstruktvalidität wird untersucht, ob die Testleistung mit der Leistung in anderen Tests, also anderen Konstrukten, zusammenhängt. Dazu werden in dieser Studie die Ergebnisse der CK-, PCK- und PK-Tests betrachtet.

Um die **Kriteriumsvalidität** zu untersuchen, werden Physiklehrkräfte der verschiedenen Schulformen (Tabelle 19), Lehrkräfte verschiedener Fächer (Tabelle 20) und Diplom-Physikerinnen und -Physiker (Tabelle 21 und 22) untersucht. Es werden die folgenden Hypothesen geprüft:

HV1: Physiklehrkräfte, die am Gymnasium unterrichten, schneiden in CK und PCK besser ab als Physiklehrkräfte, die an anderen Schulformen unterrichten.

HV2: Physiklehrkräfte, die am Gymnasium unterrichten, schneiden in CK und PCK besser ab als Lehrkräfte anderer Fächer, die ebenfalls am Gymnasium unterrichten.

HV3: Physiklehrkräfte, die am Gymnasium unterrichten, schneiden im Bereich PK nicht besser oder schlechter ab als Lehrkräfte anderer Fächer, die ebenfalls am Gymnasium unterrichten.

## Ergebnisse zur Validität

HV4: Physiklehrkräfte, die am Gymnasium unterrichten, schneiden im Bereich CK nicht besser oder schlechter ab als Diplom-Physikerinnen und -Physiker, die nicht an einer Schule unterrichten.

HV5: Physiklehrkräfte, die am Gymnasium unterrichten, schneiden im Bereich PCK besser ab als Diplom-Physikerinnen und -Physiker, die nicht an einer Schule unterrichten.

HV6: Physiklehrkräfte, die am Gymnasium unterrichten, schneiden im Bereich PK besser ab als Diplom-Physikerinnen und -Physiker, die nicht an einer Schule unterrichten.

In Tabelle 19 sind die Unterschiede zwischen Physiklehrkräften der verschiedenen Schulformen dargestellt, in Tabelle 20 die Unterschiede zwischen Physiklehrkräften und Lehrkräften anderer Fächer. Physiklehrkräfte, die am Gymnasium unterrichten, schneiden in CK und PCK signifikant besser ab als Physiklehrkräfte, die an anderen Schulformen unterrichten (HV1) und als Lehrkräfte anderer Fächer, die ebenfalls am Gymnasium unterrichten (HV2). Sie unterscheiden sich im Bereich PK nicht signifikant von Lehrkräften anderer Fächer, die ebenfalls am Gymnasium unterrichten (HV3). Damit können HV1, HV2 und HV3 angenommen werden.

**Tabelle 19:** Testleistung der Physiklehrkräfte der verschiedenen Schulformen

	CK		PCK	
	NGY	GY	NGY	GY
N	62 <sup>1</sup>	216 <sup>1</sup>	62 <sup>1</sup>	216 <sup>1</sup>
M	463	557	425	533
SD	84	91	100	83
SE	11	6	13	6
t(276)	7.30		8.58	
p	<.001		<.001	
d	1.1		1.2	

<sup>1</sup> Normalverteilung angenommen, da N>30

GY Unterrichtet zurzeit am Gymnasium

NGY Unterrichtet zurzeit an einer anderen Schulform als dem Gymnasium

N Stichprobengröße

M Durchschnittliche Personenfähigkeit

SD Standardabweichung

SE Standardfehler

t t-Wert in Abhängigkeit der Freiheitsgrade

p Signifikanz der Unterschiede

d Effektstärke der Unterschiede

**Tabelle 20:** Testleistung der Physiklehrkräfte und der Lehrkräfte anderer Fächer. Alle einbezogenen Lehrkräfte unterrichten am Gymnasium.

	CK		PCK		PK	
	Lehrkräfte Physik	Lehrkräfte anderer Fächer	Lehrkräfte Physik	Lehrkräfte anderer Fächer	Lehrkräfte Physik	Lehrkräfte anderer Fächer
N	216 <sup>1</sup>	31 <sup>1</sup>	216 <sup>1</sup>	31 <sup>1</sup>	149 <sup>1</sup>	21 <sup>2</sup>
M	557	385	533	456	512	523
SD	91	69	83	79	84	49
SE	6	12	6	14	7	11
t	t(245)=10.07		t(245)=4.84		t(168)=-0.61	
p	<.001		<.001		.542	
d	2.1		0.9		-	

<sup>1</sup> Normalverteilung angenommen, da N>30

<sup>2</sup> Normalverteilung nach Prüfung angenommen (siehe Anhang, Abschnitt 10.6)

Konventionen siehe Tabelle 19

In Tabelle 21 ist das Wissen von Physiklehrkräften dem von Diplom-Physikerinnen und -Physikern gegenübergestellt. Aus Tabelle 21 geht hervor, dass sich die beiden Gruppen weder im CK (HV4) noch im PCK (HV5) signifikant unterscheiden. HV4 kann angenommen werden, HV5 muss verworfen werden.

Das PK der Diplom-Physikerinnen und -Physiker kann nur mit Hilfe einer sehr kleinen Stichprobe von sieben Versuchspersonen, die nicht unterrichten, betrachtet werden. Das Wissen dieser Subgruppe ist in Tabelle 22 dem PK der Physiklehrkräfte gegenübergestellt. Alle in Tabelle 22 dargestellten Ergebnisse können wegen des geringen Stichprobenumfangs nur als erster Eindruck betrachtet werden. H6 muss verworfen werden, da im PK kein signifikanter Unterschied zwischen den Gruppen zu erkennen ist. Im Gegensatz zur Gesamtstichprobe kann HV5 in dieser Substichprobe angenommen werden, Physiklehrkräfte schneiden im Mittel im PCK besser ab als die getesteten Diplom-Physikerinnen und -Physiker.

**Tabelle 21:** Testleistung der Physiklehrkräfte am Gymnasium und der Diplom-Physikerinnen und -Physiker

	CK		PCK	
	GY-Lehrkräfte Physik	Dipl. Phys.	GY-Lehrkräfte Physik	Dipl. Phys.
N	216 <sup>1</sup>	22 <sup>2</sup>	216 <sup>1</sup>	22 <sup>2</sup>
M	557	526	533	517
SD	91	63	83	77
SE	6	13	6	16
t(236)	1.55		0.86	
p	.122		.394	

<sup>1</sup> Normalverteilung angenommen, da N>30

<sup>2</sup> Normalverteilung nach Prüfung angenommen (siehe Anhang)

Dipl. Phys. Diplom-Physikerinnen und -Physiker

Weitere Konventionen siehe Tabelle 19

**Tabelle 22:** Testleistung der Physiklehrkräfte am Gymnasium und der Diplom-Physikerinnen und -Physiker, die weder an einer Universität tätig sind noch lehren

	CK		PCK		PK	
	GY-Lehrkräfte	Dipl. Phys.	GY-Lehrkräfte	Dipl. Phys.	GY-Lehrkräfte	Dipl. Phys.
N	216	7	216	7	149	7
M	557	537	533	461	512	509
SD	91	73	83	48	84	83
U	637		334		462	
p	.488		.010		.615	

Eine Normalverteilung kann nicht angenommen werden.

Dipl. Phys. Diplom-Physikerinnen und -Physiker

U U-Wert

Weitere Konventionen siehe Tabelle 19

Es kann von einer vorwiegend erfüllten Kriteriumsvalidität ausgegangen werden. Alle Aspekte, die den Schulformunterschied und das unterrichtete Schulfach betreffen, können uneingeschränkt bestätigt werden. Der Vergleich mit Diplom-Physikern, die nicht unterrichten, ergibt kein eindeutiges Bild.

Zur Untersuchung der **Konstruktvalidität** werden die folgenden Hypothesen geprüft:

HV7: Die Ergebnisse der Physiklehrkräfte im CK-Test und PCK-Test korrelieren miteinander.

HV8: Die Ergebnisse der Physiklehrkräfte im PK-Test und PCK-Test korrelieren miteinander.

HV9: Die Korrelation der Ergebnisse zwischen CK- und PCK-Test ist signifikant größer als die zwischen den Ergebnissen von CK- und PK-Test.

HV10: Die Korrelation der Ergebnisse zwischen PK- und PCK-Test ist signifikant größer, als die zwischen den Ergebnissen von CK- und PK-Test.

In Tabelle 23 sind die Korrelationen zwischen CK, PCK und PK<sub>D</sub> für alle untersuchten Physiklehrkräfte dargestellt. Sowohl CK als auch PK korrelieren mit PCK, damit können HV7 und HV8 angenommen werden.

**Tabelle 23:** Testleistungskorrelationen nach Pearson für CK, PCK (N=279) und PK (N=186)

	Korrelation r	Signifikanz p
CK zu PCK	.453	<.001
PCK zu PK <sub>D</sub>	.265	<.001
CK zu PK <sub>D</sub>	.174	<.05

HV9 kann angenommen werden, da die Korrelation der Ergebnisse zwischen CK- und PCK-Test signifikant größer ist als die zwischen den Ergebnissen von CK- und PK<sub>D</sub>-Test ( $p < .001$ , einseitig getestet). HV10 muss abgelehnt werden, da die Korrelation der Ergebnisse zwischen PK<sub>D</sub>- und PCK-Test nicht signifikant größer ist, als die zwischen den Ergebnissen von CK- und PK<sub>D</sub>-Test ( $p = .180$ , einseitig getestet).

Die diskriminante Konstruktvalidität kann überwiegend bestätigt werden. Alle drei Dimensionen korrelieren miteinander. Der PK<sub>D</sub>-Test korreliert weniger stark mit den anderen beiden Testteilen als CK und PCK miteinander.

Für eine hohe externe Validität der Tests, die die Generalisierbarkeit widerspiegelt, spricht, dass ein den Lehrkräften bekanntes Format gewählt wurde und die Ergebnisse auf andere Physiklehrkräfte, Studierende und Lehrkräfte im Vorbereitungsdienst übertragen werden können (Bortz & Döring, 2006). Eingeschränkt wird sie möglicherweise durch die anzunehmende Positivauslese und die Zahlung einer Aufwandsentschädigung.

#### **7.4 ZUSAMMENFASSUNG TESTGÜTE**

Alle später kodierten Items erreichen eine zufriedenstellende Interraterübereinstimmung: Goodmans und Kruskal Gamma  $>.7$  sowie ICC<sub>2,2</sub> (unjustiert)  $>.7$ . Der PK<sub>KP</sub>-Test ist nicht Rasch-konform. Wie erwartet stellen die Bereiche CK, PCK und PK getrennte Dimensionen dar. Das dreidimensionale Modell passt signifikant ( $p < .001$ ) besser als mögliche ein- oder zweidimensionale Modelle. Für fast alle Items, die in der anschließenden Analyse berücksichtigt werden, gilt MnSq  $< 2.0$ . Die Reliabilität ist zufriedenstellend, die EAP/PV-Reliabilität liegt in der Gruppe der Physiklehrkräfte der zweiten Studie, die den CK-, PCK- und PK-Test ausgefüllt haben, bei .86 für den CK-Test, bei .77 für den PCK-Test und .71 für den in den folgenden Analysen verwendeten deklarativen Teil des PK-Tests.

Die Inhaltsvalidität der Instrumente kann als gesichert angenommen werden. Die Kriteriumsvalidität und die diskriminante Konstruktvalidität wurden überwiegend bestätigt.

Da alle drei Bereiche reliabel sind und als valide betrachtet werden können, kann die Personenfähigkeit für weitere Analysen verwendet werden.

#### **7.5 ERGEBNISSE ZU DEN PERSONENMERKMALEN**

Nachdem in den letzten Abschnitten belegt wurde, dass CK, PCK und PK drei getrennte Dimensionen sind und alle drei Dimensionen reliabel und valide erfasst wurden, wird nun der Zusammenhang zwischen CK, PCK und PK detaillierter betrachtet sowie die Korrelation der einzelnen unabhängigen Variablen mit CK und PCK dargestellt. Daraufhin wird der Zusammenhang ausgewählter unabhängiger Variablen mit CK und PCK mit Regressionen analysiert. Im weiteren Verlauf werden die CK- und PCK-Testergebnisse von Studierenden des Physiklehramts, Physiklehrkräften im Vorbereitungsdienst und Physiklehrkräften verglichen sowie der Zusammenhang zwischen den Testergebnissen für die Gruppen gegenübergestellt.

### 7.5.1 ZUSAMMENHANG VON CK, PCK UND PK

Zum Zweck der Validierung wurden die Korrelationen zwischen CK, PCK und PK bereits in Abschnitt 7.3 für alle Physiklehrkräfte behandelt. In Abbildung 16 ist der Zusammenhang von CK, PK und PCK grafisch dargestellt. Es ist zu erkennen, dass PCK enger mit CK als mit PK zusammenhängt. In der Abbildung wurde zusätzlich nach der Schulform getrennt, an der aktuell unterrichtet wird, da sich das Wissen in diesen Gruppen deutlich unterscheidet, wie in Abschnitt 7.3 im Rahmen der Validierung dargestellt wurde. In Tabelle 24 sind die Korrelationen zwischen CK, PCK und PK für alle Physiklehrkräfte und zusätzlich aufgeteilt nach der aktuellen Schulform zu finden. Wie auch in Abbildung 16 zu erkennen ist, ist die Korrelation zwischen CK und PCK für alle Physiklehrkräfte höher als für die beiden Subgruppen. Da der PK-Test nur in der zweiten Studie eingesetzt wurde und der Großteil der Lehrkräfte am Gymnasium unterrichtet, haben nur 36 Physiklehrkräfte, die an anderen Schulformen als dem Gymnasium unterrichten, den PK-Test ausgefüllt. Für die Korrelationen von PK zu CK und PK zu PCK ist festzustellen, dass sie für diese Gruppe nicht signifikant sind. Die Abbildungen mit den entsprechenden Regressionsgraden sind im Anhang (Abschnitt 10.3) zu finden.

**Tabelle 24:** Testleistungskorrelationen nach Pearson für Physiklehrkräfte. Die Korrelationen zwischen GY und NGY unterscheiden sich nicht signifikant ( $p > .2$ ).

		N	Korrelation r	Signifikanz p
CK zu PCK	Gesamt	279	.453	<.001
	GY	216	.331	<.001
	NGY	62	.316	<.05
PK <sub>D</sub> zu PCK	Gesamt	186	.265	<.001
	GY	149	.227	<.01
	NGY	36	.173	.314
CK zu PK <sub>D</sub>	Gesamt	186	.174	<.05
	GY	149	.151	.066
	NGY	36	-.073	.672

GY Unterrichtet zurzeit am Gymnasium

NGY Unterrichtet zurzeit an einer anderen Schulform als dem Gymnasium

## Ergebnisse zu den Personenmerkmalen

In Tabelle 25 ist der Einfluss von CK und PK auf PCK in Form einer Regression für die Physiklehrkräfte dargestellt. An den standardisierten Beta-Koeffizienten ist zu erkennen, dass ein Unterschied von einer Standardabweichung im CK, unter Kontrolle von PK, zu 0.46 Standardabweichungen im PCK führt. Ein Unterschied über eine Standardabweichung im PK führt dagegen nur zu 0.19 Standardabweichungen im PCK. In allen drei Dimensionen liegt eine Standardabweichung für die gesamte Stichprobe bei 100 Einheiten. Ein Unterschied von einer Einheit im CK führt zu einem Unterschied von 0.40 Einheiten im PCK, ein Unterschied einer Einheit im PK hängt mit einem Unterschied von 0.18 Einheiten im PCK zusammen. Daran ist zu erkennen, dass der Einfluss von CK auf PCK mehr als doppelt so groß ist wie der von PK auf PCK.

In Tabelle 26 ist der Einfluss von CK und PK auf PCK für Physiklehrkräfte des Gymnasiums dargestellt. Die Ergebnisse sind vergleichbar mit denen aller Physiklehrkräfte. Werden nur die 36 Physiklehrkräfte an anderen Schulformen betrachtet, wird die Regression nicht signifikant ( $R^2=.04$ ,  $F=0.7$ ,  $p=.509$ ) und kann daher nicht interpretiert werden.

**Tabelle 25:** Lineare Regression für Physiklehrkräfte (N=279): Einfluss von CK und PK auf PCK,  $R^2=.27$ ,  $F=34$

Abhängige Variable	Unabhängige Variable	Beta, unstandardisiert	SE	Beta, standardisiert	p
PCK	CK	0.40	0.06	0.46	<.001
	PK	0.18	0.06	0.19	<.01

SE Standardfehler

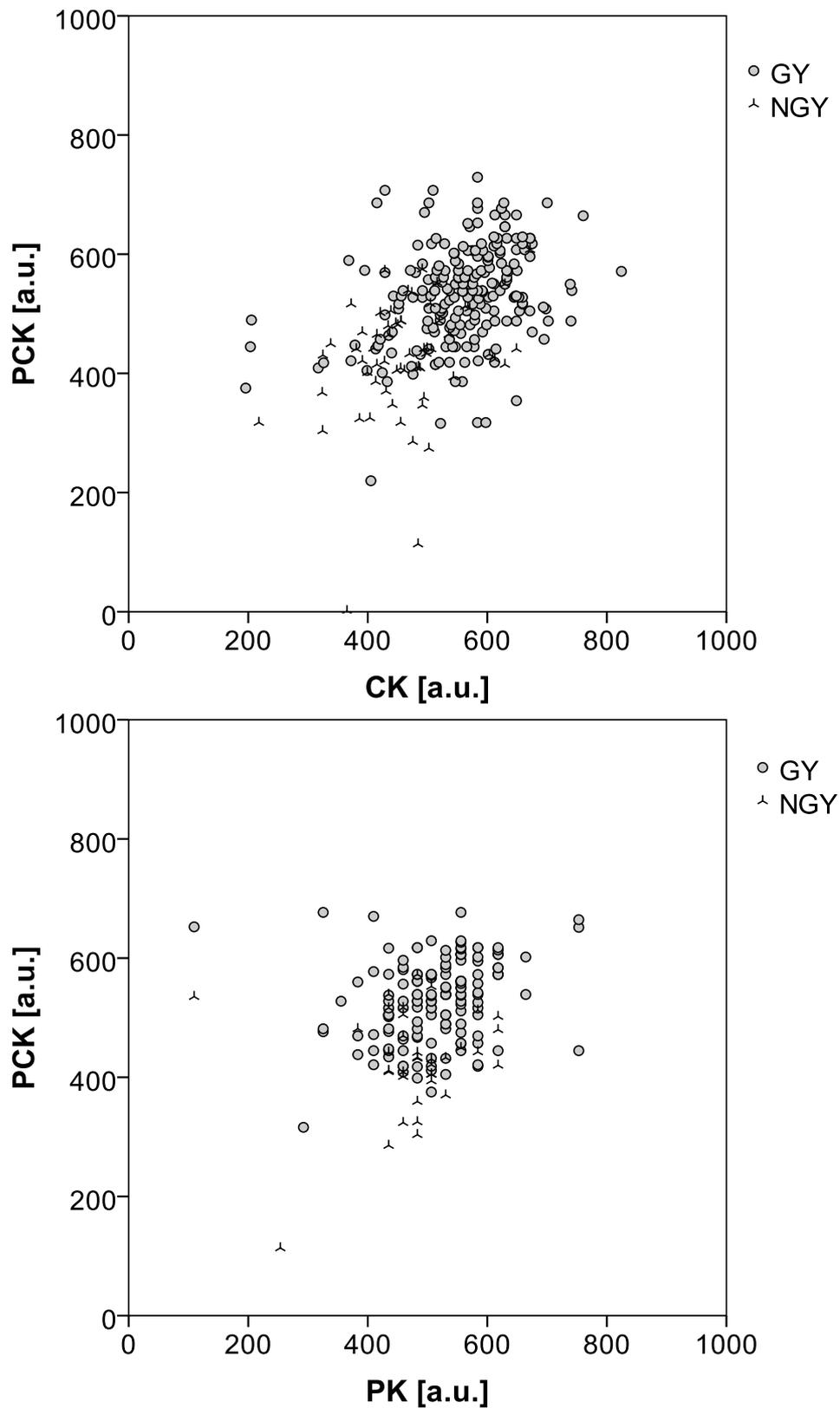
Identische Werte für backwards und forced entry Regression

**Tabelle 26:** Lineare Regression für Physiklehrkräfte, die am Gymnasium unterrichten (N=149): Einfluss von CK und PK auf PCK,  $R^2=.20$ ,  $F=18$

Abhängige Variable	Unabhängige Variable	Beta, unstandardisiert	SE	Beta, standardisiert	p
PCK	CK	0.30	0.06	0.40	<.001
	PK	0.14	0.06	0.17	<.05

SE Standardfehler

Forced entry Regression



**Abbildung 16:** Zusammenhang zwischen CK und PCK bzw. PK und PCK für alle untersuchten Physiklehrkräfte, getrennt nach der Schulform, an der aktuell unterrichtet wird. Für die gesamte Stichprobe gilt  $M=500$ ,  $SD=100$  für CK, PCK und PK.

### **7.5.2 KORRELATIONEN ZWISCHEN DEN UNABHÄNGIGEN VARIABLEN UND CK SOWIE PCK**

In Abbildung 17 wird dargestellt, wie die unabhängigen Variablen jeweils mit CK und PCK korrelieren.

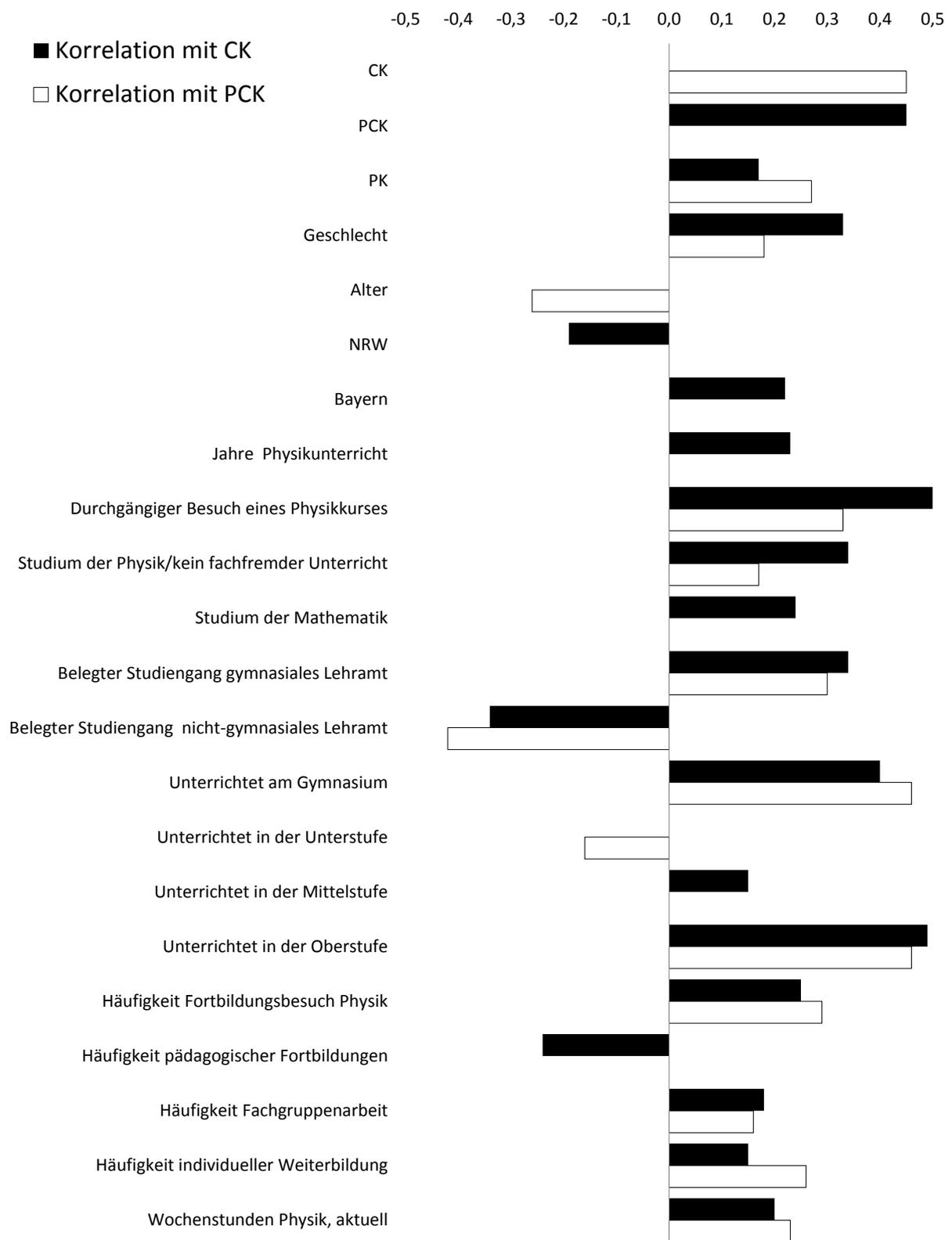
Sowohl mit CK als auch mit PCK korrelieren das Geschlecht (Lehrer schneiden im Mittel besser ab), der fachfremde Unterricht (wer Physik studiert hat, schneidet besser ab als diejenigen, die fachfremd unterrichten), der belegte Studiengang (ein Studium des gymnasialen Lehramts hängt mit besseren Ergebnissen zusammen, ein Studium des nicht-gymnasialen Lehramts mit weniger CK und PCK), ebenso wie der Unterricht am Gymnasium mit besseren Ergebnissen zusammenhängt. Einen positiven Einfluss haben der Besuch von Fortbildungen zur Physik, individuelle Weiterbildung und Fachgruppenarbeit, die Anzahl an Stunden pro Woche, in denen Physik unterrichtet wird, der Besuch eines Physikkurses in der eigenen Oberstufenzeit und der physikbezogene Unterricht an der Oberstufe.

Signifikant mit CK, aber nicht mit PCK, korrelieren das Studium der Mathematik, welches zu einem besseren Abschneiden führt, das Bundesland, in dem unterrichtet wird (Lehrkräfte in NRW schneiden im Mittel schlechter ab als Lehrkräfte in anderen Bundesländern, Lehrkräfte in Bayern besser), die Häufigkeit des Besuchs pädagogischer Fortbildungen, deren seltener Besuch mit besserem Abschneiden zusammenhängt, die Anzahl an Jahren, in denen Physik unterrichtet wurde (eine längere Zeit hängt mit einem höheren CK zusammen) und der Unterricht in der Mittelstufe, der ebenfalls mit höherem CK zusammenhängt.

Mit PCK, nicht aber CK, korrelieren der Unterricht in der Unterstufe, der mit weniger PCK zusammenhängt und das Alter, jüngere Lehrkräfte schneiden besser ab.

Weder einen signifikanten Zusammenhang zu CK noch PCK zeigen die Jahre im Schuldienst und ob ein Diplom-Studiengang statt eines Lehramtsstudiums belegt wurde.

Alle Korrelationen sind im Anhang tabellarisch dargestellt (Abschnitt 10.4).



**Abbildung 17:** Korrelationen der unabhängigen Variablen mit CK und PCK  
 Für alle eingezeichneten Korrelationen gilt  $p < .05$ . Zum Vergleich sind auch die Korrelationen der Dimensionen untereinander aufgetragen.

### 7.5.3 REGRESSION

Um zu untersuchen, welchen Einfluss die Variablen zur aktuellen Situation von Physiklehrkräften haben, wenn sie gemeinsam betrachtet werden, wird eine Regressionsanalyse durchgeführt. Wie in Abschnitt 6.3 zur Auswertung der Daten dargestellt, werden backwards-Regressionen gerechnet. Wie in Abschnitt 4.3 dargestellt, werden die folgenden Variablen untersucht:

- Unterrichtet am Gymnasium
- Fachfremder Unterricht (kein Studium der Physik)
- Häufigkeit Fortbildungsbesuch Physik
- Häufigkeit pädagogischer Fortbildungen
- Häufigkeit individueller Weiterbildung
- Häufigkeit Fachgruppenarbeit
- Wochenstunden Physik
- Alter und Geschlecht
- CK und PK für die Varianzaufklärung von PCK

Wie in Abschnitt 2.4 beschrieben, wird nur der Einfluss von CK und PK auf PCK betrachtet, nicht der von PCK und PK auf CK. Der Vollständigkeit halber sind die beiden Regressionen zur Aufklärung von CK mit den beiden anderen Wissensdimensionen PCK und PK sowie zur Aufklärung von PCK ohne die beiden anderen Wissensdimensionen CK und PK im Anhang angeführt (Abschnitt 10.5).

In Tabelle 27 und Tabelle 28 ist dargestellt, welche Variablen einen signifikanten Einfluss auf CK bzw. PCK haben und wie groß dieser ist. Mit fünf unabhängigen Variablen können 40% der Varianz des CK aufgeklärt werden. Den größten Einfluss hat der Schulzweig, an dem unterrichtet wird. Lehrkräfte, die am Gymnasium unterrichten, schneiden besser ab. Ebenso schneidet besser ab, wer im Lehramtsstudium oder im Diplomstudiengang Physik studiert hat, also nicht fachfremd unterrichtet. Physiklehrer schneiden im Mittel besser ab als Physiklehrerinnen. Einen Einfluss hat auch der Besuch allgemein pädagogischer Fortbildungen. Das CK sinkt mit der Häufigkeit ihres Besuchs. Den geringsten signifikanten Einfluss hat die Häufigkeit von Fachgruppenarbeit, wobei eine höhere Häufigkeit mit besseren Ergebnissen zusammenhängt.

Den größten Einfluss auf PCK hat CK. Da die beiden mit  $r=.453$ ,  $p<.001$ , korrelieren, klärt schon CK allein 21% von PCK auf. Den nächststärksten Einfluss hat der Schulzweig, an dem unterrichtet wird. Lehrkräfte, die am Gymnasium unterrichten, schneiden besser ab. Jüngere Lehrkräfte schneiden besser ab als ältere. Wer sich häufiger individuell weiterbildet, z. B. durch die Lektüre von Fachzeitschriften, schneidet besser ab. Insgesamt können 40% der Varianz des PCK durch diese vier Variablen aufgeklärt werden.

**Tabelle 27:** Lineare Regression: Einfluss demographischer Variablen auf CK (N=279),  $R^2_{CK}=.40$ ,  $F_{CK}=23$

Abhängige Variable	Unabhängige Variable	Beta, unstandardisiert	SE	Beta, standardisiert	Signifikanz p
CK	Unterrichtet am Gymnasium	84	15	0.36	<.001
	Geschlecht	53	12	0.26	<.001
	Unterrichtet fachfremd	-57	13	0.26	<.001
	Häufigkeit pädagogischer Fortbildungen	-13	6	-0.14	<.05
	Häufigkeit Fachgruppenarbeit	12	6	0.13	<.05

Backwards Regression

**Tabelle 28:** Lineare Regression: Einfluss demographischer Variablen auf PCK (N=279),  $R^2_{PCK}=.40$ ,  $F_{PCK}=30$

Abhängige Variable	Unabhängige Variable	Beta, unstandardisiert	SE	Beta, standardisiert	Signifikanz p
PCK	CK	0.3	0.6	0.33	<.001
	Unterrichtet am Gymnasium	63	14	0.30	<.001
	Alter	-2	0.5	-0.23	<.001
	Häufigkeit individueller Weiterbildung	14	5	0.17	<.01

Backwards Regression

## Ergebnisse zu den Personenmerkmalen

Für die Kreuzvalidierung werden mit einer linearen Regression die Beta-Koeffizienten mit 80% der Stichprobe bestimmt und festgelegt. Mit diesen wird das Wissen der verbleibenden Versuchspersonen geschätzt (siehe Abschnitt 6.3). Die Schätzung führt zu einer vergleichbar hohen Aufklärung des CK (47%) und PCK (41%) wie die Regression in der ersten Stichprobe (CK: 42%, PCK: 40%). Damit kann die Regression als valide angesehen werden.

Allerdings führen mehrere Regressionen mit 100%, 80% und 50% der Stichprobe der Physiklehrkräfte (jeweils eine Zufallsauswahl) dazu, dass weder die Prädiktoren (unabhängige Variablen mit einem signifikanten Einfluss auf CK bzw. PCK) noch die Anzahl der Prädiktoren noch die Größe ihres Einflusses konstant bleiben.

Dies führt zu einem anderen Verfahren, um die Generalisierbarkeit der Regressionen sicherzustellen. Um belastbarere Ergebnisse zu erzeugen, werden vier weitere Regressionen (backwards) mit denselben unabhängigen Variablen durchgeführt. Dazu wird die Stichprobe zufällig halbiert und in jeder der beiden Hälften wird unabhängig voneinander eine Regression durchgeführt. Die Ergebnisse werden miteinander verglichen. Zur Absicherung wird die Gesamtstichprobe erneut zufällig halbiert und es werden zwei weitere Regressionen in den Teilstichproben durchgeführt. Nur unabhängige Variablen, die in allen vier Regressionen einen signifikanten Einfluss auf CK zeigen, werden weiterhin betrachtet. Da ich nach den vorhergehenden Überlegungen davon ausgehe, dass diese Variablen relevant sind, werden sie in eine Regression mit der gesamten Stichprobe gezwungen (forced entry) (Tabachnick & Fidell, 2001). Dasselbe Verfahren wird für die Dimension PCK durchgeführt.

Auf **CK** haben die Variablen **Geschlecht**, **fachfremder Unterricht** und **Unterricht am Gymnasium** einen signifikanten Einfluss. Diese klären 35% der Varianz von CK auf. Der Einfluss der einzelnen Variablen wird in Tabelle 29 berichtet. Auf **PCK** haben die Variablen **CK**, **Alter** und **Unterricht am Gymnasium** einen signifikanten Einfluss. Diese klären 34% der Varianz von PCK auf. Über den Einfluss der einzelnen Variablen wird in Tabelle 30 berichtet.

Ich nehme an, dass die dargestellten Variablen auch in anderen Stichproben einen Einfluss auf CK bzw. PCK haben könnten, da ihr signifikanter Einfluss in vier Stichproben, die allerdings nicht voneinander unabhängig sind, mit jeweils 140 Versuchspersonen gezeigt wurde (Tabachnick & Fidell, 2001) und sie auch in der ersten dargestellten Regression (Tabellen 27 und 28) einen signifikanten Einfluss zeigen .

**Tabelle 29:** Lineare Regression: Einfluss demographischer Variablen auf CK (N=279),  $R^2_{CK} = .35$ ,  $F_{CK} = 47$

Abhän- gige Va- riable	Unabhängige Variable	Beta, unstan- dardisiert	SE	Beta, stan- dardisiert	Signifikanz
CK	Unterrichtet am Gymnasium	89	12	0.38	<.001
	Geschlecht	74	11	0.34	<.001
	Unterrichtet fachfremd	58	13	0.23	<.001

Forced entry Regression

**Tabelle 30:** Lineare Regression: Einfluss demographischer Variablen auf PCK (N=279),  $R^2_{PCK} = .34$ ,  $F_{PCK} = 44$

Abhän- gige Va- riable	Unabhängige Variable	Beta, unstan- dardisiert	SE	Beta, stan- dardisiert	Signifikanz
PCK	CK	0.3	0.1	0.34	<.001
	Unterrichtet am Gymnasium	70	13	0.30	<.001
	Alter	-2	0.5	-0.23	<.001

Forced entry Regression

#### **7.5.4 AUSBILDUNGSSCHRITTE**

Es werden drei Gruppen untersucht: Lehrkräfte, Lehrkräfte im Vorbereitungsdienst, und Studierende. Zuerst werden Mittelwertunterschiede beleuchtet, danach Unterschiede in den Korrelationen zwischen CK und PCK (siehe Abschnitt 2.5). Da Studierende und Lehrkräfte im Vorbereitungsdienst vorwiegend in NRW getestet wurden, werden ausschließlich Personen aus NRW betrachtet, um die Homogenität der Stichprobe zu erhöhen.

##### **7.5.4.1 MITTELWERTUNTERSCHIEDE ZWISCHEN DEN AUSBILDUNGSSTUFEN**

Im Folgenden wird untersucht, wie sich das fachspezifische Professionswissen zwischen Lehramtsstudierenden der Physik, Lehrkräften der Physik im Vorbereitungsdienst und Lehrkräften der Physik unterscheidet. Da sich zwischen den Lehrkräften der verschiedenen Schulformen erhebliche Unterschiede in CK und PCK zeigen, wird im weiteren Verlauf unterschieden, welcher Studiengang absolviert wurde: ein gymnasiales Lehramtsstudium, ein nicht-gymnasiales Lehramtsstudium oder ein Diplomstudium.

Tabelle 31 zeigt, wie groß die einzelnen Gruppen sind und ob CK und PCK der Personen in den Gruppen normalverteilt ist. In Tabelle 32 und Abbildung 18 wird das mittlere CK und PCK dargestellt. In Tabelle 32 werden zusätzlich zum Mittelwert Standardabweichung und Standardfehler angegeben.

Wie im Abschnitt 6.3 beschrieben, werden ANOVAs und t-Tests gerechnet, wenn das Wissen aller beteiligten Gruppen normalverteilt ist. Wenn das Wissen mindestens einer der Gruppen nicht normal verteilt ist, werden Kruskal-Wallis-Tests und Mann-Whitney-Tests verwendet, um Mittelwertunterschiede aufzuzeigen.

Nur die Hälfte der Physiklehrkräfte, die einen Diplomstudiengang absolviert haben, hat einen Abschluss in Physik, die andere Hälfte in einem anderen Fach (siehe 5.3). Daher wird im Anhang ergänzend das Wissen der beiden Subgruppen gegenübergestellt (Abschnitt 10.7).

**Tabelle 31:** Anzahl der Personen in den Gruppen

	Probandenzahl N			Normalverteilung des Wissen		
	GY studiert	NGY studiert	Diplom studiert	GY studiert	NGY studiert	Diplom studiert
Physiklehrkräfte	71	18	32	ja <sup>1</sup>	CK ja <sup>2</sup> PCK nein <sup>2</sup>	ja <sup>1</sup>
Lehrkräfte im Vor- bereitungsdienst	37	10	12	ja <sup>1</sup>	ja <sup>2</sup>	ja <sup>2</sup>
Lehramtsstudie- rende Physik	43	43	-	ja <sup>1</sup>	ja <sup>1</sup>	-

GY Gymnasiallehramt

NGY Nicht-gymnasiales Lehramt

<sup>1</sup> da N>30<sup>2</sup> Normalverteilung nach Prüfung angenommen (siehe Anhang)**Tabelle 32:** CK und PCK der Versuchspersonen aus NRW

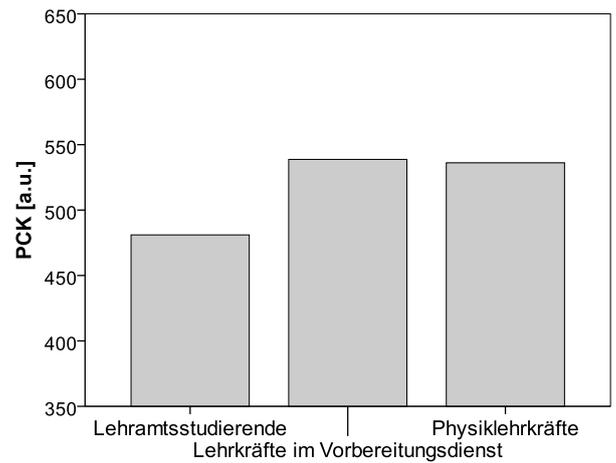
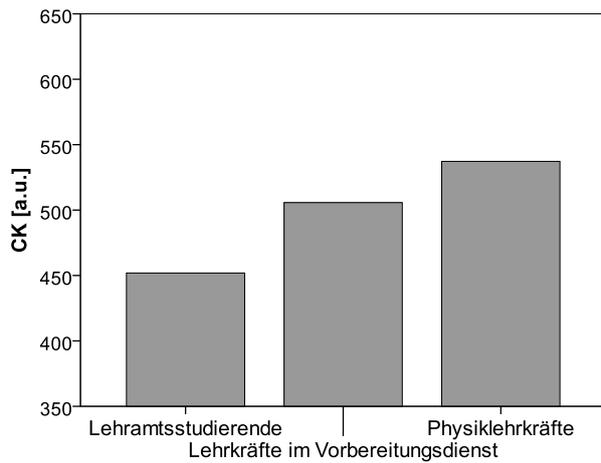
			Physiklehrkräfte	Physiklehrkräfte im Vorbereitungs- dienst	Lehramts- studierende Physik
GY stu- diert	CK	M	537	505	451
		SD	96	78	77
		SE	11	13	12
	PCK	M	536	538	481
		SD	80	109	82
		SE	9	18	13
NGY stu- diert	CK	M	450	428	411
		SD	73	70	76
		SE	17	22	12
	PCK	M	379	456	439
		SD	137	134	116
		SE	-	42	18
Diplom studiert	CK	M	501	493	-
		SD	98	76	-
		SE	17	22	-
	PCK	M	518	520	-
		SD	81	65	-
		SE	14	19	-

GY Gymnasiallehramt

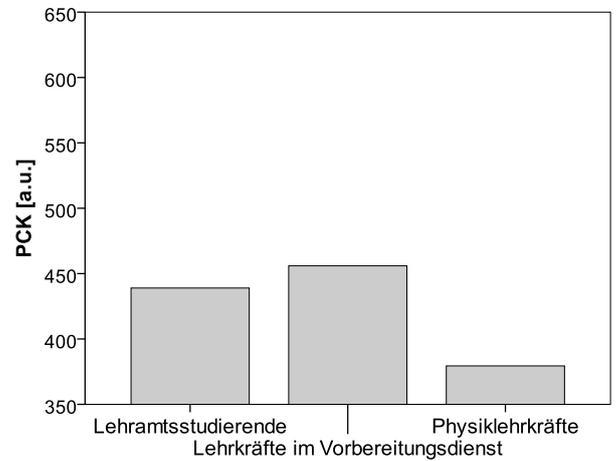
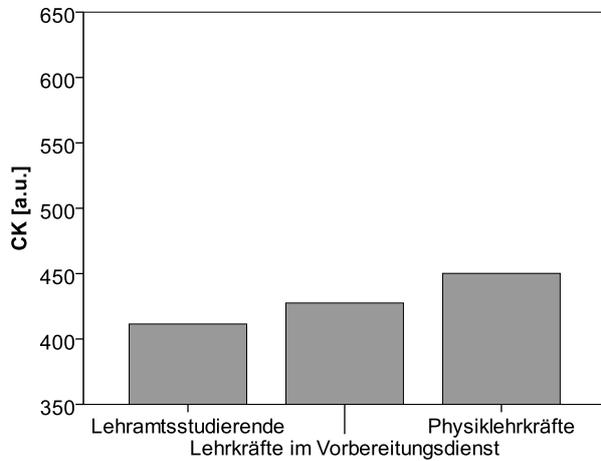
NGY Nicht-gymnasiales Lehramt

## Ergebnisse zu den Personenmerkmalen

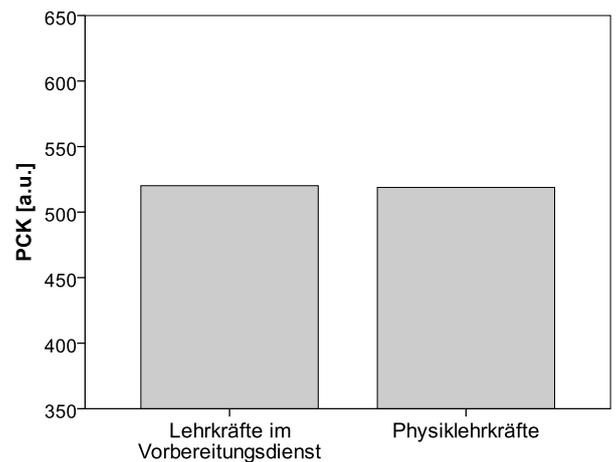
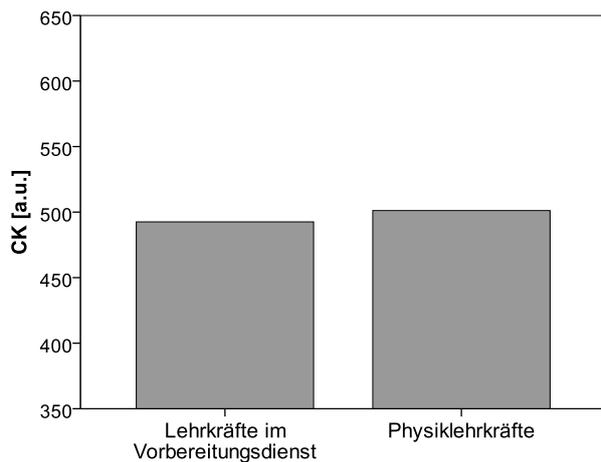
### a) Studium des gymnasialen Lehramts



### b) Studium des nicht-gymnasialen Lehramts



### c) Diplomstudium



**Abbildung 18:** Vergleich des Wissens von Lehramtsstudierenden, Lehrkräften im Vorbereitungsdienst und Physiklehrkräften aus NRW, getrennt nach dem absolvierten Studiengang

Für Personen, die ein gymnasiales Lehramt studiert haben/studieren, ergeben ANOVAs einen signifikanten Einfluss der Ausbildungsstufe auf CK ( $F(148,2)=12.93$ ,  $p<.001$ ), und PCK ( $F(148,2)=6.17$ ,  $p<.01$ ). Sowohl im CK als auch im PCK schneiden Lehrkräfte und Lehrkräfte im Vorbereitungsdienst signifikant besser als Studierende ab. Es ist kein signifikanter Unterschied zwischen Lehrkräften im Vorbereitungsdienst und Lehrkräften zu erkennen (siehe Tabelle 33).

**Tabelle 33:** Testleistungsunterschiede in CK und PCK für Personen, die ein gymnasiales Lehramt studiert haben bzw. studieren, Post-Hoc-Tests (GT2 nach Hochberg)

	CK	PCK
Lehrkräfte zu LiV	$p=.209$	$p=.998$
Lehrkräfte zu Studierenden	$p<.001$ , $d=0.99$	$p<.01$ , $d=0.68$
LiV zu Studierenden	$p<.05$ , $d=0.70$	$p<.05$ , $d=0.60$

p Signifikanz der Unterschiede

d Effektstärke der Unterschiede

LiV Lehrkräfte im Vorbereitungsdienst

Für Personen, die ein nicht-gymnasiales Lehramt studiert haben/studieren, ergibt eine ANOVA keinen signifikanten Einfluss der Ausbildungsstufe auf CK ( $F(68,2)=1.73$ ,  $p=.185$ ) und der Kruskal-Wallis-Test keinen signifikanten Einfluss der Ausbildungsstufe auf PCK ( $H(2)=1.76$ ,  $p=.422$ ).

Der Unterschied zwischen Lehrkräften im Vorbereitungsdienst und Physiklehrkräften, die einen Diplomstudiengang absolviert haben, wird mit t-Tests beleuchtet. Weder für CK noch für PCK ist ein signifikanter Unterschied zwischen den beiden Gruppen zu erkennen (CK:  $t(42)=0.27$ ,  $p=.786$ , zweiseitig; PCK:  $t(42)=0.05$ ,  $p=.960$ , zweiseitig).

Zusammenfassend ist festzustellen, dass nur bei den Gymnasiallehrkräften in CK und PCK signifikante Unterschiede zwischen den Gruppen auftreten. Studierende schneiden in dieser Gruppe signifikant schlechter ab als Lehrkräfte im Vorbereitungsdienst und Lehrkräfte.

Im Anhang sind für alle drei Ausbildungsstufen auch die Unterschiede zwischen den Ausbildungswegen (Studium für das gymnasiale Lehramt, Diplomstudium, Studium für das nicht-gymnasiale Lehramt) dargestellt (Abschnitt 10.9).

### 7.5.4.2 ZUSAMMENHANG ZWISCHEN CK UND PCK

Im Folgenden wird der Zusammenhang zwischen CK und PCK bei Lehramtsstudierenden der Physik, Lehrkräften der Physik im Vorbereitungsdienst und Lehrkräften der Physik und der studierten Schulform untersucht, da bisher nur einzelne Vergleiche gezogen wurden (siehe Abschnitt 2.5).

Die Korrelationen sind in Tabelle 34 dargestellt. Sie unterscheiden sich nicht signifikant zwischen Lehrkräften, Lehrkräften im Vorbereitungsdienst und Studierenden (siehe Tabelle 35). Besonders hervorzuheben ist, dass es keine Versuchspersonen aus dem Bereich der Lehramtsstudierenden, Lehrkräfte im Vorbereitungsdienst und Physiklehrkräfte gibt, die ein hohes CK und niedriges PCK haben, sowie keine Versuchspersonen, die ein niedriges CK und hohes PCK haben, wie in Abbildung 19 anhand der eingezeichneten Flächen zu erkennen ist.

**Tabelle 34:** Testleistungskorrelationen nach Pearson zwischen CK und PCK für Versuchspersonen aus NRW

		r	p
GY	Lehrkräfte	.407	<.001
	LiV	.306	.066
	Studierende	.339	<.05
NGY	Lehrkräfte	.271	.277
	LiV	.618	.057
	Studierende	.415	<.01

GY Gymnasiallehramt studiert

NGY Nicht-gymnasiales Lehramt studiert

LiV Lehrkräfte im Vorbereitungsdienst

r Korrelationskoeffizient

p Signifikanz der Korrelation

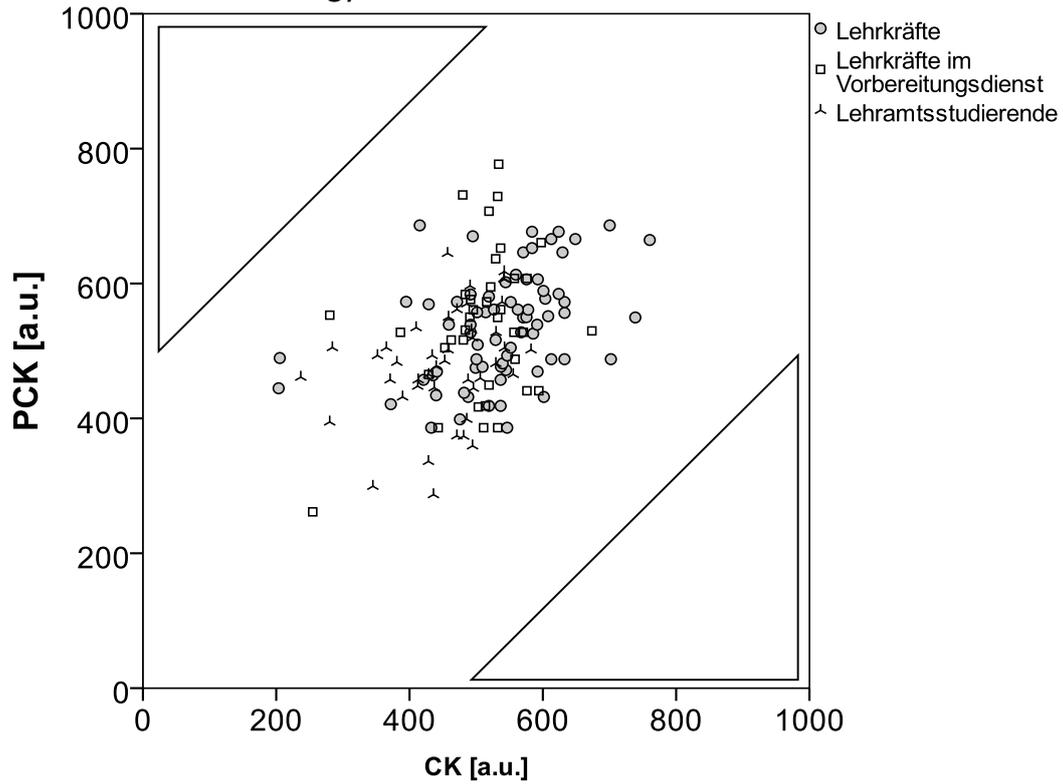
**Tabelle 35:** Unterschiede zwischen den Testleistungskorrelationen von CK zu PCK

Korrelationen zwischen		p
Lehrkräften und LiV	GY	.581
	NGY	.332
Lehrkräften und Studierenden	GY	.692
	NGY	.589
LiV und Studierenden	GY	.875
	NGY	.494

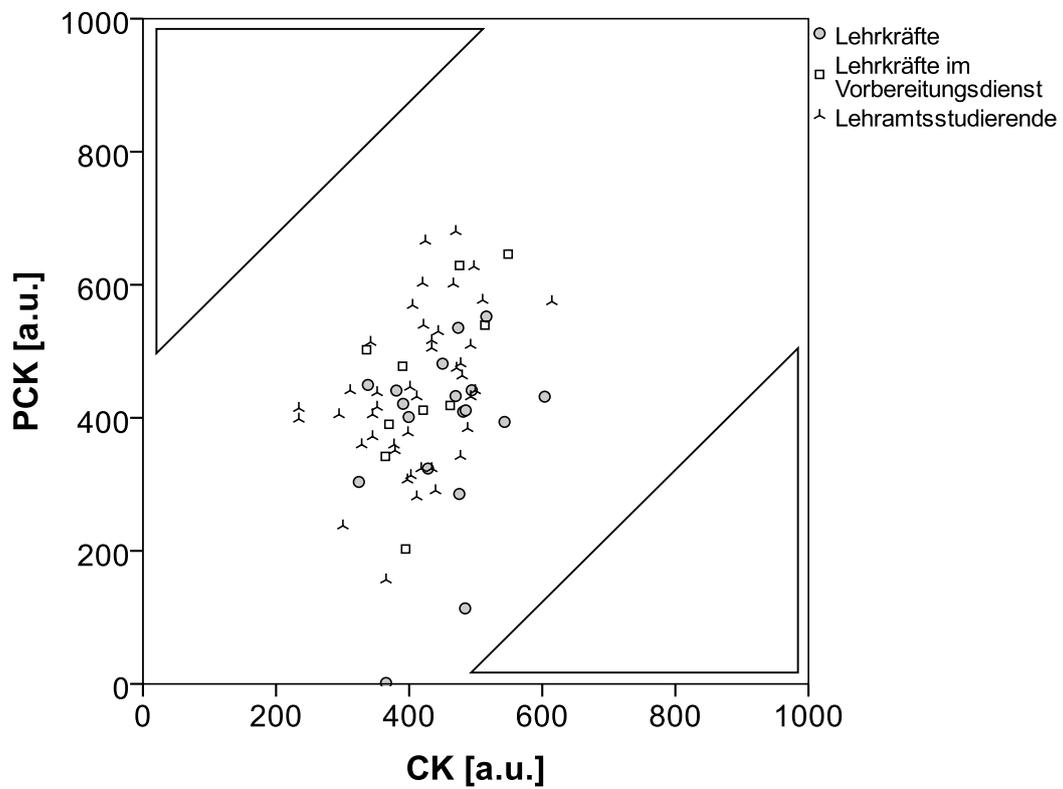
p Signifikanz der Unterschiede zwischen den Korrelationen

Weitere Konventionen siehe Tabelle 34

a) Absolviertes Studium des gymnasialen Lehramts



b) Absolviertes Studium des nicht-gymnasialen Lehramts



**Abbildung 19:** Zusammenhang zwischen CK und PCK für die Schulformen  
Die eingezeichneten Dreiecke verdeutlichen Abschnitte ohne Versuchspersonen und haben alle die gleiche Fläche.



# *8. KAPITEL*

## *DISKUSSION UND AUSBLICK*

In diesem abschließenden Kapitel wird kritisch reflektiert, inwieweit die Ziele der Studie erreicht wurden (8.1). Es wird diskutiert, ob Professionswissen und insbesondere PCK gemessen werden konnte (8.2), welche Schlüsse sich aus dem Zusammenhang von CK, PCK und PK ziehen lassen (8.3), wie einzelne Variablen mit Professionswissen zusammenhängen (8.4), wo diese Ergebnisse unsere Vorstellungen in Frage stellen oder bestätigen und wie aussagekräftig die Ergebnisse sind. Es wird außerdem dargelegt, welche Schlüsse sich aus dem Vergleich von Lehramtsstudierenden, Lehrkräften im Vorbereitungsdienst und Lehrkräften ziehen lassen (8.5).

Am Ende des Kapitels wird ein Ausblick gegeben, welche Studien sich anschließen und anschließen könnten (8.6).

## **8.1 REKAPITULATION DER ZIELE DER STUDIE**

In Kapitel 4 wurden die Modellierung und die Analyse des Professionswissens als Ziele für diese Studie definiert. Mit der Modellierung ist die Erstellung eines Modells, die Umsetzung des Modells in ein Testinstrument und die Modelltestung gemeint. Die Analyse des Professionswissens steht für die Untersuchung des Zusammenhangs des Professionswissens von Physiklehrkräften zu ihrem demographischen Hintergrund und dem Vergleich des Professionswissens von Lehramtsstudierenden, Lehrkräften im Vorbereitungsdienst und Lehrkräften der Physik.

In Kapitel 3 wurde Professionswissen modelliert, die Umsetzung in ein Testinstrument wurde in Kapitel 5.1 beschrieben, die Dimensionalität des Modells in Kapitel 7.2 getestet und die zufriedenstellende Reliabilität für die Dimensionen angegeben. Die Validität wurde in Kapitel 7.3 getestet und sichergestellt. Diese Abschnitte beschreiben das erste Forschungsziel.

Kapitel 7.5 hat das zweite Forschungsziel, die Analyse des Professionswissens, zum Inhalt. In Kapitel 7.5.2 und 7.5.3 konnte der Einfluss wichtiger demographischer Variablen auf CK und PCK durch Korrelationsanalysen und Regressionen geklärt werden. Abschließend wurden CK und PCK von Lehrkräften der verschiedenen Ausbildungsstufen in Kapitel 7.5.4 verglichen.

Die Ziele der Studie wurden damit erreicht. Es wurde ein reliables und valides Testinstrument entwickelt, das für weitere Studien zur Verfügung steht.

## **8.2 MESSUNG DES PROFESSIONSWISSENS**

Das Ziel der Studie war es, nicht nur Professionswissen zu modellieren, sondern das Modell auch in einen Test umzusetzen. Es war daher notwendig, das Modell so klein zu halten, dass alle Aspekte in einen zweistündigen Test umgesetzt werden konnten. Mehr Items hätten nur - aufgeteilt in mehrere Testhefte - in einem Multi-Matrix-Design umgesetzt werden können. Dies war aufgrund der erwarteten Stichprobengröße nicht möglich.

Damit war mit Beginn des Projektes festgelegt, dass einzelne Elemente des Professionswissens für das Modell ausgewählt werden mussten. Im Modell finden sich Wissensbereiche, -stufen, Themen und Facetten. Das, was im folgenden CK, PCK und PK

genannt wurde, ist das, was der Test erfasst, nicht das komplette Konstrukt. Das bedeutet, dass sich die Einschränkungen des Modells in den Tests und in den Ergebnissen widerspiegeln.

Auch wenn Items nicht immer eindeutig dem deklarativen, prozeduralen oder konditionalen Wissen zugeordnet werden können, fehlt keine Kategorie. Ganz anders verhält es sich bei den fachlichen Inhalten, die zwar eindeutig zugeordnet werden konnten, aber mit dem gewählten Schwerpunkt Mechanik alle anderen Themenbereiche ausschlossen. Eine Lehrkraft, die lange nicht Mechanik unterrichtet hat, wird möglicherweise schlechter abschneiden als eine, die Mechanik gerade in der Mittel- und Oberstufe unterrichtet.

Der Schwerpunkt wurde unter der Annahme gewählt, dass die Mehrheit der Physiklehrkräfte in Deutschland regelmäßig Mechanik unterrichtet. Mechanik ist ein Schlüsselthema, das auch während der universitären Ausbildung vertieft behandelt wird. Es kann allerdings nicht ausgeschlossen werden, dass die aktuellen Tests durch die Themenwahl einzelne Zielgruppen bevorzugen und dies einen Einfluss auf die Ergebnisse hat.

Die Bindung an ein Fach, und insbesondere an ein Thema, stellt außerdem eine Schwierigkeit für den internationalen Einsatz der Tests dar, da sich die Vertrautheit mit dem Thema Mechanik nicht ohne weiteres auf den angloamerikanischen Raum übertragen lässt. Dort wird das Fach „Science“ von Lehrkräften unterrichtet, die nicht in allen Grundlagenfächern spezialisiert sind (z. B. Hashweh, 1987).

An die Frage nach der Qualität des Instruments schließt sich die nach der Berechnung der Personenfähigkeit an. Um die Antworten aller befragten Personen einbeziehen zu können, wurde ein dreidimensionales Rasch-Modell gerechnet, das Unterschiede zwischen den beiden Raterinnen sowie der ersten und zweiten Studie einbezieht. Diese Komplexität verlangte einige Kompromisse. Beispielsweise musste die Berechnung der Reliabilität in einer anderen Rechnung als die Berechnung der Personenfähigkeiten geschehen und für die Reliabilität des CK-Tests konnten nur eine obere und untere Grenze, sowie eine Schätzung angegeben werden. Darüber hinaus ist die Passung, für die der MnSq steht, zwischen Items und Rasch-Modell nur ausreichend. Allerdings ist das bevorzugte dreidimensionale Modell das Beste der untersuchten. Bei einer besseren Passung hätte mit strengeren Kriterien für die Fit-Statistik gearbeitet werden können (z. B. Bond & Fox, 2007). Es scheint allerdings üblich zu sein, dass sich ein erheblicher Teil der Items als nicht Rasch-konform herausstellt (Bortz & Döring, 2006).

Begründet werden kann die begrenzte Passung zum Modell mit dem sehr weiten Konstrukt (Lehrerwissen), das untersucht wird. Es wurden sehr unterschiedliche Aufgaben gewählt, offen und geschlossen, lang und kurz, komplex und einfach, die auch noch verschiedene Wissensbereiche, -stufen und Facetten abdecken sollten. Möglicherweise sind diese tatsächlich qualitativ unterschiedlich.

Für zukünftige Arbeiten könnten mehrere Wege zielführend sein:

- Die Entwicklung standardisierter Items anhand eines Manuals, um den Einfluss weiterer Aspekte, wie der Aufgabenlänge, zu minimieren.
- Die Entwicklung mehrerer kürzerer Tests, die z. B. nur in der Facette oder im Wissensbereich variieren, um dann tatsächlich das CK, PCK oder PK in diesem Bereich zu messen. Dies wirft allerdings die Frage auf, ob eine solche Messung dem Konstrukt noch gerecht werden kann.
- Die genauere Anpassung der Aufgabenauswahl an die Zielgruppe: So könnten zum Beispiel explizit kürzere oder komplexere Aufgaben genutzt werden. Dies wird in Abschnitt 8.6 diskutiert.

### **8.2.1 MESSUNG DES PCK**

Die Validitätsprüfung hat ergeben, dass der PCK-Test ein Wissen misst, das vor allem Physiklehrkräfte auszeichnet und sich von CK und PK unterscheidet. Damit scheint sichergestellt, dass dieses Wissen für den Physikunterricht relevant ist. Dies rechtfertigt die Benennung als PCK – kurz, dem Wissen, das eine Lehrkraft zum Unterrichten ihres Faches benötigt (siehe 3.1). Ob es einen messbaren Einfluss auf Schülerleistung und – motivation hat, bleibt abzuwarten.

Wird das Modell betrachtet, ist die Relevanz der Facetten des PCK-Tests am kritischsten zu sehen. Es wurden aufgrund der beschränkten Testzeit normativ drei Facetten – Wissen über Schülervorstellungen, Experimente und Konzepte – ausgewählt, die als wichtig für den Physikunterricht gelten (siehe Abschnitt 3.2.4). Allerdings bleibt offen, ob andere Facetten möglicherweise zu anderen Ergebnissen geführt hätten. Die Wahl der Facetten hat einen Einfluss auf die Validität. Wird kein Zusammenhang zwischen Wissen und Handeln bzw. Schülerleistung und -motivation gefunden, kann dies mit der Wahl der Facetten zusammenhängen.

Es besteht keine uneingeschränkte Einigkeit darüber, ob PCK als Wissen in mehreren Facetten, als Verknüpfung mehrerer Facetten oder als Ganzes zu verstehen ist. Im ersten Falle müssten möglichst viele (oder die wichtigsten) Facetten getestet werden, um diese einzeln untersuchen zu können. Es ist möglich, dass Wissen über Experimente in Unterrichtsstunden mit Experimenten besonders relevant ist oder Wissen über Schülervorstellungen, wenn neue Konzepte eingeführt werden sollen. Der vorgestellte Test ist so kurz, dass die Facetten nicht statistisch trennbar sind, so dass diese nicht einzeln untersucht werden können.

Offen bleibt, ob sich tatsächlich die Ergebnisse mehrerer Tests zu verschiedenen Facetten zu einem Wert für PCK summieren ließen, da der Standpunkt vertreten werden kann, dass es sich bei PCK um ein Ganzes handelt, das mehr ist als die Summe seiner Teile, nicht weiter zerlegt werden kann und möglichst breit getestet werden sollte. In diesem Fall ist es eine Stärke des Testinstrumentes, dass einzelne Facetten miteinander verwoben werden. Allerdings führt ein solch großes Konstrukt unweigerlich zu einer weniger guten Messbarkeit wie an der Passung zu sehen ist. Um den Test unter dieser Annahme zu verbessern, müssten weitere Facetten, wie das Wissen über Leistungsbeurteilung, einbezogen werden bzw. die genutzten Facetten müssten breiter interpretiert werden. Es sollte unter „Wissen über Schülerverständnis“ nicht nur das Wissen über Schülervorstellungen abgefragt werden und unter „Wissen über Instruktionsstrategien“ nicht nur das Wissen über Experimente und Konzepte.

### **8.2.2 MESSUNG DES CK**

Der CK-Test misst offensichtlich Wissen, das weder PCK noch PK ist und das Lehrkräfte anderer Fächer nicht beherrschen. Hypothesenkonform haben Lehrkräfte am Gymnasium mehr CK als Lehrkräfte anderer Schulformen. Die untersuchten Diplom-Physikerinnen und -Physiker haben im Mittel weder mehr, noch weniger CK.

Allerdings bleibt die Einordnung des CK schwierig. In der Testentwicklung wurde der Schwerpunkt auf Schulwissen und vertieftes Schulwissen gelegt. Allerdings orientierte sich das Schulwissen eher an den Inhalten des gymnasialen Unterrichtes. Dies kann dazu führen, dass Lehrkräfte der Hauptschulen möglicherweise wenig Schulwissen wiedererkennen und eine Aufgabe eher als universitär einordnen würden, wohingegen sie für Lehrkräfte, die in der Oberstufe unterrichten, eindeutig an der Schule verortet ist.

## Diskussion und Ausblick

Im Laufe der Arbeit hat sich mein Verständnis von vertieftem Schulwissen von der zur Aufgabenkonstruktion genutzten Beschreibung (3.2.2) zu der folgenden Arbeitsdefinition verändert: *Wer vertieftes Schulwissen besitzt, hat ein Verständnis der grundlegenden physikalischen Konzepte, das keine höhere Mathematik benötigt, kann zwischen Schulwissen und universitärem Wissen unterscheiden und beides flexibel einsetzen und verbinden, um Lernen möglich zu machen.* Damit ist die Bekanntheit für die Definition des vertieften Schulwissens nicht mehr relevant, was die Trennung von Schulwissen und vertieftem Schulwissen erleichtern könnte. Es muss überlegt werden, ob vertieftes Schulwissen nicht Teil des PCK ist, da der Zusammenhang zum Lehren explizit gemacht werden kann.

Für den CK-Test konnten für die Reliabilität nur eine obere Grenze (.86) und als untere Grenzen die der offenen Items (.82) und der geschlossenen Items (.56) bestimmt, sowie eine Reliabilität für den gesamten Test (.86) geschätzt werden. Unter der Annahme, dass die beiden CK-Teile dasselbe Konstrukt messen und die geschlossenen Items nur aufgrund der geringen Anzahl (vier) keine höhere Reliabilität erreichen, kann die Reliabilität der offenen CK-Items als untere Grenze der Reliabilität für den gesamten CK-Test betrachtet werden.

Da die Werte für die obere Grenze und die Schätzung identisch sind und nahe bei der Reliabilität der offenen Items liegen, nehme ich weiterhin an, dass die Reliabilität von .86 als die Reliabilität des CK-Tests angegeben werden kann.

Da die Hinzunahme der geschlossenen Items die Reliabilität des CK-Tests nicht entscheidend verbessert, wäre es möglich, in weiteren Einsätzen des Tests auf sie zu verzichten. Dagegen steht der geringe Aufwand, diese zu bearbeiten und auszuwerten.

### 8.2.3 MESSUNG DES PK

Der PK<sub>D</sub>-Test wurde in dieser Studie insofern validiert, als dass sich PK von CK und PCK trennen lässt und Lehrkräfte anderer Fächer nicht besser oder schlechter abschneiden, er also nicht fachabhängig ist. Im PK<sub>D</sub>-Test konnten nur 17 der 45 Items in der Rasch-Analyse erfasst werden. Dies kann auch als Kürzungspotential für weitere Studien gesehen werden, da die Reliabilität danach immer noch zufriedenstellend ist.

Es muss die Frage gestellt werden, warum sich der PK<sub>KP</sub>-Test nicht mit der durchgeführten Rasch-Analyse erfassen lässt. Möglicherweise fragen die Items nicht alle ein und dasselbe Konzept ab. Allerdings wurden die PK-Tests auch nicht explizit für eine Rasch-Analyse entwickelt, die Daten der Pilotierung wurden nicht Rasch-analysiert, sondern nach klassischen Kriterien ausgewählt. Wird nur das PK der Versuchspersonen aus NRW, die an einem Gymnasium unterrichten, ergänzt um weitere Personen aus der ProwiN-Videostudie des Faches Physik, betrachtet ( $N_{\text{gesamt}}=107$ ), sind beide PK-Tests mit dem Programm Winsteps (Linacre, 2010) Rasch-skalierbar (Cauet et al., 2013).

### 8.2.4 MODELL

Der Kern des in der Studie genutzten Modells, eine Einteilung in CK, PCK und PK, wobei PCK-Items näher am Fach oder der Pädagogik liegen können (3.1), erscheint weiterhin schlüssig. Es lässt sich diskutieren, ob in das Modell nur Elemente aufgenommen werden sollten, die klar trennbar sind (wie die Dimensionen), oder auch, wie hier geschehen, Aspekte zur Aufgabenentwicklung. Die Rolle der Facetten und des (vertieften) Schulwissens wurde in den beiden vorherigen Abschnitten bereits diskutiert.

Das Modell deckt nicht den Erwerb von Professionswissen ab. Auch in dieser Studie wird davon ausgegangen, dass ein Test das (hypothetisch immer ähnlich strukturierte) Wissen aller Ausbildungsstufen abdecken kann. Das Modell lässt offen, wann, wo und wie dieses gemessene Wissen erworben wird.

Wird die Entwicklung des Professionswissens betrachtet, könnte davon ausgegangen werden, dass es kanonisches Professionswissen gibt (Banilower & Smith, 2012), also das Wissen, das gelehrt wird und aus dem wir die korrekten Antworten für die Testitems generieren. Dieses Wissen muss dann in individuelles Wissen umgewandelt werden, das mit Tests erfasst wird.

Der Erwerb des Wissens lässt sich in einem Kreislauf beschreiben. Individuelle Voraussetzungen führen zur Nutzung von Lerngelegenheiten. Daraufhin verändert sich das

individuelle Professionswissen. Dieses bildet wiederum die Grundlage für weiteren Wissenserwerb. Eine besondere Lerngelegenheit bieten Veranstaltungen, in denen kanonisches Professionswissen gelehrt wird. Eine weitere hervorzuhebende Lerngelegenheit bildet der Unterricht. Mit diesem Kreislauf kann auch der Einfluss von CK auf den Erwerb von PCK und der Einfluss von PCK auf den Erwerb von CK modelliert werden, da CK und PCK Voraussetzungen für die Nutzung weiterer Lerngelegenheiten sind.

### **8.3 WAS LÄSST SICH AUS DEM ZUSAMMENHANG ZWISCHEN CK, PCK UND PK SCHLIEßEN?**

Der Zusammenhang zwischen den Dimensionen wird mit Korrelationen und Regressionen beschrieben. Es muss davon ausgegangen werden, dass die tatsächlichen Korrelationen durch die dargestellten Werte unterschätzt werden, da sie aus messfehlerbehafteten WLEs (6.2.1) berechnet werden (Walter, 2005). Das gilt für alle in dieser Arbeit angegebenen Korrelationen, die daher vergleichbar bleiben. Latente Korrelationen konnten nur für den Zusammenhang von CK, PCK und PK angegeben werden (siehe Anhang, Abschnitt 10.3), nicht aber für die mit anderen unabhängigen Variablen. Eine Korrektur mit Hilfe der Reliabilität der Skalen erschien aufgrund der schwierigen Berechnung der Reliabilität des CK-Tests nicht sinnvoll. Der Zusammenhang der Dimensionen wurde ergänzend mit Streudiagrammen dargestellt (Abbildung 16).

Die Interpretation beginnt mit der Aufklärung von CK und PK an PCK, die eng mit den einzelnen Korrelationen zusammenhängt. Die Analyse zeigt, dass CK mehr als doppelt so viel PCK aufklärt wie PK<sub>D</sub>. Dies lässt mehrere Schlüsse zu:

- Der PK<sub>D</sub>-Test deckt nur einen Teil des PK ab. Wäre der PK<sub>KP</sub>-Test Rasch-modellierbar gewesen, hätte er einen weiteren Teil von PCK aufklären können.
- Im CK-Test werden ebenso wie im PCK-Test offene Aufgaben genutzt, im PK<sub>D</sub>-Test nur geschlossene. Dies hat einen Anteil an der unterschiedlichen Korrelation zwischen den Dimensionen.
- Die Ausbildung erfolgt in einer Weise, dass CK und PCK enger verknüpft sind als PK und PCK.
- PK und PCK der Lehrkräfte korrelieren nicht hoch genug. Es wäre besser, wenn PK und PCK enger verknüpft wären. Dies sollte in Aus- und Weiterbildung unterstützt werden.

Anhand der vorliegenden Daten mit nur jeweils einem CK-, PCK- und PK-Test kann nicht entschieden werden, welchen Anteil an dem Zusammenhang die Testinstrumente und das Aufgabenformat haben. Zum Vergleich müssten zwei der drei Tests unverändert übernommen werden und der dritte in dem Aufgabenformat oder den Facetten/Inhalten verändert werden. In anderen Studien sind die Korrelationen zwischen den Dimensionen sehr unterschiedlich (vgl. Tabelle 25), so dass sie keinen Anhaltspunkt bieten. Es bleibt offen, welche Korrelation zwischen den Dimensionen wünschenswert wäre. Liegt sie zu hoch, liegt der Verdacht nahe, dass die eingesetzten Tests keine unterschiedlichen Konstrukte messen. Zu niedrig darf sie auch nicht sein, da CK und PCK Dimensionen eines Konstrukts, des Professionswissens, sind. Die gemessenen Korrelationen beantworten nicht die Frage, welcher Zusammenhang zwischen den Dimensionen zu gutem Unterricht führt. Diese Frage kann nicht theoretisch beantwortet werden.

Gehen wir davon aus, dass CK als Vorwissen für den Erwerb von PCK dient, ist eine enge Verknüpfung zwischen CK und PCK erwünscht und sollte gefördert werden. Für den universitären Bereich bedeutet das, dass sich integrierte Physikkurse für Lehramtsstudierende nicht nur auf theoretische und experimentelle Physik beziehen sollten, sondern eine Integration von fachlichen und didaktischen Aspekten bieten müssten. Die Integration könnte im Bereich von Fortbildungen weitergeführt werden. Dies hätte, neben der Verbindung von CK und PCK, einen weiteren Vorteil: Lehrkräften würden immer mehrere Inhalte angeboten. Hat eine Lehrkraft ein besonderes Interesse am Fachlichen, würde ihr auch die Vermittlung nahegebracht. Sieht sie sich mehr als Pädagoge oder Pädagogin, werden trotzdem auch die fachlichen Inhalte vermittelt und mögliche Lücken geschlossen.

Es ist zu erkennen, dass die Korrelation jeweils am höchsten für die gesamte Stichprobe der Physiklehrkräfte ist. Werden die Streudiagramme (Abbildung 16) betrachtet, sind trotz der unterschiedlichen Höhe keine wesentlichen Unterschiede zwischen den Zusammenhängen von CK zu PCK und von PK zu PCK zu erkennen. In beiden Fällen ist zu erkennen, dass es keine Physiklehrkräfte gibt, die ein hohes CK bzw. PK bei gleichzeitig niedrigem PCK aufweisen. Allerdings finden sich einige wenige Lehrkräfte, die gleichzeitig hohes PCK und niedriges PK haben. Die Kombination aus hohem CK und hohem PK scheint nicht mit niedrigem PCK einherzugehen. Allerdings darf der Zusammenhang nicht überschätzt werden, CK und PK klären nur 27% des PCK auf.

Die Streudiagramme, in denen auch das Wissen der Studierenden und Lehrkräfte im Vorbereitungsdienst abgebildet wurde (Abbildung 19), unterscheiden sich nicht wesentlich von den eben besprochenen. Es finden sich weder Personen mit gleichzeitig hohem CK und niedrigem PCK, noch mit gleichzeitig hohem PCK und niedrigem CK. Hohes CK scheint also den Aufbau von PCK zu unterstützen. Tatsächlich fehlen aber Informationen über die Richtung des Zusammenhangs. Daher können CK und PCK jeweils als notwendige Voraussetzung für die andere Dimension gesehen werden. Auch hier gilt es, die entsprechenden Korrelationen zu beachten, CK und PCK klären sich, je nach Gruppe, nur zu einem Bruchteil von 10-21% gegenseitig auf (siehe Tabelle 24 und Tabelle 34).

Die Korrelationen zwischen CK und PCK unterscheiden sich erwartungsgemäß nicht signifikant zwischen den Gruppen. Die Stichproben sind zwar klein, aber die Unterschiede werden nicht einmal annähernd signifikant.

### **8.4 WIE HÄNGEN PERSONENMERKMALE MIT PROFESSIONSWISSEN ZUSAMMEN?**

Tabelle 39 stellt zum Vergleich die Richtung der erwarteten und festgestellten Zusammenhänge dar, die im Folgenden diskutiert werden.

Erwartungsgemäß korrelieren sowohl CK als auch PK mit PCK. Wider Erwarten korrelieren CK und PK miteinander, allerdings in geringerer Höhe als CK- und PCK-Test (7.3, HV9).

Ob in der eigenen **Oberstufenzeit ein Physikkurs** belegt wurde, hat einen großen Einfluss auf die Testergebnisse. Dies legt nahe, dass fachliche Lücken während des Studiums bzw. des Referendariats nicht geschlossen werden können. Eine alternative Erklärung ist, dass motiviertere Lehrkräfte bereits in der Oberstufe einen Physikkurs belegt haben und gleichzeitig bessere Ergebnisse in den Tests erzielen.

**Tabelle 36:** Richtung der erwarteten und gemessenen Korrelationen der unabhängigen Variablen mit CK und PCK

Variable	CK		PCK	
	H	E	H	E
CK			+	+
PK	<b>0</b>	+	+	+
Durchgängiger Besuch eines Physikkurses in der eigenen Oberstufenzeit	+	+	+	+
Belegter Studiengang gymnasiales Lehramt	+	+	+	+
Belegter Studiengang nicht-gymnasiales Lehramt	-	-	-	-
Belegter Studiengang Diplom	0	0	-	<b>0</b>
Studium der Physik/kein fachfremder Unterricht	+	+	+	+
Studium der Mathematik	<b>0</b>	+	0	0
Jahre im Schuldienst	0	0	-	<b>0</b>
Anzahl an Jahren, in denen bereits Physik unterrichtet wurde	<b>0</b>	+	-	<b>0</b>
Alter	0	0	-	-
Wochenstunden Physik, aktuell	+	+	+	+
Unterrichtet in der Unterstufe	0	0	<b>0</b>	-
Unterrichtet in der Mittelstufe	+	+	+	<b>0</b>
Unterrichtet in der Oberstufe	+	+	+	+
Unterrichtet am Gymnasium	+	+	+	+
Häufigkeit Fortbildungsbesuch Physik	+	+	+	+
Häufigkeit individueller Weiterbildung	+	+	+	+
Häufigkeit Fachgruppenarbeit	+	+	+	+
Häufigkeit pädagogischer Fortbildungen	<b>0</b>	-	+	<b>0</b>
Bayern	+	+	+	<b>0</b>
NRW	-	-	-	<b>0</b>
Geschlecht (männlich)	+	+	<b>0</b>	+

Nicht erwartungskonforme Ergebnisse sind fett gedruckt

- H Hypothese
  - E Ergebnis
  - +
  - 
  - 0
- Positiver Zusammenhang  
Negativer Zusammenhang  
Kein Einfluss

Bezüglich des belegten **Studiengangs** sind nur zwei Ergebnisse nicht hypothesenkonform. Physiklehrkräfte, die einen Diplomstudiengang absolviert haben, haben im Mittel nicht signifikant weniger PCK als die mit einem absolvierten Lehramtsstudiengang. Möglicherweise kann das im Test abgefragte PCK auch im Vorbereitungsdienst erlernt werden, wenn Fachwissen bereits vorhanden ist.

Das Studium der Mathematik hängt mit einem höheren CK in Physik zusammen. Dies könnte darauf zurückzuführen sein, dass Lehrkräfte zwei Fächer studieren und die betrachtete Gruppe im Wesentlichen Mathematik als zweites Schulfach unterrichtet. Das Bearbeiten des CK-Tests könnte bei hoher Vertrautheit mit der Mathematik der Sekundarstufe I leichter fallen. Allein dies könnte zu besseren Ergebnissen führen. Dazu könnte der Vorteil kommen, dass nicht zwei prinzipiell unterschiedliche Fächer verbunden werden müssen und mehr Kapazitäten für das mathematisch-naturwissenschaftliche Wissen bleiben.

Kompliziert ist die Interpretation der **Berufserfahrung**. Obwohl die Variablen dazu selbst eng korrelieren (siehe 5.3), haben sie einen unterschiedlichen Einfluss auf das Wissen. So hängen weder das Alter, noch die Jahre im Schuldienst mit CK zusammen, aber die Anzahl der Jahre, in denen Physik unterrichtet wurde. Obwohl aufgrund der veränderten universitären Ausbildung mit einem negativen Einfluss der Berufserfahrung auf PCK gerechnet wurde, hat tatsächlich nur das Alter diesen Einfluss. So müssten also junge Physiklehrkräfte mit langer Berufserfahrung am besten abschneiden – oder ältere, mit wenig Berufserfahrung, am schlechtesten. Zumindest letzteres scheint für Quereinsteiger, die Ihre Ausbildung als Lehrkraft erst in verhältnismäßig hohem Alter begonnen haben und Personen, die lange Auszeiten vom Beruf genommen haben, einleuchtend.

Eine verbesserte universitäre Ausbildung kann als Begründung für das schlechtere Abschneiden von älteren Lehrkräften in PCK herangezogen werden. Verbessert sich die Ausbildung in Zukunft nicht stetig weiter, dürfte der Einfluss des Alters in Zukunft sinken.

Die fehlende Korrelation von CK und Alter kann dahingehend interpretiert werden, dass die Fachausbildung über die Jahre weder besser noch schlechter geworden ist. CK korreliert mit den Jahren, in denen Physik unterrichtet wurde. Das könnte bedeuten, dass (vertieftes) Schulwissen an der Schule gelernt wird, was dafür spräche, dieses an

der Universität zu vertiefen. Das wird im Rahmen der Ergebnisse zu den Ausbildungsstufen diskutiert.

Wie erwartet, hängt die Anzahl an **Stunden**, in denen jede Woche Physik unterrichtet wird, positiv mit CK und PCK zusammen. Die die Anzahl der Lerngelegenheiten scheint entscheidend zu sein.

Der Einfluss der **Schulstufe** kann wie folgt beschrieben werden: Lehrkräfte, die Physik in der Unterstufe unterrichten, haben weniger PCK als jene, die nicht in der Unterstufe unterrichten. Dies könnte darauf zurückzuführen sein, dass sie in der Zeit, in der sie in der Unterstufe unterrichten, nicht in anderen Schulstufen unterrichten und ihnen Lerngelegenheiten zur Mechanik fehlen. Einen positiven Zusammenhang mit CK hat der Unterricht in der Mittel- und Oberstufe, in der auch Mechanik regelmäßig unterrichtet wird. Es bieten sich Lerngelegenheiten und das zum Unterrichten benötigte Wissen wird wiederholt. Der Unterricht in der Mittelstufe hat im Gegensatz zur Oberstufe keinen Einfluss auf PCK. Der Einfluss des Unterrichts in der Oberstufe auf CK und PCK lässt sich dadurch erklären, dass diese Lehrkräfte am Gymnasium unterrichten und eine Lehrbefähigung für die Sekundarstufe II haben, also im Regelfall einen gymnasialen Studiengang absolvierten oder ein Diplom haben.

Hypothesenkonform ist es, dass es zu besseren Ergebnissen führt, wenn am **Gymnasium** unterrichtet wird. Dies hat einen noch größeren Einfluss als der belegte Studiengang. Dieses Ergebnis wurde bereits für die Validierung genutzt. Wenn Professionswissen mit Schülerleistung und -motivation zusammenhängt, haben Schülerinnen und Schüler an Haupt- und Realschulen unter diesem Gesichtspunkt schlechtere Chancen. Abzuwarten bleibt der Einfluss neuer, fünfjähriger Bachelor-Master-Studiengänge für Lehramtsstudierende der verschiedenen Schulformen (Matzdorf, 2012).

Sowohl häufige **Fortbildungen** zum Fach Physik, individuelle Weiterbildung als auch Fachgruppenarbeit hängen mit höherem CK und PCK zusammen. Allerdings darf, wie bei allen Korrelationen, daraus kein Schluss auf die Kausalität gezogen werden. Es gibt mehrere mögliche Interpretationen:

- Lehrkräfte lernen in Fortbildungen CK und PCK.
- Lehrkräfte, die Fortbildungen besuchen, haben bereits hohes CK und PCK
- Lehrkräfte mit hohem CK oder PCK erkennen ihre Lücken besser und schließen diese gezielt mit Fortbildungen.

## Diskussion und Ausblick

Der negative Einfluss häufiger Besuche pädagogischer Fortbildungen auf CK deutet darauf hin, dass Fortbildungen gezielt ausgesucht werden. Eine mögliche Interpretation ist, dass Lehrkräfte, die sich eher als Pädagogin oder Pädagoge sehen und entsprechende Fortbildungen besuchen, Lücken im Fachlichen haben. Dies müsste durch eine Untersuchung der Überzeugungen belegt werden. Um diesen Einfluss zu verringern, sollten Fortbildungen immer Elemente aus CK, PCK und PK beinhalten, damit alle Bereiche gleichmäßig vertieft werden.

Der erwartete Unterschied zwischen den **Bundesländern** ist für den Bereich CK tatsächlich zu messen und möglicherweise auf unterschiedliche Schwerpunkte in der Ausbildung zurückzuführen. Allerdings ist kein Zusammenhang zwischen dem Bundesland und PCK zu erkennen. Daraus lassen sich schließen, dass unterschiedlich hohes CK mit vergleichbar hohem PCK einhergehen kann und dass sich, in diesem Falle, Leistungsunterschiede von Schülerinnen und Schülern nicht auf Unterschiede im PCK zurückführen lassen. Es sollte im Gedächtnis behalten werden, dass die Stichprobe der Lehrkräfte nicht repräsentativ ist, und die Ergebnisse zu Lehrkräften und Schülerinnen und Schülern nicht aus derselben Studie stammen.

Leistungsdifferenzen zwischen Lehrerinnen und Lehrern finden sich nicht nur im CK, sondern auch im PCK. Dies leitet zu den Ergebnissen der Regressionen über, da das Geschlecht unter Kontrolle des CK, des Schulzweigs und des Alters keinen signifikanten Einfluss auf PCK hat. Bei Korrelationen wird nur der Zusammenhang zweier Variablen betrachtet, ohne andere Variablen, die den Effekt möglicherweise besser erklären würden, zu beachten bzw. zu kontrollieren.

Welche Variablen haben in der generalisierbaren **Regression** einen Einfluss? Die Variablen Geschlecht, fachfremder Unterricht und Unterricht am Gymnasium haben einen Einfluss auf CK, der auf andere Stichproben übertragbar sein sollte und die Variablen CK, Alter und Unterricht am Gymnasium haben einen Einfluss auf PCK. Es fällt auf, dass der Unterricht am Gymnasium nicht nur auf CK, sondern auch auf PCK unter Kontrolle des CK einen Einfluss hat. Dieses Ergebnis widerspricht Brunner et al. (2006), Dollny (2011) und Riese (2009), bestätigt aber die in der Validierung getroffene Annahme (siehe 7.3).

Der negative Einfluss des fehlenden Physikstudiums, also fachfremden Unterrichts, auf CK bestätigt die Annahme, dass eine Lehrkraft nicht fachfremd unterrichten sollte. Die Variable fachfremder Unterricht hat keinen signifikanten Einfluss auf PCK. Dies könnte daran liegen, dass CK als Variable in die Regression zur Erklärung von PCK aufgenommen wurde. Dieselbe Erklärung lässt sich auf das Geschlecht anwenden: Unter Kontrolle des CK hat es keinen Einfluss auf PCK. Das Alter wiederum hat nur einen Einfluss auf PCK, nicht auf CK. Dies führe ich nach Dollny (2011), auf eine verbesserte universitäre Ausbildung zurück, die direkt auf PCK wirkt. Diese Annahme führt zu dem Schluss, dass die Ausbildung einen Einfluss auf das abgeprüfte Wissen hat und zu der Hoffnung, dass die universitäre Ausbildung weiter verbessert werden kann.

Die weiteren Ergebnisse in Bezug auf die Aufklärung von CK bzw. PCK stimmen nicht besonders optimistisch, wenn CK und PCK von Physiklehrkräften verbessert werden sollen. Das Geschlecht lässt sich nicht ändern. Der Schulformunterschied scheint gesellschaftlich gewollt und wird von der Wissenschaft als Validierungsmöglichkeit genutzt. Es bleibt nur der Hinweis, dass Physik nicht fachfremd unterrichtet werden sollte, da dies mit schlechterem CK zusammenhängt, welches wiederum der wichtigste Prädiktor für PCK ist.

Es zeigt sich, dass der Besuch von Fortbildungen unter Kontrolle der anderen unabhängigen Variablen keinen Einfluss auf CK und PCK hat. Die in den Korrelationen gesehenen Unterschiede werden durch andere Variablen aufgeklärt.

Hervorzuheben ist, dass in der Regression, die für verallgemeinerbar gehalten wird, jeweils mehr als ein Drittel an CK und PCK durch jeweils nur drei unabhängige Variablen aufgeklärt werden können.

Nach der Interpretation der Ergebnisse wird auf die Regression als Auswertemethode eingegangen. Regressionen werden zwar benutzt, um Vorhersagen zu machen, geben aber keine Richtung, keinen kausalen Zusammenhang an und dürfen auch nicht so interpretiert werden (Tabachnick & Fidell, 2001). Im Falle des fachfremden Unterrichts erscheint allerdings der kausale Zusammenhang (kein Studium der Physik führt zu einem geringeren Fachwissen) so klar, dass er als bestätigt angesehen werden kann. Würde sich eine fachfremd unterrichtende Lehrkraft entsprechend nachqualifizieren, würde aller Voraussicht nach ihr CK wachsen.

Für die durchgeführten Regressionen wurden aus einer großen Anzahl möglicher unabhängiger Variablen einige zur aktuellen Situation der Physiklehrkräfte ausgesucht.

## Diskussion und Ausblick

Dies betrifft Variablen, denen ein großer Einfluss vorhergesagt wurde (wie die Schulform und fachfremder Unterricht, aber auch CK für die Aufklärung von PCK) und Variablen, von denen kein großer Einfluss erwartet wurde (der Einfluss von Fortbildungen). Letztere wurden betrachtet, um ihren Einfluss unter Kontrolle der anderen Variablen beschreiben zu können. Daher ist es gut möglich, dass eine veränderte Auswahl an Variablen zu anderen Ergebnissen führen würde: Die Regression stellt eine möglichst gute Aufklärung anhand der vorgegebenen Daten dar (Tabachnick & Fidell, 2001).

Alle erhobenen unabhängigen Variablen in die Analyse einfließen zu lassen, war aufgrund der Stichprobengröße unmöglich. Die Stichprobe war bereits mit 31 (CK) bzw. 25 (PCK) Personen pro Variable für eine explorative Regression knapp kalkuliert. Gehen wir wie Tabachnick & Fidell (2001) von 40 Personen pro Variable aus, müssten, um allein die 21 (CK) bzw. 23 (PCK) Variablen zu beachten, die mit Korrelationen untersucht wurden, über 900 Physiklehrkräfte untersucht werden. Allerdings würde auch eine so große Stichprobe nicht davon befreien, einzelne Variablen auszuwählen. Da z. B. die beiden Variablen Alter und Berufserfahrung hoch miteinander korrelieren, kann ihr (gegenläufiger) Einfluss nicht in einer Regression untersucht werden.

Die verhältnismäßig geringe Anzahl von Versuchspersonen erklärt auch, warum Regressionen in verschiedenen Substichproben zu unterschiedlichen Ergebnissen führen. Um dies zu vermeiden, wäre bei der Stichprobengröße von 279 Personen eine Beschränkung auf höchstens sieben unabhängige Variablen nötig gewesen.

Um besonders viel CK bzw. PCK aufklären zu können, gibt es mehrere Möglichkeiten. Nach der Berechnung der Korrelationen zwischen den Variablen und CK bzw. PCK könnten diejenigen unabhängigen Variablen ausgewählt werden, die empirisch einen besonders großen Einfluss gezeigt haben. Alternativ könnten mehrere voneinander unabhängige Regressionen mit unterschiedlichen inhaltlichen Schwerpunkten, die jeweils andere Aspekte außer Acht lassen, durchgeführt werden. Mit der gezeigten Analyse wurde der Versuch gemacht, möglichst viele Variablen zur aktuellen Situation der Lehrkräfte einzubeziehen.

## 8.5 UNTERSCHIEDE ZWISCHEN DEN AUSBILDUNGSSTUFEN

Welche Schlüsse lassen sich aus dem Vergleich des Professionswissens von Lehramtsstudierenden, Lehrkräften im Vorbereitungsdienst und Lehrkräften ziehen?

Wird der Unterschied von CK und PCK getrennt nach Schulformen betrachtet, ist zu erkennen, dass die Stichproben für Personen, die einen Diplomstudiengang oder ein nicht-gymnasiales Lehramt studiert haben, recht klein sind. In diesen Gruppen sind keine signifikanten Einflüsse zu finden. Aufgrund der Stichprobengröße werden vorrangig die Ergebnisse der Personen interpretiert, die ein gymnasiales Lehramt studiert haben. Hier zeigt sich dasselbe Bild für CK und PCK: Lehramtsstudierende schneiden schlechter ab als Lehrkräfte und Lehrkräfte im Vorbereitungsdienst. Zwischen den beiden letztgenannten Gruppen zeigen sich keine signifikanten Unterschiede. Es scheint eindeutig zu sein, dass das Wissen, das Physiklehrkräfte haben, nicht vollständig an der Universität erlangt wird, sondern erst in der praktischen Ausbildung oder dem Arbeitsalltag. Hervorzuheben ist, dass dies nicht nur für PCK gilt, sondern auch für CK. Es bleibt zu überlegen, ob bzw. wie dieses Ergebnis die weitere Fachausbildung von angehenden Lehrkräften beeinflussen könnte und wie an den Universitäten (vertieftes) Schulwissen vermittelt werden sollte, da Studierende offensichtlich über weniger CK verfügen als Lehrkräfte. Dies wäre zum Beispiel möglich, indem Schulwissen in universitären Veranstaltungen regelmäßig aufgegriffen und Verbindungen im Sinne des vertieften Schulwissens hergestellt würden. Um im Fachwissenstest gut abzuschneiden, könnte es hilfreich sein, wenn die Versuchsperson geübt darin ist, sich mit ungewöhnlichen, neuen Fragestellungen zu beschäftigen, um Wissen flexibel einzusetzen. Dafür könnte bereits in der Ausbildung der Grundstein gelegt werden.

Entgegen der Erwartung werden bei Lehrkräften keine höheren Leistungen bezüglich CK und PCK als bei Lehrkräften im Vorbereitungsdienst gefunden. Es ist möglich, dass sich die gegenläufigen Effekte von Alter und Berufserfahrung aufheben.

Mit den kleinen Stichproben für Personen, die einen Diplomstudiengang bzw. einen nicht-gymnasialen Lehramtsstudiengang absolviert haben, können Hypothesen für weitere Untersuchungen generiert werden. Für den Fall des Diplomstudiums liegen CK bzw. PCK für Lehrkräfte und Lehrkräfte im Vorbereitungsdienst sehr nah beieinander. Das Ergebnis entspricht dem der größeren Gruppe mit einem gymnasialen Studiengang, was dafür spricht, dass angehende Lehrkräfte im Referendariat die Möglichkeit

haben, das nötige Wissen zu erwerben und sie es sich nicht erst im Schulalltag erarbeiten müssen. Weniger positiv ist die Interpretation, dass, möglicherweise aufgrund fehlender schlagkräftiger Fortbildungskonzepte, Lehrkräfte ihr Wissen nicht mehr erweitern, sobald sie im Beruf sind – obwohl sie jeden Tag die Möglichkeit zur Reflexion haben. Möglich ist auch, dass die Tests genau das Wissen abfragen, das im Referendariat verbalisiert und gelernt wird und Lehrkräfte im Schulalltag trotzdem dazulernen – dies aber nicht gemessen wird.

Für die Gruppe der Personen, die ein nicht-gymnasiales Lehramt studiert haben bzw. studieren, ergibt sich kein signifikanter Unterschied für CK oder PCK zwischen den Gruppen.

Für CK sind tendenzielle Unterschiede in der erwarteten Richtung zwischen den Gruppen zu erkennen. Die fehlende Signifikanz liegt möglicherweise darin begründet, dass 5 der 18 Lehrkräfte fachfremd unterrichten und sie daher das mittlere CK der Gruppe senken (siehe Anhang, Abschnitt 10.8).

Eine Hypothesenbildung für PCK ist schwierig; sicher ist nur, dass die ursprüngliche Wachstumshypothese aufgrund der vorliegenden Ergebnisse abgelehnt werden muss, da Lehrkräfte tendenziell schlechter abschneiden als Lehrkräfte im Vorbereitungsdienst und Lehramtsstudierende. Allerdings ist unklar, ob

- aufgrund fehlender signifikanter Unterschiede davon ausgegangen werden sollte, dass keinerlei Unterschiede vorliegen,
- tatsächlich angenommen werden müsste, dass Lehrkräfte weniger PCK haben als Lehrkräfte im Vorbereitungsdienst und Studierende oder
- die Ergebnisse die Hypothesenbildung nicht unterstützen können, da sie nicht schlüssig zu erklären sind.

Der Vergleich Grund- versus Hauptstudium und Beginn versus Ende des Referendariats sollte ebenfalls mit dem vorliegenden Testinstrument durchgeführt werden, um den Erwerb des Professionswissens noch weiter zu beleuchten. In der vorliegenden Stichprobe kann er leider aufgrund der Stichprobengröße nicht verwirklicht werden. Zudem fiel die Datenerhebung in die Umstellung der Lehrerausbildung auf Bachelor-Studiengänge, so dass der Vergleich dadurch beeinflusst würde.

## 8.6 AUSBLICK

In diesem Ausblick sollen drei Aspekte diskutiert werden: Erstens Möglichkeiten der Reanalyse der Daten, zweitens eine Erweiterung des Testinstruments und drittens eine bereits angelaufene Folgestudie.

Eine mögliche **Reanalyse** betrifft nicht die Kodierung, sondern ausschließlich die Auswertemethode. Bisher wurden die Daten der zweiten Studie nicht nach den klassischen Kriterien Trennschärfe, Itemschwierigkeit und Cronbachs Alpha analysiert, um die gezogenen Schlussfolgerungen zu validieren. Es würde eine substantielle Übereinstimmung zwischen den gewonnenen Personenfähigkeiten erwartet (Bortz & Döring, 2006). Für eine klassische Analyse müssten die Items ausgesucht werden, die eine genügend hohe Interraterübereinstimmung bieten, um nur von einer Raterin ausgewertet zu werden. Daraufhin könnten die kodierten Antworten analysiert werden. Allerdings wäre das nur für die Daten der zweiten Studie möglich, da in einer klassischen Analyse nicht mehrere Studien zusammengefasst werden können.

Die vorhandenen Antworten könnten bei einer erneuten Kodierung daraufhin untersucht werden, welche Fehlvorstellungen, die typischerweise bei Schülerinnen und Schülern gefunden werden, auch von Lehrkräften gezeigt werden. Dies ist sowohl bei den CK-Items möglich, die zum Teil bekannte Fehlvorstellungen als Distraktoren nutzen, als auch bei den PCK-Items, in denen die Lehrkräfte nicht nur nach typischen Schülervorstellungen gefragt wurden, sondern auch nach der fachlich korrekten Lösung.

In der vorliegenden Studie mussten die Antworten der Versuchspersonen auf halboffene Fragen nach einer umfangreichen Schulung aufwendig von zwei Raterinnen kodiert werden. Die große Anzahl an Antworten kann zukünftig genutzt werden, um Antwortmöglichkeiten für ein geschlossenes Aufgabenformat zu generieren. Für CK-Items können häufig auftauchende falsche Antworten als Distraktoren genutzt werden. Für PCK könnte analog zu Dollny (2011), bzw. der Entwicklung des PK<sub>KP</sub>-Tests vorgegangen werden (5.1.2.2), indem Expertinnen und Experten die gesammelten kategorisierten Antworten auf ihre Richtigkeit bewerten und damit eine Rangfolge der Antwortmöglichkeiten festlegen. In diesem Fall würden allerdings Vorteile der offenen Aufgaben (siehe 5.1.1.2) gegen eine höhere Testökonomie und Auswerteobjektivität eingetauscht.

Es bleibt zu bedenken, dass das zur Testkonstruktion genutzte Modell von Professionswissen zu Einschränkungen führt (siehe 8.2.1). Dies betrifft insbesondere die Beschränkung auf die Facetten Wissen über Schülervorstellungen, Konzepte und Experimente. Diese Einschränkungen übertragen sich auf Testinstrumente, die direkt auf dem bestehenden aufbauen. Diese Schwierigkeit könnte beispielsweise durch das Ergänzen weiterer Facetten oder Themen reduziert werden.

Die genutzten Instrumente zu CK und PCK sollten zu einem **adaptiven Testinstrument** erweitert werden, da sie unter mindestens zwei Gesichtspunkten nicht fair sind: Erstens haben Lehrkräfte einen Vorteil, die das Thema Mechanik gerade unterrichten. Zweitens sind die Aufgaben für Lehrkräfte, die an Haupt- und Realschulen unterrichten, teilweise weiter von ihrem Schulalltag entfernt, als für Lehrkräfte des Gymnasiums.

Bei der Erweiterung der Testinstrumente sollte Folgendes beachtet werden:

- Die Inhalte der Aufgaben müssen mehrere Themen der Curricula umfassen, damit für jede Lehrkraft individuelle Tests zusammengestellt werden können.
- Sollen die Aufgaben an das Schulwissen anknüpfen, müssen sie sich auf das Schulwissen der jeweils unterrichteten Schulform beziehen. Damit wird die Schwierigkeit der Aufgaben so gewählt, dass die Aufgaben nicht für eine Gruppe Routineaufgaben und für eine andere Gruppe eine ferne Erinnerung an die eigene Oberstufenzeit darstellen. Ein unterschiedliches Funktionieren der Aufgaben für verschiedene Gruppen kann so vermieden werden.
- Die Zusammenstellung der Aufgaben muss in jedem Falle reliabel und valide sein.
- Das Wissen der Versuchspersonen kann verglichen werden, wenn alle Aufgaben Rasch-skaliert sind und für alle Gruppen gleich funktionieren, also keinen DIF zeigen.

Diese Aspekte werden unter dem Begriff des fairen Testens zusammengefasst und würden die Inhaltsvalidität entscheidend verbessern. Die Erweiterung des Instrumentes könnte mit der oben beschriebenen Schließung der Aufgaben kombiniert werden. Das Projekt KiL verfolgt bereits den Ansatz, mehrere Themen innerhalb der Physik abzudecken, allerdings liegen hier Lehramtsstudierende im Fokus der Untersuchung (Kröger, Neumann, & Petersen, 2013).

Ein Schritt in die Richtung eines fairen Tests für Lehrkräfte der verschiedenen Schulformen wäre die Entwicklung eines Instrumentes für Lehrkräfte an Haupt- und Realschulen und deren Untersuchung mit einer großen Stichprobe. Dieses Instrument könnte über Ankeritems mit dem vorhandenen Instrument verbunden werden, was einen Vergleich der Gruppen möglich macht. In einem allerersten Schritt könnten die bereits vorhandenen Antworten der Lehrkräfte an Haupt- und Realschulen getrennt von denen der Lehrkräfte an Gymnasien beurteilt und ausgewertet werden.

Die Schließung und Erweiterung des in dieser Studie genutzten Testinstruments wird noch nicht umgesetzt, ist aber möglich. Bereits begonnen hat die direkt an diese Arbeit anschließende **ProwiN-Videostudie** (Fischer et al., 2011), in der der Zusammenhang zwischen Wissen und Handeln der Lehrkräfte und Schülerleistung und –motivation untersucht wird. Die Videostudie bietet die Möglichkeit, die Relevanz des getesteten Wissens für Unterricht zu klären. Die in der hier vorgestellten Studie entwickelten Fragebögen werden in gekürzter Form in der ProwiN-Videostudie Physik von den Doktorierenden Eva Cauet und Sven Liepertz eingesetzt. Das Manual wird aufgrund der Erfahrungen angepasst, insbesondere um die Interraterübereinstimmung zu verbessern. Im Rahmen dieses Projektes wird das hier vorgestellte Modell des Professionswissens auf Handeln und Schülerleistung erweitert (Kirschner & Cauet). Dafür wird in einem ersten Schritt der Zusammenhang von gelehrtem Professionswissen zu individuellem Wissen hergestellt, in einem zweiten der zwischen dem mit Papier-und-Bleistift-Tests gemessenem Wissen und dem Wissen, das, bzw. dessen Anwendung im Unterricht beobachtet werden kann. Der dritte Schritt modelliert die Verbindung von Professionswissen zu Schülerhandeln, -leistung und -motivation.



# *VERZEICHNISSE*

## 9.1 ABBILDUNGSVERZEICHNIS

<b>Abbildung 1:</b> Die Aufteilung des Professionswissens in die drei Bereiche CK, PCK und PK .....	7
<b>Abbildung 2:</b> Modelle von PCK .....	9
<b>Abbildung 3:</b> Kompetenzmodell von COACTIV mit Schwerpunkt auf dem Professionswissen (Kunter et al., 2011, S. 32) .....	10
<b>Abbildung 4:</b> Kompetenzstrukturmodell (Riese, 2009, S. 26, in Anlehnung an Baumert & Kunter, 2006; Blömeke, Felbrich, & Müller, 2008) .....	10
<b>Abbildung 5:</b> Ausschnitt aus dem Angebots-Nutzungs-Modell (nach Reusser & Pauli, 2003, S. 3).....	11
<b>Abbildung 6:</b> Die Aufteilung des Professionswissens in drei Bereiche CK, PCK und PK und die Zuordnung der Items zu den Professionswissensbereichen .....	24
<b>Abbildung 7:</b> Beispielitem aus dem CK-Test (CK 1180) .....	28
<b>Abbildung 8:</b> Das Magnusson-Modell - Facetten von PCK nach Magnusson et al. (1999, S. 99).....	31
<b>Abbildung 9:</b> Der Zusammenhang zwischen den Facetten von PCK dargestellt als PCK-Map nach Park & Chen (2012, S. 930) .....	33
<b>Abbildung 10:</b> Das Modell zur Testkonstruktion.....	36
<b>Abbildung 11:</b> Beispielitems aus dem PCK-Test (PCK S230, PCK S180) .....	45
<b>Abbildung 12:</b> Beispiel für 1. den deklarativen (PKD 4) und 2. den konditional-prozeduralen Teil des PK-Tests (PK 18) .....	49
<b>Abbildung 13:</b> Zusammenhang zwischen Alter und Lehrerfahrung .....	55
<b>Abbildung 14:</b> Design der Datenmatrix. Für jede Raterin sind links die Kodierungen der offenen Aufgaben zu sehen und rechts die der geschlossenen Aufgaben, die für beide Raterinnen identisch sind. ....	61
<b>Abbildung 15:</b> Design der Datenmatrix zur Berechnung der Reliabilität. Links sind die Kodierungen der offenen Aufgaben zu sehen, rechts die der geschlossenen Aufgaben. 9 stellt eine fehlende Information dar (Missing).....	63
<b>Abbildung 16:</b> Zusammenhang zwischen CK und PCK bzw. PK und PCK für alle untersuchten Physiklehrkräfte, getrennt nach der Schulform, an der aktuell unterrichtet wird.....	83
<b>Abbildung 17:</b> Korrelationen der unabhängigen Variablen mit CK und PCK .....	85
<b>Abbildung 18:</b> Vergleich des Wissens von Lehramtsstudierenden, Lehrkräften im Vorbereitungsdienst und Physiklehrkräften aus NRW, getrennt nach dem absolvierten Studiengang .....	92
<b>Abbildung 19:</b> Zusammenhang zwischen CK und PCK für die Schulformen .....	95
<b>Abbildung 20:</b> Zusammenhang zwischen CK und PCK bzw. PK und PCK für alle untersuchten Physiklehrkräfte, getrennt nach der Schulform, an der aktuell unterrichtet wird.....	136

## 9.2 TABELLENVERZEICHNIS

<b>Tabelle 1:</b> Korrelationen zwischen den Dimensionen des Professionswissens .....	15
<b>Tabelle 2:</b> Wachstum des Professionswissens in den Dimensionen in und zwischen den Ausbildungsstufen.....	18
<b>Tabelle 3:</b> Schulformunterschiede in den Ausbildungsstufen.....	19
<b>Tabelle 4:</b> Kurzbeschreibung und Beispiele für die drei Wissensbereiche aus den Bereichen PCK und CK. Die Fragen entsprechen nicht den im Testinstrument verwendeten Items.....	26
<b>Tabelle 5:</b> Facetten von PCK nach Park & Oliver (2008), angepasst und ergänzt um weitere Studien.....	32
<b>Tabelle 6:</b> Facetten generischen pädagogischen Wissens nach Kunter et al., 2011, .... S. 39.....	35
<b>Tabelle 7:</b> H2 - Erwartete Korrelationen der unabhängigen Variablen mit CK und PCK .....	39
<b>Tabelle 8:</b> Anzahl der Items .....	46
<b>Tabelle 9:</b> Bearbeitungszeiten .....	50
<b>Tabelle 10:</b> Anzahl, Alter und Geschlecht der Versuchspersonen. Fehlende Prozentangaben zum Geschlecht beruhen auf fehlenden Angaben.....	52
<b>Tabelle 11:</b> Anzahl der Versuchspersonen, aufgeteilt nach der Schulform, an der sie zur Zeit der Untersuchung unterrichteten und der studierten Schulform .....	52
<b>Tabelle 12:</b> Zusammenhang zwischen studierter und aktueller Schulform für alle Physiklehrkräfte, die beide Informationen angegeben haben.....	52
<b>Tabelle 13:</b> Besuchte Fortbildungen und Anzahl der besuchten Fortbildungen der Physiklehrkräfte .....	54
<b>Tabelle 14:</b> Überblick über die sechs Modelle, die gegeneinander getestet werden .	70
<b>Tabelle 15:</b> Freiheitsgrade, Final Deviance und die Anzahl der Items, die nicht Raschkonform sind für die sechs möglichen Modelle. Oben alle Physiklehrkräfte, unten die Physiklehrkräfte der 2. Studie.....	70
<b>Tabelle 16:</b> Modelle, die ohne $PK_{KP}$ verglichen werden .....	71
<b>Tabelle 17:</b> Freiheitsgrade, Final Deviance und Reliabilität der vier möglichen Modelle (Siehe Tabelle 16) unter Ausschluss nicht passender Items. ....	72
<b>Tabelle 18:</b> Reliabilität des CK-, PCK- und $PK_D$ -Tests in verschiedenen Personengruppen.....	73
<b>Tabelle 19:</b> Testleistung der Physiklehrkräfte der verschiedenen Schulformen .....	76
<b>Tabelle 20:</b> Testleistung der Physiklehrkräfte und der Lehrkräfte anderer Fächer. Alle einbezogenen Lehrkräfte unterrichten am Gymnasium. ....	77
<b>Tabelle 21:</b> Testleistung der Physiklehrkräfte am Gymnasium und der Diplom-Physikerinnen und -Physiker.....	78
<b>Tabelle 22:</b> Testleistung der Physiklehrkräfte am Gymnasium und der Diplom-Physikerinnen und -Physiker, die weder an einer Universität tätig sind noch lehren .	78
<b>Tabelle 23:</b> Testleistungskorrelationen nach Pearson für CK, PCK (N=279) und PK (N=186).....	79

<b>Tabelle 24:</b> Testleistungskorrelationen nach Pearson für Physiklehrkräfte. Die Korrelationen zwischen GY und NGY unterscheiden sich nicht signifikant ( $p > .2$ ). .....	81
<b>Tabelle 25:</b> Lineare Regression für Physiklehrkräfte (N=279): Einfluss von CK und PK auf PCK, $R^2 = .27$ , $F = 34$ .....	82
<b>Tabelle 26:</b> Lineare Regression für Physiklehrkräfte, die am Gymnasium unterrichten (N=149): Einfluss von CK und PK auf PCK, $R^2 = .20$ , $F = 18$ .....	82
<b>Tabelle 27:</b> Lineare Regression: Einfluss demographischer Variablen auf CK (N=279), $R^2_{CK} = .40$ , $F_{CK} = 23$ .....	87
<b>Tabelle 28:</b> Lineare Regression: Einfluss demographischer Variablen auf PCK (N=279), $R^2_{PCK} = .40$ , $F_{PCK} = 30$ .....	87
<b>Tabelle 29:</b> Lineare Regression: Einfluss demographischer Variablen auf CK (N=279), $R^2_{CK} = .35$ , $F_{CK} = 47$ .....	89
<b>Tabelle 30:</b> Lineare Regression: Einfluss demographischer Variablen auf PCK (N=279), $R^2_{PCK} = .34$ , $F_{PCK} = 44$ .....	89
<b>Tabelle 31:</b> Anzahl der Personen in den Gruppen .....	91
<b>Tabelle 32:</b> CK und PCK der Versuchspersonen aus NRW .....	91
<b>Tabelle 33:</b> Testleistungsunterschiede in CK und PCK für Personen, die ein gymnasiales Lehramt studiert haben bzw. studieren, Post-Hoc-Tests (GT2 nach Hochberg) .....	93
<b>Tabelle 34:</b> Testleistungskorrelationen nach Pearson zwischen CK und PCK für Versuchspersonen aus NRW .....	94
<b>Tabelle 35:</b> Unterschiede zwischen den Testleistungskorrelationen von CK zu PCK ...	94
<b>Tabelle 36:</b> Richtung der erwarteten und gemessenen Korrelationen der unabhängigen Variablen mit CK und PCK .....	107
<b>Tabelle 37:</b> Eigenschaften der CK- und PCK-Items der zweiten Studie in der Reihenfolge im Testheft .....	134
<b>Tabelle 38:</b> Latente Korrelationen für alle Physiklehrkräfte .....	
(Ausgabe aus ConQuest) .....	137
<b>Tabelle 39:</b> Testleistungskorrelationen nach Pearson und Unterschiede für Physiklehrkräfte getrennt nach aktueller Schulform .....	137
<b>Tabelle 40:</b> Korrelationen von CK, PCK und PK sowie der unabhängigen Variablen zu CK und PCK. Für dichotome Variablen Punkt-Biserial-Korrelationen, sonst Korrelation nach Pearson .....	138
<b>Tabelle 41:</b> Lineare Regression: Einfluss demographischer Variablen auf CK (N=279), $R^2_{CK} = .46$ , $F_{CK} = 30$ .....	139
<b>Tabelle 42:</b> Lineare Regression (backwards): Einfluss demographischer Variablen auf PCK (N=279), $R^2_{PCK} = .33$ , $F_{PCK} = 23$ .....	140
<b>Tabelle 43:</b> Kolmogorov-Smirnov-Test, Schiefe und Kurtosis in kleinen Gruppen ....	141
<b>Tabelle 44:</b> CK und PCK der Physiklehrkräfte mit Diplom in NRW .....	142
<b>Tabelle 45:</b> CK und PCK der Physiklehrkräfte mit Diplom .....	142
<b>Tabelle 46:</b> CK, PCK und PK der Physiklehrkräfte aus NRW mit einem nicht-gymnasialen Lehramtsstudium unterschieden danach, ob sie Physik studiert haben oder fachfremd unterrichten .....	143
<b>Tabelle 47:</b> Testleistungsunterschiede in CK für Physiklehrkräfte, Post-Hoc-Tests (GT2 nach Hochberg) .....	144

<b>Tabelle 48:</b> Testleistungsunterschiede in PCK für Physiklehrkräfte, Mann-Whitney-Tests .....	144
<b>Tabelle 49:</b> Testleistungsunterschiede in CK für Lehrkräfte im Vorbereitungsdienst, Post-Hoc-Tests (GT2 nach Hochberg) .....	145

### 9.3 LITERATURVERZEICHNIS

- Abell, S. K. (2007). Research on science teachers' knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 1105–1149). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Anderson, L., Evertson, C., & Brophy, J. (1979). An Experimental Study of Effective Teaching in First-grade Reading Groups. *The Elementary School Journal*, 79(4), 193–223.
- Banilower, E. & Smith, S. (2012). *PCK Summit - Expanded Paper*. Verfügbar unter [pck-summit.bsccs.org/sites/default/files/Smith Banilower EP.pdf](http://pck-summit.bsccs.org/sites/default/files/Smith%20Banilower%20EP.pdf)
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, (9), 469–520.
- Baumert, J., & Kunter, M. (2013). Professionelle Kompetenz von Lehrkräften. In I. Gogolin, H. Kuper, H.-H. Krüger, & J. Baumert (Eds.), *Stichwort: Zeitschrift für Erziehungswissenschaft* (pp. 277–337). Wiesbaden: Springer Fachmedien Wiesbaden.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, 47(1), 133–180.
- Baxter, J. A., & Lederman, N. G. (1999). Assessment and measurement of pedagogical content knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 147–161). Dordrecht: Kluwer.
- Bayrisches Staatsministerium für Unterricht und Kultus. *Staatliche Lehrerfortbildung*. Verfügbar unter <http://www.km.bayern.de/lehrer/fort-und-weiterbildung/staatliche-lehrerfortbildung.html>
- Bezirksregierung Düsseldorf. *Die wichtigsten Dienstpflichten im Überblick*. Verfügbar unter <http://www.brd.nrw.de/wirueberuns/Disziplinarangelegenheiten/Dienstpflichten.html#fortbildung>
- Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.). (2008). Professionelle Kompetenz angehender Lehrerinnen und Lehrer.: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare. Münster: Waxmann.
- Blömeke, S., Suhl, U., Kaiser, G., Felbrich, A., Schmotz, C., & Lehmann, R. (2010). Lerngelegenheiten und Kompetenzerwerb angehender Mathematiklehrkräfte im internationalen Vergleich. *Unterrichtswissenschaft*, 38(1), 29–50.
- Bond, T., & Fox, C. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.
- Borowski, A., Kirschner, S., Liedtke, S., & Fischer, H. E. (2011). Vergleich des Fachwissens von Studierenden, Referendaren und Lehrenden in der Physik. *PhyDid A – Physik und Didaktik in Schule und Hochschule.*, 10(1), 1–9.

- Borowski, A., Neuhaus, B. J., Tepner, O., Wirth, J., Fischer, H. E., Leutner, D., ... (2010a). Professionswissen von Lehrkräften in den Naturwissenschaften (ProwiN) – Kurzdarstellung des BMBF-Projekts. *Zeitschrift für Didaktik der Naturwissenschaften*, *16*, 341–349.
- Borowski, A., Olszewski, J., & Fischer, H. E. (2010b). Fachdidaktisches Wissen von Physikreferendaren. *Der mathematische und naturwissenschaftliche Unterricht*, *63*(5), 260–263.
- Borowski, A., & Riese, J. (2010). Physikalisch-fachdidaktisches Wissen: Was kommt in der Praxis an? *Praxis der Naturwissenschaften - Physik in der Schule*, *59*(5), 5–8.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6th ed.). Heidelberg: Springer Medizin.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4th ed.). Heidelberg: Springer.
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Dubberke, T., ... (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? *Zeitschrift für Erziehungswissenschaft*, *9*(4), 521–544.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Aufl.). Psychologie. München: Pearson Studium.
- Carstensen, C. H. (2000). *Mehrdimensionale Testmodelle mit Anwendungen aus der pädagogisch-psychologischen Diagnostik*. Kiel: IPN.
- Cauet, E., Borowski, A., & Fischer, H. (2012). Zusammenhänge zwischen dem Professionswissen von Physiklehrkräften, Unterrichtshandeln und Schülerleistung –Projektskizze. In S. Bernholt (Ed.), *Konzepte fachdidaktischer Strukturierung für den Unterricht. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Oldenburg 2011*. Berlin: Lit.
- Cauet, E., Liepertz, S., Kirschner, S., Borowski, A., & Fischer, H. E. (2013). *Professionswissen von Physiklehrkräften und Schülerleistung*, München.
- Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical Content Knowing: An Integrative Model for Teacher Preparation. *Journal of Teacher Education*, *44*(4), 263–272. doi:10.1177/0022487193044004004
- Deutsche Physikalische Gesellschaft. (2004). *Bachelor- und Master-Studiengänge im Fach Physik*. Verfügbar unter [http://www.dpg-physik.de/veroeffentlichung/stellungnahmen/mem\\_bild\\_2004\\_01.html](http://www.dpg-physik.de/veroeffentlichung/stellungnahmen/mem_bild_2004_01.html)
- Dollny, S. (2011). *Entwicklung und Evaluation eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Chemielehrkräften*. Berlin: Logos.

- Fernandez-Balboa, J. M., & Stiehl, J. (1995). The generic nature of pedagogical content knowledge among college professors. *Teaching and Teacher Education*, 11(3), 293–306.
- Field, A. (2009). *Discovering Statistics using SPSS* (3rd ed.). London: Sage Publications.
- Fischer, H. E., Borowski, A., Leutner, D., Neuhaus, B. J., Sandmann, A., Sumfleth, E., ... (2009). *ProwiN Zwischenbericht für den Zeitraum 01.06.2009 bis 31.12.2009*.
- Fischer, H. E., Borowski, A., Leutner, D., Neuhaus, B. J., Sandmann, A., Sumfleth, E., ... (2010). *ProwiN Zwischenbericht für den Zeitraum 01.01.2010 bis 31.12.2010*.
- Fischer, H. E., Borowski, A., Leutner, D., Neuhaus, B. J., Sumfleth, E., Tepner, O., ... (2011). *Professionswissen in den Naturwissenschaften (ProwiN) - Videostudie: Fortsetzungsantrag*.
- Fischer, H. E., Borowski, A., & Tepner, O. (2012). Professional knowledge of science teachers. In B. J. Fraser, K. Tobin, & C. McRobbie (Eds.), *Second International Handbook of Science Education* (pp. 435–448). New York: Springer.
- Fischler, H. (2008). Physikdidaktisches Wissen und Handlungskompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 51–55.
- Geddis, A. N. (1993). Transforming subject-matter knowledge: the role of pedagogical content knowledge in learning to reflect on teaching. *International Journal of Science Education*, 15(6), 673–683.
- Gess-Newsome, J. (1999). Pedagogical content knowledge: an introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 3–17). Dordrecht: Kluwer.
- Gess-Newsome, J. (2011). Pedagogical Content Knowledge. In J. Hattie & E. Anderman (Eds.), *International handbook of student achievement*. New York: Routledge.
- Gess-Newsome, J., Taylor, J., Carlson, J., Gardner, A., Wilson, C., & Stuhlsatz, M. (In review). Impact of educative materials and professional development on teachers' professional knowledge, practice, and student achievement. *American Educational Research Journal*.
- Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. Professional development and practice series. New York: Teachers College Press.
- Hartig, J., Jude, N., & Wagner, W. (2008). Methodische Grundlagen der Messung und Erklärung sprachlicher Kompetenzen. In E. Klieme (Ed.), *Beltz Pädagogik. Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (pp. 34–54). Weinheim, Basel: Beltz. Verfügbar unter [http://www.pedocs.de/volltexte/2010/3152/pdf/Hartig\\_Jude\\_Wagner\\_Methodische\\_Grundlagen\\_D\\_A.pdf](http://www.pedocs.de/volltexte/2010/3152/pdf/Hartig_Jude_Wagner_Methodische_Grundlagen_D_A.pdf)
- Hashweh, M. (1987). Effects of subject matter knowledge in the teaching of biology and physics. *Teaching & Teacher Education*, 3(2), 109–120.

- Hattie, John A. C. (2009). Visible learning: A synthesis of meta-analyses relating to achievement. London: Routledge.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141.
- IT.NRW: Information und Technik Nordrhein-Westfalen - Geschäftsbereich Statistik. (2013). *Prozentsatz weiblicher Physiklehrkräfte in NRW*.
- Jüttner, M. (2013). Entwicklung, Evaluation und Validierung eines Fachwissenstests und eines fachdidaktischen Wissenstests für die Erfassung des Professionswissens von Biologielehrkräften: Dissertation der Fakultät für Biologie der Ludwigs-Maximilians-Universität München.
- Jüttner, M., Boon, W., Park, S., & Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability*, 25(1), 45–67. doi:10.1007/s11092-013-9157y
- Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential for progress. *Studies in Science Education*, 45(2), 169–204.
- Kirschner, S., Borowski, A., & Fischer, H. E. (2011a). ProwiN-Test zum fachdidaktischen Wissen von Physiklehrkräften. In A. Borowski, H. E. Fischer, M. Jüttner, S. Kirschner, B. J. Neuhaus, E. Sumfleth, ... (Eds.), *ProwiN-Testinstrumente*. Essen: Universität Duisburg-Essen.
- Kirschner, S., Borowski, A., & Fischer, H. E. (2011b). ProwiN-Test zum Fachwissen von Physiklehrkräften. In A. Borowski, H. E. Fischer, M. Jüttner, S. Kirschner, B. J. Neuhaus, E. Sumfleth, ... (Eds.), *ProwiN-Testinstrumente*. Essen: Universität Duisburg-Essen.
- Kirschner, S., Borowski, A., & Fischer, H. E. (2013). *Die Modellierung des Professionswissens von Physiklehrkräften*. Verfügbar unter [http://www.gdcp.de/images/tagungsbaende/GDCP\\_Band33.pdf](http://www.gdcp.de/images/tagungsbaende/GDCP_Band33.pdf)
- Kirschner, S., & Cauet, E. Professionswissen von Physiklehrkräften: Modellierung des Zusammenhangs von Wissen und Handeln auf Seite der Lehrenden und Lernenden, *in Vorbereitung*.
- Kirschner, S., Wlotzka, U., Borowski, A., & Fischer, H. E. (2011). Das Professionswissen von Physik Lehrern - Pilotierung und Validierung. In D. Höttecke (Ed.), *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Potsdam 2010* (pp. 114–116). Berlin: Lit.
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2012). Teachers' Content Knowledge and Pedagogical Content Knowledge: The

- Role of Structural Differences in Teacher Education. *Journal of Teacher Education*, 64(1), 90–106. doi:10.1177/0022487112460398
- Kolbe, F.-U. (2004). Das Verhältnis von Wissen und Handeln. In S. Blömeke, P. Reinhold, & G. Tulodziecki (Eds.), *Handbuch Lehrerbildung* (pp. 206–232). Bad Heilbrunn: Klinkhardt.
- Krauss, S., Neubrand, M., Blum, W., & Baumert, J. (2008). The Professional Knowledge of German Secondary Mathematics Teachers: Investigations in the Context of the COACTIV Project. Verfügbar unter <http://tsg.icme11.org/document/get/397>
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M., & Jordan, A. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29(3/4), 223–258.
- Kröger, J., Neumann, K., & Petersen, S. (2013). *Messung professioneller Kompetenz im Fach Physik*. Verfügbar unter [http://www.gdcp.de/images/tagungsbaende/GDCP\\_Band33.pdf](http://www.gdcp.de/images/tagungsbaende/GDCP_Band33.pdf)
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann. Verfügbar unter <http://www.worldcat.org/oclc/704822581>
- Kunter, M., & Pohlmann, B. (2009). Lehrer. In E. Wild & J. Möller (Eds.), *Einführung in die Pädagogische Psychologie* (pp. 261–282). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
- Linacre, J. M. (2010). Winsteps® Rasch Measurement. Beaverton, Oregon: Winsteps.com.
- Linacre, J. M. (2011a). A User's Guide to FACETS: Rasch-Model Computer Programs (3.68.0th ed.): Winsteps.com.
- Linacre, J. M. (2011b). A User's Guide to W I N S T E P S ® M I N I S T E P: Rasch-Model Computer Programs (3.72.0th ed.): Winsteps.com.
- Lipowsky, F. (2004). Was macht Fortbildungen für Lehrkräfte erfolgreich? Befunde der Forschung und mögliche Konsequenzen für die Praxis. *Die Deutsche Schule*, 94(4), 462–479.
- Loughran, J., Berry, A., & Mulhall, P. (2006). *Understanding and Developing Science Teachers' Pedagogical Content Knowledge*. Rotterdam: Sense Publishers.
- Magnusson, S., Krajcik J., & Borko H. (1999). Nature, Sources, and Development of Pedagogical Content Knowledge for Science teacher. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge*. Dordrecht: Kluwer.
- Marks, R. (1990). Pedagogical Content Knowledge: From a Mathematical Case to a Modified Conception. *Journal of Teacher Education*, 41(3), 3–11.

- Matzdorf, R. (2012). Mehr Physikstudierende als je zuvor. *Physik Journal*, 11(8/9), 29–33. Verfügbar unter [http://www.dpg-physik.de/veroeffentlichung/physik\\_journal/artikel\\_pj/physikstudium/physikstudium\\_2012.pdf](http://www.dpg-physik.de/veroeffentlichung/physik_journal/artikel_pj/physikstudium/physikstudium_2012.pdf)
- Gesetz über die Ausbildung für Lehrämter an öffentlichen Schulen, Ministerium für Schule und Weiterbildung des Landes Nordrhein- Westfalen 13.11.2012.
- Mutke, S., & Tepner, O. (15.4.2013). Einfluss des Referendariats auf die Entwicklung des Professionswissens im Fach Chemie, Essen.
- Mutke, S. & Tepner, O. (2013). *Entwicklung des Professionswissens im Fach Chemie in den ersten sechs Monaten des Referendariats*. Verfügbar unter [http://www.gdcp.de/images/tagungsbaende/GDCP\\_Band33.pdf](http://www.gdcp.de/images/tagungsbaende/GDCP_Band33.pdf)
- Neumann, K., Fischer, H. E., Labudde, P., & Viiri, J. (2010). Physikunterricht: Deutschland, Finnland und die Schweiz im Vergleich. In D. Höttecke (Ed.), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Dresden 2009* (pp. 383–385). Berlin: Lit.
- OECD. (2006). PISA 2006: Naturwissenschaftliche Kompetenzen für die Welt von morgen. Kurzzusammenfassung. Verfügbar unter <http://www.oecd.org/pisa/39731064.pdf>
- Olszewski, J. (2010). The impact of physics teachers' Pedagogical content knowledge on teacher action and student outcomes. Berlin: Logos.
- Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 293–316.
- Park, S., & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49(7), 922–941. doi:10.1002/tea.21022
- Park, S., & Oliver, S. J. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38(3), 261–284.
- Park, S. & Suh, J. (2012). Are there attributes of PCK and teachers' professional knowledge that are similar across multiple papers? Verfügbar unter <http://pcksummit.bsccs.org/node/60>
- Park, S., Suh, J., Friedrichsen, P., Sickel, A., Lannin, J., Berry, A., ... (2012). *Are there attributes of PCK and teachers' professional knowledge that are similar across multiple papers?* Verfügbar unter <http://pcksummit.bsccs.org/node/60>
- Pollender, T., & Tepner, O. (2012). Kommunikationsmuster von Chemielehrkräften. In S. Bernholt (Ed.), *Konzepte fachdidaktischer Strukturierung für den Unterricht*. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Oldenburg 2011 (pp. 625–627). Berlin: Lit.

## Verzeichnisse

- Prenzel, M., & Deutsches PISA-Konsortium. (2008). PISA 2006 in Deutschland: Die Kompetenzen der Jugendlichen im dritten Ländervergleich / PISA '06. Münster: Waxmann.
- Reusser, K., & Pauli, C. (2003). Mathematikunterricht in der Schweiz und in weiteren sechs Ländern. Bericht über die Ergebnisse einer internationalen und schweizerischen Video-Unterrichtsstudie. Universität Zürich: Pädagogisches Institut.
- Riese, J. (2009). Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften. Berlin: Logos.
- Riese, J., & Reinhold, P. (2012). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen. *Zeitschrift für Erziehungswissenschaft*, 15(1), 111–143. doi:10.1007/s11618-012-0259-y
- Rollnick, M., Bennett, J., Rhemtula, M., Dharsey, N., & Ndlovu, T. (2008). The Place of subject matter knowledge in pedagogical content knowledge: A case study of South African teachers teaching the amount of substance and chemical equilibrium. *International Journal of Science Education*, 30(10), 1365–1387.
- Rost, J. (2004). Lehrbuch Testtheorie - Testkonstruktion. Aus dem Programm Huber: Psychologie-Lehrbuch. Bern [u.a.]: Huber.
- Schlagmüller, M., & Schneider, W. (2007). Würzburger Lesestrategie-Wissenstest für die Klassen 7-12 (WLST 7-12). In M. Hasselhorn, H. Marx, & W. Schneider (Eds.), *Deutsche Schultests*. Göttingen: Hogrefe.
- Schmader, T. (2002). Gender Identification Moderates Stereotype Threat Effects on Women's Math Performance. *Journal of Experimental Social Psychology*, 38(2), 194–201. doi:10.1006/jesp.2001.1500
- Schmelzing, S. (2010). Das fachdidaktische Wissen von Biologielehrkräften: Konzeptualisierung, Diagnostik, Struktur und Entwicklung im Rahmen der Biologielehrerbildung. Berlin: Logos.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK]. (2005). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK]. (2010). Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung: (Beschluss der Kultusministerkonferenz vom 16.10.2008 i. d. F. vom 16.09.2010). Verfügbar unter [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2008/2008\\_10\\_16-Fachprofile-Lehrerbildung.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf)
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.

- Shulman, L. S. (1987). Knowledge and teaching of the new reform. *Harvard Educational Review*, 57, 1–22.
- Smith, D. C., & Neale, D. C. (1989). The construction of subject matter knowledge in primary science teaching. *Teaching and Teacher Education*, 5(1), 1–20.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.
- Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education*, 4(2), 99–110.
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., ... (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 7–28.
- Terhart, E. (2002). *Standards für die Lehrerbildung: Eine Expertise für die Kultusministerkonferenz*. Verfügbar unter [http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-1151/Standards\\_fuer\\_die\\_Lehrerbildung\\_Eine\\_Expertise\\_fuer\\_die\\_Kultusministerkonferenz.pdf](http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-1151/Standards_fuer_die_Lehrerbildung_Eine_Expertise_fuer_die_Kultusministerkonferenz.pdf)
- Tesch, M. (2005). *Das Experiment im Physikunterricht: Didaktische Konzepte und Ergebnisse einer Videostudie*. Berlin: Logos.
- Thillmann, H. (12.12.2011). PK-Auswertung (E-Mail).
- van Driel, J. (12.2.2013). *Professional development of science teachers*. Plenary session nwu-Winterschool 2013, Essen.
- van Driel, J., Jong, O. de, & Verloop, N. (2002). The development of preservice chemistry teachers' PCK. *Science Education*, 86(4), 572–590.
- van Driel, J., Verloop, N., & Vos, W. de. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, 35(6), 673–695.
- Van Vorst, H. (2013). *Kontextmerkmale und ihr Einfluss auf das Schülerinteresse im Fach Chemie*. Berlin: Logos.
- Veal, W. R., & MaKinster, J. G. (1999). Pedagogical Content Knowledge Taxonomies. *Electronic Journal of Science Education*, 3(4). Verfügbar unter <http://wolfweb.unr.edu/homepage/crowther/ejse/vealmak.html>
- Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969. doi:10.1037/a0025125
- Walter, O. (2005). *Kompetenzmessung in den PISA-Studien. Simulationen zur Schätzung von Verteilungsparametern und Reliabilitäten*. Lengerich: Pabst Science Publishers.

## Verzeichnisse

- 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. (1996). Untersuchungen von Lernprozessen beim Lernen Newtonscher Dynamik im Anfangsunterricht. *Naturwissenschaften und Technik - Didaktik im Gespräch*: Vol. 25. Münster, Frankfurt (Main): Lit.
- Woitkowski, D., Riese, J., & Reinhold, P. (2011). Modellierung fachwissenschaftlicher Kompetenz angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 289–313.
- Wu, M., Adams, R., & Haldane, S. (2007). ConQuest. Berkeley: Australian Council for Educational Research.
- Wu, M., Adams, R., & Wilson, M. (1998). *Acer Conquest - Generalised Item Response Modelling Software*. Melbourne: ACER Press.

# ANHANG

10.1	Liste der CK- und PCK-Items.....	134
10.2	Liste der gelöschten PK-Items.....	135
10.3	Zusammenhang von CK, PCK und PK .....	136
10.4	Korrelationen von Personenmerkmalen zu CK und PCK .....	138
10.5	Ergänzende Regressionen.....	139
10.6	Prüfung der Normalverteilung in kleinen Stichproben (N<30) .....	141
10.7	Quereinsteiger .....	142
10.8	Fachfremder Unterricht .....	143
10.9	Mittelwertunterschiede zwischen den Studiengängen.....	144

**10.1 LISTE DER CK- UND PCK-ITEMS****Tabelle 37:** Eigenschaften der CK- und PCK-Items der zweiten Studie in der Reihenfolge im Testheft

ID	Inhalt	Format	Objektivitätskriterien	Fit-Kriterium
PCK_S020	Warum Experimente	Halboffen		
PCK_S230	Warum Einheiten	Halboffen		
PCK_S180	Stundenkilometer	Halboffen	nicht erfüllt	
PCK_0261	Lok	Halboffen		
PCK_0051	Diagramm 1	Halboffen		
PCK_0052	Diagramm 2	Halboffen		
PCK_0151	Flugbahn 1	Zeichnung		
PCK_0152	Flugbahn 2	Halboffen		
PCK_0071	Lampe 1	Halboffen		
PCK_0072	Lampe 2	Halboffen		
PCK_0031	Anknüpfen an Schülervorstellungen	Halboffen		
PCK_0180	Schülervorstellungen zu Geschwindigkeit	Halboffen		
PCK_0280	Wirkung von Kraft	Halboffen		
PCK_0230	Zusammenhänge	Halboffen	nicht erfüllt	
PCK_0320	Zeichnung Kraft	Halboffen		
PCK_0080	Wasserrakete	Halboffen		
NAW1B	Zunge 1	Geschlossen		
NAW1B	Zunge 2	Geschlossen		
NAW1E	Chemie 1	Geschlossen		
NAW2S	Chemie 2	Geschlossen		
NAW3H	Chemie 3	Geschlossen		
PCK_0040	Stundenfortsetzung Experiment	Geschlossen		nicht erfüllt
CK_1150	Rutsche	Kombiniert		nicht erfüllt
CK_1512	Schuss	Halboffen		
CK_1460	Rennstrecke	Zeichnung		
CK_1160	Flugzeug Wind	Kombiniert		
CK_1240	Puk	Geschlossen		nicht erfüllt
CK_1130	3. Axiom (Newton)	Kombiniert	nicht erfüllt	
CK_1490	E-Lehre	Geschlossen		
CK_1450	Hebel	Halboffen		
CK_1410	Ampel	Halboffen		
CK_1300	Beschleunigung	Geschlossen		nicht erfüllt
CK_1470	Looping	Halboffen		

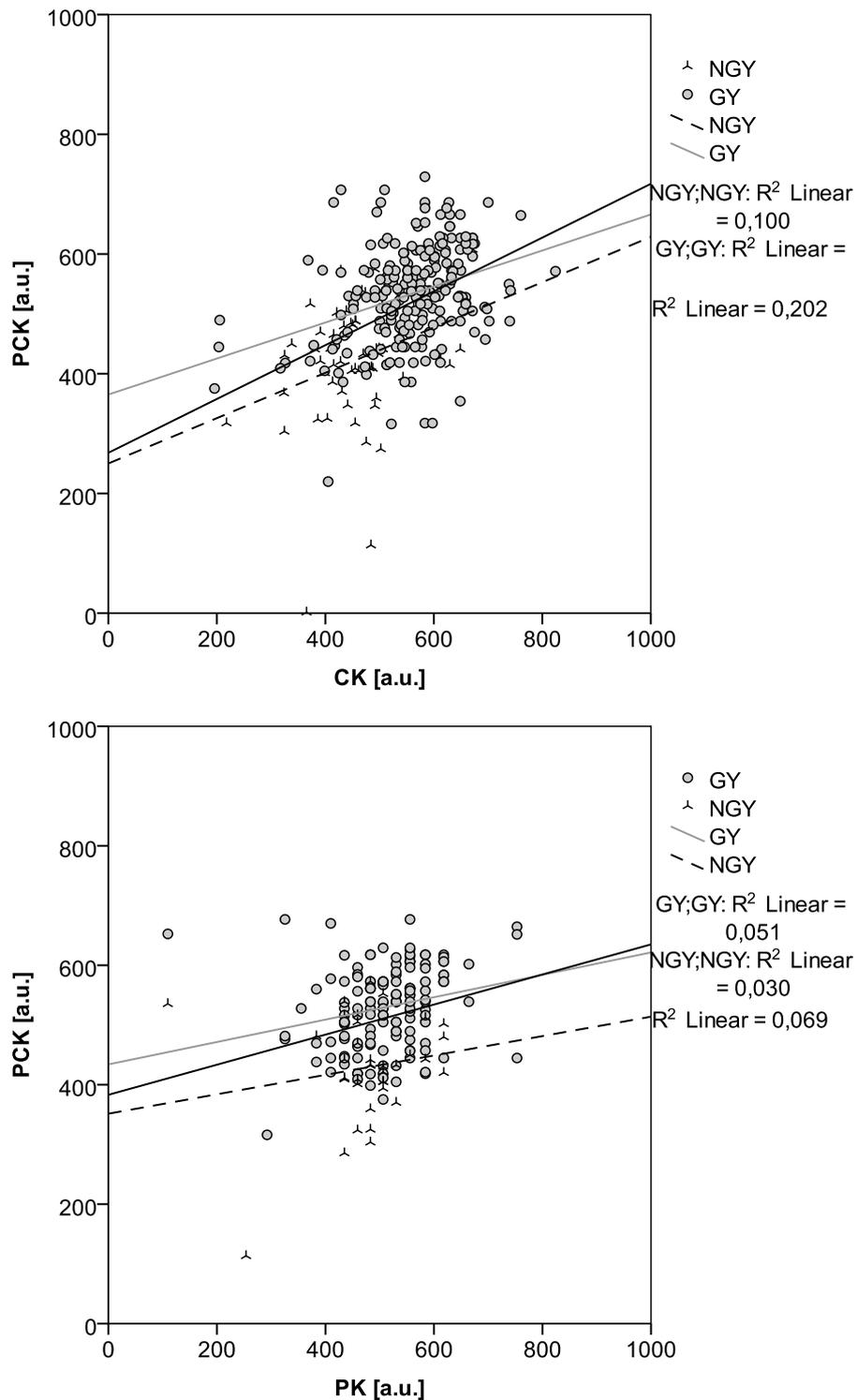
CK_1290	Schaukel	Geschlossen
CK_1180	Kepler	Kombiniert
CK_1140	Pendel	Halboffen
CK_1220	LKW	Geschlossen

## 10.2 LISTE DER GELÖSCHTEN PK-ITEMS

PK-Items, die aufgrund von Misfit ( $MnSq < 2.0$ ) ausgeschlossen werden mussten

pkd1a_rec	pkd4b_rec	pkd8a_rec
pkd1b_rec	pkd4c_rec	pkd8c_rec
pkd1c_rec	pkd4e_rec	pkd8d_rec
pkd1d_rec	pkd5a_rec	pkd10e_rec
pkd1e_rec	pkd5b_rec	pkd11a_rec
pkd2a_rec	pkd5d_rec	pkd11d_rec
pkd2b_rec	pkd5e_rec	
pkd3c_rec	pkd6a_rec	PK <sub>KP</sub> ist in dieser Analyse
pkd3d_rec	pkd6b_rec	nicht Rasch-modellier-
pkd3e_rec	pkd6c_rec	bar.
pkd4a_rec	pkd6f_rec	

### 10.3 ZUSAMMENHANG VON CK, PCK UND PK



**Abbildung 20:** Zusammenhang zwischen CK und PCK bzw. PK und PCK für alle untersuchten Physiklehrkräfte, getrennt nach der Schulform, an der aktuell unterrichtet wird.

Eingezeichnet sind die Regressionsgraden der beiden Subgruppen und der gesamten Stichprobe. Für die gesamte Stichprobe gilt  $M=500$ ,  $SD=100$  für CK, PCK und PK.

**Tabelle 38:** Latente Korrelationen für alle Physiklehrkräfte (Ausgabe aus ConQuest)

	Latente Korrelation
CK zu PCK	.595
PCK zu PK	.373
CK zu PK	.230

**Tabelle 39:** Testleistungskorrelationen nach Pearson und Unterschiede für Physiklehrkräfte getrennt nach aktueller Schulform

		N	Korrelation r	Signifi- kanz p	Unterschiede zwischen Korrelationen		
GY	CK zu PCK	216	.33	<.001	zwei- seitig, p=.293	zwei- seitig, p=.500	zwei- seitig, p=.074
	PCK zu PK <sub>D</sub>	149	.23	<.01			
	CK zu PK <sub>D</sub>	149	.15	.066			
NGY	CK zu PCK	62	.32	<.05	zwei- seitig, p=.483	zwei- seitig, p=.314	zwei- seitig, p=.066
	PCK zu PK <sub>D</sub>	36	.17	.314			
	CK zu PK <sub>D</sub>	36	-.07	.672			

GY Unterrichtet zurzeit am Gymnasium

NGY Unterrichtet zurzeit an einer anderen Schulform als dem Gymnasium

Die in der Validierung für alle Physiklehrkräfte untersuchten Unterschiede zwischen den Korrelationen sind schwächer, wenn die beiden Teilgruppen, GY und NGY, getrennt untersucht werden. In der Tendenz bleibt zu erkennen, dass der Zusammenhang zwischen CK und PCK stärker ist als der zwischen CK und PK.

## 10.4 KORRELATIONEN VON PERSONENMERKMALEN ZU CK UND PCK

**Tabelle 40:** Korrelationen von CK, PCK und PK sowie der unabhängigen Variablen zu CK und PCK. Für dichotome Variablen Punkt-Biserial-Korrelationen, sonst Korrelation nach Pearson.

Unabhängige Variable	Korrelation zu CK			Korrelation zu PCK		
	r	p	R <sup>2</sup>	r	p	R <sup>2</sup>
CK	1	<.001	100	.45	<.001	20
PCK	.45	<.001	20	1	<.001	100
PK	.17	<.05	3	.27	<.001	7
Geschlecht	.33	<.001	11	.18	<.001	3
Alter	.03	.59	0	-.26	<.001	7
NRW	-.19	<.01	4	-.01	.819	0
Bayern	.22	<.001	5	.06	.362	0
Anzahl an Jahren, in denen bereits Physik unterrichtet wurde	.23	<.01	5	-.04	.600	0
Jahre im Schuldienst	.14	.058	2	-.1	.184	1
Durchgängiger Besuch eines Physikkurses in der eigenen Oberstufenzeit	.50	<.001	25	.33	<.001	11
Studium der Physik/kein fachfremder Unterricht	.34	<.001	12	.17	<.001	3
Studium der Mathematik	.24	<.001	6	.09	.153	1
Belegter Studiengang gymnasiales Lehramt	.34	<.001	12	.30	<.001	9
Belegter Studiengang nicht-gymnasiales Lehramt	-.34	<.001	12	-.42	<.001	18
Belegter Studiengang Diplom	-.09	.15	1	.05	.41	0
Unterrichtet am Gymnasium	.40	<.001	16	.46	<.001	21
Unterrichtet in der Unterstufe	-.13	.088	2	-.16	<.05	3
Unterrichtet in der Mittelstufe	.15	<.05	2	.12	.11	1
Unterrichtet in der Oberstufe	.49	<.001	24	.46	<.001	21
Häufigkeit Fortbildungsbesuch Physik	.25	<.01	6	.29	<.001	8
Häufigkeit pädagogischer Fortbildungen	-.24	<.01	6	-.01	.928	0
Häufigkeit Fachgruppenarbeit	.18	<.05	3	.16	<.05	3
Häufigkeit individueller Weiterbildung	.15	<.05	2	.26	<.001	7
Wochenstunden Physik, aktuell	.20	<.01	4	.23	<.01	5
r	Korrelationskoeffizient					
p	Signifikanz					
	R <sup>2</sup>	Aufklärung in %				

## 10.5 ERGÄNZENDE REGRESSIONEN

Tabelle 41 zeigt eine Regression für CK unter Einbezug von PCK und PK, also mit den folgenden Variablen:

- Unterrichtet am Gymnasium
- Fachfremder Unterricht
- Häufigkeit Fortbildungsbesuch Physik
- Häufigkeit pädagogischer Fortbildungen
- Häufigkeit individueller Weiterbildung
- Häufigkeit Fachgruppenarbeit
- Wochenstunden Physik
- Alter und Geschlecht
- CK und PK

**Tabelle 41:** Lineare Regression: Einfluss demographischer Variablen auf CK (N=279),  $R^2_{CK}=.46$ ,  $F_{CK}=30$

Abhängige Variable	Unabhängige Variable	Beta, unstandardisiert	SE	Beta, standardisiert	Signifikanz p
CK	PCK	0.3	0.1	0.28	<.001
	Unterrichtet fachfremd	-60	13	-0.28	<.001
	Unterrichtet am Gymnasium	52	16	0.22	<.01
	Geschlecht	45	12	0.21	<.001
	Häufigkeit pädagogischer Fortbildungen	-16	5	-0.16	<.01

SE Standardfehler  
Backwards Regression

## Anhang

Tabelle 42 zeigt eine Regression für PCK ohne CK und PK, also mit den folgenden Variablen:

- Unterrichtet am Gymnasium
- Fachfremder Unterricht
- Häufigkeit Fortbildungsbesuch Physik
- Häufigkeit pädagogischer Fortbildungen
- Häufigkeit individueller Weiterbildung
- Häufigkeit Fachgruppenarbeit
- Wochenstunden Physik
- Alter und Geschlecht

**Tabelle 42:** Lineare Regression (backwards): Einfluss demographischer Variablen auf PCK (N=279),  $R^2_{PCK}=.33$ ,  $F_{PCK}=23$

Abhängige Variable	Unabhängige Variable	Beta, unstandardisiert	SE	Beta, standardisiert	Signifikanz p
PCK	Unterrichtet am Gymnasium	97	13	0.46	<.001
	Alter	-2	0.5	-0.22	<.001
	Geschlecht	27	12	0.15	<.05
	Häufigkeit individueller Weiterbildung	14	5	0.16	<.05

SE Standardfehler  
Backwards Regression

## 10.6 PRÜFUNG DER NORMALVERTEILUNG IN KLEINEN STICHPROBEN (N<30)

**Tabelle 43:** Kolmogorov-Smirnov-Test, Schiefe und Kurtosis in kleinen Gruppen

Gruppe		K-S-Test			Schiefe			Kurtosis		
		D	df	p	S	SE <sub>S</sub>	Z <sub>S</sub>	K	SE <sub>K</sub>	Z <sub>K</sub>
Lehrkräfte anderer Fächer	PK	0.16	21	.175	0.53	0.50	1.08	-0.50	0.97	-0.51
Dipl.-Phys.	CK	0.13	22	.200	0.44	0.49	0.89	0.15	0.95	0.15
	PCK	0.11	22	.200	0.21	0.49	0.43	-.085	0.95	-0.09
NRW: NGY studiert, Physiklehrkräfte	CK	0.16	18	.200	0.04	0.54	0.08	-0.17	1.04	-0.17
	PCK	0.26	18	<.01	-1.63	0.54	-3.04	2.89	1.04	2.79
NRW: NGY studiert, LiV	CK	0.18	10	.200	0.52	0.69	0.75	-0.86	1.33	-2.58
	PCK	0.11	10	.200	-.031	0.69	-0.05	0.17	1.33	0.12
NRW: Diplom studiert, LiV	CK	0.21	12	.142	0.98	0.64	1.54	-0.01	1.23	-0.01
	PCK	0.20	12	.200	-.039	0.64	-0.06	0.10	1.23	0.08
Physiklehrkräfte mit Diplom in Physik	CK	0.21	19	<.05	1.05	0.52	1.99	0.09	1.01	0.09
	PCK	0.15	19	0.200	-0.54	0.52	1.02	-0.32	1.01	-0.32
Physiklehrkräfte mit Diplom in einem anderen Fach	CK	0.14	19	.200	-0.50	0.52	-0.95	1.70	1.01	1.67
	PCK	0.12	19	.200	-0.10	0.52	-0.18	-1.28	1.01	-1.26
NRW: Physiklehrkräfte mit Diplom in Physik	CK	0.19	14	.200	0.73	0.60	1.22	-0.54	1.15	-0.47
	PCK	0.13	14	.200	-0.02	0.54	0.04	-1.25	1.04	-1.20
NRW: Physiklehrkräfte mit Diplom in einem anderen Fach	CK	0.14	18	.200	-0.50	0.52	0.95	1.70	1.01	1.67
	PCK	0.13	18	.200	-0.10	0.52	0.18	-1.28	1.01	1.26

Dipl. Phys.	Diplom-Physikerinnen und Physiker
LiV	Lehrkräfte im Vorbereitungsdienst
D	Teststatistik
df	Freiheitsgrade
p	Signifikanz
S	Schiefe
K	Kurtosis
SE	Standardfehler
Z <sub>S</sub>	S/SE <sub>S</sub> ,
Z <sub>K</sub>	K/SE <sub>K</sub>
<b>Z</b>	signifikanter Wert

In kleinen Stichproben gilt  $Z > 1.96$  als signifikante Abweichung von der Normalverteilung mit  $p < .05$  (Field, 2009).

**10.7 QUEREINSTEIGER****Tabelle 44:** CK und PCK der Physiklehrkräfte mit Diplom in NRW

		Diplom in Physik	Diplom in einem anderen Fach
CK	N	14	18
	M	570	448
	SD <sub>M</sub>	52	94
	SE <sub>M</sub>	14	22
PCK	N	14	18
	M	555	491
	SD <sub>M</sub>	57	21
	SE <sub>M</sub>	15	88

N Stichprobengröße

M Mittelwert

SD<sub>M</sub> Standardabweichung des MittelwertsSE<sub>M</sub> Standardfehler des Mittelwerts

Da das Wissen beider Gruppen von Quereinsteigern in NRW normalverteilt ist (siehe Tabelle 43), kann der Gruppenunterschied mit t-Tests beleuchtet werden. Der Unterschied ist für CK und PCK signifikant (CK:  $t(30)=4.37$ ,  $p<.001$ ,  $d=1.67$ , PCK:  $t(29)=2.49$ ,  $p<.05$ ,  $d=1.64$ ). Lehrkräfte mit einem Diplom im Fach Physik schneiden besser ab als jene mit einem Diplom in einem anderen Fach.

**Tabelle 45:** CK und PCK der Physiklehrkräfte mit Diplom

		Diplom in Physik	Diplom in einem anderen Fach
CK	N	19	19
	M	570	459
	SD <sub>M</sub>	55	102
	SE <sub>M</sub>	13	23
PCK	N	19	19
	M	549	495
	SD <sub>M</sub>	58	87
	SE <sub>M</sub>	13	20

Konventionen siehe Tabelle 47

Für CK kann keine Normalverteilung angenommen werden, daher wird ein Mann-Whitney-Test durchgeführt. Der Unterschied ist signifikant ( $U=45$ ,  $p<.001$ , zweiseitig,  $d=1.41$ ). Für PCK kann ein t-Test berechnet werden. Der Unterschied ist signifikant ( $t(31)=2.25$ ,  $p<.05$ ,  $d=0.74$ ). Lehrkräfte mit einem Diplom im Fach Physik schneiden besser ab als jene mit einem Diplom in einem anderen Fach.

## 10.8 FACHFREMDER UNTERRICHT

**Tabelle 46:** CK, PCK und PK der Physiklehrkräfte aus NRW mit einem nicht-gymnasialen Lehramtsstudium unterschieden danach, ob sie Physik studiert haben oder fachfremd unterrichten

		Physik studiert	Keine Physik studiert (fachfremder Unterricht)
CK	N	13	5
	M	475	384
	SD <sub>M</sub>	63	58
	SE <sub>M</sub>	17	26
PCK	N	13	5
	M	371	401
	SD <sub>M</sub>	160	58
	SE <sub>M</sub>	44	26
PK <sub>D</sub>	N	11	5
	M	408	507
	SD <sub>M</sub>	122	38
	SE <sub>M</sub>	37	17

Konventionen siehe Tabelle 47

## 10.9 MITTELWERTUNTERSCHIEDE ZWISCHEN DEN STUDIENGÄNGEN

Für Physiklehrkräfte ergibt eine ANOVA einen signifikanten Einfluss der Ausbildungswege auf CK ( $F(118,2)=6.65$ ,  $p<.01$ ). Es zeigt sich nur ein signifikanter Unterschied zwischen Lehrkräften, die ein gymnasiales bzw. nicht-gymnasiales Lehramt studiert haben (siehe Tabelle 47).

Ein Kruskal-Wallis-Tests ergibt für Physiklehrkräfte einen signifikanten Einfluss des Ausbildungsweges auf PCK ( $H(2)=27,23$ ,  $p<.001$ ). Da drei Mann-Whitney-Tests gerechnet werden müssen, um herauszufinden, zwischen welchen Gruppen die Unterschiede liegen, muss das geforderte Signifikanzniveau angepasst werden (siehe 6.3). Es muss also  $p<.5/3=.017$  gelten, um von einem signifikanten Unterschied zu sprechen. Wie in Tabelle 48 zu erkennen, schneiden Lehrkräfte, die ein nicht-gymnasiales Lehramt studiert haben, signifikant schlechter ab als die beiden anderen Gruppen.

**Tabelle 47:** Testleistungsunterschiede in CK für Physiklehrkräfte, Post-Hoc-Tests (GT2 nach Hochberg)

	CK
GY zu NGY	$p<.01$ , $d=1.03$
GY zu Diplom	$p=.204$
NGY zu Diplom	$p=.188$

GY Gymnasiallehramt studiert

NGY Nicht-gymnasiales Lehramt

Weitere Konventionen siehe Tabelle 33

**Tabelle 48:** Testleistungsunterschiede in PCK für Physiklehrkräfte, Mann-Whitney-Tests

	PCK
GY zu NGY	$U=155$ , $p<.001$ , zweiseitig, $d=1.45$
GY zu Diplom	$U=1041$ , $p=.501$ , zweiseitig
Diplom zu NGY	$U=103$ , $p<.001$ , zweiseitig, $d=1.28$

U U-Wert

Weitere Konventionen siehe Tabelle 33 und 34

Für Lehrkräfte im Vorbereitungsdienst ergibt eine ANOVA keinen signifikanten Unterschied für PCK ( $F(56,2)=2.38$ ,  $p=.102$ ). Der Unterschied für CK ist signifikant ( $F(56,2)=4.14$ ,  $p<.05$ ). Er ist zwischen Lehrkräften im Vorbereitungsdienst zu finden, die ein gymnasiales bzw. nicht-gymnasiales Lehramt studiert haben (siehe Tabelle 49).

**Tabelle 49:** Testleistungsunterschiede in CK für Lehrkräfte im Vorbereitungsdienst, Post-Hoc-Tests (GT2 nach Hochberg)

	CK
GY zu NGY	$p < .05$ , $d = 1.04$
GY zu Diplom	$p = .938$
NGY zu Diplom	$p = .145$

Konventionen siehe Tabelle 33 und 34

Für Lehramtsstudierende werden die Unterschiede mit t-Tests berechnet. Für CK ergibt sich ein signifikanter Unterschied zwischen Studierenden die ein gymnasiales bzw. ein nicht-gymnasiales Lehramt studiert haben ( $t(84) = 2.44$ ,  $p < .05$ ,  $d = 0.52$ ), für PCK ergibt sich kein signifikanter Unterschied ( $t(84) = 1.94$ ,  $p = .057$ ).

Zusammenfassend ist festzustellen, dass Lehrkräfte der nicht-gymnasialen Lehrämter im Vergleich mit Gymnasiallehrpersonen in CK oder PCK höchstens gleich gute Werte erreichen.



*DANKE*

Ich danke nicht nur meiner Familie für die Unterstützung, sondern auch meiner akademischen Familie für die Aufnahme in die Welt der Wissenschaft.

## **ICH DANKE...**

Meinen Betreuern und Gutachtern Hans E. Fischer und Andreas Borowski für ihre Ideen, gute Diskussionen und die Möglichkeit mein Projekt nicht nur bearbeiten, sondern auch national und international vorstellen zu können

Heike Theyßen für die Übernahme des Prüfungsvorsitzes

Claudia von Aufschnaiter für die Übernahme der Rolle der Prüferin und ihre Ermutigung

Julie Gess-Newsome, für einen inspirierenden und intensiven Auslandsaufenthalt

Meinen „Büromitbewohnerinnen“ Eva, Cornelia und Alina - und Katja ehrenhalber - für ihren Rückhalt und Gespräche über alles, also auch Fachdidaktik und Statistik

Stefan und Mirko für die zu kurze gemeinsame Bürozeit

Meinen Kolleginnen Andrea, Annett, Annika, Christina, Helena, Ingrid, Katharina, Lena, Meike und Tina für eine gute Zeit

Allen Mitgliedern und GründerInnen der nwu, sowie allen KollegInnen, allen MitarbeiterInnen des ProwiN-Projektes und der AG-Fischer für die freundschaftliche Zusammenarbeit

Helene und Raffaella für die Suche – und das Finden – der Lehrkräfte

Claudia und Janina für tatkräftige Unterstützung

Bill Boone und Christian Spoden für ihre Hilfe bei der Datenauswertung

Meiner neuen Arbeitsgruppe für die herzliche Aufnahme

Allen TagungsteilnehmerInnen, die meine Poster und Vorträge besucht und mit mir diskutiert haben und insbesondere allen TeilnehmerInnen des PCK-Summits

Den OrganisatorInnen und TeilnehmerInnen der Summer- und Winterschools - ESERA, NARST, nwu -, die ich besuchen durfte

Unseren Hilfskräften und Examenskandidatinnen Daniel, Evelin, Judith, Julia A., Julia B. und Sarah

Allen Lehrkräften, Lehrkräften im Vorbereitungsdienst, Studierenden und PhysikerInnen, die an der Studie teilgenommen haben

Allen Seminarleitern, die zur Testung ihrer Lehrkräfte im Vorbereitungsdienst Zeit zur Verfügung gestellt haben

Der Fachschaft Physik der Uni Frankfurt, die eine Validierungsstudie organisiert hat

Allen Lehrkräften und Seminarleitern, die bei der Entwicklung der Tests mitgewirkt haben, insbesondere Udo Wlotzka

Den folgenden FachdidaktikerInnen, die mich, neben meinen Betreuern, bei der Testentwicklung, -durchführung, -validierung und -übersetzung unterstützten:

Claudia von Aufschnaiter

Julie Gess-Newsome

Jennifer Ergönenc (geb. Olszewski)

Josef Riese

Karsten Rincke

Christian Salinga

Horst Schecker

Rita Wodzinski

Allen KorrekturleserInnen dieser Arbeit

Hanno, meiner Familie, allen Bluts- und Wasser-Verwandten, sowie Freunden für ihre Toleranz und Unterstützung