

Fachdidaktisches Wissen angehender und praktizierender Biologielehrkräfte im Kompetenzbereich Erkenntnisgewinnung

Mehrperspektivische Entwicklung, Validierung und Anwendung eines
Multiple-Choice-Instruments

DISSERTATION

ZUR ERLANGUNG DES DOKTORGRADES DER
MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT DER
CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL

Dr. rer. nat.

VORGELEGT VON

TOBIAS LIEBEREI

KIEL, 2024

Erster Gutachter:

Prof. Dr. Moritz Krell

Zweiter Gutachter:

Prof. Dr. Dirk Krüger

Tag der mündlichen Prüfung:

14.02.2025

Zusammenfassung

Das fachdidaktische Wissen einer Lehrkraft hat als zentraler Teil ihres Professionswissens und damit elementarer Bestandteil ihrer professionellen Handlungskompetenz einen entscheidenden Einfluss auf die Unterrichtsqualität und damit auch Lernleistung ihrer Schüler:innen. Aus diesem Grund sollen angehende Lehrkräfte grundlegendes fachdidaktisches Wissen bereits während ihrer universitären Ausbildung erlangen und in ihrer Tätigkeit als praktizierende Lehrkraft weiter ausbauen. Zur Evaluation entsprechender Aus- und Weiterbildungsangebote bedarf es geeigneter Erhebungsmethoden und -instrumente, welche sich aufgrund der Komplexität des zu erfassenden Wissensbereichs jedoch maßgeblich voneinander unterscheiden. So haben Multiple-Choice-Instrumente im Vergleich zu anderen Erhebungsmethoden Vorteile in Bezug auf Testökonomie und Objektivität und eignen sich damit insbesondere für den Einsatz bei großen Stichproben. Die Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von fachdidaktischem Wissen geht jedoch auch mit Herausforderungen bezüglich der Formulierung der Items, der Konstruktrepräsentation sowie psychometrischer Eigenschaften einher. Um dennoch von den Vorteilen dieser Instrumente profitieren zu können, müssen Multiple-Choice-Instrumente zur Erfassung von fachdidaktischem Wissen systematisch entwickelt und validiert werden.

Lehrkräfte brauchen fachdidaktisches Wissen zu unterschiedlichen Inhalten eines Fachs, so auch in Bezug auf Erkenntnisgewinnungskompetenzen, welche einen elementaren Bestandteil des naturwissenschaftlichen Unterrichtes darstellen. Trotz dieser zentralen Rolle wurde dieser spezifische Wissensbereich bisher kaum erforscht und bereits durchgeführte Studien unterscheiden sich hinsichtlich diverser Aspekte, etwa der betrachteten Stichprobe oder der verwendeten Erhebungsmethoden und -instrumente. Ein grundlegendes Verständnis über die Ausprägung des fachdidaktischen Wissens angehender und praktizierender Lehrkräfte im Kompetenzbereich Erkenntnisgewinnung sowie dem Zusammenhang mit konzeptionell verwandten Variablen wie der Berufserfahrung oder epistemologischen Überzeugungen ist jedoch für die Gestaltung effektiver Aus- und Weiterbildungsangebote notwendig.

Den genannten Desideraten wird im Rahmen dieser Dissertation durch die mehrperspektivische Entwicklung, Validierung und Anwendung eines Multiple-Choice-Instruments zur Erfassung von fachdidaktischem Wissen angehender und praktizierender Biologielehrkräfte im Kompetenzbereich Erkenntnisgewinnung (PCK_{SR-bio}) begegnet. Dazu wird einerseits untersucht, wann sich Multiple-Choice-Instrumente zur Erfassung von fachdidaktischem Wissen eignen und wie mit der Entwicklung dieser Instrumente einhergehende Herausforderungen hinsichtlich einer adäquaten Formulierung der Items, einer passenden Repräsentation des zugrunde liegenden Konstrukts sowie der psychometrischen Eigenschaften begegnet werden kann. Andererseits wird untersucht, wie das fachdidaktische Wissen im Kompetenzbereich Erkenntnisgewinnung angehender und praktizierender Lehrkräfte ausgeprägt ist und wie es mit der Berufserfahrung sowie anderer konzeptionell verwandter Variablen zusammenhängt.

Dazu wurden fünf empirische Studien durchgeführt. In Studie 1 wurde im Rahmen einer Interviewstudie das Antwortverhalten verschiedener Gruppen mit unterschiedlich zu erwartender Expertise bei der Bearbeitung des PCK_{SR-bio} basierend auf dem Experten-Novizen-Paradigma untersucht, um Herausforderungen hinsichtlich der Konstruktrepräsentation zu begegnen. Dabei konnte gezeigt werden, dass sich diese Gruppen hinsichtlich ihres Wissens, der Konsistenz des Antwortverhaltens sowie der Begründung zur Auswahl der Antwortoptionen unterscheiden. Die Validität der Testwertinterpretation des PCK_{SR-bio} wurde in Studie 2 basierend auf unterschiedlichen Quellen geprüft. Die Ergebnisse zeigen, dass eine valide Interpretation des Testwerts möglich ist. In Studie 3 wurde untersucht, ob Validitätsevidenz auf Grundlage von Antwortprozessen durch die

Verwendung des *PCK map approaches* gesammelt werden kann. Dies konnte empirisch gestützt werden und stärkt damit ebenfalls die Validität der Testwertinterpretation des PCK_{SR}-bio. In Studie 4 wurde der PCK_{SR}-bio in einer querschnittlichen Erhebung eingesetzt, um grundlegende Erkenntnisse bezüglich der Ausprägung des fachdidaktischen Wissens im Kompetenzbereich Erkenntnisgewinnung von praktizierenden Biologielehrkräften an Grund- und weiterführenden Schulen sowie von Hochschuldozierenden in der Lehrkräfteausbildung im Fach Biologie zu erlangen. Zusätzlich wurde der Einfluss der Berufserfahrung, der transmissiven Überzeugungen sowie der Offenheit zur Innovation als konzeptionell verwandte Variablen auf das fachdidaktische Wissen im Kompetenzbereich Erkenntnisgewinnung untersucht. Dabei zeigen sich eine unterschiedliche Ausprägung sowie verschiedene Zusammenhänge für die unterschiedlichen Gruppen. Weiterhin wurde eine englischsprachige Version des PCK_{SR}-bio in Studie 5 in einer explorativen Datenerhebung mit australischen Lehramtsstudierenden eingesetzt, um Erkenntnisse über das fachdidaktische Wissen im Kompetenzbereich Erkenntnisgewinnung von angehenden Biologielehrkräften in Australien zu erlangen.

Die Ergebnisse dieser Studien werden aus einer methodischen sowie inhaltlichen Perspektive diskutiert, um den beiden oben genannten Desideraten zu begegnen. Zusätzlich findet eine Synthese beider Perspektiven statt. Die vorliegende Dissertation stellt damit nicht nur ein Multiple-Choice-Instrument zur Erfassung von fachdidaktischem Wissen im Kompetenzbereich Erkenntnisgewinnung vor, dessen valide Testwertinterpretation kritisch diskutiert wird, sondern gibt Hinweise für die Entwicklung von Multiple-Choice-Instrumenten zur Erfassung fachdidaktischen Wissens im Allgemeinen sowie hinsichtlich der Ausprägung des fachdidaktischen Wissens angehender und praktizierender Lehrkräfte im Kompetenzbereich Erkenntnisgewinnung im Speziellen.

Summary

Pedagogical content knowledge (PCK) of teachers, as a central component of their professional knowledge and an essential part of their professional competence, plays a crucial role in influencing the effectiveness of their teaching and, therefore, their students' achievement. Thus, preservice science teachers should acquire basic PCK during their university studies and build on it during their work as practicing teachers. In order to evaluate university courses during teacher education as well as professional development programs, suitable assessment methods and instruments are needed, which differ significantly from each other due to the complexity of PCK. Multiple-choice instruments, for instance, have advantages in terms of test economy and objectivity compared to other assessment methods, making them particularly useful for assessing PCK in larger sample sizes. However, developing multiple-choice instruments to assess PCK also comes with challenges, such as the phrasing of the items, the construct representation, and the psychometric properties. In order to take advantage of the benefits of multiple-choice instruments for assessing PCK, those instruments need to be systematically developed and validated.

Science teachers need PCK for different topics, including scientific reasoning, which is an essential part of science education. However, despite its central role, PCK about scientific reasoning has barely been researched to date and studies that have already been conducted differ regarding various aspects, for example the different samples or the assessment methods and instruments used. A fundamental understanding of the PCK about scientific reasoning of preservice and in-service science teachers, as well as its relationship with conceptually related variables such as their professional experience or epistemological beliefs, is crucial for implementing effective university courses in teacher education as well as professional development programs.

These desiderata are addressed in this dissertation through a multi-perspective development, validation, and application of a multiple-choice instrument for assessing preservice and in-service biology teachers' PCK about scientific reasoning (PCK_{SR}-bio). On the one hand, it is investigated when multiple-choice instruments are suitable for assessing PCK and how challenges regarding an adequate phrasing of the items, an appropriate representation of the underlying construct, as well as the psychometric properties of the instrument can be addressed. On the other hand, it is examined how PCK about scientific reasoning of preservice and in-service teachers is characterized and how it is related to their professional experience as well as to other variables.

Five empirical studies were conducted for this purpose. Study 1 investigated the response behavior of two groups with different expected expertise in answering the PCK_{SR}-bio, based on the expert-novice paradigm, in order to address challenges regarding the construct representation. It was shown that these groups differ in terms of their knowledge, the consistency of their response behavior, and the justification for selecting a response option. The validity of the test score interpretation of the PCK_{SR}-bio was evaluated based on different sources in Study 2. The results show that a valid interpretation of the test score is possible. Study 3 examined whether validity evidence based on response processes can be collected by using the *PCK map approach*, which is empirically supported and thus also strengthens the validity of the test score interpretation of the PCK_{SR}-bio. In Study 4, the PCK_{SR}-bio was used in a cross-sectional study to gain insights into the PCK_{SR} of in-service biology teachers at elementary and secondary schools as well as teacher educators in biology education. In addition, the influence of their professional experience, their beliefs about transmission, as well as their openness to innovation on PCK about scientific reasoning was investigated. The results showed different characteristics and relationships across the different groups. Furthermore, in Study 5, the PCK_{SR}-bio was used in an exploratory study with preservice science teachers from Australia to gain insights into their PCK about scientific reasoning.

The findings of these studies are evaluated from both a methodological and content-related perspective to address the two outlined desiderata. Additionally, a synthesis of these perspectives is provided. This dissertation thus not only presents a multiple-choice instrument for assessing PCK and critically discuss the validity of its test score interpretation, but also provides insights into the development of multiple-choice instruments for assessing PCK in general, as well as with regard to the characteristics of PCK about scientific reasoning of preservice and in-service teachers in particular.

Inhaltsverzeichnis

Zusammenfassung	I
Summary	III
Abkürzungsverzeichnis	XI
Abbildungsverzeichnis	XIII
Tabellenverzeichnis	XV
1 Einleitung	1
2 Theoretischer Hintergrund	5
2.1 Fachdidaktisches Wissen	5
2.1.1 Integrative und transformative Konzeptualisierung fachdidaktischen Wissens	5
2.1.2 Unterschiedliche Modelle zum fachdidaktischen Wissen	6
2.1.2.1 Modell von Magnusson et al. (1999)	7
2.1.2.2 Pentagon-Modell	9
2.1.2.3 Consensus-Modell	11
2.1.2.4 Refined-Consensus-Modell	14
2.1.3 Unterschiedliche Formen fachdidaktischen Wissens	16
2.1.4 Ausgewählte empirische Befunde	16
2.1.4.1 Zentrale Komponenten fachdidaktischen Wissens	17
2.1.4.2 Einfluss auf die Unterrichtsqualität und Lernleistung von Schüler:innen	17
2.1.4.3 Entwicklung von fachdidaktischem Wissen	18
2.1.5 Nationale und internationale curriculare Verankerung in der Lehrkräfteausbildung	20
2.1.6 Erfassung von fachdidaktischem Wissen	21
2.1.6.1 Multiple Choice-Instrumente	23
2.1.6.2 Der <i>PCK map approach</i>	25
2.2 Erkenntnisgewinnungskompetenzen	27
2.2.1 Konzeptualisierung von Erkenntnisgewinnungskompetenzen	29
2.2.2 Nationale und internationale curriculare Verankerung im naturwissenschaftlichen Unterricht	32
2.2.3 Erkenntnisgewinnungskompetenzen im Studium	33
2.2.4 Erkenntnisgewinnungskompetenzen im Unterricht	34
2.2.4.1 Experimentieren im naturwissenschaftlichen Unterricht	36
2.2.4.2 Hypothesen formulieren im naturwissenschaftlichen Unterricht	36
2.2.4.3 Modellieren im naturwissenschaftlichen Unterricht	37
2.3 Fachdidaktisches Wissen im Kompetenzbereich Erkenntnisgewinnung	38

2.4	Validierung von Testinstrumenten	40
2.4.1	Validitätsevidenz auf Grundlage des Testinhalts	41
2.4.2	Validitätsevidenz auf Grundlage von Antwortprozessen	42
2.4.3	Validitätsevidenz auf Grundlage der internen Struktur	43
2.4.4	Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen	43
2.4.5	Validitätsevidenz auf Grundlage von Folgen der Testung.....	44
3	Zielsetzung der Dissertation und Zusammenfassung der durchgeführten Studien	45
3.1	Studie 1: Findings from the Expert-Novice Paradigm on Differential Response Behavior Among Multiple-Choice Items of a Pedagogical Content Knowledge Test – Implications for Test Development (Kapitel 4)	46
3.2	Studie 2: Considering Multiple Sources of Validity Evidence Can Help to Address Challenges in the Development of Pedagogical Content Knowledge (PCK) Multiple-Choice Items (Kapitel 5)	47
3.3	Studie 3: Using the PCK Map Approach to Obtain Validity Evidence Based on Response Processes for a Multiple-Choice Instrument (Kapitel 6).....	47
3.4	Studie 4: Pedagogical Content Knowledge (PCK) about Scientific Reasoning in Different Groups of Science Teachers and Its Relation to Other Variables (Kapitel 7)	48
3.5	Studie 5: Preservice Science Teachers’ Knowledge About How to Teach Scientific Reasoning (Kapitel 8)	49
4	Studie 1: Findings from the Expert-Novice Paradigm on Differential Response Behavior Among Multiple-Choice Items of a Pedagogical Content Knowledge Test – Implications for Test Development.....	57
4.1	Introduction.....	58
4.2	Theoretical Background.....	59
4.2.1	Pedagogical Content Knowledge	59
4.2.2	Assessing PCK with Paper-Pencil Tests.....	60
4.2.2.1	Challenges in Developing Multiple-Choice Items	61
4.2.3	Analyzing Response Processes within the Framework of the Expert-Novice Paradigm ..	62
4.2.4	Aim of the Study and Hypotheses	64
4.3	Materials and Methods	64
4.3.1	Sample	64
4.3.2	Data Collection	65
4.3.2.1	Interview Protocol	65
4.3.2.2	Multiple-Choice Test.....	65
4.3.3	Scoring	67
4.3.4	Statistical Analyses	69

4.4	Results	70
4.4.1	RQ1: To what extent do experts achieve higher proportion of correct responses in a PCK _{SR} test than novices?	70
4.4.2	RQ2: To what extent do experts' answers in a PCK _{SR} test vary less than novices' ones?	71
4.4.3	RQ3: To what extent do experts and novices justify their choice of response options in a PCK _{SR} test differently?	73
4.5	Discussion	76
4.5.1	RQ1: To what extent do experts achieve higher proportion of correct responses in a PCK _{SR} test than novices?	77
4.5.2	RQ2: To what extent do experts' answers in a PCK _{SR} test vary less than novices' ones?	77
4.5.3	RQ3: To what extent do experts and novices justify their choice of response options in a PCK _{SR} test differently?	77
4.5.4	Conclusions.....	78
4.5.5	Implications	78
4.5.6	Limitations and Future Work.....	78
4.6	References	79
5	Studie 2: Considering Multiple Sources of Validity Evidence Can Help to Address Challenges in the Development of Pedagogical Content Knowledge (PCK) Multiple-Choice Items	87
5.1	Introduction.....	88
5.2	Theoretical Background.....	89
5.2.1	Pedagogical Content Knowledge and its Assessment	89
5.2.1.1	Cognitive Operations and Task Formats.....	90
5.2.1.2	Construct Representation.....	91
5.2.1.3	Psychometric Properties.....	91
5.2.2	Different Sources of Validity Evidence	91
5.2.3	Research Question.....	92
5.3	Item Development (Evidence Based on Test Content)	93
5.3.1	Systematic Development.....	93
5.3.2	Initial Item Development.....	94
5.3.3	Revision of the Items.....	95
5.4	Evidence Based on Response Processes.....	95
5.4.1	Methods	95
5.4.1.1	Sample	95
5.4.1.2	Data Collection	95
5.4.1.3	Data Analysis.....	95
5.4.2	Results	96

5.5	Evidence Based on Internal Structure	97
5.5.1	Methods	97
5.5.1.1	Sample	97
5.5.1.2	Data Collection	97
5.5.1.3	Data Analysis.....	97
5.5.2	Results	97
5.6	Evidence Based on Relations to Other Variables	98
5.6.1	Methods	98
5.6.1.1	Sample	98
5.6.1.2	Data Collection	98
5.6.1.3	Data Analysis.....	99
5.6.2	Results	99
5.7	Overall Discussion.....	100
5.7.1	Conclusion and Outlook	101
5.7.2	Recommendations.....	102
5.8	References	102
6	Studie 3: Using the PCK Map Approach to Obtain Validity Evidence Based on Response Processes for a Multiple-Choice Instrument.....	107
6.1	Introduction.....	108
6.2	Theoretical Background.....	108
6.2.1	Pedagogical Content Knowledge	108
6.2.2	Assessing PCK	109
6.2.3	Validity Evidence Based on Response Processes.....	110
6.2.4	The PCK Map Approach	110
6.2.5	Aim of this Study.....	110
6.3	Methods	111
6.3.1	Instrument Development	111
6.3.2	Data Collection	112
6.4	Data Analysis	112
6.5	Results	113
6.6	Discussion	114
6.6.1	Recommendations.....	114
6.7	References	115
7	Studie 4: Pedagogical Content Knowledge (PCK) about Scientific Reasoning in Different Groups of Science Teachers and Its Relation to Other Variables.....	119
7.1	Introduction.....	120
7.2	Theoretical Background.....	121

7.2.1	Scientific Reasoning.....	121
7.2.1.1	Teaching Scientific Reasoning.....	121
7.2.2	Pedagogical Content Knowledge.....	122
7.2.2.1	Assessment of PCK.....	123
7.2.3	Pedagogical Content Knowledge About Scientific Reasoning.....	124
7.2.3.1	Differences Between the Groups of Teachers.....	125
7.2.4	Aim of the Study and Hypotheses	126
7.3	Methods	127
7.3.1	Data Collection	127
7.3.2	Sample	127
7.3.3	Assessment of PCK: <i>PCK_{SR}-bio</i> by Lieberei, Großmann, et al. (2024)	128
7.3.4	Assessment of Other Variables	129
7.3.4.1	Teachers' Epistemological Beliefs about Transmission: <i>Q4TB</i> by Schlichter (2012) ..	129
7.3.4.2	Teachers' Openness to Innovation: <i>FPIB</i> by Waßmann (2014)	129
7.3.4.3	Teachers' Professional Experience	129
7.3.5	Data Analysis	129
7.4	Results	130
7.4.1	How Does the <i>PCK_{SR}</i> of Elementary School Teachers, Secondary School Teachers, and Teacher Educators Differ?	130
7.4.2	To What Extent Can Teachers' <i>PCK_{SR}</i> Be Predicted by Their Epistemological Beliefs About Transmission, Their Professional Experience, and Their Openness to Innovation?	130
7.5	Discussion	134
7.5.1	Limitations	136
7.5.2	Conclusions and Implications	137
7.6	References	137
8	Studie 5: Preservice Science Teachers' Knowledge About How to Teach Scientific Reasoning	145
8.1	Introduction.....	146
8.2	Theoretical Background.....	147
8.2.1	Scientific Reasoning Competencies.....	147
8.2.2	Teachers' Knowledge About Scientific Reasoning and How to Teach it	147
8.2.3	Scientific Reasoning in the Australian Context.....	148
8.2.4	Research Aim	149
8.3	Exploratory Study on PCK about SRC in Science Teacher Education.....	149
8.3.1	Instrument Development	149
8.3.2	Data Collection	150
8.3.3	Data Analysis	151

8.4	Results	151
8.5	Discussion & Conclusion	153
8.5.1	Limitations	155
8.5.2	Suggestions for Initial Teacher Education	155
8.6	References	156
9	Zusammenfassung und Interpretation der zentralen Ergebnisse der durchgeführten Studien	159
9.1	Studie 1: Findings from the Expert-Novice Paradigm on Differential Response Behavior Among Multiple-Choice Items of a Pedagogical Content Knowledge Test – Implications for Test Development (Kapitel 4)	159
9.2	Studie 2: Considering Multiple Sources of Validity Evidence Can Help to Address Challenges in the Development of Pedagogical Content Knowledge (PCK) Multiple-Choice Items (Kapitel 5)	159
9.3	Studie 3: Using the PCK Map Approach to Obtain Validity Evidence Based on Response Processes for a Multiple-Choice Instrument (Kapitel 6).....	160
9.4	Studie 4: Pedagogical Content Knowledge (PCK) about Scientific Reasoning in Different Groups of Science Teachers and Its Relation to Other Variables (Kapitel 7)	161
9.5	Studie 5: Preservice Science Teachers' Knowledge About How to Teach Scientific Reasoning (Kapitel 8)	161
10	Übergreifende Diskussion und Ausblick	163
10.1	Methodische Perspektive	163
10.2	Inhaltliche Perspektive	168
10.3	Synthese beider Perspektiven	170
10.4	Limitationen.....	172
10.5	Implikationen und Ausblick	173
10.5.1	Implikationen für die Forschung.....	173
10.5.2	Implikationen für die Praxis.....	174
10.5.3	Ausblick.....	175
11	Literaturverzeichnis	177
12	Anhang	195
12.1	Multiple-Choice-Instrument zur Erfassung fachdidaktischen Wissens im Kompetenzbereich Erkenntnisgewinnung (PCK _{SR} -bio)	197
12.2	Englische Version des PCK _{SR} -bio	209
12.3	Erweiterung des PCK _{SR} -bio.....	215
12.4	Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1	227
12.5	Publikationen im Rahmen der Dissertation.....	235
12.6	Vorträge und Poster im Rahmen der Dissertation	237
	Danksagung	239

Abkürzungsverzeichnis

ACARA	Australian Curriculum, Assessment and Reporting Authority
AERA	American Educational Research Association
AITSL	Australian Institute for Teaching and School Leadership
ANOVA	Varianzanalyse (<i>analysis of variances</i>)
APA	American Psychological Association
BCTC	British Columbia Teachers' Council
CK	Fachwissen (<i>content knowledge</i>)
CoRe	Content Representation
cPCK	Kollektives PCK (<i>collective pedagogical content knowledge</i>)
ePCK	Angewandtes PCK (<i>enacted pedagogical content knowledge</i>)
FPIB	Fragebogen zur Erfassung psychologischer Innovationsblockaden
GDSU	Gesellschaft für Didaktik des Sachunterricht
H	Hypothese (<i>hypothesis</i>)
KAs	Wissen über Beurteilung (<i>knowledge of assessment of science learning</i>)
KISR	Wissen über Instruktionsstrategien (<i>knowledge of instructional strategies for teaching science</i>)
KMK	Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland
KR20	Kuder-Richardson-Formel 20 (<i>Kuder-Richardson formula 20</i>)
KSC	Wissen über das Fachcurriculum (<i>knowledge of science curriculum</i>)
KSU	Wissen über die Lerngruppe (<i>knowledge of students' understanding</i>)
MC	Multiple-Choice
MC–PCK	Multiple-Choice-Instrumente zur Erfassung von PCK
MINEDUC	Ministry of Education
MBWK SH	Ministerium für Bildung, Wissenschaft und Kultur Schleswig-Holstein
NCME	National Council on Measurement in Education
NGSS	Next Generation Science Standards
NSW	New South Wales
OECD	Organization for Economic Co-operation and Development
OTS	Wissen über kontextuelle Rahmenbedingungen (<i>orientations to teaching science</i>)
PCK	Fachdidaktisches Wissen (<i>pedagogical content knowledge</i>)
PCK&S	PCK und Fähigkeiten (<i>pedagogical content knowledge and skill</i>)
PCK-IBI	Pedagogical Content Knowledge in Biology Inventory
PCK _{SR}	Fachdidaktisches Wissen im Kompetenzbereich Erkenntnisgewinnung (<i>pedagogical content knowledge about scientific reasoning</i>)
PCK _{SR} -bio	Multiple-Choice-Instrument zur Erfassung von fachdidaktischem Wissen im Kompetenzbereich Erkenntnisgewinnung
PISA	Program for International Student Assessment
PK	Pädagogisches Wissen (<i>pedagogical knowledge</i>)
POSITT	Pedagogy of Science Inquiry Teaching Test
POSTT	Pedagogy of Science Teaching Tests

pPCK	Persönliches PCK (<i>personal PCK</i>)
pPCK _{SR}	Persönliches PCK im Kompetenzbereich Erkenntnisgewinnung (<i>personal PCK about scientific reasoning</i>)
Q4TB	Questionnaire of Four Teachers' Beliefs
RCM	Refined-Consensus-Modell (<i>refined consensus model</i>)
RQ	Forschungsfrage (<i>research question</i>)
SDDS	Scientific Discovery as Dual Search-Modell
SR	Wissenschaftliches Denken (<i>scientific reasoning</i> ; im Rahmen dieser Dissertation synonym mit Erkenntnisgewinnung verwendet)
SRC	Erkenntnisgewinnungskompetenzen (<i>scientific reasoning competencies</i>)
VIF	Varianzinflationsfaktor (<i>variance inflation factor</i>)

Abbildungsverzeichnis

Abbildung 2.1: Modell von Magnusson et al. (1999; übersetzt)	8
Abbildung 2.2: Pentagon Modell (adaptiert und übersetzt nach Park & Chen, 2012; Park & Oliver, 2008).....	10
Abbildung 2.3: Consensus-Modell (übersetzt nach Gess-Newsome, 2015).....	13
Abbildung 2.4: Refined-Consensus-Modell (adaptiert und übersetzt nach Carlson et al., 2019)	15
Abbildung 2.5: Rahmenmodell für den Kompetenzbereich Erkenntnisgewinnung (Mayer, 2007, S. 178).....	29
Abbildung 2.6: Strukturmodell zum wissenschaftlichen Denken (Mayer, 2007, S. 181).....	31
Figure 4.1: Pentagon-model of PCK; KSU and KISR as well as their connections are highlighted	60
Figure 4.2: Sample item (item #12).....	67
Figure 4.3: Boxplots of the proportions of correct answers	71
Figure 4.4: Comparison of the number of justifications assigned to <i>Deep Thinking</i> and <i>Surface Thinking</i>	72
Figure 4.5: Sum of the codings (± 1 standard deviation) of (A) <i>Surface Thinking</i> and (B) <i>Deep Thinking</i>	75
Figure 5.1: Sample item (Item #12).....	94
Figure 5.2: Test scores of the experimental group and of the control group in the pre- and posttest	99
Figure 6.1: Pentagon model of PCK. KSU and KISR, as well as their connections, are highlighted.....	109
Figure 6.2: Sample item of the PCK _{SR} -bio (Item #12)	112
Figure 6.3: PCK maps, divided into the four PCK components of the PCK _{SR} -bio.....	113
Figure 7.1: Overview of the professional experience of the different groups.....	131
Figure 8.1: Overview of the two central components of PCK: KISR and KSU, as well as their connections	148
Figure 8.2: Sample item (Item #12) of the PCK _{SR} -bio	150
Figure 8.3: Proportion of correct answers ($\pm SD$) for items addressing the different PCK components	153

Tabellenverzeichnis

Tabelle 2.1: Übersicht über die fachdidaktischen Kompetenzen zum Ende der universitären Ausbildung (KMK, 2024a, S. 4) und zugehörige PCK-Komponenten	20
Tabelle 2.2: Mögliche Erhebungsmethoden und -instrumente zur Analyse von unterschiedlichen Bereichen des PCK	23
Tabelle 2.3: Übersicht über die fünf Validitätsquellen nach AERA et al. (2014).....	41
Tabelle 3.1: Übersicht über die durchgeführten Studien	50
Table 4.1: Two major challenges in the development of multiple-choice PCK instruments	62
Table 4.2: Systematic overview of the 12 multiple-choice items, developed to capture three skills of scientific reasoning, two PCK components, and their connections, respectively.....	66
Table 4.3: Short version of the coding system	69
Table 4.4: Descriptive statistics for each item of the PCK _{SR} items and the full test score	70
Table 4.5: Overview of the codings.....	74
Table 5.1: Overview of the studies conducted, addressing the four sources of validity evidence.....	92
Table 5.2: Theoretical framework of the MC instrument developed to assess PCK _{SR}	93
Table 5.3: Category system for analyzing the justifications of the selection of response options.....	96
Table 5.4: Total number of justifications based on “PCK argumentation” for the respective items....	97
Table 5.5: Overview of the internal consistency and the test scores of the groups in different stages of teacher education	98
Table 6.1: Systematic overview of the 12 multiple-choice items developed to capture three skills of scientific reasoning, two PCK components, and two PCK connections	111
Table 6.2: Coding system for analysing the justifications given for the selection of response options based on the categories KSU and KISR.....	113
Table 7.1: Potential assessment methods or instruments for analyzing different layers of PCK	124
Table 7.2: Overview of the sample regarding the proportion of teachers who had completed in-service training, had a PhD, and were teaching or had studied another science subject	128
Table 7.3: Systematic overview of the PCK _{SR} -bio	128
Table 7.4: Overview of the PCK _{SR} test score and the internal consistency of the groups	130
Table 7.5: Overview of the groups’ epistemological beliefs about transmission and openness to innovation.....	131
Table 7.6: Correlation analysis for the related variables and the PCK _{SR} test score for each group....	132
Table 7.7: Regression coefficients of PCK _{SR} test score for the different groups.....	133
Table 7.8: Regression coefficients of PCK _{SR} test score including the groups of teachers as dummy-coded variables.....	133
Table 8.1: Systematic overview of the 12 multiple-choice items of the PCK _{SR} -bio, developed to assess two PCK components and their connections for three scientific reasoning activities.....	149
Table 8.2: Descriptive statistics for each item (<i>M</i> = mean proportion of correct answers, <i>SD</i> = standard deviation) answered by 36 preservice science teachers (secondary science).....	152
Tabelle 12.1: Übersicht über die Systematik des PCK _{SR} -bio.....	197
Tabelle 12.2: Lösungsbogen des PCK _{SR} -bio	204
Tabelle 12.3: Übersicht über die Systematik der neu entwickelten Items des PCK _{SR} -bio	215
Tabelle 12.4: Lösungsbogen der neu entwickelten Items des PCK _{SR} -bio	222
Tabelle 12.5: Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1	227

1 Einleitung

Die Frage, was *guten* Unterricht ausmacht, stellt einen zentralen Schwerpunkt in der fachdidaktischen Forschung dar. Es gibt eine Vielzahl empirisch gestützter Einflussfaktoren, welche die Qualität des Unterrichts und damit die Lernleistung von Schüler:innen (positiv sowie negativ) beeinflussen (vgl. Hattie, 2023). Dabei spielt die Lehrkraft eine entscheidende Rolle: Als ein zentraler Einflussfaktor auf die Effektivität ihres Unterrichts wird ihre professionelle Handlungskompetenz¹ angesehen (Blömeke et al., 2022; Hattie, 2023; Hill et al., 2005; Kunter et al., 2013), welche unterschiedliche Aspekte umfasst und die Fähigkeiten, Fertigkeiten und Einstellungen beschreibt, die eine Lehrkraft benötigt, um den Anforderungen ihres Berufs gerecht zu werden (Baumert & Kunter, 2006). Ein in Deutschland entwickeltes und im deutschsprachigen sowie internationalen Raum etabliertes Modell zur Konzeptualisierung professioneller Handlungskompetenzen stammt aus der *COACTIV*-Studie (Baumert & Kunter, 2013; Baumert et al., 2011) und umfasst die vier Kompetenzaspekte Professionswissen (Wissen, das eine Lehrkraft benötigt, um ihren Beruf erfolgreich auszuüben), motivationale Orientierungen (Antriebe und Haltungen, welche die berufliche Tätigkeit beeinflussen, zum Beispiel Selbstwirksamkeitserwartungen oder intrinsische Motivation), selbstregulative Fähigkeiten (Fähigkeiten zur Steuerung und Reflexion des eigenen Denkens und Handelns) sowie Überzeugungen und Werthaltungen (grundlegende Prinzipien und Überzeugungen, die das Handeln leiten; Baumert & Kunter, 2006).

Insbesondere das Professionswissen als zentraler Aspekt der professionellen Handlungskompetenz spielt in Bezug auf erfolgreiches Unterrichten und damit den Lernerfolg der Schüler:innen eine wichtige Rolle (Baumert & Kunter, 2006; Kulgemeyer et al., 2020; van Driel et al., 2023). Basierend auf den Überlegungen von Shulman (1986, 1987) werden dabei verschiedene Wissensbereiche unterschieden, darunter pädagogisches Wissen (*pedagogical knowledge*; PK), Fachwissen (*content knowledge*; CK) sowie fachdidaktisches Wissen (*pedagogical content knowledge*; PCK; z. B. Baumert & Kunter, 2006). Besonders letzterem kommt dabei eine zentrale Rolle zu. So stellte sich PCK beispielsweise in einer Metaanalyse von Coe et al. (2014) als wichtigster Einflussfaktor auf die Lernleistung von Schüler:innen heraus. Aus diesem Grund sollten angehende Lehrkräfte fundamentales PCK bereits während ihrer universitären Ausbildung erwerben, was national sowie international curricular verankert ist (z. B. Australien: AITSL, 2017; Deutschland: KMK, 2024a). Dies ist insbesondere vor dem Hintergrund des Lehrkräftemangels (Geis-Thöne, 2022; Klemm, 2020; KMK, 2022) relevant, da empfohlen wird, (noch) nicht vollständig qualifizierte Personen (etwa Lehramtsstudierende im Master of Education) als Vertretungslehrkräfte einzusetzen (SWK, 2024), weshalb diese bereits über ein grundlegendes PCK verfügen sollten. Da sich das PCK einer Lehrkraft jedoch hauptsächlich durch praktische Unterrichtserfahrung und Reflexion dieser (weiter)entwickelt (z. B. van Driel & Berry, 2019), sollten die im Studium erlangten Grundlagen im Referendariat sowie der anschließenden Tätigkeit als praktizierende Lehrkraft kontinuierlich weiterentwickelt werden.

Um diesen Erwerb im Rahmen der Aus- und Weiterbildung zu evaluieren und den Zusammenhang von PCK mit anderen konzeptionell verwandten Variablen zu untersuchen, bedarf es geeigneter Erhebungsmethoden und -instrumente zur Erfassung dieses Wissensbereichs. PCK stellt ein komplexes Konstrukt dar, welches unterschiedlich konzeptualisiert werden (z. B. Neumann et al., 2019) und in unterschiedlichen Formen auftreten kann (z. B. She & Chan, 2023). Zusätzlich gibt es eine Vielzahl unterschiedlicher Modelle zur Beschreibung von PCK (z. B. Carlson et al., 2019; Gess-Newsome, 2015; Magnusson et al., 1999; Park & Oliver, 2008). Aus diesem Grund existieren ebenfalls diverse

¹ Der Begriff „Kompetenz“ wird im Rahmen dieser Dissertation als kontextspezifisches Kontinuum aus kognitiven, motivational-affektiven Dispositionen und beobachtbarem Verhalten zur Problemlösung verstanden, welches bereichsspezifisch durch Erfahrung erworben und verändert werden kann (vgl. Blömeke et al., 2015; Klieme & Leutner, 2006; Weinert, 2001).

Erhebungsmethoden und -instrumente, die sich maßgeblich voneinander unterscheiden und verschiedene Vorteile mit sich bringen. Während sich die Beobachtung des Verhaltens einer (angehenden) Lehrkraft im Unterricht primär dazu eignet, das in einer spezifischen Situation gezeigte PCK zu erfassen (z. B. Chan & Hume, 2019), eignen sich Paper-Pencil-Tests dazu, diesen Wissensbereich situationsunabhängig zu messen. Paper-Pencil-Tests mit offenen Antwortformaten können dabei genutzt werden, die dem Antwortverhalten zugrunde liegenden Denkprozesse zu erfassen (Martinez, 1999) und das PCK einer Lehrkraft somit in seiner Tiefe abzubilden, während geschlossene Paper-Pencil-Tests wie Multiple-Choice-Instrumente mit Vorteilen in Bezug auf die Testökonomie und Objektivität einhergehen (z. B. Haladyna, 2004). Letztere eignen sich aus diesem Grund insbesondere als Erhebungsinstrumente bei großen Stichproben (Hill et al., 2008). Multiple-Choice-Instrumente zur Erfassung von PCK werden jedoch auch kritisch gesehen, da sie mit Herausforderungen in Bezug auf eine adäquate Formulierung der Items (z. B. Schmelzing et al., 2013), einer passende Repräsentation des zugrunde liegenden Konstrukts (z. B. Smith & Banilower, 2015) sowie den psychometrischen Eigenschaften (durch die Komplexität und den Anwendungsbezug des zugrunde liegenden Konstrukts, vgl. Taber, 2018) verbunden sind und deshalb von einigen Forschenden bewusst vermieden werden (z. B. Rosenkränzer et al., 2017).

Bei der Erfassung von PCK sollte darüber hinaus beachtet werden, dass es sich nicht um ein generisches Konstrukt handelt, Lehrkräfte also über spezifisches PCK zu unterschiedlichen Themen und Disziplinen des naturwissenschaftlichen Unterrichts verfügen müssen (z. B. Gess-Newsome, 2015). Ein elementarer Bestandteil des naturwissenschaftlichen (und damit auch des Biologie-)Unterrichts sind dabei Erkenntnisgewinnungskompetenzen (z. B. Mayer, 2007), welche curricular nicht nur in Deutschland als ein eigener Kompetenzbereich (z. B. KMK, 2024b), sondern auch international (z. B. in Australien; ACARA, 2022) verankert sind. Um ihren Schüler:innen den Erwerb dieser essenziellen Kompetenzen zu ermöglichen, müssen Lehrkräfte nicht nur selbst über diese verfügen (z. B. Khan & Krell, 2019), sondern auch wissen, wie sie diese optimal vermitteln können. Sie brauchen also entsprechendes fachdidaktisches Wissen im Kompetenzbereich Erkenntnisgewinnung (*PCK about scientific reasoning*; PCK_{SR}).

Bisher wurde PCK_{SR} allerdings nur in wenigen Studien empirisch untersucht, welche sich darüber hinaus maßgeblich hinsichtlich verschiedener Aspekte wie den betrachteten Stichproben (z. B. Studierende: Hilfert-Rüppell et al., 2021; Lehrkräfte an weiterführenden Schulen: Lin & Wang, 2024) oder den verwendeten Erhebungsmethoden und -instrumenten (z. B. Beobachtungen in O. de Jong & van der Valk, 2007; Paper-Pencil-Tests mit gemischtem Aufgabenformat: Lin & Wang, 2024) unterscheiden. Daher fehlt es bisher an grundlegendem Wissen über die Ausprägung des PCK_{SR} angehender und praktizierender Lehrkräfte, die Entwicklung dieses Wissensbereichs sowie den Zusammenhang zu anderen, konzeptionell verwandten Variablen wie ihren epistemologischen Überzeugungen. Erkenntnisse darüber sind jedoch Voraussetzung für ein Verständnis des Konstrukts und werden für die Evaluierung und Gestaltung effektiver Aus- und Weiterbildungsangebote benötigt. Es bedarf daher geeigneter Erhebungsinstrumente, die eine valide Testwertinterpretation als PCK_{SR} für angehende und praktizierende Lehrkräfte ermöglichen.

Somit ergeben sich zwei Desiderate: Einerseits ist die Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK mit Herausforderungen verbunden, welchen bei der Entwicklung und Validierung entsprechender Testinstrumente berücksichtigt werden müssen. Andererseits mangelt es bisher an Erkenntnissen über das PCK_{SR} angehender und praktizierender Biologielehrkräfte. Diesen Desideraten wird im Rahmen dieser Dissertation durch die mehrperspektivische Entwicklung, Validierung und Anwendung eines Multiple-Choice-Instruments zur Erfassung des PCK_{SR} angehender und praktizierender Biologielehrkräfte (PCK_{SR}-bio) begegnet. Dazu werden die Ergebnisse der durchgeführten Studien aus einer methodischen Perspektive diskutiert, um zu untersuchen, wann sich

Multiple-Choice-Instrumente zur Erfassung von PCK eignen und wie mit der Entwicklung dieser Instrumente verbundenen häufig genannten Herausforderungen hinsichtlich einer adäquaten Formulierung der Items, einer passenden Konstruktrepräsentation sowie der psychometrischen Eigenschaften begegnet werden kann. Darüber hinaus werden die Ergebnisse aus einer inhaltlichen Perspektive diskutiert, um zu untersuchen, wie das PCK_{SR} bei angehenden und praktizierenden Biologielehrkräften ausgeprägt ist und wie es mit der Berufserfahrung und anderen konzeptionell verwandten Variablen zusammenhängt. Zusätzlich erfolgt eine Synthese beider Perspektiven, welche eine holistische Betrachtung der Ergebnisse ermöglicht.

Im Folgenden wird dazu zunächst eine Übersicht über die theoretischen Grundlagen gegeben (Kapitel 2). Darauf aufbauend wird die Zielsetzung dieser Dissertation vorgestellt sowie eine Zusammenfassung der durchgeführten Studien gegeben (Kapitel 3). Im Anschluss werden diese durchgeführten Studien präsentiert (Kapitel 4 bis 8). Es folgt eine Zusammenfassung der zentralen Ergebnisse (Kapitel 9) sowie eine abschließende übergreifende Diskussion mit Ausblick (Kapitel 10).

2 Theoretischer Hintergrund

Um die beschriebenen Desiderate im Rahmen dieser Dissertation untersuchen zu können, müssen zunächst die theoretischen Grundlagen gelegt werden. Dazu erfolgt zunächst eine ausführliche Betrachtung des fachdidaktischen Wissens (Kapitel 2.1) und eine Konzeptualisierung von Erkenntnisgewinnungskompetenzen (Kapitel 2.2). Darauf aufbauend wird das fachdidaktische Wissen im Kompetenzbereich Erkenntnisgewinnung definiert (Kapitel 2.3). Zuletzt wird eine Übersicht über den Validierungsprozess von Testinstrumenten gegeben (Kapitel 2.4).

2.1 Fachdidaktisches Wissen

Das fachdidaktische Wissen einer Lehrkraft hat einen zentralen Einfluss auf die Unterrichtsgestaltung und damit den Lernerfolg ihrer Schüler:innen (z. B. Kunter et al., 2013; Mahler et al., 2017) und ist deshalb ein elementarer Bestandteil der fachdidaktischen Forschung geworden (z. B. van Driel et al., 2023). Da es sich um ein komplexes Konstrukt handelt, sollen im Folgenden zunächst unterschiedliche Konzeptualisierungen (Kapitel 2.1.1) sowie verschiedene Modelle (Kapitel 2.1.2) und Formen (Kapitel 2.1.3) von PCK betrachtet werden. Im Anschluss erfolgt eine Vorstellung ausgewählter empirischer Befunde (Kapitel 2.1.4) sowie die Betrachtung der nationalen und internationalen curricularen Verankerung in der Lehrkräfteausbildung (Kapitel 2.1.5). Zuletzt werden unterschiedliche Methoden und Instrumente zur Erfassung von PCK vorgestellt (Kapitel 2.1.6).

2.1.1 Integrative und transformative Konzeptualisierung fachdidaktischen Wissens

Im Laufe der Jahre haben sich verschiedene Sichtweisen auf die Konzeptualisierung² von PCK etabliert, wobei grundlegend zwei Ansätze unterschieden werden (Gess-Newsome, 1999; V. Kind, 2009; Neumann et al., 2019).

Bei einer integrativen Konzeptualisierung wird PCK als das Wissen beschrieben, welches eine Lehrkraft im Unterricht anwendet und aus der Kombination unterschiedlicher Wissensbereiche (etwa PK und CK) entsteht (Gess-Newsome, 1999; V. Kind, 2009; Neumann et al., 2019). PCK stellt somit keinen eigenständigen Wissensbereich dar, sondern kann viel eher als diejenige Schnittmenge von PK und CK aufgefasst werden, welche die Vermittlung spezifischer Inhalte an eine Lerngruppe beschreibt (Gess-Newsome, 1999; V. Kind, 2009). In empirischen Untersuchungen zeichnet sich eine integrative Konzeptualisierung dadurch aus, dass PCK nicht als eigener Wissensbereich betrachtet wird, sondern andere Wissensbereiche wie PK oder CK untersucht werden (Chan & Hume, 2019).

Im Gegensatz dazu beschreiben transformative Konzeptualisierungen PCK als eigenständigen Wissensbereich, der durch eine Synthese der zugrunde liegenden Wissensbereiche entsteht (Gess-Newsome, 1999; V. Kind, 2009; Neumann et al., 2019). PCK stellt somit einen für den Unterricht spezifischen Wissensbereich dar (V. Kind, 2009; Neumann et al., 2019) und entspricht mehr als der Schnittmenge von PK und CK. In empirischen Studien zeichnet sich eine transformative Konzeptualisierung durch die Betrachtung von PCK als eigenständigen Wissensbereich aus, welcher häufig unterschiedliche Komponenten umfasst (Chan & Hume, 2019).

² Synonym werden auch die Begriffe „Modellierungen“ (z. B. in Gess-Newsome, 1999; Kind, 2009) oder „Perspektiven“ (z. B. in Neumann et al., 2019) verwendet. Da diese Begriffe im Rahmen dieser Dissertation anderweitig genutzt werden, wurde der Begriff „Konzeptualisierungen“ (z. B. in Chan & Hume, 2019) zur Vermeidung von Missverständnissen gewählt.

Gess-Newsome (1999) beschreibt diese Konzeptualisierungen als zwei Extrema eines Kontinuums. Auch andere Autor:innen betonen die Unterschiede zwischen diesen Ansätzen, aber ebenso die Möglichkeit zur Kombination beider Konzeptualisierungen (vgl. Neumann et al., 2019). In empirischen Studien werden hauptsächlich transformative Konzeptualisierungen zugrunde gelegt, wobei sich auf unterschiedliche PCK-Modelle berufen wird (Chan & Hume, 2019). Im Folgenden sollen einige dieser zentralen Modelle zum PCK vorgestellt werden, beginnend mit den fundamentalen Überlegungen von Shulman.

2.1.2 Unterschiedliche Modelle zum fachdidaktischen Wissen

Zuerst beschrieben wurde das Konzept des fachdidaktischen Wissens von Lee Shulman (Berry et al., 2016; Neumann et al., 2019; Pohlmann, 2019; van Driel & Berry, 2019), welcher in zwei wegweisenden Artikeln auf das Desiderat einging, dass in der damaligen Forschung, Lehrkräfteausbildung und Bildungspolitik ausschließlich das Fachwissen und pädagogische Wissen von Lehrkräften betrachtet wurde (Shulman, 1986, 1987). Dies erweiterte Shulman um das fachdidaktische Wissen, welches er als den Wissensbereich definierte, in welchem sich Lehrkräfte von Fachwissenschaftler:innen unterscheiden (Shulman, 1986, 1987):

„[Pedagogical content knowledge] represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners, and presented for instruction. Pedagogical content knowledge is the category most likely to distinguish the understanding of the content specialist from that of the pedagogue.” (Shulman, 1987, S. 8)

Nach Shulman (1987) umfasst das Professionswissen von Lehrkräften insgesamt sieben unterschiedliche Bereiche:

- (1) Fachwissen (*[subject matter] content knowledge*³): Verständnis über die Inhalte eines Fachs sowie den Zusammenhang dieser;
- (2) (Allgemeines) pädagogisches Wissen (*general pedagogical knowledge*): Fachunabhängige Strategien zur Organisation von Unterricht und Umgang mit Schüler:innen;
- (3) Fachdidaktisches Wissen (*pedagogical content knowledge*): Ein „Amalgam“⁴ aus fachlichem und pädagogischem Wissen, umfasst Kenntnisse darüber, wie das inhaltsbezogene Fachwissen einer Lehrkraft für Schüler:innen verständlich gemacht werden kann, insbesondere unter Berücksichtigung des Vorwissens dieser;
- (4) Curriculares Wissen (*curricular knowledge*): Fächerübergreifende und fachspezifische Kenntnisse über Lehrpläne und curriculare Vorgaben;
- (5) Wissen über die Eigenschaften der Lerngruppe (*knowledge of learners and their characteristics*): Zum Beispiel Wissen über Unterschiede der Lerngruppe in Bezug auf Motivation oder soziale Hintergründe;

³ Synonym wird auch der Begriff „*subject matter knowledge*“ verwendet (z. B. in Großschedl et al., 2019). Shulman bezeichnete diesen Wissensbereich 1986 als „*subject matter content knowledge*“, 1987 als „*content knowledge*“. Im Rahmen dieser Dissertation wird hinsichtlich der Konsistenz einheitlich die englischsprachige Abkürzung „CK“ verwendet.

⁴ Die von Shulman (1987) verwendete Bezeichnung von PCK als „Amalgam“ wird jedoch als uneindeutig kritisiert. So kann PCK als Amalgam sowohl hinsichtlich einer integrativen Konzeptualisierung (PCK als die Kombination von CK und PK) oder aber einer transformativen Konzeptualisierung (PCK entsteht aus Transformation von Fachwissen für Unterrichtskontexte) verstanden werden (Neumann et al., 2019). In seiner ursprünglichen Definition beschreibt Shulman PCK als eigenständigen Wissensbereich, welcher aus der Verknüpfung der Komponenten PK und CK entsteht, aber von diesen getrennt ist, sodass er einem transformativen Ansatz folgt (Neumann et al., 2019).

- (6) Wissen über Bildungskontexte (*knowledge of educational contexts*): Umfasst gesellschaftliche und politische Rahmenbedingungen und
- (7) Wissen über Bildungsziele und -zwecke sowie deren philosophische und historische Grundlagen (*knowledge of educational ends, purposes, and values, and their philosophical and historical grounds*).

Trotz der einflussreichen Rolle dieser ursprünglichen Überlegungen wurde Shulmans Definition von PCK in den folgenden Jahren aus unterschiedlichen Gründen kritisiert. So fehlte es zur damaligen Zeit an empirischer Grundlage und es wurde argumentiert, dass PCK mehr als nur das Wissen über Instruktionsstrategien sowie Vorstellungen der Lerngruppe umfassen sollte (Berry et al., 2016). Darüber hinaus wurde die Definition als ausschließlich statischer Wissensbereich, welcher unabhängig vom Unterrichtskontext entwickelt werden kann, kritisiert (Berry et al., 2016). Shulman beanstandete dies später selbst und betonte zusätzlich, dass affektive Eigenschaften der Lehrkraft wie Emotionen und Überzeugungen sowie kontextuelle, soziale und kulturelle Rahmenbedingungen in seiner ursprünglichen Definition vernachlässigt werden, obwohl diese ebenfalls einen Einfluss auf das unterrichtliche Handeln haben können (Shulman, 2015). Zusätzlich kritisiert er an seiner ursprünglichen Definition die unzureichende Beschreibung des Zusammenhangs zwischen dem PCK einer Lehrkraft und dem Lernerfolg ihrer Schüler:innen (Shulman, 2015).

Die ursprünglichen Überlegungen von Shulman dienten trotz oder gerade wegen dieser Kritik als Ausgangspunkt für die Entwicklung einer Vielzahl unterschiedlicher Modelle.

2.1.2.1 Modell von Magnusson et al. (1999)

Das Modell von Magnusson et al. (1999) stellt eines der ersten und am häufigsten genutzten PCK-Modelle dar (Chan & Hume, 2019; Gess-Newsome, 2015; van Driel et al., 2023). Es baut auf Vorarbeiten von Grossman (1990) und Tamir (1988) auf und umfasst fünf Komponenten, die das PCK einer Lehrkraft formen (Magnusson et al., 1999; siehe auch Abbildung 2.1):

- (1) Einstellungen zum Unterrichten in den Naturwissenschaften (*orientations toward science teaching*) beschreibt die Überzeugungen und das damit einhergehende Wissen einer Lehrkraft über den Zweck und das Ziel des naturwissenschaftlichen Unterrichts, welche einen entscheidenden Einfluss auf die Planung, Durchführung und Reflexion des Unterrichts haben⁵.
- (2) Wissen über das naturwissenschaftliche Curriculum (*knowledge of science curriculum*) entspricht dem von Shulman (1986, 1987) beschriebenen curricularen Wissen. Es umfasst sowohl das Wissen der Lehrkraft über vorgeschriebene Ziele und Zielsetzungen des Curriculums (*mandated goals and objectives*) als auch Kenntnisse über spezifische Lehrpläne und Materialien (*specific curricular programs and materials*).
- (3) Wissen über das Verständnis der Lerngruppe in den Naturwissenschaften (*knowledge of students' understanding of science*) umfasst einerseits Wissen über die Lernanforderungen der Schüler:innen (*knowledge of requirements for learning*), andererseits auch Wissen über gängige Schwierigkeiten, die Schüler:innen in Bezug auf relevante Themen des Unterrichts aufweisen (*knowledge of areas of students difficulty*).

⁵ Auch wenn die Überzeugungen einer Lehrkraft als eigener Teil ihrer professionellen Handlungskompetenz angesehen werden können (z. B. Baumert & Kunter, 2006), haben sie einen maßgeblichen Einfluss auf die anderen PCK-Komponenten (Magnusson et al., 1999) und können deshalb als integraler Bestandteil des PCK einer Lehrkraft angesehen werden (Friedrichsen et al., 2011).

- (4) Wissen über die Beurteilung im naturwissenschaftlichen Unterricht (*knowledge of assessment in science*) umfasst sowohl Kenntnisse darüber, *welche* Aspekte innerhalb einer Lerneinheit bewertet werden sollten (Dimensionen des naturwissenschaftlichen Lernens, die für die Bewertung wichtig sind; *knowledge of the dimensions of science learning that are important to assess*) als auch Kenntnisse darüber, *wie* diese Aspekte bewertet werden können (Methoden, mit denen das Gelernte bewertet werden kann; *knowledge of the methods by which the learning can be assessed*).
- (5) Wissen über Instruktionsstrategien (*knowledge of instructional strategies*) umfasst zwei Kategorien: Wissen über disziplinspezifische Strategien (für den naturwissenschaftlichen Unterricht spezifische Instruktionsstrategien; *knowledge of subject-specific strategies*)⁶ und das Wissen über themenspezifische Strategien (für das Unterrichten konkreter Themen spezifische Instruktionsstrategien; *knowledge of topic-specific strategies*).

Mit diesem Modell stellen Magnusson et al. (1999) zwei Ideen zur Topologie von PCK vor: Zunächst implizieren die unterschiedlichen Komponenten, dass es verschiedene Aspekte des PCK gibt, die Lehrkräfte beherrschen müssen. Lehrkräfte müssen deshalb für unterschiedliche Themenbereiche über Kenntnisse bezogen auf jede der genannten Komponenten verfügen, um diese erfolgreich unterrichten zu können (Magnusson et al., 1999). Darüber hinaus impliziert der Zusammenhang der Komponenten, dass eine fehlende Kohärenz zu einer unzureichenden Entwicklung von PCK führen kann, während eine Verbesserung des Wissens hinsichtlich einer Komponente nicht automatisch zu einem insgesamt höheren PCK führt (Magnusson et al., 1999). Das von Magnusson et al. (1999) vorgestellte Modell ist in Abbildung 2.1 dargestellt.

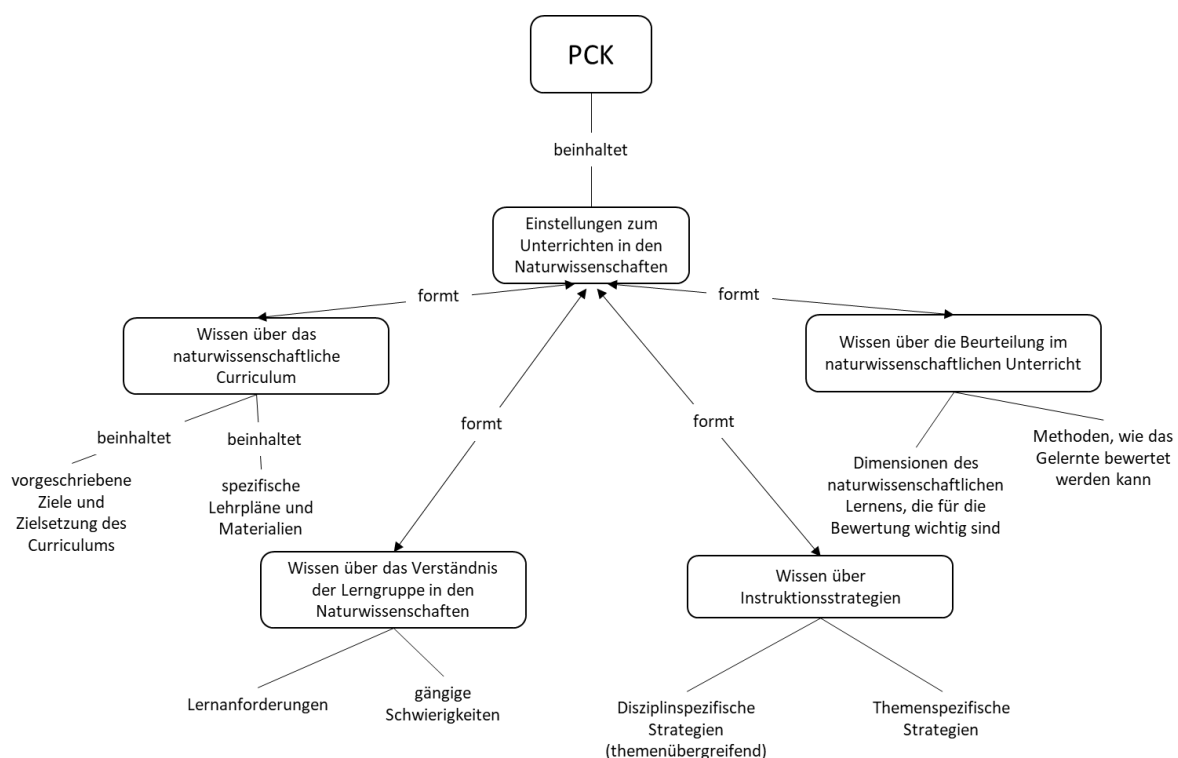


Abbildung 2.1: Modell von Magnusson et al. (1999; übersetzt)

⁶ Es wurde sich bewusst gegen die Verwendung der deutschen Übersetzung „fachspezifisch“ entschieden, da es sich um Wissen handelt, welches auch fächerübergreifend für den naturwissenschaftlichen Unterricht relevant ist. Um Missverständnisse zu vermeiden, wird im Rahmen dieser Dissertation stattdessen analog zum Refined-Consensus-Modell (s. Kapitel 2.1.2.4) die Bezeichnung „disziplinspezifisch“ verwendet.

Das Modell von Magnusson et al. (1999) beschreibt eine transformative Konzeptualisierung von PCK, da es dieses als einen eigenständigen Wissensbereich darstellt (V. Kind, 2009) und der Einsatz der definierten PCK-Komponenten im Unterricht insbesondere durch die Einstellungen einer Lehrkraft zum Unterrichten in den Naturwissenschaften bestimmt wird (Neumann et al., 2019). Die genannten Komponenten bilden die Grundlage einer Vielzahl weiterer Modelle zur Topologie des PCK, darunter das Pentagon-Modell von Park und Oliver (2008).

2.1.2.2 Pentagon-Modell

So präzisieren Park und Oliver (2008) in ihrem Pentagon-Modell (*pentagon model*) die von Magnusson et al. (1999) genannten Komponenten wie folgt (die deutsche Übersetzung der Komponenten orientiert sich an Pohlmann, 2019):

- (1) Wissen über kontextuelle Rahmenbedingungen (*orientations to teaching science*) beschreibt Überzeugungen hinsichtlich des Zwecks des Unterrichts,
- (2) Wissen über (fachspezifische) Instruktionsstrategien (*knowledge of instructional strategies for teaching science*) beschreibt Wissen über disziplin- und themenspezifische Strategien zur Vermittlung von Kompetenzen,
- (3) Wissen über Beurteilung (*knowledge of assessment of science learning*) beschreibt Wissen über verschiedene Arten der Leistungsbewertung,
- (4) Wissen über das Fachcurriculum (*knowledge of science curriculum*) beschreibt das Wissen über die Rolle des Curriculums beim Unterrichten eines spezifischen Fachs und
- (5) Wissen über die Lerngruppe (*knowledge of students' understanding in science*; KSU) beschreibt Wissen über Vorstellungen und Vorwissen der Lerngruppe.

Diese fünf Komponenten wurden in eine pentagonale Form gebracht, wobei PCK, anders als im Modell von Magnusson et al. (1999), im Zentrum steht (Park & Oliver, 2008; siehe auch Abbildung 2.2). Durch diese Darstellung wird insbesondere der bereits von Magnusson et al. (1999) beschriebene Zusammenhang zwischen den einzelnen Komponenten deutlich (Park & Oliver, 2008): Durch die Entwicklung einer Komponente kann auch gleichzeitig die Entwicklung anderer Komponenten gefördert werden, wodurch sich das PCK einer Lehrkraft insgesamt verbessert. Allerdings impliziert dies auch, dass die Entwicklung einzelner Komponenten alleine nicht automatisch einen positiven Effekt auf das Unterrichten hat, sodass für ein effektives Unterrichten eine Integration aller Komponenten notwendig ist (Park & Oliver, 2008). Zusätzlich eignet sich die Darstellung als pentagonale Form insbesondere hinsichtlich der Visualisierung PCK-bezogener Daten mittels des *PCK map approaches* (Chan, 2022; Park & Chen, 2012; Park & Suh, 2019; siehe auch Kapitel 2.1.6.2).

Über diese Topologie hinaus betonen Park und Oliver (2008) im Pentagon-Modell den dynamischen Erwerb von PCK: Lehrkräfte übernehmen nicht einfach nur bereits vorhandenes Wissen von Ausbilder:innen oder Kolleg:innen, sondern entwickeln dieses auch kontinuierlich und selbstständig basierend auf ihren eigenen Erfahrungen weiter. Dies trifft auch auf andere Wissensbereiche wie PK und CK zu, welche sich ebenfalls durch Erfahrung und Reflexion weiterentwickeln (konstruktives Verständnis von Wissenserwerb). Hinsichtlich des PCK einer Lehrkraft beschreiben Park und Oliver (2008) den Erwerb als eine Wechselbeziehung zwischen zwei Arten von Wissen:

- (1) Wissen, welches durch Reflexion während des Unterrichtens erworben wird (*reflection in action*): Lehrkraft erlangt oder modifiziert ihr PCK während der Handlung in einer spezifischen Unterrichtssituation und
- (2) Wissen, welches durch Reflexion nach dem Unterrichten erworben wird (*reflection on action*): Lehrkraft erlangt oder modifiziert ihr PCK durch Reflexion ihres Handelns im Anschluss des Unterrichts.

Zusätzlich wird PCK von Park und Oliver (2008) als idiosynkratisch beschrieben: Auch wenn unterschiedliche Lehrkräfte über vergleichbare Bestandteile von PCK verfügen können, wendet jede Lehrkraft individuelles PCK an, welches von ihren eigenen Erfahrungen und Überzeugungen sowie dem spezifischen Unterrichtskontext abhängt. Dazu konnte ebenfalls empirische Evidenz gefunden werden (Park & Chen, 2012). Das Pentagon-Modell von Park und Oliver (2008) ist in Abbildung 2.2 dargestellt.

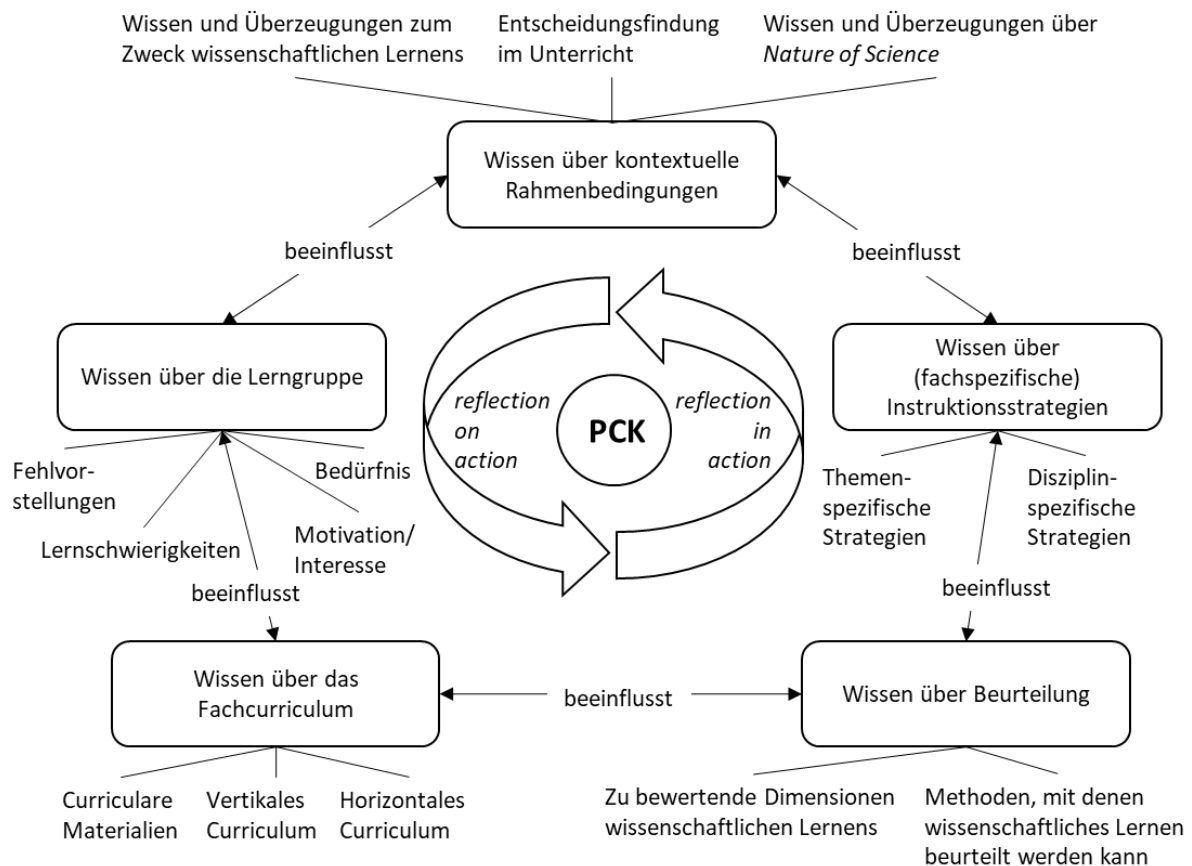


Abbildung 2.2: Pentagon Modell (adaptiert und übersetzt nach Park & Chen, 2012; Park & Oliver, 2008)

Neben dem Pentagon-Modell wurden im Laufe der Jahre eine Vielzahl weiterer Modelle bezogen auf die Topologie von PCK entwickelt (zum Beispiel von Hill et al., 2008; Mavhunga & Rollnick, 2013; Tepner et al., 2012), welche sich insbesondere hinsichtlich der betrachteten Komponenten unterscheiden. So betrachten Tepner et al. (2012) neben dem „Wissen über Schülerkognitionen“ auch das „Wissen über instruktionale Strategien“, wobei sie sich dabei auf das „Wissen über den Einsatz von theoretischen und gegenständlichen Modellen“ sowie dem „Wissen über die Planung und Durchführung von Experimenten“ fokussieren, um der besonderen Rolle der Erkenntnisgewinnung im naturwissenschaftlichen Unterricht gerecht zu werden. Hill et al. (2008) unterscheiden hingegen neben den Komponenten „Wissen über den Inhalt und die Lerngruppe“ (*knowledge of content and students*) und „Wissen über den Inhalt und das Unterrichten“ (*knowledge of content and teaching*; entspricht dem Wissen über Instruktionsstrategien) auch das „Wissen über das Curriculum“ (*knowledge of curriculum*). Im Gegensatz dazu betrachten Mavhunga und Rollnick (2013) neben den Komponenten „Vorwissen der Lerngruppe“ (*learners prior knowledge*), „konzeptuelle Instruktionsstrategien“ (*conceptual teaching strategies*) sowie „curriculare Bedeutsamkeit“ (*curricular saliency*) zusätzlich die Komponenten „Wissen darüber, was ein Thema leicht oder schwierig zu verstehen macht“ (*what*

makes a topic easy or difficult to understand) sowie „Wissen über Darstellungen einschließlich Analogien“ (*representations including analogies*).

Diese Betrachtung unterschiedlicher Komponenten hat jedoch einen entscheidenden Einfluss auf die empirische Untersuchung von PCK (Chan & Hume, 2019; Depaepe et al., 2013) und damit auch die mit den Untersuchungen einhergehenden Ergebnisse. Um dem zu begegnen, entwickelte eine Vielzahl internationaler Wissenschaftler:innen gemeinsam das Consensus-Modell.

2.1.2.3 Consensus-Modell

Um ein einheitliches Modell als Basis für zukünftige Forschung und somit ein gemeinsames Verständnis von PCK zu schaffen, wurde das Consensus-Modell (*consensus model*) von internationalen Wissenschaftler:innen bei einem *PCK Summit* im Jahr 2012 entwickelt (Berry et al., 2015). Das Consensus-Modell unterscheidet sich grundlegend von dem von Magnusson et al. (1999) vorgestellten Modell und stellt PCK, ebenso wie das Pentagon-Modell, in das Zentrum, betont zusätzlich aber auch die übergreifende Rolle des Professionswissens (Gess-Newsome, 2015; siehe auch Abbildung 2.3).

Ausgangspunkt des Consensus-Modells ist das Professionswissen der Lehrkraft (*teacher professional knowledge bases*), welches Wissen über Bewertung (*assessment knowledge*), pädagogisches Wissen (*pedagogical knowledge*), Fachwissen (*content knowledge*), Wissen über die Lerngruppe (*knowledge of students*), curriculares Wissen (*curricular knowledge*) sowie weitere darüber hinausreichende Wissensbereiche umfasst (Gess-Newsome, 2015). Es handelt sich um generische (und damit nicht inhaltspezifische), normative Wissensbereiche, die dem aktuellen Stand der Forschung entsprechen und in der Aus- und Weiterbildung von Lehrkräften vermittelt werden (Gess-Newsome, 2015).

Aus dem Professionswissen leitet sich das einer Lehrkraft zur Verfügung stehende themenspezifische professionelle Wissen (*topic-specific professional knowledge*) ab, sobald es um die Vermittlung eines spezifischen Themas geht (Gess-Newsome, 2015). Dieses Wissen umfasst Kenntnisse über die Auswahl effektiver Instruktionsstrategien und geeigneter Darstellungen im Unterricht, die Organisation des Fachinhalts, mögliche Vorstellungen und das Vorwissen der Lerngruppe sowie die Integration von naturwissenschaftlichen Denk- und Arbeitsweisen in den Unterricht (Gess-Newsome, 2015). Im Gegensatz zum PCK, welches eine individuelle Wissensbasis einer Lehrkraft darstellt (Magnusson et al., 1999; Park & Oliver, 2008; Shulman, 1986, 1987), stellt das themenspezifische professionelle Wissen viel eher eine themenspezifische kanonische Grundlage dar (Gess-Newsome, 2015).

Überzeugungen und Einstellungen (zum Beispiel epistemologisch, also transmissiv oder konstruktivistisch) sowie der affektive Zustand der Lehrkraft (Motivation, Unzufriedenheit, Effizienz oder Risikobereitschaft) haben einen Einfluss auf ihr Wissen und ihre Fähigkeiten, weshalb sie als Verstärker oder Filter für ihren Wissenserwerb sowie ihr Handeln wirken (Gess-Newsome, 2015). Auch die Erfahrung und das Vorwissen einer Lehrkraft haben einen entscheidenden Einfluss auf die Implementation ihres Professionswissens: Während Lehrkräfte mit viel Erfahrung (Expert:innen) neue Ideen und Methoden, welche sie zum Beispiel in Fortbildungen erlangt haben, eher in den Unterricht integrieren können, sind Lehrkräfte mit wenig Erfahrung (Noviz:innen) hauptsächlich mit der Klassenführung beschäftigt, sodass sie neue erworbene Praktiken seltener effektiv anwenden können (Borko & Livingston, 1989; Gess-Newsome, 2015)⁷.

⁷ Diese Ergebnisse stützen sich auf das Experten-Novizen-Paradigma (de Groot, 1965), für eine genauere Betrachtung siehe Kapitel 2.4.2.

PCK selbst kann nach dem Consensus-Modell erst im Rahmen des Unterrichtskontexts erfasst werden. Zusätzlich wird im Consensus-Modell zwischen „PCK“ und „PCK und Fähigkeiten“ (*PCK and skill; PCK&S*), einem mit PCK im Zusammenhang stehenden Konstrukt, unterschieden (Gess-Newsome, 2015):

„[...] PCK is the knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes“ (Gess-Newsome, 2015, S. 38)

„[...] PCK&S is the act of teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes“ (Gess-Newsome, 2015, S. 38)

PCK wird nach diesem Modell also als das individuell verfügbare, durch Erfahrung erlangte themenspezifische Wissen einer Lehrkraft beschrieben, welches nicht zwangsläufig dem kollektiven Wissen vieler Expert:innen entspricht (Gess-Newsome, 2015). Im Gegensatz dazu beschreibt PCK&S die Handlungen, die eine Lehrkraft in einer spezifischen Unterrichtssituation ausführt (Gess-Newsome, 2015).

Die getrennte Betrachtung von PCK und PCK&S macht einerseits deutlich, dass Lehrkräfte, die über spezifisches PCK verfügen, dieses nicht automatisch anwenden (bedingt durch die Filter/Verstärker) sowie andererseits, dass beispielsweise die Auswahl effektiver Instruktionsstrategien im Unterricht nicht zwangsläufig auf ihrem PCK beruht (Gess-Newsome, 2015). Vor dem Hintergrund der Forschung ermöglicht dies Wissenschaftler:innen insbesondere eine getrennte Beschreibung des zu messenden Konstrukts, bei der PCK erstmals nicht nur als Wissensbasis für die Planung und Durchführung von Unterricht definiert wird, sondern auch als eine Fähigkeit, die beim Unterrichten zum Tragen kommt (Gess-Newsome, 2015).

In der Unterrichtspraxis spielt neben dem PCK einer Lehrkraft auch der Unterrichtskontext (*classroom context*) eine entscheidende Rolle, welcher Faktoren umfasst, die von der Lehrkraft nicht immer kontrolliert werden können (zum Beispiel das Curriculum oder die Anzahl der zu betreuenden Klassen; Gess-Newsome, 2015). Auch der Lernerfolg der Schüler:innen (*student outcomes*) wird im Consensus-Modell explizit gemacht, wobei dieser nicht automatisch aus der Instruktion durch die Lehrkraft entsteht, sondern ebenfalls durch Verstärker und Filter beeinflusst wird (zum Beispiel externe und demografische Einflussfaktoren, Vorwissen und Vorstellungen; Gess-Newsome, 2015). Umgekehrt können die Schüler:innen auch einen Einfluss auf das Geschehen im Klassenraum haben (Gess-Newsome, 2015). Das Consensus-Modell ist in Abbildung 2.3 dargestellt.

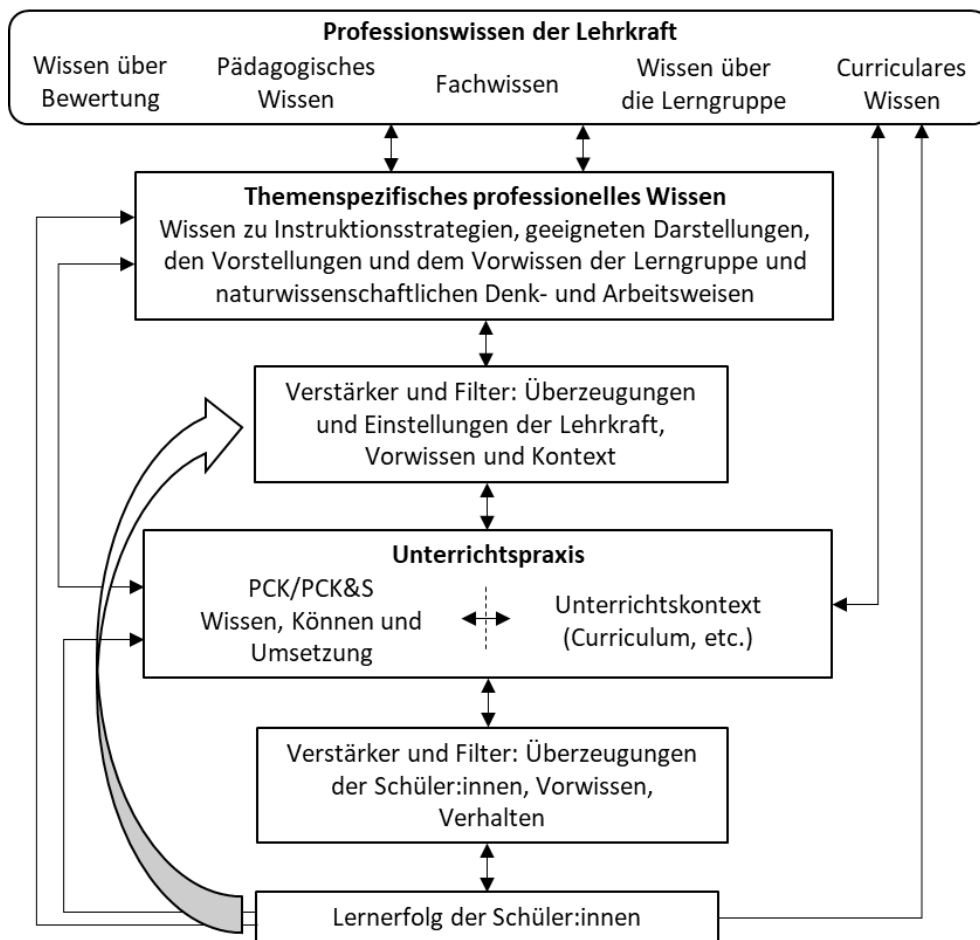


Abbildung 2.3: Consensus-Modell (übersetzt nach Gess-Newsome, 2015)

Das Consensus-Modell ist rekursiv, der Lernerfolg der Schüler:innen sowie die Unterrichtspraxis haben einen Einfluss auf das themenspezifische professionelle Wissen sowie das Professionswissen der Lehrkraft (Gess-Newsome, 2015). Darüber hinaus zeichnet es sich insbesondere durch die Einbettung von dynamischem, individuellem PCK mit Einfluss auf die Unterrichtspraxis aus, welches im Kontrast zum kanonischen professionellen Wissen steht (Gess-Newsome, 2015).

Auch wenn das Consensus-Modell in Übereinstimmung von Expert:innen entwickelt wurde, zeigen sich dennoch Uneinigkeiten (Chan & Hume, 2019). So wird kritisiert, dass die Komplexität des Konstrukts in dem Modell nicht ausreichend dargestellt wird (Carlson et al., 2019). Darüber hinaus ist die Unterscheidung zwischen PCK und PCK&S nicht eindeutig trennbar, was zu Unstimmigkeit hinsichtlich der Zuordnung elementarer Komponenten (zum Beispiel Wissen über Instruktionsstrategien oder Wissen über die Lerngruppe) führt (Carlson et al., 2019). Zusätzlich ist PCK&S als die Handlung einer Lehrkraft im Unterricht definiert („*act of teaching*“; Gess-Newsome, 2015, S. 38), weshalb es nicht das Wissen einer Lehrkraft, sondern viel eher die Fähigkeit zur Anwendung von PCK beschreibt.

2.1.2.4 Refined-Consensus-Modell

Als Rückmeldung auf die Kritik am Consensus-Modell wurde in einem erneuten *PCK Summit* im Jahr 2017 das bestehende Consensus-Modell von internationalen Wissenschaftler:innen überarbeitet und als Refined-Consensus-Modell (*refined consensus model*) vorgestellt (Hume et al., 2019; siehe auch Abbildung 2.4).

Im Refined-Consensus-Modell werden drei Bereiche (*realms*) unterschieden (Carlson et al., 2019): kollektives PCK (*collective PCK*; cPCK), persönliches PCK (*personal PCK*; pPCK) und angewandtes PCK (*enacted PCK*; ePCK). Das ePCK steht dabei im Zentrum und beschreibt die pädagogischen Überlegungen einer Lehrkraft in Form individuellen Wissens und spezifischer Fähigkeiten, welche sie in einer konkreten Unterrichtssituation mit einer spezifischen Lerngruppe einsetzt, um ausgewählte Inhalte zu vermitteln (Carlson et al., 2019). Es umfasst neben dem Wissen und den Überzeugungen, die eine Lehrkraft in dieser spezifischen Unterrichtssituation zeigt (*reflection in action*) auch Wissen über die Planung und Reflexion des Unterrichts (*reflection on action*; Carlson et al., 2019) und entspricht damit dem durch Reflexion erworbenen Wissen aus dem Pentagon-Modell (vgl. Park & Oliver, 2008). Lehrkräfte brauchen spezifisches ePCK bezogen auf das Planen, Unterrichten und Reflektieren als unterschiedliche Prozesse, welche sich nicht nur auf eine vollständige Unterrichtsstunde, sondern auch auf einzelne Situationen währenddessen (zum Beispiel der Reflexion einer Schüler:innenantwort und anschließende Reaktion) beziehen (Alonzo et al., 2019). Bei diesen grundlegenden dynamischen und pädagogischen Überlegungen greift eine Lehrkraft auf Teile ihres gesamten PCK zurück, welches als pPCK bezeichnet wird (Carlson et al., 2019).

Das pPCK beschreibt das kumulative, im Laufe der Zeit gesammelte PCK einer einzelnen Lehrkraft und umfasst neben dem Wissen, welches durch eigene Lehrerfahrungen erlangt wurde, auch solches, welches im beruflichen Austausch (zum Beispiel mit Kolleg:innen) oder durch eigenständige Recherche (zum Beispiel das Lesen von Fachliteratur) gewonnen wurde (Carlson et al., 2019). Es dient als Grundlage für das in einer spezifischen Unterrichtssituation gezeigte ePCK (Carlson et al., 2019). So wählt eine Lehrkraft in einer solchen Situation beispielsweise diejenige Instruktionsstrategie aus ihrem Repertoire aus, von der sie ausgeht, dass sie am geeignetsten für den Umgang mit der entsprechenden Lerngruppe ist (Carlson et al., 2019). Der Wissensaustausch zwischen ePCK und pPCK erfolgt dabei in beide Richtungen (Carlson et al., 2019): Das neue Wissen, das eine Lehrkraft in einer Unterrichtssituation sammelt, ergänzt ihr bisheriges pPCK, während das ePCK, welches sie in einer Unterrichtssituation anwendet, von ihrem pPCK abhängt und durch ihre Einstellungen und Überzeugungen gefiltert wird.

Unabhängig von der Lehrkraft können sowohl der Unterrichtskontext (zum Beispiel das Curriculum) als auch die Eigenschaften der Schüler:innen als Verstärker oder Filter für ihren Lernerfolg agieren (analog zum Consensus-Modell; Gess-Newsome, 2015), weshalb eine Lehrkraft gute Kenntnisse über ihre Lerngruppe benötigt (Carlson et al., 2019). Diese Verstärker und Filter trennen das individuelle pPCK zusätzlich von dem kollektiven cPCK und mediierten damit ebenfalls das Handeln einer Lehrkraft (Carlson et al., 2019).

Das cPCK beschreibt das gesamte vorhandene PCK (zum Beispiel zu geeigneten Instruktionsstrategien oder gängigen Vorstellungen von Schüler:innen zu diversen Themen), welches als Wissensgrundlage für den naturwissenschaftlichen Unterricht dient (Carlson et al., 2019). Es handelt sich damit um das Wissen, über das mehr als eine einzelne Person verfügt und das zwischen Forschenden, Lehrkräften und anderen Expert:innen geteilt wird (Carlson et al., 2019). Dabei beschreibt cPCK sowohl das Wissen einer spezifischen Gruppe (zum Beispiel das Kollegium an einer Schule; lokales kollektives PCK; *local collective PCK*) als auch das Wissen, welches von der Öffentlichkeit geteilt wird und oft auf den Ergebnissen aktueller Forschung beruht (kanonisches PCK; *canonical PCK*; Carlson et al., 2019). Es reicht dabei von disziplinspezifischem PCK (*disciplin-specific PCK*) über themenspezifisches PCK

(*topic-specific PCK*) bis zu konzeptspezifischem PCK (*concept-specific PCK*), wobei sich dieses Kontinuum auch auf die anderen Bereiche (ePCK und pPCK) übertragen lässt (Carlson et al., 2019).

Weitere Wissensbereiche bilden die Grundlage für das cPCK und umfassen neben Fachwissen das pädagogisches Wissen, Wissen über die Schüler:innen, curriculare Wissen und Wissen über die Bewertung (Carlson et al., 2019). Diese generischen und damit fachunabhängigen Bereiche unterscheiden sich vom Fachwissen, welches eine zentrale Rolle spielt und nicht nur Kenntnisse über die zu unterrichtenden Themen, sondern auch über die Arbeitsweisen des jeweiligen Fachs umfasst (Carlson et al., 2019). Das Refined-Consensus-Modell ist in Abbildung 2.4 dargestellt.

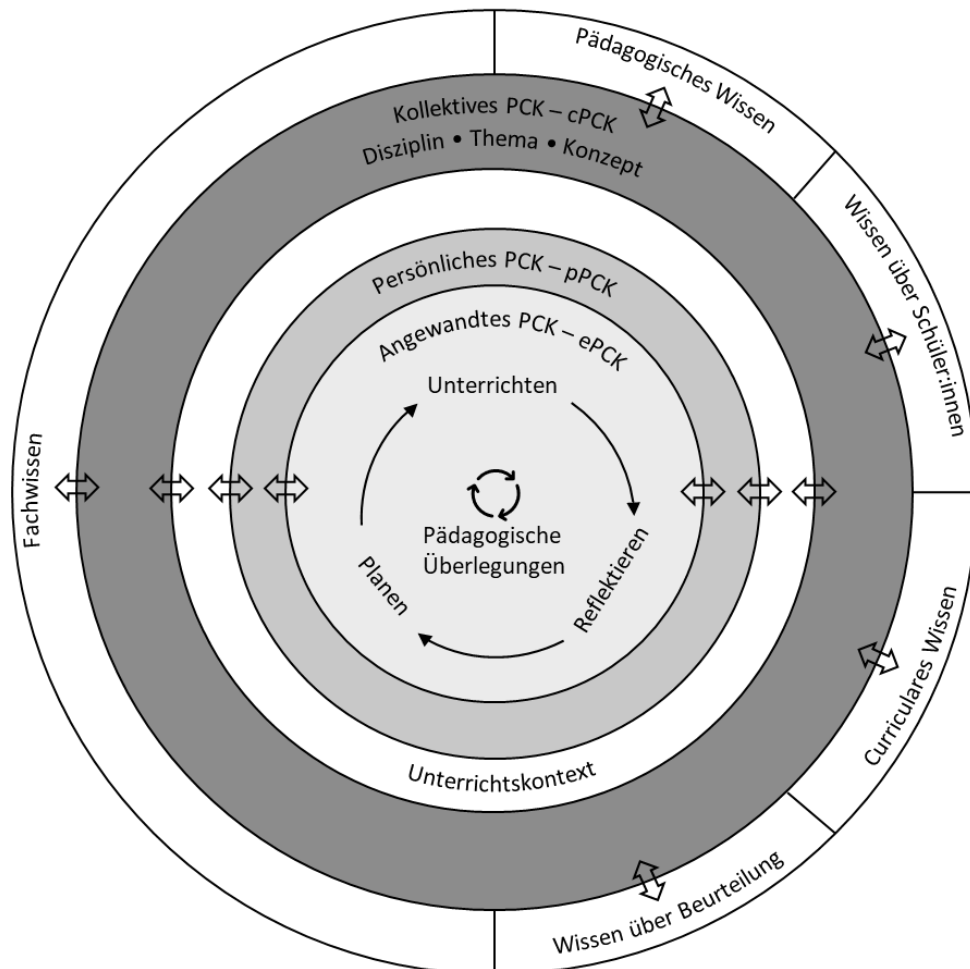


Abbildung 2.4: Refined-Consensus-Modell (adaptiert und übersetzt nach Carlson et al., 2019)

Die zentrale Rolle von ePCK in Form eines Zyklus aus Planen, Unterrichten und Reflektieren für den Erwerb von pPCK einer Lehrkraft konnte bereits empirisch gestützt werden (Behling et al., 2022b; Forsler et al., 2023). Das Refined-Consensus-Modell nimmt darüber hinaus insbesondere Rücksicht auf die Rolle der sozialen Interaktion mit Kolleg:innen und kontextuellen Faktoren bei der Entwicklung von PCK und zeichnet sich durch einen Fokus auf den Unterrichtskontext als Verstärker oder Filter des PCK einer Lehrkraft aus (Park, 2019). Dennoch empfehlen Forschende, weitere Filter zwischen den Ebenen des Refined-Consensus-Modells hinzuzufügen (zum Beispiel epistemische Überzeugungen oder Fachwissen), da diese ebenfalls einen Einfluss auf die Entwicklung von PCK haben können (Behling et al., 2022a).

Zusammengefasst lässt sich erkennen, dass PCK-Modelle wie das von Magnusson et al. (1999) oder das Pentagon-Modell primär die Topologie des PCK hinsichtlich der Struktur und Verbindung unterschiedlicher PCK-Komponenten beschreiben, während der Fokus im Consensus-Modell sowie Refined-Consensus-Modell auf der Einbettung von PCK im Unterrichtskontext liegt. Die Modelle können damit in Ergänzung zueinander betrachtet werden.

2.1.3 Unterschiedliche Formen fachdidaktischen Wissens

Ergänzend zu den beschriebenen Modellen herrscht weitgehende Einigkeit darüber, dass PCK in unterschiedlichen Formen beschrieben werden kann, wobei viele Autor:innen zwischen dynamischem (auch prozeduralem) und statischem (auch deklarativem) PCK unterscheiden (Alonzo & Kim, 2016; Barendsen & Henze, 2019; Berry et al., 2016; Jüttner et al., 2013; Schmelzing et al., 2013; She & Chan, 2023): Dynamisches PCK beschreibt das Wissen, das in einer bestimmten Unterrichtssituation gezeigt wird, deklaratives PCK das gesamte PCK, über das eine Lehrkraft unabhängig von einer spezifischen Unterrichtssituation verfügt. Diese unterschiedlichen Formen finden sich auch im Consensus-Modell als PCK (statisch) sowie PCK&S (dynamisch) (Gess-Newsome, 2015) und im Refined-Consensus-Modell als ePCK (dynamisch) sowie pPCK und cPCK (statisch; Alonzo et al., 2019; Carlson et al., 2019; She, 2023) wieder. Diese Unterscheidung ist insbesondere bei der Entwicklung und Auswahl geeigneter Erhebungsmethoden und -instrumente relevant (siehe Kapitel 2.1.6). Im Folgenden werden zunächst jedoch ausgewählte empirische Befunde sowie die curriculare Verankerung vorgestellt.

2.1.4 Ausgewählte empirische Befunde

Das Professionswissen als Kern der professionellen Handlungskompetenz ist Bestandteil zahlreicher Untersuchungen, wobei insbesondere die zentralen Wissensbereiche PK, CK und PCK in den letzten Jahrzehnten umfassend erforscht wurden. So konnte für diverse Fächer, darunter in der Biologie (z. B. Großschedl et al., 2019), der Mathematik (z. B. Baumert et al., 2010) oder der Physik (z. B. Sorge et al., 2019) empirische Evidenz dafür gefunden werden, dass es sich um voneinander getrennte, aber miteinander korrelierende Wissensbereiche handelt.

Darüber hinaus herrscht weitgehende Einigkeit darüber, dass Lehrkräfte über für einen Lerngegenstand spezifisches PCK verfügen müssen, häufig als „*topic-specific*“ (Abell, 2008; Chan & Hume, 2019; Depaepe et al., 2013), im deutschsprachigen Raum auch als „themenspezifisch“ (z. B. Pohlmann, 2019), bezeichnet. Diese Bezeichnung kann jedoch zu Missverständnissen führen, da Lehrkräfte nicht nur über PCK zu spezifischen Themen ihres Unterrichtsfachs (im Sinne von fachlichen Inhalten) verfügen müssen (zum Beispiel PCK zu Evolution: Becerra et al., 2022; PCK zu nachhaltiger Entwicklung: Forsler et al., 2023), sondern auch zu disziplinspezifischen Konzepten wie naturwissenschaftlicher Erkenntnisgewinnung und damit zusammenhängenden Denk- und Arbeitsweisen (zum Beispiel PCK zum Modellieren: Günther et al., 2019; PCK zur Durchführung wissenschaftlicher Untersuchungen: van Uum et al., 2021; PCK zu wissenschaftlichem Denken: Lin & Wang, 2024). Dies geht mit unterschiedlichen Instruktionsstrategien einher und wurde bereits in dem Modell von Magnusson et al. (1999) als die Kategorien „Wissen über themenspezifische Strategien“ und „Wissen über disziplinspezifische Strategien“ unterschieden (vgl. Kapitel 2.1.2.1). Auch im Rahmen des Refined-Consensus-Modells beschreiben Carlson et al. (2019) PCK als ein Kontinuum von konzeptspezifischem zu disziplinspezifischem Wissen⁸.

⁸ Im Rahmen dieser Dissertation wurde sich dennoch dafür entschieden, den Begriff „themenspezifisch“ zu verwenden, da er in der fachdidaktischen Forschung etabliert ist. Wenn nicht explizit anders beschrieben, umfasst der Begriff damit auch disziplinspezifische Inhalte.

2.1.4.1 Zentrale Komponenten fachdidaktischen Wissens

Da in empirischen Untersuchungen häufig unterschiedliche PCK-Modelle zugrunde gelegt werden, unterscheiden sich diese Untersuchungen auch hinsichtlich der betrachteten PCK-Komponenten (Chan & Hume, 2019). Zwei Komponenten können dabei als zentral angesehen werden (van Driel et al., 2023): Das Wissen über Instruktionsstrategien (*knowledge of instructional strategies for teaching science*; KISR) sowie das Wissen über die Lerngruppe (*knowledge of students understanding*; KSU). Dies konnte ebenfalls empirisch gestützt werden. So führten Park und Chen (2012) eine qualitative Studie mit vier Biologielehrkräften durch, bei der sie Unterrichtsbeobachtungen und halbstrukturierte Interviews sowie die verwendeten Unterrichtsmaterialien und die damit zusammenhängenden Antworten der Schüler:innen analysierten. Dabei zeigte sich, dass die Komponenten KISR und KSU einen entscheidenden Bestandteil des PCK von Lehrkräften ausmachten. Auch Alonzo et al. (2012) führten eine qualitative Studie durch, bei der mehrere videografierte Unterrichtsstunden zum Thema Optik von zwei Physiklehrkräften untersucht wurden. Es konnte gezeigt werden, dass die zwei Komponenten KISR und KSU eine zentrale Rolle in der Unterrichtspraxis spielten (Alonzo et al., 2012).

Darüber hinaus zeigt sich, dass KISR und KSU die am häufigsten untersuchten Komponenten in Studien zu PCK sind (Chan & Hume, 2019). Die Komponenten sollten jedoch nicht losgelöst voneinander betrachtet werden, da insbesondere auch ihre Verbindung zentral ist (vgl. van Driel et al., 2023; siehe auch Pentagon-Modell, Kapitel 2.1.2.2). So beschreibt die Verbindung KISR→KSU die Wirkung ausgewählter Instruktionsstrategien auf das Lernen einer Lerngruppe und KSU→KISR, wie unterrichtet werden muss, um angemessen mit den Vorstellungen einer Lerngruppe umzugehen.

2.1.4.2 Einfluss auf die Unterrichtsqualität und Lernleistung von Schüler:innen

Empirische Untersuchungen zum Einfluss auf die Unterrichtsqualität kommen zu uneinheitlichen Ergebnissen. Eine Vielzahl von Studien konnte einen positiven Zusammenhang zwischen dem PCK von Lehrkräften und dem Lernerfolg ihrer Schüler:innen für unterschiedliche naturwissenschaftliche Fächer zeigen. In einer Studie von She et al. (2024) erfassten diese das PCK von 147 Biologielehrkräften mit einem offenen Antwortformat, bei dem die Lehrkräfte Fragen zu Videoclips und Transkripten von Unterrichtssituationen beantworten sollten. Dabei konnte ein signifikant positiver Zusammenhang des PCK mit der Lernleistung der Schüler:innen gezeigt werden (She et al., 2024). Sadler et al. (2013) führten eine Studie mit 181 Physiklehrkräften durch, in der die Lehrkräfte ein Multiple-Choice-Instrument beantworteten, in welchem sie typische Vorstellungen von Schüler:innen identifizieren mussten. Es zeigte sich, dass die Schüler:innen von Lehrkräften, welche neben einem fundierten CK ebenfalls über ein hohes Wissen über die Vorstellungen der Lerngruppe verfügten, einen höheren Lernerfolg erzielten (Sadler et al., 2013). Jin et al. (2015) führten eine Studie mit 25 Biologielehrkräften durch, welche eine auf Lernprogression basierende Unterrichtseinheit zu dem Thema „Wachstum und Gasaustausch bei Pflanzen“ unterrichteten. Die Ergebnisse zeigen einen positiven Einfluss des PCK der Lehrkräfte auf die Lernleistung ihrer Schüler:innen (Jin et al., 2015). Auch Mahler et al. (2017) konnten in einer Studie mit 48 Biologielehrkräften und ihren 1036 Schüler:innen einen signifikant positiven Zusammenhang zwischen dem PCK einer Lehrkraft und dem Lernerfolg ihrer Schüler:innen zeigen. Darüber hinaus konnten auch Kunter et al. (2013) im Rahmen der COACTIV-Studie (Baumert et al., 2011) einen positiven Effekt von PCK auf die Unterrichtsqualität messen. Sie untersuchten dabei 181 Mathematiklehrkräfte und ihre Schüler:innen ($n = 4353$) in einer longitudinalen Betrachtung über ein Jahr (Kunter et al., 2013). Baumert et al. (2010) konnten für diese Stichprobe zusätzlich zeigen, dass der positive Einfluss des PCK auf den Lernerfolg der Schüler:innen durch die Bereitstellung von kognitiver Aktivierung sowie der individueller Lernunterstützung durch die Lehrkraft mediiert wurde.

Vergleichbare Ergebnisse erzielten auch Förtsch et al. (2016) in der Auswertung videobasierter Beobachtungen einer Unterrichtsstunde zum Thema Neurobiologie von 39 Biologielehrkräften. Dabei zeigte sich ein indirekter Effekt des PCK auf die Lernleistung der Schüler:innen, welcher durch die kognitive Aktivierung dieser durch die Lehrkraft mediiert wurde (Förtsch et al., 2016). Auch Keller et al. (2017) konnten in einer Studie mit 77 Physiklehrkräften zeigen, dass das PCK einer Lehrkraft die Lernleistung ihrer Schüler:innen positiv beeinflusst, wobei dieser Effekt ebenfalls durch die kognitive Aktivierung der Schüler:innen durch die Lehrkraft mediiert wurde.

Andere Studien konnten diesen Effekt von PCK auf die Lernleistung der Schüler:innen nicht nachweisen. In einer Studie von Gess-Newsome et al. (2019) zeigte sich, dass das PCK von 50 betrachteten Biologielehrkräften keinen signifikanten Einfluss auf die Lernleistung ihrer Schüler:innen hatte. Auch Yang et al. (2020) konnten keinen signifikanten Zusammenhang zwischen dem PCK von 204 an einer interdisziplinären Fortbildung teilnehmenden Lehrkräften und der Lernleistung ihrer Schüler:innen feststellen.

Empirische Studien geben außerdem Hinweise darauf, dass das PCK den potenziell positiven Einfluss des CK auf die Unterrichtsqualität⁹ mediiert. Eine Studie von Kulgemeyer und Riese (2018) zeigte, dass das CK von angehenden Physiklehrkräften ($n = 109$) nur dann einen positiven Einfluss auf ihre Erklärleistung hatte, wenn diese ein hohes PCK aufwiesen, was darauf schließen lässt, dass PCK eine wichtige Voraussetzung für die adressatengerechte Vermittlung des CK einer Lehrkraft ist. Weitere empirische Befunde legen nahe, dass das CK einer Lehrkraft eine wichtige Voraussetzung für die Entwicklung ihres PCK ist (Großschedl et al., 2015; Krauss et al., 2008; Riese & Reinhold, 2012; Schiering et al., 2022).

2.1.4.3 Entwicklung von fachdidaktischem Wissen

Eine häufig beschriebene Voraussetzung für die Entwicklung des PCK von Lehrkräften ist ihre Berufserfahrung sowie die Reflexion dieser (z. B. van Driel & Berry, 2019). Empirische Studien kommen diesbezüglich jedoch zu teilweise widersprüchlichen Ergebnissen.

In einer Studie von Luft und Zhang (2014) wurden jährlich Interviews mit 76 neu eingestellten Lehrkräften über einen Verlauf von drei Jahren geführt, in denen diese die Planung und Durchführung ausgewählter Unterrichtsstunden beschrieben. Die Daten wurden im Anschluss mit Fokus auf die PCK-Komponenten KSU und KISR ausgewertet (Luft & Zhang, 2014). Die Ergebnisse zeigten eine signifikante Entwicklung des PCK der Lehrkräfte über den Verlauf der drei Jahre, wobei im ersten Jahr die stärkste Zunahme erfolgte und sich zwischen dem zweiten und dritten Jahr ein leichter Rückgang erkennen ließ (Luft & Zhang, 2014). Schiering et al. (2022) konnten in einer Studie mit angehenden Physiklehrkräften vier unterschiedliche Kompetenzniveaus von PCK identifizieren, welche von inhaltsunspezifischem Wissen zu inhaltspezifischen Strategien zur Planung und Durchführung von Unterricht reichten. Dazu beantworteten die Teilnehmenden einen Paper-Pencil-Test mit gemischtem Aufgabenformat, mit welchem die von Magnusson et al. (1999) genannten Komponenten KSU, KISR, „Wissen über das naturwissenschaftliche Curriculum“ sowie „Wissen über die Beurteilung im naturwissenschaftlichen Unterricht“ zu unterschiedlichen Inhaltsbereichen erfasst wurden (Schiering et al., 2022). Es wurden insgesamt 427 Rückmeldungen ausgewertet, wobei einige Teilnehmenden im Rahmen einer Längsschnitterhebung zu unterschiedlichen Zeitpunkten mehrfach teilnahmen (Schiering et al., 2022). Die Ergebnisse zeigen, dass in den niedrigen Kompetenzniveaus insbesondere das CK der Lehrkräfte einen Einfluss auf die Entwicklung ihres PCK hat, zum Erreichen von höheren Niveaus jedoch auch Unterrichtserfahrung notwendig ist (Schiering et al., 2022). In einer anderen Studie untersuchten

⁹ Ein entsprechender positiver Einfluss von CK auf die Unterrichtsqualität und damit die Lernleistung der Schüler:innen konnte in einigen der betrachteten Studien gezeigt werden (z. B. Gess-Newsome, 2019), in anderen wiederum nicht (z. B. Mahler et al., 2017).

E. Lee et al. (2007) die Entwicklung des PCK von 24 Lehrkräften im ersten Jahr ihrer Lehrtätigkeit. Dazu führten sie Interviews zu Beginn und am Ende des Schuljahres sowie mehrere Unterrichtsbeobachtungen in dessen Verlauf durch und bestimmten drei Niveaustufen (limitiert, grundlegend, kompetent), denen sie das PCK der Lehrkräfte bezogen auf die PCK-Komponenten KSU und KISR zuordneten (E. Lee et al., 2007). Es zeigte sich eine statistisch signifikante Verbesserung des PCK der Lehrkräfte in der Kategorie KSU, nicht aber in der Kategorie KISR (E. Lee et al., 2007). Ein Großteil der Lehrkräfte blieb darüber hinaus auf einem limitierten oder grundlegenden PCK-Niveau (E. Lee et al., 2007).

Andere Studien konnten hingegen keinen Zusammenhang zwischen dem PCK von Lehrkräften und ihrer Berufserfahrung nachweisen. Friedrichsen et al. (2009) untersuchten Unterrichtsentwürfe mit anschließenden Folgeinterviews von vier angehenden Biologielehrkräften, von denen zwei über eine zweijährige Berufserfahrung verfügten, die anderen hingegen über keine. Grundlage für die Auswertung der Daten bildeten die von Magnusson et al. (1999) beschriebenen PCK-Komponenten (Friedrichsen et al., 2009). Es zeigte sich, dass beide Gruppen über ähnlich gering ausgeprägtes PCK verfügten und bei der Planung ihres Unterrichts hauptsächlich auf ihr PK zurückgriffen (Friedrichsen et al., 2009). S. Kirschner et al. (2017) führten eine Studie mit 171 Biologie-, 110 Chemie-, sowie 172 Physiklehrkräften durch, bei der diese einen fächerspezifischen Paper-Pencil-Test mit gemischtem Antwortformat beantworteten. Dabei wurden die PCK-Komponenten „Wissen über Experimente“, „Wissen über Modelle/Konzepte“ sowie KSU betrachtet (S. Kirschner et al., 2017). Für die Chemielehrkräfte konnte kein Zusammenhang zwischen ihrem PCK und ihrer Berufserfahrung gezeigt werden, für die Biologie- und Physiklehrkräfte hingegen sogar ein negativer Zusammenhang (S. Kirschner et al., 2017).

Eine mögliche Erklärung für die unterschiedlichen Ergebnisse der betrachteten Studien stellen die verschiedenen Erhebungsmethoden und -instrumente sowie die Betrachtung unterschiedlicher PCK-Komponenten dar. Aber auch andere Faktoren können neben der Berufserfahrung einen Einfluss auf die Entwicklung des PCK einer Lehrkraft haben und so zu konträren Ergebnissen führen. Eine Studie von Kulgemeyer et al. (2021) gibt beispielsweise Hinweise darauf, dass die Reflexionsfähigkeit von Lehrkräften eine Voraussetzung für die Entwicklung ihres pPCK ist.

Darüber hinaus konnten Kleickmann et al. (2013) in einer Querschnittstudie zeigen, dass es große Unterschiede zwischen dem PCK angehender Mathematiklehrkräfte zu Beginn und zum Ende des Studiums sowie zum Ende des Referendariats gab. Lehrkräfte mit hoher Berufserfahrung (durchschnittlich ca. 21 Jahre) wiesen hingegen nur ein geringfügig höheres PCK im Vergleich zu der Gruppe der Referendar:innen auf (Kleickmann et al., 2013). Diese Unterschiede im PCK angehender Lehrkräfte konnten auch in anderen empirischen Studien gezeigt werden (z. B. Großschedl et al., 2019), ebenso wie die Entwicklung über den Verlauf der universitären Ausbildung in Längsschnittstudien (z. B. Mahler et al., 2024), wobei insbesondere Lerngelegenheiten für die Entwicklung des PCK relevant scheinen (Blömeke & Delaney, 2012; Riese & Reinhold, 2012). Schiering et al. (2021) konnten zusätzlich zeigen, dass die Anzahl der Lerngelegenheiten den Einfluss des CK auf das PCK von Studierenden moderiert, was impliziert, dass bei einer geringen Anzahl an Lerngelegenheiten im Studium insbesondere das Fachwissen eine entscheidende Rolle auf die Entwicklung von PCK hat.

Damit angehende Lehrkräfte dieses grundlegende PCK bereits während ihrer universitären Ausbildung entwickeln können, ist dies sowohl in den deutschen Bildungsstandards (KMK, 2024a) als auch in internationalen Curricula (z. B. AITSL, 2017) verankert.

2.1.5 Nationale und internationale curriculare Verankerung in der Lehrkräfteausbildung

Die in der deutschen Lehramtsausbildung zu erlangenden Kompetenzen sind in den „Ländergemeinsamen inhaltlichen Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung“ festgelegt (KMK, 2024a). Für den universitären Teil der Lehramtsausbildung sollen Lehrkräfte unterschiedliche Kompetenzen erlangen, welche den fachdidaktischen Wissenskomponenten „Wissen über Instruktionsstrategien“, „Wissen über die Lerngruppe“ und „Wissen über Bewertung“ (vgl. Magnusson et al., 1999; Park & Oliver, 2008) entsprechen (KMK, 2024a; s. Tabelle 2.1). Im anschließenden Vorbereitungsdienst soll dieses Wissen weiter vertieft (zum Beispiel „Unterricht in heterogenen Lerngruppen planen, durchführen und analysieren“; KMK, 2024a, S. 4) und um weitere Wissensbereiche (zum Beispiel dem Organisationswissen; „die Fähigkeit in multiprofessionellen Teams zu kooperieren“; KMK, 2024a, S. 4) ergänzt werden.

Tabelle 2.1: Übersicht über die fachdidaktischen Kompetenzen zum Ende der universitären Ausbildung (KMK, 2024a, S. 4) und zugehörige PCK-Komponenten

Genannte Kompetenzen	PCK-Komponente
Studienabsolvent:innen...	
...haben ein solides und strukturiertes Wissen über fachdidaktische Positionen und Strukturierungsansätze und können fachwissenschaftliche bzw. fachpraktische Inhalte auf ihre Bildungswirksamkeit hin und unter didaktischen Aspekten analysieren,	Wissen über Instruktionsstrategien
...sind in der Lage, komplexe Sachverhalte adressatengerecht, auch in einfacher Sprache, darzustellen,	Wissen über die Lerngruppe
...kennen und nutzen Ergebnisse fachdidaktischer und lernpsychologischer Forschung über das Lernen in ihren Fächern bzw. Fachrichtungen,	Wissen über Instruktionsstrategien
...kennen die Grundlagen fach- bzw. fachrichtungs- und anforderungsgerechter Leistungsbeurteilung,	Wissen über Beurteilung
...haben fundierte Kenntnisse über Merkmale von Schülerinnen und Schülern, die den Lernerfolg fördern oder hemmen können und darüber, wie daraus Lernumgebungen differenziert zu gestalten sind.	Wissen über die Lerngruppe

Anmerkung. Die Komponenten sind auf Basis des Modells von Magnusson et al. (1999) sowie des Pentagon-Modells (Park & Oliver, 2008) zugeordnet.

Auch international ist der Erwerb von PCK in der Lehrkräfteausbildung curricular verankert (z. B. Chile: BCTC, 2019; Kanada: Mineduc, 2012). In Australien¹⁰ sind Kompetenzen, die angehende Lehrkräfte während ihres Studiums sowie in der darauffolgenden außeruniversitären Aus- und Weiterbildung erlangen sollen, in den „*Australian Professional Standards for Teaching*“ verankert (AITSL, 2017). In Bezug auf das fachdidaktische Wissen ist in diesen Standards explizit vermerkt, dass (angehende) Lehrkräfte über Wissen hinsichtlich der Lerngruppe („*know students and how they learn*“; Standard 1) sowie hinsichtlich des Fachinhalts und des Unterrichts dieses Fachinhalts („*know the content and how to teach it*“; Standard 2) verfügen sollten (AITSL, 2017). Dies entspricht den zentralen PCK-Komponenten KSU und KISR. Darüber hinaus sollen sie Wissen in den Bereichen „Berufspraxis“ (*professional practice*) und „professionelles Handeln“ (*professional engagement*) erlangen (AITSL, 2017).

Auf das Grundstudium mit Lehramtsperspektive folgt in Deutschland in der Regel ein konsekutiver Master, sodass sich fachwissenschaftliche, pädagogische und fachdidaktische Inhalte über die gesamte universitäre Ausbildung verteilen. In dem im Anschluss folgenden schulpraktischen Teil der Lehramtsausbildung in Form eines Vorbereitungsdienstes (auch Referendariat genannt) unterrichten die Anwärter:innen bereits selbst in der Schule, werden dabei aber von ausgebildeten Lehrkräften mentoriert. In Australien unterscheidet sich die Lehramtsausbildung. Hier absolvieren angehende Lehrkräfte zunächst ein fachwissenschaftliches Grundstudium (Bachelor) und entscheiden sich im Anschluss für einen lehramtsbezogenen Master, in welchem ein inhaltlicher Schwerpunkt auf die Vermittlung von pädagogischem und fachdidaktischem Wissen gelegt wird. Nach Abschluss des Studiums können die Absolvent:innen direkt mit der Tätigkeit als praktizierende Lehrkräfte beginnen.

Um die geforderte Entwicklung von PCK in der Lehrkräfteausbildung (und Praxis) zu evaluieren, bedarf es geeigneter Erhebungsmethoden und -instrumente. Die verschiedenen Konzeptualisierungen sowie die Verwendung unterschiedlicher Modelle und das daraus resultierende Fehlen einer einheitlichen Definition von PCK führt zu einer unterschiedlichen Erfassung des Konstrukts (Chan & Hume, 2019; Smith & Banilower, 2015), was ebenfalls mit der Verwendung unterschiedlicher Erhebungsmethoden einhergeht (Chan & Hume, 2019; Depaepe et al., 2013). Im folgenden Kapitel werden deshalb die Erfassung von PCK und damit einhergehende unterschiedliche Erhebungsmethoden und -instrumente betrachtet.

2.1.6 Erfassung von fachdidaktischem Wissen

Als Wissensbasis ist PCK ein latentes Konstrukt, welches nicht direkt beobachtet werden kann, sondern über Aus- oder Bewertungen von (geschriebenen) Antworten auf Fragen oder Interviews oder der Bewertung des unterrichtlichen Handelns operationalisiert werden muss (Smith & Banilower, 2015). Dabei sollte nicht nur die Themenspezifität des Konstrukts berücksichtigt werden, sondern bei einer transformativen Konzeptualisierung zusätzlich beachtet werden, dass PCK mehr als den Schnittpunkt von PK und CK darstellt (Abell, 2008; Smith & Banilower, 2015). Insbesondere wenn das PCK einer Lehrkraft nicht nur beschrieben, sondern auch als korrekt oder inkorrekt bewertet werden soll, bedarf es eines einheitlichen kanonischen Wissens, welches als Grundlage für diese Bewertung dienen kann (Smith & Banilower, 2015). Dies stellt vor dem Hintergrund einer fehlenden einheitlichen Definition eine besondere Herausforderung dar (Smith & Banilower, 2015).

¹⁰ Exemplarisch für die internationale curriculare Verankerung werden die Standards in Australien näher betrachtet, da im Rahmen dieser Dissertation eine Studie mit australischen Studierenden durchgeführt wurde (Studie 5; Kapitel 8).

Es gibt diverse Möglichkeiten zur Erfassung des PCK von Lehrkräften, welche Chan und Hume (2019) in einem systematischen Literaturreview untersuchten. Dabei konnten sie zwei fundamental unterschiedliche Ansätze feststellen (Chan & Hume, 2019): die Auswertung von selbst gegebenen Antworten (*self-reports*)¹¹ und die Analyse des Verhaltens während des Unterrichtens (*performance in teaching tasks*). Für die Auswertung selbst gegebener Antworten bieten sich dabei unterschiedliche Arten von Instrumenten und Erhebungsmethoden an. So können geschlossene Paper-Pencil-Tests, in denen die Antworten der Teilnehmenden als korrekt oder inkorrekt bewertet werden (wobei das Urteil von Expert:innen oder kanonisches PCK als Bewertungsgrundlage genutzt wird), offene Fragebögen oder Interviews mit anschließender qualitativer Auswertung (zum Beispiel Inhaltsanalyse) oder Selbstberichte, in denen das eigene unterrichtliche Handeln oder die eigenen Fähigkeiten durch eine Lehrkraft (schriftlich oder in Interviews) selbstständig eingeschätzt wird, verwendet werden (Chan & Hume, 2019). Diese Erhebungsmethoden haben den Nachteil, dass sie als weniger authentisch und damit realitätsferner wahrgenommen werden können, insbesondere im Vergleich zur Analyse des Verhaltens einer Lehrkraft im Unterricht (Chan & Hume, 2019). Aus dem PCK, das eine Lehrkraft durch die Erfassung selbst gegebener Antworten zeigt, lässt sich also nicht zwangsläufig schließen, dass sie dieses PCK auch in realen Unterrichtssituationen anwendet (Chan & Hume, 2019), insbesondere, da dieses durch kontextuelle Faktoren beeinflusst werden kann (Behling et al., 2022a; Carlson et al., 2019; Gess-Newsome, 2015). Die Erfassung selbst gegebener Antworten hat allerdings den Vorteil, das PCK einer Lehrkraft situationsunabhängig abzubilden.

Beide Ansätze unterscheiden sich damit insbesondere darin, welche Form von PCK erhoben wird (Chan & Hume, 2019): Mit der Erfassung selbst gegebener Antworten kann PCK erfasst werden, über das eine Lehrkraft situationsunabhängig verfügt (deklaratives PCK; pPCK und cPCK), während die Analyse des Verhaltens während des Unterrichts dazu genutzt werden kann, zu erfassen, wie die Lehrkraft dieses PCK in der Praxis anwendet (dynamisches PCK; ePCK). Bei der Auswahl von geeigneten Erhebungsinstrumente sollte deshalb insbesondere darauf geachtet werden, welcher Bereich des PCK erfasst wird (Stender & Brückmann, 2020). Soll das dynamische ePCK von Lehrkräften erfasst werden, eignen sich je nach betrachtetem Prozess unterschiedliche Erhebungsmethoden, etwa die Analyse von Unterrichtsentwürfen hinsichtlich der Planung (z. B. Großmann & Krüger, 2023), (videobasierte) Unterrichtsbeobachtungen hinsichtlich des Unterrichtens (z. B. Alonzo et al., 2012) oder Interviews zur Auswertung der Reflexion im Anschluss an eine Unterrichtssituation (z. B. Park & Suh, 2019). Zur Erfassung des deklarativen pPCK eignen sich wiederum Paper-Pencil-Tests, wobei zwischen Instrumenten mit offenen (z. B. Krauss et al., 2008; Loughran et al., 2004), geschlossenen (Multiple-Choice-Instrumenten; z. B. Sadler et al., 2013; PCK_{SR}-bio, s. Kapitel 12.1) und gemischten Antwortformaten (z. B. Großschedl et al., 2019; Jüttner et al., 2013) unterschieden werden kann. Das ebenfalls dynamische cPCK lässt sich, soweit es als das kollektive PCK einer Gruppe (statt des kanonischen PCK der gesamten Community; Carlson et al., 2019) verstanden wird, durch einen Paper-Pencil-Tests, welcher gemeinsam in der betrachteten Gruppe bearbeitet wird, erfassen (Buldu & Buldu, 2021; Forsler et al., 2023)¹². Eine Übersicht über diese möglichen Erhebungsmethoden und -instrumente ist in Tabelle 2.2 dargestellt.

¹¹ Chan und Hume (2019) verwenden den Begriff "*self-reports*", welcher im Deutschen auch als "Selbstbericht" missverstanden werden kann. Da Chan und Hume (2019) mit dem Begriff "*self-reports*" jedoch nicht nur das Erfassen von Selbstauskünften, zum Beispiel auf Ratingskalen oder in Interviews, sondern auch objektive Erfassungen durch offene oder geschlossene Instrumente beschreiben, wird zur Vermeidung von Missverständnissen im Rahmen dieser Dissertation die Bezeichnung "selbst gegebene Antworten" verwendet.

¹² Andere Forschende, zum Beispiel Stender und Brückmann (2020), verstehen cPCK als die Übereinstimmung des PCK individueller Lehrkräfte mit dem kanonischen PCK und pPCK als die individuellen Ausprägungen des PCK einer Lehrkraft. Dieses unterschiedliche Verständnis hat einen Einfluss auf die Wahl der Erhebungsmethoden zur Erfassung der jeweiligen Bereiche und wird in Kapitel 10.1 kritisch diskutiert.

Tabelle 2.2: Mögliche Erhebungsmethoden und -instrumente zur Analyse von unterschiedlichen Bereichen des PCK

Bereich	Form	Mögliche Auswertungsmethoden oder -instrumente	Exemplarische Referenz(en)	
ePCK	Planen	Dynamisch	Analyse von Unterrichtsentwürfen	Großmann und Krüger (2023)
	Unterrichten	Dynamisch	(Videobasierte) Unterrichtsbeobachtungen	Alonzo et al. (2012)
	Reflektieren	Dynamisch	Interviews	Park und Suh (2019)
pPCK	Deklarativ		Paper-Pencil-Tests mit offenem Antwortformat	Krauss et al. (2008); Loughran et al. (2004)
			Paper-Pencil-Tests mit geschlossenem Antwortformat	Sadler et al. (2013); PCK _{SR} -bio (Kapitel 12.1)
			Paper-Pencil-Tests mit gemischtem Antwortformat	Großschedl et al. (2019); Jüttner et al. (2013)
cPCK	Deklarativ		Paper-Pencil-Tests, die von Gruppen bearbeitet werden	Buldu und Buldu (2021); Forsler et al. (2023)

Anmerkung. Tabelle übersetzt und adaptiert aus Studie 4 (siehe Kapitel 7, Table 7.1).

Im Rahmen dieser Dissertation wird ein Fokus auf Multiple-Choice-Instrumente zur Erfassung von PCK gelegt. Dazu werden zunächst mit der Entwicklung dieser Instrumente einhergehende Herausforderungen näher betrachtet. Im Anschluss (Kapitel 2.1.6.2) erfolgt zusätzlich eine Betrachtung des *PCK map approaches* als gängige Auswertungsmethode für PCK-bezogene Daten.

2.1.6.1 Multiple Choice-Instrumente

Paper-Pencil-Tests zur Erfassung des deklarativen pPCK von Lehrkräften können unterschiedliche Aufgabenformate aufweisen. So eignen sich Instrumente mit offenen Antwortoptionen insbesondere dazu, das PCK einer Lehrkraft in seiner Tiefe abzubilden (Chan & Hume, 2019; Smith & Banilower, 2015) und damit ein breites Spektrum an Denkprozessen, die der Antwort zugrunde liegen, zu erfassen (Martinez, 1999). Solche offenen Paper-Pencil-Tests erfordern allerdings ein hohes Maß an Interpretation und gehen häufig ebenfalls mit einer langen Bearbeitungszeit sowie einem hohen Zeitaufwand bei der qualitativen Auswertung einher (Hollingworth et al., 2007; Smith & Banilower, 2015). Instrumente mit geschlossenen Antwortoptionen (zum Beispiel Multiple-Choice-Instrumente) haben dahingehend Vorteile in Bezug auf Testökonomie und Objektivität (Haladyna, 2004) und können dadurch insbesondere bei großen Stichproben eingesetzt werden (Hill et al., 2008). Instrumente mit gemischten Aufgabenformaten (z. B. Großschedl et al., 2019; Jüttner et al., 2013) kombinieren die Vorteile beider Aufgabenformate, sind allerdings oftmals ebenfalls mit einer hohen Bearbeitungszeit verbunden (zum Beispiel 50 Minuten für 24 Items in Jüttner et al., 2013).

Die Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK geht allerdings mit unterschiedlichen Herausforderungen einher. So muss bei der Formulierung von Multiple-Choice-Items und entsprechenden Antwortoptionen im Allgemeinen auf eine eindeutige Formulierung geachtet werden (Haladyna, 2004). Insbesondere bei der Entwicklung von Multiple-Choice-Items zur Erfassung von PCK muss sichergestellt werden, dass die Attraktoren (korrekte Antwortoptionen) mit Sicherheit korrekt und die Distraktoren (inkorrekte Antwortoptionen) mit Sicherheit inkorrekt sind (Hill et al., 2008; Schmelzing et al., 2013) und damit dem aktuellen kanonischen cPCK entsprechen (Smith & Banilower, 2015). Auch Hill et al. (2008) betont die Schwierigkeit der Entwicklung eindeutig inkorrektor Distraktoren, die aber dennoch als mögliche Auswahloptionen authentisch erscheinen.

Eine weitere Herausforderung, die mit der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK verbunden ist, bezieht sich auf die Konstruktrepräsentation. So können Multiple-Choice-Instrumente nur eine begrenzte Anzahl an Komponenten erfassen, weshalb insbesondere das durch Unterrichtserfahrung gesammelte PCK einer Lehrkraft nicht vollständig abgebildet werden kann (Schmelzing et al., 2013). Gleichzeitig muss das Konstrukt jedoch angemessen erfasst werden, zumal sich die erfassten PCK-Komponenten je nach Definition unterscheiden (Smith & Banilower, 2015). Bei der Entwicklung von Testinstrumenten muss deshalb insbesondere beachtet werden, dass das zugrunde liegende Konstrukt nicht über- oder unterrepräsentiert wird (AERA et al., 2014; Smith & Banilower, 2015). So werden bei einer Unterrepräsentation nicht alle relevanten Aspekte des Konstrukts durch den Test erfasst (AERA et al., 2014). Um dem vorzubeugen, ist es im Entwicklungsprozess notwendig, dem Instrument eindeutig definierte PCK-Komponenten zugrunde zu legen (Smith & Banilower, 2015) und das zugrunde liegende Konstrukt auf diese Weise eindeutig zu definieren (Messick, 1996). Eine Überrepräsentation des Konstrukts beschreibt hingegen, dass das Testergebnis des Instruments von Faktoren beeinflusst wird, die nicht zu dem zu erfassenden Konstrukt gehören (AERA et al., 2014). In Bezug auf Multiple-Choice-Instrumente zur Erfassung von PCK kann eine Überrepräsentation dazu führen, dass statt PCK unter Umständen eher CK oder die Überzeugungen zum Unterrichten einer Lehrkraft erfasst werden (Smith & Banilower, 2015).

Eine weitere Herausforderung bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK kann in Bezug zu den psychometrischen Eigenschaften eines Testinstruments auftreten. So weisen Multiple-Choice-Instrumente, die komplexe und anwendungsbezogene Konstrukte messen, häufig eine niedrige interne Konsistenz¹³ im Vergleich zu herkömmlichen Skalen oder Fragebögen auf (Taber, 2018). Dies spiegelt sich bei einigen bereits veröffentlichten Multiple-Choice-Instrumenten zur Erfassung von PCK wider (zum Beispiel $\alpha = .53$ in Förtsch et al., 2016), wobei es auch Instrumente gibt, die als Paper-Pencil-Tests mit gemischtem Antwortformat zumindest teilweise aus Multiple-Choice-Items bestehen und eine hohe interne Konsistenz aufweisen (z. B. Großschedl et al., 2019: $\alpha = .92$). Im Allgemeinen gibt es jedoch kein eindeutiges und akzeptiertes Limit bezüglich der internen Konsistenz, ab dem Testinstrumente als akzeptabel gelten (Schmitt, 1996; Taber, 2018). Auch Instrumente, die eine niedrige interne Konsistenz aufweisen, können eine valide Interpretation des Testwertes ermöglichen (Cronbach, 1951; Schmitt, 1996; Taber, 2018). Darüber hinaus zeigt sich, dass die interne Konsistenz bei Multiple-Choice-Instrumenten zur Erfassung von PCK mit höherem Ausbildungsstand steigt (Kleickmann et al., 2013), was damit zusammenhängen kann, dass die Konsistenz des Antwortverhaltens auch von dem vorhandenen Wissen der befragten Person abhängt (z. B. Krepf et al., 2018; s. Kapitel 2.4.2).

¹³ Die interne Konsistenz beschreibt eine Möglichkeit zur stabilen Schätzung der Reliabilität (Grad der Messgenauigkeit eines Tests), bei der die mittlere Korrelation aller einzelnen Items miteinander untersucht wird (Döring, 2023). Gängige Maße sind die Kuder-Richardson Formel (bei dichotomen Items, zum Beispiel Wissenstests mit korrekter und inkorrektor Antwort) oder Cronbachs Alpha (bei polytomen Items, zum Beispiel Ratingskalen mit mehr als zwei Antwortmöglichkeiten).

Wegen dieser Herausforderungen entscheiden sich manche Forschende bewusst dazu, keine Multiple-Choice-Instrumente zur Erfassung von PCK zu verwenden (Rosenkränzer et al., 2017) oder stellen sogar in Frage, ob sich PCK mit Multiple-Choice-Instrumenten adäquat erfassen lässt (Smith & Banilower, 2015). Vor diesem Hintergrund soll im Rahmen dieser Dissertation untersucht werden, in welchen Anwendungsgebieten sich Multiple-Choice-Instrumente zur Erfassung von PCK eignen und wie damit verbundenen häufig genannten Herausforderungen hinsichtlich einer adäquaten Formulierung der Items, einer passenden Konstruktrepräsentation sowie den psychometrischen Eigenschaften begegnet werden kann.

Neben diesen unterschiedlichen Erhebungsmethoden gibt es auch verschiedene Möglichkeiten, die erfassten Daten auszuwerten. Spezifisch für die Auswertung PCK-bezogener Daten eignet sich der *PCK map approach*, welcher im Folgenden näher betrachtet wird.

2.1.6.2 Der *PCK map approach*

Der *PCK map approach* stellt eine häufig verwendete Auswertungsmethode dar, die speziell dazu verwendet werden kann, Daten bezogen auf das PCK von Lehrkräften zu analysieren (Chan, 2022; Park & Chen, 2012; Park & Suh, 2019). In der deutschsprachigen Forschung wurde diese Methode erstmals von Pohlmann (2019) unter dem Begriff „PCK-Kartenansatz“ vorgestellt¹⁴.

Auf Grundlage des Pentagon-Modells (Park & Oliver, 2008; s. Kapitel 2.1.2.2) bietet der *PCK map approach* die Möglichkeit, durch quantitative Inhaltsanalyse erhobene Ausprägungen der fünf im Modell betrachteten PCK-Komponenten und insbesondere ihre Verbindungen zu visualisieren (Park & Chen, 2012; Park & Suh, 2019). Typischerweise erfolgt die Analyse der Daten dabei in drei Schritten (Park & Chen, 2012; Park & Suh, 2019): Zunächst werden die Daten qualitativ analysiert, indem die zugrunde liegenden PCK-Komponenten identifiziert werden (zum Beispiel durch Kodierung). PCK-Komponenten, die gemeinsam in einer Analyseseinheit¹⁵ auftreten, werden im Anschluss als Verbindung zwischen diesen Komponenten gezählt und summiert. Dabei werden keine qualitativen Unterschiede hinsichtlich der Stärke der Ausprägungen einzelner Verbindungen vorgenommen (Park & Chen, 2012). Im letzten Schritt werden die Daten in eine pentagonale Form gebracht, wobei die Häufigkeit der genannten Verbindungen zwischen zwei Komponenten durch unterschiedlich starke Linien visualisiert wird. Diese *PCK maps* können anschließend hinsichtlich ihrer Muster kontrastiert oder zum Zweck der Triangulation mit den Ergebnissen anderer Methoden zur Auswertung derselben Daten vergleichend analysiert werden (Park & Suh, 2019).

Der *PCK map approach* stellt ein etabliertes Verfahren dar, welches zur Auswertung von PCK-bezogenen Daten verwendet werden kann, die auf unterschiedliche Weisen erhoben wurden (Chan, 2022). So können Daten basierend auf (videobasierten) Unterrichtsbeobachtungen (z. B. in Aydin & Boz, 2013; Bayram-Jacobs et al., 2019; Park & Chen, 2012; Suh & Park, 2017), Interviews (z. B. in Aydin & Boz, 2013; Ottogalli & Bermudez, 2024; Park & Chen, 2012; Suh & Park, 2017) oder Unterrichtsplänen (z.B. in Bayram-Jacobs et al., 2019; Großmann & Krüger, 2022; Park & Chen, 2012) sowie weiteren Methoden erhoben werden (vgl. Chan, 2022). Dabei werden neben dem *PCK map approach* weitere Erhebungs- und Auswertungsmethoden genutzt, was insbesondere vor dem Hintergrund der Triangulation sinnvoll ist (Chan, 2022).

¹⁴ Da in anderen deutschsprachigen Studien dennoch der Begriff „*PCK map approach*“ verwendet wird (z. B. in Großmann und Krüger, 2022), wurde sich dafür entschieden, im Rahmen dieser Dissertation ebenfalls die englischsprachige Bezeichnung zu verwenden.

¹⁵ Der Umfang einer Analyseseinheit wird im Voraus festgelegt und umfasst beispielsweise einen Satz oder eine Sinneinheit.

Mit Verwendung des *PCK map approaches* gehen jedoch auch einige Einschränkungen einher. So führt die Summierung der Anzahl der Verbindungen zwischen den PCK-Komponenten mit gleicher Stärke ohne qualitative Betrachtung zu einer vereinfachten Darstellung ebendieser (Park & Chen, 2012). Aydin und Boz (2013) gehen auf diese Einschränkung ein, indem sie die Stärke der Verbindungen qualitativ in Bezug auf die Nützlichkeit hinsichtlich des Lernerfolgs der Schüler:innen kodieren. Darüber hinaus vernachlässigt der *PCK map approach* kontextuelle Einflussfaktoren auf das PCK von Lehrkräften (Park & Suh, 2019). Zusätzlich ist diese Art der Datenauswertung zeitintensiv, was dazu führt, dass ein Großteil der Studien mit kleinen Stichproben ($n < 50$) durchgeführt wird (Chan, 2022).

2.2 Erkenntnisgewinnungskompetenzen

Naturwissenschaftliche Erkenntnisgewinnung ist ein elementarer Bestandteil der naturwissenschaftlichen Bildung (Driver et al., 2010; Khan & Krell, 2019; Mathesius et al., 2016; Mayer, 2007; Osborne, 2013) und entsprechende Kenntnisse sind eine notwendige Voraussetzung dafür, dass sich Schüler:innen mit aktuellen gesellschaftlichen Themen wie dem Klimawandel oder globalen Pandemien auseinandersetzen können (Gresch et al., 2013; J. Lederman et al., 2024; Zeidler et al., 2005). Schüler:innen sollten daher nicht nur über Wissen zu den Inhalten eines naturwissenschaftlichen Fachs verfügen, sondern auch ein Verständnis darüber erlangen, wie diese Erkenntnisse zustande kommen und wie zuverlässig sie eingeschätzt werden können (Bybee, 2013; J. Lederman et al., 2024; OECD, 2022). Entsprechende Kompetenzen sollten demnach im naturwissenschaftlichen Unterricht gefördert werden, weshalb diese als einer von vier Kompetenzbereichen in den deutschen Bildungsstandards (z. B. KMK, 2024b) sowie ebenfalls international in den Curricula diverser Länder verankert sind (z. B. Australien: ACARA, 2022; siehe Kapitel 2.2.2).

Im Zusammenhang mit naturwissenschaftlicher Erkenntnisgewinnung existieren eine Vielzahl unterschiedlicher Begriffe, welche verschiedene Konzepte beschreiben, aber nicht immer klar voneinander getrennt oder eindeutig definiert werden. Während in der deutschsprachigen Forschung die Begriffe naturwissenschaftliche Grundbildung (z. B. Fischler et al., 2018), wissenschaftliches Denken (z. B. Mayer, 2007) oder naturwissenschaftliche Denk- und Arbeitsweisen (z. B. Krüger & Upmeyer zu Belzen, 2021) im Zusammenhang mit naturwissenschaftlicher Erkenntnisgewinnung verwendet werden, sind in der englischsprachigen Forschung Begriffe wie *scientific literacy* (z. B. Osborne, 2023), *scientific reasoning* beziehungsweise *scientific reasoning competencies* (z. B. Krell, Vorholzer & Nehring, 2022), *scientific practices* (z. B. Osborne, 2014b) oder *scientific inquiry* (z. B. Chen & She, 2015) gängig. Im Folgenden sollen diese Begriffe deshalb zunächst definiert und miteinander in Beziehung gesetzt werden.

Scientific literacy, in der deutschsprachigen Forschung auch naturwissenschaftliche Grundbildung genannt¹⁶, beschreibt die Fähigkeit, naturwissenschaftliche Denk- und Arbeitsweisen im Alltag anwenden zu können, um unter Rücksichtnahme wissenschaftlicher Aspekte persönliche oder gesellschaftliche Entscheidungen zu treffen (Fischler et al., 2018; Osborne, 2023; Schwartz et al., 2023). *Scientific literacy* beschreibt damit also die Fähigkeit, sich reflektiert mit wissenschaftlichen Themen auseinander zu setzen, welche mit Kompetenzen in Bezug auf das wissenschaftliche Erklären von Phänomenen, der Durchführung und Evaluierung wissenschaftlicher Untersuchungen sowie die wissenschaftliche Interpretation von Daten und Erkenntnissen einhergeht (OECD, 2019). Dabei kommen dem Konzept *scientific literacy* unterschiedliche Bedeutungen zu, so zum Beispiel die zentrale Rolle von Wissenschaft in der Gesellschaft oder das Verständnis und die Verwendung wissenschaftlicher Fachsprache (Osborne, 2023). Es kann als eine notwendige Voraussetzung für die gesellschaftliche Teilhabe angesehen werden (Fischler et al., 2018), indem sich Schüler:innen durch wissenschaftliches Denken (*scientific reasoning*) mit der Welt auseinandersetzen können (J. Lederman et al., 2024) und ist deshalb ein grundlegendes Ziel des naturwissenschaftlichen Unterrichts (zum Beispiel in Deutschland: KMK, 2024b, 2024c, 2024d; Australien: ACARA, 2022).

¹⁶ Die Bezeichnung „naturwissenschaftliche Grundbildung“ wird allerdings auch kritisch gesehen, da *scientific literacy* nicht nur das Wissen über wissenschaftliche Inhalte und Konzepte umfasst, sondern auch die Fähigkeit zur Anwendung damit zusammenhängender Praktiken, welche jedoch über den klassischen Bildungsbegriff hinausgehen (Fischler et al., 2018). Aus diesem Grund wurde sich dafür entschieden, im Rahmen dieser Dissertation den englischsprachigen Begriff „*scientific literacy*“ zu verwenden.

Wissenschaftliches Denken (*scientific reasoning*)¹⁷ beschreibt eine auf Dispositionen beruhende Fähigkeit, die für die Anwendung empirischer Erkenntnismethoden zur Lösung naturwissenschaftlicher Probleme erforderlich ist (P. Kind & Osborne, 2017; Krell, Khan et al., 2022; Lawson, 2004) und umfasst neben deklarativem auch prozedurales und epistemisches Wissen (s. Kapitel 2.2.1). Diese Fähigkeiten im Bereich wissenschaftlichen Denkens sind zur erfolgreichen Durchführung wissenschaftlicher Untersuchungen (*scientific inquiry*) erforderlich (Bicak et al., 2021; Chen & She, 2015). In der deutschsprachigen Forschung wird dabei typischerweise zwischen Denk- und Arbeitsweisen unterschieden (Lübeck, 2020; Upmeier zu Belzen & Krüger, 2019; Wellnitz & Mayer, 2013). Arbeitsweisen (*scientific practices*) beschreiben ausführbare Tätigkeiten zur Erlangung neuer Erkenntnisse (zum Beispiel Experimentieren, Modellieren; Gropengießer, 2023; Osborne, 2014b; Wellnitz & Mayer, 2016). Denkweisen beschreiben hingegen Teilprozesse, die in Arbeitsweisen vorkommen können (zum Beispiel Hypothesen formulieren; Krell & Krüger, 2022; Upmeier zu Belzen & Krüger, 2019; Wellnitz & Mayer, 2016).

Die genannten Konzepte stehen also miteinander in Beziehung: Eine Person mit hoher *scientific literacy* kann ihre Fähigkeit zum wissenschaftlichen Denken (*scientific reasoning*) nutzen, um durch die Anwendung ausgewählter Arbeitsweisen (*scientific practices*) und den damit verbundenen Denkweisen fundierte Lösungen auf naturwissenschaftliche Problemstellungen zu erlangen. Im naturwissenschaftlichen Unterricht können diese Kompetenzen durch die Durchführung von wissenschaftlichen Untersuchungen (*scientific inquiry*) vermittelt werden.

Im deutschsprachigen Raum beschreibt Mayer (2007) für den Kompetenzbereich Erkenntnisgewinnung ein Rahmenmodell bestehend aus den drei zentralen Dimensionen praktische Arbeitstechniken (*practical work*), wissenschaftliche Erkenntnismethoden (*scientific inquiry*) sowie Charakteristika der Naturwissenschaften (*nature of science*). Diese Dimensionen sind durch die Konstrukte manuelle Fähigkeiten (*practical skills*), wissenschaftliches Denken (*scientific reasoning*) sowie Wissenschaftsverständnis (*epistemological beliefs*) systematisch miteinander verbunden (Mayer, 2007; siehe Abbildung 2.5).

¹⁷ Ebenfalls kann der Begriff „*scientific thinking*“ genutzt werden, welcher jedoch weiter gefasst werden kann und neben der Anwendung empirischer Erkenntnismethoden auch das Denken über Wissenschaft sowie damit zusammenhängende Denkprozesse umfasst (Dunbar und Klahr 2013).

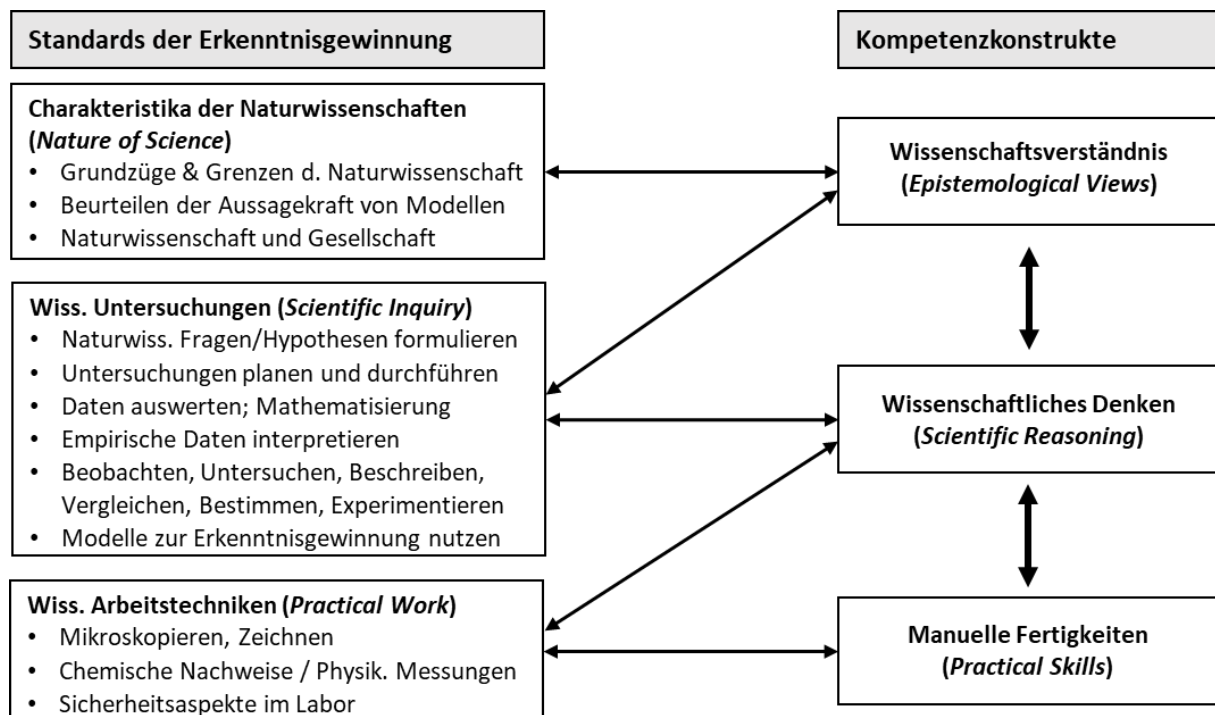


Abbildung 2.5: Rahmenmodell für den Kompetenzbereich Erkenntnisgewinnung (Mayer, 2007, S. 178)

Erkenntnisgewinnungskompetenzen umfassen damit all diejenigen Fähigkeiten, die zur wissenschaftlichen Problemlösung sowie zur Reflexion dieses Prozesses notwendig sind (P. Kind & Osborne, 2017; Krell, Khan et al., 2022; Lawson, 2004; Morris et al., 2012). Im Rahmen dieser Dissertation werden die Begriffe „Erkenntnisgewinnungskompetenz“ und „*scientific reasoning competencies*“ synonym verwendet¹⁸. Im Folgenden erfolgt deshalb zunächst eine Konzeptualisierung von Erkenntnisgewinnungskompetenzen (Kapitel 2.2.1) sowie eine Betrachtung der nationalen und internationalen curricularen Verankerung (Kapitel 2.2.2). Im Anschluss wird die Rolle von Erkenntnisgewinnungskompetenzen im Lehramtsstudium (Kapitel 2.2.3) sowie im naturwissenschaftlichen Unterricht (Kapitel 2.2.4) herausgestellt.

2.2.1 Konzeptualisierung von Erkenntnisgewinnungskompetenzen

Erkenntnisgewinnungskompetenzen umfassen neben deklarativem Wissen (*declarative knowledge*)¹⁹ auch prozedurales (*procedural knowledge*) und epistemisches Wissen (*epistemic knowledge*). Deklaratives Wissen beschreibt dabei das Wissen über naturwissenschaftliche Fachinhalte und -konzepte („wissen, was“; „*knowing what*“; P. Kind & Osborne, 2017). Dieses Wissen ist eine notwendige Grundlage, um Erkenntnisgewinnungskompetenz in Bezug zu einem spezifischen fachwissenschaftlichen Inhalt zu fördern (Osborne, 2014a). Sollen Schüler:innen zum Beispiel Hypothesen zu Einflussfaktoren der Samenkeimung entwickeln, müssen sie zunächst wissen, welche potenziellen Einflussfaktoren relevant sind (zum Beispiel Licht, Wasser, ...). Prozedurales Wissen

¹⁸ Damit wird der Begriffe „*scientific reasoning competencies*“ weiter als in dem aufgeführten Modell von Mayer (2007) gefasst. In manchen der durchgeführten Studien (zum Beispiel Studie 1 oder Studie 2) wird zusätzlich der Begriff „*scientific reasoning*“ synonym zu der Bezeichnung „Erkenntnisgewinnung“ verwendet. Diese begriffliche Inkonsistenz wird in Kapitel 10.4 kritisch diskutiert.

¹⁹ Synonym zu dem Begriff „*declarative knowledge*“ werden auch Begriffe wie „*content knowledge*“ in Krell, Vorholzer & Nehring (2022) und OECD (2019) oder „*disciplinary knowledge*“ in Osborne (2023) verwendet.

beschreibt hingegen das Verständnis darüber, wie naturwissenschaftliche Verfahren und Konzepte angewandt werden („wissen, wie“; „*knowing how*“; P. Kind & Osborne, 2017). In Bezug auf die Hypothesenbildung zu Einflussfaktoren der Samenkeimung müssen Schüler:innen neben dem fachinhaltlichen Wissen also auch über fachmethodisches Wissen in Bezug auf die Hypothesenbildung verfügen, also ein Verständnis darüber haben, wie Hypothesen formuliert werden. Zuletzt brauchen Schüler:innen auch epistemisches Wissen, welches das Wissen bezüglich der Relevanz der Verfahren und Konzepte der naturwissenschaftlichen Erkenntnisgewinnung beschreibt („wissen, warum“; „*knowing why*“; P. Kind & Osborne, 2017). Dieser Wissensbereich ist nicht unbedingt notwendig für das erfolgreiche Ausführen von naturwissenschaftlichen Verfahren, aber unverzichtbar, wenn sichergestellt werden soll, dass es sich dabei nicht um die Ausführung auswendig gelernter Prozesse handelt, sondern die Schüler:innen über ein Verständnis hinsichtlich der entsprechenden Verfahren verfügen (Berland et al., 2016).

Theoretischer Ausgangspunkt für die Beschreibung von Erkenntnisgewinnungskompetenzen bildet die Wissenschaftstheorie von Karl Popper (1935), welche davon ausgeht, dass Hypothesen nicht belegt, sondern nur durch empirische Forschung gestützt oder durch Falsifikation widerlegt werden können (Falsifikationismus; Godfrey-Smith, 2003; Lawson, 2004; Wellnitz & Mayer, 2016). Naturwissenschaftliche Erkenntnisgewinnung wird auf diese Weise oftmals als hypothetisch-deduktive Vorgehensweise modelliert (Godfrey-Smith, 2003; Wellnitz & Mayer, 2013, 2016).

Der Falsifikationismus als Wissenschaftstheorie bildet einen wichtigen Teilaspekt des von Klahr und Dunbar (1988) beschriebenen Scientific Discovery as Dual Search-Modells (SDDS-Modell), welches als kognitionswissenschaftliches Modell die Grundlage für die Konzeptualisierung von Erkenntnisgewinnungskompetenzen bildet. Das SDDS-Modell erlaubt eine evidenzbasierte Beschreibung der Erkenntnisgewinnung, beginnend mit der Hypothesenbildung, dem hypothesengeleiteten Experimentieren sowie der Analyse von Evidenzen (Engelmann et al., 2016; Hammann, 2007; Klahr, 2000). Dabei ist das Lösen komplexer naturwissenschaftlicher Probleme zurückzuführen auf die Suche im Hypothesenraum sowie die Suche im Experimentraum (Edelsbrunner & Dablander, 2018; Hammann, 2007; Klahr, 2000). Die Hauptkomponenten des Modells umfassen die Suche im Hypothesenraum (Beginn des Erkenntnisprozesses, Bildung einer überprüfbaren Hypothese aus [meist eingeschränktem] domänenspezifischem Wissen), das Testen von Hypothesen (Suche im Experimentraum als Teilaspekt, aus dem die experimentellen Ergebnisse stammen, die für die Beantwortung der Hypothese benötigt werden) und der Analyse von Evidenzen (Entscheidung, ob die vorliegende Hypothese vorläufig akzeptiert, verworfen oder weiter geprüft werden muss; Edelsbrunner & Dablander, 2018; Hammann, 2007; Klahr, 2000).

Die Nutzbarkeit des SDDS-Modells für die Konzeptualisierung von Erkenntnisgewinnungskompetenzen wird jedoch kritisch gesehen, zum Beispiel in Bezug auf Differenzierungen der bestehenden (van Joolingen & T. de Jong, 1997) oder Ergänzung um weitere mögliche Problemräume (Schunn & Klahr, 2000). Mayer (2007) schlägt deshalb eine Ergänzung um die prozedurale Variable „Fragen formulieren“ sowie die Aufnahme von kognitiven Fähigkeiten neben dem deklarativen Wissen vor (Abbildung 2.6). Diese Denkweisen können für jede Arbeitsweise in unterschiedliche Kompetenzniveaus differenziert werden (s. Wellnitz & Mayer, 2016).

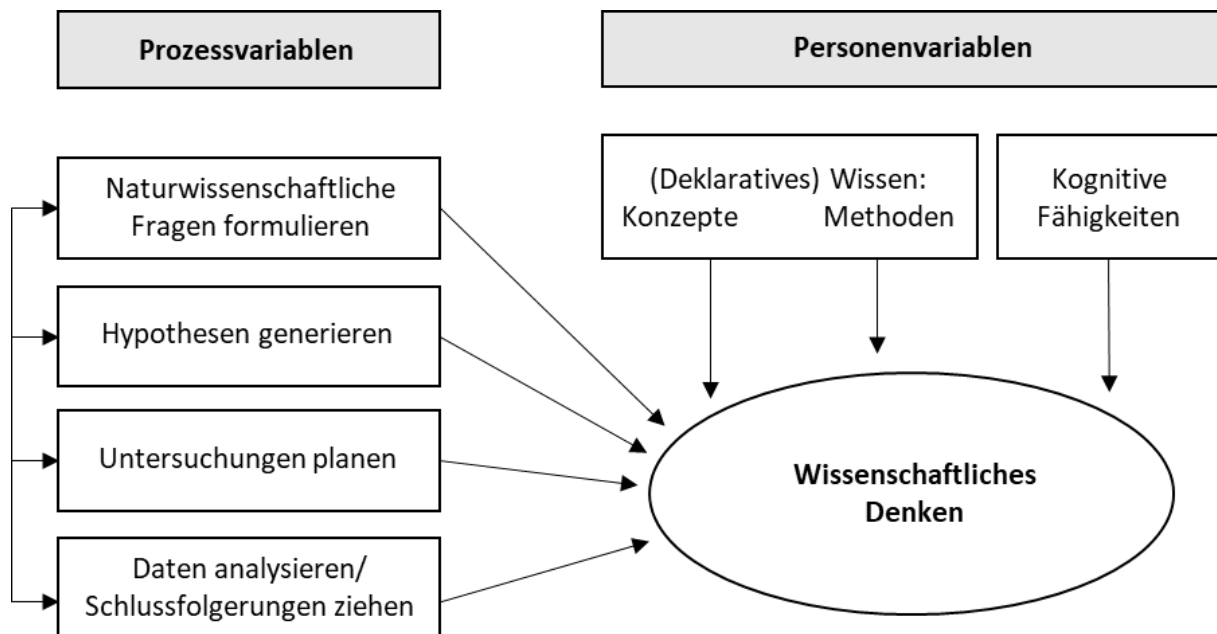


Abbildung 2.6: Strukturmodell zum wissenschaftlichen Denken (Mayer, 2007, S. 181)

Im aktuellen internationalen Diskurs sind Erkenntnisgewinnungskompetenzen jedoch nicht eindeutig konzeptualisiert (Edelsbrunner & Dablander, 2018; Engelman et al., 2016; Krell, Vorholzer & Nehring, 2022; Opitz et al., 2017). Beispielsweise beschreiben P. Kind und Osborne (2017) Erkenntnisgewinnungskompetenzen als die grundlegenden Fähigkeiten, die erforderlich sind, um ein kontextualisiertes wissenschaftliches Problem durch die Verwendung des eigenen Wissens und die Anwendung unterschiedlicher Fähigkeiten zu lösen. Mathesius et al. (2016) beschreiben Erkenntnisgewinnungskompetenzen als diejenigen Kompetenzen, die für das Verständnis der Prozesse des naturwissenschaftlichen Wissenserwerbs notwendig sind. Mayer (2007) beschreibt den naturwissenschaftlichen Erkenntnisprozess als wissenschaftsbasierten Problemlöseprozess, der durch Personen- und Situationsvariablen (zum Beispiel kognitive Fähigkeiten einer Person oder die Art, wie ein Problem präsentiert wird) beeinflusst werden kann.

Erkenntnisgewinnungskompetenzen werden dabei oft als komplexes Konstrukt beschrieben, das eine Vielzahl von Fähigkeiten und damit zusammenhängendem Wissen umfasst (Krell, Vorholzer & Nehring, 2022; Rönnebeck et al., 2016). Aktuelle Studien unterscheiden sich deshalb häufig hinsichtlich vielfältiger Konzeptualisierungen, sowohl in Bezug auf verschiedene Arbeitsweisen als auch hinsichtlich damit verbundener unterschiedlicher Teilprozesse (Krell, Vorholzer & Nehring, 2022). Während die meisten Konzeptualisierungen darin übereinstimmen, dass es verschiedene Teilprozesse gibt, die als separate Dimensionen naturwissenschaftlichen Denkens modelliert werden können (Krell, Vorholzer & Nehring, 2022), unterscheiden sich diese hinsichtlich (1) der Fähigkeiten, die sie umfassen; (2) darin, ob *scientific reasoning* als eindimensionales Konstrukt oder als mehrdimensionales Konstrukt, das unterschiedliche Fähigkeiten umfasst, angesehen wird und (3) ob es als fachspezifisch (*domain-specific*) oder fachübergreifend (*domain-general*) angesehen werden kann (Opitz et al., 2017).

Wellnitz und Mayer (2013, 2016) führen das Experimentieren, Beobachten und Vergleichen als grundlegende Methoden der Erkenntnisgewinnung auf. Andere Forschende betonen zusätzlich auch die zentrale Rolle des Modellierens für die Erkenntnisgewinnung (Krell et al., 2016; Krüger & Upmeyer zu Belzen, 2021; Upmeyer zu Belzen et al., 2021). Rönnebeck et al. (2016) stellen in einem Review verschiedene Denk- und Arbeitsweisen heraus, die in empirischen Studien zu *scientific inquiry* untersucht wurden. Dabei zeigte sich, dass in den untersuchten Studien besonders häufig Denkweisen

wie das Argumentieren von Forschungsergebnissen, das Entwickeln von Erklärungen und das Formulieren von Hypothesen sowie die Arbeitsweisen Experimentieren und Modellieren untersucht wurden (Rönnebeck et al., 2016). Die Konzeptualisierungen sowie Operationalisierung dieser Aktivitäten unterscheiden sich jedoch maßgeblich zwischen den betrachteten Studien (Rönnebeck et al., 2016).

Ungeachtet dieser Unstimmigkeit hinsichtlich der verschiedenen Konzeptualisierung spielen Erkenntnisgewinnungskompetenzen eine zentrale Rolle im naturwissenschaftlichen Unterricht, weshalb sie sowohl in den nationalen Bildungsstandards (z. B. KMK, 2024b) als auch in internationalen Curricula (z. B. ACARA, 2022) verankert sind.

2.2.2 Nationale und internationale curriculare Verankerung im naturwissenschaftlichen Unterricht

Die nationalen Bildungsstandards für den Mittleren Schulabschluss in den naturwissenschaftlichen Fächern (Biologie: KMK, 2024b; Chemie: KMK, 2024c und Physik: KMK, 2024d) sowie für die Allgemeine Hochschulreife in den naturwissenschaftlichen Fächern (Biologie: KMK, 2020a; Chemie: KMK, 2020b und Physik: KMK, 2020c) umfassen neben der Sachkompetenz²⁰ als Inhaltsdimension auch die Erkenntnisgewinnungskompetenz, Kommunikationskompetenz und Bewertungskompetenz als Prozessdimensionen. Für das Fach Biologie umfasst die Sachkompetenz das Kennen biologischer Phänomene, Begriffe, Prinzipien und Fakten sowie die Anwendung auf fach- und alltagsbezogene Sachverhalte, die Erkenntnisgewinnungskompetenz Kenntnisse naturwissenschaftlicher Denk- und Arbeitsweisen, die Kommunikationskompetenz das fach- und sachgerechte Erschließen und Austauschen von Informationen und die Bewertungskompetenz das Erkennen und Bewerten von biologischen Sachverhalten in verschiedenen Kontexten (KMK, 2024b). Dabei beziehen sich die Kompetenzbereiche auf die Basiskonzepte „Struktur und Funktion“, „Stoff und Energieumwandlung“, „Information und Kommunikation“, „Steuerung und Regelung“, „Individuelle Entwicklung“ sowie „Evolutive Entwicklung“ (KMK, 2024b).

Erkenntnisgewinnungskompetenz ist in den Bildungsstandards für das Fach Biologie für den Mittleren Schulabschluss wie folgt definiert:

„Die Erkenntnisgewinnungskompetenz der Lernenden zeigt sich in der Kenntnis von naturwissenschaftlichen Denk- und Arbeitsweisen verbunden mit der Fähigkeit, diese zu beschreiben, zu erklären, für Erkenntnisprozesse zu nutzen und deren Möglichkeiten und Grenzen zu reflektieren.“ (KMK, 2024b, S. 8)

Ähnliche Definitionen befinden sich auch in den Bildungsstandards für die Allgemeine Hochschulreife im Fach Biologie (KMK, 2020a) sowie in den Bildungsstandards anderer naturwissenschaftlicher Fächer (Chemie: KMK, 2020b, 2024c; Physik: KMK, 2020c, 2024d). Damit finden sich sowohl die Unterscheidung zwischen deklarativem, prozeduralem und epistemischem Wissen als auch die Beachtung verschiedener Denk- und Arbeitsweisen als essenzielle Bestandteile der wissenschaftlichen Konzeptualisierung von Erkenntnisgewinnungskompetenzen in den Bildungsstandards wieder.

²⁰ In der ursprünglichen Version der Bildungsstandards (z. B. KMK, 2005) auch „Kompetenzbereich Fachwissen“ genannt.

Für den Primarbereich existieren keine national einheitlichen Bildungsstandards, jedoch umfasst der „Perspektivrahmen Sachunterricht“ der Gesellschaft für Didaktik des Sachunterrichts (GDSU, 2002) Kompetenzen der naturwissenschaftlichen Perspektive²¹, welche die Grundlagen für die Entwicklung der Erkenntnisgewinnungskompetenzen in der weiterführenden Schule bilden. So sollen die Schüler:innen Kompetenzen bezüglich der Bereiche „Fragehaltungen aufbauen, Probleme identifizieren und Verfahren der Problemlösung anwenden“, „Entwickeln und Formulieren von Vermutungen“, „Recherchieren von Informationen“, „Entwerfen, Durchführen und Auswerten von Versuchen“, „Darstellen von Ergebnissen“, „Übertragen von Erkenntnissen aus Versuchen auf weitere Naturphänomene“, „an Beispielen erfahren, dass mit dem experimentellen Verfahren Wissen aufgebaut werden kann und Sachverhalte intersubjektiv feststellbar sind“ und „an Beispielen erfahren, wie man ein Experiment als eine Frage an die Natur ausdenkt, ausführt und auswertet“ entwickeln (GDSU, 2002, S. 15).

Auch international sind Erkenntnisgewinnungskompetenzen oftmals curricular verankert (zum Beispiel Australien: ACARA, 2022; Chile: Mineduc, 2015, 2019; Kanada: Ministry of Education, 2008a, 2008b; Vereinigte Staaten: NGSS Lead States, 2013). Im Gegensatz zu Deutschland, wo unterschiedliche Curricula für die verschiedenen Fächer vorliegen, existiert in Australien ein einheitliches Curriculum (genannt F-10 Curriculum), welches die drei Dimensionen „*learning areas*“ (Lerngebiete), „*general capabilities*“ (allgemeine Fähigkeiten) und „*cross-curriculum priorities*“ (fächerübergreifende Schwerpunkte) unterscheidet (ACARA, 2022). Das naturwissenschaftliche Lerngebiet umfasst neben den Bereichen „*science understanding*“ (Wissenschaftsverständnis) und „*science as a human endeavour*“ (Wissenschaft als menschliches Bestreben) auch den Bereich „*science inquiry skills*“ (wissenschaftliche Fähigkeiten), welcher grundlegende Erkenntnisgewinnungskompetenzen umfasst (ACARA, 2022). Dort werden die fünf Unterbereiche „*questioning and predicting*“ (Hinterfragen und Vorhersagen), „*planning and conducting*“ (Planen und Durchführen), „*processing, modelling and analysing*“ (Verarbeiten, Modellieren und Analysieren), „*evaluating*“ (Auswerten) und „*communicating*“ (Kommunizieren) unterschieden (ACARA, 2022). Damit liegt der Fokus insbesondere auf Denkweisen, Arbeitsweisen werden in den Bereichen nicht explizit genannt.

Um diese Erkenntnisgewinnungskompetenzen bei ihren Schüler:innen zu vermitteln, müssen Lehrkräfte jedoch auch selbst über diese verfügen (Khan & Krell, 2019; Mathesius et al., 2016; Schwartz et al., 2004), weshalb sie ebenfalls ein Bestandteil in der universitären Lehrkräfteausbildung darstellen.

2.2.3 Erkenntnisgewinnungskompetenzen im Studium

Ebenso wie das fachdidaktische Wissen (siehe Kapitel 2.1.5) sind Erkenntnisgewinnungskompetenzen curricular in den „Ländergemeinsamen inhaltlichen Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung“ verankert (KMK, 2024a). Am Ende ihres Studiums sollen angehende Lehrkräfte im Fach Biologie „vertraut mit basalen Arbeits- und Erkenntnismethoden der Biologie [sein] und [...] über Kenntnisse und Fertigkeiten sowohl im hypothesengeleiteten Experimentieren und Modellieren, im kriteriengelernten Beobachten als auch im hypothesengeleiteten Vergleichen sowie im Handhaben von (schulrelevanten) Geräten [verfügen]“ (KMK, 2024a, S. 22). Ähnliche Anforderungen werden auch an in anderen naturwissenschaftlichen Fächern wie Chemie und Physik gestellt (KMK, 2024a).

²¹ Da das Fach Sachunterricht im Primarbereich nicht nur aus naturwissenschaftlichen, sondern auch aus gesellschaftswissenschaftlichen Fachinhalten und -methoden besteht, umfasst der Perspektivrahmen noch weitere Perspektiven, wie etwa die „sozial- und kulturwissenschaftliche Perspektive“ (GDSU, 2002).

Empirische Studien, welche die Entwicklung von Erkenntnisgewinnungskompetenzen im Verlauf des Studiums untersuchten, kommen zu uneindeutigen Ergebnissen. So konnte in manchen Untersuchungen gezeigt werden, dass Lehramtsstudierende mit naturwissenschaftlichen Fächern signifikant höhere Erkenntnisgewinnungskompetenzen aufwiesen, wenn sie weiter in ihrem Studium fortgeschritten waren, allerdings mit mäßigen Effekten (Hartmann et al., 2015; Mathesius et al., 2016). Auch in einer Längsschnittstudie von Krüger et al. (2020) mit 644 angehenden Biologielehrkräften zeigte sich ein signifikanter, aber moderater Zuwachs ihrer Erkenntnisgewinnungskompetenzen über den Verlauf der universitären Ausbildung. Andere Studien konnten dies nicht stützen. So zeigte eine Studie von Krell et al. (2020) mit australischen Studierenden, dass diese zu Beginn ihres Masterstudiums ein grundlegendes Niveau an Erkenntnisgewinnungskompetenzen aufwiesen, welche sich jedoch im ersten Jahr des Masterstudiums nicht signifikant verbesserten.

Ein möglicher Einflussfaktor für die Entwicklung von Erkenntnisgewinnungskompetenzen im Studium könnte die Anzahl der Lerngelegenheiten, in denen explizit entsprechende Kompetenzen adressiert werden, darstellen (Hartmann et al., 2015; Krell, Khan et al., 2022). Dies konnte jedoch nicht in allen Studien gezeigt werden. Eine Studie von Khan und Krell (2019) kam zu dem Ergebnis, dass sich die Erkenntnisgewinnungskompetenzen von angehenden Lehrkräften durch die Teilnahme an einem Kurs, welcher sich inhaltlich mit Themen der Erkenntnisgewinnung befasste, nicht signifikant verbesserte. Es zeigte sich jedoch, dass Teilnehmende mit zwei Studienabschlüssen (zum Beispiel Bachelor und Master of Education) nach dem Kurs signifikant höhere Erkenntnisgewinnungskompetenzen aufwiesen, was für Teilnehmende mit weniger Studienabschlüssen nicht galt (Khan & Krell, 2019). Dies gibt Hinweise darauf, dass angehende Lehrkräfte mit mehr Vorwissen stärker von solchen spezifischen Lerngelegenheiten profitieren (Khan & Krell, 2019). Diese Annahme wird gestützt durch die Ergebnisse anderer Studien, die zeigen konnten, dass angehende Lehrkräfte mit zwei naturwissenschaftlichen Fächern höhere Erkenntnisgewinnungskompetenzen aufwiesen als Studierende, die nur ein naturwissenschaftliches Fach studieren (Hartmann et al., 2015)²².

Für die Entwicklung von Erkenntnisgewinnungskompetenzen angehender Lehrkräfte in der universitären Ausbildung bedarf es folglich geeigneter Lerngelegenheiten. (Angehende) Lehrkräfte sollten jedoch nicht nur selbst über entsprechende Kompetenzen verfügen, sondern auch wissen, wie sie diese optimal im naturwissenschaftlichen Unterricht vermitteln können.

2.2.4 Erkenntnisgewinnungskompetenzen im Unterricht

Erkenntnisgewinnungskompetenzen können auf verschiedene Weisen im naturwissenschaftlichen Unterricht gefördert werden. So können sie implizit (Erkenntnisgewinnungskompetenzen werden nicht direkt durch die Lehrkraft adressiert und müssen von den Schüler:innen selbstständig entwickelt werden) oder explizit (Erkenntnisgewinnungskompetenzen werden in Instruktionen der Lehrkraft adressiert) vermittelt werden (Khishfe & Abd-El-Khalick, 2002; Vorholzer et al., 2020). In empirischen Studien konnte vielfach gezeigt werden, dass explizite Instruktionsstrategien zur Förderung von Erkenntnisgewinnungskompetenzen wirksamer im Vergleich zu impliziten Strategien sind (Chen & She, 2015; Khishfe & Abd-El-Khalick, 2002; S. W.-Y. Lee, 2024; Matlen & Klahr, 2013; Vorholzer et al., 2020). Um dabei fachmethodisches Wissen in einer Unterrichtsstunde explizit erarbeiten zu können, müssen Schüler:innen über dieses Wissen dekontextualisiert von einer spezifischen Situation verfügen, um es im Anschluss in Bezug auf eine unbekannte Problemstellung zu rekontextualisieren (Krell & Krüger, 2022). Lernen Schüler:innen zum Beispiel die Variablenkontrollstrategie am Beispiel

²² Ähnliche Synergieeffekte ließen sich ebenfalls für das fachdidaktische Wissen angehender Biologielehrkräfte zeigen: Studierende, die Biologie und Chemie auf Lehramt studierten, wiesen höheres PCK im Vergleich zu Studierenden, die eines der Fächer kombiniert mit einem anderen studierten, auf (Welter et al., 2022).

von Experimenten zu Einflussfaktoren auf die Photosyntheserate, sollten sie in der Lage sein, diese Strategie auch in Bezug auf andere Experimente anzuwenden. Lehrkräfte sollten für die Planung einer Unterrichtsstunde zur Erkenntnisgewinnung also dazu in der Lage sein, das entsprechende fachmethodische Wissen zu identifizieren, welches sie bei ihren Schüler:innen fördern wollen (Krell & Krüger, 2022).

Darüber hinaus ist bei der erfolgreichen Vermittlung von Erkenntnisgewinnungskompetenzen auch der Öffnungsgrad entscheidend, also der Umfang der Entscheidungen, die von den Schüler:innen selbstständig getroffen werden können, was der Unterstützung entspricht, die diese erhalten (Abrams et al., 2008; Baur et al., 2022; Priemer, 2011). Während ein niedriger Öffnungsgrad mit der Anleitung durch die Lehrkraft verbunden ist, haben Schüler:innen bei einem hohen Öffnungsgrad die Möglichkeit, selbstständig naturwissenschaftliche Denk- und Arbeitsweisen auszuführen. Dabei kann und sollte der Öffnungsgrad auch schrittweise erhöht werden (Baur et al., 2022; Vorholzer & Aufschnaiter, 2019).

In den vergangenen Jahren wurde diskutiert, ob sich explizite Instruktionsstrategien in Form von direkter Instruktion durch die Lehrkraft (verbunden mit einem niedrigen Öffnungsgrad) oder in Form von schülerzentrierten Instruktionen (zum Beispiel forschendem Lernen [*inquiry-based instruction*]; verbunden mit einem hohen Öffnungsgrad) besser eignen, um Erkenntnisgewinnungskompetenzen zu fördern (siehe zum Beispiel Diskussion in T. de Jong et al., 2023; Sweller et al., 2024; Zhang et al., 2022). Empirische Befunde legen nahe, dass schülerzentrierte Instruktionen, solange sie von der Lehrkraft entsprechend der Kompetenzen der Schüler:innen angeleitet werden, zu einem besseren Lernerfolg führen (Aditomo & Klieme, 2020; Lazonder & Harmsen, 2016). Ein hoher Öffnungsgrad kann bei niedrigem Vorwissen unter Umständen gleichzeitig zu einem geringeren Lernerfolg führen (Lazonder & Harmsen, 2016), insbesondere wenn die Schüler:innen mit dem selbstständigen Ausführen naturwissenschaftlicher Denk- und Arbeitsweisen überfordert sind (P. A. Kirschner et al., 2006). Deswegen wird empfohlen, den Öffnungsgrad sukzessive mit steigendem fachmethodischen Wissen der Schüler:innen zu erhöhen, um den Erwerb von Erkenntnisgewinnungskompetenzen zu ermöglichen (Vorholzer et al., 2022). Empirische Studien zeigen jedoch, dass Lehrkräfte naturwissenschaftliche Denk- und Arbeitsweisen häufig in Verbindung mit Anleitungen ausführen und den Öffnungsgrad nicht erhöhen (z. B. Börling & Labudde, 2014).

Zusätzlich konnte in empirischen Studien gezeigt werden, dass Lehrkräfte Schwierigkeiten haben, Stunden zur Förderung der Erkenntnisgewinnungskompetenz zu planen (Großmann & Krüger, 2024; Petermann & Vorholzer, 2023; Vo et al., 2015) oder diese Kompetenzen explizit zu adressieren (Vorholzer et al., 2022) und deshalb primär die Sachkompetenz ihrer Schüler:innen fördern (Akuma & Callaghan, 2018). Darüber hinaus zeigt sich, dass Lehrkräfte selbst oftmals nicht über ausreichend Erkenntnisgewinnungskompetenzen verfügen, um diese adäquat zu fördern (Capps & Crawford, 2013; Capps et al., 2016) oder sich unsicher in der Vermittlung dieser Kompetenzen im Vergleich zur Sachkompetenz fühlen (Petermann et al., 2024). Da Erkenntnisgewinnungskompetenzen von Schüler:innen während der Schullaufbahn häufig unzureichend entwickelt werden (J. Lederman, Bartels et al., 2023; J. Lederman, N. G. Lederman, Bartels, Jimenez, Acosta et al., 2021; J. Lederman, N. G. Lederman, Bartels, Jimenez, Akubo et al., 2019) unterstreicht dies die Notwendigkeit, effektive Unterrichtstrategien zur Förderung von Erkenntnisgewinnungskompetenzen in Aus- und Weiterbildungsangeboten für angehende und praktizierende Lehrkräfte zu adressieren (J. Lederman et al., 2024). Im Folgenden soll deshalb exemplarisch für die Denk- und Arbeitsweisen

Experimentieren, Hypothesen formulieren und Modellieren eine geeignete Einbindung in den Unterricht betrachtet werden²³.

2.2.4.1 Experimentieren im naturwissenschaftlichen Unterricht

Experimente sollten im naturwissenschaftlichen Unterricht nicht ausschließlich dazu genutzt werden, Fachwissen durch praktisches Arbeiten zu veranschaulichen, sondern ebenfalls dazu verwendet werden, Kompetenzen im Bereich Erkenntnisgewinnung in Form fachmethodischen Wissens zu vermitteln (Krell & Krüger, 2022; Sannert, Petermann et al., 2024; Sannert, Voitle et al., 2024). Mit Experimenten sollte deshalb der kausale Zusammenhang von zwei oder mehr Variablen in Form einer Ursache-Wirkung-Beziehung untersucht werden, wobei die relevanten Einflussvariablen systematisch kontrolliert werden (Krell & Krüger, 2022).

Experimentelle Kompetenz umfasst dabei unterschiedliche Teilprozesse in Bezug auf verschiedene Denkweisen wie die Bildung einer Forschungsfrage und Hypothesen, der Versuchsplanung und -durchführung sowie der Datenauswertung und -interpretation (Baur et al., 2019; Gut-Glanzmann & Mayer, 2018). Zur Förderung von Kompetenzen in Bezug auf das Experimentieren sollte deshalb das Ziel verfolgt werden, dass die Schüler:innen diese Teilprozesse sowie die Eigenschaften eines Experiments kennen (Baur et al., 2019). Darüber hinaus sollten sie lernen, die Grenzen der Aussagekraft eines Experiments einschätzen und Experimente selbstständig durchführen zu können (Baur et al., 2019). Dadurch können Schüler:innen neben deklarativem Wissen auch prozedurales und epistemisches Wissen erlangen.

2.2.4.2 Hypothesen formulieren im naturwissenschaftlichen Unterricht

Das Formulieren von Hypothesen als zentrale Denkweise ist in Bezug zu verschiedenen Arbeitsweisen relevant (Krell & Krüger, 2022; Upmeyer zu Belzen & Krüger, 2019; Wellnitz & Mayer, 2016). Es beschreibt die Formulierung von auf Vorwissen basierenden Annahmen über die Operationalisierung von abhängigen und unabhängigen Variablen zur empirischen Untersuchung des Zusammenhangs / der Unterschiede dieser Variablen (Krell & Krüger, 2022). Hypothesen unterscheiden sich für die unterschiedlichen Arbeitsweisen (Upmeyer zu Belzen & Krüger, 2019): Während beim Experimentieren häufig Kausalhypothesen überprüft werden, können beim Modellieren zusätzlich auch Zusammenhangs- oder Unterschiedshypothesen überprüft werden.

Im naturwissenschaftlichen Unterricht wird der Begriff Hypothese jedoch häufig auch alltagssprachlich als Erklärungsversuch eines Phänomens, zum Beispiel im Rahmen des Unterrichtseinstiegs, verwendet (Großmann & Krüger, 2024). Viel eher sollte jedoch die Funktion von Hypothesen als auf Vorwissen basierende, empirisch überprüfbare sowie potenziell falsifizierbare Voraussage herausgestellt werden, wodurch ihnen eine zentrale Rolle im naturwissenschaftlichen Erkenntnisgewinn zukommt (Großmann & Krüger, 2024; Krell & Krüger, 2022). Hinsichtlich der Umsetzung im Unterricht zeigt sich, dass (angehende) Lehrkräfte häufig Schwierigkeiten mit der korrekten Einbindung von Hypothesen haben und sich viel eher darauf fokussieren, fachinhaltliches Wissen unter Verwendung (vermeintlicher) Hypothesen statt wie geplant Erkenntnisgewinnungskompetenzen zu fördern (Großmann & Krüger, 2024).

²³ Diese Denk- und Arbeitsweisen wurden an dieser Stelle exemplarisch ausgewählt, da sie den Dimensionen des im Rahmen dieser Dissertation entwickelten Instruments entsprechen. Eine kritische Diskussion über die Auswahl genau dieser Denk- und Arbeitsweisen erfolgt in Kapitel 10.4.

2.2.4.3 Modellieren im naturwissenschaftlichen Unterricht

Modelle werden im naturwissenschaftlichen Unterricht häufig zur Visualisierung von Objekten oder Konzepten verwendet (Godfrey-Smith, 2007; Krell et al., 2016; Nielsen & Nielsen, 2021; Upmeier zu Belzen et al., 2019). In dieser Funktion können sie als Repräsentationen zur Vermittlung von fachinhaltlichem Wissen genutzt werden (Upmeier zu Belzen et al., 2019). Darüber hinaus sollten Modelle jedoch auch miteinander verglichen und als Werkzeug zur Erkenntnisgewinnung genutzt werden, indem mit ihnen Phänomene vorhergesagt werden (Chiu & Lin, 2019; Göhner & Krell, 2022; Knuuttila, 2011; Krell & Krüger, 2022; Krüger & Upmeier zu Belzen, 2021).

Modellierkompetenz²⁴ umfasst neben dem Wissen über Modelle auch das Wissen über den Modellierprozess (beinhaltet die Modellierpraktiken und das Modellierprodukt; Nielsen & Nielsen, 2021; Upmeier zu Belzen et al., 2021) und kann in unterschiedlichen Niveaustufen ausgeprägt sein (s. Krüger & Upmeier zu Belzen, 2021; Upmeier zu Belzen & Krüger, 2010). Zur Förderung der Modellierkompetenz wird empfohlen, einen Fachinhalt zu wählen, der für die Schüler:innen verständlich ist und, insbesondere, wenn diese über noch kaum ausgeprägte Kompetenzen verfügen, leicht zugängliche Modelle zu wählen (Krell et al., 2016). Weiterhin sollten in einer einzelnen Unterrichtsstunde nur ausgewählte Teilkompetenzen gefördert und das Erlernete regelmäßig mit den Schüler:innen reflektiert und in verschiedenen Kontexten umgesetzt werden (Dekontextualisierung; Krell et al., 2016).

²⁴ In der deutschsprachigen Forschung wird der Begriff „Modellierkompetenz“ in Abgrenzung zu dem Begriff „Modellkompetenz“ verwendet und umfasst eine zusätzliche Niveaustufe hinsichtlich des Wissens über Modelle basierend auf dem Konzept des abduktiven Schließens (Krüger und Upmeier zu Belzen, 2021).

2.3 Fachdidaktisches Wissen im Kompetenzbereich Erkenntnisgewinnung

Um Erkenntnisgewinnungskompetenzen angemessen bei ihren Schüler:innen zu fördern, müssen Lehrkräfte folglich nicht nur selbst über Kompetenzen in diesem Bereich verfügen, sondern auch wissen, wie sie ihren Schüler:innen diese optimal vermitteln können. Im Rahmen dieser Dissertation wird deshalb von fachdidaktischem Wissen im Kompetenzbereich Erkenntnisgewinnung (*PCK about scientific reasoning; PCK_{SR}*) gesprochen, welches basierend auf der Definition von Gess-Newsome (2015) wie folgt definiert wird:

PCK_{SR} describes the knowledge of, reasoning behind, and planning required to teach scientific reasoning in different educational settings to specific students to improve their scientific reasoning competencies.

PCK_{SR} umfasst damit insbesondere Wissen über gängige Vorstellungen der Schüler:innen sowie ihr Vorwissen zu naturwissenschaftlichen Denk- und Arbeitsweisen und geeignete Instruktionsstrategien, um Erkenntnisgewinnungskompetenzen erfolgreich zu vermitteln (zum Beispiel wie epistemisches und prozedurales Wissen mit Experimenten gefördert werden kann oder wie Modelle zusätzlich zu Mitteln der Visualisierung auch als epistemische Werkzeuge benutzt werden können).

Es handelt sich demnach um disziplinspezifisches PCK (Carlson et al., 2019; Magnusson et al., 1999; siehe auch Kapitel 2.1.4), welches sich nicht auf ein fachinhaltliches Thema, sondern auf davon unabhängiges fachmethodisches Wissen bezieht. Damit ist PCK_{SR} nicht nur relevant für das Unterrichten im Fach Biologie, sondern auch in den anderen naturwissenschaftlichen Fächern Chemie und Physik. Trotz der zentralen Rolle von Erkenntnisgewinnungskompetenzen im naturwissenschaftlichen Unterricht und der damit verbundenen Notwendigkeit, dass Lehrkräfte über PCK_{SR} verfügen müssen, ist dieser Wissensbereich bisher nur wenig erforscht. Aktuelle empirische Studien, welche das PCK_{SR} (angehender) Lehrkräfte untersuchen, unterscheiden sich hauptsächlich bezüglich der folgenden vier Punkte:

- (1) Die betrachtete Stichprobe: Studien werden mit Studierenden (z. B. in Hilfert-Rüppell et al., 2021), Lehrkräften an Grundschulen (z. B. in van Uum et al., 2021) oder Lehrkräften an weiterführenden Schulen (z. B. in Lin & Wang, 2024; Sannert, van Driel & Krell, 2024) durchgeführt.
- (2) Die erfassten Denk- und Arbeitsweisen: Häufig erfassen Studien unterschiedliche Dimensionen der Erkenntnisgewinnung (z. B. Modellieren: Günther et al., 2019; Justi & van Driel, 2006) oder Konzepte, die damit im Zusammenhang stehen (z. B. die Durchführung wissenschaftlicher Untersuchungen: van Uum et al., 2021; wissenschaftliches Denken: Lin & Wang, 2024).
- (3) Die erfassten Bereiche des PCK: Studien unterscheiden sich hinsichtlich dessen, ob sie das dynamische ePCK (z. B. in O. de Jong & van der Valk, 2007) oder das statische pPCK (z. B. in Lin & Wang, 2024) von Lehrkräften erfassen.
- (4) Die verwendeten Auswertungsmethoden und -instrumente: Insbesondere in Abhängigkeit der erfassten Bereiche des PCK werden in empirischen Studien unterschiedliche Erhebungsmethoden verwendet, so zum Beispiel Beobachtungen (z. B. in O. de Jong & van der Valk, 2007), offene Paper-Pencil Tests (z. B. in Sannert, van Driel & Krell, 2024) oder gemischte Antwortformate (z. B. in Lin & Wang, 2024).

Diese Unterschiede in der Erhebung von PCK_{SR} erschweren einen Vergleich der Ergebnisse sowie das Ziehen übergeordneter Schlussfolgerungen. Ein grundlegendes Wissen über die Ausprägung sowie die Entwicklung von PCK_{SR} bei angehenden und praktizierenden Lehrkräften ist jedoch notwendig, um entsprechende Aus- und Fortbildungsangebote effektiv zu gestalten. Dies beinhaltet auch Erkenntnisse über den Zusammenhang von PCK_{SR} mit konzeptionell verwandten Variablen.

Da es sich bei PCK_{SR} um disziplinspezifisches PCK handelt, lässt sich vermuten, dass es ebenfalls mit CK (d. h. Erkenntnisgewinnungskompetenzen) zusammenhängt (z. B. Sorge et al., 2019; s. Kapitel 2.1.4). Darüber hinaus konnte für PCK bereits ein Zusammenhang mit epistemologischen Überzeugungen (Großschedl et al., 2019; Kunter et al., 2013) sowie ein Einfluss dieser auf die Entwicklung von PCK (Wu et al., 2021) gezeigt werden. Gleichzeitig konnten Suh und Park (2017) in einer empirischen Studie zeigen, dass Lehrkräfte eher einen forschungsbasierten Unterricht (schülerzentrierter und damit konstruktivistischer Unterrichtsansatz; Minner et al., 2010) durchführten, wenn dieser mit ihren Überzeugungen übereinstimmte. Darüber hinaus zeigt sich, dass Lehrkräfte, die offen für Innovation sind, eher schülerzentriert und forschungsbasiert unterrichten (Kelly & Staver, 2005; Suh & Park, 2017). Da Studien zeigen, dass insbesondere konstruktivistischer forschungsbasierter Unterricht in Kombination mit der Reflexion des Forschungsprozesses effektiv für die Entwicklung von Erkenntnisgewinnungskompetenzen ist (Aditomo & Klieme, 2020; Engelmann et al., 2016; Lazonder & Harmsen, 2016), ist zu erwarten, dass ein positiver Zusammenhang zwischen konstruktivistischen Überzeugungen und PCK_{SR} (bzw. ein negativer Zusammenhang zwischen transmissiven Überzeugungen und PCK_{SR}) sowie ein positiver Zusammenhang zwischen der Offenheit zur Innovation und PCK_{SR} besteht.

Um diese Zusammenhänge zu überprüfen, bedarf es eines Erhebungsinstruments zur Erfassung von PCK_{SR} , welches eine valide Testwertinterpretation erlaubt.

2.4 Validierung von Testinstrumenten

Damit ein Testinstrument dazu verwendet werden kann, Aussagen über das zu erfassende Konstrukt zu liefern, muss sichergestellt werden, dass die Schlussfolgerungen aus der Interpretation des Testwerts die Realität des zugrunde liegenden Konstrukts korrekt abbilden (AERA et al., 2014; Cook & Hatala, 2016; Döring, 2023; Kane, 2013a, 2013b; Messick, 1995; Shadish et al., 2002). In der aktuellen Auflage der „*Standards for Educational and Psychological Testing*“ wird dieses Gütekriterium, auch als Validität bezeichnet, wie folgt definiert:

“Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.” (AERA et al., 2014, S. 11)

Validierung beschreibt einen laufenden Prozess, bei dem basierend auf unterschiedlichen Evidenzquellen untersucht wird, ob die Interpretation des Testwerts eines Instruments gestützt werden kann (Cook & Hatala, 2016; Kane, 2013b) und stellt damit die Grundlage der Testentwicklung dar (AERA et al., 2014). Ein Testinstrument kann demnach nicht selbst validiert werden, Validität beschreibt nicht die Eigenschaft eines Tests, sondern viel eher die Gültigkeit der Testwertinterpretation (AERA et al., 2014; Kane, 2013b; Messick, 1995)²⁵.

Nach traditionellem Verständnis kann zwischen Inhaltsvalidität (*content validity*; Repräsentation des zugrunde liegenden Konstrukts durch das entwickelte Instrument), Kriteriumsvalidität (*criterion validity*; Grad, mit dem der Testscore eines entwickelten Instruments mit einem externen Kriterium korreliert) und Konstruktvalidität (*construct validity*; Fähigkeit eines entwickelten Instruments, das zugrunde liegende Konstrukt zu erfassen) unterschieden werden (Cronbach & Meehl, 1955; Döring, 2023; Fischer et al., 2023; Messick, 1989). Diese Unterscheidung geht auf Cronbach und Meehl (1955) zurück, welche die zum damaligen Zeitpunkt bereits bekannte Inhaltsvalidität und Kriteriumsvalidität weiter ausdifferenzierten und um die Konstruktvalidität ergänzten. Die Betrachtung getrennter Validitätsarten wird jedoch kritisch gesehen (z. B. AERA et al., 2014), weshalb Messick (1989) sie als ein einheitliches Konzept rahmte.

Einen strukturierten Ansatz zur Validierung der Testwertinterpretation eines Instruments, welcher die schrittweise Entwicklung und Überprüfung von Argumenten umfasst, beschreibt Kane (2013a, 2013b) als „*argument based approach to validation*“: Dabei geht es darum, zunächst ein „Interpretations-Verwendungsargument“ (*interpretation/use argument*) zu formulieren, welches eine Aussage darüber trifft, wie der Testwert zu interpretieren ist. Im Anschluss zielt ein „Validitätsargument“ (*validity argument*) darauf ab, dieses Interpretations-Verwendungsargument zu stützen oder zu widerlegen, und damit eine Aussage bezüglich der Plausibilität dieses Arguments zu treffen (Kane, 2013a, 2013b). Ergänzend dazu empfiehlt Gerasimova (2024) zur Erhöhung der Transparenz des Validierungsprozesses, eine fortlaufend aktualisierte Übersicht der Evidenzen anzulegen sowie die formulierten Behauptungen vor Beginn der Überprüfung zu teilen.

Die AERA et al. (2014) unterscheiden in der aktuellen Ausgabe der „*Standards for Educational and Psychological Testing*“ fünf Validitätsquellen (*sources of validity evidence*), auf Basis derer Argumente für die Validität der Testwertinterpretation eines Instruments gesammelt werden können. Diese sind in Tabelle 2.3 dargestellt und werden im Folgenden jeweils kurz vorgestellt.

²⁵ Wenn im Rahmen dieser Dissertation von einer Validierung von Testinstrumenten gesprochen wird, ist damit folglich die Überprüfung der Validität der Testwertinterpretation des entsprechenden Instruments gemeint.

Tabelle 2.3: Übersicht über die fünf Validitätsquellen nach AERA et al. (2014)

Validitätsquellen	Beschreibung
Evidenz auf Grundlage des Testinhalts	Inhaltliche Passung des Testinhalts zu dem zu testenden Merkmal
Evidenz auf Grundlage von Antwortprozessen	Kognitive Prozesse während der Testbearbeitung
Evidenz auf Grundlage der internen Struktur	Passung des Antwortverhaltens zu der Konstruktion des Fragebogens
Evidenz auf Grundlage der Beziehung zu anderen Variablen	Beziehung des erfassten Merkmals zu konzeptionell zusammenhängenden Variablen
Evidenz auf Grundlage der Folgen der Testung	Identifikation von beabsichtigten und unbeabsichtigten Konsequenzen des Testgebrauchs

2.4.1 Validitätsevidenz auf Grundlage des Testinhalts

Validitätsevidenz auf Grundlage des Testinhalts (*evidence based on test content*) beschreibt die Passung des entwickelten Instruments zu dem zugrunde liegenden Konstrukt (AERA et al., 2014) und entspricht damit der Inhaltsvalidität nach dem klassischem Validitätskonzept (McCoach et al., 2013c; Sireci & Parker, 2006). Da dies den Ausgangspunkt des Validierungsprozesses darstellt, sollte ihm die höchste Priorität im Entwicklungsprozess zukommen (McCoach et al., 2013c; Sireci & Faulkner-Bond, 2014).

In einem Review stellte Sireci (1998) vier Bestandteile der Validitätsevidenz auf Grundlage des Testinhalts heraus: Domänendefinition (*domain definition*), Domänenrelevanz (*domain relevance*), Domänenrepräsentation (*domain representation*) und der Eignung des Testkonstruktionsverfahrens (*appropriate test construction procedures*). Die Domänendefinition beschreibt die Operationalisierung des zugrunde liegenden Konstrukts und erfolgt meist durch eine detaillierte Beschreibung sowie Festlegung der Dimensionen des Instruments, häufig in Zusammenarbeit mit einer Gruppe von Expert:innen (Sireci & Faulkner-Bond, 2014). Domänenrelevanz beschreibt, wie relevant einzelne Items des Instruments für die Erfassung des zugrunde liegenden Konstrukts sind, Domänenrepräsentation hingegen, wie umfassend dieses von allen Items des Instruments erfasst wird (Sireci & Faulkner-Bond, 2014). Zuletzt beschreibt die Eignung des Testkonstruktionsverfahrens diejenigen Prozesse, auf dessen Basis sichergestellt wird, dass das entwickelte Instrument das zugrunde liegende Konstrukt misst und es nicht zu einer Über- oder Unterrepräsentation kommt (Sireci & Faulkner-Bond, 2014; siehe auch Kapitel 2.1.6.1 in Bezug auf die Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK).

Bezüglich des zugrunde liegenden Konstrukts kann zwischen der konzeptuellen (*conceptual definition*; Beschreibung des zugrunde liegenden Konstrukts) und operationalen Definition (*operational definition*; Operationalisierung des zugrunde liegenden Konstrukts in messbare Items) unterschieden werden (McCoach et al., 2013c). Um Validitätsevidenz auf Grundlage des Testinhalts zu gewährleisten, müssen diese Definitionen zusammenpassen, wodurch die Items des Tests eine valide Interpretation des zugrunde liegenden Konstrukts ermöglichen (McCoach et al., 2013c). Eine Nichtpassung kann zu Unter- oder Überrepräsentation des Konstrukts führen (McCoach et al., 2013c).

Um die Validitätsevidenz auf Grundlage des Testinhalts zu überprüfen, werden häufig Expert:innen befragt, welche die Passung der entwickelten Items zu dem zugrunde liegenden Konstrukt durch Rating oder Matching einschätzen sollen (Sireci & Faulkner-Bond, 2014). Damit hängt die Qualität dieser Validierungsansätze jedoch entscheidend von der Auswahl dieser Expert:innen ab (Sireci & Faulkner-Bond, 2014).

2.4.2 Validitätsevidenz auf Grundlage von Antwortprozessen

Validitätsevidenz auf Grundlage von Antwortprozessen (*evidence based on response processes*) beschreibt die kognitiven Prozesse der Teilnehmenden während der Testbearbeitung (AERA et al., 2014). Wird ein Testscore berechnet (insbesondere bei Multiple-Choice-Instrumenten), gibt dieser zwar Aufschluss über das zugrunde liegende Konstrukt, es kann jedoch keine Aussage über die Gedankenprozesse und die Begründung zur Auswahl einer Antwortoption getätigt werden (Kane & Mislevy, 2017; Padilla & Benítez, 2014). Eine Analyse der Antwortprozesse hilft deshalb dabei, ein Verständnis darüber zu erlangen, welche zugrunde liegenden kognitiven, affektiven und motivationalen Mechanismen einen Einfluss auf die Beantwortung durch die Teilnehmenden haben (Hubley & Zumbo, 2017). Da dies eng mit der Validitätsevidenz auf Grundlage des Testinhalts verknüpft ist, werden entsprechende Studien bereits häufig während der Instrumentenentwicklung durchgeführt (Padilla & Benítez, 2014). Die Überprüfung dieser Validitätsevidenz sollte insbesondere dann integraler Bestandteil der Testentwicklung sein, wenn komplexe Konstrukte erfasst (Lane, 2017) oder Testinstrumente zu bisher wenig untersuchten Konstrukten (Newton, 2018) entwickelt werden sollen. Die Analyse von Antwortprozessen kann dabei nicht nur zur Ermittlung möglicher Ursachen für konstruktirrelevante Varianz verwendet werden, sondern auch, um Modelle zu den beabsichtigten Antwortprozessen zu entwickeln und zu prüfen (Wise, 2017).

Grundlegend werden zwei Ansätze zur Analyse von Antwortprozessen unterschieden (Kane & Mislevy, 2017): Während Prozessmodellinterpretationen (*process-model interpretations*) die Analyse kognitiver und mentaler Beantwortungsprozesse auf Basis eines theoretischen Modells zur Überprüfung der kognitiven Prozesse beschreiben, umfassen Merkmalsinterpretationen (*trait interpretations*) die Analyse der zugrunde liegenden Fähigkeiten (zum Beispiel das benötigte Wissen), die für die Beantwortung der Items notwendig sind. Beide Ansätze können zur Überprüfung der Validitätsevidenz auf Grundlage von Antwortprozessen verwendet werden (Ercikan & Pellegrino, 2017; Hubley & Zumbo, 2017).

Um Validitätsevidenz auf Grundlage von Antwortprozessen zu überprüfen, können zum Beispiel die Interpretation der Items durch die Teilnehmenden, ihre Antwortstrategien, das von ihnen angewendete Wissen oder die Art und Weise, wie sie die zur Verfügung stehenden Informationen eines Items nutzen, analysiert werden (Ercikan & Pellegrino, 2017). Die Daten können dabei auf unterschiedliche Weise erfasst werden, darunter Interviews (zum Beispiel in Zusammenhang mit lautem Denken, z. B. Krell et al., 2012), aber auch Methoden wie Eye-Tracking (z. B. Mathesius et al., 2018) oder die Analyse der Beantwortungszeit (Padilla & Benítez, 2014).

Zur Untersuchung von Antwortprozessen bei der Bearbeitung von Testinstrumenten kann sich darüber hinaus auf das Experten-Novizen-Paradigma (de Groot, 1965) berufen werden. Dabei bezeichnet der Begriff Expert:in eine Person, welche wiederholt hohe Leistungen in repräsentativen Aufgaben eines Fachbereichs erfüllt (Ericsson, 2006). Vor dem Hintergrund der Testbearbeitung ist von Expert:innen damit eine höhere Leistung (in Leistungstests also ein höherer Testscore) zu erwarten als von Noviz:innen, da sie über das erforderliche Wissen verfügen, um die Items eines Testinstruments erfolgreich bearbeiten zu können (Ericsson, 2006). Darüber hinaus ist davon auszugehen, dass Expert:innen die Items eines Tests konsistenter beantworten, da sie ihr Wissen auf unterschiedliche

Kontexte anwenden können (Hogan et al., 2003; Krepf et al., 2018; Wolff et al., 2015). Noviz:innen weisen hingegen eine höhere Varianz in ihrem Antwortverhalten auf. Auch lassen sich Noviz:innen bei der Beantwortung von Testitems häufig von Oberflächenmerkmalen (Merkmale, die nicht in Zusammenhang mit dem zu erfassenden Konstrukt stehen) beeinflussen, während Expert:innen über das zur Beantwortung notwendige Wissen verfügen, um die Tiefenmerkmale (Merkmale, die die zugrunde liegenden kognitiven Prozesse beschreiben) der Items zu verstehen (Hogan et al., 2003; Krepf et al., 2018; Wolff et al., 2015). Diese Oberflächen- und Tiefenmerkmale sind deshalb auch in Bezug auf psychometrischen Eigenschaften des Testinstruments relevant (z. B. Krell, 2018; Nehm & Ridgway, 2011; Opfer et al., 2012). Diese Analyse von Unterschieden im Antwortverhalten basierend auf dem Experten-Novizen-Paradigma kann darüber hinaus auch dazu verwendet werden, Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen zu überprüfen (s. Kapitel 2.4.4).

2.4.3 Validitätsevidenz auf Grundlage der internen Struktur

Validitätsevidenz basierend auf der internen Struktur (*validity evidence based on internal structure*) beschreibt die Passung des Antwortverhaltens einer befragten Person zu der Konstruktion des Instruments (AERA et al., 2014). Zentral sind dabei insbesondere die Dimensionalität, die Messinvarianz sowie die interne Konsistenz des Testinstruments (Rios & Wells, 2014).

Bezüglich der Dimensionalität eines Testinstruments ist relevant, inwiefern sich die Dimensionen des latenten Konstrukts auch im Antwortverhalten, also in der Korrelation der Items, zeigen (McCoach et al., 2013d; Rios & Wells, 2014). Ein gängiges Vorgehen zur Überprüfung der Dimensionalität stellt die Faktorenanalyse (konfirmatorisch oder explorativ) dar, bei der die Anzahl der Dimensionen eines Instruments anhand des Antwortverhaltens der Proband:innen bestimmt wird (McCoach et al., 2013d; Rios & Wells, 2014). Alternativ kann jedoch auch latente Klassenanalyse, Item-Response-Theorie oder Rasch-Modellierung verwendet werden (McCoach et al., 2013a).

Die Messinvarianz stellt die Eigenschaft eines Testinstruments dar, das zugrunde liegende Konstrukt in verschiedenen Gruppen unter verschiedenen Bedingungen stabil zu messen (Rios & Wells, 2014). Dies ist insbesondere eine notwendige Voraussetzung für die Untersuchung von Gruppenunterschieden, da das Testinstrument hierbei für unterschiedliche Gruppen vergleichbare Ergebnisse (zum Beispiel in Bezug auf die Itemschwierigkeit) liefern muss (Rios & Wells, 2014).

Die interne Konsistenz beschreibt eine Möglichkeit zur stabilen Schätzung der Reliabilität (Döring, 2023). Eine niedrige interne Konsistenz kann dabei Hinweise darauf geben, dass das zugrunde liegende Konstrukt nicht angemessen erfasst wird, während eine zu hohe interne Konsistenz darauf schließen lässt, dass manche Items redundant sind (Taber, 2018). Die Definition eines festgelegten Limits, ab welchem Testinstrumente als akzeptabel hinsichtlich ihrer internen Konsistenz interpretiert werden können, sehen einige Autor:innen dabei kritisch (Schmitt, 1996; Taber, 2018; siehe auch Kapitel 2.1.6.1 in Bezug auf die Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK).

2.4.4 Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen

Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen beschreibt den Zusammenhang des Testscores eines betrachteten Instruments mit den Testscores, welche mit Instrumenten erfasst wurden, die konzeptionell zusammenhängende Variablen beschreiben (AERA et al., 2014). Für die Validität der Testwertinterpretation ist dies eine wichtige Voraussetzung, da die erwarteten Zusammenhänge zwischen dem betrachteten Konstrukt und auf theoretischer Grundlage verwandter Variablen für neue Testinstrumente ebenfalls empirisch nachgewiesen werden sollten (McCoach et

al., 2013b). Damit entspricht diese Validitätsevidenz der Konstruktvalidität nach dem traditionellen Validitätskonzept (McCoach et al., 2013b; Sireci & Parker, 2006).

Bei der Untersuchung dieser Validitätsevidenz kann zwischen konvergenten (positive Korrelationen mit Testscores von Instrumenten, die zusammenhängende Konstrukte messen) und diskriminanten Methoden (schwache oder keine Korrelationen mit Testscores von Instrumenten, die theoretisch nicht zusammenhängende Konstrukte messen) unterschieden werden (AERA et al., 2014; McCoach et al., 2013b). Darüber hinaus kann hinsichtlich des Zeitpunkts der Messung zwischen konkurrenter Validität (zeitgleiche Messung mit einem vergleichbaren Testinstrument und Überprüfung des Zusammenhangs) und prädikativer Validität (Vorhersage eines zukünftigen Kriteriums durch ein Testinstrument; McCoach et al., 2013b) unterschieden werden.

Eine weitere Möglichkeit, Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen zu überprüfen, ist ein Kontrastgruppenvergleich (*known-groups analysis*; McCoach et al., 2013b). Dabei wird untersucht, inwiefern sich zwei Gruppen, von denen auszugehen ist, dass sie sich bezüglich des zu erfassenden Merkmals unterscheiden, ebenfalls hinsichtlich des mit dem betrachteten Instrument erzielten Testscores unterscheiden (McCoach et al., 2013b). Diese Annahme über die Unterschiede zwischen den Gruppen muss theoretisch fundiert sein (McCoach et al., 2013b), zum Beispiel basierend auf dem Experten-Novizen-Paradigma (de Groot, 1965).

2.4.5 Validitätsevidenz auf Grundlage von Folgen der Testung

Validitätsevidenz auf Grundlage von Konsequenzen der Testung (*evidence based on the consequences of testing*) beschreibt die Identifikation von beabsichtigten und unbeabsichtigten Konsequenzen des Testgebrauchs (AERA et al., 2014). Auf Basis dieser Validitätsevidenz sollte überprüft werden, ob gewünschte positiven Effekte durch die Anwendung des Testinstruments tatsächlich eintreten, sodass der Test wie bei der Entwicklung beabsichtigt eingesetzt werden kann (AERA et al., 2014). Gleichzeitig kann der Einsatz eines Testinstruments damit auch unbeabsichtigte, bei der Entwicklung nicht vorgesehene, Konsequenzen nach sich ziehen (AERA et al., 2014). Auch wenn diese nicht intendierten Konsequenzen nicht automatisch negative Auswirkungen haben müssen (AERA et al., 2014), wird diese Validitätsevidenz häufig dafür verwendet, die Nutzung neu entwickelter Testinstrumente kritisch zu überprüfen (Sireci & Parker, 2006). Unbeabsichtigte Konsequenzen können dabei durch Fehler im Entwicklungsprozess oder der Testnutzung entstehen (AERA et al., 2014).

3 Zielsetzung der Dissertation und Zusammenfassung der durchgeführten Studien

Das fachdidaktische Wissen als elementarer Bestandteil des Professionswissens einer (angehenden) Lehrkraft (z. B. Baumert & Kunter, 2006) hat einen entscheidenden Einfluss auf ihr unterrichtliches Handeln und folglich auch die Lernleistung ihrer Schüler:innen (z. B. Coe et al., 2014). Es umfasst dabei das Wissen über geeignete Instruktionsstrategien sowie gängige Vorstellungen der Lerngruppe (sowie weitere Komponenten; z. B. van Driel et al., 2023), welches sowohl kontextspezifisch, fachspezifisch oder disziplinspezifisch sein kann (Carlson et al., 2019). Zur Erfassung von PCK, etwa mit dem Ziel der Evaluation der Entwicklung in der Aus- und Weiterbildung, bedarf es geeigneter Erhebungsinstrumente. So eignen sich insbesondere Multiple-Choice-Instrumente durch Vorteile in Bezug auf die Testökonomie und Objektivität (z. B. Haladyna, 2004) für den Einsatz bei großen Stichproben, werden zur Erfassung von PCK jedoch auch kritisch gesehen (z. B. Rosenkränzer et al., 2017), da sie mit Herausforderungen hinsichtlich einer adäquaten Formulierung der Items (z. B. Schmelzing et al., 2013), einer passenden Repräsentation des zugrunde liegenden Konstrukts (z. B. Smith & Banilower, 2015) sowie der psychometrischen Eigenschaften (vgl. Taber, 2018) verbunden sind.

Im naturwissenschaftlichen Unterricht spielt die Erkenntnisgewinnung eine zentrale Rolle (z. B. Mayer, 2007), weshalb die Vermittlung entsprechender Kompetenzen curricular verankert ist (z. B. Deutschland: KMK, 2024b; Australien: ACARA, 2022). Lehrkräfte müssen dabei nicht nur selbst über Erkenntnisgewinnungskompetenzen verfügen (z. B. Khan & Krell, 2019), sondern auch über entsprechendes PCK_{SR}, um diese Kompetenzen adäquat bei ihren Schüler:innen zu fördern. Studien deuten jedoch darauf hin, dass Lehrkräfte häufig selbst nicht ausreichend über Erkenntnisgewinnungskompetenzen verfügen (z. B. Capps et al., 2016) und Schwierigkeiten bei der Planung zur Förderung (z. B. Großmann & Krüger, 2024) sowie einer expliziten Adressierung dieser (z. B. Vorholzer et al., 2022) im Unterricht haben. Dennoch existieren bisher nur wenige Studien, die das PCK_{SR} angehender und praktizierender Lehrkräfte erforschen und sich dabei zusätzlich hinsichtlich diverser Aspekte, darunter der betrachteten Stichprobe sowie verwendeten Erhebungsmethoden und -instrumente, unterscheiden. Dadurch fehlt es an einheitlichen Erkenntnissen, welche jedoch Voraussetzung für ein grundlegendes Verständnis dieses spezifischen Wissensbereichs sind und als Ausgangspunkt für die Gestaltung effektiver Aus- und Weiterbildungsangebote genutzt werden können.

Um diesen beiden Desideraten zu begegnen, sollen im Rahmen dieser Dissertation die folgenden zwei übergeordneten Fragestellungen untersucht werden:

- (I) Wann eignen sich Multiple-Choice-Instrumente zur Erfassung von PCK und wie kann mit der Entwicklung dieser Instrumente einhergehenden Herausforderungen hinsichtlich einer adäquaten Formulierung der Items, einer passenden Repräsentation des zugrunde liegenden Konstrukts sowie der psychometrischen Eigenschaften begegnet werden?
- (II) Wie ist das PCK_{SR} angehender und praktizierender Lehrkräfte ausgeprägt und wie hängt es mit der Berufserfahrung sowie anderer konzeptionell verwandter Variablen zusammen?

Dazu soll ein Multiple-Choice-Instrument zur Erfassung von PCK_{SR} (PCK_{SR}-bio) systematisch entwickelt, validiert und eingesetzt werden. Die Ergebnisse der fünf durchgeführten empirischen Studien werden zur Beantwortung der übergeordneten Fragestellung (I) aus einer methodischen Perspektive und zur Beantwortung der übergeordneten Fragestellung (II) aus einer inhaltlichen Perspektive diskutiert. Zusätzlich erfolgt eine Synthese beider Perspektiven, um eine holistische Betrachtung der Ergebnisse zu ermöglichen.

In Studie 1 wird deshalb zunächst das Antwortverhalten bei der Bearbeitung des PCK_{SR}-bio auf Grundlage des Experten-Novizen-Paradigmas analysiert, um ein besseres Verständnis über Schwierigkeiten bezüglich der Konstruktrepräsentation zu erlangen. In der darauffolgenden Studie 2 wird untersucht, wie die Berücksichtigung unterschiedlicher Validitätsquellen dabei helfen kann, Schwierigkeiten bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK zu begegnen. In Studie 3 wird darüber hinaus betrachtet, ob sich der *PCK map approach* für die Überprüfung von Validitätsevidenz auf Grundlage von Antwortprozessen eignet. In Studie 4 wird der PCK_{SR}-bio in einer Querschnittstudie mit Lehrkräften an Grund- und weiterführender Schule sowie Hochschuldozierenden in der Lehrkräfteausbildung zur Untersuchung von Gruppenunterschieden und Zusammenhängen mit konzeptionell verwandten Variablen eingesetzt. Zuletzt erfolgt in Studie 5 eine explorative Betrachtung der Ausprägung des PCK_{SR} angehender Lehrkräften in Australien.

Die Studien 1, 2 und 3 fokussieren sich damit stärker auf die Beantwortung der übergeordneten Fragestellung (I) aus einer methodischen Perspektive, während in den Studien 4 und 5 ein stärkerer Fokus auf die inhaltliche Perspektive zur Beantwortung von Fragestellung (II) gelegt wird. In den folgenden Abschnitten soll eine kurze Übersicht über die im Rahmen dieser Dissertation durchgeführten Studien gegeben werden.

3.1 Studie 1: Findings from the Expert-Novice Paradigm on Differential Response Behavior Among Multiple-Choice Items of a Pedagogical Content Knowledge Test – Implications for Test Development (Kapitel 4)

In der aktuellen wissenschaftlichen Literatur werden Herausforderungen diskutiert, die mit der Entwicklung von Testinstrumenten zur Erhebung von PCK einhergehen. So sollte beachtet werden, dass es sich um ein themenspezifisches (bzw. disziplinspezifisches) Konstrukt handelt (z. B. Gess-Newsome, 2015), welches verschiedene miteinander in Beziehung stehende Komponenten umfasst (Magnusson et al., 1999; Park & Chen, 2012; van Driel et al., 2023). Darüber hinaus sollten Testinstrumente im Entwicklungsprozess begründet ausgewählt werden: Während offene Paper-Pencil-Tests den Vorteil haben, ein breites Spektrum an der Antwort zugrunde liegende Denkprozesse zu erfassen (Martinez, 1999), gehen sie mit einem hohen Grad an Interpretation bei der Auswertung einher (Hollingworth et al., 2007). Geschlossene Paper-Pencil-Tests wie Multiple-Choice-Instrumente haben hingegen Vorteile in Bezug auf die Objektivität und Testökonomie (Haladyna, 2004), können jedoch keinen Aufschluss über die kognitiven Prozesse der Proband:innen während der Beantwortung geben. Die Entwicklung von Multiple-Choice-Instrumenten geht deshalb mit der Herausforderung einher, dass die zugehörigen Items so entwickelt werden müssen, dass sie das zugrunde liegende Konstrukt valide erfassen und weder unter- noch überrepräsentieren (Smith & Banilower, 2015). Um ein besseres Verständnis über diese Schwierigkeiten zu erlangen, wurde der PCK_{SR}-bio als exemplarisch gewähltes Instrument genutzt, um das Antwortverhalten von Proband:innen unter Bezugnahme des Experten-Novizen-Paradigmas (de Groot, 1965) zu untersuchen. Dazu beantworteten 10 Expert:innen (Wissenschaftler:innen im Bereich Biologiedidaktik, die sich mit den Themen PCK und/oder naturwissenschaftliche Erkenntnisgewinnung beschäftigen) und 10 Noviz:innen (Biologie-Lehramtsstudierende zu Beginn des Bachelors) die Items des PCK_{SR}-bio und begründeten nach jeder Antwort, warum sie sich für die gewählte Antwortoption und gegen die anderen entschieden. Im Anschluss wurde untersucht, wie sich der Testscore, die Konsistenz ihres Antwortverhaltens sowie die Begründungen zur Wahl einer Antwortoption zwischen den Expert:innen und Noviz:innen unterschieden.

3.2 Studie 2: Considering Multiple Sources of Validity Evidence Can Help to Address Challenges in the Development of Pedagogical Content Knowledge (PCK) Multiple-Choice Items (Kapitel 5)

Wie bereits im Rahmen von Studie 1 beschrieben, haben Multiple-Choice-Instrumente zur Erfassung von PCK im Vergleich zu geschlossenen Paper-Pencil-Instrumenten Vorteile in Bezug auf Testökonomie und Objektivität (Haladyna, 2004), gehen jedoch auch mit Herausforderungen hinsichtlich der Konstruktrepräsentation einher (Smith & Banilower, 2015). Auch hinsichtlich psychometrischer Eigenschaften treten bei der Erfassung komplexer, anwendungsbezogener Wissensbereiche wie PCK Herausforderungen auf, zum Beispiel bezüglich einer niedrigen internen Konsistenz (Taber, 2018). Um zu untersuchen, inwiefern die Berücksichtigung unterschiedlicher Validitätsquellen dabei helfen kann, diesen Herausforderungen zu begegnen, wurde der PCK_{SR}-bio basierend auf unterschiedlichen Validitätsquellen evaluiert. Dazu wurde zunächst der systematische Entwicklungsvorgang des PCK_{SR}-bio beschrieben (Validitätsevidenz auf Grundlage des Testinhalts). In Bezug auf Validitätsevidenz auf Grundlage von Antwortprozessen wurden im Anschluss die Daten der zehn Biologie-Lehramtsstudierenden im Bachelor aus Studie 1 erneut ausgewertet, wobei untersucht wurde, ob die Studierenden bei der Wahl einer Antwortoption häufiger auf PCK oder auf oberflächliches, intuitives Wissen rekurrierten. Darüber hinaus wurde eine Querschnittstudie mit 164 angehenden Lehrkräften ($n = 29$ Bachelor, $n = 114$ Master und $n = 21$ Referendariat) durchgeführt, bei der die interne Konsistenz der jeweiligen Gruppen untersucht wurde (Validitätsevidenz auf Grundlage der internen Struktur). Zur Überprüfung der Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen wurde zusätzlich der Zusammenhang zwischen dem Testscore des PCK_{SR}-bio und dem Studiensemester untersucht, da vermutet wurde, dass angehende Lehrkräfte ein höheres PCK_{SR} aufweisen, wenn sie weiter in ihrer universitären Ausbildung fortgeschritten sind (vgl. Krüger et al., 2020). Darüber hinaus wurde der Zusammenhang des PCK_{SR} mit dem CK (d. h. Erkenntnisgewinnungskompetenzen) von 67 Biologie-Lehramtsstudierenden im Master überprüft, wobei ebenfalls von einem positiven Zusammenhang ausgegangen wurde (vgl. Großschedl et al., 2019). Zuletzt wurde eine Interventionsstudie zur Überprüfung der Veränderungssensitivität des PCK_{SR}-bio durchgeführt. Dazu bearbeitete eine Versuchsgruppe von 11 Masterstudierenden das Instrument zu Beginn sowie am Ende einer Semesterveranstaltung, welche sich mit dem Thema „Kompetenzorientierte Lernaufgaben entwickeln in der Erkenntnisgewinnung“ befasste. Der Unterschied im Testscore wurde ermittelt und mit einer Kontrollgruppe von 8 Masterstudierenden verglichen, welche eine Semesterveranstaltung mit dem Thema „Sachkompetenz mit digitalen Medien fördern“ belegten.

3.3 Studie 3: Using the PCK Map Approach to Obtain Validity Evidence Based on Response Processes for a Multiple-Choice Instrument (Kapitel 6)

Wie bereits beschrieben geht die Verwendung von Multiple-Choice-Instrumenten, so auch zur Erfassung von PCK, mit der Einschränkung einher, dass keine direkten Informationen über die zugrunde liegenden kognitiven Prozesse der Teilnehmenden bei der Testbearbeitung erhoben werden. Dies stellt jedoch eine wichtige Quelle für die Validität von Testinstrumenten dar (AERA et al., 2014). Um die Validität der Testwertinterpretation basierend auf dieser Evidenz zu überprüfen, werden deshalb häufig Daten über Antwortprozesse in Interviews oder vergleichbaren Methoden gesammelt (Ercikan & Pellegrino, 2017) und mit einer Vielzahl unterschiedlicher Möglichkeiten ausgewertet (Padilla & Benítez, 2014). Ein möglicher Ansatz, der speziell zur Auswertung von PCK-bezogenen Daten entwickelt wurde, ist der *PCK map approach* (z. B. Park & Suh, 2019). Dabei wird das Pentagon-Modell (Park & Oliver, 2008; siehe auch Kapitel 2.1.2.2) als analytisches Rahmenmodell verwendet, um die individuellen Ausprägungen der PCK-Komponenten und insbesondere auch ihre Verbindungen untereinander zu visualisieren (z. B. Park & Suh, 2019). Im Rahmen dieser Studie soll deshalb

untersucht werden, wie der *PCK map approach* dazu genutzt werden kann, Validitätsevidenz auf Grundlage von Antwortprozessen mit Hilfe von in Interviews erhobenen Begründungen zur Auswahl von Antwortoptionen des PCK_{SR}-bio zu sammeln. Dazu wurden die Daten der zehn Biologie-Lehramtsstudierenden im Bachelor aus Studie 1 basierend auf dem *PCK map approach* erneut ausgewertet und vergleichend zwischen den Items der erfassten PCK-Komponenten oder ihren Verbindungen analysiert.

3.4 Studie 4: Pedagogical Content Knowledge (PCK) about Scientific Reasoning in Different Groups of Science Teachers and Its Relation to Other Variables (Kapitel 7)

Da Erkenntnisgewinnungskompetenzen zentraler Bestandteil des naturwissenschaftlichen Unterrichts sind (z. B. Mayer, 2007), sollten Lehrkräfte an Grund- und weiterführenden Schulen sowie Hochschuldozierende in der naturwissenschaftlichen Lehrkräfteausbildung, welche diese Lehrkräfte auf universitärer Ebene ausbilden, nicht nur selbst über diese Kompetenzen verfügen (z. B. Khan & Krell, 2019), sondern auch wissen, wie sie diese optimal bei Schüler:innen fördern können. Die beschriebenen Gruppen (im Folgenden als „Lehrende“ bezeichnet) müssen somit über PCK_{SR} verfügen. Trotz der zentralen Rolle dieses Wissensbereichs wurden bisher nur wenige Studien durchgeführt, die das PCK_{SR} von naturwissenschaftlichen Lehrkräften untersuchen und sich darüber hinaus hauptsächlich hinsichtlich (1) der untersuchten Stichprobe, (2) der erfassten Dimensionen der Erkenntnisgewinnung und damit zusammenhängender Konstrukte, (3) der betrachteten Ebenen des PCK und damit einhergehend auch (4) der verwendeten Auswertungsmethoden und -instrumente unterscheiden. Dies macht einen Vergleich der Forschungsergebnisse, insbesondere zwischen unterschiedlichen Gruppen von Lehrenden, schwierig. Es ist jedoch davon auszugehen, dass sich die Gruppen hinsichtlich ihres PCK_{SR} unterscheiden, da sie sich maßgeblich hinsichtlich des Alters und den damit verbundenen Erkenntnisgewinnungskompetenzen ihrer Zielgruppe, der Lehrpläne, nach denen sie unterrichten (z. B. GDSU, 2002; KMK, 2024b) sowie ihrer universitären Ausbildung (KMK, 2024a) unterscheiden. Darüber hinaus ist bisher noch unklar, inwiefern das PCK_{SR} naturwissenschaftlicher Lehrkräfte von konzeptionell verwandten Variablen wie ihren transmissiven Überzeugungen, ihrer Offenheit zur Innovation sowie ihrer Berufserfahrung abhängt. Um diese Gruppenunterschiede sowie die potenzielle Abhängigkeit von den genannten konzeptionellen Variablen zu untersuchen, wurde eine Querschnittstudie mit Sachkundelehrkräften an Grundschulen ($n = 113$), Biologielehrkräften an weiterführenden Schulen ($n = 201$) sowie Dozierenden in der universitären Lehramtsausbildung im Fach Biologie ($n = 41$) durchgeführt. Die Teilnehmenden beantworteten neben dem PCK_{SR}-bio auch ein Testinstrument zur Erfassung ihrer transmissiven Überzeugungen (Q4TB; Schlichter, 2012) sowie ihrer psychologischen Innovationsblockade (FPIB; Waßmann, 2014) und gaben Informationen zu ihrer Berufserfahrung an, wodurch grundlegende Kenntnisse hinsichtlich des Unterschieds des PCK_{SR} zwischen verschiedenen Lehrenden der Biologie sowie der Vorhersage ihres PCK_{SR} durch konzeptionell verwandte Variablen erlangt wurden.

3.5 Studie 5: Preservice Science Teachers' Knowledge About How to Teach Scientific Reasoning (Kapitel 8)

Auch wenn sich PCK hauptsächlich durch praktische Erfahrungen und Reflexion dieser entwickelt (z. B. van Driel & Berry, 2010), sollten angehende Lehrkräfte diesbezüglich bereits grundlegende Kenntnisse während des Studiums erlangen. Dies ist nicht nur in den nationalen ländergemeinsamen Anforderungen in der Lehrerbildung (KMK, 2019), sondern auch im australischen Curriculum für die Lehrkräfteausbildung (AITSL, 2017) verankert. Im Rahmen dieser Studie soll deshalb erstmals die Ausprägung des PCK_{SR} angehender Lehrkräfte während der universitären Lehrkräfteausbildung in Australien untersucht werden. Dazu wurde eine explorative Studie mit 36 Studierenden im Master of Education, welche mindestens ein naturwissenschaftliches Fach (Biologie, Chemie, Naturwissenschaften oder Physik) studierten, durchgeführt. Die Teilnehmenden beantworteten dazu eine englischsprachige Version des PCK_{SR}-bio. Im Anschluss wurde explorativ untersucht, inwiefern sich der Testscore von Studierenden im ersten und zweiten Jahr des Masterstudiums sowie zwischen Studierenden, welche Biologie studierten und solchen, welche nicht Biologie (aber mindestens ein weiteres naturwissenschaftliches Fach) studierten, unterschieden. Darüber hinaus wurde die Itemschwierigkeit der Items einzeln sowie gruppiert in Bezug auf die zwei im Instrument betrachteten PCK-Komponenten und ihre Verbindungen sowie den drei betrachteten Denk- und Arbeitsweisen untersucht. Zuletzt wurde verglichen, wie sich der Testscore der australischen Studierenden von dem der deutschen Studierenden zu einem vergleichbaren Zeitpunkt in der Lehrkräfteausbildung ($n = 114$ Masterstudierende aus Studie 2) unterschied.

Tabelle 3.1: Übersicht über die durchgeführten Studien

Studie 1 (Kapitel 4)

Veröffentlichung / Manuskript	Lieberei, T., Welter, V. D. E., Großmann, L. & Krell, M. (2023). Findings from the expert-novice paradigm on differential response behavior among multiple-choice items of a pedagogical content knowledge test – implications for test development. <i>Frontiers in Psychology</i> , 14, 1240120. https://doi.org/10.3389/fpsyg.2023.1240120
Status	Veröffentlicht
Zielsetzung	Analyse des Antwortverhaltens bei der Bearbeitung eines Multiple-Choice-Tests zur Erfassung von PCK basierend auf dem Experten-Novizen-Paradigma, um ein besseres Verständnis über Schwierigkeiten während der Testentwicklung in Bezug auf die Konstruktrepräsentation zu erlangen
Forschungsfrage(n)	1. Inwiefern erreichen Expert:innen in einem PCK_{SR}-Test eine höhere durchschnittliche Lösungswahrscheinlichkeit als Noviz:innen? 2. Inwiefern variieren die Antworten von Expert:innen in einem PCK_{SR}-Test weniger als die Antworten von Noviz:innen? 3. Inwiefern begründen Expert:innen und Noviz:innen ihre Wahl der Antwortoptionen in einem PCK_{SR}-Test unterschiedlich?
Design	Interviewstudie
Stichprobe	• 10 Expert:innen (Wissenschaftler:innen im Bereich Biologiedidaktik, die sich mit den Themen PCK und/oder naturwissenschaftliche Erkenntnisgewinnung beschäftigen) • 10 Noviz:innen (Biologie-Lehramtsstudierende zu Beginn des Bachelors)
Datenauswertung	• Quantitative Auswertung des Antwortverhaltens hinsichtlich des erreichten Testscores und der Varianz • Qualitative Auswertung hinsichtlich der Begründung zur Wahl einer Antwortoption
Primäre Perspektive	methodisch
Verwendete Instrumente	PCK _{SR} -bio (s. Kapitel 12.1)

Tabelle 3.1: Übersicht über die durchgeführten Studien (Fortsetzung)

Studie 2 (Kapitel 5)

Veröffentlichung / Manuskript	Lieberei, T. , Großmann, L., Welter, V. D. E., Krüger, D. & Krell, M. (angenommen). Considering multiple sources of validity evidence can help to address challenges in the development of pedagogical content knowledge (PCK) multiple-choice items. <i>Research in Science Education</i> .
Status	Angenommen
Zielsetzung	Betrachtung unterschiedlicher Validitätsquellen, um Herausforderungen bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK zu begegnen und Hinweise auf eine valide Testwertinterpretation zu erlangen
Forschungsfrage(n)	Wie kann die Berücksichtigung unterschiedlicher Validitätsquellen dazu beitragen, gängigen Herausforderungen in Bezug auf kognitive Operationen und Aufgabenformate, der Repräsentation des zugrunde liegenden Konstrukts sowie psychometrischer Eigenschaften bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK zu begegnen?
Design	Durchführung unterschiedlicher qualitativer und quantitativer Studien zur Überprüfung verschiedener Validitätsevidenzen: <i>Validitätsevidenz auf Grundlage des Testinhalts:</i> • Systematische Instrumentenentwicklung <i>Validitätsevidenz auf Grundlage von Antwortprozessen:</i> • Qualitative Interviewstudie <i>Validitätsevidenz auf Grundlage der internen Struktur:</i> • Querschnittstudie <i>Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen:</i> • Korrelationsstudien • Interventionsstudie

Tabelle 3.1: Übersicht über die durchgeführten Studien (Fortsetzung)

Studie 2 (Kapitel 5)

Stichprobe	<i>Validitätsevidenz auf Grundlage von Antwortprozessen:</i> • 10 Biologie-Lehramtsstudierende zu Beginn des Bachelors (Daten aus Studie 1) <i>Validitätsevidenz auf Grundlage der internen Struktur:</i> • 164 angehende Lehrkräfte ($n = 29$ Bachelor, $n = 114$ Master und $n = 21$ Referendariat) <i>Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen:</i> • Stichprobe aus Querschnittstudie ($n = 164$ angehende Lehrkräfte) • 67 Biologie-Lehramtsstudierende im Master • Versuchsgruppe mit 11, Kontrollgruppe mit 8 Biologie-Lehramtsstudierenden im Master
Datenauswertung	<i>Validitätsevidenz auf Grundlage des Testinhalts:</i> • Beschreibung der systematischen Entwicklung des Instruments <i>Validitätsevidenz auf Grundlage von Antwortprozessen:</i> • Analyse des rekurrierten Wissens bei der Begründung zur Wahl einer Antwortoption <i>Validitätsevidenz auf Grundlage der internen Struktur:</i> • Analyse der internen Konsistenz der unterschiedlichen Gruppen <i>Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen:</i> • Analyse des Zusammenhangs zwischen dem PCK_{SR} und dem Bildungsabschluss sowie dem CK • Durchführung einer Interventionsstudie zur Testung der Veränderungssensitivität
Primäre Perspektive	methodisch
Verwendete Instrumente	• PCK_{SR}-bio (s. Kapitel 12.1) • KoWADiS (Hartmann et al., 2015)

Tabelle 3.1: Übersicht über die durchgeführten Studien (Fortsetzung)

Studie 3 (Kapitel 6)

Veröffentlichung / Manuskript	Lieberei, T., Welter, V. D. E. & Krell, M. (2024). Using the PCK map approach to obtain validity evidence based on response processes for a multiple-choice instrument. In G. Kaya, M. Sardag & I. Donmez (Hrsg.), <i>Proceedings Book Series-V of the ESERA 2023 Conference: Connecting Science Education with Cultural Heritage</i> (S. 114–126). Nobel Bilimsel Eserler.
Status	Veröffentlicht
Zielsetzung	Betrachtung, inwiefern sich der <i>PCK map approach</i> dazu eignet, Validitätsevidenz auf Grundlage von Antwortprozessen hinsichtlich der Entwicklung von Multiple-Choice-Instrumenten zu sammeln
Forschungsfrage(n)	Kann der <i>PCK map approach</i> dazu verwendet werden, Validitätsevidenz auf Grundlage von Antwortprozessen zu erlangen?
Design	Qualitative Interviewstudie
Stichprobe	10 Biologie-Lehramtsstudierende zu Beginn des Bachelors (Daten aus Studie 1)
Datenauswertung	Analyse der Auswahl zur Begründung der Items basierend auf dem <i>PCK map approach</i>
Primäre Perspektive	methodisch
Verwendete Instrumente	PCK _{SR} -bio (s. Kapitel 12.1)

Tabelle 3.1: Übersicht über die durchgeführten Studien (Fortsetzung)

Studie 4 (Kapitel 7)

Veröffentlichung / Manuskript	Lieberei, T., Welter, V. D. E. & Krell, M. (eingereicht). Pedagogical content knowledge (PCK) about scientific reasoning in different groups of science teachers and its relation to other variables. <i>Journal of Research in Science Teaching</i> .
Status	Eingereicht
Zielsetzung	Untersuchung von Unterschieden im PCK _{SR} verschiedener Lehrender in der Biologie sowie der Abhängigkeit des PCK _{SR} von konzeptionell verwandten Variablen
Forschungsfrage(n)	1. Wie unterscheidet sich das PCK_{SR} von Lehrkräften an Grundschulen und weiterführenden Schulen sowie von Hochschuldozierenden in der Lehrkräfteausbildung? 2. Wie wird das PCK_{SR} von Lehrkräften durch das Alter ihrer Zielgruppe, ihren transmissiven Überzeugungen, ihrer Offenheit zur Innovation und ihrer Berufserfahrung beeinflusst?
Design	Querschnittstudie
Stichprobe	• 113 Grundschullehrkräfte mit dem Fach Sachunterricht • 201 Lehrkräfte an weiterführender Schule mit dem Fach Biologie • 41 Dozierende in der universitären Lehrkräfteausbildung im Fach Biologie
Datenauswertung	• Untersuchung statistischer Gruppenunterschiede bezüglich des PCK_{SR} • Untersuchung der Abhängigkeit des PCK_{SR} durch konzeptionell verwandte Variablen mittels linearer Regression
Primäre Perspektive	inhaltlich
Verwendete Instrumente	• PCK_{SR}-bio (s. Kapitel 12.1) • Q4TB (Schlichter, 2012) • FPIB (Waßmann, 2014)

Tabelle 3.1: Übersicht über die durchgeführten Studien (Fortsetzung)

Studie 5 (Kapitel 8)

Veröffentlichung / Manuskript	Lieberei, T. , Dawborn-Gundlach, M., van Driel, J. & Krell, M. (2024). Preservice science teachers' knowledge about how to teach scientific reasoning. <i>Teaching Science</i> , 70(2), 30-43
Status	Veröffentlicht
Zielsetzung	Explorative Untersuchung des PCK _{SR} australischer Lehramtsstudierender
Forschungsfrage(n)	-
Design	Quantitative Datenerhebung
Stichprobe	36 Studierende im Master of Teaching mit mindestens einem naturwissenschaftlichen Fach (Biologie, Chemie, Naturwissenschaften oder Physik)
Datenauswertung	• Analyse von Unterschieden im Testscore zwischen Studierenden im ersten und zweiten Jahr des Masters sowie zwischen Studierenden, die Biologie studierten und solchen, die nicht Biologie studierten • Analyse der Schwierigkeit einzelner Items sowie der Items in Bezug auf die zwei erfassten PCK-Komponenten und ihren Verbindungen sowie die drei erfassten Denk- und Arbeitsweisen • Vergleich des Testscores mit dem einer vergleichbaren Stichprobe Studierender aus Deutschland ($n = 114$ Masterstudierende aus Studie 2)
Primäre Perspektive	inhaltlich
Verwendete Instrumente	PCK _{SR} -bio (englischsprachige Version, s. Kapitel 12.2)

4 Studie 1: Findings from the Expert-Novice Paradigm on Differential Response Behavior Among Multiple-Choice Items of a Pedagogical Content Knowledge Test – Implications for Test Development²⁶

Abstract

Pedagogical content knowledge (PCK) is one core dimension of teachers' professional knowledge and comprises knowledge about conceptual ideas of learners and appropriate instructions. However, several challenges regarding the assessment of PCK are discussed in the literature: For example, PCK is a topic-specific construct and contains differentiable subdomains, which must be considered during test development. In addition, the choice of test type needs to be considered. While open-ended instruments can capture a broader range of cognitions, they often require a high level of interpretation; in contrast, multiple-choice instruments have advantages in terms of objectivity and test economy. Some challenges of assessing PCK are particularly related to multiple-choice instruments, such as an insufficient focus on specific components or the accidental assessment of teachers' beliefs instead of PCK. To better understand and explain these challenges in developing multiple-choice PCK instruments, we exemplarily used an instrument to assess PCK about scientific reasoning and considered the assumptions of the expert-novice paradigm to analyse differential response behavior between $n = 10$ researchers in the field of biology education (experts) and $n = 10$ undergraduate pre-service biology teachers (novices). As expected, experts scored significantly higher than novices. At the same time, experts answered the items more consistently than novices, i.e., showed less variance. However, the difference found was statistically insignificant. Regarding the explanations for choosing a response option, experts more often correctly identified the quintessence of the items, which means that they more often understand the items as intended and argued based on their PCK. On the other hand, novices focused more on surface characteristics, i.e., they argued rather with surface knowledge like intuition or personal experience than choosing the response option based on their PCK. These crucial differences in how experts and novices understand the items of the used PCK instrument and how they respond based on their understanding affect different test characteristics. In conclusion, we recommend ensuring that instruments address only a few, specific PCK aspects, considering the target group of a test, and take into account that target groups with larger variability among their responses require a higher number of items to achieve satisfactory discrimination and reliability indices.

Keywords: expert-novice paradigm, test instrument, pedagogical content knowledge (PCK), scientific reasoning, pre-service science teachers

²⁶ Lieberei, T., Welter, V. D. E., Großmann, L. & Krell, M. (2023). Findings from the expert-novice paradigm on differential response behavior among multiple-choice items of a pedagogical content knowledge test – implications for test development. *Frontiers in Psychology*, 14, 1240120.
<https://doi.org/10.3389/fpsyg.2023.1240120>

4.1 Introduction

Teachers' professional competence has a crucial influence on the effectiveness of their teaching (Blömeke et al., 2022; Hattie, 2008). Besides motivational orientations, beliefs, values, goals, and self-regulation, professional knowledge represents a central aspect of this professional competence (Baumert & Kunter, 2013). Following Shulman (1986, 1987), professional knowledge comprises several domains, of which the most prominent include (1) pedagogical knowledge (PK; knowledge about basic principles of education, teaching, learning, and the learners), (2) content knowledge (CK; knowledge about the discipline to be taught), and (3) pedagogical content knowledge (PCK; an "amalgam" of CK and PK, including knowledge about conceptual ideas of learners and appropriate instruction; Kind & Chan, 2019). These knowledge domains are separable but correlated, as empirically demonstrated for several disciplines, including early literacy (König et al., 2022), mathematics education (Baumert et al., 2010), physics education (e.g., Sorge, Kröger, et al., 2019), or biology education (e.g., Großschedl et al., 2019). Moreover, studies have shown that high levels of PCK among mathematics (Kunter et al., 2013), physics (Keller et al., 2017), and biology (Mahler et al., 2017) teachers have a beneficial effect on the quality of their classroom practice and thus on student learning. Hence, PCK is considered an essential resource for planning, implementing, and reflecting on instruction (Alonzo et al., 2019; Carlson et al., 2019). PCK develops primarily over time as a result of learning and teaching experiences, as well as reflection on them (van Driel & Berry, 2019). In this regard, however, the foundations should already be acquired during (predominantly theoretically oriented) university teacher training, which is also demanded by the standards for teacher education in several countries (e.g., Australia: NSW Education Standards Authority, 2018; Canada: BCTC, 2019; Chile: MINEDUC, 2012; Germany: KMK, 2019). Indeed, findings suggest a growth of pre-service teachers' PCK during university studies (e.g., Kleickmann et al., 2013; Großschedl et al., 2019).

Because PCK is such an essential aspect of teachers' professional competence, several attempts have been made to develop test instruments for measuring this construct as objectively, reliably, and validly as possible (e.g., Großschedl et al., 2019; Loughran et al., 2004; Park et al., 2018; She & Chan, 2023). Chan and Hume (2019) distinguish between two main approaches: (a) Paper-pencil tests (e.g., multiple-choice instruments) and (b) performance analyses, that is, measures which focus on teachers' behavior in PCK-indicative situations (e.g., when planning a lesson). Performance analyses are both more time-consuming and associated with objectivity problems, but they are considered more action-valid (Chan & Hume, 2019). Concerning paper-pencil tests, open-ended instruments can capture a rather broad range of cognitions (Martinez, 1999), they often require a high level of interpretation in the scoring. Therefore, most test practitioners appreciate the advantages of multiple-choice instruments in terms of test economy and objectivity (Haladyna, 2004). However, developing multiple-choice tests is often challenging, for example, in terms of an accurate construct definition (i.e., an under- or overrepresentation of the construct must be avoided), the formulation of clear and unambiguous item stems and response options, or the development of a sufficient number of items to create a reliable instrument (Haladyna, 2004). Therefore, scholars have questioned the possibility to develop multiple-choice instruments to validly assess PCK (e.g., Smith & Banilower, 2015). This poses a challenge because closed-ended instruments are needed to investigate and establish quantitative relationships between PCK components among each other as well as between PCK and other variables.

Against this background, our study aims to investigate challenges that may arise when developing multiple-choice instruments to assess PCK in more detail. For this purpose, we exemplarily use a test instrument for assessing science teachers' PCK about scientific reasoning (PCK_{SR}) in order to explore differential response behavior within the framework of the expert-novice paradigm (de Groot, 1965). This paradigm assumes that experts more often correctly identify the quintessence of tasks due to a more elaborate knowledge network, whereas novices tend to focus on the tasks' surface characteristics (Hogan et al., 2003). Accordingly, we consider the response behavior of researchers in

biology education (experts) as well as that of pre-service biology teachers (novices). Based on theoretical assumptions, we analysed the PCK_{SR} test scores, their variance, and the experts' and novices' explanations for choosing a response option in terms of surface and deep characteristics of the test items. This approach focusing on differential response behavior allows deriving specific recommendations for the construction of PCK-assessing multiple-choice instruments with better psychometric properties.

4.2 Theoretical Background

The following sections describe the construct of PCK as well as its assessment in more detail, followed by a statement of how the expert-novice paradigm and the analysis of response behavior can lead to a better understanding of the functioning of test instruments.

4.2.1 Pedagogical Content Knowledge

As mentioned above, Shulman (1986, 1987) introduced PCK as an “amalgam” of PK and CK. Since this initial definition, several researchers have further developed the construct of PCK (Carlson et al., 2019; Gess-Newsome, 2015; Magnusson et al., 1999; Neumann et al., 2019). A more recent definition specifies it as “the knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes” (Gess-Newsome, 2015, p. 36). Other scholars emphasise that PCK should be conceptualised as encompassing both (theoretical) knowledge and (practical) skills (Chan & Hume, 2019). In a broader sense, PCK thus covers several components that interact with each other. The “pentagon model of PCK” (Figure 4.1), for example, distinguishes five components of PCK: *Orientations to Teaching Science* (e.g., beliefs about the purpose of teaching science), *Knowledge of Instructional Strategies for Teaching* (e.g., subject-specific and topic-specific strategies), *Knowledge of Assessment of Science Learning* (e.g., about different kinds of performance evaluation), *Knowledge of Curriculum* (e.g., about its implications for teaching a specific subject), and *Knowledge of Students' Understanding* (e.g., about [mis]conceptions and prior knowledge; Park & Oliver, 2008). Based on this model, a high level of PCK is characterised by close connections between these components (Park & Chen, 2012). Vice versa, less close or insufficient connections between components indicate poorly developed PCK (Park & Chen, 2012). The two components of *Knowledge of Students' Understanding in Science* (KSU) and *Knowledge of Instructional Strategies for Teaching Science* (KISR), as well as their connection, are considered most crucial for effective teaching (Alonzo et al., 2012; Jin et al., 2015; Park & Chen, 2012). While the connection KISR→KSU reflects the impact of instructions on students' conceptions, the connection KSU→KISR refers to addressing students' (mis)conceptions when designing instruction.

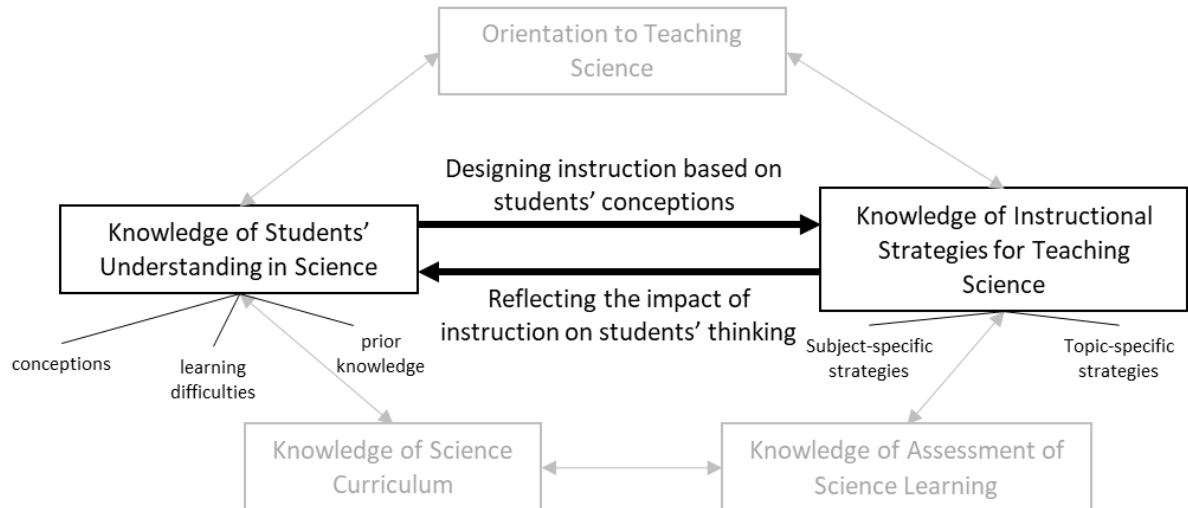


Figure 4.1: Pentagon-model of PCK; KSU and KISR as well as their connections are highlighted
Note. Adapted from Park and Chen (2012); Park and Oliver (2008).

Compared to Park and Oliver (2008), Carlson et al. (2019) argue that the construct of PCK does not merely consist of components, but of different realms of knowledge. In addition to collective PCK (cPCK), the canonical knowledge held by the community of researchers or practitioners in science education, they differentiate between personal PCK (pPCK), which describes the entire PCK an individual teacher possesses, and enacted PCK (ePCK), that is, the part of a teacher's pPCK used during planning, teaching, and reflecting (Alonzo et al., 2019). In view of this, it is very important to always consider the specific realm in which teachers' PCK is assessed when developing instruments (She & Chan, 2023; Sorge, Stender, & Neumann, 2019). Consequently, assessing ePCK requires instruments that provide information about teachers' PCK used during planning, teaching or reflecting (Alonzo et al., 2019). For instance, scholars suggest analysing the quality of teachers' written lesson plans for indicators of their PCK (Großmann & Krüger, 2023) or to use video-based instruments to assess science teachers' PCK from their comments on authentic classroom teaching (She & Chan, 2023). In contrast, instruments that claim to measure pPCK require a cross-situational design. As a consequence, the different realms in which PCK is considered allow for different conclusions about the depth and the breadth of teachers' PCK, respectively (Chan & Hume, 2019). In this manuscript, we will specifically address paper-pencil tests, which can be used to assess pPCK or cPCK (Carlson et al., 2019).

4.2.2 Assessing PCK with Paper-Pencil Tests

Several paper-pencil tests with different task formats exist to assess PCK. For example, some instruments consist of open-ended questions (e.g., CoRe by Loughran et al. [2004]), self-report rating scales (e.g., an instrument by Tepner and Dollny [2014]), multiple-choice items (e.g., POSITT by Schuster et al. [2006]), or mixed-task formats (e.g., PCK-IBI by Großschedl et al. [2019]). In instruments with open-ended questions, teachers are asked to answer in their own words. Such free-response tests aim more at drawing conclusions about the depth of teachers' PCK (Chan & Hume, 2019; Smith & Banilower, 2015). In contrast, closed-ended multiple-choice questions require participants to select one correct response option from several given ones (Chan & Hume, 2019).

While open-ended questions instruments can capture a broader range of cognitions (Martinez, 1999), most test practitioners appreciate the advantages of multiple-choice instruments in terms of test economy and objectivity (Haladyna, 2004). Nevertheless, some researchers argue that PCK, due to its complexity, cannot be captured by multiple-choice items (e.g., Park & Oliver, 2008; Smith & Banilower, 2015). For this reason, some authors have developed instruments that comprise both closed- and open-ended questions to draw conclusions about both, the breadth and the depth of PCK (Großschedl et al., 2019; Jüttner et al., 2013). However, as a consequence, most of these instruments require a quite long test duration (e.g., 24 items that take about 50 minutes to solve in Jüttner et al. [2013]), and the data analysis requires equally long time due to the integration of open-ended items.

4.2.2.1 Challenges in Developing Multiple-Choice Items

If tests consisting exclusively of multiple-choice items are used for reasons of test economy, a major challenge in the development of those instruments is to avoid a possible under- or overrepresentation of the construct. While the former is defined as “the degree to which a test fails to capture important aspects of the construct”, the latter is associated with construct-irrelevant variance, which “refers to the degree to which test scores are affected by processes that are extraneous to the test’s intended purpose” (AERA et al., 2014 p. 12). Therefore, it is recommended to make sure that test items focus on narrowly defined aspects of PCK (Smith & Banilower, 2015). Such a clear focus and an adequate and unambiguous definition of PCK and its constituting aspects also help to avoid a possible underrepresentation of the target construct (Messick, 1996). At the same time, however, due to the definition of PCK as an “amalgam” of CK and PK (Shulman, 1987), there is also a risk of overrepresentation of the construct in terms of construct-irrelevant CK and/or PK variance, which means that “the assessment is too broad” (Messick, 1996, p. 5). In this regard, some multiple-choice instruments for assessing PCK have been criticised for including either tasks that require CK rather than PCK to solve them, or for focusing on teachers’ beliefs about teaching and learning by asking them about appropriate instructional interventions even though empirical evidence for the superiority of an instructional strategy over another strategy is often lacking (Smith & Banilower, 2015). For example, Schmitt (2016) used 19 items to assess chemistry teachers’ PCK about experimentation. Seventeen items describe a classroom situation and possible ways of (re)acting are to be rated on a five-point scale according to their appropriateness for the given situation. In contrast, the remaining two items of the instrument refer to scientific inquiry, asking teachers to sort the steps of scientific inquiry and to correctly assign specific statements to them. The final instrument has a reliability of $\alpha = .76$, while validity was evaluated via expert ratings (Schmitt, 2016). Although the psychometric properties are acceptable, the two items referring to scientific inquiry rather capture CK than PCK, as they address declarative knowledge about the concept of inquiry and not knowledge about how to teach it (i.e., *scientific reasoning competencies*; Khan & Krell, 2019). Another example is Schuster et al.’s (2006) Pedagogy of Science Inquiry Teaching Test (POSITT) for the assessment of science teachers’ PCK about inquiry-based science teaching. The POSITT consists of multiple-choice items that relate to specific instructional settings. The response options for each item include different instruction-related actions (Schuster et al., 2006). In a second publication on this instrument, now called Pedagogy of Science Teaching Tests (POSTT), the authors changed its application: It is now to be used for formative assessment of science teachers’ orientations toward inquiry-based instruction (Cobern et al., 2014). This shift to a different construct during test development, from PCK to orientations, again highlights the challenge with items that were originally designed to capture PCK but are more likely to measure different constructs (Smith & Banilower, 2015).

To summarise, there are important benefits in multiple-choice items in terms of test economy and objectivity (Haladyna, 2004; Martinez, 1999). However, avoiding under- and overrepresentation of PCK is challenging (see Table 4.1). The result might be an instrument with only poor reliability and/or of questionable validity. Smith and Banilower (2015) therefore argued, that the development of

appropriate instruments for the assessment of teachers' PCK will remain a challenge for science education. However, PCK is one of the core constructs in science teacher education, determining teaching quality and student learning outcome (Keller et al., 2017; Kunter et al., 2013; Mahler et al., 2017). Therefore, test instruments are highly needed to evaluate the success of teacher education programs (e.g., Karal & Alev, 2016) or teacher professional development (e.g., Sannert & Krell, 2023), for instance. Hence, exploring the challenges in developing multiple-choice instruments is likely to provide test developers with valuable insights and guidance for the successful development of multiple-choice instruments. We aim to contribute to this exploration with our study by analyzing differential response behavior in the framework of the expert-novice paradigm.

Table 4.1: Two major challenges in the development of multiple-choice PCK instruments

Challenge	Description	Reference
Underrepresentation	The test does not cover all target aspects of PCK	e.g., Smith & Banilower (2015)
Overrepresentation	The test covers more than the target aspects of PCK (e.g., CK, beliefs)	e.g., Schuster et al. (2006)

4.2.3 Analyzing Response Processes within the Framework of the Expert-Novice Paradigm

Response processes refer to the cognitive, affective, and motivational mechanisms underlying responses to test items (Hubley & Zumbo, 2017). Thus, these mechanisms include both intentional, easily verbalizable processes and more unconscious, automatic reactions to test items. Think-aloud settings or interviews are particularly useful for capturing cognitive response processes. These methodological approaches analyse, for example, participants' item interpretations, their response strategies, their applied knowledge, or the way they use item information (Ercikan & Pellegrino, 2017).

In general, two (overlapping) approaches to analyzing response processes can be distinguished: process-model interpretations and trait interpretations (Kane & Mislevy, 2017). Process-model interpretations are based on an exploration of the respondents' task-solving process. In contrast, trait interpretations aim to investigate whether respondents apply those traits (e.g., knowledge, skills) claimed necessary to solve a task, or whether other traits are necessary (Kane & Mislevy, 2017). Both approaches can be used to evaluate the validity of test score interpretations (Ercikan & Pellegrino, 2017; Hubley & Zumbo, 2017).

In our study, we refer to the expert-novice paradigm to derive hypotheses about the test performance of two groups of respondents. Experts are defined as "individuals who exhibit reproducibly superior performance on representative, authentic tasks in their field" (Ericsson, 2006, p. 688). This definition implies that experts are expected to be better at solving tasks in their field than novices because of their superior knowledge and skills. Differences in how tasks and problems are dealt with can therefore be investigated on the basis of the expert-novice paradigm, which is commonly used in educational research. For example, Wolff et al. (2015) investigated differences in experts' (experienced teachers) versus novices' (pre-service teachers) representations of situations requiring classroom management. For this purpose, the descriptions of specific classroom situations were analysed by using a coding scheme that included different kinds of representations, for example, thematic focus or cognitive processing. The results showed a significant difference in cognitive processing between experts and novices, with experts focusing on characteristics such as learning output and novices focusing rather on surface characteristics such as student behavior (Wolff et al., 2015). In another study by Krepf et al.

(2018), the activated knowledge of experts (several criteria, e.g., at least 15 years of teaching experience) and novices (pre-service teachers) was analysed during their evaluation of a videotaped lesson. While the experts activated their PCK, the novices (at best) activated limited CK and PK, which, moreover, were rarely related to each other. In addition, the experts could better solve the problem and describe their ways of doing so (Krepf et al., 2018). A related approach was chosen for the development of the PCK-IBI by Großschedl et al. (2019): The authors compared PCK test scores between experts (experienced teachers) and novices (pre-service teachers) to evaluate the criterion validity of their instrument. They found significantly higher levels of PCK for the experts as well.

However, there are not only quantitative differences between experts and novices, but also those related to the quality of knowledge and its use. For example, novices seem more likely to be distracted by surface characteristics of a problem or task, whereas experts more often understand the core idea, have a deeper understanding, and the ability to transfer this understanding to related problems (Hogan et al., 2003). In the specific context of PCK, experts and novices differ in terms of lesson planning, teaching, and reflection: While experts can adapt their instruction to the learning group's knowledge comparatively easily, novices often use more rigid forms of interaction due to a lack of flexible, transferable knowledge (Hogan et al., 2003). Another study by Forsythe et al. (2022) reports on video analysis/reflection tasks in science teacher education. Here, the authors provided a framework for the analysis of surface and deep characteristics. They assumed that novices "might identify events that are superficial and disconnected from the lesson learning goals. [...] Expert noticing focuses on events that are pivotal to achieving these goals" (p. 632).

Studies in other fields have also made the distinction between surface and deep characteristics of tasks. For the subject of chemistry, for example, Kozma and Russell (1997) showed that novices (undergraduate chemistry students) were often distracted by surface characteristics when solving chemical tasks, whereas experts (professional chemists) used their knowledge in terms of a deeper and more flexible understanding of chemical phenomena. Besides, surface and deep characteristics are important in educational research on the psychometric properties of items and test instruments (e.g., Krell, 2018; Opfer et al. 2012; Schnotz & Baadte, 2015). There is evidence suggesting that surface characteristics influence item difficulty in test-taking by novices, whereas experts can better recognise the deep characteristics of items and therefore respond more consistently to related ones (Clough & Driver 1986; Nehm & Ridgway 2011). We take up these results to establish a distinction between *Surface Thinking* and *Deep Thinking*. *Surface Thinking* means that a respondent focuses on the surface characteristics of items when working on a test, whereas *Deep Thinking* means that the respondent recognises the core idea of the task and relates to it in their response.

In summary, one can make several assumptions in light of the expert-novice paradigm about the response processes of experts and novices when solving multiple-choice items of a PCK test. These assumptions relate primarily to differences in the perception and cognitive processing of the items. For example, experts can be expected to achieve significantly higher test scores than novices because they have a higher level of PCK (Ericsson, 2006). In addition, novices can be expected to show more variance in their responses than experts, who are, vice versa, more likely able to answer items consistently by transferring their knowledge to related ones (Hogan et al., 2003; Krepf et al., 2018; Wolff et al., 2015). Finally, it can be assumed that experts tend to apply *Deep Thinking* more than novices, who, by contrast, are expected to apply more *Surface Thinking* when working on test items. This assumption is based on the consideration that experts have the PCK required to identify the core idea of an item (Forsythe et al., 2022; Hogan et al., 2003).

4.2.4 Aim of the Study and Hypotheses

To better understand and explain the challenges in developing multiple-choice tests to assess science teachers' PCK, we exemplarily used an instrument to measure PCK_{SR} (Section 3.2) and considered the assumptions mentioned above to analyse differences in the response processes of experts and novices. This will provide valuable insights and guidance for the successful development of multiple-choice instruments. Specifically, we investigated the following research questions (RQ) and hypotheses (H):

RQ1: To what extent do experts achieve a higher proportion of correct responses in a PCK_{SR} test than novices?

H1: Experts reach a statistically significant higher proportion of correct responses in a PCK_{SR} test than novices.

RQ2: To what extent do experts' answers in a PCK_{SR} test vary less than novices' ones?

H2: Novices show significantly more variability in their answers in a PCK_{SR} test than experts.

RQ3: To what extent do experts and novices justify their choice of response options in a PCK_{SR} test differently?

H3: When experts justify their choice of response options in a PCK_{SR} test, their explanations are significantly more often indicative of *Deep Thinking*, while novices' explanations are significantly more often indicative of *Surface Thinking*.

4.3 Materials and Methods

The following sections give an overview of sample characteristics, the data collection, scoring, and the statistical methods applied to address our research questions and test our hypotheses. Our study was reviewed and approved by our local ethics committee (ID 2021_LI56) before any data collection took place.

4.3.1 Sample

A total of $N = 20$ participants took part in this study, who could be assumed to belong to one of two groups that differ significantly from each other on the construct of PCK_{SR}. Accordingly, the two groups of participants in the study consisted of experts and novices. The experts were researchers in the field of biology education who were familiar with the construct of PCK_{SR}, which means that they have obtained a PhD as well as conducted research and published in the field of PCK and scientific reasoning. The experts were individually invited to participate in this study via e-mail. The novices were undergraduate pre-service biology teachers from the authors' university and were invited via e-mail lists of their university courses. Participation was voluntary.

Based on the considerations above (see Section 4.2.3), we expected undergraduate pre-service biology teachers to be PCK_{SR} novices because they just started their teacher university program and did not have the necessary learning opportunities to develop PCK since they are at the beginning of their university studies. At the same time, the experts in the field of biology education already did research and published about PCK and scientific reasoning, which makes it necessary to deal with it intensively.

A total of $n = 10$ experts agreed to participate in this study which is why we also randomly selected $n = 10$ novices to reach two groups of equal size. We considered this sample size as large enough to explore the challenges in developing multiple-choice PCK tests.

4.3.2 Data Collection

A one-on-one interview was conducted and recorded with each participant via videoconference. The participants answered a total of 12 multiple-choice items from a test designed to assess PCK_{SR} during the interviews.

4.3.2.1 Interview Protocol

At the beginning of the interview, the participants were informed about the purpose of the study, data collection, and data management. After consenting to take part in the study, it was explained that the task would be to answer 12 multiple-choice items from a test designed to assess PCK_{SR} and to verbally explain the responses.

After answering an item, the participants justified why they chose that particular response option and explained why they did not choose another option, respectively. If a participant could not decide on a response option, they were asked to explain why no option seemed right. Subsequently, they did not have to select a response option.

This process was repeated analogously for all items. After all items were answered, the interview ended.

4.3.2.2 Multiple-Choice Test

All items of the multiple-choice test comprised one attractor (correct response option) and two distractors (incorrect response options), as suggested by Rodriguez (2005). After completing the test, the participants were asked to justify their choice of response options and to explain why they did not choose another option, respectively.

As the multiple-choice instrument was designed to assess PCK_{SR}, it refers to teaching scientific reasoning, an essential part of science and biology education (Khan & Krell, 2019; Mathesius et al., 2016). Biology teachers are expected to foster their students' scientific reasoning competencies to enable them to participate in discussions and decision-making on current social challenges (Osborne, 2013). Therefore, scientific reasoning competencies are emphasised in science educational standards and curricula in various countries worldwide (e.g., Australia: ACARA, 2022; Canada: Ministry of Education, 2008a, b; Chile: MINEDUC, 2015, 2019; Germany: KMK, 2005, 2020). An essential prerequisite for promoting such competencies among students is that their teachers have an appropriate level of PCK_{SR} which relates to methodological and pedagogical competencies for teaching scientific reasoning (Günther et al., 2019; Justi & van Driel, 2006).

Based on Smith and Banilower's (2015) recommendation to focus on capturing narrowly defined PCK aspects, the items of our PCK_{SR} test refer to only two components of PCK (KSU, KISR) as well as their connections (KISR→KSU, KSU→KISR). Empirically, KSU and KISR have been shown the most crucial components for effective science teaching (Alonzo et al., 2012; Jin et al., 2015; Park & Chen, 2012), which is why existing test instruments mainly focus on one (e.g., Sadler et al., 2013) or both of them (e.g., Großschedl et al., 2019; Park et al., 2018). As scientific reasoning involves numerous different activities, there is not yet a standard definition of the construct (Krell et al., 2022). Rönnebeck et al. (2016) found, among others, three highly relevant scientific reasoning skills: (1) planning and conducting investigations, (2) generating hypotheses, and (3) using scientific models. These skills also form the dimensions of our PCK_{SR} instrument (Table 4.2).

Table 4.2: Systematic overview of the 12 multiple-choice items, developed to capture three skills of scientific reasoning, two PCK components, and their connections, respectively

	Planning and Conducting Investigations	Generating Hypotheses	Using Scientific Models
KISR	item #1	item #2	item #3
KSU	item #4	item #5	item #6
KISR→KSU	item #7	item #8	item #9
KSU→KISR	item #10	item #11	item #12

After clarifying the construct to be assessed by choosing the two PCK components (Park & Oliver, 2008) and three crucial scientific reasoning skills (Rönnebeck et al., 2016), the items were developed in an elaborated sequence: First, the general design of the items was conceptualised. Each item of our PCK_{SR} instrument consists of a stem, the prompt, and three response options (Figure 4.2). The prompts are identical for all related items (e.g., the KISR items 1, 2, and 3) to reduce the possible influence of construct-irrelevant sources of variance (Messick, 1996). Moreover, all stems are nearly of equal length and contain identical pieces of information, for example, an outline of students' skills, the topic, or the intended learning outcome (for an example, see item #12 in Figure 4.2). After the item stems had been determined, they were first presented in the form of open-ended questions to a sample of $n = 14$ pre-service biology teachers in order to generate response options that were as ecologically valid as possible and that corresponded to the knowledge and skills of the target group (Haladyna, 2004). Afterward, we developed both correct (attractors) and wrong response options (distractors). They are all about the same length and of similar phrasing. Such standardization is intended to increase the likelihood that respondents will think about the semantic content of a response option, rather than being attracted by superficial cues. Subsequently, the items were administered to two samples of pre-service biology teachers to collect empirical validity evidence (e.g., evidence based on internal structure; AERA et al., 2014). The item development and the result of the validation study are reported in Lieberei et al. (2023), which is why we do not report this in more detail within the current article. Figure 4.2 shows a sample item that relates to the connection KSU→KISR on the “using scientific models” dimension. The full instrument is available upon request from the last author.

Stem: A biology teacher has worked with his students on the structure of DNA using models. In a subsequent lesson, he asks his students what models are and what function they have in biology. One student answers: “Models are visualisations of something, for example, DNA. We can use them to learn how DNA is built.”

Prompt: Please tick the ways in which the teacher should take this idea about models and extend it through an understanding that is as scientifically and methodologically appropriate as possible.

Options:

He should...

- (1) ... ask his students to make their own DNA models to highlight the potential of models as representations.
- (2) ... practise “model critique” with his students by working out similarities and differences between DNA models and DNA.
- (3) ... reflect on models by elaborating on the importance of DNA models in the discovery of DNA structure.

Figure 4.2: Sample item (item #12)

Note. In response options 1 and 2, models are portrayed as mere visualizations according to the classical conception of students (Upmeier zu Belzen et al., 2019). For this reason, the first two response options do not describe appropriate ways to guide students’ ideas toward an adequate scientific understanding. In contrast, the third answer option describes an instructional intervention that emphasises the role of models as epistemic tools in the context of scientific reasoning (Göhner & Krell, 2022; Knuuttila, 2011).

4.3.3 Scoring

To address RQ1 and RQ2, we analysed whether participants selected the attractor or one of the two distractors. If a participant decided not to answer an item (e.g., because they considered the response options ambiguous or were unsure about the correct answer but did not want to fail), this was treated as a missing response that was excluded from subsequent data analyses. To address RQ3, all interviews were first transcribed word-by-word, as described in Dresing and Pehl (2018). In the second step, we assigned the participants’ justifications for choosing or not choosing response options for each item. If a justification referred to more than one response option at once, it was assigned to all affected ones. If a justification could not be assigned to any response option, it was assigned to the item in general. In a third step, we further analysed the reorganised transcripts by means of a coding scheme in a structured qualitative content analysis (Schreier, 2014). Its development was carried out deductively-inductively (Schreier, 2012). In this process, we first deductively developed an initial version in line with the distinction between surface and deep characteristics (Hogan et al., 2003; Krepf et al., 2018), which exclusively included the main categories of *Surface Thinking* and *Deep Thinking*. This preliminary coding scheme was then tested step by step during the interviews with the participants to check whether it reflected their answers in a sufficiently exhaustive way. If an answer could no longer be easily classified into the existing categories (e.g., because something new was addressed), the coding scheme was expanded by creating a new (sub)category (Schreier, 2012). Subsequently, 10% of the interviews (one novice and one expert interview) were coded by both the first author of this article and a trained student assistant. Discrepancies between the two coders were discussed afterward to refine the category system.

The final coding scheme comprised the two main categories *Surface Thinking* and *Deep Thinking* as well as 16 subcategories (Table 4.3). Some of the ten inductively developed subcategories of *Surface Thinking* have also been used in previous studies (e.g., *CK* or *PK*; Krepf et al., 2018), while others are theory-based (e.g., *Item Stem*, based on Kunter and Trautwein [2013]). The six subcategories of *Deep Thinking* were deductively derived from the pentagon model (Section 2.1; Park & Chen, 2012; Park & Oliver, 2008). The complete version of the coding scheme can be found in Section 12.4.

As part of the structured qualitative content analysis, the participants' responses were assigned to the 16 subcategories of the coding scheme. Here, the shortest coding units were less than one sentence, while the longest were less than one paragraph. If a participant's justification referred to more than one response option, it was coded in the same category for all affected response options. If a participant had changed their choice of a response option, only the justification for their final choice was coded. Statements that were no justifications for choosing a response option (e.g., an analysis of the item context) were not coded. We then built sum scores for each participant and each item (i.e., 1 = subcategory was coded, 0 = subcategory was not coded). Therefore, a maximum sum score of 12 could be achieved on each subcategory.

Inter-coder agreement was ensured by following recommendations of Wirtz and Caspar (2002). To do so, 20% of the transcripts were coded independently by a second rater (the same trained student assistant as before). To increase the probability of representativeness, $n = 2$ interviews of novices and experts each were randomly selected (O'Connor & Joffe, 2020). The intercoder agreement was then calculated separately for each subcategory (Table 4.3). For most subcategories, Cohen's Kappa indicated a substantial ($.61 \leq \kappa \leq .80$) to almost perfect ($\kappa \geq .81$) intercoder agreement (Landis & Koch, 1977). For the two subcategories *Item Stem – Topic* and *Item Stem – Teaching Methods*, a moderate agreement was reached ($.41 \leq \kappa \leq .60$), while in two cases a slight ($.00 \leq \kappa \leq .20$) or poor agreement ($\kappa < .00$) was found (*Misconception* and *Knowledge of Instructional Strategies and Representations of Teaching Science: Topic-Specific-Strategies*); the latter likely as a result of only a few codings in the respective subcategories (see Table 4.5 for the number of codings in the total sample). On average, an intercoder agreement of $\kappa = .72$ resulted for the subcategories of *Surface Thinking* while one of $\kappa = .87$ resulted for the subcategories of *Deep Thinking*. All discrepancies between the two coders could be resolved by discussion.

Table 4.3: Short version of the coding system

Main Category	Sub Category	Intercoder Agreement (κ)
Surface Thinking	Item Stem – Student Characteristics	.82
	Item Stem – Topic	.48
	Item Stem – Teaching Methods	.54
	Personal Experience	1.00
	Intuition	.79
	Wrong Interpretation	.70
	Misconception	-.02
	Pedagogical Knowledge (PK)	.76
	Content Knowledge (CK)	.79
	No Relation to Scientific Reasoning	.62
Deep Thinking	Orientations to Teaching Science	not coded
	Knowledge of Students' Understanding	.83
	Knowledge of Science Curriculum	1
	Knowledge of Instructional Strategies and Representations of Teaching Science: Topic-Specific-Strategies	.00
	Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies	.79
	Knowledge of Assessment of Science Learnings	not coded

Note. Categories that did not occur in the selection of transcripts for the intercoder agreement are marked as “not coded”. The complete version of the coding system can be found in Section 12.4.

4.3.4 Statistical Analyses

To address RQ1 and RQ2, we first calculated the proportion of correct responses by dividing the number of correct answers (max. 12) by the total number of items answered (max. 12). This procedure was chosen to account for the fact that some participants had not answered some questions (possibly for reasons unknown to us; Section 4.3.3). We then tested the proportion of correct answers calculated in this way for normal distribution in both groups before running a one-tailed *t*-test for independent groups (RQ1) as well a Levene's test for equality of variances (RQ2).

To address RQ3, the sum scores in each subcategory of *Surface Thinking* and *Deep Thinking* were again summed up to build a total score for these two main categories. These total scores were normally distributed in both groups in the case of *Deep Thinking*, but not in that of *Surface Thinking* (Shapiro-Wilk-test). For this reason, a Mann-Whitney-U test was conducted to address RQ3.

The interview coding and qualitative analyses were done with MAXQDA 2022; the quantitative analyses with IBM SPSS Statistics 26. Effect sizes were calculated using an online tool by Lenhard and Lenhard (2022).

4.4 Results

In the following, the results of the statistical analyses are presented separately for each of the research questions, additionally illustrated in each case by corresponding text examples from the interview transcripts. Table 4.4 provides an overview of the descriptive statistics for each item of the PCK_{SR} test. Item #12 turned out to be the item with the largest difference between experts and novices in terms of the proportion of correct answers, while item #7 involved the largest difference in the variability of answers (in terms of standard deviation).

Table 4.4: Descriptive statistics for each item of the PCK_{SR} items and the full test score

		Items												Test score
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	
Experts	<i>M</i>	0.70	0.44	0.38	0.80	0.90	0.89	0.89	0.80	0.70	0.33	0.11	0.89	0.67
	<i>SD</i>	0.46	0.50	0.48	0.40	0.30	0.31	0.31	0.40	0.46	0.47	0.31	0.31	0.12
Novices	<i>M</i>	0.50	0.60	0.20	0.60	0.70	0.80	0.60	0.50	0.70	0.40	0.30	0.30	0.52
	<i>SD</i>	0.50	0.49	0.40	0.49	0.46	0.40	0.49	0.50	0.46	0.49	0.46	0.46	0.20
Total sample	<i>M</i>	0.60	0.53	0.28	0.70	0.80	0.84	0.74	0.65	0.70	0.37	0.21	0.58	0.59
	<i>SD</i>	0.49	0.50	0.45	0.46	0.40	0.36	0.44	0.48	0.46	0.48	0.41	0.49	0.18

Note. *M* = mean score, *SD* = standard deviation.

4.4.1 RQ1: To what extent do experts achieve higher proportion of correct responses in a PCK_{SR} test than novices?

The *t*-test revealed that the experts reached a statistically significant higher proportion of correct answers (*M* = 0.67, *SD* = 0.12) than the novices (*M* = 0.52, *SD* = 0.20), *t*(18) = 2.03, *p* = .029. This corresponds to a large effect, *d* = 0.91 (Cohen, 1988).

The following parts of the interview transcripts, which refer to item 12 (Figure 4.2), can be used to consider potential reasons for the group differences found. Most of the experts (8 out of 9 [1 missing answer]) identified the attractor correctly and were able to justify their choices accordingly, while most of the novices (7 out of 10) ruled out the attractor as the correct answer.

Two examples of experts' justifications:

I would say the third one is [the right answer]. [...] In my opinion, this item addresses that this student did not understand that the model is an element in the process of scientific inquiry. [The third response option] [...] specifically addresses this. Therefore I think this answer is most convincing. - (Expert 4) [Coding: Knowledge of Students' Understanding and Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies]

Here, students have an [...] idea of models that is not at all predominant in biology. The point is that the students learn that models are not only there for illustration, but are also used to gain knowledge and I see this in answer option (3) [...]. There, the gain in knowledge becomes more apparent and the students realize [...] [that] the model was previously used to elucidate the structure in the first place. - (Expert 9) [Coding: Knowledge of Students' Understanding and Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies]

Two examples of novices' justifications:

'Reflect on models'- I don't find that important. So I would rather take (2) [because] I really don't know what [the way in] (3) is supposed to do. [...] I don't know, I find that somehow strange. And then reflecting on models- I don't find that important either. So I would rather take (2). [...] And with (3)- [...]. Yes, I don't really know what the point is. [...] Yes, I don't know, this is somehow strange. - (Novice 4) [Coding: No Relation to Scientific Reasoning and Intuition]

I would definitely not choose the third one, because I think it only promotes the misconception. [This is because] it says 'reflect on models by elaborating on the importance of DNA models in the discovery of DNA structure' which is exactly what they already know. [...] This would confirm the student in her idea that models are [visualizations of something and] used to see the structure [of DNA], because by doing [the way in answer option (3)] in the lesson, only the structure of the models is addressed. So I think you have to go into it further. - (Novice 1) [Coding: Wrong Interpretation]

4.4.2 RQ2: To what extent do experts' answers in a PCK_{SR} test vary less than novices' ones?

Levene's test revealed that the variability in the experts' responses ($s^2 = 0.04$) was not significantly different from that in the novices' responses ($s^2 = 0.02$), $F(1, 18) = 0.94$, $p = .340$. However, a clear trend can be seen in nominal terms (Figure 4.3).

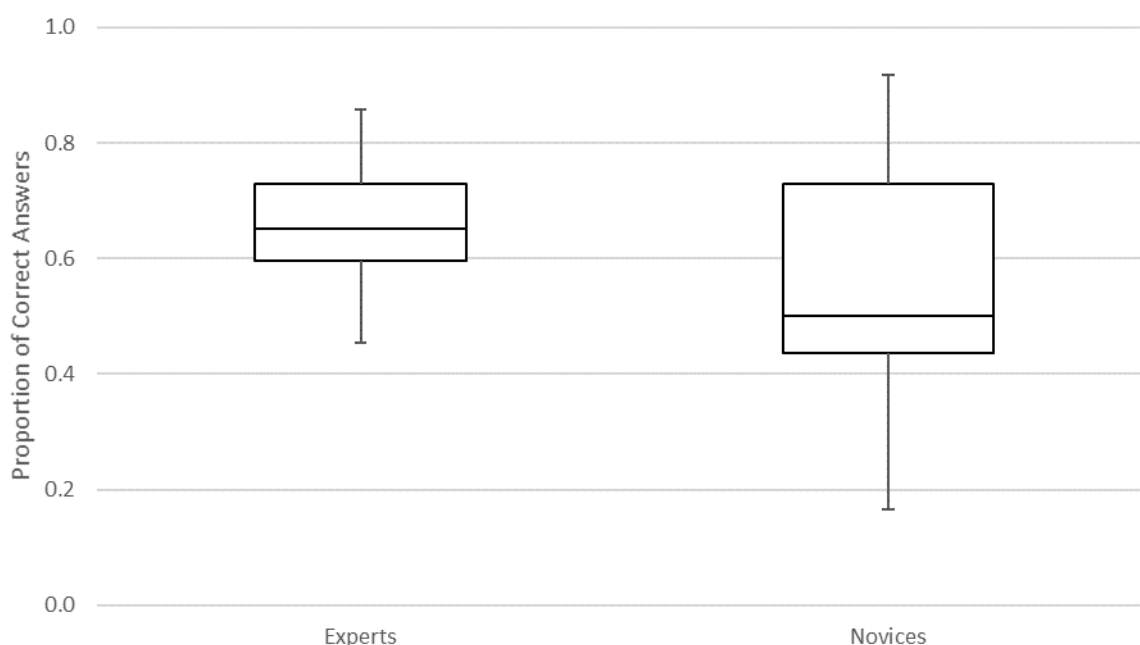


Figure 4.3: Boxplots of the proportions of correct answers

This result corresponds to the pattern of justifications for choosing a response option. Whereas the experts' justifications could be mainly assigned to only two subcategories of *Deep Thinking* (*Knowledge of Students' Understanding* and *Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies*), the novices' justifications point to numerous different considerations in their choices of response options (Figure 4.4). These differences probably reflect the higher variability in the proportion of correct answers in the novices group.

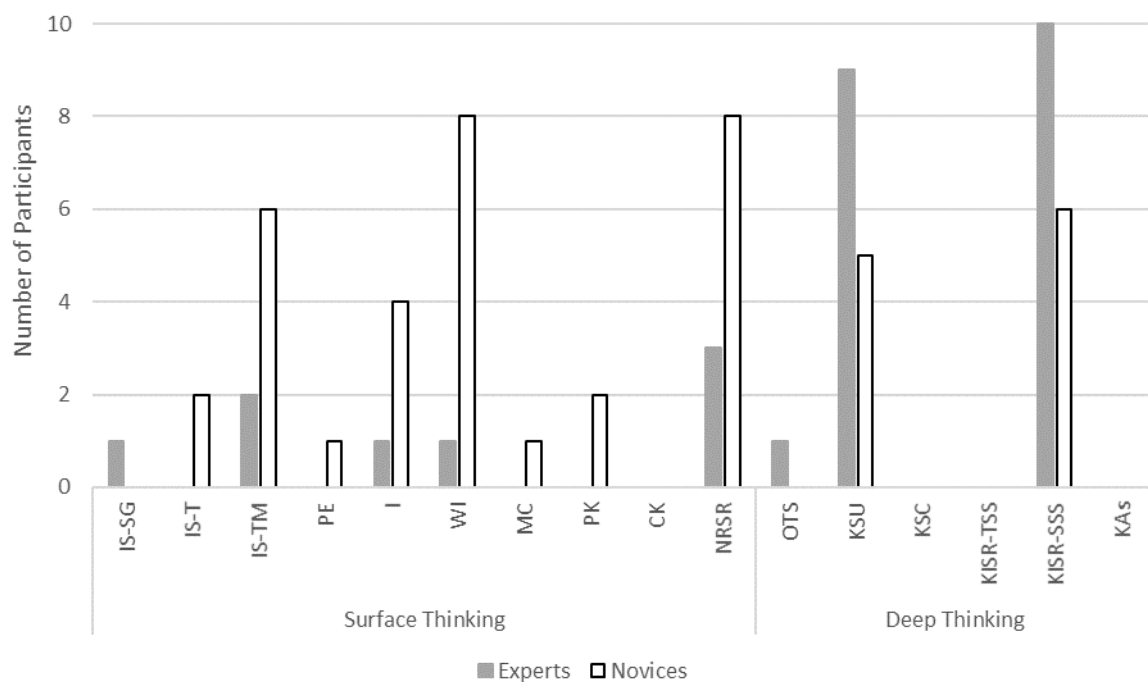


Figure 4.4: Comparison of the number of justifications assigned to *Deep Thinking* and *Surface Thinking* Note. The number of experts and novices whose justifications for choosing a response option was assigned to a specific subcategory of *Deep Thinking* and *Surface Thinking*. IS-SG Item Stem – Study Group; IS-T Item Stem – Topic; IS-TM Item Stem – Teaching Methods; PE Personal Experience; I Intuition; WI Wrong Interpretation; MC Misconception; NRSR No Relation to Scientific Reasoning, OTS Orientations to Teaching Science; KSU Knowledge of Students’ Understanding; KSC Knowledge of Science Curriculum; KISR-TSS Knowledge of Instructional Strategies and Representations of Teaching Science: Topic-Specific-Strategies; KISR-SSS Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies; KAs Knowledge of Assessment of Science Learnings.

The following parts of the interview transcripts, which again refer to item #12 (Figure 4.2), may indicate potential reasons for the group differences found.

Two examples of experts’ justifications:

Answer option (2) would be too simple in my opinion. The students would only learn what they already know, that [models] are visualizations of something which we can use to learn how DNA is built. In answer option (2), they would only discuss, for example, aspects like redundancy and size, but that would have no additional knowledge gain, because they could figure this out on their own.
- (Expert 1) [Coding: Knowledge of Students’ Understanding and Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies]

The point in answer option (3) is that models should not only be seen as a teaching method, as a visualization, but that it should really help for scientific inquiry, [...] which means that the students should learn the purpose for which models are used in science [...]. - (Expert 5) [Coding: Knowledge of Students’ Understanding and Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies]

Four examples of novices' justifications:

[The way described in answer option (1) would not be the right way.] When the student sees the visualization, her own DNA model would be very similar to the model that she already knows and she would still have the same problem with understanding [the function of models]. - (Novice 9) [Coding: Knowledge of Students' Understanding and Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies]

Okay. (...) I would take (2) in any case. [...] What she says is correct, you can see how the DNA is built, depending on how simplified it is. And then I would just go a step further and ask [the student] [...] what's missing in the model, [so that she can practice 'model critique']. - (Novice 4) [Coding: Item Stem – Teaching Methods]

Answer option (3), [...] that is simply reflecting on the learning process and seeing how important the model was for the understanding of the structure. But I think in the end you get the same result as with answer option (2), but there the learning process [...] is not as given as with answer option (3) because [...] you give them the information there and they work it out for themselves in answer option (2). I would exclude answer option (3) because the students are already given the knowledge acquisition and they cannot make it for themselves in their own thought process. - (Novice 8) [Coding: Wrong Interpretation and Item Stem – Teaching Methods]

I also did not choose answer option (3), [...] even though it is an important and correct way. In order to take the student's idea, I think that's missing the point a bit, because it's already about the discovery of the DNA structure, where I don't know how much has already been said about the research of DNA with the students. [...] Actually, you teach the students what the current knowledge is and [...] how you got to that. They should not throw [the knowledge] all together and then go into an exam with the knowledge from the 18th century. They should really know what is assumed nowadays. That's why I think that practicing 'model critique' [in answer option (2)] is much more beneficial to extent the students' idea through an understanding that is as scientifically and methodologically appropriate as possible than answer option (1) or (3). - (Novice 3) [Coding: No Relation to Scientific Reasoning]

4.4.3 RQ3: To what extent do experts and novices justify their choice of response options in a PCK_{SR} test differently?

A total of 695 codings were done in the analysis of the $N = 20$ interviews ($n = 435$ codings of *Surface Thinking*, $n = 260$ codings of *Deep Thinking*). A detailed overview of the codings can be found in Table 4.5. The most frequently coded subcategory of *Surface Thinking* was *Wrong Interpretation* for both experts and novices. The most frequently coded subcategory for *Deep Thinking* was KSU, also for both experts and novices.

Table 4.5: Overview of the codings

Category	Number of Codings		
	Experts	Novices	Total sample
Surface Thinking			
Item Stem – Student Characteristics	19	24	43
Item Stem – Topic	2	4	6
Item Stem – Teaching Methods	22	36	58
Personal Experience	1	11	12
Intuition	22	66	88
Wrong Interpretation	44	68	112
Misconception	1	4	5
Pedagogical Knowledge (PK)	8	25	33
Content Knowledge (CK)	10	6	16
No Relation to Scientific Reasoning	22	40	62
Sum	151	284	435
Deep Thinking			
Orientations to Teaching Science	12	0	12
Knowledge of Students' Understanding	70	53	123
Knowledge of Science Curriculum	2	0	2
Knowledge of Instructional Strategies and Representations of Teaching Science: Topic-Specific-Strategies	2	1	3
Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies	77	43	120
Knowledge of Assessment of Science Learnings	0	0	0
Sum	163	97	260

The total sum scores were used to compare experts and novices with regard to the categories of *Surface Thinking* and *Deep Thinking*. The Mann-Whitney-U test revealed a significantly higher level of *Deep Thinking* for experts ($Sum = 163$) than for novices ($Sum = 97$), $U = 5.50$, $z = 3.39$, $p < .001$, $d = 2.28$, whereas novices showed *Surface Thinking* ($Sum = 284$) significantly more often than experts ($Sum = 151$), $U = 0.00$, $z = 3.79$, $p < .001$, $d = 3.16$ (Figure 4.5).

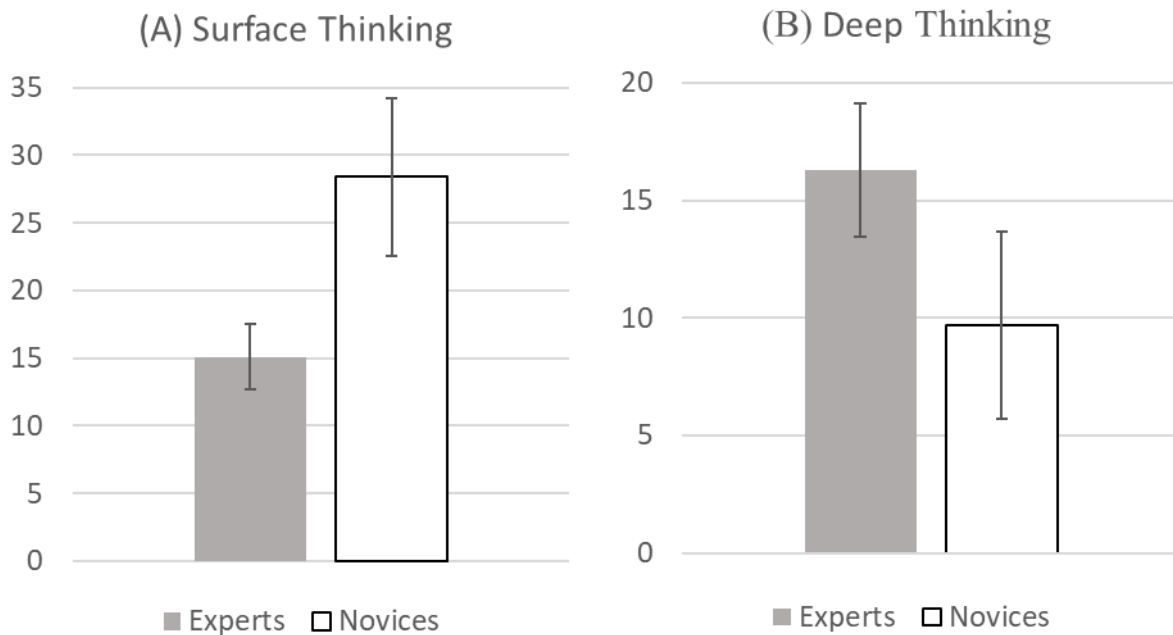


Figure 4.5: Sum of the codings (± 1 standard deviation) of (A) *Surface Thinking* and (B) *Deep Thinking*. *Note.* Sum of the codings for novices and experts compared. The total sum scores for (A) *Surface Thinking* are generally higher than those for (B) *Deep Thinking* because of the different numbers of subcategories (10 versus only 6; see Table 4.5).

The different levels of experts' and novices' *Surface Thinking* and *Deep Thinking* can also be illustrated in the participants' justifications for choosing a response option of item #12. Two experts, for example, said:

However, I would now tend towards answer option three, [...] because this addresses exactly the point that the student does not perceive. She states that models are pure visualizations of reality, but does not consider that insights can also be gained through such model considerations. - (Expert 8) [Coding: Knowledge of Students' Understanding and Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies]

Making DNA models to highlight the potential [in answer option (1)] rather promotes the perspective [of models as] representations which would reinforce the idea [of the student]. Practicing 'model critique' [in answer option (2)] is also a medial perspective and would not lead further in terms of scientific methodology. I'll take answer option (3) then. There you have a historical example, Watson and Crick, where you can draw the importance of DNA models [in the discovery of DNA structure] over time, [by showing] the changes with more increase in knowledge, also from other researcher. - (Expert 6) [Coding: Knowledge of Students' Understanding and Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies]

Here the experts argue with both, *Knowledge of Students' Understanding* and *Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies*. This corresponds exactly to the theoretical background of the item, which was intended to address the interaction KSU→KISR (Table 4.2). One expert also addressed *Orientations to Teachings Science* in their justification:

That's easy, [I don't choose] answer option (1). Models as representations, this would promote the idea of the student, because the visualization character of models is promoted by this, which is level 1 in the competence model [by Upmeier zu Belzen et al. (2019)]. [For answer option (2)] [...], this is similar to what I said earlier. It is not about practicing 'model critique' within the meaning of an understanding that is scientifically and methodologically as appropriate as possible, because that is also not on level 3 of the competence model [by Upmeier zu Belzen et al. (2019)]. But [the lesson] should be about reflecting on the purpose of models. In this case, what purpose the model has in order to discover the phenomenon of DNA structure. - (Expert 3) [Coding: Knowledge of Students' Understanding, and Knowledge of Instructional Strategies and Representations of Teaching Science: Subject-Specific-Strategies, and Orientations to Teaching Science]

Regarding *Surface Thinking*, a novice, for example, justified their answer as follows:

But I think that practice 'model critique' is just a bit more plausible and perhaps sticks a bit more [than the way in answer option (3)]. I don't know why at the moment. - (Novice 7) [Coding: Intuition]

This is a representative example of the subcategory *Intuition* because it is not a justification based on the PCK_{SR} of the participant. Instead, the novice cannot justify why they chose this answer option. Many novices, on the other hand, did not interpret the item in the intended way. For example:

I don't think they have to reflect on models and work out the importance of DNA models in the discovery of DNA structures [as in answer option (3)]. That is already the idea that is in the students' heads, that you can learn with it how the DNA is built and I think that's also a very intuitive idea. So I think they don't have to reflect on it. - (Novice 1) [Coding: Wrong Interpretation]

Other novices justify without reference to scientific reasoning. For example, the following participant only argues that model criticism is the right way. They do not argue with scientific reasoning competencies.

I would choose answer option (2), simply because the student currently has the idea that models represent the truth and how the DNA is structured. But just because of the size that can't be right. And I think it's important to make it clear to the students that [a model] is not a biologically accurate visualization of something [...] [and] therefore give them a critical view of it [...]. - (Novice 3) [Coding: No Relation to Scientific Reasoning]

4.5 Discussion

With our study, we aimed to contribute to a better understanding of the challenges in developing multiple-choice instruments to assess science teachers' PCK. For this purpose, we exemplarily used an instrument for the assessment of PCK_{SR} and considered assumptions from the expert-novice paradigm to analyse differences in the response processes of experts and novices. Our research questions referred to potential group differences regarding the proportion of correct responses (RQ1), the variability in that proportion (RQ2), and the frequency to which justifications for choosing a response option indicate rather *Surface Thinking* or *Deep Thinking* (RQ3). While the corresponding hypotheses

1 and 3 could be empirically supported by the results, the difference postulated in hypothesis 2 was only found at a nominal, but not at a statistically significant level.

4.5.1 RQ1: To what extent do experts achieve higher proportion of correct responses in a PCK_{SR} test than novices?

The significantly higher proportion of correct responses among the experts is in line with previous studies which showed that the knowledge of novices is on a significantly lower level (Großschedl et al., 2019; Hogan et al., 2003; Krepf et al., 2018). This assumption is also supported by the analysis of the qualitative data. For item construction in instruments to measure PCK, this means that the target group must be given appropriate consideration. Methodically, the proportion of correct answers corresponds to the mean item difficulty (Schweizer & DiStefano, 2016). According to test-theoretical recommendations, the individual items' difficulties should range between .20 (20% of participants solve the item correctly) and .80 (80% of participants solve the item correctly), whereas, across all items, a medium item difficulty is to be aimed for (Schweizer & DiStefano, 2016). Given that, our PCK_{SR} instrument seems perfectly suited to capture the construct in undergraduate pre-service teachers (Table 4.4), but it seems less suited to assess experts' PCK_{SR} because their average probability of correctly answering an item is .67, which is clearly above the recommendation of a medium item difficulty. In this context, however, an immediate question arises about the proportion of correct answers likely to be achieved by individuals with a PCK_{SR} level ranging between that of novices and experts (e.g., master's students or teachers who have just completed their teacher education program). Even though PCK develops primarily over time as a result of practical experiences and reflection on them (van Driel & Berry, 2019), some studies have shown an increase in PCK among students during the university studies (e.g., Kleickmann et al., 2013). It is therefore to be expected that novices will also achieve a higher test score as they progress in their university studies. For the development of PCK multiple-choice instruments for pre-service science teachers, it also makes a difference what level of education they have.

4.5.2 RQ2: To what extent do experts' answers in a PCK_{SR} test vary less than novices' ones?

The (statistically insignificant) larger variability in the novices' answers is in line with previous studies which have already shown experts more likely capable to answer items consistently by transferring their knowledge to related ones (Hogan et al., 2003; Krepf et al., 2018; Wolff et al., 2015). The analysis of the qualitative data also supports this assumption. This is an expected result since novices usually do not have sufficient PCK to answer the items consistently. The novices in this study were undergraduate pre-service biology teachers and, consequently, had only few previous learning opportunities related to PCK as part of their university studies. It is to be expected that they will also become more consistent in their response behavior as their level of education increases. This could also be shown in previous studies (Kleickmann et al., 2013).

This finding is important for instrument development, because a higher variance results in a higher number of items needed to reach appropriate discrimination and internal consistency (Kline, 2015). Hence, instruments to assess experts' PCK would require fewer items than those administered among novices.

4.5.3 RQ3: To what extent do experts and novices justify their choice of response options in a PCK_{SR} test differently?

The significantly higher levels of experts' *Deep Thinking* and novices' *Surface Thinking* are in line with previous studies showing that surface characteristics influence item difficulty in test-taking by novices, whereas experts can better recognise the deep characteristics of an item (Clough & Driver 1986; Nehm & Ridgway 2011). From an assessment perspective, the influence of item surface characteristics on item difficulty is a construct-irrelevant source of variance (Messick, 1996), which, in turn, threatens

the validity of test score interpretation. Hence, the comparatively high level of *Surface Thinking* among novices may indicate limited validity of the PCK_{SR} test scores in this group, whereas experts' test scores may be validly interpreted as indicators for their PCK_{SR}. These findings highlight the need to conduct sample-specific validation studies for PCK instruments and to not conceptualise validity as a fixed, ubiquitous characteristic of a test instrument (AERA et al., 2014).

In summary, the findings thus consistently support the validity of our PCK_{SR} instrument. Nevertheless, we will continue to optimise it based on revision suggestions from the experts. We expect that this will further improve the psychometric quality.

4.5.4 Conclusions

In conclusion, this study aimed to understand and explain challenges regarding an under- or overrepresentation of PCK in the development of multiple-choice PCK instruments, which poses a challenge to the science education research community as closed-ended instruments are required to quantify relationships among PCK components as well as with other variables. Based on the expert-novice paradigm, it was investigated to what extent experts provide a higher proportion of correct responses and show less variability in their answers than novices, as well as to what extent experts and novices justify their choice of response options to an exemplary used PCK_{SR} multiple-choice instrument. It could be statistically shown that experts reach a significantly higher proportion of correct answers than novices. Furthermore, it could be shown that experts' justifications were significantly more often indicative of *Deep Thinking*, while novices' justifications were significantly more often indicative of *Surface Thinking*. However, a difference in the answers' variability between experts and novices was only found at a nominal, but not at a statistically significant level. Keeping these conclusions in mind can support the development and interpretation of PCK instruments. The results of our study show that the analysis of response processes using the expert-novice paradigm can help to understand certain challenges in the context of test construction and to optimise this process accordingly.

4.5.5 Implications

Finally, the following implications can be given: First, care should be taken to ensure that instruments address only a few, specific PCK aspects (Smith & Banilower, 2015). Second, our results suggest it may be helpful to also consider the target group of a test, even during piloting or other preliminary research. Because PCK develops primarily through practical experience and reflection on it (van Driel & Berry, 2019), instruments designed for novices (i.e., undergraduates) need to be rather simple. In contrast, instruments for experts (i.e., experienced teachers) need to be more difficult. Furthermore, target groups with larger variability among their responses require a higher number of items in order to achieve satisfactory discrimination and reliability indices. Finally, because of the novices' tendency toward *Surface Thinking*, formal task characteristics should be kept constant across items or varied systematically to avoid or control for construct-irrelevant sources of variance.

4.5.6 Limitations and Future Work

The main limitation of our study concerns the sample, both its size and its composition. On the one hand, the findings for H2 indicate that the overall sample was too small to detect relevant effects of medium size. Correspondingly, the power was only less than .30. Conversely, to achieve a satisfactory power of .80, the two groups would have had to include at least $n = 50$ participants each. However, due to the large time investment required to conduct the study and, in particular, to code the interviews, it was not possible to include more participants. On the other hand, the results for H3, precisely the implausibly high effect sizes of more than 2 and 3 standard deviations, respectively, suggest that the two groups were each too homogeneous. We expect these group differences to diminish with increasing sample size and larger heterogeneity of participants. Against this background,

our results are not generalizable in their present form. Nevertheless, our study has provided important initial evidence that may be useful in the construction of test instruments to assess PCK.

A second limitation concerns the overall generalizability of our results. Even if we assume that our findings might be useful in constructing other PCK tests, we can (of course) not guarantee that they are equally relevant to tasks other than scientific reasoning due to the topic-specificity of the PCK construct itself.

In the next step, we will evaluate the instrument based on the feedback from the experts. In addition, the validity of the test score interpretation will be examined considering sources of validity evidence.

Acknowledgments

The authors would like to thank Kerstin Landgraf for her support in data analysis and Cornelia Gerigk for English language editing.

4.6 References

- ACARA (2022). *Science. F-10 Version 9.0. Scope and sequence*.
<https://v9.australiancurriculum.edu.au/content/dam/en/curriculum/ac-version-9/downloads/science/science-scope-and-sequence-f-10-v9.docx>
- AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: Enacted and personal PCK. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 273–288). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_12
- Alonzo, A. C., Kobarg, M., & Seidel, T. (2012). Pedagogical content knowledge as reflected in teacher-student interactions: Analysis of two video cases. *Journal of Research in Science Teaching*, 49(10), 1211–1239. <https://doi.org/10.1002/tea.21055>
- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 25–38). Springer. https://doi.org/10.1007/978-1-4614-5149-5_2
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusman, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- BCTC (2019). *Professional standards for BC educators*.
https://www2.gov.bc.ca/assets/gov/education/kindergarten-to-grade-12/teach/teacher-regulation/standards-for-educators/edu_standards.pdf
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, 79, 101600. <https://doi.org/10.1016/j.learninstruc.2022.101600>

- Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., . . . Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–94). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_2
- Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 3–76). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_1
- Clough, E. E. & Driver, R. (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Science Education*, 70(4), 473–496. <https://doi.org/10.1002/sce.3730700412>
- Cobern, W. W., Schuster, D., Adams, B., Skjold, B. A., Muğaloğlu, E. Z., Bentz, A., & Sparks, K. (2014). Pedagogy of science teaching tests: Formative assessments of science teaching orientations. *International Journal of Science Education*, 36(13), 2265–2288. <https://doi.org/10.1080/09500693.2014.918672>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- de Groot, A. D. (1965). *Thought and choice in chess*. Mouton Publishers.
- Dresing, T., & Pehl, T. (2018). *Praxisbuch Interview, Transkription & Analyse: Anleitungen und Regelsysteme für qualitativ Forschende* [Interview, transcription, & analysis practice book: Instructions and rule systems for qualitative researchers].
- Ercikan, K., & Pellegrino, J. W. (2017). Validation of score meaning using examinee response processes for the next generation of assessments. In K. Ercikan & J. W. Pellegrino (Eds.), *Validation of score meaning for the next generation of assessments: The use of response processes* (pp. 1–8). Routledge.
- Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511816796.038>
- Forsythe, M. E., Criswell, B. A., Arias, A. M., Ellis, J. A., Escalada, L., Johnson, H. J., Palmeri, A. B., Riccio, J., & Parker, M. E. (2022). The framework for analyzing video in science teacher education (FAVSTE). *Journal of Science Teacher Education*, 33(6), 621–640. <https://doi.org/10.1080/1046560X.2021.1970698>
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). Routledge.
- Göhner, M., & Krell, M. (2022). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*, 52(2), 395–414. <https://doi.org/10.1007/s11165-020-09945-7>
- Großmann, L., & Krüger, D. (2023). Identifying performance levels of enacted pedagogical content knowledge in trainee biology teachers' lesson plans. In G. S. Carvalho, A. S. Afonso, & Z. Anastácio (Eds.), *Fostering Scientific Citizenship in an Uncertain World: Vol. 13. Contributions from Science Education Research* (pp. 95–116). Springer. https://doi.org/10.1007/978-3-031-32225-9_7
- Großschedl, J., Welter, V., & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402–439. <https://doi.org/10.1002/tea.21482>

- Günther, S. L., Fleige, J., Upmeyer zu Belzen, A., & Krüger, D. (2019). Using the case method to foster preservice biology teachers' content knowledge and pedagogical content knowledge related to models and modeling. *Journal of Science Teacher Education*, 30(4), 321–343.
<https://doi.org/10.1080/1046560X.2018.1560208>
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Routledge.
<https://doi.org/10.4324/9780203825945>
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hogan, T., Rabinowitz, M., & Craven, J. A. (2003). Representation in teaching: Inferences from research of expert and novice teachers. *Educational Psychologist*, 38(4), 235–247.
https://doi.org/10.1207/S15326985EP3804_3
- Hubley, A., & Zumbo, B. (2017). Response processes in the context of validity: Setting the stage. In B. Zumbo & A. Hubley (Eds.), *Understanding and investigating response processes in validation research: Vol. 69. Social Indicators Research Series* (pp. 1–12). Springer.
https://doi.org/10.1007/978-3-319-56129-5_1
- Jin, H., Shin, H., Johnson, M. E., Kim, J., & Anderson, C. W. (2015). Developing learning progression-based teacher knowledge measures. *Journal of Research in Science Teaching*, 52(9), 1269–1295. <https://doi.org/10.1002/tea.21243>
- Justi, R., & van Driel, J. (2006). The use of the interconnected model of teacher professional growth for understanding the development of science teachers' knowledge on models and modelling. *Teaching and Teacher Education*, 22(4), 437–450.
<https://doi.org/10.1016/j.tate.2005.11.011>
- Jüttner, M., Boone, 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.
<https://doi.org/10.1007/s11092-013-9157-y>
- Kane, M. T., & Mislevy, R. (2017). Validating score interpretations based on response processes. In K. Ercikan & J. W. Pellegrino (Eds.), *Validation of score meaning for the next generation of assessments: The use of response processes* (pp. 11–24). Routledge.
- Karal, I. S., & Alev, N. (2016). Development of pre-service physics teachers' pedagogical content knowledge (PCK) throughout their initial training. *Teacher Development*, 20(2), 162–180.
<https://doi.org/10.1080/13664530.2015.1124138>
- Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, 54(5), 586–614. <https://doi.org/10.1002/tea.21378>
- Khan, S., & Krell, M. (2019). Scientific reasoning competencies: A case of preservice teacher education. *Canadian Journal of Science, Mathematics and Technology Education*, 19(4), 446–464. <https://doi.org/10.1007/s42330-019-00063-9>
- Kind, V., & Chan, K. K. H. (2019). Resolving the amalgam: Connecting pedagogical content knowledge, content knowledge and pedagogical knowledge. *International Journal of Science Education*, 41(7), 964–978. <https://doi.org/10.1080/09500693.2019.1584931>
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, 64(1), 90–106.
<https://doi.org/10.1177/0022487112460398>
- Kline, P. (2015). *A handbook of test construction (psychology revivals). Introduction to psychometric design*. Routledge. <https://doi.org/10.4324/9781315695990>
- KMK (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss Beschluss vom 16.12.2004* [Educational standards in the subject biology for the intermediate school leaving. Resolution of 16.12.2004].
https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Biologie.pdf

- KMK (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung* [Common content-specific requirements of the federal states for the subject sciences and subject didactics in teacher education].
https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf
- KMK (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife* [Educational standards in the subject biology for the general higher education entrance qualification].
https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42(2), 262–271.
<https://doi.org/10.1016/j.shpsa.2010.11.034>
- König, J., Hanke, P., Glutsch, N., Jäger-Biela, D., Pohl, T., Becker-Mrotzek, M., Schabmann, A., & Waschewski, T. (2022). Teachers' professional knowledge for teaching early literacy: Conceptualization, measurement, and validation. *Educational Assessment, Evaluation and Accountability*, 34(4), 483–507. <https://doi.org/10.1007/s11092-022-09393-z>
- Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching*, 34(9), 949–968. [https://doi.org/10.1002/\(SICI\)1098-2736\(199711\)34:9%3C949::AID-TEA7%3E3.0.CO;2-U](https://doi.org/10.1002/(SICI)1098-2736(199711)34:9%3C949::AID-TEA7%3E3.0.CO;2-U)
- Krell, M. (2018). *Schwierigkeitserzeugende Aufgabenmerkmale bei Multiple-Choice-Aufgaben zur Experimentierkompetenz im Biologieunterricht: Eine Replikationsstudie* [Difficulty-generating item characteristics in multiple-choice items for experimental literacy in biology education: A replication study]. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1), 1–15.
<https://doi.org/10.1007/s40573-017-0069-0>
- Krell, M., Vorholzer, A., & Nehring, A. (2022). Scientific reasoning in science education: From global measures to fine-grained descriptions of students' competencies. *Education Sciences*, 12(2), 97. <https://doi.org/10.3390/educsci12020097>
- Krepf, M., Plöger, W., Scholl, D., & Seifert, A. (2018). Pedagogical content knowledge of experts and novices—what knowledge do they activate when analyzing science lessons? *Journal of Research in Science Teaching*, 55(1), 44–67. <https://doi.org/10.1002/tea.21410>
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>
- Kunter, M., & Trautwein, U. (2013). *Psychologie des Unterrichts* [Psychology of teaching]. Schöningh.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lenhard, W., & Lenhard, A. (2022). *Berechnung von Effektstärken* [Computation of effect sizes]. Psychometrica. <https://www.psychometrica.de/effektstaerke.html>
- Lieberei, T., Großmann, L., Welter, V.D.E., Krüger, D., Krell, M. (2023). *Entwicklung und Evaluation eines Instruments zur Erhebung fachdidaktischen Wissens angehender Biologielehrkräfte im Kompetenzbereich Erkenntnisgewinnung: PCK_SR-bio* [Development and evaluation of an instrument to assess pedagogical content knowledge about scientific reasoning: PCK_SR-bio]. Paper submitted for presentation at Internationale Tagung der Fachsektion Didaktik der Biologie (FDdB) 2023, Ludwigsburg, Germany.
- Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching*, 41(4), 370–391. <https://doi.org/10.1002/tea.20007>
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for Science Teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Kluwer Academic Publishers.
https://doi.org/10.1007/0-306-47217-1_4

- Mahler, D., Großschedl, J., & Harms, U. (2017). Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. *International Journal of Science Education*, 39(2), 213–237. <https://doi.org/10.1080/09500693.2016.1276641>
- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34(4), 207–218. https://doi.org/10.1207/s15326985ep3404_2
- Mathesius, S., Hartmann, S., Upmeyer zu Belzen, A., & Krüger, D. (2016). Scientific reasoning as an aspect of pre-service biology teacher education: Assessing competencies using a paper-pencil-test. In T. Tal & A. Yarden (Eds.), *The future of biology education research* (pp. 93–110).
- Messick, S. (1996). Validity of performance assessments. In G. W. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp. 1–18). National Center for Education Statistics.
- MINEDUC (2012). *Estándares orientadores para carreras de pedagogía en Educación Media* [Guiding standards for secondary education pedagogy careers]. https://www.cpeip.cl/wp-content/uploads/2018/09/Est%C3%A1ndares_Media.pdf
- MINEDUC (2015). *Bases Curriculares: 7° básico a 2° medio* [Curricular bases: 7th basic to 2nd middle grade]. https://www.curriculumnacional.cl/614/articles-34949_Bases.pdf
- MINEDUC (2019). *Bases Curriculares: 3° y 4° medio* [Curricular bases: 3rd and 4th grades]. https://www.curriculumnacional.cl/614/articles-91414_bases.pdf
- Ministry of Education (2008a). *The Ontario curriculum: Grades 11 and 12. Science*. https://www.edu.gov.on.ca/eng/curriculum/secondary/2009science11_12.pdf
- Ministry of Education (2008b). *The Ontario curriculum: Grades 9 and 10. Science*. https://www.edu.gov.on.ca/eng/curriculum/secondary/science910_2008.pdf
- Nehm, R. H. & Ridgway, J. (2011). What do experts and novices “see” in evolutionary problems? *Evolution: Education and Outreach*, 4(4), 666–679. <https://doi.org/10.1007/s12052-011-0369-7>
- Neumann, K., Kind, V., & Harms, U. (2019). Probing the amalgam: The relationship between science teachers' content, pedagogical and pedagogical content knowledge. *International Journal of Science Education*, 41(7), 847–861. <https://doi.org/10.1080/09500693.2018.1497217>
- NSW Education Standards Authority (2018). *Australian professional standards for teachers: Teacher accreditation*. <https://educationstandards.nsw.edu.au/wps/wcm/connect/9ba4a706-221f-413c-843b-d5f390c2109f/australian-professional-standards-teachers.pdf?MOD=AJPERES>
- O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 1–13. <https://doi.org/10.1177/1609406919899220>
- Opfer, J. E., Nehm, R. H. & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. *Journal of Research in Science Teaching*, 49(6), 744–777. <https://doi.org/10.1002/tea.21028>
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265–279. <https://doi.org/10.1016/j.tsc.2013.07.006>
- 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. <https://doi.org/10.1002/tea.21022>
- Park, S., & Oliver, J. S. (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. <https://doi.org/10.1007/s11165-007-9049-6>
- Park, S., Suh, J., & Seo, K. (2018). Development and validation of measures of secondary science teachers' PCK for teaching photosynthesis. *Research in Science Education*, 48(3), 549–573. <https://doi.org/10.1007/s11165-016-9578-y>
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3–13. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>

- Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020–1049. <https://doi.org/10.3102/0002831213477680>
- Sannert, R., & Krell, M. (2023). A professional development program to foster science teachers' professional competence, enhance classroom practice, and improve student outcomes related to scientific reasoning. *Progress in Science Education*, 6(2), 47–62. <https://doi.org/10.25321/PRISE.2023.1401>
- Schmitt, A. K. (2016). *Entwicklung und Evaluation einer Chemielehrerfortbildung zum Kompetenzbereich Erkenntnisgewinnung* [Development and evaluation of a chemistry teacher training program on scientific reasoning]. Logos Verlag Berlin.
- Schnotz, W. & Baadte, C. (2015). Surface and deep structures in graphics comprehension. *Memory & Cognition*, 43(4), 605–618. <https://doi.org/10.3758/s13421-014-0490-2>
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE.
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten [Variants of qualitative content analysis: A guide through the thicket of terminology]. *Forum Qualitative Sozialforschung*, 15(1), 18.
- Schuster, D., Cobern, W., Applegate, B., Schwartz, R., Vellom, P., & Undreiu, A. (2006). Assessing pedagogical content knowledge of inquiry science teaching: Developing an assessment instrument to support the undergraduate preparation of elementary teachers to teach science as inquiry. In D. Deeds & B. Callen (Eds.), *Proceedings of National STEM Assessment Conference* (pp. 247-265).
- Schweizer, K., & DiStefano, C. (2016). *Principles and Methods of Test Construction. Standards and Recent Advances*. Hogrefe Publishing.
- Seidel, T., & Reiss, K. (2014). *Lerngelegenheiten im Unterricht* [Learning opportunities in the classroom]. In T. Seidel & A. Krapp (Eds.), *Pädagogische Psychologie* (pp. 253–275). Julius Beltz.
- She, J., & Chan, K. K. H. (2023). Situated and dynamic versus declarative and static forms of pedagogical content knowledge: An evaluation of the differences in test reactions and performance. *Journal of Research in Science Teaching*, 60(3), 568-607. <https://doi.org/10.1002/tea.21810>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Smith, P. S., & Banilower, E. R. (2015). Assessing PCK: A new application of the uncertainty principle. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (pp. 88–103). Routledge.
- Sorge, S., Kröger, J., Petersen, S., & Neumann, K. (2019). Structure and development of pre-service physics teachers' professional knowledge. *International Journal of Science Education*, 41(7), 862–889. <https://doi.org/10.1080/09500693.2017.1346326>
- Sorge, S., Stender, A., & Neumann, K. (2019). The development of science teachers' professional competence. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 149–164). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_6
- Tepner, O., & Dollny, S. (2014). *Entwicklung eines Testverfahrens zur Analyse fachdidaktischen Wissens* [Development of a test procedure for the analysis of pedagogical content knowledge]. In D. Krüger, I. Parchmann, & H. Schecker (Eds.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (pp. 311–323). Springer Spektrum. https://doi.org/10.1007/978-3-642-37827-0_25

- Upmeier zu Belzen, A., van Driel, J., & Krüger, D. (2019). Introducing a framework for modeling competence. In A. Upmeier zu Belzen, D. Krüger, & J. van Driel (Eds.), *Towards a competence-based view on models and modeling in science education: Vol. 12. Models and modeling in science education* (pp. 3–19). Springer. https://doi.org/10.1007/978-3-030-30255-9_1
- van Driel, J., & Berry, A. (2019). Pedagogical content knowledge in preservice teacher education. In M. A. Peters (Ed.), *Encyclopedia of teacher education* (pp. 1-6). Springer Singapore. https://doi.org/10.1007/978-981-13-1179-6_176-1
- Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen* [Intercoder agreement and intercoder reliability]. Hogrefe Verlag für Psychologie.
- Wolff, C. E., van den Bogert, N., Jarodzka, H., & Boshuizen, H. P. A. (2015). Keeping an eye on learning: Differences between expert and novice teachers' representations of classroom management events. *Journal of Teacher Education*, 66(1), 68–85. <https://doi.org/10.1177/0022487114549810>

5 Studie 2: Considering Multiple Sources of Validity Evidence Can Help to Address Challenges in the Development of Pedagogical Content Knowledge (PCK) Multiple-Choice Items²⁷

Abstract

The use of multiple-choice (MC) instruments to assess pedagogical content knowledge (PCK) has advantages regarding test economy and objectivity, but it also poses challenges, for example, in terms of adequately capturing the intended construct. To help address these challenges, we report on the development and evaluation of a new instrument to assess science teachers' PCK of scientific reasoning in biology contexts (PCK_{SR-bio}); for the evaluation, we considered multiple sources of validity evidence. First, 12 MC items were developed to assess crucial PCK components for three scientific reasoning skills. Subsequently, the correlation of corresponding content knowledge (CK) with the PCK_{SR-bio} score was tested with 67 master's students. In addition, the instrument was used in a cross-sectional study with 164 students ($n = 29$ bachelor, $n = 114$ master, $n = 21$ school-based preservice teachers), and the internal consistency as well as the correlation of the test score with the educational level was determined. An analysis of the response processes of 10 bachelor's students showed that they more often referred to PCK when selecting an attractor and more often (rather intuitively) to other knowledge when selecting a distractor. In the cross-sectional study, the internal consistency was relatively low but increased with higher educational level. A correlation was found between the test score and CK but not between the test score and the educational level. Our results show that considering multiple sources of validity evidence can help to address common challenges in developing MC-PCK instruments. The results and limitations are discussed, and recommendations are made for the development of MC instruments to assess PCK in general.

Keywords: pedagogical content knowledge (PCK), preservice science teachers, multiple-choice instrument, test development, validation, scientific reasoning

²⁷ Lieberei, T., Großmann, L., Welter, V. D. E., Krüger, D. & Krell, M. (angenommen). Considering multiple sources of validity evidence can help to address challenges in the development of pedagogical content knowledge (PCK) multiple-choice items. *Research in Science Education*.

5.1 Introduction

Science teachers' professional competence and particularly their professional knowledge has a crucial influence on the effectiveness of their teaching (Blömeke et al., 2022). According to Shulman (1986), a distinction is often made between different domains of professional knowledge such as pedagogical knowledge (PK; knowledge about generic principles of education), content knowledge (CK; knowledge about the specific subject matter being addressed in the teaching), and pedagogical content knowledge (PCK; knowledge about learners' conceptual ideas about a particular topic and instructional strategies for teaching it to a particular group of students).

Various studies have shown that especially a high level of teachers' PCK positively affects the quality of their classroom practice and of student learning (Kunter et al., 2013). To ensure that science teachers develop this central knowledge, preservice science teachers should acquire basic PCK during initial teacher education, which is why it is anchored in the standards for teacher education in several countries around the globe (e.g., Germany: KMK, 2024). To evaluate the quality of these teacher education programs, it is necessary to evaluate the development of preservice teachers' PCK during initial teacher education, for example, with paper-pencil tests that can be used for larger groups in particular. However, the development of those instruments comes with challenges, which have been discussed in the literature, regarding cognitive operations and task formats, the construct representation, as well as psychometric properties (Lieberei et al., 2023; Smith & Banilower, 2015; Rosenkränzer et al., 2017).

PCK instruments can have different task formats, with a main distinction being made between open-ended and multiple-choice (MC) instruments. While open-ended formats can provide more insight into the scope and depth of respondents' knowledge (Martinez, 1999), MC instruments are particularly advantageous regarding test economy and objectivity (Haladyna, 2004). Two-tier MC instruments or instruments with mixed task formats combine these advantages but require a high level of interpretation for data analysis and a longer test duration. However, regarding MC-PCK tests, care must be taken in the development and evaluation process to ensure that the target construct is adequately represented (Smith & Banilower, 2015). Furthermore, instruments for assessing complex and application-related constructs often show relatively low reliability compared to conventional scales or questionnaires (Taber, 2018), which is also the case for some MC-PCK instruments (e.g., Förtsch et al., 2016). To deal with these challenges, the Standards for Educational and Psychological Testing (AERA et al., 2014) provide guidance on which validity aspects can and should be considered in the context of test development. In our study, on the basis of these standards, we aimed to address the challenges regarding the development of MC-PCK instruments that are discussed in the literature.

In order to do this, we developed and evaluated an instrument to assess PCK about scientific reasoning (PCK_{SR}) in biology contexts based on the two crucial PCK components "knowledge of instructional strategies for teaching science" (KISR) and "knowledge of students' understanding" (KSU), as well as their connections: the PCK_{SR}-bio. PCK_{SR} is a major part of science education (Khan & Krell, 2019), for which no instruments exist yet. In this paper, we take up on the existing discussion on the use of MC instruments to assess PCK by presenting a structural approach, based on the sources of validity evidence (AERA et al., 2014), that was used to test the development of and to evaluate the PCK_{SR}-bio. By doing this, we aimed to investigate how considering different sources of validity evidence can help to address common challenges in developing MC-PCK instruments in general.

5.2 Theoretical Background

The following sections describe PCK and its assessment, as well as different sources of validity evidence.

5.2.1 Pedagogical Content Knowledge and its Assessment

Since its original definition, the construct of PCK has been further differentiated by several researchers. A more recent definition describes PCK as “the knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes” (Gess-Newsome, 2015, p. 36). Shulman (1986) already described the two PCK components KSU und KISR, which also play, alongside further components, a crucial role in other models, for example in the model by Magnusson et al. (1999) or in the pentagon-model (Park & Chen, 2012). While KISR includes knowledge about subject-specific and topic-specific instructional strategies, KSU describes knowledge about students’ conceptions, learning difficulties, and prior knowledge. These and other components are separable but correlated parts of PCK, which means that, for example, the more teachers know about their students’ learning difficulties, the better they can teach the relevant topic and vice versa (van Driel et al., 2023). PCK can also be conceptualized as interconnected layers of knowledge, as, for example, described in the refined consensus model. In this model, different (but related) layers of PCK are distinguished (Carlson et al., 2019): While enacted PCK (ePCK) describes the dynamic PCK of an individual teacher shown in a specific teaching situation, personal PCK (pPCK) describes the comprehensive PCK available to an individual teacher, which can be more static. The pPCK of multiple professionals shared in groups is described as collective PCK (cPCK) and also includes canonical cPCK (i.e., the findings of science education research), which can serve as the basis for a science teacher’s individual pPCK (Carlson et al., 2019).

It is assumed that CK is a prerequisite for developing PCK (Friedrichsen et al., 2009). In line with this assumption, correlations between CK and PCK have been shown empirically (e.g., Baumert et al., 2010). The development of ePCK is mainly attributable to practical teaching experience and reflection on that experience (van Driel & Berry, 2019); previous findings have suggested that the development of pPCK only benefits from practical experience if teachers possess a certain level of reflection skills (Kulgemeyer et al., 2021). Nevertheless, a basic level of pPCK should already be reached during initial teacher education. It has been empirically shown that preservice science teachers in the master’s program have a higher level of PCK than bachelor’s students (e.g., Kleickmann et al., 2013). However, because PCK is topic-specific (Gess-Newsome, 2015), it can be assumed that PCK may develop differently for different topics. Therefore, when interpreting empirical findings on PCK and its development, the topic area covered by the instrument has to be considered.

Science teachers not only need PCK about science topics (e.g., PCK about evolution; Becerra et al., 2022); they also need discipline-specific PCK (Carlson et al., 2019; Kenyon et al., 2011). This includes PCK about epistemic elements of science such as PCK about scientific inquiry (Davis & Krajcik, 2005) or PCK about modeling (Justi & van Driel, 2005). It is therefore necessary for science teachers to have knowledge about scientific practices and how to teach them, which is why we argue that they need PCK about scientific reasoning (PCK_{SR}) in order to successfully plan and conduct science lessons and implement different epistemic elements of science in the science classroom (Osborne, 2014).

Instruments should therefore specifically address PCK about different science topics or PCK about scientific practices, for example, PCK_{SR}. Scientific reasoning, as a major goal of science education, is described as a complex construct involving numerous practices, but it does not yet have a standard definition (Krell et al., 2022). However, the skills “planning and conducting investigations”, “generating hypotheses”, and “using scientific models” were, among others, identified as highly relevant by Rönnebeck et al. (2016). Regarding these skills, PCK_{SR} describes the knowledge of, reasoning behind,

and planning required to teach scientific reasoning in different educational settings to specific students to improve their scientific reasoning competencies (see Gess-Newsome, 2015). PCK_{SR} , therefore, includes knowledge about students' conceptions and about instructional strategies related to scientific reasoning, which a science teacher should know in order to be able to appropriately teach scientific reasoning (e.g., how to foster students' procedural and epistemic knowledge with experiments or how to use models as epistemic tools instead of as visualizations). This means that PCK_{SR} describes PCK about scientific practices (Osborne, 2014) that are not specifically related to a particular science topic or discipline.

Although scientific reasoning is a major goal of science education, no instruments exist to specifically assess PCK_{SR} . Therefore, appropriate instruments are needed. However, the development of PCK instruments holds challenges in terms of cognitive operations and task format, the construct representation, as well as psychometric properties.

5.2.1.1 Cognitive Operations and Task Formats

One challenge in developing MC-PCK instruments is the selection of an appropriate task format and the cognitive operations that come with it; this selection depends on the layer of PCK considered. While methods for performance assessment need to be developed to assess dynamic and action-related teaching (i.e., ePCK), this study focuses on paper-pencil instruments that can be used to assess the comprehensive and nonsituational PCK a teacher possesses (i.e., pPCK; Chan & Hume, 2019) and on an evaluation of it on the basis of canonical cPCK. Paper-pencil PCK instruments can have widely different task formats. On the one hand, open-ended questions (e.g., CoRe by Loughran et al., 2004) can capture a broad range of cognitive operations regarding what the respondents think (Martinez, 1999), but they require a high level of interpretation for data analysis. Also, these instruments often come with a long test duration, and data analysis is time consuming as well. On the other hand, MC instruments (e.g., POSITT by Schuster et al., 2006) have advantages in terms of objectivity and test economy but come with challenges as well, for example, regarding the phrasing of clear and unambiguous items (Haladyna, 2004). There are also instruments with mixed task formats (e.g., PCK-IBI by Großschedl et al., 2019; unnamed instrument by Jüttner et al., 2013), which, however, are also associated with a long test duration, depending on the proportion of open-ended questions (e.g., about 50 minutes for 24 items in Jüttner et al., 2013). To evaluate the validity of the test score interpretation of these instruments, Jüttner et al. (2013) as well as Großschedl et al. (2019) considered different types of validity. For example, Jüttner et al. (2013) evaluated construct validity by using Wright Maps and they suggest that beginning teachers could be compared with experienced teachers (criterion validity) and think-aloud interviews could be performed (content validity). In contrast, Großschedl et al. (2019) evaluated construct validity by performing factor analyses, content validity by carrying out expert ratings, and criterion validity by examining the correlation with external variables. Approaches like this are also recommended by AERA et al. (2014), although they emphasize the concept of validity as a coherent construct based on different forms of evidence rather than as distinct types. As another alternative, two-tier MC instruments provide information about what respondents think, but they forgo the advantages of MC instruments regarding test economy and objectivity. Currently, the possible digitization of assessment methods comes with new challenges as it can impact the authenticity and, therefore, the validity of the assessment method. It is especially important that instruments assessing action-related teaching provide action-related contexts; this is only possible to a limited extent with digital instruments (Kulgemeyer et al., 2023).

5.2.1.2 Construct Representation

Another challenge is to adequately cover the construct with the items of the instrument, especially because PCK is a complex construct with diverse components (see Chan & Hume, 2019). As mentioned in Lieberei et al. (2023), on the one hand, there is the risk of underrepresenting the construct if the instrument does not cover all aspects of PCK or connections between the components considered. A clear focus and an adequate and unambiguous definition of PCK are needed; therefore, it is recommended to focus on narrowly defined PCK aspects when developing MC instruments (Smith & Banilower, 2015). On the other hand, an overrepresentation of the construct must also be avoided because it could lead to CK or beliefs about teaching being assessed instead of PCK (Smith & Banilower, 2015), which could lead to an assessment that is “too broad” (Messick, 1996, p. 5).

5.2.1.3 Psychometric Properties

MC instruments for assessing complex and application-related constructs often show relatively low internal consistency compared to conventional scales or questionnaires (Taber, 2018). Indeed, some already published MC–PCK instruments show a relatively low internal consistency (e.g., $\alpha = .53$ in Förtsch et al., 2016), while other instruments that consist at least in part of MC items have a relatively high internal consistency (e.g., $\alpha = .92$ in Großschedl et al., 2019). Furthermore, for MC–PCK instruments, it has been shown empirically that a higher educational level of respondents is associated with higher internal consistency (e.g., Kleickmann et al., 2013). However, there is no clear and established limit to when the internal consistency of instruments can be considered acceptable (Taber, 2018).

Considering these challenges in developing MC–PCK instruments, some researchers argue that PCK, due to its complexity, cannot be validly captured by MC items (e.g., Smith & Banilower, 2015; Rosenkränzer et al., 2017). In contrast, we assume that assessing PCK (operationalized as selected components, e.g., KISR and KSU, for a specific topic or discipline) with MC instruments is possible, as long as different sources of validity evidence (AERA et al., 2014) are considered in the instrument development process.

5.2.2 Different Sources of Validity Evidence

Obtaining validity evidence involves evaluating the plausibility of the instrument’s claims based on the test results (Kane, 2013) and “refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of test” (AERA et al., 2014, p. 11). The validation process involves obtaining empirical and theoretical evidence for a valid test score interpretation. For this procedure, the Standards for Educational and Psychological Testing (AERA et al., 2014) recommend five sources of validity evidence, which are described in the following sections.

Evidence based on test content describes the fit of the test content to the test items (AERA et al., 2014). It therefore forms the foundation of the validity argument (Sireci & Faulkner-Bond, 2014) and is characterized by two key aspects (McCoach et al., 2013b): the conceptual definition, which includes a substantive description of the construct (i.e., its attributes and defining features), and the operational definition, which outlines the characteristics of the instrument (i.e., how the construct is measured and which items and scales are used).

Evidence based on response processes describes the mental processes of the participants during the test (AERA et al., 2014) and encompass cognitive, affective, and motivational mechanisms influencing how participants interact with test items (Hubley & Zumbo, 2017). These processes can be either intentional and easily verbalizable or less conscious, automatic reactions to test items. Response processes can be captured in think-aloud settings or interviews (Ercikan & Pellegrino, 2017).

Evidence based on internal structure describes the fit of the response behavior to the theoretical construct of the instrument (AERA et al., 2014). Internal consistency is a central aspect of evidence based on internal structure (Rios & Wells, 2014) and is often reported as Cronbach's alpha. While some researchers define a clear lower limit for acceptable internal consistency, some authors suggest that instruments should not be discarded solely because they do not meet a certain level (Schmitt, 1996; Taber, 2018). Internal consistency also depends on the test takers' knowledge (Krepf et al., 2018), which means that higher expertise is often associated with more consistent response behavior.

Evidence based on relations to other variables describes the relationship of the test score of the considered instrument to the test score of instruments that assess theoretically related variables (AERA et al., 2014). A distinction can be made between convergent (positive correlations with variables that assess a closely related construct) and discriminant (a weak or no correlation with theoretically unrelated constructs) evidence (McCoach et al., 2013a).

Evidence based on the consequences of testing describes the identification of intended and unintended consequences of test use (AERA et al., 2014). This validity evidence will not be considered further in this paper because no suitable consequences for validity investigations can be drawn based on the test results.

In the following, we will outline the extent to which the consideration of multiple sources of validity evidence can help address the challenges identified in developing MC instruments for the assessment of PCK in general, using the example of the PCK_{SR}-bio.

5.2.3 Research Question

The following research question is addressed in this paper:

To what extent can considering different sources of validity evidence help to address common challenges in developing MC–PCK instruments regarding cognitive operations and task formats, the construct representation, as well as psychometric properties?

In order to answer this research question, the PCK_{SR}-bio was developed and evaluated systematically in different studies that addressed the sources of validity evidence mentioned above (Table 5.1).

Table 5.1: Overview of the studies conducted, addressing the four sources of validity evidence

Source of Validity Evidence	Studies Conducted
Evidence Based on Test Content	Systematic development of the instrument on a theoretical basis. First pilot study with 67 preservice biology teachers. Revision of the distractors based on 14 preservice biology teachers' conceptions
Evidence Based on Response Processes	Interview study with 10 preservice biology teachers about their justifications for selecting a response option
Evidence Based on Internal Structure	Cross-sectional data collection with 164 preservice biology teachers (bachelor, master, in-service training) to test the variability of test scores
Evidence Based on Relations to Other Variables	Examining the correlation between preservice biology teachers' test scores on the PCK _{SR} -bio and their CK ($N = 60$) as well as their semester of study ($N = 164$) in two data collections. Interventional study to check whether the PCK _{SR} -bio is sensitive to change

The following sections provide an overview of the initial development of the PCK_{SR}-bio, describing validity evidence based on test content as well as the data collection and statistical methods applied in the different studies that were conducted to investigate the validity of the test score interpretation of the PCK_{SR}-bio based on the three remaining sources of validity evidence. Our studies were conducted in Germany and reviewed and approved by the local ethics committee (ID 2021_LI56, ID 2022_23_LI, ID 2023_18_LI). All statistical analyses presented were conducted with IBM SPSS Statistics 28.

5.3 Item Development (Evidence Based on Test Content)

This section describes the item development process of the PCK_{SR}-bio, which covers validity evidence based on test content.

5.3.1 Systematic Development

The PCK_{SR}-bio was developed and validated in several steps, based on the definition of PCK_{SR} (see Section 5.2.1). Because we aimed to assess the comprehensive PCK_{SR} a teacher possesses independently of a specific teaching situation, we developed an MC-PCK instrument in order to assess this discipline-specific part of teachers' pPCK. As it is recommended to focus on narrowly defined PCK components when developing such instruments (Smith & Banilower, 2015), the items of the PCK_{SR}-bio refer only to the crucial PCK components KISR (Items #1, #2, and #3) and KSU (Items #4, #5, and #6), as well as their connections, KISR→KSU and KSU→KISR. The connection KISR→KSU (Items #7, #8, and #9) thereby reflects the impact of instruction on students' conceptions. The connection KSU→KISR (Items #10, #11, and #12) refers to designing instruction that takes students' conceptions into consideration. Most other PCK instruments also refer to these PCK components (Chan & Hume, 2019). Because a high level of PCK is understood as the ability to connect different PCK components (instead of having isolated knowledge about the separate components; e.g., Schneider & Plasman, 2011), we explicitly targeted the connections to avoid construct underrepresentation. As there is no standard definition of scientific reasoning yet (Krell et al., 2022), we decided to focus on three highly relevant scientific reasoning skills: planning and conducting investigations, generating hypotheses, and using scientific models (Rönnebeck et al., 2016). Together with the PCK components and their connections, these three skills form the dimensions of the 4x3 matrix that was used as the framework of the PCK_{SR}-bio (Table 5.2). One item was assigned to each intersection, which led to a total of 12 items. The PCK_{SR}-bio addresses biology-related contexts as teaching situations in the item stem (Figure 5.1) but can also be answered by preservice teachers of other science subjects. Because the items address scientific practices, biology CK in general is not necessarily needed to answer them (Lieberei et al., 2024).

Table 5.2: Theoretical framework of the MC instrument developed to assess PCK_{SR}

	Planning and Conducting Investigations	Generating Hypotheses	Using Scientific Models
KISR	Item #1	Item #2	Item #3
KSU	Item #4	Item #5	Item #6
KISR→KSU	Item #7	Item #8	Item #9
KSU→KISR	Item #10	Item #11	Item #12

5.3.2 Initial Item Development

One of the major challenges in developing MC–PCK instruments is the lack of empirical evidence on the effectiveness of certain instructional strategies in addressing certain students’ conceptions (Smith & Banilower, 2015). In the development process, we referred to theoretical and empirical literature to ensure that the items aligned with recent findings in science education. For instance, Item #12 (Figure 5.1) describes a student’s conception of models and modeling. The participants are expected to select the instructional strategy that is most appropriate to target this conception in order to extend it through an understanding that is as scientifically and methodologically appropriate as possible. It has repeatedly been shown that students, preservice, and in-service teachers often regard models as visualizations of scientific phenomena (Upmeier zu Belzen et al., 2019). Beyond this view, models are epistemic tools in science (Göhner & Krell, 2022; Knuuttila, 2011). Hence, Answer Option 3 is the attractor here as this option targets the role of models in gaining new scientific knowledge (in contrast to Answer Options 1 and 2, where models are portrayed as mere visualizations, in line with the most common conception of students; Upmeier zu Belzen, 2019). To answer this item correctly, participants therefore need knowledge of students’ conceptions of models and modeling and of appropriate instructional strategies for dealing with them (Justi & van Driel, 2005).

Stem: A biology teacher has worked with his students on the structure of DNA using models. In a subsequent lesson, he asks his students what models are and what function they have in biology. One student answers: “Models are visualisations of something, for example, DNA. We can use them to learn how DNA is built.”

Prompt: Please tick the ways in which the teacher should take this idea about models and extend it through an understanding that is as scientifically and methodologically appropriate as possible.

Options:

He should...

- (1) ... ask his students to make their own DNA models to highlight the potential of models as representations.
- (2) ... practise “model critique” with his students by working out similarities and differences between DNA models and DNA.
- (3) ... reflect on models by elaborating on the importance of DNA models in the discovery of DNA structure.

Figure 5.1: Sample item (Item #12)

To ensure that the items comprised two distractors that were definitely wrong and one attractor that was definitely correct, we discussed all items with peer researchers in science education. In this step-wise process and on the basis of the colleagues’ feedback, we eliminated answer options that targeted beliefs about instruction. Moreover, we ensured that all answer options were more or less equal in length and that trigger words appeared in all three answer options to avoid making the attractor most attractive due to linguistic reasons (Haladyna, 2004). For example, in our initial draft of Item #12, the acronym DNA was only used in Answer Options 2 and 3, which might have impacted the probability of participants choosing Answer Option 1. Finally, we had an initial set of 12 items; a complete version of the instrument with all items can be found in Section 12.2.

5.3.3 Revision of the Items

In a first pilot study in April 2021, a total of 67 preservice biology teachers (in the master's program) participated. We analyzed the distractors based on the participants' answers to the 12 items. For that purpose, we calculated the proportion of participants that chose the attractor or one of the two distractors. In cases where one (or even both) of the distractors was rarely chosen, we were requested to modify the distractors. In May 2021, we provided all 12 items to 14 preservice biology teachers in an open-ended format for that purpose. In doing so, we were able to get authentic insights into the target group's thinking and were able to collect distractors that were probably more authentic and appropriate than those that the research team had initially developed.

After having identified which distractors performed well and which were rarely chosen, we analyzed the participants' answers to the open-ended questions in an iterative process and aimed to identify answer options that could be used as new distractors to replace the low-performing distractors. We compared the answers to each other and to the existing distractors and finally replaced some distractors based on the answers given by the 14 preservice biology teachers.

On the basis of this procedure, we assume that the items adequately address the PCK_{SR} construct.

5.4 Evidence Based on Response Processes

This section describes the study that was conducted between January 2022 and March 2022 to ensure validity evidence based on response processes.

5.4.1 Methods

In order to investigate the response behavior of the target group (preservice science teachers), we reanalyzed data from an already published interview study in which the PCK_{SR}-bio was first tested (Lieberei et al., 2023); we placed a specific focus on validity evidence.

5.4.1.1 Sample

A total of 10 undergraduate preservice biology teachers, who were invited via e-mail, participated in this study. Participation was voluntary, and participants received financial compensation for their participation.

5.4.1.2 Data Collection

The participants answered all MC items of the PCK_{SR}-bio and, after each item, justified their selection of a response option as well as explained why they did not choose another option.

5.4.1.3 Data Analysis

Each interview lasted approximately one hour and was recorded and transcribed word for word. Subsequently, the justifications were assigned to the respective response options of the MC items. Using a newly developed category system, the participants' justifications were examined in a structural qualitative content analysis (Schreier, 2012) in terms of whether they tended to justify their selection of a response option based on their PCK, on surface characteristics of the items, or on intuition. Justifications were coded as "PCK argumentation" if one of the PCK components, KSU or KISR, or both were recognizable. Justifications were coded as "non-PCK argumentation" if the selection of a response option was not justified based on PCK but, instead, on surface characteristics of the items or on intuition. The category system was inductively extended in these cases. The category system used for the reanalysis of the data in this study is shown in Table 5.3. The full category system can be found in Lieberei et al. (2023).

Table 5.3: Category system for analyzing the justifications of the selection of response options

Category		Description	Empirical Evidence
PCK argumentation	KSU	Responses relating to students' (mis)conceptions and prior knowledge	<i>However, I would now tend towards answer option three, [...] because this addresses exactly the point that the student does not perceive.</i>
	KISR	Responses relating to subject-specific and topic-specific strategies	<i>She states that models are pure visualizations of reality but does not consider that insights can also be gained through such model considerations.</i>
Non-PCK argumentation		Responses relating to surface characteristics of the items or intuition	<i>But I think that the model criticism is just a bit more plausible and perhaps sticks a bit more. I don't know why at the moment.</i>

Note. The empirical evidence refers to Item #12 (see Figure 5.1).

To ensure intercoder agreement, following the recommendations of Wirtz and Caspar (2002), 20% of the transcripts were coded independently by a trained student assistant. To increase the probability of representativeness, 2 interviews were randomly selected (O'Connor & Joffe, 2020). The overall intercoder agreement was $\kappa = .87$ for "PCK argumentation" and $\kappa = .72$ for "non-PCK argumentation". Disagreements were resolved by discussion.

To analyze the data, a quantitative comparison of the frequency of justifications was made based on "PCK argumentation" and "non-PCK argumentation". Sensitivity (the proportion of justifications based on "PCK argumentation" when choosing the attractor) and specificity (the proportion of justifications based on "non-PCK argumentation" when choosing one of the two distractors) were examined. Furthermore, justifications based on "PCK argumentation" for choosing the response option, as well as for not choosing either of the other two options, were compared for items addressing the same PCK components (e.g., Items #1, #2, and #3 for KISR; Table 5.2).

As the items were expected to provide measures of preservice biology teachers' PCK (or KSU, KISR, or their connections), we expected to obtain a significantly higher proportion of answers that indicated PCK-related reasoning rather than non-PCK-related reasoning; this could then be interpreted as validity evidence based on response processes.

5.4.2 Results

The analysis of the 10 interviews revealed that when the participants chose the attractor, the justification was more often based on "PCK argumentation" than on "non-PCK argumentation" (46 vs. 15 justifications; sensitivity: .75). Furthermore, the selection of a distractor was justified based on "PCK argumentation" only one time; all remaining justifications were based on "non-PCK argumentation" (56 justifications; specificity: .98).

Regarding the participants' justification based on "PCK argumentation" for choosing a response option as well as for not choosing either of the other two options, it appeared that the participants most often argued with the PCK component that corresponded to the PCK component addressed in the respective item (Table 5.4).

Table 5.4: Total number of justifications based on “PCK argumentation” for the respective items

Items related to ...	KISR argumentation only	KSU argumentation only	Both KISR and KSU argumentation
... KISR	15	0	2
... KSU	0	18	1
... KISR→KSU	4	0	14
... KSU→KISR	5	0	14

Thus, the results supported our assumption that the selection of a response option was more frequently made based on “PCK argumentation”, while the selection of a distractor was almost exclusively made based on “non-PCK argumentation”. Furthermore, as expected, the participants most often argued with the PCK component that corresponded to the PCK component addressed in the respective item. This supports the validity of the test score interpretation.

5.5 Evidence Based on Internal Structure

This section presents the study that was conducted to ensure validity evidence based on internal structure.

5.5.1 Methods

The variability of test scores on the PCK_{SR}-bio was empirically tested in a cross-sectional data collection to find validity evidence based on internal structure.

5.5.1.1 Sample

A total of 164 preservice biology teachers participated in this survey, of which 29 were in the bachelor’s program, 114 were in the master’s program, and 21 were in in-service training. In Germany, this is the last part of teacher education after the master’s degree, when preservice teachers already teach at schools, guided by mentors, to develop more practical knowledge (Neumann et al., 2017). Notably, the bachelor’s and the master’s programs are concurrent teacher education programs, including both discipline-specific and pedagogical studies as well as a small number of placements in schools lasting a few weeks.

5.5.1.2 Data Collection

The survey took place between July 2022 and July 2023 at three different universities in Germany. The participants answered the items of the PCK_{SR}-bio and provided demographic information such as their semester of study.

5.5.1.3 Data Analysis

The internal consistency was subsequently calculated using the Kuder-Richardson 20 formula to test the variability of test scores. As PCK_{SR} is a complex construct and we covered it with 12 items, we expected to find a comparably low internal consistency. At the same time, we assumed that the internal consistency would increase as the level of experience rose because we assumed that the participants would become more consistent in their responses (Kleickmann et al., 2013; Krepf et al., 2018).

5.5.2 Results

The analysis of the test scores of the 164 participants revealed an internal consistency of .44 for the entire sample (Table 5.5). However, a closer look at the subgroups in the different stages of teacher education revealed that a higher level of education was also associated with higher internal consistency (.38 for bachelor’s students, .44 for master’s students, .58 for participants in in-service

training). At the same time, the average test score remained the same across these subgroups, while its variance increased slightly but did not differ significantly ($p = .989$; Levene's test).

Table 5.5: Overview of the internal consistency and the test scores of the groups in different stages of teacher education

Group	N	Internal Consistency	Test Score	
			M	SD
Bachelor	29	.38 ($n = 26$)	.50	.17
Master	114	.44 ($n = 109$)	.50	.19
In-Service Training	21	.58 ($n = 18$)	.50	.20
Total	164	.44 ($n = 153$)	.50	.19

Note. The proportion of correct answers was calculated to also include participants who did not answer all items. For internal consistency, only participants who answered all items were included.

Thus, the results support our assumption that the instrument has a comparatively low internal consistency, which, however, increases with a higher educational level of the participants.

5.6 Evidence Based on Relations to Other Variables

This section presents the studies that were conducted to ensure validity evidence based on relations to other variables.

5.6.1 Methods

The CK of the preservice biology teachers, as well as their semester of study, were considered in two data collections to find validity evidence based on relations to other variables. In addition, an interventional study with two groups of preservice biology teachers in the master's program was conducted to test whether the instrument is sensitive to change in PCK_{SR} .

5.6.1.1 Sample

To investigate the relationship between PCK_{SR} and CK, 60 preservice biology teachers in the 2nd semester of their master's program were surveyed. In order to check the relation between the participants' PCK_{SR} -bio test score and their semester of study, the participants in the data collection reported in Section 5.5 (i.e., $N = 164$ preservice biology teachers) also provided their semester of study. To check whether the PCK_{SR} -bio is sensitive to change, we conducted an interventional study with an experimental group with 11 and a control group with 8 participants.

5.6.1.2 Data Collection

Concerning the relationship between PCK_{SR} and CK, the participants answered the items of the PCK_{SR} -bio and a test about their scientific reasoning competencies (i.e., CK; Khan & Krell, 2019).

Regarding the relation between PCK_{SR} and the semester of study, the bachelor's and master's students provided information about their semester of study (up to a maximum of 12); participants in in-service training were assigned a semester number of 13.

As part of the interventional study, the experimental group took a course about promoting students' scientific reasoning competencies with competence-oriented learning tasks. The control group attended a course about promoting students' subject matter knowledge with digital media, in which the topic of scientific reasoning and how to teach it was not addressed. Both courses were electives

within a mandatory module in the master's program and were provided as in-person courses. The participants' test score on the PCK_{SR}-bio was assessed at the beginning (April 2023) and the end (July 2023) of these courses.

5.6.1.3 Data Analysis

In order to analyze the relationship between PCK_{SR} and CK, a correlation analysis was conducted. We expected to find a significant positive correlation between the preservice teachers' PCK and CK (comparable to the results of Großschedl et al., 2019). Regarding the relation between PCK_{SR} and the semester of study, the test score on the PCK_{SR}-bio was normally distributed but the semester of study was not ($p < .05$; Shapiro-Wilk test). Therefore, a Spearman correlation analysis was conducted. We assumed that there would be a small increase in knowledge over the course of study (comparable to the increase in scientific reasoning competencies, e.g., Krüger et al., 2020). In the interventional study, the test score was not normally distributed in the experimental or the control group (Shapiro-Wilk test, $p < .05$). Therefore, a Wilcoxon-W test was conducted to address the differences between pre- and posttest. As the experimental group addressed scientific reasoning-related content in their course, we expected to find an improvement in the test score of this group but not in that of the control group; this would be an indicator that our instrument is sensitive to change.

5.6.2 Results

A significant correlation was found between the test score on the PCK_{SR}-bio and the CK of the participants, $r = .41$, $p < .001$. However, no significant correlation was found between the semester of study and the test score, $r = .12$, $p = .180$.

In the interventional study, the experimental group obtained a score of $M = .52$ ($SD = .14$) in the pretest and of $M = .58$ ($SD = .21$) in the posttest (i.e., $\Delta M = .06$). The test score did not significantly change, $z = .897$, $p = .370$. The control group achieved a test score of $M = .43$ ($SD = .10$) in the pretest and of $M = .45$ ($SD = .12$) in the posttest (i.e., $\Delta M = .02$). Again, the test score did not significantly change, $z = .425$, $p = .671$. The test scores are shown in Figure 5.2. These results will be discussed below.

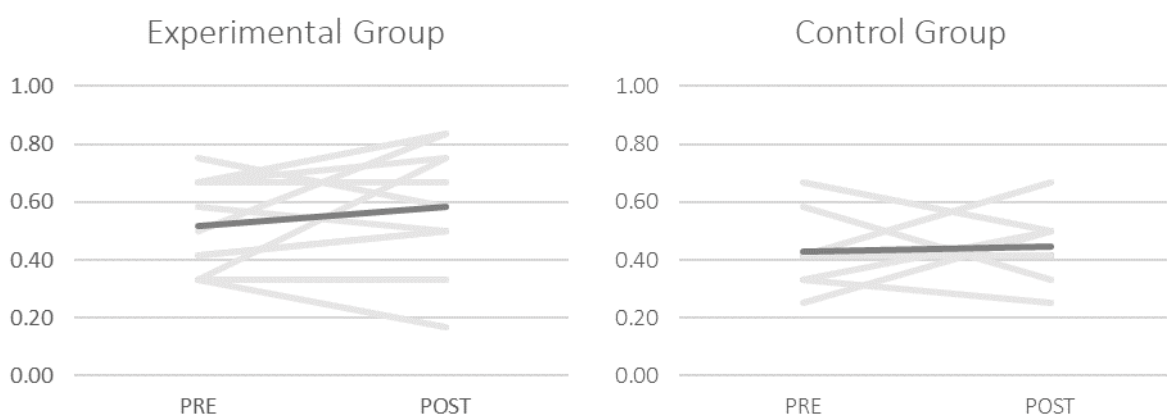


Figure 5.2: Test scores of the experimental group and of the control group in the pre- and posttest
Note. The lighter lines describe the individual development of the test scores of the different participants in each group. The darker lines describe the average development of the test score in each of the groups.

5.7 Overall Discussion

PCK, as a central part of science teachers' professional knowledge, positively affects the quality of their classroom practice and student learning (e.g., Kunter et al., 2013). Therefore, preservice science teachers should already acquire basic PCK during initial teacher education, which is why suitable paper-pencil instruments are needed to evaluate the development of preservice teachers' PCK. However, the development of those instruments comes with challenges, namely, in terms of cognitive operations and task formats, the construct representation, as well as psychometric properties. This study aimed to address these challenges by testing the validity of the test score interpretation of an instrument to assess PCK about scientific reasoning (PCK_{SR}-bio), based on four sources of validity evidence, namely, test content, response processes, internal structure, and relations to other variables.

Regarding evidence based on test content, we systematically developed the instrument based on a theoretical framework that comprised the two major PCK facets KSU and KISR, their connections (van Driel et al., 2023), and the three highly relevant scientific reasoning practices, planning and conducting investigations, generating hypotheses, and using scientific models (Rönnebeck et al., 2016). In order to ensure that the PCK_{SR}-bio adequately represents the underlying construct, the items were developed based on this theoretical framework (see Table 5.2). The distractors were developed based on common conceptions and were identified by assessing the responses of 14 preservice science teachers to open-ended questions. We developed one item for each intersection of the PCK components and their connections with the scientific reasoning skills (see Table 5.2). This might have impacted the representation of the underlying construct because it is no longer fully represented if single items have to be excluded (e.g., because they limit internal consistency). Nevertheless, due to this systematic development and evaluation of the instrument, we can confirm that the items of the PCK_{SR}-bio adequately address the construct of PCK_{SR}, which supports the validity of the test score interpretation (see Section 5.3).

Regarding evidence based on response processes, we were able to show that the 10 preservice biology teachers chose an attractor based on "PCK argumentation" (KSU, KISR, or both) more often than based on "non-PCK argumentation" (surface characteristics or intuition). In contrast, a distractor was almost exclusively selected based on "non-PCK argumentation". Furthermore, if the preservice biology teachers justified their selection of a response option based on PCK, they most often argued with the PCK component that corresponded to the PCK component addressed in the respective item (KISR for KISR items, KSU for KSU items, and both components at the same time for the connections; Table 5.4). These findings supported our assumption that a significantly higher proportion of answers would indicate PCK-related reasoning than non-PCK-related reasoning. However, when interpreting the results, the sample size of 10 participants has to be considered. Furthermore, justifications were counted as a connection between KSU and KISR if they occurred together in the justification for choosing a response option. Therefore, no statement can be made about the direction or strength of the connection. Feedback provided by the participants in the interview study (Lieberei et al., 2023) was used to optimize the language of the instrument to make it more understandable to the target audience.

Regarding evidence based on internal structure, the instrument's internal consistency was relatively low, which means that the participants' responses were relatively inconsistent on the whole. Particularly in the development of instruments for assessing complex constructs such as PCK, the question arises as to what lower limit can still be regarded as acceptable because a low level of internal consistency does not necessarily lead to an impairment of the validity of test score interpretation (Schmitt, 1996; Taber, 2018). For example, Taber (2018) showed that values of .45 are sometimes interpreted as "sufficient" and "acceptable" in studies published in leading science journals; these values correspond with the internal consistency that participants in the master's program (.44) and in-

service training (.58) showed in our study. However, as there is no standardized lower limit, evidence for the validity of test score interpretation should always be considered based on several sources. While an internal consistency that is too low can indicate that the underlying construct is not assessed accurately, an internal consistency that is very high can indicate that some items may be redundant (Taber, 2018). Cronbach (1951) argued that, although high internal consistency is desirable, it is the valid interpretation of the test score that is crucial; this is still possible for instruments that have a rather low internal consistency. In our study, the internal consistency increased with a higher educational level. This supports our assumption that internal consistency increases as the level of experience rises (Krepf et al., 2018), which is consistent with other empirical findings on MC–PCK instruments (e.g., Kleickmann et al., 2013). Furthermore, the internal consistency for the group of participants in in-service training is comparable to that of other PCK instruments (e.g., Förtsch et al., 2016). Our results thus show that a valid interpretation of the test score of the PCK_{SR}-bio is possible even if the internal consistency of the instrument used is relatively low. One way to increase internal consistency is to develop more items (Fischer, 2023; Taber, 2018), although this is associated with a higher test duration.

Regarding evidence based on relations to other variables, a significant positive correlation was found between the test score on the PCK_{SR}-bio and the CK (in this case, scientific reasoning competencies) of the preservice biology teachers, which is in line with theoretical assumptions and previous empirical findings (e.g., $q = .34$ [mean effect] in Großschedl et al., 2019). However, no significant correlation was found between the test score on the PCK_{SR}-bio and the semester of study. This contradicted our assumption that PCK_{SR} increases slightly over the course of study. As CK is a prerequisite for PCK (Friedrichsen et al., 2009), we expected to find a small increase over the course of study, comparable to that found in studies on scientific reasoning competencies (e.g., Krüger et al., 2020). However, empirical studies have shown that even in-service teachers often have weaknesses in planning lessons about scientific reasoning (Capps & Crawford, 2013). The PCK_{SR}-bio assesses only a very specific area of knowledge. Therefore, there might be a lack of learning opportunities to acquire this knowledge, especially during initial teacher education.

To test whether the PCK_{SR}-bio is sensitive to change, we conducted an intervention study with two groups: The experimental group took part in a course in which content on PCK about scientific reasoning was explicitly discussed. In contrast, the control group participated in a course in which this content was not addressed. It was shown that the test scores did not significantly change for either group after the intervention. However, descriptively, a small but greater increase in the test score was seen in the experimental group ($\Delta M = .06$) compared to the control group ($\Delta M = .06$; three times as high). A larger sample would be needed, especially because of the rather low internal consistency of the PCK_{SR}-bio, to investigate whether this difference was not significant because of the small group size, which was due to the fact that only a few students in both groups participated in both the pre- and posttest. Furthermore, in the pretest, the test score of the control group was substantially lower than that of the experimental group. So far the results support the assumption that the PCK_{SR}-bio is only moderately sensitive to change.

5.7.1 Conclusion and Outlook

In summary, with some limitations, we found validity evidence for each of the sources. Regarding the overall research question, this means that we were able to address common challenges in developing MC–PCK instruments. In contrast to the assumption that MC instruments are not appropriate for assessing PCK, for the case of the PCK_{SR}-bio, we were able to find evidence from different sources for the validity of the test score interpretation as PCK_{SR}. Through systematic development, we were able to ensure that the construct was adequately represented. This is a crucial result because it supports the assumption that a valid interpretation of the test scores of MC–PCK instruments is possible despite their low internal consistency.

In order to increase the internal consistency, it would be possible to develop more items (e.g., one for each intersection, see Table 5.1), although this would increase the test duration. To further investigate whether the test score did not differ between the groups of bachelor's students, master's students, and participants in in-service training because they did not have enough learning opportunities to acquire PCK_{SR} during initial teacher education, a focus group interview study in which students currently completing their master's degree are asked to report on their learning opportunities for this topic is currently being conducted.

5.7.2 Recommendations

The main goal of this study was to address common challenges in developing MC–PCK instruments by considering multiple sources of validity evidence. Therefore, we would like to make the following recommendations that might help other researchers in the field to develop MC–PCK instruments:

- (1) Using an approach to develop and evaluate an instrument that is based on the sources of validity evidence specified by AERA et al. (2014) can help to address common challenges in developing MC–PCK instruments regarding an adequate representation of the target construct. In particular, PCK instruments should focus on narrowly defined PCK aspects to avoid an under- or overrepresentation (Smith & Banilower, 2015). At the same time, the PCK components should not only be considered separately; their connections should also be considered.
- (2) Our results support the assumption that a valid interpretation of the test score is possible despite the low internal consistency of the instrument. Therefore, an instrument should not be excluded just because it does not meet a lower limit, and an internal consistency that is too high should also be viewed critically (Taber, 2018; Schmitt, 1996). Because there is no standardized lower limit for internal consistency, when developing MC instruments, especially for assessing complex constructs such as PCK, the validity of the test score interpretation has to be ensured based on different sources of validity evidence.
- (3) When choosing a PCK instrument, we recommend considering the topic area it covers. In particular, preservice science teachers may not have the opportunity to acquire PCK about a specific topic during initial teacher education.

Acknowledgments

The authors would like to thank Gráinne Newcombe for English language editing.

5.8 References

- AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Becerra, B., Núñez, P., Vergara, C., Santibáñez, D., Krüger, D. & Cofré, H. (2022). Developing an instrument to assess pedagogical content knowledge for evolution. *Research in Science Education*, 53(8), 213-229. <https://doi.org/10.1007/s11165-022-10042-0>
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box. *Learning and Instruction*, 79, 101600. <https://doi.org/10.1016/j.learninstruc.2022.101600>

- Capps, D., & Crawford, B. (2013). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education*, 24(3), 497–526. <https://doi.org/10.1007/s10972-012-9314-z>
- Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., . . . Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–94). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_2
- Chan, K. K. H. & Hume, A. (2019). Towards a consensus Model. In A. Hume, R. Cooper & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 3–76). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_1
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Davis, E. A. & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, 34(3), 3–14. <https://doi.org/10.3102/0013189X034003003>
- Ercikan, K., & Pellegrino, J. (2017). Validation of score meaning using examinee response processes for the next generation of assessments. In K. Ercikan & J. Pellegrino (Eds.), *Validation of score meaning for the next generation of assessments: The use of response processes*. (pp. 1–8). Routledge.
- Fischer, H., Boone, W., & Neumann, K. (2023). Quantitative research designs and approaches. In N. Lederman, D. Zeidler, & J. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 28–59). Routledge. <https://doi.org/10.4324/9780367855758-3>
- Förtsch, C., Werner, S., von Kotzebue, L., & Neuhaus, B. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666. <https://doi.org/10.1080/09500693.2016.1257170>
- Friedrichsen, P., Abell, S. K., Pareja, E., Brown, P., Lankford, D., & Volkmann, M. (2009). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching*, 46(4), 357–383. <https://doi.org/10.1002/tea.20283>
- Germany. In J. Pedersen, T. Isozaki, & T. Hirano (Eds.), *Model science teacher preparation programs* (pp. 29–52). Information Age.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). Routledge.
- Göhner, M. & Krell, M. (2022). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*, 52(2), 395–414. <https://doi.org/10.1007/s11165-020-09945-7>
- Großschedl, J., Welter, V. D. E. & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402–439. <https://doi.org/10.1002/tea.21482>
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Routledge. <https://doi.org/10.4324/9780203825945>
- Hubley, A., & Zumbo, B. (2017). Response processes in the context of validity: Setting the stage. In B. Zumbo & A. Hubley (Eds.), *Understanding and investigating response processes in validation research: Vol. 69. Social Indicators Research Series* (pp. 1–12). Springer. https://doi.org/10.1007/978-3-319-56129-5_1
- Justi, R. & van Driel, J. (2005). The development of science teachers' knowledge on models and modelling: Promoting, characterizing, and understanding the process. *International Journal of Science Education*, 27(5), 549–573. <https://doi.org/10.1080/0950069042000323773>

- Jüttner, M., Boone, W., Park, S., & Neuhaus, B. (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. <https://doi.org/10.1007/s11092-013-9157-y>
- Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457. <https://doi.org/10.1080/02796015.2013.12087465>
- Kenyon, L., Davis, E. A. & Hug, B. (2011). Design approaches to support preservice teachers in scientific modeling. *Journal of Science Teacher Education*, 22(1), 1–21. <https://doi.org/10.1007/s10972-010-9225-9>
- Khan, S. & Krell, M. (2019). Scientific reasoning competencies: A case of preservice teacher education. *Canadian Journal of Science, Mathematics and Technology Education*, 19(4), 446–464. <https://doi.org/10.1007/s42330-019-00063-9>
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, 64(1), 90–106. <https://doi.org/10.1177/0022487112460398>
- KMK (2024). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. [Common content-specific requirements of the federal states for the subject sciences and subject didactics in teacher education]. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42(2), 262–271. <https://doi.org/10.1016/j.shpsa.2010.11.034>
- Krell, M., Vorholzer, A. & Nehring, A. (2022). Scientific reasoning in science education: From global measures to fine-grained descriptions of students' competencies. *Education Sciences*, 12(2), 97. <https://doi.org/10.3390/educsci12020097>
- Krepf, M., Plöger, W., Scholl, D., & Seifert, A. (2018). Pedagogical content knowledge of experts and novices—what knowledge do they activate when analyzing science lessons? *Journal of Research in Science Teaching*, 55(1), 44–67. <https://doi.org/10.1002/tea.21410>
- Krüger, D., Hartmann, S., Nordmeier, V. & Upmeyer zu Belzen, A. (2020). Measuring scientific reasoning competencies. In O. Zlatkin-Troitschanskaia, H. A. Pant & C. Lautenbach (Eds.), *Student learning in German higher education: Innovative modelling and measurement approaches and research results* (pp. 261–280). Springer. https://doi.org/10.1007/978-3-658-27886-1_13
- Kulgemeyer, C., Kempin, M., Weißbach, A., Borowski, A., Buschhüter, D., Enkrott, P., Reinhold, P., Riese, J., Schecker, H., Schröder, J. & Vogelsang, C. (2021). Exploring the impact of pre-service science teachers' reflection skills on the development of professional knowledge during a field experience. *International Journal of Science Education*, 43(18), 3035–3057. <https://doi.org/10.1080/09500693.2021.2006820>
- Kulgemeyer, C., Riese, J., Vogelsang, C., Buschhüter, D., Borowski, A., Weißbach, A., Jordans, M., Reinhold, P. & Schecker, H. (2023). How authenticity impacts validity: Developing a model of teacher education assessment and exploring the effects of the digitisation of assessment methods. *Zeitschrift für Erziehungswissenschaft*, 26(3), 601–625.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>
- Lieberei, T., Dawborn-Gundlach, M., van Driel, J. & Krell, M. (2024). Preservice science teachers' knowledge about how to teach scientific reasoning. *Teaching Science*, 70(2), 30–43.
- Lieberei, T., Welter, V. D. E., Großmann, L. & Krell, M. (2023). Findings from the expert-novice paradigm on differential response behavior among multiple-choice items of a pedagogical content knowledge test – implications for test development. *Frontiers in Psychology*, 14, 1240120. <https://doi.org/10.3389/fpsyg.2023.1240120>

- Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching*, 41(4), 370–391. <https://doi.org/10.1002/tea.20007>
- Magnusson, S., Krajcik, J. & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47217-1_4
- Martinez, M. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34(4), 207–218. https://doi.org/10.1207/s15326985ep3404_2
- McCoach, D. B., Gable, R., & Madura, J. (2013a). Evidence based on relations to other variables: Bolstering the empirical validity arguments for constructs. In D. B. McCoach, R. Gable, & J. Madura (Eds.), *Instrument development in the affective domain: School and corporate applications* (3rd ed., pp. 209–248). Springer. https://doi.org/10.1007/978-1-4614-7135-6_6
- McCoach, D. B., Gable, R., & Madura, J. (2013b). Evidence based on test content. In D. B. McCoach, R. Gable, & J. Madura (Eds.), *Instrument development in the affective domain* (3rd ed., pp. 91–107). Springer. https://doi.org/10.1007/978-1-4614-7135-6_3
- Messick, S. (1996). Validity of performance assessments. In G. W. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp. 1–18). National Center for Education Statistics.
- Neumann, K., Härtig, H., Harms, U., & Parchmann, I. (2017). Science teacher preparation in Germany. In J. Pedersen, T. Isozaki, & T. Hirano (Eds.), *Model science teacher preparation programs: An international comparison of what works best* (pp. 29–52). Information Age Publishing.
- O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 1-13. <https://doi.org/10.1177/1609406919899220>
- Osborne, J. (2014). Teaching Scientific Practices: Meeting the challenge of change. *Journal of Science Teacher Education*, 25(2), 177–196. <https://doi.org/10.1007/s10972-014-9384-1>
- 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. <https://doi.org/10.1002/tea.21022>
- Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26(1), 108–116. <https://doi.org/10.7334/psicothema2013.260>
- Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>
- Rosenkränzer, F., Hörsch, C., Schuler, S. & Riess, W. (2017). Student teachers' pedagogical content knowledge for teaching systems thinking: Effects of different interventions. *International Journal of Science Education*, 39(14), 1932–1951. <https://doi.org/10.1080/09500693.2017.1362603>
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>
- Schneider, R. M., & Plasman, K. (2011). Science teacher learning progressions: A review of science teachers' pedagogical content knowledge development. *Review of Educational Research*, 81(4), 530–565. <https://doi.org/10.3102/0034654311423382>
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE.
- Schuster, D., Cobern, W., Applegate, B., Schwartz, R., Vellom, P., & Undreiu, A. (2006). Assessing pedagogical content knowledge of inquiry science teaching: Developing an assessment instrument to support the undergraduate preparation of elementary teachers to teach science as inquiry. In D. Deeds & B. Callen (Eds.), *Proceedings of National STEM Assessment Conference*. (pp. 247-265).
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107. <https://doi.org/10.7334/psicothema2013.256>

- Smith, P., & Banilower, E. R. (2015). Assessing PCK: A new application of the uncertainty principle. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (pp. 88–103). Routledge.
- Taber, K. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296.
<https://doi.org/10.1007/s11165-016-9602-2>
- Upmeier zu Belzen, A., van Driel, J., & Krüger, D. (2019). Introducing a framework for modeling competence. In A. Upmeier zu Belzen, D. Krüger, & J. van Driel (Eds.), *Towards a competence-based view on models and modeling in science education: Vol. 12. Models and modeling in science education* (pp. 3–19). Springer. https://doi.org/10.1007/978-3-030-30255-9_1
- van Driel, J., & Berry, A. (2019). Pedagogical content knowledge in preservice teacher education. In M. A. Peters (Ed.), *Encyclopedia of teacher education* (pp. 1-6). Springer Singapore.
https://doi.org/10.1007/978-981-13-1179-6_176-1
- van Driel, J. H., Hume, A., & Berry, A. (2023). Research on science teacher knowledge and its development. In N. G. Lederman, D. L. Zeidler, & J. S. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 1123–1161). Routledge. <https://doi.org/10.4324/9780367855758-41>
- Verloop, N., van Driel, J., & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research*, 35(5), 441–461.
[https://doi.org/10.1016/S0883-0355\(02\)00003-4](https://doi.org/10.1016/S0883-0355(02)00003-4)
- Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen* [Intercoder agreement and intercoder reliability]. Hogrefe Verlag für Psychologie.

6 Studie 3: Using the PCK Map Approach to Obtain Validity Evidence Based on Response Processes for a Multiple-Choice Instrument²⁸

Abstract

As a central part of their professional knowledge, science teachers' pedagogical content knowledge (PCK) positively affects the quality of classroom practice and, thus, students' learning. To examine the development of PCK during science teacher education, test instruments of high psychometric quality are required. These instruments should also be designed to be economically applicable to large samples, which is usually true for multiple-choice instruments rather than for those with an open-ended response format. However, one difficulty of multiple-choice instruments is that they do not provide direct information about underlying cognitive processes in item responses, although these are a crucial source of validity evidence, especially in knowledge tests. In this study, we examined how Park and Chen's PCK map approach can be used to obtain validity evidence using interview-based justifications given by participants for choosing answer options in a multiple-choice instrument used to assess PCK about scientific reasoning. The results of 10 undergraduate students showed that their justifications for choosing an answer option mainly referred to the PCK components that the respective items intended to capture. At the same time, the participants did not mention PCK components not covered by the items. These findings illustrate how the PCK map approach can be used to obtain validity evidence based on response processes. The findings and limitations are discussed, and recommendations for using the PCK map approach to investigate validity evidence are given.

Keywords: PCK map approach, validity evidence, response processes

²⁸ Lieberei, T., Welter, V. D. E. & Krell, M. (2024). Using the PCK map approach to obtain validity evidence based on response processes for a multiple-choice instrument. In G. Kaya, M. Sardag & I. Donmez (Hrsg.), *Proceedings book series-V of the ESERA 2023 conference: Connecting science education with cultural heritage* (S. 114–126). Nobel Bilimsel Eserler.

6.1 Introduction

Science teachers' professional competence, particularly their professional knowledge, has a crucial influence on the effectiveness of their teaching (Blömeke et al., 2022; Hattie, 2023). Science teachers' professional knowledge is comprised of different domains (Kind & Chan, 2019): In order to teach a specific topic, science teachers need not only pedagogical knowledge (PK; knowledge about the basic principles of education, teaching, learning, and the learners) and content knowledge (CK; knowledge about the topics within the discipline to be taught); they also need pedagogical content knowledge (PCK; knowledge about learners' conceptual ideas about a particular topic and instructional strategies for teaching it to a particular group of students). Even though PCK develops over time as a result of learning and teaching experience, as well as reflection on that experience (van Driel & Berry, 2019), preservice science teachers are expected to acquire basic PCK during their university studies; this is specified in standards for teacher education (e.g., Germany: KMK, 2019). In order to evaluate the acquisition of preservice science teachers' PCK throughout their university studies, appropriate test instruments are needed. For this purpose, multiple-choice instruments are particularly useful compared to open-ended questions because they have advantages regarding test economy and objectivity (Haladyna, 2004). However, developing multiple-choice instruments, specifically to assess PCK, also comes with challenges, for example, in terms of adequate construct coverage (Smith & Banilower, 2015). Therefore, multiple-choice PCK instruments should focus on capturing narrowly defined PCK aspects (Smith & Banilower, 2015), especially when participants' cognitive processes in item response are unknown. To ensure that the items of a multiple-choice instrument capture the construct adequately, when developing these instruments, it is important to provide validity evidence based on the response processes observed in participants' item responses (AERA et al., 2014). Data relating to this source of validity evidence are often collected through cognitive interviews (Padilla & Benítez, 2014), which can be evaluated in different ways. Concerning PCK, a common method for analysing data collected in interviews is the PCK map approach by Park and Chen (2012), in which participants' PCK is visualised based on quantitative content analysis.

Against this background, we aimed to explore the extent to which the PCK map approach can also be used to obtain validity evidence based on response processes. To demonstrate the application of this procedure, we used interview-based justifications for the selection of individual response options in a multiple-choice PCK instrument. For this purpose, we used the PCK_{SR}-bio, a multiple-choice instrument for assessing preservice science teachers' PCK about scientific reasoning with items concerning biology-related classroom situations (Lieberei et al., 2023).

6.2 Theoretical Background

The following sections describe PCK, its assessment, how validity evidence can be obtained based on response processes, and the PCK map approach.

6.2.1 Pedagogical Content Knowledge

The concept of PCK as knowledge of how to teach specific topics and what learners understand was first introduced by Lee Shulman in 1986 and has since been further developed by various scientists. PCK describes knowledge about particular ways to teach specific topics to enhance students' learning and is therefore understood as a topic-specific construct (Gess-Newsome, 2015). Due to its complexity, different models have attempted to define PCK in more depth by distinguishing between its various interacting components. A common model is the pentagon model of PCK by Park and Oliver (2008), which comprises five components (see Figure 6.1): Orientation to Teaching Science (e.g., beliefs about the purpose of teaching science; OTS), Knowledge of Instructional Strategies for Teaching Science (e.g., subject-specific and topic-specific strategies; KISR), Knowledge of Assessment of Science Learning (e.g., knowledge about different kinds of performance evaluation; KAs), Knowledge of Curriculum (e.g.,

knowledge about its implications for teaching a specific subject; KSC), and Knowledge of Students' Understanding (e.g., knowledge about students' [mis]conceptions and prior knowledge; KSU). These components are connected, and a high level of PCK is characterised by strong connections between the components (Park & Chen, 2012). The two components, KSU and KISR, as well as their connections, have been empirically shown to be most crucial (Alonzo et al., 2012; Jin et al., 2015; Park & Chen, 2012). The connection KISR→KSU describes the impact of instruction on students' thinking, and the connection KSU→KISR describes the designing of instruction based on students' conceptions.

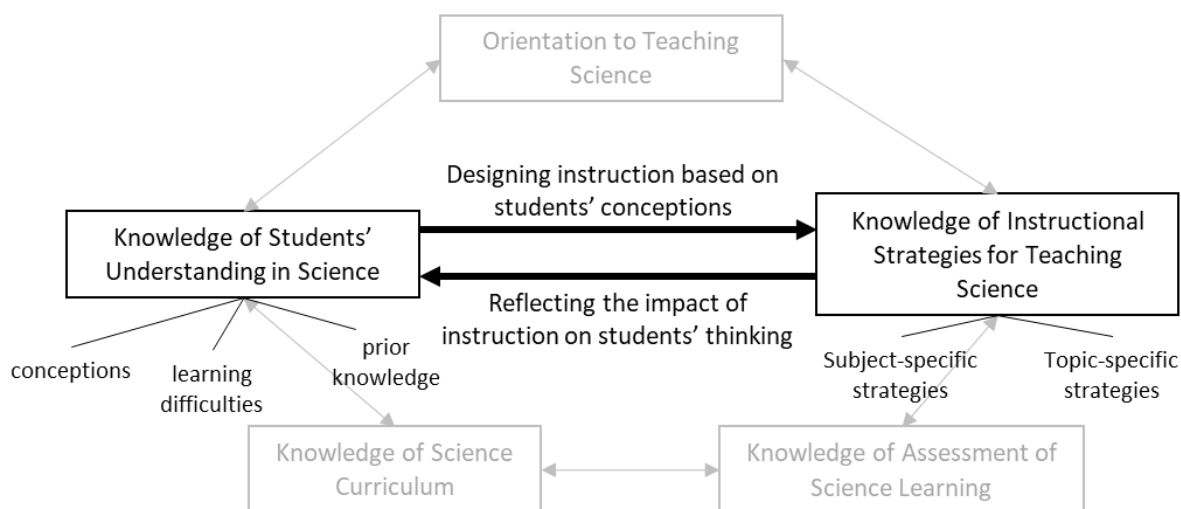


Figure 6.1: Pentagon model of PCK. KSU and KISR, as well as their connections, are highlighted

Note: The figure is taken from Lieberei et al. (2023).

Preservice science teachers are expected to acquire basic PCK during their university studies, which is why suitable test instruments are needed to monitor the acquisition of PCK throughout their university studies. This leads to the question of what type of test instruments are most appropriate for assessing preservice science teachers' PCK.

6.2.2 Assessing PCK

Several paper-pencil PCK instruments have already been published; they have vastly different task formats. Although instruments with open-ended questions (e.g., *Content Representation* by Loughran et al., 2004) have the potential to capture a broader range of respondents' cognitions (Martinez, 1999), instruments with multiple-choice items (e.g., *Pedagogy of Science Inquiry Teaching Test* by Schuster et al., 2006) have advantages in terms of objectivity and test economy (Haladyna, 2004). Therefore, multiple-choice instruments can be used to assess a larger groups' knowledge with less effort than open-ended instruments, which is why they are especially appropriate for assessing preservice science teachers' PCK. At the same time, assessing PCK with multiple-choice instruments also comes with challenges regarding an adequate construct representation (Smith & Banilower, 2015). As PCK is a complex construct with diverse components (e.g., Park & Chen, 2012), the risk of underrepresenting the construct should be avoided. Therefore, focusing on narrowly defined PCK aspects is recommended when developing multiple-choice instruments (Smith & Banilower, 2015). At the same time, an overrepresentation of the construct can lead to the assessment of CK or beliefs about teaching instead of PCK (Smith & Banilower, 2015). Therefore, ensuring that the construct is captured adequately is especially important. This is mainly the case for multiple-choice instruments in which the cognitive processes involved in participants' item responses are unknown. Therefore, when developing

multiple-choice instruments, it is important to ensure the validity of the test score interpretation based on response processes.

6.2.3 Validity Evidence Based on Response Processes

The Standards for Educational and Psychological Testing (AERA et al., 2014) describe that validity evidence that is based on the response processes observed during testing “can provide evidence concerning the fit between the construct and the detailed nature of the performance or response actually engaged in by test takers” (p. 15). The goal is to gain insights into cognitive, affective, and motivational mechanisms underlying participants’ responses to the test items (Hubley & Zumbo, 2017), including intentional processes and reactions (Ercikan & Pellegrino, 2017). In the development process of multiple-choice instruments, it is crucial to develop the items in such a way that they capture the components they intend to capture because participants’ answers (i.e., the selection of one response option) do not provide any information about the underlying cognitive processes.

Validity evidence based on response processes is often collected through think-aloud protocols or in interviews (Padilla & Benítez, 2014), which is why there are various methods for analysing the data collected. Regarding PCK, one of these methods is the PCK map approach.

6.2.4 The PCK Map Approach

The PCK map approach is a common method used for analysing PCK-related data (Chan, 2022; Park & Chen, 2012; Park & Suh, 2019). Using the pentagon model as an analytic framework, PCK maps can provide a quantitative content analysis-based visualisation of the expression of individual PCK components and, in particular, their connections (Park & Chen, 2012). Three steps are typically carried out when PCK maps are used as a visualisation (Park & Chen, 2012; Park & Suh, 2019): First, an in-depth analysis is conducted in which the PCK components of the pentagon model are identified in the given data (e.g., by coding an interview protocol). This is followed by the enumerative approach, in which the PCK components that were mentioned together are interpreted as being connected. Each connection is thereby counted with the same “strength”, even if they occur in varying degrees in the data (Park & Chen, 2012). In the final step, the data is analysed and clustered to identify common patterns.

However, the PCK map approach has some limitations. Counting the connections with the same “strength” results in a simplified visualisation of individual connections (Aydin & Boz, 2013). Furthermore, this approach cannot represent contextual factors that influence PCK (Park & Suh, 2019). In addition, the PCK map approach is usually only used for smaller samples of up to 50 participants (Chan, 2022). Despite these limitations, the PCK map approach is an established procedure that can be used to evaluate data collected in various ways, for example, in classroom observations, interviews, or lesson plans (Chan, 2022). Several data sources are often used, and there are also various applications for quantifying and visualising the data (Chan, 2022). However, to the best of our knowledge, no study has yet investigated how the PCK map approach can be used to obtain validity evidence based on response processes.

6.2.5 Aim of this Study

We explored the extent to which the PCK map approach can be used to obtain validity evidence based on the response processes observed in multiple-choice instruments. We thus formulated this research question:

Can the PCK map approach be used to obtain validity evidence based on response processes?

To answer this research question, we used a multiple-choice instrument to assess preservice science teachers’ PCK about scientific reasoning: the PCK_{SR}-bio.

6.3 Methods

The following sections describe the development process of the PCK_{SR}-bio instrument as well as the data collection and analysis.

6.3.1 Instrument Development

The PCK_{SR}-bio was developed in a theory-based manner to assess preservice science teachers' PCK about scientific reasoning (Lieberei et al., 2023). In order to enable students to participate in current discussions and understand relevant decision-making processes, science teachers need to teach scientific reasoning competencies to their students (Osborne, 2013), which is why fostering these competencies is a major goal in science education (Khan & Krell, 2019; Mathesius et al., 2016). For this reason, science teachers need to know common (mis)conceptions of students about scientific reasoning as well as appropriate instructional strategies for teaching scientific reasoning. Because there is not yet a standard definition of scientific reasoning (Krell et al., 2022), we selected three highly relevant scientific reasoning skills, which Rönnebeck et al. (2016), among others, pointed out: (1) planning and conducting investigations, (2) generating hypotheses, and (3) using scientific models.

However, because PCK instruments should not only be designed in a topic-specific manner but also be limited to the measurement of narrowly defined PCK components (Smith & Banilower, 2015), the PCK_{SR}-bio focuses on the most relevant PCK components: KSU, KISR, and their connections, KISR→KSU and KSU→KISR.

Based on this framework, a total of 12 items were assigned to each intersection of the PCK components and their interactions with the scientific reasoning skills (see Table 6.1).

Table 6.1: Systematic overview of the 12 multiple-choice items developed to capture three skills of scientific reasoning, two PCK components, and two PCK connections

	Planning and Conducting Investigations	Generating Hypotheses	Using Scientific Models
KISR	Item #1	Item #2	Item #3
KSU	Item #4	Item #5	Item #6
KISR→KSU	Item #7	Item #8	Item #9
KSU→KISR	Item #10	Item #11	Item #12

A sample item of the PCK_{SR}-bio is shown in Figure 6.2. The complete instrument is available upon request from the last author. The items address biology-related classroom situations but can also be answered by preservice science teachers of other subjects as they relate to scientific reasoning in general.

Stem: A biology teacher has worked with his students on the structure of DNA using models. In a subsequent lesson, he asks his students what models are and what function they have in biology. One student answers: “Models are visualisations of something, for example, DNA. We can use them to learn how DNA is built.”

Prompt: Please tick the ways in which the teacher should take this idea about models and extend it through an understanding that is as scientifically and methodologically appropriate as possible.

Options:

He should...

- (1) ... ask his students to make their own DNA models to highlight the potential of models as representations.
- (2) ... practise “model critique” with his students by working out similarities and differences between DNA models and DNA.
- (3) ... reflect on models by elaborating on the importance of DNA models in the discovery of DNA structure.

Figure 6.2: Sample item of the PCK_{SR}-bio (Item #12)

All items are structured similarly and are based on theoretical assumptions and empirical research findings on students’ (mis)conceptions about scientific reasoning skills and appropriate instructional strategies to foster related competencies. For instance, Item #12 addresses the connection KSU→KISR related to using scientific models. Because models are often seen as visualisations of biological phenomena (Upmeyer zu Belzen et al., 2019), and the student mentioned in the item has this conception as well, it is necessary to choose an instructional strategy in which the students learn that models can also be used as epistemic tools in science (Göhner & Krell, 2022; Knuuttila, 2011). Therefore, Answer Option 3 is correct as it describes the only way to promote such an understanding (in the context of DNA models).

6.3.2 Data Collection

To examine how the PCK map approach can be used to obtain validity evidence based on response processes, we asked 10 undergraduate preservice biology teachers at two German universities to answer the multiple-choice items of the PCK_{SR}-bio. After each answer, the preservice biology teachers justified why they chose a particular answer option instead of the other two options. The interviews were conducted online, and participation was voluntary. Our local ethics committee reviewed and approved our study (ID 2021_LI56) before any data collection took place.

6.4 Data Analysis

Subsequently, we conducted a structural qualitative content analysis (Schreier, 2012) of the preservice biology teachers’ justifications for choosing the attractor. For this purpose, we used a deductively developed coding system based on the pentagon model and comprising all five components. As required for the PCK map approach, justifications that referred to a single PCK component were coded once for this component (no matter how many times it was addressed in the justification given for choosing the answer option for one item). If two or more PCK components were referred to in a justification, this was counted as a connection between them. The part of the coding system with the categories KSU and KISR is shown in Table 6.2. The entire coding system can be found in Lieberei et al. (2023).

Table 6.2: Coding system for analysing the justifications given for the selection of response options based on the categories KSU and KISR

Category	Description	Empirical Evidence
KSU	Responses relating to students' (mis)conceptions and prior knowledge	<i>However, I would now tend towards answer option three, [...] because this addresses exactly the point that the student does not perceive.</i>
KISR	Responses relating to subject-specific and topic-specific strategies	<i>She states that models are pure visualisations of reality but does not consider that insights can also be gained through such model considerations.</i>

Note. The empirical evidence refers to Item #12 (see Figure 6.2).

Following the recommendations of Wirtz and Caspar (2002), 25 % of the interviews were second-coded by a trained student assistant. The overall intercoder agreement according to Cohen's Kappa was .87, indicating an almost perfect agreement (Landis & Koch, 1977).

Next, we quantified the codings for those items that each referred to the same PCK component or connection (i.e., Items #1, #2, and #3 for KISR; Items #4, #5, and #6 for KSU; Items #7, #8, and #9 for KISR→KSU; and Items #10, #11, and #12 for KSU→KISR). We visualised the data in four separate PCK maps, one for each PCK component and their connections included in the test. By doing so, we were able to check whether the participants' justifications referred to precisely that PCK component that the test intended to capture.

6.5 Results

The four PCK maps referring to each of the PCK_{SR}-bio's PCK components and their connections are shown in Figure 6.3.

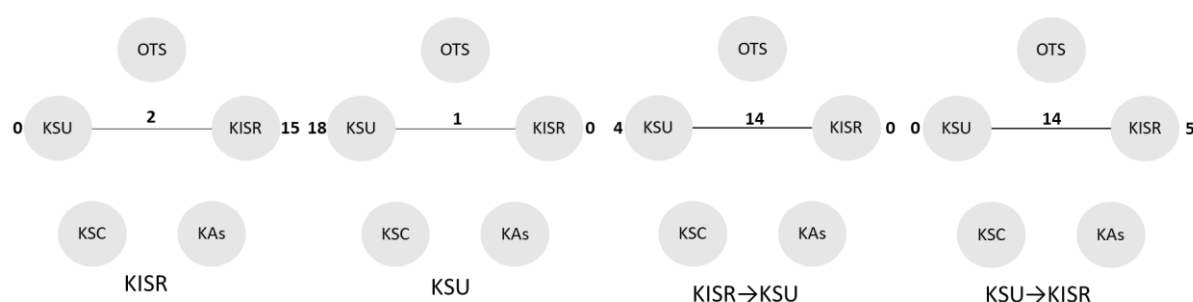


Figure 6.3: PCK maps, divided into the four PCK components of the PCK_{SR}-bio

Note. OTS = Orientations to Teaching Science; KISR = Knowledge of Instructional Strategies for Teaching Science; KAs = Knowledge of Assessment of Science Learning; KSC = Knowledge of Curriculum; KSU = Knowledge of Students' Understanding.

In their justifications for choosing an answer option, the participants mainly referred to the PCK components or their connections that the corresponding test items were intended to capture. For the items referring to KISR, most justifications were coded as KISR ($n = 15$), while only two items were coded as having a connection between KSU and KISR. For the items referring to KSU, most justifications were assigned to KSU ($n = 14$), while only one was assigned as having a connection between KSU and KISR. However, for the items referring to the two connections, KISR→KSU and KSU→KISR, most

justifications were assigned to those components ($n = 14$ each). Only four justifications were assigned to KSU for the KISR→KSU connection, and only five justifications were assigned to KISR for the KSU→KISR connection. Moreover, the participants did not mention those PCK components of the pentagon model or their connections that were not covered by the test items.

6.6 Discussion

This study examined how the PCK map approach can be used to obtain validity evidence based on response processes. As a well-established procedure, the PCK map approach is often used to evaluate qualitative data collected in interviews, classroom observations, etc. (Chan, 2022). In this way, the characteristics of the PCK of individual participants or groups can be visualised to examine and compare individuals' PCK with each other (Chan, 2022). In addition to this, we wanted to explore the extent to which the PCK map approach can also be used to obtain validity evidence based on response processes.

The analysis of 10 interviews conducted with undergraduate preservice biology teachers showed that the participants' justifications could be assigned to those PCK components and their connections that the corresponding items were intended to capture. This indicates that the items of the PCK_{SR}-bio are understood as intended by the participants as they assess precisely the two PCK components measured, which strengthens the validity of the test score interpretation of the PCK_{SR}-bio. At the same time, no justifications referring to other PCK components and, thus, their connections were captured by the corresponding items. This finding indicates that the PCK_{SR}-bio seems to adequately capture the underlying construct and avoids an under- or overrepresentation of PCK, which is important in PCK multiple-choice instruments (Smith & Banilower, 2015). This result also supports the validity of the test score interpretation.

However, when interpreting the results, some limitations must be considered. First, it is important to keep in mind that the interviews were conducted with only a small number of participants, who were also highly homogeneous. Even though smaller samples are common in studies using the PCK map approach (Chan, 2022), the small sample size means that it is not possible to generalise the findings of this study. This is also due to the fact that the PCK_{SR}-bio was only used as an exemplary instrument. A second limitation is related to the PCK map approach. Because the approach only assesses whether the PCK components occur together in a justification, regardless of how strongly or in what context they occur, the direction of the connections remains unknown. This limitation was also mentioned by Aydin and Boz (2013), who stated that the PCK map approach enables only a simplified visualisation of individual connections. When interpreting the results for the items referring to the KISR→KSU and KSU→KISR connections in this study, it cannot be assumed that the participants showed precisely this direction of the connection.

Nevertheless, the findings of this study support our assumption that the PCK map approach can be used to obtain validity evidence based on response processes.

6.6.1 Recommendations

On the basis of the findings of our study, we recommend using the PCK map approach to obtain validity evidence based on response processes (Chan, 2022). This approach is particularly useful for testing newly developed multiple-choice instruments because it ensures the validity of the test score interpretation so that the instrument can subsequently be used with all its benefits regarding test economy and objectivity (Haladyna, 2004).

For the PCK_{SR}-bio, which was used exemplarily in this study as an instrument for assessing PCK about scientific reasoning, the validity of the test score interpretation was demonstrated after the justifications were coded with a newly developed coding system and the PCK map approach.

Acknowledgments

The authors would like to thank Kerstin Landgraf for her support in data analysis and Gráinne Newcombe for English language editing.

6.7 References

- AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Alonzo, A. C., Kobarg, M., & Seidel, T. (2012). Pedagogical content knowledge as reflected in teacher-student interactions: Analysis of two video cases. *Journal of Research in Science Teaching*, 49(10), 1211–1239. <https://doi.org/10.1002/tea.21055>
- Aydin, S., & Boz, Y. (2013). The nature of integration among PCK components: A case study of two experienced chemistry teachers. *Chemistry Education Research and Practice*, 14(4), 615–624. <https://doi.org/10.1039/C3RP00095H>
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, 79, 101600. <https://doi.org/10.1016/j.learninstruc.2022.101600>
- Chan, K. K. H (2022). A critical review of studies using the pedagogical content knowledge map approach. *International Journal of Science Education*, 44(3), 487–513. <https://doi.org/10.1080/09500693.2022.2035011>
- Ercikan, K., & Pellegrino, J. W. (2017). Validation of score meaning using examinee response processes for the next generation of assessments. In K. Ercikan & J. W. Pellegrino (Eds.), *Validation of score meaning for the next generation of assessments: The use of response processes* (pp. 1–8). Routledge.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). Routledge.
- Göhner, M., & Krell, M. (2022). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*, 52(2), 395–414. <https://doi.org/10.1007/s11165-020-09945-7>
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed). Routledge. <https://doi.org/10.4324/9780203825945>
- Hattie, J. (2023). *Visible learning: The sequel. A synthesis of over 2,100 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9781003380542>
- Hubley, A., & Zumbo, B. (2017). Response processes in the context of validity: Setting the stage. In B. Zumbo & A. Hubley (Eds.), *Understanding and investigating response processes in validation research: Vol. 69. Social Indicators Research Series* (pp. 1–12). Springer. https://doi.org/10.1007/978-3-319-56129-5_1
- Jin, H., Shin, H., Johnson, M. E., Kim, J., & Anderson, C. W. (2015). Developing learning progression-based teacher knowledge measures. *Journal of Research in Science Teaching*, 52(9), 1269–1295. <https://doi.org/10.1002/tea.21243>
- Khan, S., & Krell, M. (2019). Scientific reasoning competencies: A case of preservice teacher education. *Canadian Journal of Science, Mathematics and Technology Education*, 19(4), 446–464. <https://doi.org/10.1007/s42330-019-00063-9>
- Kind, V., & Chan, K. K. H (2019). Resolving the amalgam: Connecting pedagogical content knowledge, content knowledge and pedagogical knowledge. *International Journal of Science Education*, 41(7), 964–978. <https://doi.org/10.1080/09500693.2019.1584931>

- KMK. (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung* [Common content-specific requirements of the federal states for the subject sciences and subject didactics in teacher education]. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42(2), 262–271. <https://doi.org/10.1016/j.shpsa.2010.11.034>
- Krell, M., Vorholzer, A., & Nehring, A. (2022). Scientific reasoning in science education: From global measures to fine-grained descriptions of students' competencies. *Education Sciences*, 12(2), 97. <https://doi.org/10.3390/educsci12020097>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lieberei, T., Welter, V. D. E., Großmann, L., & Krell, M. (2023). Findings from the expert-novice paradigm on differential response behavior among multiple-choice items of a pedagogical content knowledge test – Implications for test development. *Frontiers in Psychology*, 14, Article 1240120. <https://doi.org/10.3389/fpsyg.2023.1240120>
- Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching*, 41(4), 370–391. <https://doi.org/10.1002/tea.20007>
- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34(4), 207–218. https://doi.org/10.1207/s15326985ep3404_2
- Mathesius, S., Hartmann, S., Upmeyer zu Belzen, A., & Krüger, D. (2016). Scientific reasoning as an aspect of pre-service biology teacher education: Assessing competencies using a paper-pencil-test. In T. Tal & A. Yarden (Eds.), *The future of biology education research*. (pp. 93–110).
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265–279. <https://doi.org/10.1016/j.tsc.2013.07.006>
- Padilla, J.-L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136–144. <https://doi.org/10.7334/psicothema2013.259>
- 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. <https://doi.org/10.1002/tea.21022>
- Park, S., & Oliver, J. S. (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. <https://doi.org/10.1007/s11165-007-9049-6>
- Park, S., & Suh, J. K. (2019). The PCK map approach to capturing the complexity of enacted PCK (ePCK) and pedagogical reasoning in science teaching. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 185–197). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_8
- Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE.
- Schuster, D., Cobern, W., Applegate, B., Schwartz, R., Vellom, P., & Undreiu, A. (2006). Assessing pedagogical content knowledge of inquiry science teaching: Developing an assessment instrument to support the undergraduate preparation of elementary teachers to teach science as inquiry. In D. Deeds & B. Callen (Eds.), *Proceedings of National STEM Assessment Conference* (pp. 247–265).
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>

- Smith, P. S., & Banilower, E. R. (2015). Assessing PCK: A new application of the uncertainty principle. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (pp. 88–103). Routledge.
- Upmeier zu Belzen, A., van Driel, J., & Krüger, D. (2019). Introducing a framework for modeling competence. In A. Upmeier zu Belzen, D. Krüger, & J. van Driel (Eds.), *Towards a competence-based view on models and modeling in science education: Vol. 12. Models and modeling in science education* (pp. 3–19). Springer. https://doi.org/10.1007/978-3-030-30255-9_1
- van Driel, J., & Berry, A. (2019). Pedagogical content knowledge in preservice teacher education. In M. A. Peters (Ed.), *Encyclopedia of teacher education* (pp. 1-6). Springer Singapore. https://doi.org/10.1007/978-981-13-1179-6_176-1
- Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen* [Intercoder agreement and intercoder reliability]. Hogrefe Verlag für Psychologie.

7 Studie 4: Pedagogical Content Knowledge (PCK) about Scientific Reasoning in Different Groups of Science Teachers and Its Relation to Other Variables²⁹

Abstract

Scientific reasoning (SR), as a major goal in science education, describes the skills and knowledge required to solve scientific problems. Teachers at elementary and secondary schools, as well as teacher educators at universities need to know how to promote these competencies to their students (pedagogical content knowledge; PCK). Even though different studies have been conducted on PCK about SR (PCK_{SR}), comparing the results is difficult not only because various components and layers of PCK were considered but also because the studies were conducted with different samples (i.e., elementary and secondary school teachers as well as teacher educators). However, because these groups usually differ regarding the age of the students they teach, the curricular and learning goals the teachers pursue, as well as the university courses that constitute their teacher training, it can be assumed that the PCK_{SR} of these groups of teachers also differs from each other. In addition, PCK can be influenced by variables such as teachers' epistemological beliefs, openness to innovation, or professional experience. Against this background, we investigated differences between the PCK_{SR} of science teachers at elementary ($n = 113$) and secondary schools ($n = 201$), as well as teacher educators in biology education ($n = 41$), using a multiple-choice instrument. In addition, a regression analysis was conducted in order to investigate to what extent teachers' PCK_{SR} can be predicted by other variables (teachers' epistemological beliefs about transmission, their openness to innovation, and their professional experience). It was found that the groups differed significantly regarding their PCK_{SR}. In addition, the related variables predicted the PCK_{SR} for each group differently. The results not only provide fundamental insights into how the PCK_{SR} of different groups of teachers differs and how it can be predicted by related variables, but also on the development of PCK_{SR} for practicing teachers as well as the assessment of PCK in general.

Keywords: pedagogical content knowledge (PCK), scientific reasoning, elementary school teachers, secondary school teachers, university educators

²⁹ Lieberei, T., Welter, V. D. E. & Krell, M. (eingereicht). Pedagogical content knowledge (PCK) about scientific reasoning in different groups of science teachers and its relation to other variables. *Journal of Research in Science Teaching*.

7.1 Introduction

Scientific reasoning (SR) is a central goal of science education and describes the ability to solve scientific problems by applying scientific skills and knowledge (P. Kind & Osborne, 2017; Lawson, 2004). SR comprises different dimensions, such as planning and conducting investigations or using scientific models (Krell et al., 2023). Science teachers at elementary and secondary schools, as well as teacher educators at university, need to possess SR competencies themselves in order to promote students' SR competencies effectively (Khan & Krell, 2019; Schwartz et al., 2004). However, they must not only have these competencies themselves but also know how to teach SR to their students (Lieberei, Dawborn-Gundlach, et al., 2024). This knowledge about the conceptual ideas of learners and appropriate strategies for dealing with them is called pedagogical content knowledge (PCK). Alongside their pedagogical knowledge (PK; knowledge about the basic principles of education, teaching, learning, and the learners) and their content knowledge (CK; knowledge about the discipline to be taught), PCK is a central part of teachers' professional knowledge (V. Kind & Chan, 2019).

In recent decades, there has been a growing emphasis on researching PCK (van Driel et al., 2023). This research includes studies on PCK related to different scientific topics and practices, for example PCK about SR (PCK_{SR}; e.g., Hilfert-Rüppell et al., 2021; Lieberei, Großmann, et al., 2024) or PCK about specific dimensions of SR (e.g., PCK about modeling; Günther et al., 2019; Justi & van Driel, 2005). Overall, these studies indicate that science teachers possess low PCK_{SR} (or PCK about specific dimensions of SR), although it can be improved by professional development courses. However, comparing the results from different studies poses a difficulty because there seems to be a lack of valid instruments for the assessment of PCK_{SR} (Sannert, Baur, & Krell et al., 2024) and the studies differ with regard to the various components and layers of PCK. Furthermore, the existing studies were conducted with different samples (e.g., teachers at elementary schools: van Uum et al., 2021; teachers at secondary schools: Sannert, van Driel, & Krell, 2024; teacher educators: Lieberei et al., 2023). Nevertheless, because the different groups of teachers differ mainly regarding the age of the students they teach (i.e., students at the elementary, secondary, or university level), the curricular and learning goals the teachers pursue (e.g., Germany: KMK, 2020, 2024b; MBWK SH, 2019), as well as the university courses that constitute their teacher training (KMK, 2024a), it can be assumed that they also differ in their PCK_{SR}.

Against this background, we aimed to take a closer look at how the PCK_{SR} of science teachers at elementary and secondary schools, as well as of teacher educators at universities, differs. To examine this, their PCK_{SR} was assessed using a validated multiple-choice instrument (PCK_{SR}-bio; Lieberei, Großmann, et al., 2024). In addition, because teachers' PCK can be predicted by, for example, their epistemological beliefs (Großschedl et al., 2019), their openness to innovation (Suh & Park, 2017), or their professional experience (van Driel & Berry, 2019), we want to gain insights into how teachers' PCK_{SR} is related to these variables. With this approach, we first wanted to gain a basic understanding of PCK_{SR} as a construct with regard to differences between different groups of teachers as well as its relation to other variables. In addition, we aimed to draw implications regarding the assessment of PCK in general as well as the development of PCK_{SR} for practicing teachers.

7.2 Theoretical Background

The following section describes SR in science education, the importance of teachers' PCK, and the construct of PCK_{SR}.

7.2.1 Scientific Reasoning

SR³⁰ plays a major role in science education (Khan & Krell, 2019; Mathesius et al., 2016) and is therefore emphasized in science educational standards all around the globe (e.g., Australia: ACARA, 2022; Canada: Ministry of Education, 2008b, 2008a; Chile: Mineduc, 2015, 2019; Germany: KMK, 2020, 2024b; United States: NGSS Lead States, 2013). The term SR competencies describes the essential abilities required to successfully use specific skills and knowledge to solve scientific problems (P. Kind & Osborne, 2017; Lawson, 2004). It comprises knowledge related to science (declarative knowledge; “knowing what”), knowledge about scientific procedures and concepts (procedural knowledge; “knowing how”), and knowledge about epistemic constructs and values (epistemic knowledge; “knowing why”; P. Kind & Osborne, 2017). Hence, SR refers to empirical inquiry (e.g., planning and conducting investigations, using scientific models) and to theoretical thinking (e.g., generating hypotheses; Rönnebeck et al., 2016).

7.2.1.1 Teaching Scientific Reasoning

In science education, SR can be taught either implicitly (i.e., SR is not specifically addressed) or explicitly (i.e., SR is addressed in instructional approaches; Vorholzer et al., 2020). Even though it has been empirically shown that explicit instructional approaches can be more effective in teaching SR than implicit approaches (e.g., Matlen & Klahr, 2013), SR is often not addressed explicitly in science lessons (Vorholzer et al., 2022).

In recent years, the question of whether explicit approaches in form of direct or inquiry-based instruction is more suitable for developing SR competencies has been discussed (T. de Jong et al., 2023; Sweller et al., 2024; Zhang et al., 2022). Research has started to show that inquiry-based instruction, as long as it is guided by the teacher, leads to better learning outcomes than direct instruction (Aditomo & Klieme, 2020; Lazonder & Harmsen, 2016). At the same time, however, independent (open) inquiry often leads to poor learning outcomes (Aditomo & Klieme, 2020). Although science teachers with more solid views about inquiry-based instruction are more likely implement them in their teaching (Capps & Crawford, 2013a), many science teachers have misconceptions about how to give inquiry-based instruction, which results in them not implementing it adequately (Capps et al., 2016). Professional development courses can be used to help science teachers increase their professional knowledge about how to provide inquiry-based instruction (Capps & Crawford, 2013b; Sannert, Baur, & Krell, 2024).

Whether a student-centered approach in the form of (guided) inquiry-based learning or a teacher-centered approach in the form of direct instruction is chosen to teach SR depends partially on teachers' epistemological beliefs (Lotter et al., 2007). Teachers can have two fundamentally different beliefs about knowledge development in students (Schmeisser et al., 2013). While *transmission* describes the attempt to directly transfer the teachers' knowledge to their students, who take a passive role in the learning process, *constructivism* describes the self-directed acquisition of knowledge in a learning environment where students take an active role in the learning process (Schmeisser et al., 2013; Staub & Stern, 2002). These two epistemological beliefs therefore describe two opposites.

³⁰ SR is often used synonymously with other terms such as scientific inquiry. Even though both concepts describe comparable abilities and knowledge (Krell et al., 2022), in this manuscript, we use the term “inquiry-based instruction” to describe teaching methods that focus on the process of gaining scientific knowledge (Bicak et al., 2021). SR describes the abilities that students need to have in order to carry out these inquiry-based activities in science lessons.

Inquiry-based instruction is a student-centered and, thus, constructivist teaching approach (Minner et al., 2010). It has been shown that teachers sustainably implement inquiry-based instruction when it aligns with their beliefs, especially regarding students' conceptions and suitable instructional strategies (Suh & Park, 2017). Therefore, it can be assumed that teachers with more constructivist and less transmissive epistemological beliefs are more likely to implement inquiry-based instruction in their science lessons. Especially regarding teaching SR, previous studies have suggested that guided inquiry-based and constructivist teaching approaches, accompanied by explicit reflection on the inquiry process, are effective in developing students' SR competencies (Aditomo & Klieme, 2020; Engelmann et al., 2016; Lazonder & Harmsen, 2016). Hence, it can be expected that teachers with more constructivist epistemological beliefs possess more elaborate PCK_{SR} in an inquiry-based and constructivist way.

Furthermore, teachers need to be open to innovation in order to teach in a student-centered way with inquiry-based instruction instead of traditional teacher-centered approaches (Kelly & Staver, 2005; Suh & Park, 2017). It can be assumed that teachers' openness to innovation also has an influence on the instructional strategies selected and, thus, on their PCK_{SR}.

7.2.2 Pedagogical Content Knowledge

As mentioned above, PCK is seen as a central part of teachers' professional knowledge (V. Kind & Chan, 2019). Since PCK was introduced by Shulman (1986, 1987), various authors have differentiated it further (e.g., Carlson et al., 2019; Magnusson et al., 1999; Park & Oliver, 2008). A more recent definition specifies PCK as "the knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes" (Gess-Newsome, 2015, p. 36). PCK develops over time mainly as a result of practical teaching experience as well as reflection on it (van Driel & Berry, 2019), which implies that teachers' professional experience has an influence on their PCK. Teachers with higher PCK not only have a deeper understanding of how to teach a specific topic but also the ability to transfer their knowledge flexibly across different teaching situations (i.e., a higher consistency of PCK; Hogan et al., 2003; Lieberei et al., 2023). Hence, teachers with a high level of PCK recognize deep structures of learning processes (i.e., the procedure required to solve a problem, for example students' conceptions and suitable instructional strategies to deal with them), even if these processes differ in terms of surface features (e.g., the topic of the lesson; Chi & VanLehn, 2012).

Although PCK models often encompass several PCK components (e.g., there are five components in the PCK pentagon model; Park & Oliver, 2008), the two components *knowledge of students' understanding* (KSU) and *knowledge of instructional strategies* (KISR) are regarded as central components of PCK (van Driel et al., 2023); this has been empirically shown various times (Alonzo et al., 2012; Jin et al., 2015; Park & Chen, 2012).

A common model that is used to describe PCK is the refined consensus model, in which a distinction is made between different (but connected) layers of knowledge and experience (Carlson et al., 2019). The layer *enacted PCK* (ePCK) describes the PCK an individual teacher shows in a specific teaching situation, depending on the prior knowledge of the learning group, the learning goal, as well as other factors that influence teaching (Carlson et al., 2019). A teaching situation comprises planning, teaching, and reflecting, which is why ePCK can also be differentiated for these actions (Alonzo et al., 2019). For every teaching situation that is different from others and cannot be repeated, teachers show different ePCK (Alonzo et al., 2019; Behling et al., 2022b). Therefore, ePCK is situation-specific and cannot be described as independent of a specific situation (Behling et al., 2022a).

The entire PCK possessed by an individual teacher is described as their *personal PCK* (pPCK). This knowledge can be acquired through teaching experience and interactions with others (e.g., colleagues;

Carlson et al., 2019). During teaching, when a teacher acquires new ePCK, this potentially increases or modifies their pPCK (Alonzo et al., 2019; Irmer et al., 2023), and vice versa, teachers apply parts of their pPCK as ePCK in a teaching situation (Behling et al., 2022b; Irmer et al., 2023). Each teacher possesses unique pPCK, which includes their explicit knowledge (which can be verbalized) and their implicit knowledge (e.g., experiential knowledge; Alonzo et al., 2019).

Collective PCK (cPCK) describes the overall existing PCK shared in the professional community (Carlson et al., 2019). It includes current knowledge about student prior knowledge about various topics and suitable instructional strategies for dealing with them (Carpendale & Hume, 2019). Because cPCK is more than one person can possess, teachers take up parts of cPCK to build their own pPCK (Behling et al., 2022a; Irmer et al., 2023). While cPCK is typically defined as shared PCK in the professional community (Carlson et al., 2019), some authors describe cPCK as the knowledge that a group of teachers, for example, colleagues from the same school, collectively possesses and shares through interactions (Buldu & Buldu, 2021; Carpendale & Hume, 2019). Hence, a consensus seems to be missing about the appropriate reference group of cPCK.

Carlson and Daehler (2019) describe cPCK as a continuum ranging from discipline-specific to topic-specific knowledge. While topic-specific knowledge describes PCK about a specific topic (e.g., evolution or electricity; Behling et al., 2022a; Irmer et al., 2023), discipline-specific knowledge describes PCK about comprehensive practices (e.g., nature of science or SR). Some authors refer to the latter as PCK about science practices (Kenyon et al., 2011) or PCK about topic-specific activities (Magnusson et al., 1999). This continuum from discipline-specific to topic-specific PCK can also be applied to the other layers of PCK (i.e., pPCK and ePCK; Carlson et al., 2019).

A further distinction is made between dynamic (also called procedural) and declarative (also called static) PCK (Alonzo & Kim, 2016; Barendsen & Henze, 2019; Jüttner et al., 2013; Schmelzing et al., 2013; She, 2023; She & Chan, 2023). While dynamic PCK describes situated PCK that is shown in a specific situation (“knowledge in action”), declarative PCK describes a static form of PCK that a person has in total (“knowledge on action”). In relation to the refined consensus model, this means that ePCK (regarding planning, teaching, and reflecting) describes more dynamic PCK because it occurs in a specific situation, while pPCK and cPCK describe more declarative knowledge, which a teacher or a group of professionals holds (Alonzo et al., 2019; She, 2023).

7.2.2.1 Assessment of PCK

When choosing suitable instruments to assess PCK, the PCK layer should be considered. To assess teachers’ ePCK, instruments are needed that assess dynamic PCK that is shown in a specific teaching situation, for example, video-based analysis (e.g., Alonzo et al., 2012). In contrast, paper-pencil tests (with open-ended or closed items) are more suitable for assessing declarative and static PCK as they can be used to assess teachers’ PCK more broadly (i.e., pPCK or cPCK). An overview of potential instruments for assessing different layers of PCK is given in Table 7.1. As the aim of this study was to compare the PCK_{SR} of different groups of science teachers, it was most suitable to assess their pPCK_{SR}. In the following, we refer to pPCK_{SR} as PCK_{SR}.

Table 7.1: Potential assessment methods or instruments for analyzing different layers of PCK

Layer		Form	Potential Assessment Methods or Instruments	Example References
ePCK	Planning	Dynamic	Analysis of lesson plans	e.g., Großmann & Krüger (2023)
	Teaching	Dynamic	Video-based lesson observations	e.g., Alonzo et al. (2012)
	Reflection	Dynamic	Interviews	e.g., Park & Suh (2019)
pPCK		Declarative	Open-ended paper-pencil tests	e.g., Loughran et al. (2004)
			Closed paper-pencil tests	e.g., Lieberei, Großmann, et al. (2024)
			Paper-pencil tests with mixed task formats	e.g., Großschedl et al. (2019)
cPCK ¹		Declarative	Paper-pencil tests completed by groups of professionals	e.g., Buldu & Buldu (2021)

Note. ¹This depends on the definition of cPCK. If cPCK is understood as the PCK a group of professionals holds (e.g., Buldu & Buldu, 2021; Carpendale & Hume, 2019), it can be captured by paper-pencil tests. If cPCK is understood as the overall existing PCK of the professional community (e.g., Behling et al., 2022a, 2022b), it cannot be assessed validly at all.

Paper-pencil tests can be divided into instruments with open-ended items, instruments with closed items, and instruments with mixed task formats. While instruments with open-ended items can capture a broader range of cognitions on what respondents think (Martinez, 1999), instruments with closed items (e.g., multiple-choice instruments) have advantages in terms of objectivity and test economy (Haladyna, 2004). However, the use of multiple-choice instruments to assess PCK is seen particularly critically. Some authors emphasize that multiple-choice instruments often over- or underrepresent the construct of PCK (Smith & Banilower, 2015). Other points of criticism are the difficulty of developing attractors that are definitely correct while developing distractors that are definitely incorrect (Schmelzing et al., 2013). Furthermore, teachers may have teaching experience with a topic that is not covered by the answer options and which is, therefore, not assessed. In such cases, the respective instrument would not provide a valid indicator of the individual teachers' PCK (Schmelzing et al., 2013). Because of such criticisms, some researchers exclude multiple-choice instruments from their studies (e.g., Rosenkränzer et al., 2017). However, there are multiple-choice PCK instruments for which the test score interpretation has been comprehensively validated and which are, therefore, suitable for assessing teachers' PCK (e.g., Lieberei, Großmann et al., 2024; Sadler et al., 2013).

7.2.3 Pedagogical Content Knowledge About Scientific Reasoning

Based on the definition of PCK (Gess-Newsome, 2015), PCK_{SR} is the knowledge of, reasoning behind, and planning required to teach SR in different educational settings to specific students to improve their SR competencies (Lieberai, Großmann, et al., 2024). PCK_{SR} includes knowledge about students' prior knowledge related to SR and suitable instructional strategies to teach SR (e.g., how to foster students' procedural and epistemic knowledge about experimentation or how to use models not only as visualizations, but also as epistemic tools; Lieberei, Großmann, et al., 2024).

Some studies have already investigated PCK_{SR}. However, they (1) differ in their samples (e.g., preservice science teachers in Hilfert-Rüppell et al., 2021; in-service elementary school teachers in van Uum et al., 2021; in-service secondary school teachers in Sannert, van Driel, & Krell, 2024; teacher educators in Lieberei et al., 2023), (2) target specific dimensions of SR (e.g., PCK about modeling in Günther et al., 2019), (3) refer to different layers of PCK (e.g., ePCK in O. de Jong & van der Valk, 2007; pPCK in Lieberei, Großmann, et al., 2024), and (4) make use of various instruments and other assessment methods (e.g., observations in O. de Jong & van der Valk, 2007; open-ended instruments in Sannert, van Driel, & Krell, 2024; multiple-choice instruments in Lieberei, Großmann, et al., 2024). These differences make a comparison of the findings, especially between different groups of teachers (elementary and secondary school teachers as well as teacher educators at university), very difficult. However, because these groups of teachers differ regarding the age of the students and the content they teach as well as their teacher training, it can be assumed that they also differ in their PCK_{SR}.

7.2.3.1 Differences Between the Groups of Teachers

Elementary and secondary school teachers as well as the teacher educators at universities mainly differ regarding the age of their students (from children at a young age at elementary schools, through adolescents and young adults at secondary schools, to adults at universities). The students they teach therefore differ regarding their conceptions as well as their prior knowledge about SR. Therefore, the teacher groups differ in terms of the curricular and learning goals they pursue and, hence, in the content they teach.

In Germany, the context of this study, there are national educational standards for science subjects such as biology at secondary school level (KMK, 2020, 2024b) and for teacher education (KMK, 2024a), as well as federal educational standards for elementary school teachers (e.g., Schleswig-Holstein: MBWK SH, 2019). While the learning standards for elementary school intend to give students an initial contact with scientific methods (e.g., MBWK SH, 2019), the secondary school learning standards expect students to progressively build on these skills and develop their SR competencies (KMK 2020, 2024b). In contrast, the teacher education standards focus not only on developing preservice science teachers' CK (e.g., knowledge regarding SR), but also explicitly on their PK and PCK (KMK, 2024a).

Another difference between the teacher groups in the German context is that teachers at elementary schools teach the subject "Sachunterricht", which includes the science subjects as well as social science subjects (e.g., history). This also means that elementary school teachers do not necessarily need to have studied a natural science subject to teach "Sachunterricht". In contrast, secondary school teachers teach the subjects they studied (i.e., biology, chemistry, or physics) or integrated science as a subject for lower secondary level students.

For this reason, elementary and secondary school teachers also need to complete different university courses as part of their teacher training programs, which differs in terms of both subject and pedagogical content (KMK, 2024a). While university courses for elementary school teachers focus more on PK, university courses for secondary school teachers have a stronger focus on CK (KMK, 2024a). It is also possible to work as a teacher without having previously studied to become a teacher (i.e., as a career changer), but it is then necessary to acquire the basics of PK and PCK during retraining.

Teacher educators have usually studied to become school teachers, but they stayed at university or came back after their in-service training to work as researchers and teacher educators. In Germany, it is also possible to work as a teacher educator in biology education with a science degree (and not a degree in biology/science education).

Because of these differences between the groups of teachers regarding the age of the students being taught, the different learning standards, and the different university courses in teacher training, it can be assumed that they also differ regarding their PCK_{SR}.

7.2.4 Aim of the Study and Hypotheses

Against this background, we wanted to examine how the PCK_{SR} of different science teachers (at elementary school, secondary school, and university) differs. In addition, we wanted to investigate how the age of the students being taught as well as other variables predict the PCK_{SR} of each of the three groups. For reasons concerning test economy, we decided to focus on three variables for which it has already been shown that they have an influence on PCK and SR: teachers' epistemological beliefs about transmission (Capps & Crawford, 2013a; Großschedl et al., 2019), their openness to innovation (Kelly & Staver, 2005; Suh & Park, 2017), and their professional experience (Carlson et al., 2019; van Driel & Berry, 2019). With this study, we first wanted to explore and understand PCK_{SR} as a construct. We hoped to be able to draw implications from these findings not only for group differences and relations to other variables, but also regarding the assessment of PCK in general and the development of practicing teachers' PCK_{SR} .

To investigate the PCK_{SR} of different groups of teachers and its relation to other variables, the following two research questions (RQ) were addressed:

RQ1: How does the PCK_{SR} of elementary school teachers, secondary school teachers, and teacher educators differ?

RQ2: To what extent can teachers' PCK_{SR} be predicted by their epistemological beliefs about transmission, their professional experience, and their openness to innovation?

To address RQ1, we used the multiple-choice instrument PCK_{SR} -bio to assess the teachers' $pPCK_{SR}$ (see Section 7.3.3). This instrument encompasses 12 items. In each item, a short teaching scenario is described, and the participants need to identify the most appropriate instructional strategy (for the items related to KISR) or the most likely student conception (for the items related to KSU). The correct answer options in the KISR items describe guided inquiry-based and constructivist teaching approaches, which have been shown to be effective in developing students' SR competencies, and the correct answer options in the KSU items refer to most typical student conceptions found in empirical studies.

Because elementary school teachers teach the subject "Sachunterricht" and differ in their teacher training as well as in the curricular and learning goals they pursue (KMK, 2020, 2024b; MBWK SH, 2019), we assumed that secondary school teachers are more familiar with the addressed instructional strategies and students' conceptions. Therefore, we expected secondary school teachers to outperform elementary school teachers. At the same time, we expected teacher educators to outperform secondary school teachers because they are usually more aware of the current knowledge of the research community (cPCK) and therefore likely have higher $pPCK_{SR}$ (Behling et al., 2022b; Chan & Hume, 2019; Irmer et al., 2023). In addition, we expected the groups of teachers with a higher PCK_{SR} test score to have a deeper understanding of teaching SR and the ability to transfer this knowledge to the teaching scenarios described in the PCK_{SR} -bio (see Hogan et al., 2003), which leads to them showing a more consistent PCK. This has already been demonstrated for the PCK_{SR} of preservice science teachers and teacher educators (Lieberei et al., 2023). Summarizing, our hypotheses (H) are as follows:

H1.1: Science teachers have higher PCK_{SR} the older their students are (i.e., $\text{test score}_{\text{elementary teachers}} < \text{test score}_{\text{secondary teachers}} < \text{test score}_{\text{teacher educators}}$).

H1.2: Science teachers have a more consistent PCK_{SR} the older their students are (i.e., $\text{internal consistency}_{\text{elementary teachers}} < \text{internal consistency}_{\text{secondary teachers}} < \text{internal consistency}_{\text{teacher educators}}$).

To address RQ2, we assessed teachers' epistemological beliefs about transmission and their openness to innovation using validated instruments, as well as their professional experience operationalized in years of teaching experience using a rating scale (see Section 7.3.4). Teachers with rather transmissive epistemological beliefs were expected to work in a less student-centered way when promoting SR competencies (Capps & Crawford, 2013a), which is why we expected their epistemological beliefs about transmission to be a significant predictor (negative) of their test scores. This has already been shown for PCK in biology (Großschedl et al., 2019). At the same time, because we expected teachers who are open to innovative teaching approaches to be more open to inquiry-based instructions and, therefore, PCK_{SR} (Kelly & Staver, 2005; Suh & Park, 2017), we expected teachers' openness to innovation to be a significant predictor (positive) of their PCK_{SR}. We also expected the teachers to develop PCK_{SR} over time based on their teaching experience, their interactions with colleagues, and their work in research (Carlson et al., 2019; van Driel & Berry, 2019), and, therefore, we expected their professional experience to be a significant predictor (positive) of their PCK_{SR}. However, in a previous study with the PCK_{SR}-bio, no positive correlation was found between preservice science teachers' test scores and their educational level (Lieberei, Großmann, et al., 2024). This led to the following hypotheses:

H2.1: Teachers' epistemological beliefs about transmission are a significant predictor of their PCK_{SR} (based on a negative relation).

H2.2: Teachers' openness to innovation is a significant predictor of their PCK_{SR} (based on a positive relation).

H2.3: Teachers' professional experience is a significant predictor of their PCK_{SR} (based on a positive relation).

7.3 Methods

The following section gives an overview of the data collection, the sample, the instruments used, and the data analysis. Our study was reviewed and approved by our local ethics committee before any data collection took place (ethics ID: 2023_19_LI).

7.3.1 Data Collection

The data collection took place between June 2023 and January 2024 in five federal states in the north, south, and middle of Germany. All participants were contacted via e-mail. Participation was voluntary and anonymous, the participants took part in the study by completing an online questionnaire (Leiner, 2019). As a thank-you for their participation, teachers received free teaching materials about teaching SR and individual feedback about their PCK_{SR} on request. The participants could quit the questionnaire at any time, which is why not all constructs received the same number of responses.

7.3.2 Sample

A total of 355 biology teachers participated in the study ($n = 113$ elementary school teachers, $n = 201$ secondary school teachers, $n = 41$ teacher educators in biology education teacher training programs at university). Most participants identified as female ($n = 104$ for elementary school teachers, $n = 155$ for secondary school teachers, $n = 31$ for teacher educators). The majority of elementary and secondary school teachers (more than 90%) had completed in-service training (18-month training after their Master's degree where the preservice teachers already teach at schools to become a certified teacher; Neumann et al., 2017). The number of teacher educators who had completed in-service training was much lower (34%). A more detailed overview of the sample is shown in Table 7.2.

Table 7.2: Overview of the sample regarding the proportion of teachers who had completed in-service training, had a PhD, and were teaching or had studied another science subject

	<i>N</i>	In-Service Training	PhD ¹	Second Science Subject ²
Elementary School Teachers	113	106 (94%)	3 (3%)	0 ³
Secondary School Teachers	201	182 (91%)	23 (13%)	100 (50%)
Teacher Educators	41	14 (34%)	34 (83%)	5 (12%)
Total	355	302	59	105

Note. ¹Includes completed and ongoing PhD projects (subject-related or pedagogical-related focus).

²Physics or chemistry. ³Because elementary school teachers in Germany teach the subject “Sachunterricht”, which includes all science subjects, they do not teach one specific second science subject.

Some secondary school teachers taught mainly at lower secondary level (Grades 5—10; $n = 53$), some mainly at higher secondary level (Grades 11–13; $n = 69$), and the rest was balanced between the two levels ($n = 72$). Some of the teacher educators had already conducted research on PCK ($n = 12$), a smaller proportion had conducted research on SR but not on PCK ($n = 3$), few had conducted research on both PCK and SR ($n = 3$), and around half of the teacher educators had not conducted research on either of these two constructs ($n = 23$).

7.3.3 Assessment of PCK: PCK_{SR-bio} by Lieberei, Großmann, et al. (2024)

In lines with the aims of this study, in order to assess the PCK_{SR} of different teachers, the PCK_{SR-bio} multiple-choice instrument was used. The instrument was developed and evaluated by Lieberei, Großmann, et al. (2024) and includes items related to the PCK components KISR, KSU, as well as their connections ($KISR \rightarrow KSU$, $KSU \rightarrow KISR$). For each of the PCK components and their connections, one item was developed for each of the SR activities (1) planning and conducting investigations, (2) generating hypotheses, and (3) using scientific models. This led to a total of 12 items. An overview of the framework is provided in Table 7.3.

Table 7.3: Systematic overview of the PCK_{SR-bio}

	Planning and Conducting Investigations	Generating Hypotheses	Using Scientific Models
KISR	Item #1	Item #2	Item #3
KSU	Item #4	Item #5	Item #6
$KISR \rightarrow KSU$	Item #7	Item #8	Item #9
$KSU \rightarrow KISR$	Item #10	Item #11	Item #12

Each item of the PCK_{SR-bio} addresses biology-related classroom situations, describing information about the learning group and the topic of the lesson. A question about appropriate instructional strategies or students’ knowledge must be answered (one attractor; two distractors; Rodriguez, 2005). The items were designed from the perspective that student-centered inquiry-based instruction is most effective in teaching SR competencies. The proportion of correct answers (i.e., 1.0 = 100% correct answers) was calculated to indicate teachers’ PCK_{SR} .

7.3.4 Assessment of Other Variables

Against the background of RQ2, teachers' epistemological beliefs about transmission, their openness to innovation, and their professional experience were assessed with the rating scales described below.

7.3.4.1 Teachers' Epistemological Beliefs about Transmission: *Q4TB* by Schlichter (2012)

A subscale from Schlichter's (2012) instrument was used to assess teachers' epistemological beliefs about transmission. The scale comprises 10 items that have to be answered on a 5-point rating scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither disagree nor agree*, 4 = *agree*, 5 = *strongly agree*). The mean score was calculated afterwards.

7.3.4.2 Teachers' Openness to Innovation: *FPIB* by Waßmann (2014)

To assess the teachers' openness to innovation, an instrument by Waßmann (2014) was used. The *FPIB* was designed to assess the psychological obstacles to innovation, which describes personal capabilities that influence how much resistance is put up against initiated innovation processes (psychological and socio-sociological obstacles; Waßmann, 2014). It comprises 40 items, whereby one item was omitted in this study after consultation with the developer due to a questionable wording. The items were answered on a 5-point rating scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither disagree nor agree*, 4 = *agree*, 5 = *strongly agree*). In order to measure teachers' openness to innovation, which describes the opposite of the psychological blockage of innovation, the items were inverted. The mean value was calculated afterwards.

7.3.4.3 Teachers' Professional Experience

In order to investigate their professional experience, we asked elementary and secondary school teachers to state how long they had been working at schools for on a 4-point rating scale (1 = *less than 3 years*; 2 = *3–5 years*; 3 = *5–10 years*, 4 = *more than 10 years*). Teacher educators also provided their professional experience in years, which was later recoded in correspondence with the 4-point scale to have the same ordinally scaled variable for all groups.

7.3.5 Data Analysis

Regarding RQ1, the Shapiro-Wilk test revealed that the PCK_{SR} test score was normally distributed for teacher educators ($p > .05$), but not for elementary and secondary school teachers. Therefore, a graphical analysis was conducted additionally. The Q-Q plots, as well as the skew and kurtosis, indicated that the test score for the other two groups was also (close to) normally distributed. There was a homogeneity of variance (Levene's test, $p > .05$), which is why a one-way ANOVA was conducted. A Tukey post hoc analysis was then used to test H1.1. The internal consistency was calculated using the Kuder-Richardson formula 20 (KR20) and was compared to test H1.2. We used the R package *cocron* (Diedenhofen & Musch, 2016) to investigate whether the differences are significant.

Regarding RQ2, teachers' epistemological beliefs about transmission, as well as their openness to innovation, were normally distributed for all groups as assessed by the Shapiro-Wilk test and graphical analyses. There was a homogeneity of variance (Levene's test, $p > .05$) for both variables, which is why a one-way ANOVA was conducted. A Pearson-correlation analysis between teachers' epistemological beliefs about transmission, their openness to innovation, and their PCK_{SR} test score, as well as a Spearman correlation between teachers' professional experience, their test score, and the other related variables, was carried out to further explore the relationship between the variables. As no significant correlation was found between teachers' professional experience and their test scores (see H2.3), multiple linear regression analyses focused on testing H2.1 and H2.2, that is, to investigate the extent to which the PCK_{SR} of the different groups of teachers could be predicted by their epistemological beliefs about transmission and their openness to innovation.

All statistical analyses were conducted with IBM SPSS Statistics 28, and the effect size measures were calculated using an online tool by Lenhard and Lenhard (2022) and interpreted based on Cohen (1988).

7.4 Results

In the following, the results of the statistical analyses are presented separately for the two research questions.

7.4.1 How Does the PCK_{SR} of Elementary School Teachers, Secondary School Teachers, and Teacher Educators Differ?

The elementary school teachers had the lowest PCK_{SR} test score ($M = .47$, $SD = .14$), followed by the secondary school teachers ($M = .52$, $SD = .16$) and the teacher educators ($M = .62$, $SD = .18$; Table 7.4). The one-way ANOVA revealed that the test score was significantly different between the groups, $F(2,352) = 13.81$, $p < .001$, $d = 0.94$ (large effect). Tukey post hoc analyses revealed that the test score of the teacher educators was significantly higher than that of the secondary school teachers ($p < .001$), which in turn was significantly higher than the elementary school teachers' test score ($p = .022$). These findings, therefore, supported our assumption that science teachers have a higher level of PCK_{SR} as the age of their students increases (H1.1).

Table 7.4: Overview of the PCK_{SR} test score and the internal consistency of the groups

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>KR20</i> ¹
Elementary School Teachers	113	.47	.14	.15
Secondary School Teachers	201	.52	.16	.33
Teacher Educators	41	.62	.18	.45

Note. ¹Kuder-Richardson formula 20.

With regard to H1.2, it was found that the teacher educators showed a higher internal consistency in their answers ($KR20 = .45$) than the secondary school science teachers ($KR20 = .33$), who—in turn—showed a higher internal consistency in their answers than the elementary school teachers ($KR20 = .15$). However, these differences are statistically not significant, $\chi^2(2) = 2.88$, $p = .23$. Therefore, our assumption that science teachers have more consistent PCK_{SR} as the age of their students increases (H1.2) can only be supported on a descriptive level.

7.4.2 To What Extent Can Teachers' PCK_{SR} Be Predicted by Their Epistemological Beliefs About Transmission, Their Professional Experience, and Their Openness to Innovation?

In order to answer RQ2, three variables were considered: teachers' epistemological beliefs about transmission, their openness to innovation, and their professional experience.

An overview of the participants' epistemological beliefs about transmission (a high score implies rather transmissive epistemological beliefs) and openness to innovation (a high score implies a high level of openness to innovation) in each group is given in Table 7.5. The groups differed significantly regarding their epistemological beliefs about transmission, $F(2,334) = 13.61$, $p < .001$, $d = 0.68$ (large effect). Tukey post hoc analyses revealed that the epistemological beliefs of the elementary school teachers was significantly higher than those of the university teachers ($p < .001$) and secondary school teachers ($p < .001$). However, the differences between secondary school teachers and university teachers were

not statistically significant ($p = .740$). Furthermore, the groups did not differ regarding their openness to innovation, $F(2,338) = 2.31$, $p = .680$.

Table 7.5: Overview of the groups' epistemological beliefs about transmission and openness to innovation

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>KR20</i>
Epistemological Beliefs about Transmission				
Elementary School Teachers	110	4.19	0.57	.86
Secondary School Teachers	190	3.81	0.70	.89
Teacher Educators	37	3.73	0.67	.87
Openness to Innovation				
Elementary School Teachers	109	3.50	0.37	.87
Secondary School Teachers	186	3.51	0.40	.87
Teacher Educators	39	3.65	0.37	.87

Regarding their professional experience, most elementary and secondary school teachers had been working for more than 10 years ($n_{\text{elementary school teachers}} = 67$ and $n_{\text{secondary school teachers}} = 117$). The number of teacher educators who had been working for more than 10 years ($n = 13$) was similar to the number that had been working for between 5 and 10 years ($n = 14$). An overview is given in Figure 7.1. There was a statistically significant difference in their professional experience between the groups of teachers, $\chi^2(6) = 17.35$, $p = .008$, $\phi = .16$ (small effect).

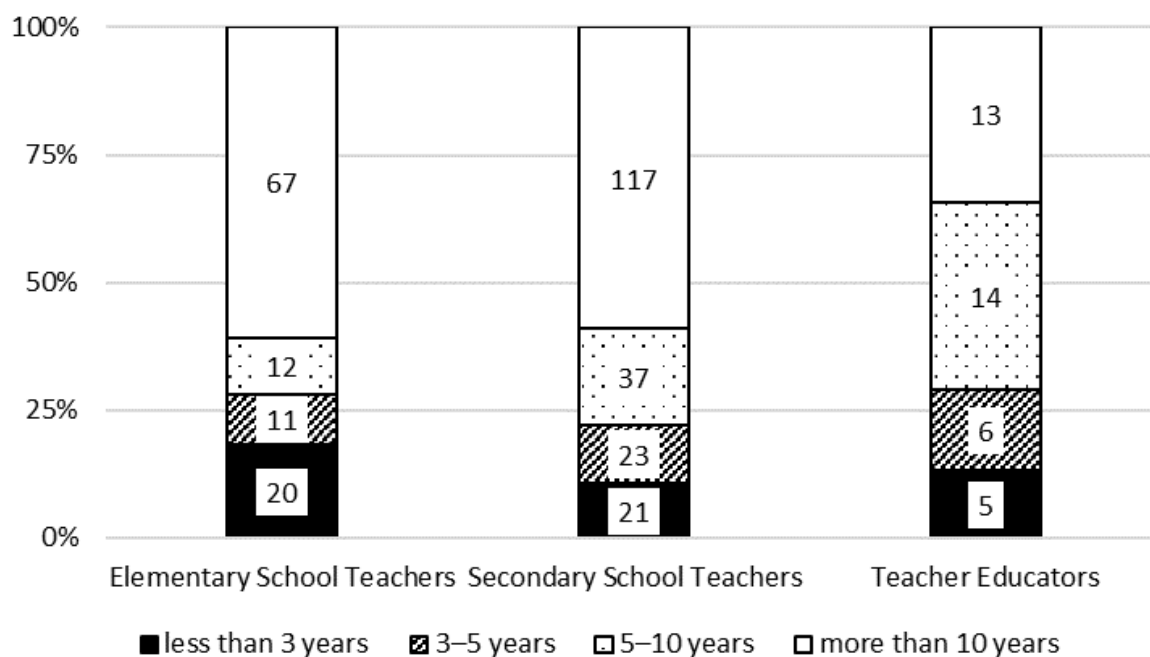


Figure 7.1: Overview of the professional experience of the different groups

A correlation analysis between the related variables and the PCK_{SR} test score was conducted for each group. For the elementary school teachers and the teacher educators, a significant negative correlation between epistemological beliefs about transmission and the PCK_{SR} test score was observed; this was marginally significant for the secondary school teachers (i.e., $p = .054$). A significant negative correlation between their openness to innovation and the PCK_{SR} test score was observed for the secondary school teachers. An overview is shown in Table 7.6.

Table 7.6: Correlation analysis for the related variables and the PCK_{SR} test score for each group

Group	Variable	1.	2.	3. ¹	4.
Elementary School Teachers	1. Epistemological Beliefs about Transmission	-			
	2. Openness to Innovation	-.15	-		
	3. Professional Experience ¹	-.04	.21*	-	
	4. PCK _{SR} Test Score	-.20*	.03	.07	-
Secondary School Teachers	1. Epistemological Beliefs about Transmission	-			
	2. Openness to Innovation	-.04	-		
	3. Professional Experience ¹	.12	.22**	-	
	4. PCK _{SR} Test Score	-.12 ²	-.17*	-.03	-
Teacher Educators	1. Epistemological Beliefs about Transmission	-			
	2. Openness to Innovation	-.14	-		
	3. Professional Experience ¹	.01	-.18	-	
	4. PCK _{SR} Test Score	-.42**	.10	-.13	-

Note. ¹Spearman correlation. ² $p = .054$. * $p < .05$ (one-tailed). ** $p < .01$ (one-tailed).

Because there was no significant correlation between teachers' professional experience and their PCK_{SR} test score, we could not expect this related variable to be a predictor of PCK_{SR}, which is why we did not include it in the regression models. Therefore, H2.3 had to be rejected. A multiple linear regression was then conducted for each of the three groups in order to test how the two remaining variables (teachers' epistemological beliefs about transmission and their openness to innovation) predict the teachers' PCK_{SR}. There was no violation of our assumptions. The models had no autocorrelation, as indicated by the value of the Durbin-Watson statistic (1.89 for elementary school teachers, 1.85 for secondary school teachers, and 1.65 for teacher educators). In addition, there was no multicollinearity, as indicated by a variance inflation factor (VIF) lower than 10 (Field, 2013). The multiple linear regression analysis resulted in a significant model for the group of secondary school teachers, $F(2,186) = 4.41$, $p = .013$, $adj. R^2 = .05$, and teacher educators, $F(2,34) = 3.86$, $p = .036$, $adj. R^2 = .18$, but not for the group of elementary school teachers, $F(2,105) = 2.48$, $p = .088$. An overview is shown in Table 7.7.

Table 7.7: Regression coefficients of PCK_{SR} test score for the different groups

Variable	<i>B</i>	<i>SE</i>	β	<i>p</i>	VIF
Elementary School Teachers (<i>adj. R</i> ² = .03)					
Constant	0.708	0.174		< .001	
Epistemological Beliefs about Transmission	-0.053	0.024	-.214	.029	1.02
Openness to Innovation	-0.003	0.036	-.008	.935	1.02
Secondary School Teachers (<i>adj. R</i> ² = .05)					
Constant	0.871	0.117		< .001	
Epistemological Beliefs about Transmission	-0.027	0.016	-.121	.094	1.00
Openness to Innovation	-0.070	0.028	-.180	.013	1.00
Teacher Educators (<i>adj. R</i> ² = .18)					
Constant	1.128	0.343		.002	
Epistemological Beliefs about Transmission	-0.109	0.040	-.426	.010	1.02
Openness to Innovation	-0.023	0.078	-.046	.769	1.02

Teachers' epistemological beliefs about transmission were a significant predictor (negative) for the group of elementary school teachers ($\beta = -.214$) as well as teacher educators ($\beta = -.426$), which supported our assumption (H2.1). However, for these two groups, teachers' openness to innovation was not a significant predictor. In contrast, teachers' openness to innovation was a significant predictor (negative) of secondary school teachers' PCK_{SR} ($\beta = -.180$), which was contrary to our assumption (H2.2).

To examine the extent to which the related variables predicted teachers' PCK_{SR} for the total sample when the age of their students was controlled for, we included the group as a dummy-coded variable in a multiple linear regression for the total sample (i.e., $N = 355$). We decided to include the groups of elementary school teachers and teacher educators as dummy variables because the group of secondary school teachers seemed to differ regarding the correlation between the test score and the other variables (see Table 7.6). Again, no violation of the assumptions was shown (Durban-Watson statistic: 1.88; VIF < 10; Field, 2013). The multiple linear regression analysis resulted in a significant model, $F(4,329) = 11.80$, $p < .001$, *adj. R*² = .12 (Table 7.8).

Table 7.8: Regression coefficients of PCK_{SR} test score including the groups of teachers as dummy-coded variables

Variable	<i>b</i>	<i>SE</i>	β	<i>p</i>	VIF
Constant	0.843	0.093		< .001	
Epistemological Beliefs about Transmission	-0.044	0.013	-.187	.001	1.09
Openness to Innovation	-0.044	0.021	-.106	.043	1.03
Elementary School Teachers (Dummy)	-0.033	0.019	-.096	.082	1.14
Teacher Educators (Dummy)	0.115	0.027	.227	< .001	1.08

Teachers' epistemological beliefs about transmission were again a significant predictor (negative) of their PCK_{SR}, with a standardized β of $-.182$, which supported our assumption (H2.1). However, contrary to our assumption (H2.2), in this model, teachers' openness to innovation was found to be a significant negative predictor of their PCK_{SR}, with a standardized β of $-.110$. Similar to our findings concerning RQ1, the teacher educators achieved a significantly higher test score than that of the secondary school teachers ($\beta = -.227$). The elementary school teachers achieved a lower test score than that of the secondary school teachers, which, however, was not statistically significant in this analysis.

7.5 Discussion

The aim of our study was to examine differences in the PCK_{SR} of elementary and secondary school teachers as well as teacher educators (RQ1) and to investigate to what extent teachers PCK_{SR} can be predicted by their epistemological beliefs about transmission, openness to innovation, and professional experience (RQ2).

Regarding RQ1, it was shown that the PCK_{SR} test score of the teacher educators was significantly higher than that of the secondary school teachers, who, in turn, had a significantly higher test score than that of the elementary school teachers. This supports our assumption (H1.1). As we assessed a part of teachers' pPCK that is related to their cPCK (Behling et al., 2022a; Irmer et al., 2023), this was an expected result. We expect teacher educators, in contrast to school teachers, to be more likely to be aware of current research findings about students' conceptions and suitable strategies for dealing with those conceptions (which is part of cPCK; Carlson et al., 2019; Carpendale & Hume, 2019) in their work. Because we used a multiple-choice instrument, which assesses declarative PCK, it can be assumed that this knowledge can be acquired through research and that teacher educators therefore perform better. At the same time, it can be assumed that teachers at elementary and secondary schools have higher ePCK about SR, as they have more practical experience than teacher educators. In addition, while teachers at elementary and secondary schools need to know how to teach SR, teacher educators are required to explicitly teach basics about students' conceptions and suitable instructional strategies (i.e., pPCK) to preservice science teachers (e.g., in Germany: KMK, 2024a).

The differences between elementary school teachers and secondary school teachers may be attributable to the differences between the groups regarding the content they teach as well as the curricular and learning goals they pursue. It can be assumed that elementary school teachers deal less intensively with the topics and SR activities addressed in the PCK_{SR}-bio in their teacher training (KMK, 2024a), which may result in them achieving a lower test score than in-service teachers. Furthermore, because elementary school teachers—in contrast to secondary school teachers—do not teach biology as a specific subject, but rather the subject “Sachunterricht”, it can be assumed that they also deal less intensively with the SR activities addressed in the instrument in their teaching. This is also supported by other empirical studies on the synergy of science subjects. These studies show that preservice biology teachers who study another science subject had a significantly higher PCK than preservice teachers with a second nonscientific subject because the subject content partly complement each other (Welter et al., 2022). With regard to our study, this supports our assumption because we expect teachers at secondary schools, who deal with SR more frequently and in more depth in their teacher training and teaching, to have a deeper understanding compared to teachers at elementary schools.

The differences in the PCK_{SR} of the different groups also become apparent with regard to the internal consistency of their answers to the PCK_{SR}-bio. Although not statistically significant, the internal consistency of the elementary school teachers was the lowest ($KR20 = .15$), followed by that of the secondary school teachers ($KR20 = .33$). The teacher educators had the highest internal consistency ($KR20 = .45$). Because we expected teachers with higher PCK_{SR} to have a deeper understanding of teaching SR and the ability to transfer this knowledge to the various teaching scenarios described in

the items of the PCK_{SR}-bio (see Hogan et al., 2003), our findings support our assumption (H1.2). Our findings are in line with previous findings on the internal consistency of preservice science teachers that showed an increasing internal consistency at higher stages of teacher education, although the test score did not differ between the groups in those previous findings (Lieberei, Großmann, et al., 2024). However, the internal consistency of their answers was, overall, relatively low, especially for elementary and secondary school teachers.

Internal consistency is typically regarded as a quality criterion that often leads to the exclusion of items that make the instrument fall below a certain limit (Taber, 2018). Especially for instruments that assess complex constructs, this approach has been criticized because although a high level of internal consistency is desirable, a valid interpretation of the test score is crucial (Cronbach, 1951). Therefore, it has been recommended that attention be paid to the causes of low internal consistency (Schmitt, 1996; Taber, 2018). Indeed, the consistency of response behavior can actually inform researchers about the nature of the construct and the proficiency of the respondents (Leach et al., 2000). Teachers with a higher PCK_{SR} are expected to have a deeper understanding of how to teach SR appropriately and also the ability to transfer this knowledge to different situations (Lieberei et al., 2023).

For the PCK_{SR}-bio, the validity of the test score interpretation has already been shown by considering multiple sources of validity evidence, including evidence based on test content, response processes, internal structure, and relations to other variables (Lieberei, Großmann, et al., 2024). Hence, despite the low internal consistency, especially for the groups of elementary and secondary school teachers in this study, a valid interpretation of the test score is possible. Our results show that teacher educators at universities not only have a higher PCK_{SR}, but also a better ability to flexibly apply their knowledge to different teaching situations compared to teachers at secondary school, who in turn have a better ability than teachers at elementary schools. Additionally, our results show that it is necessary to take a critical look at internal consistency. Furthermore, it can be fruitful to consider reasons of low internal consistency, which can provide insights into the nature of the construct and participants' response behavior.

Regarding RQ2, it first needs to be emphasized that the findings depend on our definition of PCK_{SR} and how it is operationalized in the PCK_{SR}-bio. As mentioned above, the correct answer options describe guided inquiry-based and constructivist teaching approaches, which have been shown to be effective in developing students' SR competencies, or typical student conceptions found in empirical studies, respectively. The correlation analysis revealed that there was no significant relationship between the professional experience and the PCK_{SR} test score of teachers in all three groups. Therefore, H2.3 had to be rejected. This finding contradicted our assumption that teachers develop PCK over time as a result of their teaching experience, their interactions with colleagues, and their work in research (Carlson et al., 2019; van Driel & Berry, 2019). However, previous findings have also suggested that a certain level of reflection skills is needed in order for teachers to develop pPCK from practical experience and reflect on it (Kulgemeyer et al., 2021). This also depends on the topic that is taught. Because teachers often do not promote SR skills in science lessons, even if they think they do (Capps et al., 2016), it is not surprising that their pPCK does not develop further, even with increasing professional experience. Similar results have already been shown for preservice science teachers, where no increase in PCK_{SR} could be shown for different stages of teacher education. This may be due to a lack of learning opportunities during their teacher training (Lieberei, Großmann, et al., 2024). Regarding teachers at elementary and secondary schools, this means that professional development courses are particularly necessary in order to effectively improve their PCK_{SR}. Empirical studies support the positive effects of professional development courses regarding SR competencies and PCK_{SR} (e.g., Capps & Crawford, 2013b; Sannert, van Driel, & Krell, 2024). For teacher educators, the lack of

correlation between their professional experience and their PCK_{SR} test score found in this study was possibly due to their research areas, as not all teacher educators dealt with PCK or SR in their work.

The multiple linear regression models showed that teachers' epistemological beliefs about transmission were a significant predictor (negative) of their PCK_{SR}, which supported our assumption (H2.1). However, teachers' openness to innovation was also a significant predictor (negative) of their PCK_{SR}, which was contrary to our assumption (H2.2). Furthermore, the related variables predicted the PCK_{SR} differently for the groups. As expected, teachers' epistemological beliefs about transmission were a significant predictor (negative) of the PCK_{SR} of elementary school teachers and teacher educators. For secondary school teachers, this was only seen on a descriptive level (Table 7.6). Because we expected teachers with those epistemological beliefs to teach in a less student-centered way when promoting SR competencies (Capps & Crawford, 2013a), this supports our assumption (H2.1). These findings are in line with previous studies investigating how teachers' epistemological beliefs about transmission predict PCK in biology (Großschedl et al., 2019). Regardless of the group, it can be seen that teachers' epistemological beliefs about transmission predict their PCK_{SR}. Therefore, when promoting PCK_{SR}, for example in professional development courses, teachers' epistemological beliefs should also be considered.

Openness to innovation was not a positive predictor of the PCK_{SR} of elementary school teachers and teacher educators, which did not support our assumption (H2.2). This could be because we assessed a part of teachers' pPCK. Teachers who are open to innovative teaching approaches could be expected to have more knowledge and experience regarding inquiry-based instruction (Kelly & Staver, 2005; Suh & Park, 2017). This could have a stronger influence on direct teaching (and thus their ePCK) than their theoretical knowledge base does (pPCK). Surprisingly, their openness to innovation was a significant predictor (negative) for secondary science teachers, in contrast to our assumption (H2.2). This could be due to the instrument used (see Section 7.3.4.2). The *FPIB* was designed to assess the psychological obstacles to innovation, which describes psychological as well as socio-sociological obstacles (Waßmann, 2014). Such obstacles could have an influence when selecting teaching methods after working with colleagues. Teachers at secondary schools often have misconceptions about SR and how to teach it (Capps et al., 2016; Sannert, van Driel, & Krell, 2024). If these teachers share their misconceptions with colleagues who are open to innovation, they may be more open to adopting these misconceptions, which could result in lower PCK_{SR}.

7.5.1 Limitations

When interpreting the results, some limitations have to be considered. First of all, the internal consistency was rather low, especially for elementary and secondary school teachers. As discussed above, from a traditional perspective, this would mean that the instrument should be discarded. However, in line with the opinions of several scholars, our results indicate that a low level of internal consistency does not necessarily undermine the validity of the test score interpretation (Schmitt, 1996; Taber, 2018).

Furthermore, the analysis on how the related variables predict teachers' PCK_{SR} explained around 12% of the total variance of the teachers' PCK_{SR} (medium effect size; Fritz et al., 2012). For reasons of test economy, we decided on the selected variables based on theoretical assumptions and previous empirical findings. However, it can also be assumed that there are other related variables that influence teachers' PCK (e.g., subject-specific interest; Großschedl et al., 2019) and therefore also influence teachers' PCK_{SR}. Further investigations are necessary to gain more insights into which variables are related to teachers' PCK_{SR} and whether they differ for different groups of teachers.

Finally, it should be noted that the findings in this study apply to PCK_{SR}, which is only a small part of teachers' PCK. Therefore, the results are not generalizable to other topics or practices, especially due to the topic-specificity of PCK. On the basis of our findings, we recommend that PCK is examined in relation to specific topics or practices when investigating both its relation to other variables and potential group differences.

7.5.2 Conclusions and Implications

With this study, we aimed to examine the differences in PCK_{SR} of different groups of teachers as well as how it is related to other variables. Beyond these fundamental findings, we can draw three implications.

First, our results show that it is fruitful to consider the internal consistency when comparing the PCK of different groups and not to automatically exclude the instrument if a certain limit is not reached. A low internal consistency is maybe due to the nature of the construct and the proficiency of the participants and, therefore, can provide information about the depth of the participants' knowledge. However, care should be taken to ensure the validity of the test score interpretation.

Furthermore, the multiple-choice instrument we used seem to be suitable to assess and compare the PCK_{SR} of different teachers. However, when interpreting studies about PCK, care should be taken about which layer of PCK and which topic is considered. Our results show that teacher educators' pPCK_{SR} was higher than elementary and secondary school teachers' pPCK_{SR}. For other layers of PCK (e.g., enacted PCK_{SR}), these findings may be different. At the same time, the age of the students taught, the curricular and learning goals the teachers pursue, as well as different university courses that constitute their teacher training should also be considered when comparing different groups of teachers as these factors may influence their PCK differently.

In addition, our results indicate that elementary and secondary school teachers, overall, have low PCK_{SR}, which, however, does not seem to improve as a result of practical teaching experience. For this reason, professional development courses seem particularly important in order to enable teachers to acquire these competencies. Regarding other related variables, further investigations are necessary to examine their influence on teachers' PCK_{SR} and whether this is different for different groups of teachers.

Acknowledgments

The authors would like to thank Stefan Waßmann for his support with the data analysis of the *FPIB* as well as Gráinne Newcombe and Anna Maria Rotermund for English language editing.

7.6 References

- ACARA (2022). *Science. F-10 Version 9.0. Scope and sequence*.
<https://v9.australiancurriculum.edu.au/content/dam/en/curriculum/ac-version-9/downloads/science/science-scope-and-sequence-f-10-v9.docx>
- Aditomo, A., & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes: Evidence from high and low-performing education systems. *International Journal of Science Education*, 42(4), 504–525.
<https://doi.org/10.1080/09500693.2020.1716093>

- Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: Enacted and personal PCK. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 273–288). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_12
- Alonzo, A. C., & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, 53(8), 1259–1286. <https://doi.org/10.1002/tea.21271>
- Alonzo, A. C., Kobarg, M., & Seidel, T. (2012). Pedagogical content knowledge as reflected in teacher-student interactions: Analysis of two video cases. *Journal of Research in Science Teaching*, 49(10), 1211–1239. <https://doi.org/10.1002/tea.21055>
- Barendsen, E., & Henze, I. (2019). Relating teacher PCK and teacher practice using classroom observation. *Research in Science Education*, 49(5), 1141–1175. <https://doi.org/10.1007/s11165-017-9637-z>
- Behling, F., Förtsch, C., & Neuhaus, B. J. (2022a). The refined consensus model of pedagogical content knowledge (PCK): Detecting filters between the realms of PCK. *Education Sciences*, 12(9), 592. <https://doi.org/10.3390/educsci12090592>
- Behling, F., Förtsch, C., & Neuhaus, B. J. (2022b). Using the plan–teach–reflect cycle of the refined consensus model of PCK to improve pre-service biology teachers' personal PCK as well as their motivational orientations. *Education Sciences*, 12(10), 654. <https://doi.org/10.3390/educsci12100654>
- Bicak, B. E., Borchert, C. E., & Höner, K. (2021). Measuring and fostering preservice chemistry teachers' scientific reasoning competency. *Education Sciences*, 11(9), 496. <https://doi.org/10.3390/educsci11090496>
- Buldu, E., & Buldu, M. (2021). Investigating pre-service early childhood teachers' cPCK and pPCK on the knowledge used in scientific process through CoRe. *SAGE Open*, 11(2). <https://doi.org/10.1177/21582440211025564>
- Capps, D. K., & Crawford, B. A. (2013a). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education*, 24(3), 497–526. <https://doi.org/10.1007/s10972-012-9314-z>
- Capps, D. K., & Crawford, B. A. (2013b). Inquiry-based professional development: What does it take to support teachers in learning about inquiry and nature of science? *International Journal of Science Education*, 35(12), 1947–1978. <https://doi.org/10.1080/09500693.2012.760209>
- Capps, D. K., Shemwell, J. T., & Young, A. M. (2016). Over reported and misunderstood? A study of teachers' reported enactment and knowledge of inquiry-based science teaching. *International Journal of Science Education*, 38(6), 934–959. <https://doi.org/10.1080/09500693.2016.1173261>
- Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., . . . Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (pp. 77–94). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_2
- Carpendale, J., & Hume, A. (2019). Investigating practising science teachers' pPCK and ePCK development as a result of collaborative CoRe design. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 225–252). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_10
- Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 3–76). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_1

- Chi, M. T. H., & VanLehn, K. A. (2012). Seeing deep structure from the interactions of surface features. *Educational Psychologist*, 47(3), 177–188.
<https://doi.org/10.1080/00461520.2012.695709>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
<https://doi.org/10.4324/9780203771587>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- de Jong, O. & van der Valk, A. E. (2007). Science teachers' PCK and teaching practice: Learning to scaffold students' open-inquiry learning. In R. Pintó & D. Couso (Eds.), *Contributions from science education research* (pp. 107–118). Springer. https://doi.org/10.1007/978-1-4020-5032-9_8
- de Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C. E., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K., & Zacharia, Z. C. (2023). Let's talk evidence – the case for combining inquiry-based and direct instruction. *Educational Research Review*, 39, 100536. <https://doi.org/10.1016/j.edurev.2023.100536>
- Diedenhofen, B. & Musch, J. (2016). cocron: A web interface and R package for the statistical comparison of cronbach's alpha coefficients. *International Journal of Internet Science*, 11(1), 51–60.
- Engelmann, K., Neuhaus, B. J., & Fischer, F. (2016). Fostering scientific reasoning in education – meta-analytic evidence from intervention studies. *Educational Research and Evaluation*, 22(5-6), 333–349. <https://doi.org/10.1080/13803611.2016.1240089>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology. General*, 141(1), 2–18.
<https://doi.org/10.1037/a0024338>
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). Routledge.
- Großmann, L., & Krüger, D. (2023). Identifying performance levels of enacted pedagogical content knowledge in trainee biology teachers' lesson plans. In G. S. Carvalho, A. S. Afonso, & Z. Anastácio (Eds.), *Fostering Scientific Citizenship in an Uncertain World: Vol. 13. Contributions from Science Education Research* (pp. 95–116). Springer. https://doi.org/10.1007/978-3-031-32225-9_7
- Großschedl, J., Welter, V. D. E., & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402–439. <https://doi.org/10.1002/tea.21482>
- Günther, S. L., Fleige, J., Upmeyer zu Belzen, A., & Krüger, D. (2019). Using the case method to foster preservice biology teachers' content knowledge and pedagogical content knowledge related to models and modeling. *Journal of Science Teacher Education*, 30(4), 321–343.
<https://doi.org/10.1080/1046560X.2018.1560208>
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Routledge.
<https://doi.org/10.4324/9780203825945>
- Hilfert-Rüppell, D., Meier, M., Horn, D., & Höner, K. (2021). Professional knowledge and self-efficacy expectations of pre-service teachers regarding scientific reasoning and diagnostics. *Education Sciences*, 11(10), 629. <https://doi.org/10.3390/educsci11100629>
- Hogan, T., Rabinowitz, M., & Craven, J. A. (2003). Representation in teaching: Inferences from research of expert and novice teachers. *Educational Psychologist*, 38(4), 235–247.
https://doi.org/10.1207/S15326985EP3804_3
- Irmer, M., Traub, D., Böhm, M., Förtsch, C., & Neuhaus, B. J. (2023). Using video-based simulations to foster pPCK/ePCK—New thoughts on the refined consensus model of PCK. *Education Sciences*, 13(3), 261. <https://doi.org/10.3390/educsci13030261>

- Jin, H., Shin, H., Johnson, M. E., Kim, J., & Anderson, C. W. (2015). Developing learning progression-based teacher knowledge measures. *Journal of Research in Science Teaching*, 52(9), 1269–1295. <https://doi.org/10.1002/tea.21243>
- Justi, R., & van Driel, J. (2005). The development of science teachers' knowledge on models and modelling: Promoting, characterizing, and understanding the process. *International Journal of Science Education*, 27(5), 549–573. <https://doi.org/10.1080/0950069042000323773>
- Jüttner, M., Boone, 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. <https://doi.org/10.1007/s11092-013-9157-y>
- Kelly, M. P., & Staver, J. R. (2005). A case study of one school system's adoption and implementation of an elementary science program. *Journal of Research in Science Teaching*, 42(1), 25–52. <https://doi.org/10.1002/tea.20043>
- Kenyon, L., Davis, E. A., & Hug, B. (2011). Design approaches to support preservice teachers in scientific modeling. *Journal of Science Teacher Education*, 22(1), 1–21. <https://doi.org/10.1007/s10972-010-9225-9>
- Khan, S., & Krell, M. (2019). Scientific reasoning competencies: A case of preservice teacher education. *Canadian Journal of Science, Mathematics and Technology Education*, 19(4), 446–464. <https://doi.org/10.1007/s42330-019-00063-9>
- Kind, P., & Osborne, J. (2017). Styles of scientific reasoning: A cultural rationale for science education? *Science Education*, 101(1), 8–31. <https://doi.org/10.1002/sce.21251>
- Kind, V., & Chan, K. K. H. (2019). Resolving the amalgam: Connecting pedagogical content knowledge, content knowledge and pedagogical knowledge. *International Journal of Science Education*, 41(7), 964–978. <https://doi.org/10.1080/09500693.2019.1584931>
- KMK (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife* [Educational standards in the subject biology for the general higher education entrance qualification]. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf
- KMK (2024a). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung* [Common content-specific requirements of the federal states for the subject sciences and subject didactics in teacher education]. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf
- KMK (2024b). *Weiterentwickelte Bildungsstandards in den Naturwissenschaften für das Fach Biologie (MSA)* [Educational standards in the subject biology for the intermediate school leaving]. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2024/2024_06_13-WeBiS_Biologie_MSA.pdf
- Krell, M., Vorholzer, A., & Nehring, A. (2022). Scientific reasoning in science education: From global measures to fine-grained descriptions of students' competencies. *Education Sciences*, 12(2), 97. <https://doi.org/10.3390/educsci12020097>
- Krell, M., Khan, S., Vergara, C., Cofré, H., Mathesius, S., & Krüger, D. (2023). Pre-service science teachers' scientific reasoning competencies: Analysing the impact of contributing factors. *Research in Science Education*, 53(1), 59–79. <https://doi.org/10.1007/s11165-022-10045-x>
- Kulgemeyer, C., Kempin, M., Weißbach, A., Borowski, A., Buschhüter, D., Enkrott, P., Reinhold, P., Riese, J., Schecker, H., Schröder, J., & Vogelsang, C. (2021). Exploring the impact of pre-service science teachers' reflection skills on the development of professional knowledge during a field experience. *International Journal of Science Education*, 43(18), 3035–3057. <https://doi.org/10.1080/09500693.2021.2006820>
- Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education*, 2(3), 307–338. <https://doi.org/10.1007/s10763-004-3224-2>

- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, 86(3), 681–718.
<https://doi.org/10.3102/0034654315627366>
- Leach, J., Millar, R., Ryder, J., & Séré, M.-G. (2000). Epistemological understanding in science learning: The consistency of representations across contexts. *Learning and Instruction*, 10(6), 497–527.
[https://doi.org/10.1016/S0959-4752\(00\)00013-X](https://doi.org/10.1016/S0959-4752(00)00013-X)
- Leiner, D. J. (2019). *SoSci Survey* (Version 3.1.06) [Computer software]. <https://www.sosicisurvey.de>
- Lenhard, W., & Lenhard, A. (2022). *Berechnung von Effektstärken* [Computation of effect sizes]. Psychometrica. <https://www.psychometrica.de/effektstaerke.html>
- Lieberei, T., Dawborn-Gundlach, M., van Driel, J., & Krell, M. (2024). Preservice science teachers' knowledge about how to teach scientific reasoning. *Teaching Science*, 70(2), 30-43.
- Lieberei, T., Großmann, L., Welter, V. D. E., Krüger, D & Krell, M. (2024). Considering multiple sources of validity evidence can help to address challenges in the development of pedagogical content knowledge (PCK) multiple-choice items. [Manuscript accepted].
- Lieberei, T., Welter, V. D. E., Großmann, L., & Krell, M. (2023). Findings from the expert-novice paradigm on differential response behavior among multiple-choice items of a pedagogical content knowledge test – implications for test development. *Frontiers in Psychology*, 14, 1240120. <https://doi.org/10.3389/fpsyg.2023.1240120>
- Lotter, C., Harwood, W. S., & Bonner, J. J. (2007). The influence of core teaching conceptions on teachers' use of inquiry teaching practices. *Journal of Research in Science Teaching*, 44(9), 1318–1347. <https://doi.org/10.1002/tea.20191>
- Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching*, 41(4), 370–391. <https://doi.org/10.1002/tea.20007>
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Kluwer Academic Publishers.
https://doi.org/10.1007/0-306-47217-1_4
- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34(4), 207–218. https://doi.org/10.1207/s15326985ep3404_2
- Mathesius, S., Hartmann, S., Upmeyer zu Belzen, A., & Krüger, D. (2016). Scientific reasoning as an aspect of pre-service biology teacher education: Assessing competencies using a paper-pencil-test. In T. Tal & A. Yarden (Eds.), *The future of biology education research*. (pp. 93–110).
- Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science*, 41(3), 621–634. <https://doi.org/10.1007/s11251-012-9248-z>
- Mineduc (2015). *Bases Curriculares: 7° básico a 2° medio* [Curricular basics: 7th basic to 2nd middle grade]. https://www.curriculumnacional.cl/614/articles-34949_Bases.pdf
- Mineduc (2019). *Bases Curriculares: 3° y 4° medio* [Curricular bases: 3rd and 4th grades]. https://www.curriculumnacional.cl/614/articles-91414_bases.pdf
- Minner, D. D., Levy, A. J. & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496. <https://doi.org/10.1002/tea.20347>
- Ministry of Education (2008a). *The Ontario Curriculum: Grades 9 and 10. Science*.
https://www.edu.gov.on.ca/eng/curriculum/secondary/science910_2008.pdf
- Ministry of Education (2008b). *The Ontario Curriculum: Grades 11 and 12. Science*.
https://www.edu.gov.on.ca/eng/curriculum/secondary/2009science11_12.pdf
- MBWK SH (2019). *Fachanforderungen Sachunterricht Primarstufe/Grundschule* [Curriculum for elementary schools].
https://fachportal.lernnetz.de/files/Fachanforderungen%20und%20Leitf%C3%A4den/Grundschule_Primarstufe/Fachanforderungen_%20barrierefrei/Fachanforderungen_Sachunterricht_GS_barrierearm%5B6553%5D.pdf

- Neumann, K., Härtig, H., Harms, U. & Parchmann, I. (2017). Science teacher preparation in Germany. In J. E. Pedersen, T. Isozaki & T. Hirano (Eds.), *Model science teacher preparation programs: An international comparison of what works* (pp. 29–52). Information Age Publishing.
- NGSS Lead States (2013). *Next Generation Science Standards*. National Academies Press.
<https://nap.nationalacademies.org/catalog/18290/next-generation-science-standards-for-states-by-states> <https://doi.org/10.17226/18290>
- 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. <https://doi.org/10.1002/tea.21022>
- Park, S., & Oliver, J. S. (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. <https://doi.org/10.1007/s11165-007-9049-6>
- Park, S., & Suh, J. K. (2019). The PCK map approach to capturing the complexity of enacted PCK (ePCK) and pedagogical reasoning in science teaching. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 185–197). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_8
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3–13.
<https://doi.org/10.1111/j.1745-3992.2005.00006.x>
- Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – a literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>
- Rosenkränzer, F., Hörsch, C., Schuler, S., & Riess, W. (2017). Student teachers' pedagogical content knowledge for teaching systems thinking: Effects of different interventions. *International Journal of Science Education*, 39(14), 1932–1951.
<https://doi.org/10.1080/09500693.2017.1362603>
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020–1049.
<https://doi.org/10.3102/0002831213477680>
- Sannert, R., Baur, A., & Krell, M. (2024). Design features and effects of professional development programmes for promoting students' competencies in scientific inquiry: A systematic review of empirical studies. [Manuscript submitted].
- Sannert, R., van Driel, J., & Krell, M. (2024). Using content representations to measure components of teachers' PCK about scientific reasoning. [Manuscript submitted].
- Schlichter, N. (2012). *Lehrerüberzeugungen zum Lehren und Lernen* [Teachers' beliefs about learning and teaching]. [Doctoral dissertation, Georg-August-Universität Göttingen].
<http://doi.org/10.53846/goediss-3049>
- Schmeisser, C., Krauss, S., Bruckmaier, G., Ufer, S., & Blum, W. (2013). Transmissive and constructivist beliefs of in-service mathematics teachers and of beginning university students. In Y. Li & J. N. Moschkovich (Eds.), *Proficiency and beliefs in learning and teaching mathematics* (pp. 51–67). SensePublishers. https://doi.org/10.1007/978-94-6209-299-0_5
- Schmelzing, S., van Driel, J. H., Jüttner, M., Brandenbusch, S., Sandmann, A., & Neuhaus, B. J. (2013). Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the "cardiovascular system". *International Journal of Science and Mathematics Education*, 11(6), 1369–1390. <https://doi.org/10.1007/s10763-012-9384-6>
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353.
<https://doi.org/10.1037/1040-3590.8.4.350>
- Schwartz, R., Lederman, N. G., & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education*, 88(4), 610–645. <https://doi.org/10.1002/sce.10128>

- She, J. (2023). *Clarifying the relationship between science teachers pedagogical content knowledge and student learning outcomes*. [Doctoral dissertation, The University of Hong Kong].
- She, J., & Chan, K. K. H. (2023). Situated and dynamic versus declarative and static forms of pedagogical content knowledge: An evaluation of the differences in test reactions and performance. *Journal of Research in Science Teaching*, 60(3), 568-607. <https://doi.org/10.1002/tea.21810>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Smith, P. S., & Banilower, E. R. (2015). Assessing PCK: A new application of the uncertainty principle. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (pp. 88–103). Routledge.
- Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94(2), 344–355. <https://doi.org/10.1037/0022-0663.94.2.344>
- Suh, J. K., & Park, S. (2017). Exploring the relationship between pedagogical content knowledge (PCK) and sustainability of an innovative science teaching approach. *Teaching and Teacher Education*, 64, 246–259. <https://doi.org/10.1016/j.tate.2017.01.021>
- Sweller, J., Zhang, L., Ashman, G., Cobern, W., & Kirschner, P. A. (2024). Response to de Jong et al.'s (2023) paper "Let's talk evidence – The case for combining inquiry-based and direct instruction". *Educational Research Review*, 42, 100584. <https://doi.org/10.1016/j.edurev.2023.100584>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- van Driel, J., & Berry, A. (2019). Pedagogical content knowledge in preservice teacher education. In M. A. Peters (Ed.), *Encyclopedia of teacher education* (pp. 1-6). Springer Singapore. https://doi.org/10.1007/978-981-13-1179-6_176-1
- van Driel, J., Hume, A., & Berry, A. (2023). Research on science teacher knowledge and its development. In N. G. Lederman, D. L. Zeidler, & J. S. Lederman (Eds.), *Handbook of research on science education* (pp. 1123–1161). Routledge. <https://doi.org/10.4324/9780367855758-41>
- van Uum, M. S. J., Peeters, M., & Verhoeff, R. P. (2021). Professionalising primary school teachers in guiding inquiry-based learning. *Research in Science Education*, 51(S1), 81–108. <https://doi.org/10.1007/s11165-019-9818-z>
- Vorholzer, A., von Aufschnaiter, C., & Boone, W. J. (2020). Fostering upper secondary students' ability to engage in practices of scientific investigation: A comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education*, 50(1), 333–359. <https://doi.org/10.1007/s11165-018-9691-1>
- Vorholzer, A., Petermann, V., Weber, J., Upmeyer zu Belzen, A., & Tiemann, R. (2022). Explicit instruction on procedural and epistemic knowledge – is it happening? A video-based exploration of classroom practice. *Research in Science & Technological Education*, 42(3), 513-532. <https://doi.org/10.1080/02635143.2022.2153245>
- Waßmann, S. (2014). *Innovationsfähigkeit im demografischen Wandel: Entwicklung und Validierung des Assessment-Tools NovaDemo zur Erfassung der Innovationsfähigkeit von Einzelpersonen und Arbeitsgruppen* [Innovative ability in demographic change: Development and validation of the NovaDemo assessment tool for assessing the innovative ability of individuals and working groups]. [Doctoral dissertation, Karlsruher Institut für Technologie]. KIT Scientific Publishing. <https://doi.org/10.5445/KSP/1000043686>

- Welter, V. D. E., Herzog, S., Harms, U., Steffensky, M. & Großschedl, J. (2022). School subjects' synergy and teacher knowledge: Do biology and chemistry teachers benefit equally from their second subject? *Journal of Research in Science Teaching*, 59(2), 285–326.
<https://doi.org/10.1002/tea.21728>
- Zhang, L., Kirschner, P. A., Cobern, W. W., & Sweller, J. (2022). There is an evidence crisis in science educational policy. *Educational Psychology Review*, 34(2), 1157–1176.
<https://doi.org/10.1007/s10648-021-09646-1>

8 Studie 5: Preservice Science Teachers' Knowledge About How to Teach Scientific Reasoning³¹

Abstract

In a modern and increasingly complex society, students are expected to acquire skills and knowledge to deal with societal challenges, for example, climate change. These include scientific reasoning competencies (SRC), which describe abilities to solve scientific problems by applying scientific skills and knowledge and thus play a crucial role in science education. To foster SRC, science teachers are expected not only to possess these competencies themselves but also to be aware of how their students acquire SRC and have instructional strategies to teach SRC. Preservice science teachers should already have acquired this pedagogical content knowledge (PCK) on a basic level by the time they graduate from university. To investigate this development of PCK about SRC in initial teacher education, we conducted an exploratory study with a sample of 36 preservice science teachers in a Master of Teaching program at a public university in Victoria, Australia. We used the PCK_{SR}-bio, a multiple-choice instrument for assessing PCK for different scientific reasoning activities regarding biology-related classroom situations. We did not find differences between the PCK about SRC of preservice science teachers in the first year of their master's studies and those in the second year of their master's studies, which indicates that their level does not significantly increase during the first year of their Master of Teaching program. An analysis of the participants' responses to the items revealed that they had comparable knowledge regarding each of the scientific reasoning activities addressed and were best at answering questions that addressed students' misconceptions. This is relevant for teachers who supervise early-career teachers as they should know what knowledge early-career teachers bring with them from initial teacher training and what they still need to learn. Furthermore, a comparison between the preservice science teachers in Australia and the preservice biology teachers in Germany, both in a similar phase of their initial teacher education, revealed that the participants in Germany achieved a significantly higher average test score than the participants in Australia. The findings are discussed and recommendations for the improvement of science teacher education related to SRC and potentially for in-service professional development are made.

³¹ Lieberei, T., Dawborn-Gundlach, M., van Driel, J. & Krell, M. (2024). Preservice science teachers' knowledge about how to teach scientific reasoning. *Teaching Science*, 70(2), 30-43.

8.1 Introduction

Addressing current societal challenges (e.g., climate change, global pandemics) in science education requires not only scientific knowledge but also an understanding of how scientific knowledge develops and its reliability. Related competencies should be promoted in science lessons, which is why students should not only obtain content knowledge about scientific topics at school but also be taught how this knowledge can be acquired and evaluated (OECD, 2022). Such competencies refer to the ability to solve scientific problems by applying scientific skills and knowledge (Kind & Osborne, 2017; Krell et al., 2020; Lawson, 2004) and are called scientific reasoning competencies (SRC). As SRC play a central role in 21st century skills in science education (Osborne, 2013), they are covered by the Australian science curriculum (ACARA, 2022). Consequently, SRC are also part of science teachers' professional knowledge because science teachers need these skills to teach their future students appropriately (Khan & Krell, 2019). Beyond that, science teachers need to know how to teach SRC to their students, which includes, among others, having knowledge about common students' (mis)conceptions of scientific reasoning and knowledge about appropriate instructional strategies to foster students' SRC (Alonzo et al., 2012; Jin et al., 2015; Park & Chen, 2012). Preservice science teachers are expected to acquire this knowledge during their university teacher training (AITSL, 2017).

In order to investigate the acquisition of SRC in teacher education programs, this article reports on a study that was conducted in the context of a long-standing international collaboration on SRC in science teacher education between researchers from Australia (Dawborn-Gundlach, van Driel) and Germany (Krell, Lieberei). A previous article published in *Teaching Science* reported findings from an empirical study on preservice science teachers' SRC in Australia (Krell et al., 2020). In that study, we found that the preservice science teachers had a fundamental understanding of SRC at the beginning of their Master of Teaching program. However, their SRC did not significantly improve during the first year of their teacher education program. This article follows up on those findings and provides theoretical insights into the relevance of preservice science teachers' knowledge about how to teach SRC (i.e., pedagogical content knowledge [PCK] about SRC). Because preservice science teachers not only need to develop SRC but also need to develop knowledge of how to teach SRC, this study focussed on their PCK.

To investigate preservice science teachers' PCK about SRC during their studies in teacher education before they start as practising early-career science teachers, we conducted an exploratory study with a multiple-choice test on PCK about SRC (PCK_{SR}-bio). Furthermore, we aimed to compare our findings with findings obtained from a sample of preservice science teachers in Germany in a similar phase of teacher education. To the best of our knowledge, this is the first study to explore preservice science teachers' PCK about SRC in Australia. Hence, the findings provide helpful insights into preservice science teachers' PCK about SRC at the end of their university teacher training before they become practising teachers. These insights can be used to further improve science teacher education related to SRC; they can also give in-service teachers who supervise early-career teachers who have just completed their teacher education an indication about what PCK about SRC can be expected in early-career teachers. Furthermore, the findings can provide insights into how in-service professional development programs can be improved.

8.2 Theoretical Background

The following sections provide (1) a definition of SRC, (2) a conceptualisation of preservice and in-service teachers' knowledge about SRC and how to teach it, and (3) the Australian curriculum context regarding SRC.

8.2.1 Scientific Reasoning Competencies

SRC are conceptualised as the essential capabilities required to solve a scientific problem in a specific context by applying a set of skills, abilities, and knowledge (Kind & Osborne, 2017; Krell et al., 2020; Lawson, 2004). Besides scientific content knowledge (declarative understanding of domain-specific concepts: "knowing what"), SRC also comprise procedural knowledge (understanding of procedures and constructs used in science: "knowing how") as well as epistemic knowledge (understanding of epistemic constructs and values: "knowing why"; Kind & Osborne, 2017). Therefore, SRC include different skills and abilities that are needed when applying content, procedural, and epistemic knowledge to solve a scientific problem (Khan & Krell, 2019; Kind & Osborne, 2017). They include empirical inquiry, such as planning and conducting investigations or using scientific models, as well as theoretical thinking, such as generating hypotheses (Rönnebeck et al., 2016).

Students need SRC to deal with societal challenges and live in an increasingly complex society; they need to be scientifically literate (Osborne, 2023). In the Programme for International Student Assessment (PISA), scientific literacy is understood as "the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen" (OECD, 2019, p. 100). According to this conceptualisation, acquiring fundamental scientific knowledge and understanding scientific reasoning activities, which include questioning, predicting, designing, and conducting experiments, is essential for developing scientific literacy (OECD, 2019; Osborne, 2013). In order to promote students' scientific literacy, it is essential to support the development of their SRC as they play an important role in understanding how scientific knowledge can be obtained and interpreted.

8.2.2 Teachers' Knowledge About Scientific Reasoning and How to Teach it

In order to support their students appropriately, science teachers need to have SRC themselves (Khan & Krell, 2019), which should already be developed over the course of their university teacher training. In a previous study, we found that preservice science teachers at an Australian public university had a fundamental understanding of SRC at the beginning of their Master of Teaching program, which did not significantly improve during the first year of the teacher education program (Krell et al., 2020).

In addition to their own SRC, science teachers also need to know how to develop SRC in their students. SRC can be addressed either implicitly (i.e., SRC are not addressed specifically but must be discovered by students when learning scientific content) or explicitly (i.e., SRC are addressed in instructional approaches; Vorholzer et al., 2020). Even though SRC are rarely addressed explicitly in science lessons (Vorholzer et al., 2022), it has been empirically shown that explicit instructional approaches are more effective in promoting students' SRC than implicit approaches are (Matlen & Klahr, 2013).

Science teachers need to know how to teach SRC to their students, that is, they need to have PCK about SRC. PCK refers to a specific form of teaching knowledge about how to transform specific content, for example, SRC, to enhance student learning (Gess-Newsome, 2015). PCK encompasses different components of knowledge, two of which are highlighted as the central ones (Alonzo et al., 2012; Jin et al., 2015; Park & Chen, 2012): Knowledge of Instructional Strategies for Teaching (KISR) and Knowledge of Students' Understanding (KSU). KISR describes subject-specific and topic-specific strategies to teach a scientific subject or topic, and KSU describes teachers' knowledge of the (mis)conceptions and prior knowledge of their students. These areas of knowledge are also connected (see Figure 8.1). While the connection KISR→KSU reflects the impact of instruction on students' conceptions, the connection KSU→KISR refers to addressing students' (mis)conceptions when designing instruction.

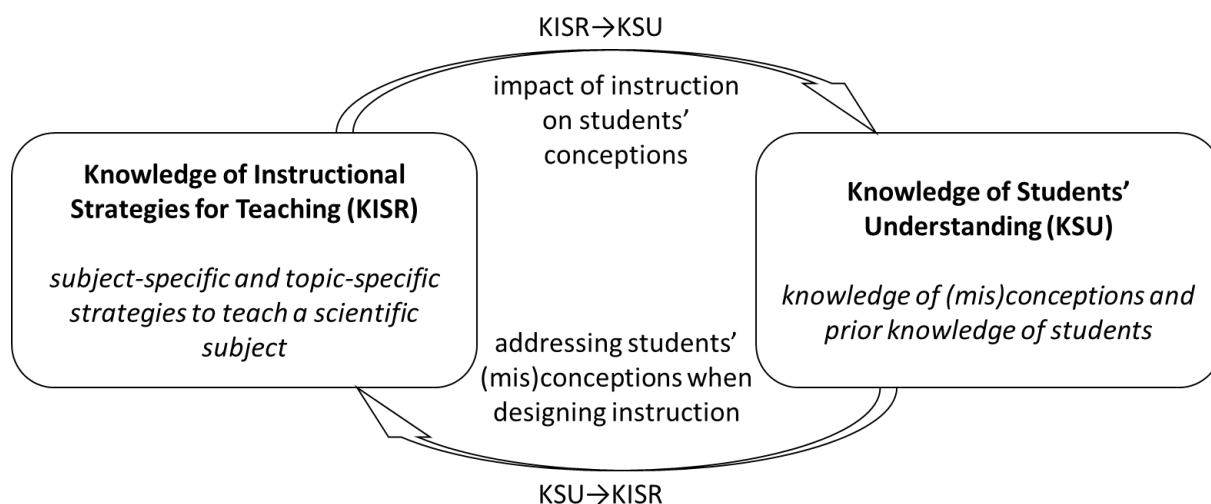


Figure 8.1: Overview of the two central components of PCK: KISR and KSU, as well as their connections

Even though teachers' PCK mainly develops over time as a result of learning and teaching experience as well as reflection on it (van Driel & Berry, 2019), preservice teachers are already expected to acquire basic PCK during initial teacher education at university (AITSL, 2017). Empirical studies have shown that the development of preservice science teachers' PCK depends on various factors, such as their content knowledge or the learning opportunities for PCK they have (e.g., Großschedl et al., 2019).

8.2.3 Scientific Reasoning in the Australian Context

As mentioned above, both students' SRC and preservice science teachers' PCK are addressed in the Australian curriculum. Regarding SRC, the Australian F-10 curriculum comprises the three dimensions of *learning areas*, *general capabilities*, and *cross-curriculum priorities*, which together "set out essential knowledge, understanding and skills that enable all young Australians to learn, contribute, and shape their world now and in the future" (ACARA, 2022). The curriculum for the learning area of science includes the three interrelated strands *science understanding*, *science as a human endeavour*, and *science inquiry skills*. To prepare students for a world in which SRC are needed to make informed decisions, the Australian science curriculum aims to enable students to understand central scientific concepts and scientific processes as well as the role of science in society (ACARA, 2022).

SRC are addressed for different scientific reasoning activities (e.g., conducting investigations, modelling) in the *science inquiry* strand, such as *questioning and predicting*; *planning and conducting*; *processing and analysing data and information*; *evaluating*; and *communicating* (ACARA, 2022). Students should have achieved different learning goals for these strands by the end of each school year, starting with the foundation year up to Year 10. Throughout their time in school, students are expected to acquire SRC by becoming increasingly independent in using different (empirical and theoretical) scientific reasoning activities.

Regarding preservice science teachers, the development of their teaching practice and expertise is specified in the Australian Professional Standards for Teaching (AITSL, 2017). Concerning their professional knowledge, it is explicitly stated that preservice teachers should know students and how they learn (Standard 1; reflecting the PCK component KSU) as well as know the content and how to teach it (Standard 2; reflecting the PCK component KISR; AITSL, 2017).

To achieve these goals as stated in the curriculum (ACARA, 2022), preservice science teachers need to have SRC themselves as well as PCK about SRC; both should already have been acquired by the time they graduate from university (AITSL, 2017). Regarding their SRC, in a previous study, we found that preservice science teachers had a fundamental understanding at the beginning of their Master of Teaching program, which, however, did not seem to improve significantly during the first year of their

teacher education program (Krell et al., 2020). However, no study yet has investigated preservice science teachers' PCK about SRC in Australia.

8.2.4 Research Aim

The aim of this study was to explore preservice science teachers' PCK about SRC during their studies in teacher education. For this purpose, we administered a multiple-choice test on PCK about SRC (PCK_{SR}-bio) in a sample of preservice science teachers from one university in Victoria. Furthermore, we compared the findings with those from a sample of preservice science teachers in Germany in a similar phase of teacher education. With this approach, the findings can provide insights not only into how to further improve science teacher education related to PCK about SRC but also into what in-service teachers who supervise early-career science teachers can expect from them regarding their PCK about SRC. Furthermore, these findings can be used to develop professional development programs on scientific reasoning tailored to the needs of science teachers.

8.3 Exploratory Study on PCK about SRC in Science Teacher Education

The following sections describe the development of the instrument used as well as the data collection and analysis.

8.3.1 Instrument Development

To explore preservice science teachers' PCK about SRC, we used the PCK_{SR}-bio, a multiple-choice instrument to assess PCK about SRC. The PCK_{SR}-bio was initially developed in German and was comprehensively evaluated with preservice biology teachers in different phases of teacher education in Germany (Lieberei et al., 2024). For the use with preservice science teachers in Australia, the PCK_{SR}-bio was independently translated into English by two professional translation institutes. The authors of this article then compared and discussed both translations and collectively developed the most appropriate translation for each item. After a first data collection, the items were marginally revised in terms of language to further improve understandability.

The items address the two central PCK components, KSU and KISR, as well as their connections KISR→KSU (reflecting the impact of instruction on students' thinking) and KSU→KISR (reflecting instruction that is designed based on students' conceptions; see Figure 8.1). As a theoretical framework of scientific reasoning, the PCK_{SR}-bio focuses on three scientific reasoning activities: (1) planning and conducting investigations, (2) generating hypotheses, and (3) using scientific models, which Rönnebeck et al. (2016), among others, pointed out as highly relevant. On the basis of this framework, we created one item for each interconnection between the four PCK elements (two components and two connections) and the three scientific reasoning activities; this led to a total of 12 items (see Table 8.1).

Table 8.1: Systematic overview of the 12 multiple-choice items of the PCK_{SR}-bio, developed to assess two PCK components and their connections for three scientific reasoning activities

	Planning and Conducting Investigations	Generating Hypotheses	Using Scientific Models
KISR	Item #1	Item #2	Item #3
KSU	Item #4	Item #5	Item #6
KISR→KSU	Item #7	Item #8	Item #9
KSU→KISR	Item #10	Item #11	Item #12

Each item of the PCK_{SR}-bio consists of a stem that describes a realistic classroom situation and contains relevant information about the learning group, for example, students' knowledge, the topic, or the targeted learning outcome. After that follows a prompt, identical for all items of the same PCK component (or connection), as well as three response options (one attractor; two distractors). A sample item is shown in Figure 8.2. Even though the PCK_{SR}-bio addresses biology-related classroom situations, we assume that preservice science teachers who specialise in other science learning areas can also answer it as it relates to scientific reasoning in general.

Stem: A biology teacher has worked with his students on the structure of DNA using models. In a subsequent lesson, he asks his students what models are and what function they have in biology. One student answers: "Models are visualisations of something, for example, DNA. We can use them to learn how DNA is built."

Prompt: Please tick the ways in which the teacher should take this idea about models and extend it through an understanding that is as scientifically and methodologically appropriate as possible.

Options:

He should...

(1) ... ask his students to make their own DNA models to highlight the potential of models as representations.

(2) ... practise "model critique" with his students by working out similarities and differences between DNA models and DNA.

(3) ... reflect on models by elaborating on the importance of DNA models in the discovery of DNA structure.

Figure 8.2: Sample item (Item #12) of the PCK_{SR}-bio

Note. The whole instrument can be found in Section 12.2. The correct answer is Option (3) because it is the only way to extend the idea of the student through an understanding that is as scientifically and methodologically appropriate as possible by emphasising the role of models as epistemic tools in the context of scientific reasoning (Göhner & Krell, 2022; Knuuttila, 2011). In contrast, in Options (1) and (2), models are used as visualisations, which do not support the teacher's goal of extending students' understanding of models in a way that is as scientifically and methodologically appropriate as possible (Upmeyer zu Belzen et al., 2019).

8.3.2 Data Collection

We collected data from preservice science teachers enrolled in the two-year Master of Teaching program at a public university in Victoria, which involves theory on campus and teaching placements in schools. A total of 36 preservice science teachers who were studying to become secondary school science teachers took part in the study. To maximise the sample, two rounds of data collection took place: The first one was in November 2022 with second-year master's students, and the second one was in spring 2023 with first-year master's students. All participants were in their initial teacher education program and had previous experience in undergraduate science or engineering courses. They were invited to participate in the study via the university's Learning Management System as well as in person. The participation was voluntary and anonymous, and the study was approved by the local ethics committee (ID 2022_23_LI). Only participants who fully completed the test were considered.

The participants answered the multiple-choice items in an online test (Leiner, 2019) and provided demographic information. In total, 41.7% of the preservice science teachers identified as male, 52.8% identified as female, and 5.6% identified as diverse. All participants studied one or two science learning areas ($n = 16$ biology, $n = 10$ chemistry, $n = 12$ physics, and $n = 27$ science), and some studied other learning areas (e.g., mathematics, $n = 11$).

8.3.3 Data Analysis

To investigate the preservice science teachers' PCK, we analysed their answers to the multiple-choice items of the PCK_{SR}-bio. For each analysis, we first tested whether the data was normally distributed by conducting a Shapiro-Wilk test. As most data turned out to be not normally distributed, we conducted nonparametric Mann-Whitney U tests to further examine differences in the test scores (i.e., the proportion of correct answers).

We first examined whether the first-year master's students and the second-year master's students differed in their test scores. Furthermore, we investigated whether the preservice science teachers' biological knowledge affected the test score because the PCK_{SR}-bio addresses biology-related classroom situations. For this purpose, we compared the test scores of the preservice science teachers who were studying biology as one of their learning areas (referred to from now on as "preservice biology teachers") with those of the preservice teachers who were not studying biology (referred to from now on as "preservice nonbiology teachers").

To further explore the preservice science teachers' PCK about SRC, we examined their responses to each item on a descriptive level (i.e., comparing the mean proportion of correct answers and the standard deviation for each item). In a more in-depth analysis, we then looked at the proportion of correct answers to the items that addressed the different scientific reasoning activities (e.g., Items #1, #4, #8, and #11 for planning and conducting investigations, see Table 8.1) as well as the different PCK components (e.g., Items #1, #2, and #3 for KISR, see Table 8.1).

Lastly, the test score of the 36 preservice science teachers in Australia was compared with that of 114 preservice biology teachers in Germany who also were in a two-year Master of Education program. The German data were collected as part of the evaluation process of the PCK_{SR}-bio (Lieberei et al., 2024). In contrast to Australia, in Germany, the bachelor's and the master's programs are concurrent teacher education programs, including both discipline-specific and pedagogical courses, as well as a small number of placements in schools lasting a few weeks.

All data analyses were done with IBM® SPSS® Statistics 28. Effect-size measures were calculated using an online tool by Lenhard and Lenhard (2022) and the interpretation of the effect size was based on Cohen (1988).

8.4 Results

Overall, the 36 preservice science teachers reached an average test score of .41 ($SD = .15$). We did not find significant differences between the test scores of the first-year master's students ($n = 11$, $M = .40$, $SD = .14$) and those of the second-year master's students ($n = 25$, $M = .41$, $SD = .15$), $U = 136.00$, $z = .05$, $p = .973$, or between the preservice biology ($n = 16$, $M = .41$, $SD = .15$) and the preservice nonbiology teachers ($n = 20$, $M = .40$, $SD = .14$), $U = 155.50$, $z = .15$, $p = .888$.

To further explore the preservice science teachers' PCK about SRC, we examined the responses of the entire group of 36 to each item (because there were no differences – see above), as shown in Table 8.2.

Table 8.2: Descriptive statistics for each item (M = mean proportion of correct answers, SD = standard deviation) answered by 36 preservice science teachers (secondary science)

	Items												Test Score
	KISR			KSU			KISR→KSU			KSU→KISR			
	IN	HY	MO	IN	HY	MO	IN	HY	MO	IN	HY	MO	
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	
<i>M</i>	0.33	0.44	0.47	0.42	0.28	0.89	0.28	0.47	0.36	0.50	0.33	0.08	.41
<i>SD</i>	0.48	0.50	0.51	0.50	0.45	0.32	0.45	0.51	0.49	0.51	0.48	0.28	.15

Note. IN = planning and conducting investigations; HY = generating hypotheses, MO = using scientific models.

It can be seen that almost all items had an average proportion of correct answers above the probability of correct guessing, which was .33. The three exceptions were Items #5, #7, and #12. Especially Item #12 (KSU→KISR for using scientific models, $M = .08$) turned out to be rather difficult. In contrast, Item #6 (KSU for using scientific models, $M = .89$) was particularly easy.

A closer look at the items addressing the scientific reasoning activities shows that the items addressing the activities of planning and conducting investigations ($M = .38$, $SD = .21$) and generating hypotheses ($M = .38$, $SD = .27$) were harder to answer for the preservice science teachers than the items addressing the activity of using scientific models ($M = .45$, $SD = .18$). However, the differences between any two scientific reasoning activities were not statistically significant.

A comparison of the items addressing the different PCK components revealed clear differences (see Figure 8.3): Items addressing KSU were the easiest to solve ($M = .53$, $SD = .28$), whereas items addressing KISR had a lower proportion of correct answers ($M = .42$, $SD = .29$). The proportion of correct answers for the connections was even lower ($M = .37$, $SD = .34$ for KISR→KSU and $M = .31$, $SD = .25$ for KSU→KISR). The differences between KSU and KISR→KSU ($U = 466$, $z = 2.15$, $p = .031$; $d = 0.50$, medium effect) as well as between KSU and KSU→KISR ($U = 385.5$, $z = 3.17$, $p = .002$; $d = 1.10$, large effect) were statistically significant. The differences between any two of the remaining PCK components were not statistically significant.

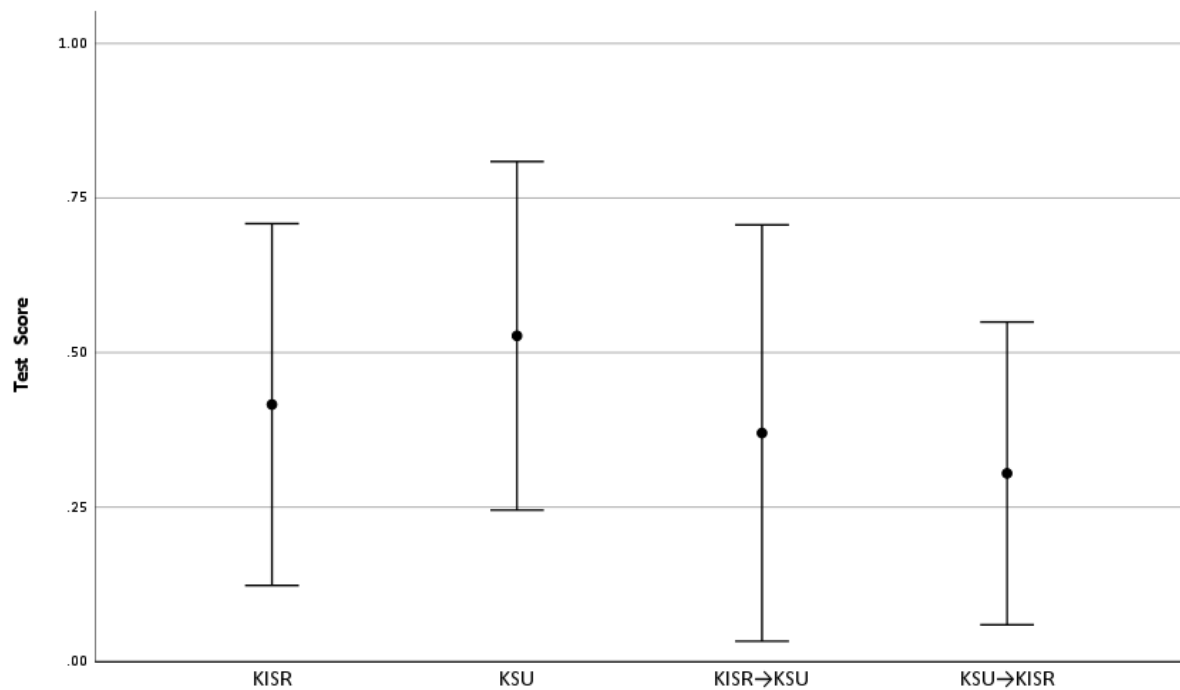


Figure 8.3: Proportion of correct answers ($\pm$ SD) for items addressing the different PCK components

Regarding the difference between the preservice science teachers in Australia and the preservice biology teachers in Germany (both in the master's program, secondary science), it can be seen that the participants in Germany achieved a significantly higher average test score ($M = .50$, $SD = .19$) than the participants in Australia ($M = .41$, $SD = .15$), $U = 1340.50$, $z = 3.15$, $p = .002$, corresponding to a large effect, $d = 0.53$ (Cohen, 1988).

8.5 Discussion & Conclusion

With our study, we aimed to follow up on previous findings on preservice science teachers' SRC (Krell et al., 2020) by exploring preservice science teachers' PCK about SRC during their studies in teacher education before they become practising early-career science teachers. In order to achieve this, we examined the answers to the PCK_{SR}-bio of 36 preservice science teachers in Australia enrolled in the two-year Master of Teaching program. The PCK_{SR}-bio is an instrument for assessing PCK about SRC that addresses biology-related classroom situations. In addition, we compared their test scores with those from a sample of preservice science teachers in Germany in a similar phase of teacher education.

First-year master's students and second-year master's students did not significantly differ in their test scores in our study. This indicates that the Australian preservice science teachers already have basic PCK about SRC at the start of their master's studies, which, however, does not significantly increase in the first year of their university teacher training. This is comparable to the findings of our previous study about preservice science teachers' SRC (Krell et al., 2020).

Furthermore, the preservice biology teachers did not significantly differ in their test scores from preservice nonbiology teachers. These findings support our assumption that the PCK_{SR}-bio, even though it addresses biology-related classroom situations, can also be answered by preservice teachers specialising in other science learning areas because the items mainly relate to scientific reasoning in general. As SRC address the skills and abilities required to solve a scientific problem in the context of different scientific subjects (Kind & Osborne, 2017; Krell et al., 2020; Lawson, 2004), (preservice) science teachers of all science learning areas need to know how to teach SRC to their (future) students. The findings show that this was the case for the participants in this study. Hence, the biology classroom context can be considered to be a surface characteristic of the items, which does not necessarily need

to be considered in order to solve the items correctly (Lieberei et al., 2023). Participants with less PCK about SRC are more likely to be influenced by surface characteristics of the items such as the classroom context when answering the items, whereas participants with more PCK about SRC are more likely to answer the items based on their knowledge.

Regarding the responses of the preservice science teachers to each item revealed that almost all items had a comparable average proportion of correct answers between .28 and .50. As each item includes three answer options, the probability of correct guessing is .33. Hence, on average, the items were rather difficult to solve. There were only two exceptions: one item was particularly easy to answer for the preservice science teachers (Item #6, KSU for using scientific models) and one was rather hard (Item #12, KSU→KISR for using scientific models). Therefore, no specific scientific reasoning activity stood out, as Items #6 and #12 address the same scientific reasoning activity. This indicates that the items of the PCK_{SR}-bio are balanced. However, the items that stand out indicate that specific PCK components appear to be particularly easy or difficult for the related scientific reasoning activity. Therefore, Item #6 may have yielded a comparatively high proportion of correct answers because the preservice science teachers already have comprehensive knowledge of common students' (mis)conceptions about modelling. In contrast, the rather low proportion of correct answers to Item #12 indicated that it is specifically difficult for preservice science teachers to develop appropriate instruction based on students' conceptions about modelling.

A closer look at the items addressing the scientific reasoning activities of planning and conducting investigations (see Table 8.2), generating hypotheses, and using scientific models revealed no statistically significant differences in the proportion of correct answers between any of the three activities. This indicates that the preservice science teachers had comparable PCK for these three activities. However, on a descriptive level, the items addressing the scientific reasoning activity of using scientific models were easier to solve for the preservice science teachers ($M = .45$) than the items addressing the other two scientific reasoning activities ($M = .38$ each). These findings support the assumption that PCK is a topic-specific construct (Gess-Newsome, 2015) as the preservice science teachers had different levels of PCK for different activities regarding SRC, which correspond to different topics. As all scientific reasoning activities assessed are stated in the Australian curriculum (ACARA, 2022), preservice science teachers must know how to teach them appropriately. A low test score on specific scientific reasoning activities of the PCK_{SR}-bio should be taken as an indication that teacher education courses could put a stronger emphasis on how to deal with common mis(conceptions) of students and on teaching suitable instructional strategies regarding the respective scientific reasoning activities.

Concerning the items addressing the different PCK components, it can be seen that the items addressing KSU were the easiest to solve ($M = .53$), the items addressing KISR had a considerably lower proportion of correct answers ($M = .42$), and the items addressing the two connections KISR→KSU ($M = .37$) and KSU→KISR ($M = .31$) had an even lower proportion of correct answers. However, even though PCK develops mainly over time due to learning and teaching experience as well as reflection on it (van Driel & Berry, 2019), preservice science teachers are already expected to acquire professional knowledge regarding KSU and KISR during their university teacher training (see Standards 1 and 2 in the Australian Professional Standards for Teaching; AITSL, 2017). As the participants in this study were in their initial teacher education program and only had previous experience in undergraduate science or engineering courses, they had limited practical teaching experience. Therefore, our findings indicate that KSU is probably a PCK component that preservice science teachers can develop in their (mainly theoretical) initial teacher education more easily than KISR or, especially, than the connections between the two components, which probably require practical teaching experience in order to develop. In-service teachers who supervise early-career teachers should be aware of this.

Finally, the comparison between the preservice science teachers in Australia and the preservice biology teachers in Germany, both in their master's program, revealed that the participants in Germany achieved a significantly higher average test score ($M = .50$) than the participants in Australia ($M = .42$). These findings are probably due to the fact that teacher education in Germany is structured differently as the bachelor's and the master's programs are concurrent teacher education programs, including both discipline-specific and pedagogical courses as well as a small number of placements in schools lasting a few weeks. Therefore, it is to be expected that the preservice biology teachers in Germany have had more learning opportunities to acquire PCK during their university teacher training than the preservice science teachers in Australia in this sample, who had studied for a science degree in their bachelor's program and were focusing on teacher education in their master's program.

8.5.1 Limitations

When interpreting the results, some limitations have to be considered. The main limitation of the study concerns the sample both in its size and in its composition, with 36 participants from one university. Therefore, it is not possible to generalise the findings of this study. However, as participation in research is typically on a voluntary basis, it is challenging to achieve large samples. Nevertheless, we are convinced that our findings provide an initial insight into preservice science teachers' PCK about SRC. A second limitation concerns the framework of the PCK_{SR}-bio. The instrument covers the central PCK components KSU and KISR as well as their connections, but PCK also includes other components (e.g., knowledge of the science curriculum; Park & Chen, 2012). Similarly, scientific reasoning also comprises other relevant activities (e.g., identifying research questions; Rönnebeck et al., 2016). Thus, the PCK_{SR}-bio does not provide a full overview of the participants' PCK about SRC. However, as the test duration would increase with more items, a focus on clearly defined and central PCK components and scientific reasoning activities was necessary.

8.5.2 Suggestions for Initial Teacher Education

Despite these limitations, our study provides initial findings on preservice science teachers' PCK about SRC in Australia. On the basis of our findings, we can make the following suggestions for initial teacher education and also, potentially, for in-service professional development.

First, the PCK_{SR}-bio can be used by teacher educators of different scientific subjects to monitor their students' PCK about SRC at various stages of their program. With this, it is possible to check whether the preservice science teachers have difficulties with certain scientific reasoning activities, which can help the educators to adapt the program so that their students can acquire PCK in all relevant areas.

Furthermore, our findings support the assumption that practical experience is needed for PCK development. Preservice science teachers are expected to have only a basic level of PCK at the end of their university teacher training, mainly about common students' (mis)conceptions. This knowledge must be further developed in early-career teachers, with a specific focus on instructional strategies and the connection between the two PCK components. Therefore, early-career teachers must learn which conceptions students develop through certain instructional strategies and which instructional strategies should be used to appropriately deal with students' (mis)conceptions. Consequently, in-service teachers who supervise early-career teachers should be aware of what PCK about SRC can realistically be expected from early-career teachers and how this knowledge can be further expanded.

Finally, as PCK is often not developed systematically after initial teacher education, professional development programs could help to foster PCK about SRC in in-service teachers. Our findings suggest that these programs should specifically focus on reflecting on the impact of instruction on students' thinking (KISR→KSU) and reflecting instruction that is designed based on students' conceptions (KSU→KISR). Furthermore, in such programs, the PCK_{SR}-bio or selected items could be used for summative or formative evaluation.

Acknowledgments

This article partly arose from research funded by the Federal Ministry of Education and Research of the Federal Republic of Germany in cooperation with the German Academic Exchange Service, Grant Number: 57559255.

The authors would like to thank Emma Stevenson, Emily Rochette, Cristina Guarrella, and Kristen Hebden for their support in data collection as well as Gráinne Newcombe for English language editing.

8.6 References

- ACARA (2022). *Science. F-10 Version 9.0. Scope and sequence*.
<https://v9.australiancurriculum.edu.au/content/dam/en/curriculum/ac-version-9/downloads/science/science-scope-and-sequence-f-10-v9.docx>
- AITSL (2017). *Australian professional standards for teachers*. <https://www.aitsl.edu.au/standards>
- Alonzo, A. C., Kobarg, M., & Seidel, T. (2012). Pedagogical content knowledge as reflected in teacher-student interactions: Analysis of two video cases. *Journal of Research in Science Teaching*, 49(10), 1211–1239. <https://doi.org/10.1002/tea.21055>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
<https://doi.org/10.4324/9780203771587>
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (pp. 28–42). Routledge.
- Göhner, M., & Krell, M. (2022). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*, 52(2), 395–414.
<https://doi.org/10.1007/s11165-020-09945-7>
- Großschedl, J., Welter, V. D. E. & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402–439. <https://doi.org/10.1002/tea.21482>
- Jin, H., Shin, H., Johnson, M. E., Kim, J., & Anderson, C. W. (2015). Developing learning progression-based teacher knowledge measures. *Journal of Research in Science Teaching*, 52(9), 1269–1295. <https://doi.org/10.1002/tea.21243>
- Khan, S., & Krell, M. (2019). Scientific reasoning competencies: A case of preservice teacher education. *Canadian Journal of Science, Mathematics and Technology Education*, 19(4), 446–464. <https://doi.org/10.1007/s42330-019-00063-9>
- Kind, P., & Osborne, J. (2017). Styles of scientific reasoning: A cultural rationale for science education? *Science Education*, 101(1), 8–31. <https://doi.org/10.1002/sce.21251>
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42(2), 262–271.
<https://doi.org/10.1016/j.shpsa.2010.11.034>
- Krell, M., Dawborn-Gundlach, M., & van Driel, J. (2020). Scientific reasoning competencies in science teaching. *Teaching Science*, 66(2), 32–42.
- Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education*, 2(3), 307–338.
<https://doi.org/10.1007/s10763-004-3224-2>
- Leiner, D. J. (2019). SoSci Survey (Version 3.1.06) [Computer software]. Available at
<https://www.sosicurvey.de>
- Lenhard, W., & Lenhard, A. (2022). *Berechnung von Effektstärken* [Computation of effect sizes]. Psychometrica. <https://www.psychometrica.de/effektstaerke.html>
- Lieberei, T., Großmann, L., Welter, V. D. E., Krüger, D., & Krell, M. (2024). Considering multiple sources of validity evidence can help to address challenges in the development of pedagogical content knowledge (PCK) multiple-choice items. [Manuscript accepted]

- Lieberei, T., Welter, V. D. E., Großmann, L., & Krell, M. (2023). Findings from the expert-novice paradigm on differential response behavior among multiple-choice items of a pedagogical content knowledge test - implications for test development. *Frontiers in Psychology*, 14, 1240120. <https://doi.org/10.3389/fpsyg.2023.1240120>
- Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science*, 41(3), 621–634. <https://doi.org/10.1007/s11251-012-9248-z>
- OECD (2019). *PISA 2018: Assessment and analytic framework*. PISA. OECD Publishing.
- OECD (2022). *PISA 2024: Strategic vision and direction for science*.
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265–279. <https://doi.org/10.1016/j.tsc.2013.07.006>
- Osborne, J. (2023). Science, scientific literacy, and science education. In N. G. Lederman, D. L. Zeidler, & J. S. Lederman (Eds.), *Handbook of research on science education* (pp. 785–816). Routledge. <https://doi.org/10.4324/9780367855758-30>
- 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. <https://doi.org/10.1002/tea.21022>
- Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>
- Upmeier zu Belzen, A., van Driel, J., & Krüger, D. (2019). Introducing a framework for modeling competence. In A. Upmeier zu Belzen, D. Krüger, & J. van Driel (Eds.), *Towards a competence-based view on models and modeling in science education: Vol. 12. Models and modeling in science education* (pp. 3–19). Springer. https://doi.org/10.1007/978-3-030-30255-9_1
- Vorholzer, A., von Aufschnaiter, C., & Boone, W. J. (2020). Fostering upper secondary students' ability to engage in practices of scientific investigation: A comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education*, 50(1), 333–359. <https://doi.org/10.1007/s11165-018-9691-1>
- Vorholzer, A., Petermann, V., Weber, J., Upmeier zu Belzen, A., & Tiemann, R. (2022). Explicit instruction on procedural and epistemic knowledge – is it happening? A video-based exploration of classroom practice. *Research in Science & Technological Education*, 42(3), 513–532. <https://doi.org/10.1080/02635143.2022.2153245>
- van Driel, J., & Berry, A. (2019). Pedagogical content knowledge in preservice teacher education. In M. A. Peters (Ed.), *Encyclopedia of teacher education* (pp. 1–6). Springer Singapore. https://doi.org/10.1007/978-981-13-1179-6_176-1

9 Zusammenfassung und Interpretation der zentralen Ergebnisse der durchgeführten Studien

Die zentralen Ergebnisse der fünf im Rahmen dieser Dissertation durchgeführten Studien werden im Folgenden vorgestellt. Eine übergreifende Diskussion und damit einhergehende Limitationen sowie Implikationen für die Forschung und Praxis erfolgen in Kapitel 10.

9.1 Studie 1: Findings from the Expert-Novice Paradigm on Differential Response Behavior Among Multiple-Choice Items of a Pedagogical Content Knowledge Test – Implications for Test Development (Kapitel 4)

Die erste Studie verfolgte das Ziel, das Antwortverhalten von Expert:innen und Noviz:innen bei der Bearbeitung des PCK_{SR}-bio vergleichend auf Grundlage des Experten-Novizen-Paradigmas zu analysieren, um ein besseres Verständnis über Schwierigkeiten bei der Testentwicklung in Bezug auf die Konstruktrepräsentation zu erlangen.

Die 10 Expert:innen erreichten einen signifikant höheren Testscore im Vergleich zu den 10 Noviz:innen. Unterschiede bezüglich der Variabilität des Antwortverhaltens waren deskriptiv sichtbar (die Expert:innen wiesen eine niedrigere Variabilität auf als die Noviz:innen), jedoch nicht statistisch signifikant. Bei der Begründung zur Auswahl einer Antwortoption beriefen sich Noviz:innen zudem häufiger auf Oberflächenmerkmale der Items, das heißt, sie begründeten ihre Wahl eher mit oberflächlichem Wissen wie Intuition oder persönlicher Erfahrung (*surface thinking*). Expert:innen hingegen verstanden häufiger die zugrunde liegenden Tiefenmerkmale der Items, begründeten die Auswahl einer Antwortoption also basierend auf ihrem PCK (*deep thinking*).

Die Ergebnisse zeigen, dass die Zielgruppe, in welcher das Instrument hauptsächlich eingesetzt werden soll, bereits bei der Konstruktion von Testinstrumenten berücksichtigt werden sollte und unterstreichen somit die Notwendigkeit der Durchführung stichprobenspezifischer Validierungsstudien für PCK-Instrumente. Dies gilt insbesondere auch in Bezug auf die Konsistenz des Antwortverhaltens, da bei einer höheren Varianz auch eine höhere Anzahl an Items notwendig ist, um eine akzeptable interne Konsistenz sowie ausreichende Trennschärfen zu erzielen. Darüber hinaus sollten Oberflächenmerkmale der Items konstant gehalten oder systematisch verändert werden, um konstruktirrelevante Ursachen für die Varianz des Antwortverhaltens von Noviz:innen zu vermeiden oder zu kontrollieren.

9.2 Studie 2: Considering Multiple Sources of Validity Evidence Can Help to Address Challenges in the Development of Pedagogical Content Knowledge (PCK) Multiple-Choice Items (Kapitel 5)

Die zweite Studie verfolgte das Ziel, Herausforderungen bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK in Bezug auf kognitive Operationen und Aufgabenformate, der Repräsentation des zugrunde liegenden Konstrukts sowie psychometrischen Eigenschaften durch die Berücksichtigung unterschiedlicher Validitätsquellen zu begegnen. Dazu wurde das betrachtete Instrument, der PCK_{SR}-bio, zunächst systematisch entwickelt, um auf diese Weise Validitätsevidenz auf Grundlage des Testinhalts sicherzustellen. Die Items wurden dabei mit Fokus auf zwei eindeutig definierte PCK-Komponenten (KISR und KSU) und ihren Verbindungen sowie drei zentralen Denk- und Arbeitsweisen gewählt (Experimentieren, Hypothesen formulieren, Modellieren) und auf Grundlage theoretischer und empirischer Forschungsergebnisse entwickelt. Im Anschluss erfolgte eine

mehrfache Revision der Items durch eine Gruppe von Expert:innen, welche eine Anpassung linguistischer Eigenschaften der Antwortoptionen (circa gleiche Länge, dieselben Fachbegriffe) sowie Überarbeitung der Distraktoren basierend auf erfassten Vorstellungen von Studierenden zur Erhöhung der Authentizität umfasste.

In Bezug auf die Validitätsevidenz auf Grundlage von Antwortprozessen zeigte sich, dass die Studierenden bei der Wahl eines Attraktors häufiger auf PCK rekurrierten, während sie bei der Wahl eines Distraktors fast ausschließlich auf oberflächliches, eher intuitives Wissen rekurrierten. Darüber hinaus wurde die Auswahl einer Antwortoption, bei der auf PCK rekurriert wurde, häufig mit derjenigen PCK-Komponente begründet, die in dem jeweiligen Item adressiert war.

In der Querschnittstudie zeigte sich im Rahmen der Überprüfung der Validitätsevidenz basierend auf der internen Struktur eine niedrige interne Konsistenz, wobei diese innerhalb der Gruppen mit steigendem Bildungsabschluss ebenfalls anstieg. Hinsichtlich der Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen konnte eine signifikant positive Korrelation zwischen dem PCK_{SR} und dem CK der Studierenden, jedoch nicht zwischen dem PCK_{SR} und dem Studiensemester gezeigt werden. In der Interventionsstudie zeigte sich darüber hinaus keine signifikante Veränderung im Testscore der Versuchsgruppe, im Vergleich zur Kontrollgruppe ließ sich jedoch deskriptiv ein stärkerer Anstieg des PCK_{SR} erkennen.

Diese Ergebnisse stützen damit die Validität der Testwertinterpretation als PCK_{SR} und zeigen, dass den genannten Herausforderungen bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK durch die Berücksichtigung unterschiedlicher Validitätsquellen begegnet werden kann. Insbesondere wird durch die Ergebnisse dieser Studie deutlich, dass eine geringe interne Konsistenz nicht automatisch zum Ausschluss eines Instruments führen sollte, da eine valide Erfassung des zugrunde liegenden Konstrukts dennoch möglich ist.

9.3 Studie 3: Using the PCK Map Approach to Obtain Validity Evidence Based on Response Processes for a Multiple-Choice Instrument (Kapitel 6)

In Studie 3 wurde betrachtet, wie der *PCK map approach* dazu verwendet werden kann, Validitätsevidenz auf Grundlage von Antwortprozessen zu sammeln. Die Auswertung der Daten der zehn Lehramtsstudierenden im Bachelor aus Studie 1 zeigte, dass die Teilnehmenden bei Beantwortung der Items am häufigsten mit denjenigen PCK-Komponenten und ihren Verbindungen argumentierten, die von dem jeweiligen Items erfasst wurden. Darüber hinaus wurde bei der Begründung zur Auswahl einer Antwortoption nicht mit anderen PCK-Komponenten des Pentagon-Modells oder ihren Verbindungen argumentiert. Die Daten ließen sich als *PCK map* darstellen.

Die Ergebnisse stärken damit nicht nur die Validität der Testwertinterpretation des PCK_{SR} -bio, indem sie eine adäquate Konstruktrepräsentation stützen, sondern zeigen ebenfalls, dass sich der *PCK map approach* im Entwicklungsprozess von Multiple-Choice-Instrumenten zur Erfassung von PCK eignet, um Validitätsevidenz auf Grundlage von Antwortprozessen zu sammeln.

9.4 Studie 4: Pedagogical Content Knowledge (PCK) about Scientific Reasoning in Different Groups of Science Teachers and Its Relation to Other Variables (Kapitel 7)

In Studie 4 wurde der PCK_{SR}-bio zur Erfassung des PCK_{SR} verschiedener Lehrender der Biologie (Lehrkräfte an Grund- und weiterführender Schule sowie Hochschuldozierende in der Lehrkräfteausbildung) eingesetzt. Darüber hinaus sollte die Abhängigkeit des PCK_{SR} der Lehrkräfte von konzeptionell verwandten Variablen (transmissiven Überzeugungen, Offenheit zur Innovation, Berufserfahrung) untersucht werden.

Grundschullehrkräfte wiesen von den betrachteten Gruppen den geringsten Testscore auf, gefolgt von Lehrkräften an weiterführenden Schulen. Hochschuldozierende in der Lehrkräfteausbildung wiesen den höchsten Testscore auf. Gleichzeitig antworteten die Hochschuldozierenden in der Lehrkräfteausbildung konsistenter als Lehrkräfte an weiterführenden Schulen, welche wiederum konsistenter als Lehrkräfte an Grundschulen antworteten. Bei der Betrachtung der konzeptionell verwandten Variablen zeigte sich in keiner der Gruppen eine signifikante Korrelation des PCK_{SR}-Testscores mit der Berufserfahrung, weshalb diese als Einflussfaktor ausgeschlossen werden konnte. Transmissive Überzeugungen stellten sich wie erwartet als ein negativer, die Offenheit zur Innovation anders als erwartet ebenfalls als ein negativer Einflussfaktor heraus. Dabei unterschied sich der Einfluss in den betrachteten Gruppen: Bei Lehrkräften an Grundschulen und Hochschuldozierenden konnte ein signifikant negativer Einfluss transmissiver Überzeugungen gezeigt werden, welcher bei Lehrkräften an weiterführenden Schulen nur auf einem deskriptivem Level erkennbar war. Gleichzeitig konnte bei Lehrkräften an Grundschulen und Hochschuldozierenden kein Einfluss der Offenheit zur Innovation auf den PCK_{SR}-Testscore gezeigt werden, bei Lehrkräften an weiterführenden Schulen zeigte sich hingegen ein überraschenderweise negativer Einfluss.

Die Ergebnisse verdeutlichen, dass die Berücksichtigung der internen Konsistenz sinnvoll sein kann, um inhaltliche Aussagen über ein Konstrukt zu treffen. Darüber hinaus geben sie Hinweise darauf, dass sich das PCK_{SR} von Lehrkräften nicht automatisch mit steigender Berufserfahrung verbessert, was die Notwendigkeit entsprechender Weiterbildungsangebote unterstreicht. Der unterschiedliche Einfluss der konzeptionell verwandten Variablen auf das PCK_{SR} bietet eine erste grundlegende Übersicht über das Konstrukt und kann als Ausgangspunkt für weitere Forschung sowie für die Berücksichtigung bei der Entwicklung effektiver Aus- und Weiterbildungsangebote dienen.

9.5 Studie 5: Preservice Science Teachers' Knowledge About How to Teach Scientific Reasoning (Kapitel 8)

Studie 5 verfolgte das Ziel, das PCK_{SR} australischer Lehramtsstudierender zu explorieren. Dabei konnte weder ein signifikanter Unterschied im PCK_{SR}-Testscore zwischen Studierenden im ersten und zweiten Jahr des Masterstudiums noch zwischen Studierenden, die Biologie studierten und denjenigen, die nicht Biologie studierten, nachgewiesen werden. Die Lösungswahrscheinlichkeit eines Großteils der Items lag über der Ratewahrscheinlichkeit von 33 %, wobei Items, die das Experimentieren oder Hypothesen formulieren adressierten, im Durchschnitt eine höhere Itemschwierigkeit aufwiesen als Items zum Modellieren. Diese Unterschiede waren jedoch statistisch nicht signifikant. Eine Analyse der Itemschwierigkeit nach adressierten PCK-Komponenten und ihren Verbindungen zeigt, dass Items über KSU leichter zu beantworten waren als Items über KISR, welche wiederum leichter zu beantworten waren als die Verbindungen der Komponenten (in beide Richtungen). Dabei waren nur die Unterschiede zwischen KSU und den zwei Verbindungen statistisch signifikant. Zuletzt zeigte sich, dass

die vergleichbare Stichprobe an Biologie-Masterstudierenden aus Deutschland einen signifikant höheren Testscore aufwies als die Masterstudierenden aus Australien.

Die Ergebnisse deuten darauf hin, dass die australischen Masterstudierenden bereits ein grundlegendes Verständnis von PCK_{SR} haben, welches sich im Verlauf des ersten Jahres des Studiums jedoch nicht signifikant verbessert. Der erreichte Testscore scheint darüber hinaus nicht davon abzuhängen, ob ein biologisches Studium absolviert wurde. Dies entspricht den Erwartungen, da der PCK_{SR} -bio darauf abzielt, trotz adressierter Unterrichtssituationen mit Fachinhalten der Biologie disziplinspezifisches PCK zu erfassen, über welches Lehrkräfte aller naturwissenschaftlicher Fächer verfügen sollten. Zusätzlich implizieren die Ergebnisse, dass insbesondere das Wissen über Vorstellungen der Lerngruppe eher in den theoretischen Kontexten der universitären Ausbildung vermittelt werden kann, während es zum Aufbau von Wissen über Instruktionsstrategien und insbesondere den Verbindungen der Komponenten eher Praxiserfahrung benötigt.

10 Übergreifende Diskussion und Ausblick

Das PCK von Lehrkräften ist eine zentrale Voraussetzung für die Unterrichtsqualität und damit die Lernleistung ihrer Schüler:innen (z. B. Coe et al., 2014). Auch wenn sich PCK mit steigender Berufserfahrung hauptsächlich durch praktische Erfahrung und Reflexion dieser entwickelt (van Driel & Berry, 2019), sollten angehende Lehrkräfte Grundlagen dieses Wissensbereichs bereits während ihrer universitären Ausbildung erlangen und diese als praktizierende Lehrkräfte weiter ausbauen. Um diesen Wissenserwerb zu evaluieren, bedarf es geeigneter Erhebungsinstrumente, welche PCK inhaltspezifisch und valide erfassen.

Auch wenn Erkenntnisgewinnungskompetenzen einen elementaren Bestandteil des naturwissenschaftlichen Unterrichts und damit auch des Fachs Biologie ausmachen (z. B. Mayer, 2007), existieren bisher nur wenige Studien zu dem PCK_{SR} angehender und praktizierender Lehrkräfte. Diese unterscheiden sich darüber hinaus maßgeblich hinsichtlich verschiedener Aspekte, darunter etwa der betrachteten Stichproben (z. B. Studierende: Hilfert-Rüppell et al., 2021; Lehrkräfte an weiterführenden Schulen: Lin & Wang, 2024) oder der verwendeten Erhebungsmethoden und -instrumente (z. B. Beobachtungen: O. de Jong & van der Valk, 2007; gemischte Aufgabenformate: Lin & Wang, 2024). Damit fehlt es an einheitlichen Erkenntnissen zu diesem spezifischen Wissensbereich, welche für ein grundlegendes Verständnis von PCK_{SR} jedoch zentral sind und dazu genutzt werden können, effektive Aus- und Weiterbildungsangebote zu gestalten.

Um diesen Desideraten zu begegnen, wurde im Rahmen dieser Dissertation ein Multiple-Choice-Instrument zur Erfassung des PCK_{SR} angehender und praktizierender Lehrkräfte entwickelt, validiert und angewendet. Die Ergebnisse der fünf dazu durchgeführten Studien wurden ausführlich in den jeweiligen Manuskripten diskutiert (Kapitel 4 bis 8). Im Folgenden werden diese Ergebnisse zusätzlich übergreifend zur Beantwortung der übergeordneten Fragestellung (I) aus einer methodischen (Kapitel 10.1) sowie zur Beantwortung der übergeordneten Fragestellung (II) aus einer inhaltlichen Perspektive (Kapitel 10.2) diskutiert. Daraufhin folgt die Synthese beider Perspektiven (Kapitel 10.3). Im Anschluss werden mit der Interpretation der Ergebnisse einhergehende zentrale Limitationen genannt (Kapitel 10.4) und Implikationen für die Forschung und Praxis abgeleitet sowie abschließend ein kurzer Ausblick gegeben (Kapitel 10.5).

10.1 Methodische Perspektive

Zur Beantwortung der übergeordneten Fragestellung (I) werden die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien aus einer methodischen Perspektive diskutiert, um zu untersuchen, wann sich Multiple-Choice-Instrumente zur Erfassung von PCK eignen und wie mit ihrer Entwicklung einhergehende häufig genannte Herausforderungen hinsichtlich der adäquaten Formulierung der Items, einer passenden Repräsentation des zugrunde liegenden Konstrukts sowie der psychometrischen Eigenschaften des Instruments durch die systematische Entwicklung und Validierung begegnet werden kann.

Zur Erfassung von PCK gibt es eine Vielzahl unterschiedlicher Erhebungsmethoden und -instrumente, welche sich je nach betrachteter Form bzw. in Abhängigkeit des betrachteten Bereichs unterscheiden (s. Kapitel 2.1.6). Im Vergleich zur Analyse des Verhaltens einer Lehrkraft im Unterricht werden Paper-Pencil-Tests, zu denen auch Multiple-Choice-Instrumente gehören, als weniger authentisch kritisiert, da den Unterricht beeinflussende Kontextfaktoren nicht berücksichtigt werden können (Chan & Hume, 2019). Auch wenn sich Paper-Pencil-Tests damit nicht zur Erfassung des dynamischen ePCK einer Lehrkraft in einer spezifischen Unterrichtssituation (in Bezug auf das Planen, Unterrichten und Reflektieren) eignen sollen, haben sie den Vorteil, das statische pPCK von Lehrkräften

situationsübergreifend zu erfassen. Insbesondere für angehende Lehrkräfte, welche im Rahmen ihrer universitären Ausbildung zumeist nur über wenig praktische Unterrichtserfahrung (und damit über kaum ausgeprägtes ePCK) verfügen, eignen sich daher Erhebungsinstrumente dieser Art.

Paper-Pencil-Tests können darüber hinaus unterschiedliche Antwortformate aufweisen. So haben Instrumente mit offenen Antwortformaten den Vorteil, das PCK von Lehrkräften in seiner Tiefe zu erfassen (Chan & Hume, 2019; Smith & Banilower, 2015) und so ein breiteres Spektrum an Denkprozessen, welche der Antwort zugrunde liegen, darzustellen (Martinez, 1999). Auf diese Weise kann auch das aus individuellen Unterrichtserfahrungen entwickelte pPCK einer Lehrkraft abgebildet werden (Schmelzing et al., 2013). Diese Art der Instrumente geht jedoch häufig mit einer langen Testbearbeitung sowie einem hohen Maß an Interpretation bei der Auswertung einher (Hollingworth et al., 2007; Smith & Banilower, 2015). Geschlossene Paper-Pencil-Tests wie Multiple-Choice-Instrumente haben hingegen Vorteile in Bezug auf die Testökonomie und Objektivität (Haladyna, 2004), wodurch sie sich insbesondere für den Einsatz bei großen Stichproben (Hill et al., 2008) und damit auch der Evaluation des PCK angehender und praktizierender Lehrkräfte eignen.

Trotz dieser Vorteile in Bezug auf die Testökonomie und Objektivität werden Multiple-Choice-Instrumente zur Erfassung von PCK kritisch gesehen (z. B. Rosenkränzer et al., 2017), da sie mit Herausforderungen in Bezug auf eine adäquate Formulierung der Items (z. B. Schmelzing et al., 2013), einer passenden Repräsentation des zugrunde liegenden Konstrukts (z. B. Smith & Banilower, 2015) sowie der psychometrischen Eigenschaften (vgl. Taber, 2018) verbunden sind. Um zu untersuchen, wie diesen Herausforderungen begegnet werden kann, wurde im Rahmen dieser Dissertation der PCK_{SR}-bio, ein Multiple-Choice-Instrument zur Erfassung von PCK im Kompetenzbereich Erkenntnisgewinnung, systematisch entwickelt und validiert.

Um sicherzustellen, dass ein Multiple-Choice-Instrument eine valide Interpretation des Testwerts erlaubt, sollte bei der Entwicklung zunächst eine eindeutige Formulierung der Items berücksichtigt werden (Haladyna, 2004). Im Entwicklungsprozess des PCK_{SR}-bio wurde deshalb auf eine in etwa gleiche Länge der Antwortoptionen sowie eine einheitliche Verwendung von Fachbegriffen geachtet, damit keine der Antwortoptionen aus linguistischen Gründen attraktiver erscheint. Insbesondere bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK muss darüber hinaus die Korrektheit der Attraktoren sowie die Inkorrektheit der (dennoch authentisch erscheinenden) Distraktoren auf Basis des aktuellen kanonischen cPCK (vgl. Carlson et al., 2019) sichergestellt werden (Hill et al., 2008; Schmelzing et al., 2013; Smith & Banilower, 2015). Deshalb wurden die Antwortoptionen des PCK_{SR}-bio zunächst basierend auf theoretischer und empirischer Grundlage entwickelt und anschließend in einer Gruppe von Expert:innen diskutiert, um konstruktirrelevante Einflussfaktoren auszuschließen. Basierend auf den Rückmeldungen von Masterstudierenden auf die als offene Fragen formulierten Items wurden zusätzlich diejenigen Distraktoren, die im Rahmen einer Pilotierungsstudie selten gewählt wurden, revidiert. Durch dieses in Studie 2 beschriebene Vorgehen kann damit nicht nur die Validitätsevidenz auf Grundlage des Testinhalts für den PCK_{SR}-bio gestützt, sondern auch sichergestellt werden, dass die Items eindeutig formuliert und die Attraktoren mit Sicherheit korrekt sowie die Distraktoren mit Sicherheit inkorrekt sind, letztere aber dennoch authentisch für die Proband:innen erscheinen. Eine Übersicht des vollständigen Instruments inklusive begründeter Lösungen zu jedem Item befindet sich im Anhang (Kapitel 12.1).

Um den Herausforderungen in Bezug auf die Konstruktrepräsentation zu begegnen, wurde sich bei der Entwicklung des PCK_{SR}-bio dafür entschieden, die zentralen naturwissenschaftlichen Denk- und Arbeitsweisen „Experimentieren“, „Hypothesen formulieren“ und „Modellieren“ als elementare Bestandteile der Erkenntnisgewinnungskompetenz zugrunde zu legen. Um das PCK der (angehenden) Lehrkräfte als eigenständigen Wissensbereich mit unterschiedlichen Komponenten zu erfassen, wurde eine transformative Konzeptualisierung des Konstrukts gewählt. Zusätzlich wurde bei der Entwicklung

des PCK_{SR}-bio den Empfehlungen von Smith und Banilower (2015) gefolgt und sich zur Vermeidung einer Unterrepräsentation des zugrunde liegenden Konstrukts auf eindeutig definierte PCK-Komponenten bezogen, weshalb KISR und KSU als die beiden zentralen (z. B. van Driel et al., 2023) und empirisch am häufigsten untersuchten Komponenten (Chan & Hume, 2019) ausgewählt wurden. Die Komponenten zeichnen sich darüber hinaus dadurch aus, dass sie nicht getrennt voneinander betrachtet werden sollten, sondern in Verbindung zueinander stehen (z. B. van Driel et al., 2023), was ebenfalls in häufig referierten Modellen zur Topologie von PCK betont wird (zum Beispiel in dem Modell von Magnusson et al., 1999 oder im Pentagon-Modell von Park & Oliver, 2008). Um zusätzlich diese Eigenschaft des zugrunde liegenden Konstrukts mit dem PCK_{SR}-bio abzubilden, wurden bei der Entwicklung die Verbindungen zwischen den betrachteten Komponenten (KISR→KSU; KSU→KISR) berücksichtigt. Dies stellt eine Besonderheit des Instruments dar, da diese Verbindungen zwischen PCK-Komponenten bisher primär durch qualitative Ansätze wie dem *PCK map approach* (z. B. Park & Suh, 2019) untersucht wurden.

Gleichzeitig sollte bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK eine Überrepräsentation des zugrunde liegenden Konstrukts vermieden werden, um die Erfassung konstruktirrelevanter Faktoren wie CK oder Überzeugungen zum Unterrichten zu vermeiden (Smith & Banilower, 2015). Dazu wurden die Items des PCK_{SR}-bio bei der Entwicklung zunächst durch eine Gruppe von Expert:innen mehrfach revidiert. Darüber hinaus erfolgte eine empirische Überprüfung der Konstruktrepräsentation. So zeigte sich in Studie 1, dass die betrachteten Noviz:innen und Expert:innen die Wahl der Antwortoptionen nicht mit CK und nur in einem Fall mit ihren Überzeugungen begründeten, was darauf schließen lässt, dass das zugrunde liegende Konstrukt nicht überrepräsentiert ist. Zusätzlich konnte in den Studien 2 und 3 gezeigt werden, dass die angehenden Lehrkräfte bei der Beantwortung des PCK_{SR}-bio die Auswahl eines Distraktors fast ausschließlich in Bezug auf Oberflächenmerkmalen begründeten, während sie bei der Auswahl eines Attraktors häufiger auf ihr PCK rekurrierten. Wenn sie darüber hinaus die Auswahl einer Antwortoption mit PCK begründeten, bezogen sie sich hauptsächlich auf diejenige PCK-Komponente (oder Verbindung), die durch die entsprechenden Items erfasst wurde, was die Annahme stützt, dass der PCK_{SR}-bio das zugrunde liegende Konstrukt angemessen repräsentiert. Für die Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK lässt sich daraus schließen, dass eine passende Konstruktrepräsentation sowohl durch eine systematische Entwicklung als auch daran anschließende empirische Überprüfung sichergestellt werden kann.

Eine weitere Herausforderung besteht hinsichtlich der psychometrischen Eigenschaften des Instruments. So weisen Multiple-Choice-Instrumente zur Messung von komplexen und anwendungsbezogenen Konstrukten häufig eine niedrigere interne Konsistenz im Vergleich zu Skalen oder Fragebögen auf (Taber, 2018), wobei in der Literatur kein eindeutiger Grenzwert definiert ist, ab dem ein Instrument als akzeptabel gilt (Schmitt, 1996; Taber, 2018). Auch der PCK_{SR}-bio wies in den im Rahmen dieser Dissertation durchgeführten Studien eine niedrigere interne Konsistenz auf, was einen Hinweis darauf gibt, dass das zugrunde liegende Konstrukt nicht zielgenau gemessen wird (Taber, 2018). Dies führt jedoch nicht automatisch zu einer Beeinträchtigung der Validität der Testwertinterpretation (Schmitt, 1996; Taber, 2018). Bereits Cronbach (1951) betonte, dass eine hohe interne Konsistenz erstrebenswert, die Validität der Testwertinterpretation jedoch als zentrales Gütekriterium entscheidend sei und auch bei Instrumenten mit niedriger interner Konsistenz erfüllt sein kann.

Beide Gütekriterien sind darüber hinaus nicht unabhängig voneinander. Das Reliabilitäts-Validitäts-Dilemma (*reliability/validity dilemma* oder *reliability-validity paradox*) beschreibt den einzugehenden Kompromiss, dass eine höhere Validität der Testwertinterpretation häufig mit einer niedrigeren internen Konsistenz einhergeht und vice versa (Banta & Palomba, 2014; Brennan, 1996; Clifton, 2020; Fendler, 2016; Slomp & Fuite, 2004). So kann bei der Erfassung komplexer Konstrukte eine umfangreiche Konstruktabdeckung zur Vermeidung von Unterrepräsentation zu einer Verringerung der internen Konsistenz führen (Clifton, 2020; Slomp & Fuite, 2004). Auch kann die Formulierung der Items einen Einfluss haben: Während komplexe und authentische Formulierungen die Validität erhöhen können, können sie ebenfalls dafür sorgen, dass es an der notwendigen Standardisierung fehlt, die für das Erreichen einer hohen internen Konsistenz notwendig ist (Banta & Palomba, 2014). Darüber hinaus kann auch die Verwendung von negativ gepolten Items sowie die Randomisierung der Reihenfolge der Items die Validität erhöhen, jedoch ebenfalls mit einer Verringerung der internen Konsistenz einhergehen (Clifton, 2020). Forschende sollten sich deshalb im Entwicklungsprozess von Multiple-Choice-Instrumenten zur Erfassung von PCK bewusst sein, dass das Begegnen der Herausforderungen in Bezug auf eine adäquate Formulierung der Items sowie der passenden Konstruktrepräsentation zu einer niedrigen internen Konsistenz führen kann. Von Änderungen des Instruments mit dem Ziel, die Reliabilität zu erhöhen, wird abgeraten, da dies im Gegenzug zu einer Verringerung der Validität führen kann (Brennan, 1996).

Für den PCK_{SR}-bio konnte insbesondere in Studie 2 gezeigt werden, dass eine valide Testwertinterpretation als PCK_{SR} möglich ist. So konnte Validitätsevidenz auf Grundlage des Testinhalts, von Antwortprozessen, der internen Struktur sowie der Beziehung zu anderen Variablen gesammelt werden. Darüber hinaus stützen auch die Ergebnisse der anderen Studien die Validität der Testwertinterpretation: Die gefundenen Unterschiede hinsichtlich des erreichten Testscores, der internen Konsistenz sowie des Antwortverhaltens zwischen Expert:innen und Noviz:innen aus Studie 1 entsprachen den zuvor formulierten Annahmen, was eine Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen darstellt. Auch durch die Verwendung des *PCK map approaches* in Studie 3 konnte die Validität der Testwertinterpretation basierend auf Antwortprozessen stützen, wobei dies eine alternative Darstellung der entsprechenden Evidenz aus Studie 2 entspricht. Da empfohlen wird, den *PCK map approach* in Kombination mit weiteren Erhebungs- und Auswertungsmethoden zu verwenden (Park & Suh, 2019), eignet er sich insbesondere zur Triangulation (Chan, 2022), weshalb die Ergebnisse der durchgeführten Studie die Nützlichkeit dieses Ansatzes im Validierungsprozess unterstreichen.

In Studie 4 entsprachen die Unterschiede im PCK_{SR}-Testscore zwischen den betrachteten Lehrenden der Biologie (Lehrkräfte an Grund- und weiterführender Schule sowie Hochschuldozierende in Lehrkräfteausbildung) den Erwartungen. Gemeinsam mit dem auf theoretischer Grundlage erwarteten Zusammenhang zwischen dem PCK_{SR} der Lehrenden und ihren transmissiven Überzeugungen (z. B. Großschedl et al., 2019) stützt dies die Validität der Testwertinterpretation als Validitätsevidenz auf Grundlage der Beziehung zu anderen Variablen. In Studie 5 konnte darüber hinaus für die englischsprachige Version des PCK_{SR}-bio gezeigt werden, dass angehende Lehrkräfte im Masterstudium in Australien, welche Biologie als Fach studierten, keine Unterschiede in ihrem PCK_{SR} im Vergleich zu Studierenden, welche ein anderes naturwissenschaftliches Fach studierten, aufwiesen. Da die in den Items beschriebenen Unterrichtssituationen einen biologischen Kontext adressieren, gibt dies Hinweise darauf, dass biologisches Fachwissen der Proband:innen keinen Einfluss auf ihren PCK_{SR}-Testscore hat. In Anbetracht dessen, dass es sich bei PCK_{SR} um disziplinspezifisches PCK handelt, über welches Lehrkräfte aller naturwissenschaftlichen Fächer verfügen sollten, unterstützen die Ergebnisse damit die Validität der Testwertinterpretation des PCK_{SR}-bio und geben Hinweise darauf, dass biologisches Fachwissen keinen konstruktirrelevanten Einflussfaktor darstellt.

Trotz dieser zahlreichen Evidenzen sollte berücksichtigt werden, dass nicht alle Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien die Validität der Testwertinterpretation des PCK_{SR}-bio stützen. In der in Studie 2 durchgeführten Interventionsstudie zeigte sich keine signifikante Veränderung des PCK_{SR}-Testscores der Versuchsgruppe nach der Teilnahme an einer Semesterveranstaltung, in welcher die im PCK_{SR}-bio adressierten Inhalte explizit vermittelt wurden. Da eine zu erwartende Verbesserung des PCK_{SR} (z. B. Großschedl et al., 2019; Kleickmann et al., 2013; Mahler et al., 2024) nicht gemessen werden konnte, gibt dies Hinweise darauf, dass der Lernfortschritt angehender Biologielehrkräfte während der universitären Ausbildung nicht adäquat durch den PCK_{SR}-bio abgebildet werden kann und das Instrument damit nicht veränderungssensitiv ist. Auch der fehlende Zusammenhang zwischen der Berufserfahrung und dem PCK_{SR}-Testscore der praktizierenden Lehrkräfte in Studie 5 bestärkt diese Vermutung.

Zusätzlich haben auch die geringen Stichprobengrößen in Studie 1 (je $n = 10$ Noviz:innen und Expert:innen) sowie in Studie 5 ($n = 36$ angehende Lehrkräfte) einen Einfluss auf die Robustheit der entsprechenden Validitätsevidenz in Bezug auf die Generalisierbarkeit der Ergebnisse. Ebenso sollte die Anzahl der Items sowie die Konzeption des PCK_{SR}-bio bei Einschätzung der Validität der Testwertinterpretation einschränkend berücksichtigt werden. So wurde jedem Schnittpunkt der PCK-Komponenten KISR und KSU sowie ihren Verbindungen KISR→KSU und KSU→KISR mit den naturwissenschaftlichen Denk- und Arbeitsweisen „Experimentieren“, „Hypothesen formulieren“ und „Modellieren“ jeweils nur ein Item zugeordnet (s. Tabelle 12.1). Auch wenn das zugrunde liegende Konstrukt durch die systematische Entwicklung des Instruments angemessen repräsentiert wird, ist es nicht möglich, einzelne Items auszuschließen (zum Beispiel aufgrund unzureichender Itemkennwerte), ohne das Konstrukt zu unterrepräsentieren und so die Validität der Testwertinterpretation einzuschränken (McCoach et al., 2013c). Um dieser Einschränkung zu begegnen, wurden über den Rahmen dieser Dissertation hinaus zusätzliche Items entwickelt, welche im Ausblick (Kapitel 10.5.3) vorgestellt werden.

Zusammenfassend lässt sich erkennen, dass auch wenn in den im Rahmen dieser Dissertation durchgeführten Studien vielfach die Validität der Testwertinterpretation des PCK_{SR}-bio gestützt werden konnte, ebenso die genannten Einschränkungen bei der Interpretation des Testscores als PCK_{SR} berücksichtigt werden müssen. Da die Validierung eines Testinstruments einen laufenden Prozess beschreibt (Cook & Hatala, 2016; Kane, 2013b), geben die vorgestellten Ergebnisse Anlass zur weiteren systematischen Untersuchung der Validität der Testwertinterpretation des PCK_{SR}-bio.

Vor dem Hintergrund der übergeordneten Fragestellung (I) geben die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien Hinweise darauf, dass den genannten Herausforderungen bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK hinsichtlich einer adäquaten Formulierung der Items, einer passenden Repräsentation des zugrunde liegenden Konstrukts sowie der psychometrischen Eigenschaften des Instruments durch die systematische Entwicklung und Validierung des PCK_{SR}-bio begegnet werden konnten. Auch wenn manche Autor:innen die Verwendung von Multiple-Choice-Instrumenten zur Erfassung von PCK kritisch sehen (Hill et al., 2008; Rosenkränzer et al., 2017; Schmelzing et al., 2013; Smith & Banilower, 2015), zeigen die vorgestellten Ergebnisse, dass eine systematische Entwicklung und Validierung helfen kann, den beschriebenen Herausforderungen bei der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK zu begegnen. Bei Verwendung des PCK_{SR}-bio zur Erfassung des deklarativen pPCK im Kompetenzbereich Erkenntnisgewinnung angehender und praktizierender Lehrkräfte sollte sich jedoch bewusst gemacht werden, dass die Validität der Testwertinterpretation durch die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien nicht ausschließlich gestützt werden konnte.

10.2 Inhaltliche Perspektive

Zur Beantwortung der übergeordneten Fragestellung (II) werden die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien aus einer inhaltlichen Perspektive diskutiert, um zu untersuchen, wie das PCK_{SR} angehender und praktizierender Lehrkräfte ausgeprägt ist und wie es mit der Berufserfahrung und anderen konzeptionell verwandten Variablen zusammenhängt.

Hinsichtlich der Ausprägung des PCK_{SR} angehender Lehrkräfte zeigen die Ergebnisse der im Rahmen von Studie 2 durchgeführten Querschnittstudie, dass die betrachteten Gruppen bereits über ein grundlegendes PCK_{SR} verfügten, welches sich zu verschiedenen Zeitpunkten der Ausbildung (Bachelor, Master und Referendariat) nicht unterschied, wobei ihr Antwortverhalten mit steigendem Ausbildungsstand konsistenter wurde. Ebenfalls zeigte sich unerwarteterweise kein Zusammenhang zwischen dem Testscore und der Semesterzahl der angehenden Lehrkräfte. Auch in Studie 5 konnte kein signifikanter Unterschied im PCK_{SR} zwischen angehenden Lehrkräften im ersten und zweiten Jahr des Masterstudiums gezeigt werden. Diese Ergebnisse geben Hinweise darauf, dass angehende Lehrkräfte zu Beginn ihres Studiums bereits über ein grundlegendes Verständnis über PCK_{SR} verfügen, sich dieses im Laufe des Studiums jedoch nicht maßgeblich weiterentwickelt und stehen damit im Widerspruch zu den Ergebnissen anderer Studien, welche einen Zuwachs des PCK über den Verlauf des Studiums (z. B. Großschedl et al., 2019; Kleickmann et al., 2013), auch in längsschnittlichen Studien (z. B. Mahler et al., 2024), nachweisen konnten. Da für den Erwerb von PCK spezifische Lerngelegenheiten in der universitären Ausbildung eine relevante Rolle zu spielen scheinen (Blömeke & Delaney, 2012; Riese & Reinhold, 2012), legen die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien die Vermutung nahe, dass mehr explizite Lerngelegenheiten für den Erwerb von PCK_{SR} in der universitären Lehrkräfteausbildung notwendig sind, sowohl in Deutschland als auch in Australien.

Zusätzlich konnte in der im Rahmen von Studie 2 durchgeführten Interventionsstudie kein signifikanter Anstieg des PCK_{SR} durch die Teilnahme an einer Semesterveranstaltung, welche explizit PCK_{SR} thematisierte, gezeigt werden. Eine mögliche Ursache dafür könnte sein, dass eine einzelne Lehrveranstaltung nicht zur effektiven Entwicklung von PCK_{SR} ausreicht. Ergänzend dazu zeigte sich in einer Studie von Schiering et al. (2021), dass das CK angehender Lehrkräfte insbesondere dann einen Einfluss auf die Entwicklung von PCK hat, wenn wenig fachdidaktische Lehrveranstaltungen im Rahmen der universitären Ausbildung angeboten werden. Da in anderen Studien nur ein moderater Anstieg von Erkenntnisgewinnungskompetenzen über den Verlauf des Studiums gezeigt werden konnte (Hartmann et al., 2015; Krüger et al., 2020; Mathesius et al., 2016), ebenfalls in Abhängigkeit der angebotenen Lerngelegenheiten (Hartmann et al., 2015; Krell, Khan et al., 2022), könnte dieser Einfluss nicht stark genug sein, um eine signifikante Veränderung des PCK_{SR} zu bewirken. Dies betont insbesondere die Relevanz expliziter Lerngelegenheiten für PCK zu spezifischen Themen in der universitären Lehramtsausbildung.

Bei dieser Interpretation sollte jedoch kritisch berücksichtigt werden, dass in der durchgeführten Semesterveranstaltung explizit die im PCK_{SR}-bio adressierten Inhalte vermittelt wurden, weshalb von einer Entwicklung des PCK_{SR} der Versuchsgruppe auszugehen ist. Diese fehlende Veränderung lässt damit gleichfalls die Vermutung zu, dass der PCK_{SR}-bio Unterschiede im Lernfortschritt nicht sensibel messen kann, was darin resultieren würde, dass eine valide Testwertinterpretation als PCK_{SR} nicht uneingeschränkt möglich ist. Als Schlussfolgerung daraus lässt sich kritisch hinterfragen, inwiefern mit dem PCK_{SR}-bio inhaltlich gültige Aussagen über die Entwicklung des PCK_{SR} getroffen werden können.

Die Erwartungen hinsichtlich der Ausprägung des PCK_{SR} unterschiedlicher Lehrender der Biologie konnten dahingegen in Studie 4 bestätigt werden, was die Vermutung unterstreicht, dass insbesondere Hochschuldozierende ein hohes pPCK im Kompetenzbereich Erkenntnisgewinnung aufweisen, da sie sich durch ihre Arbeit stärker mit dem aktuellen kanonischen cPCK auseinander setzen als praktizierende Lehrkräfte, von denen jedoch vermutet werden kann, dass sie stattdessen über höheres ePCK im Kompetenzbereich Erkenntnisgewinnung verfügen (Behling et al., 2022b; Chan & Hume, 2019; Irmer et al., 2023). Der Unterschied zwischen den praktizierenden Lehrkräften unterschiedlicher Schulformen impliziert darüber hinaus, dass PCK_{SR}, als für das Unterrichten naturwissenschaftlicher Kompetenzen notwendiger Wissensbereich, bei Lehrkräften an weiterführenden Schulen stärker ausgeprägt ist, da dieser Themenbereich einen größeren Bestandteil des Biologieunterrichts im Vergleich zum Sachkundeunterricht ausmacht. Entsprechend unterscheiden sich die damit einhergehenden curricularen Anforderungen (GDSU, 2002; KMK, 2024b). Es ist deshalb zu vermuten, dass Lehrkräfte an Grundschulen ein höheres PCK in Hinsicht auf andere, speziell für den Sachunterricht relevante Themen und Disziplinen aufweisen.

Ein Zusammenhang zwischen dem PCK_{SR}-Testscore und der Berufserfahrung konnte für die Lehrenden in Studie 4 jedoch nicht nachgewiesen werden. Andere empirische Studien, die diesen Zusammenhang untersuchten, kamen zu teilweise widersprüchlichen Ergebnissen, welche möglicherweise auf Unterschiede in den verwendeten Erhebungsmethoden und -instrumenten oder betrachteten PCK-Komponenten zurückzuführen sind. Viele Studien konnten einen positiven Zusammenhang zwischen dem PCK und der Berufserfahrung von Lehrkräften zeigen (z. B. E. Lee et al., 2007; Luft & Zhang, 2014; Schiering et al., 2022), einige Studien hingegen nicht (Friedrichsen et al., 2009). S. Kirschner et al. (2017) konnten sogar einen negativen Zusammenhang zeigen und vermuteten, dass Lehrkräfte mit wenig Berufserfahrung mehr Wissen hinsichtlich des in der universitären Ausbildung gelehrt aktuellen kanonischen cPCK aufweisen, welches Lehrkräften mit hoher Berufserfahrung während ihrer bereits länger zurückliegenden Ausbildung noch nicht vermittelt wurde. Dies könnte ebenfalls ein Grund für den fehlenden Zusammenhang in Studie 4 sein. Darüber hinaus zeigen diverse Studien, dass Lehrkräfte häufig selbst nicht über ausreichend Erkenntnisgewinnungskompetenzen verfügen (Capps & Crawford, 2013; Capps et al., 2016) oder Schwierigkeiten bei der Planung von Stunden zur Förderung von Erkenntnisgewinnungskompetenzen haben (Großmann & Krüger, 2024; Petermann & Vorholzer, 2023; Vo et al., 2015). Da das CK von Lehrkräften jedoch mit ihrem PCK zusammenhängt (z. B. Großschedl et al., 2019) und dies in Studie 2 ebenfalls für PCK_{SR} gezeigt werden konnte, lassen die Ergebnisse darauf schließen, dass Lehrkräfte unter Umständen gar nicht das erforderlichen CK aufweisen, um PCK_{SR} zu entwickeln. Dies wird gestützt durch die Ergebnisse einer Studie von Kulgemeyer et al. (2021), welche die zentrale Rolle der Reflexionsfähigkeit bei der Entwicklung von pPCK aus eigener Unterrichtserfahrung (ePCK) betonen. Auch bei dieser Interpretation ist jedoch kritisch zu berücksichtigen, dass dieses den Erwartungen widersprechende Ergebnis ebenfalls die Vermutung zulässt, dass eine valide Testwertinterpretation des PCK_{SR}-bio nicht uneingeschränkt möglich ist. Da mit dem Instrument ebenfalls kein Lernfortschritt im Rahmen der universitären Ausbildung gemessen werden konnte, stützen diese Ergebnisse die Annahme, dass der PCK_{SR}-bio nicht sensibel für Veränderungen ist, woraus die Notwendigkeit einer grundlegenden Überarbeitung des Instruments resultiert.

Hinsichtlich des Zusammenhangs mit anderen konzeptionell verwandten Variablen konnte in Studie 4 ein signifikant negativer Einfluss transmissiver Überzeugungen auf das PCK_{SR} von Hochschuldozierenden und Lehrkräften an Grundschulen statistisch nachgewiesen werden, für Lehrkräfte an weiterführenden Schulen hingegen nur deskriptiv. Dies geht mit den Erwartungen einher (Capps & Crawford, 2013; Großschedl et al., 2019) und betont den zentralen Einfluss der Überzeugungen von Lehrkräften, welche bei der Gestaltung von Aus- und Weiterbildungsangeboten folglich berücksichtigt werden sollten. Hinsichtlich der Offenheit zur Innovation zeigte sich ein

überraschenderweise negativer Zusammenhang für Lehrkräfte an weiterführenden Schulen, was die Notwendigkeit weiterer Forschung in diesem Bereich unterstreicht, um die Zusammenhänge von PCK_{SR} mit konzeptionell verwandten Variablen besser zu verstehen. Auch die Unterschiede zwischen den betrachteten Gruppen geben Anlass zur weiteren Untersuchung dieser. Zusätzlich sind auch Interaktionseffekte oder der Zusammenhang mit weiteren konzeptionell verwandten Variablen (z. B. Abiturnote, Großschedl et al., 2019; Zweitfach, Mahler et al., 2024; Welter et al., 2022) denkbar, welche im Rahmen dieser Dissertation jedoch nicht untersucht wurden.

Die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien geben damit vor dem Hintergrund der übergeordneten Fragestellung (II) einen Einblick bezogen auf die Ausprägung des PCK_{SR} angehender und praktizierender Lehrkräfte sowie den Zusammenhang mit der Berufserfahrung und anderen konzeptionell verwandten Variablen, sind jedoch vor dem Hintergrund der möglicherweise eingeschränkten Validität der Testwertinterpretation des PCK_{SR}-bio hinsichtlich ihrer Aussagekraft kritisch zu betrachten. Während PCK_{SR} als spezifischer Wissensbereich bei Lehrkräften unterschiedlicher Schulformen verschieden ausgeprägt zu sein scheint, geben die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien Hinweise darauf, dass sich das PCK_{SR} über den Verlauf der Lehramtsausbildung sowie Berufspraxis nicht automatisch weiterentwickeln, was die Notwendigkeit expliziter Aus- und Fortbildungsangebote betont.

10.3 Synthese beider Perspektiven

Die im Rahmen dieser Dissertation erlangten Ergebnisse hinsichtlich der Entwicklung, Validierung sowie Anwendung eines Multiple-Choice-Instruments zur Erfassung von PCK_{SR} angehender und praktizierender Lehrkräfte wurden zur Beantwortung der übergeordneten Fragestellung (I) aus einer methodischen und zur Beantwortung der übergeordneten Fragestellung (II) aus einer inhaltlichen Perspektive diskutiert. Dabei zeigt sich, dass die beiden Perspektiven nicht ausschließlich getrennt voneinander zu betrachten sind: Ergebnisse der primär aus methodischer Perspektive durchgeführten Studien 1, 2 und 3 konnten ebenfalls zur Beantwortung der übergeordneten Fragestellung (II), Ergebnisse der primär aus inhaltlicher Perspektive durchgeführten Studien 4 und 5 konnten ebenfalls zur Beantwortung der übergeordneten Fragestellung (I) verwendet werden. Für eine holistische Betrachtung der im Rahmen dieser Dissertation erlangten Ergebnisse ist somit eine Synthese beider Perspektiven notwendig.

Dies lässt sich ebenfalls hinsichtlich der internen Konsistenz veranschaulichen. Einerseits stellt diese ein wichtiges Gütekriterium zur Schätzung des Grads der Messgenauigkeit eines Instruments dar (Döring, 2023). Andererseits kann die interne Konsistenz aber auch dazu genutzt werden, inhaltliche Aussagen über die Konsistenz des Wissens der getesteten Personen zu tätigen und basierend auf entsprechenden Erwartungen die Validität der Testwertinterpretation zu überprüfen. So zeigen die Ergebnisse aus Studie 1 und 4, dass Personen, die einen höheren PCK_{SR}-Testscore erreichten, ebenfalls konsistenter antworteten. Dies geht mit der theoretischen Annahme einher, dass Personen mit einer höheren Expertise ihr Wissen auf unterschiedliche Kontexte anwenden können (Hogan et al., 2003; Krepf et al., 2018; Wolff et al., 2015) und deshalb eine niedrigere Varianz in ihrem Antwortverhalten aufzeigen als Personen mit niedrigerer Expertise. Da die interne Konsistenz eines Testinstruments auch von der Gesamtzahl der Items abhängt (Fischer et al., 2023; Taber, 2018), bedeutet dies für die Entwicklung von Testinstrumenten, dass eine höhere Anzahl an Items notwendig sein kann, um aussagekräftige Ergebnisse von Zielgruppen mit einer eher niedrig zu erwartenden Expertise zu erlangen. Um dies bezogen auf den PCK_{SR}-bio zu untersuchen, wurden über den Rahmen dieser Dissertation hinaus weitere Items entwickelt (s. Kapitel 10.5.3).

Die Synthese beider im Rahmen dieser Dissertation betrachteten Perspektiven lässt darüber hinaus weitere relevante Rückschlüsse für die Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK zu. So zeigen die Ergebnisse aus Studie 1 und 4, dass es sinnvoll sein kann, die primäre Zielgruppe bereits während der Entwicklung eines Instruments zu berücksichtigen: Instrumente, die sich an Personen mit geringer Expertise hinsichtlich des zu erfassenden Konstrukts richten, sollten einen niedrigeren Schwierigkeitsgrad aufweisen als Instrumente, die sich primär an Personen mit hoher Expertise hinsichtlich des zu erfassenden Konstrukts richten, um das PCK der jeweiligen Gruppe eindeutiger abbilden zu können. Damit zusammenhängend gibt die fehlende Veränderung des PCK_{SR}-Testscores in der im Rahmen von Studie 2 durchgeführten Interventionsstudie Hinweise darauf, dass die Items einen zu hohen Schwierigkeitsgrad für die angehenden Lehrkräfte aufweisen und hinsichtlich dessen optimiert werden könnten. In Studie 1 zeigte sich darüber hinaus, dass formale Charakteristika der Items (Oberflächenmerkmale) ebenfalls einen Einfluss auf das Antwortverhalten von Personen mit geringer Expertise haben können. Mehrfache Überarbeitungen und Revisionsschleifen im Entwicklungsprozess können deshalb dabei helfen, diese Merkmale zu eliminieren oder konstant zu halten, um zu verhindern, dass Personen mit geringer Expertise ihre Entscheidung für eine Antwortoption aus konstruktirrelevanten Gründen treffen.

Im Rahmen dieser Dissertation erfolgte eine Überprüfung von Oberflächenmerkmalen für den PCK_{SR}-bio sowohl durch eine mehrfache Revision der Items durch eine Gruppe von Expert:innen während des Entwicklungsprozesses als auch durch die sprachlichen Überarbeitung in Anschluss an Studie 1, bei denen die impliziten und expliziten Rückmeldungen der Noviz:innen und Expert:innen dazu verwendet wurden, die Verständlichkeit der Items zu erhöhen. Zusätzlich dazu konnte im Rahmen einer Masterarbeit gezeigt werden, dass Oberflächenmerkmale der Items des PCK_{SR}-bio hinsichtlich ihrer Textkomplexität (zum Beispiel die durchschnittliche Satzlänge in Wörtern oder die durchschnittliche Silbenzahl pro Wort) keinen signifikanten Einfluss auf die Schwierigkeit des PCK_{SR}-bio hatten (Gluth, 2023).

Die im Rahmen dieser Dissertation erlangten Erkenntnisse verdeutlichen damit die Relevanz des Zusammenhangs der methodischen sowie inhaltlichen Perspektive und die Notwendigkeit einer daraus resultierenden differenzierten Betrachtung der Ergebnisse. So beeinflusst die methodische Perspektive maßgeblich die Aussagekraft der Ergebnisse (inhaltliche Perspektive), da diese von dem entwickelten Instrument und seiner Passung zum zugrundeliegenden Konstrukt, der internen Konsistenz sowie der Validität der Testwertinterpretation abhängt. Den Erwartungen widersprechende Ergebnisse können damit Hinweise darauf geben, dass eine valide Interpretation des Testwerts als das zugrunde liegende Konstrukt nicht uneingeschränkt möglich ist, was in einer Revision des verwendeten Instruments resultiert. Umgekehrt zeigt sich jedoch ebenfalls, dass die inhaltliche Perspektive einen Einfluss auf die methodische Perspektive haben kann. Von Personen, die über wenig Expertise in dem zu erfassenden Gebiet verfügen, kann kein konsistentes Antwortverhalten erwartet werden, was wiederum einen Einfluss auf die interne Konsistenz der Messung haben kann. Darüber hinaus ist die Entwicklung von Wissen nur möglich, wenn geeignete Lerngelegenheiten vorhanden sind und genutzt werden. Den Erwartungen widersprechende Ergebnisse können damit also auch Hinweise auf die Ausprägung des betrachteten Konstrukts sowie den Zusammenhang zu konzeptionell verwandten Variablen geben.

10.4 Limitationen

Eine Diskussion der Limitation der im Rahmen dieser Dissertation durchgeführten Studien erfolgte bereits in den jeweiligen Kapiteln 4 bis 8. Im Folgenden sollen deshalb übergreifende Limitationen dieser Dissertation diskutiert werden.

Als Grundlage für die Entwicklung des PCK_{SR}-bio wurde eine transformative Konzeptualisierung von PCK gewählt, weshalb sich auf die Erfassung spezifischer Komponenten und ihrer Verbindungen fokussiert wurde. Auch wenn diese als zentral angesehen werden (z. B. van Driel et al., 2023), umfasst PCK noch weitere Komponenten, etwa das Wissen über das Curriculum oder das Wissen über Bewertung (z. B. Park & Oliver, 2008), welche mit dem PCK_{SR}-bio jedoch nicht erfasst werden können. Auch hinsichtlich der Erkenntnisgewinnungskompetenzen können mit dem PCK_{SR}-bio nur zwei ausgewählte Arbeitsweisen sowie eine ausgewählte Denkweise erfasst werden, obwohl dieses Konstrukt noch weitere zentrale Arbeitsweisen (etwa das Beobachten und Vergleichen, z. B. Wellnitz & Mayer, 2016; auch curricular verankert, z. B. KMK, 2024b) oder Denkweisen (etwa das Formulieren von Fragestellungen, z. B. Rönnebeck et al., 2016) umfasst. Die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien erlauben deshalb nur eingeschränkte Aussagen hinsichtlich des PCK_{SR}, bezogen auf ebendiese erfassten PCK-Komponenten sowie Denk- und Arbeitsweisen. Auch wenn es durch eine fehlende einheitliche Definition von Erkenntnisgewinnungskompetenzen (vgl. Krell, Vorholzer & Nehring, 2022) nicht möglich ist, dieses Konstrukt vollumfassend zu erfassen und bei der Erfassung von PCK mit Multiple-Choice-Instrumenten darüber hinaus empfohlen wird, sich auf wenige, eindeutig definierte PCK-Komponenten zu beziehen (Smith & Banilower, 2015), sollte dies bei der Interpretation der Ergebnisse berücksichtigt werden.

Zusätzlich hängt der Anwendungsbereich von Multiple-Choice-Instrumenten zur Erfassung von PCK entscheidend von der Definition dieses Konstrukts (etwa hinsichtlich der betrachteten Bereiche nach dem Refined-Consensus-Modell; Carlson et al., 2019) ab. Dies wird im Vergleich mit der Definition anderer Autor:innen deutlich. So beschreiben beispielsweise Stender und Brückmann (2020) cPCK als die kanonische Grundlage, über die auch einzelne Personen (in Teilen) verfügen können, wohingegen sie pPCK als die Gesamtheit der persönlichen Erfahrungen beschreiben, welche nicht hinsichtlich ihrer Korrektheit bewertet werden kann. Diese Definition geht mit der Anwendung unterschiedlicher Auswertungsinstrumente und -methoden einher; Multiple-Choice-Instrumente eignen sich ausschließlich zur Erfassung von cPCK, offene Paper-Pencil-Tests sowie andere Auswertungsmethoden hingegen zur Erfassung von pPCK (Stender & Brückmann, 2020). Bei der Interpretation der Ergebnisse dieser Dissertation muss deshalb die zugrunde liegende Definition von PCK beachtet werden, insbesondere beim Vergleich mit den Ergebnissen anderer Studien.

Daran anschließend sollte beachtet werden, dass im Rahmen dieser Dissertation lediglich das statische pPCK im Kompetenzbereich Erkenntnisgewinnung erfasst wurde. Auch wenn dieses in Verbindung mit dem dynamischen ePCK steht (z. B. Carlson et al., 2019), kann nicht sichergestellt werden, dass Lehrkräfte dieses mit Multiple-Choice-Instrumenten erfassbare PCK ebenfalls in der Unterrichtspraxis zeigen (Chan & Hume, 2019).

Weiterhin sollte bei der Interpretation beachtet werden, dass nicht alle Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien den Erwartungen entsprachen, weshalb eine valide Testwertinterpretation nicht uneingeschränkt angenommen werden sollte. Dies legt weitere Untersuchungen nahe, um Rückschlüsse hinsichtlich der Notwendigkeit einer Revision des Instruments zu ziehen. Die niedrige interne Konsistenz des PCK_{SR}-bio in den unterschiedlichen Studien gibt darüber hinaus Hinweise auf eine niedrige Trennschärfe der Items, sodass das zugrunde liegende Konstrukt unter Umständen nicht zielgenau gemessen wird (Taber, 2018). Möglicherweise resultiert die niedrige interne Konsistenz in den im Rahmen dieser Dissertation durchgeführten Studien daher, dass die

Distraktoren insbesondere für Personen mit niedriger Expertise attraktiv scheinen, da sie authentische Instruktionsstrategien beschreiben. Damit könnten insbesondere die Oberflächenmerkmale der Items einen Einfluss auf das Antwortverhalten von Personen mit niedriger Expertise haben. Eine Möglichkeit zur Erhöhung der Reliabilität kann die Entwicklung weiterer Items sein (Fischer et al., 2023), was jedoch mit einer längeren Bearbeitungsdauer einhergeht und deshalb insbesondere bei Multiple-Choice-Instrumenten, deren Vorteil in der Testökonomie liegt, kritisch abgewogen werden sollte. Da der PCK_{SR}-bio jedoch eine überschaubare Anzahl an Items aufweist, wurden über den Rahmen dieser Dissertation hinaus weitere Items entwickelt, um den Einfluss auf die interne Konsistenz zu prüfen (s. Kapitel 10.5.3).

Zuletzt sollte die Inkonsistenz zwischen den durchgeführten Studien berücksichtigt werden. Dies bezieht sich zum einen auf die synonyme Nutzung unterschiedlicher Begriffe (zum Beispiel *scientific reasoning* oder *scientific reasoning competencies* für Erkenntnisgewinnungskompetenzen), die aus dem Fehlen einer einheitlichen Konzeptualisierung resultiert. Zum anderen muss die Inkonsistenz hinsichtlich des PCK_{SR}-bio beachtet werden, da dieser im Laufe der Dissertation (geringfügig) sprachlich überarbeitet wurde. Ein Einfluss auf die Ergebnisse der entsprechenden Studien ist unwahrscheinlich, jedoch nicht auszuschließen. Auch die Übersetzung des PCK_{SR}-bio für die Anwendung in Studie 5 kann einen Einfluss auf die Validität der Testwertinterpretation und damit die Ergebnisse haben. Dies ist insbesondere vor dem Hintergrund relevant, dass ein Großteil der Validitätsevidenz ausschließlich für die deutschsprachige Version des Instruments gesammelt wurde und nicht automatisch für die englischsprachige Version angenommen werden kann.

Trotz dieser Einschränkungen liefern die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien wichtige Aspekte hinsichtlich der Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK sowie eines grundlegenden Verständnisses von PCK_{SR}.

10.5 Implikationen und Ausblick

Im Folgenden sollen grundlegende Implikationen basierend auf den Ergebnissen der im Rahmen dieser Dissertation durchgeführten Studien für die Forschung und Praxis vorgestellt sowie ein Ausblick gegeben werden.

10.5.1 Implikationen für die Forschung

Insbesondere vor dem Hintergrund der übergeordneten Fragestellung (I) machen die Ergebnisse der durchgeführten Studien die Notwendigkeit einer transparenten Definition von PCK aufgrund der Komplexität als zugrunde liegendes Konstrukt deutlich. So führen einige Autor:innen die unterschiedlichen Ergebnisse diverser Studien auf konzeptionelle und methodische Unterschiede zurück (z. B. Mahler et al., 2017). Die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien zeigen, dass es notwendig ist, PCK als das zugrunde liegende Konstrukt eindeutig zu definieren und die primäre Zielgruppe bei der Entwicklung eines Instruments zu beachten. Forschende sollten sich bei der Auswahl und Entwicklung von Erhebungsmethoden und -instrumenten demnach bewusst sein, welche Form von PCK sie erfassen möchten. Zusätzlich sollte PCK als das zugrunde liegende Konstrukt eindeutig definiert sowie der Entwicklungsprozess von Testinstrumenten zur Erfassung dieses Wissensbereichs transparent gemacht werden, da dies mit unterschiedlichen Erhebungsmethoden einhergeht, welche wiederum mit verschiedenen Herausforderungen verbunden sind und dadurch Auswirkungen auf die Ergebnisse haben können.

Die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien zeigen weiterhin, dass es sinnvoll ist, auch die Verbindung zwischen zwei Komponenten mit Multiple-Choice-Instrumenten zu erfassen, welche bisher hauptsächlich mit qualitativen Ansätzen untersucht wurden (zum Beispiel dem *PCK map approach*; Park & Suh, 2019). Dies ermöglicht nicht nur eine umfassendere Repräsentation von PCK als zugrunde liegendes Konstrukt, sondern auch, dieses differenzierter zu betrachten, da sich hohes PCK insbesondere durch die Verbindung dieser Komponenten auszeichnet.

Darüber hinaus zeigen die Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien, dass die interne Konsistenz nicht nur als Gütekriterium zur Schätzung des Grads der Messgenauigkeit eines Instruments verwendet werden sollte. So eignet sich die Betrachtung der internen Konsistenz ebenfalls dazu, inhaltliche Aussagen hinsichtlich der Flexibilität des Wissens in Abhängigkeit der Expertise der betrachteten Proband:innen zu tätigen. Dabei sollte insbesondere die Synthese aus methodischer und inhaltlicher Perspektive beachtet werden: Während eine geringe interne Konsistenz dazu führen kann, dass das zu erfassende Konstrukt nicht zuverlässig gemessen wird und damit einen Einfluss auf die Aussagekraft der Ergebnisse haben kann, ist nicht zu erwarten, dass eine Stichprobe bei Verwendung des Instruments eine hohe interne Konsistenz aufweist, wenn die betrachteten Proband:innen selbst nicht über ausreichend Wissen verfügen, welches durch das Instrument abgebildet werden kann.

10.5.2 Implikationen für die Praxis

Basierend auf den Ergebnissen der im Rahmen dieser Dissertation durchgeführten Studien können auch Implikationen für die Praxis abgeleitet werden. So zeigen die Ergebnisse keinen Zusammenhang zwischen dem PCK_{SR} angehender und praktizierender Lehrkräfte mit ihrer Berufserfahrung. Dies gibt Hinweise auf das Fehlen expliziter Lerngelegenheiten im Studium, welche jedoch für den Erwerb dieses spezifischen Wissensbereichs notwendig zu sein scheinen. Für praktizierende Lehrkräfte macht dies die Notwendigkeit spezifischer Weiterbildungsangebote zur Förderung von PCK_{SR} deutlich. Dabei ist insbesondere eine explizite Adressierung von Erkenntnisgewinnungskompetenzen unter Berücksichtigung der Vorstellungen der Lerngruppe notwendig (Sannert, Petermann, et al. 2024). Bei der Gestaltung entsprechender Angebote sollten darüber hinaus die Schulform (Grund- oder weiterführende Schule) sowie weitere konzeptionell verwandte Variablen wie die epistemologischen Überzeugungen berücksichtigt werden, da dies einen Einfluss auf die Ausprägung des PCK_{SR} angehender und praktizierender Lehrkräfte haben kann.

Der PCK_{SR}-bio kann ebenfalls als formative Bewertungsmethode eingesetzt werden, um gezielt Rückmeldung hinsichtlich des aktuellen Wissensstands angehender oder praktizierender Lehrkräfte zu erlangen. Dabei ist jedoch zu beachten, dass eine valide Testwertinterpretation als PCK_{SR} nach aktuellem Stand nicht uneingeschränkt möglich ist und die Ergebnisse der durchgeführten Studien Hinweise darauf geben, dass das Instrument nicht veränderungssensibel ist und damit eine Entwicklung des PCK_{SR} bei Personen mit niedriger Expertise nicht adäquat abbilden kann. In jedem Fall kann der PCK_{SR}-bio jedoch zur Selbstevaluation verwendet werden, um das eigene PCK_{SR} kontinuierlich zu reflektieren und gezielt Ansatzpunkte für dessen Verbesserung zu identifizieren. Dazu ist es möglich, die Items zunächst selbstständig zu beantworten und die eigenen Antworten im Anschluss mit den begründeten Lösungen zu vergleichen (Tabelle 12.2).

10.5.3 Ausblick

Die im Rahmen dieser Dissertation vorgestellten Ergebnisse bieten einen Überblick über die Erfassung von PCK mit Multiple-Choice-Instrumenten sowie ein grundlegendes Verständnis über PCK_{SR}. Aufbauend auf diesen Erkenntnissen wurde an den im Rahmen dieser Dissertation durchgeführten Studien angeknüpft, welches im Folgenden kurz vorgestellt werden soll.

Da der PCK_{SR}-bio in den durchgeführten Studien eine niedrige interne Konsistenz aufwies, wurden insgesamt zwölf zusätzliche Items entwickelt (ein weiteres zu jedem Schnittpunkt der Systematik; s. Tabelle 12.3), welche in den im Rahmen dieser Dissertation vorgestellten Studien jedoch noch nicht eingesetzt wurden. Mitsamt dieser Ergänzung umfasst die erweiterte Version des Instruments damit insgesamt 24 Items. Bei der Entwicklung wurde sich an dem systematischen Entwicklungsprozess der bisherigen Items des PCK_{SR}-bio orientiert, weshalb die neuen Items ebenfalls basierend auf theoretischer und empirischer Grundlage entwickelt und mehrfach durch eine Gruppe von Expert:innen revidiert wurden. Zusätzlich wurden die biologischen Unterrichtskontexte der neuen Items bewusst so gewählt, dass sie diejenigen Basiskonzepte (vgl. KMK, 2024b) abdecken, die in der im Rahmen dieser Dissertation verwendeten Version des Instruments unterrepräsentiert waren. Auf diese Weise soll der Einfluss des Unterrichtskontexts als Oberflächenmerkmal reduziert werden. Im Anschluss wurden die ursprünglichen Items gemeinsam mit den neu entwickelten Items im Rahmen einer Masterarbeit pilotiert, wobei sich kein signifikanter Unterschied im Gesamtttestscore zwischen Studierenden und Lehrkräften sowie zwischen Lehrkräften mit naturwissenschaftlichem Zweitfach und ohne feststellen ließ (Meinshausen, 2024). Hinsichtlich der internen Konsistenz zeigte sich darüber hinaus ein großer Unterschied im Antwortverhalten zwischen den angehenden Lehrkräften im Bachelor ($\alpha = .20$; $n = 15$) und Master ($\alpha = .49$; $n = 11$). Im Vergleich zu Letzteren wies die Stichprobe praktizierender Lehrkräfte hingegen eine etwas geringere interne Konsistenz auf ($\alpha = .44$; $n = 225$; Meinshausen, 2024). Es zeigt sich damit, dass die interne Konsistenz der erweiterten Version mit 24 Items für angehende Lehrkräfte im Bachelor niedriger und für angehende Lehrkräfte im Master geringfügig höher ist als in der vergleichbaren Stichprobe aus Studie 2. Für praktizierende Lehrkräfte ist die interne Konsistenz im Vergleich zu der Stichprobe aus Studie 4 deutlich höher. Bei diesem Vergleich müssen jedoch die unterschiedlichen Stichprobengrößen berücksichtigt werden. Die Erweiterung des Instruments scheint damit eine sinnvolle Möglichkeit darzustellen, die interne Konsistenz des PCK_{SR}-bio bei praktizierenden Lehrkräften zu erhöhen, geht allerdings mit einer höheren Testbearbeitungszeit einher. Darüber hinaus ist eine Überprüfung der Validität der Testwertinterpretation der neuen Items in weiteren Studien notwendig.

Ebenfalls wurde in Anschluss an die im Rahmen dieser Dissertation erzielten Ergebnisse untersucht, welche Lerngelegenheiten angehende Lehrkräfte für den Erwerb von PCK_{SR} in der universitären Lehrkräfteausbildung wahrnehmen können. Dazu wurde eine Fokusgruppen-Interviewstudie mit Masterstudierenden durchgeführt, welche sich dazu äußerten, zu welchen Kompetenzbereichen sie im Rahmen ihres Lehramtsstudiums am meisten gelernt hatten und in welchen Lehrveranstaltungen sie explizit PCK_{SR} erlangen konnten. Eine vergleichbare Studie ist ebenfalls mit australischen Studierenden geplant. Entsprechende Ergebnisse sollen Hinweise über das vermutete Fehlen geeigneter Lerngelegenheiten zur Förderung von PCK_{SR} in der universitären Lehrkräfteausbildung geben.

Da einige Ergebnisse der im Rahmen dieser Dissertation durchgeführten Studien den Erwartungen widersprochen haben und damit die Validität der Testwertinterpretation einschränken, stellt eine Revision des Instruments beziehungsweise die Adaption des Instruments für spezifische Stichproben ein weiteres mögliches Vorgehen dar. Auf diese Weise kann der PCK_{SR}-bio beispielsweise mit dem Ziel optimiert werden, den Lernfortschritt (angehender) Lehrkräfte veränderungssensibel abzubilden, beispielsweise durch eine Anpassung der Schwierigkeit der Items. Auch ist es möglich, zusätzliche Items in Bezug auf weitere PCK-Komponenten (und ihre Verbindungen) oder weitere naturwissenschaftliche Denk- und Arbeitsweisen zu entwickeln, um das zugrunde liegende Konstrukt umfassender abzubilden.

Auch wurden die Erkenntnisse hinsichtlich der Entwicklung des PCK_{SR} im Rahmen dieser Dissertation ausschließlich in Querschnittstudien erlangt. Insbesondere um die Veränderung des PCK_{SR} angehender und praktizierender Lehrkräfte über den Verlauf der universitären Ausbildung, der Berufspraxis sowie dem Übergang zwischen diesen Phasen zu untersuchen, bedarf es Längsschnittstudien, durch die Entwicklungsverläufe individuell abgebildet und analysiert werden können.

Im Rahmen dieser Dissertation wurde zusammenfassend ein Multiple-Choice-Instrument zur Erfassung von PCK_{SR} angehender und praktizierender Biologielehrkräfte entwickelt, dessen Validität der Testwertinterpretation basierend auf zahlreichen Evidenzquellen gestützt werden kann, jedoch nicht uneingeschränkt angenommen werden sollte. Zusätzlich geben die Ergebnisse der durchgeführten Studien aus einer methodischen Perspektive Hinweise für die Entwicklung von Multiple-Choice-Instrumenten zur Erfassung von PCK im Allgemeinen und aus einer inhaltlichen Perspektive erste Einblicke über die Ausprägung des PCK_{SR} angehender und praktizierender Lehrkräfte sowie dessen Zusammenhang mit konzeptionell verwandten Variablen. Zusätzlich zeigt sich, dass eine Synthese beider Perspektiven für eine holistische Betrachtung der Ergebnisse notwendig ist.

11 Literaturverzeichnis

Anmerkung: Die Literaturangaben der einzelnen Manuskripte (Kapitel 4 bis 8) befinden sich am Ende der jeweiligen Kapitel.

- Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education*, 30(10), 1405–1416.
<https://doi.org/10.1080/09500690802187041>
- Abrams, E., Southerland, S. A. & Silva, P. (2008) (Hrsg.). *Contemporary research in education. Inquiry in the classroom: Realities and opportunities*. Information Age Publishing.
- ACARA (2022). *Science. F-10 Version 9.0. Scope and sequence*.
<https://v9.australiancurriculum.edu.au/content/dam/en/curriculum/ac-version-9/downloads/science/science-scope-and-sequence-f-10-v9.docx>
- Aditomo, A. & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes: Evidence from high and low-performing education systems. *International Journal of Science Education*, 42(4), 504–525.
<https://doi.org/10.1080/09500693.2020.1716093>
- AERA, APA & NCME (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- AITSL. (2017). *Australian Professional Standards for Teachers*. <https://www.aitsl.edu.au/standards>
- Akuma, F. V. & Callaghan, R. (2018). A systematic review characterizing and clarifying intrinsic teaching challenges linked to inquiry-based practical work. *Journal of Research in Science Teaching*, 56(5), 619–648. <https://doi.org/10.1002/tea.21516>
- Alonzo, A. C., Berry, A. & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: Enacted and personal PCK. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 273–288). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_12
- Alonzo, A. C. & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, 53(8), 1259–1286. <https://doi.org/10.1002/tea.21271>
- Alonzo, A. C., Kobarg, M. & Seidel, T. (2012). Pedagogical content knowledge as reflected in teacher-student interactions: Analysis of two video cases. *Journal of Research in Science Teaching*, 49(10), 1211–1239. <https://doi.org/10.1002/tea.21055>
- Aydin, S. & Boz, Y. (2013). The nature of integration among PCK components: A case study of two experienced chemistry teachers. *Chemistry Education Research and Practice*, 14(4), 615–624. <https://doi.org/10.1039/C3RP00095H>
- Banta, T. W. & Palomba, C. A. (2014). *Assessment essentials: Planning, implementing, and improving assessment in higher education* (2. Aufl.). Wiley.
<http://gbv.ebib.com/patron/FullRecord.aspx?p=1782543>
- Barendsen, E. & Henze, I. (2019). Relating teacher PCK and teacher practice using classroom observation. *Research in Science Education*, 49(5), 1141–1175.
<https://doi.org/10.1007/s11165-017-9637-z>
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Baumert, J. & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Mathematics teacher education: Bd. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (S. 25–38). Springer.
https://doi.org/10.1007/978-1-4614-5149-5_2

- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2011). Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die mathematische Kompetenz von Schülerinnen und Schülern (COACTIV) – Ein Forschungsprogramm. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 7–25). Waxmann.
- Baur, A., Emden, M. & Bewersdorff, A. (2019). Welche Unterrichtsprinzipien sollten für den Aufbau von Kompetenzen zum Experimentieren Beachtung finden? Eine Ableitung auf Basis multiperspektivisch begründeter Unterrichtsziele. *Zeitschrift für Didaktik der Biologie*, 23, 10–24. <https://doi.org/10.4119/zdb-1738>
- Baur, A., Xenofontos, N. & Papaevripidou, M. (2022). Inquiry-based learning (Forschendes Lernen). In A. Baur, N. Baumgartner-Hirscher & A. Lehtinen (Hrsg.), *Differenzierung beim Inquiry-based Learning im naturwissenschaftlichen Unterricht: Ein Differenzierungstool für das Experimentieren im Sinne des Forschenden Lernens* (S. 34–59). Beltz Juventa.
- Bayram-Jacobs, D., Henze, I., Evagorou, M., Shwartz, Y., Aschim, E. L., Alcaraz-Dominguez, S., Barajas, M. & Dagan, E. (2019). Science teachers' pedagogical content knowledge development during enactment of socioscientific curriculum materials. *Journal of Research in Science Teaching*, 56(9), 1207–1233. <https://doi.org/10.1002/tea.21550>
- BCTC (2019). *Professional Standards for BC Educators*. https://www2.gov.bc.ca/assets/gov/education/kindergarten-to-grade-12/teach/teacher-regulation/standards-for-educators/edu_standards.pdf
- Becerra, B., Núñez, P., Vergara, C., Santibáñez, D., Krüger, D. & Cofré, H. (2022). Developing an instrument to assess pedagogical content knowledge for evolution. *Research in Science Education*, 53(8), 213–229. <https://doi.org/10.1007/s11165-022-10042-0>
- Behling, F., Förtsch, C. & Neuhaus, B. J. (2022a). The refined consensus model of pedagogical content knowledge (PCK): Detecting filters between the realms of PCK. *Education Sciences*, 12(9), 592. <https://doi.org/10.3390/educsci12090592>
- Behling, F., Förtsch, C. & Neuhaus, B. J. (2022b). Using the plan–teach–reflect cycle of the refined consensus model of PCK to improve pre-service biology teachers' personal PCK as well as their motivational orientations. *Education Sciences*, 12(10), 654. <https://doi.org/10.3390/educsci12100654>
- Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S. & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, 53(7), 1082–1112. <https://doi.org/10.1002/tea.21257>
- Berry, A., Depaepe, F. & van Driel, J. (2016). Pedagogical content knowledge in teacher education. In J. Loughran & M. L. Hamilton (Hrsg.), *International handbook of teacher education* (S. 347–386). Springer. https://doi.org/10.1007/978-981-10-0366-0_9
- Berry, A., Friedrichsen, P. & Loughran, J. (Hrsg.). (2015). *Re-examining pedagogical content knowledge in science education*. Routledge. <https://doi.org/10.4324/9781315735665>
- Bicak, B. E., Borchert, C. E. & Höner, K. (2021). Measuring and fostering preservice chemistry teachers' scientific reasoning competency. *Education Sciences*, 11(9), 496. <https://doi.org/10.3390/educsci11090496>
- Blömeke, S. & Delaney, S. (2012). Assessment of teacher knowledge across countries: a review of the state of research. *ZDM Mathematics Education*, 44(3), 223–247. <https://doi.org/10.1007/s11858-012-0429-7>
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies: Competence Viewed as a Continuum. *Zeitschrift für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>

- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G. & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, 79, 101600. <https://doi.org/10.1016/j.learninstruc.2022.101600>
- Borko, H. & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473–498. <https://doi.org/10.3102/00028312026004473>
- Börling, J. & Labudde, P. (2014). Practical work in physics instruction: an opportunity to learn? In H. E. Fischer, K. Neumann & V. Jouni (Hrsg.), *Quality of instruction in physics: Comparing Finland, Germany and Switzerland* (S. 111–127). Waxmann.
- Brennan, R. (1996). Generalizability of performance assessments. In G. W. Phillips (Hrsg.), *Technical issues in large-scale performance assessment* (S. 19–58).
- Buldu, E. & Buldu, M. (2021). Investigating pre-service early childhood teachers' cPCK and pPCK on the knowledge used in scientific process through CoRe. *SAGE Open*, 11(2). <https://doi.org/10.1177/21582440211025564>
- Bybee, R. W. (2013). *The case for STEM education: Challenges and opportunities*. National Science Teachers Association.
- Capps, D. K. & Crawford, B. A. (2013). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education*, 24(3), 497–526. <https://doi.org/10.1007/s10972-012-9314-z>
- Capps, D. K., Shemwell, J. T. & Young, A. M. (2016). Over reported and misunderstood? A study of teachers' reported enactment and knowledge of inquiry-based science teaching. *International Journal of Science Education*, 38(6), 934–959. <https://doi.org/10.1080/09500693.2016.1173261>
- Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., . . . Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 77–94). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_2
- Chan, K. K. H. (2022). A critical review of studies using the pedagogical content knowledge map approach. *International Journal of Science Education*, 44(3), 487–513. <https://doi.org/10.1080/09500693.2022.2035011>
- Chan, K. K. H. & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 3–76). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_1
- Chen, C.-T. & She, H.-C. (2015). The effectiveness of scientific inquiry with/without integration of scientific reasoning. *International Journal of Science and Mathematics Education*, 13(1), 1–20. <https://doi.org/10.1007/s10763-013-9508-7>
- Chiu, M.-H. & Lin, J.-W. (2019). Modeling competence in science education. *Disciplinary and Interdisciplinary Science Education Research*, 1, 12. <https://doi.org/10.1186/s43031-019-0012-y>
- Clifton, J. D. W. (2020). Managing validity versus reliability trade-offs in scale-building decisions. *Psychological Methods*, 25(3), 259–270. <https://doi.org/10.1037/met0000236>
- Coe, R., Aloisi, C., Higgins, S. & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research*. Sutton Trust.
- Cook, D. A. & Hatala, R. (2016). Validation of educational assessments: A primer for simulation and beyond. *Advances in simulation*, 1, 31. <https://doi.org/10.1186/s41077-016-0033-y>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>

- Cronbach, L. J. & Meehl, P. J. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- de Groot, A. D. (1965). *Thought and choice in chess*. Mouton Publishers.
- de Jong, O. & van der Valk, A. E. (2007). Science teachers' PCK and teaching practice: Learning to scaffold students' open-inquiry learning. In R. Pintó & D. Couso (Hrsg.), *Contributions from science education research* (S. 107–118). Springer. https://doi.org/10.1007/978-1-4020-5032-9_8
- de Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C. E., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K. & Zacharia, Z. C. (2023). Let's talk evidence – The case for combining inquiry-based and direct instruction. *Educational Research Review*, 39, 100536. <https://doi.org/10.1016/j.edurev.2023.100536>
- Depaepe, F., Verschaffel, L. & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, 34, 12–25. <https://doi.org/10.1016/j.tate.2013.03.001>
- Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (6. Aufl.). Springer. <https://doi.org/10.1007/978-3-662-64762-2>
- Driver, R., Leach, J., Millar, R. & Scott, P. (2010). *Young people's images of science*. Open University Press.
- Dunbar, K. N. & Klahr, D. (2013). Scientific thinking and reasoning. In K. J. Holyoak (Hrsg.), *Oxford library of psychology. The Oxford handbook of thinking and reasoning* (S. 701–718). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199734689.013.0035>
- Edelsbrunner, P. A. & Dablander, F. (2018). The psychometric modeling of scientific reasoning: A review and recommendations for future avenues. *Educational Psychology Review*, 31(1), 1–34. <https://doi.org/10.1007/s10648-018-9455-5>
- Engelmann, K., Neuhaus, B. J. & Fischer, F. (2016). Fostering scientific reasoning in education – meta-analytic evidence from intervention studies. *Educational Research and Evaluation*, 22(5-6), 333–349. <https://doi.org/10.1080/13803611.2016.1240089>
- Ercikan, K. & Pellegrino, J. W. (2017). Validation of score meaning using examinee response processes for the next generation of assessments. In K. Ercikan & J. W. Pellegrino (Hrsg.), *Validation of score meaning for the next generation of assessments: The use of response processes* (S. 1–8). Routledge.
- Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Hrsg.), *The cambridge handbook of expertise and expert performance* (S. 683–703). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816796.038>
- Fendler, L. (2016). Ethical implications of validity-vs.-reliability trade-offs in educational research. *Ethics and Education*, 11(2), 214–229. <https://doi.org/10.1080/17449642.2016.1179837>
- Fischer, H. E., Boone, W. J. & Neumann, K. (2023). Quantitative research designs and approaches. In N. G. Lederman, D. L. Zeidler & J. Lederman (Hrsg.), *Handbook of research on science education* (S. 28–59). Routledge. <https://doi.org/10.4324/9780367855758-3>
- Fischler, H., Gebhard, U. & Rehm, M. (2018). Naturwissenschaftliche Bildung und Scientific Literacy. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 11–29). Springer. https://doi.org/10.1007/978-3-662-56320-5_2
- Forsler, A., Nilsson, P. & Walan, S. (2023). Capturing and developing teachers' pedagogical content knowledge in sustainable development using content representation and video-based reflection. *Research in Science Education*. <https://doi.org/10.1007/s11165-023-10149-y>
- Förtsch, C., Werner, S., Kotzebue, L. von & Neuhaus, B. J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666. <https://doi.org/10.1080/09500693.2016.1257170>

- Friedrichsen, P. J., Abell, S. K., Pareja, E. M., Brown, P. L., Lankford, D. M. & Volkmann, M. J. (2009). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching*, 46(4), 357–383. <https://doi.org/10.1002/tea.20283>
- Friedrichsen, P., van Driel, J. H. & Abell, S. K. (2011). Taking a closer look at science teaching orientations. *Science Education*, 95(2), 358–376. <https://doi.org/10.1002/sce.20428>
- GDSU. (2002). *Perspektivrahmen Sachunterricht*. <https://gdsu.de/content/perspektivrahmen-sachunterricht>
- Geis-Thöne, W. (2022). *Lehrkräftebedarf und -angebot: Bis 2035 steigende Engpässe zu erwarten*. <https://www.iwkoeln.de/studien/wido-geis-thoene-bis-2035-steigende-engpaesse-zu-erwarten.html>
- Gerasimova, D. (2024). Argument-based approach to validity: Developing a living document and incorporating preregistration. *Journal of Educational Measurement*, 61(2), 252–273. <https://doi.org/10.1111/jedm.12385>
- Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Hrsg.), *Examining pedagogical content knowledge* (S. 3–17). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47217-1_1
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen & J. Loughran (Hrsg.), *Re-examining pedagogical content knowledge in science education* (S. 28–42). Routledge.
- Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A. L., Wilson, C. D. & Stuhlsatz, M. A. M. (2019). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, 41(7), 944–963. <https://doi.org/10.1080/09500693.2016.1265158>
- Gluth, K. (2023). *Erfassung von fachdidaktischem Wissen angehender Biologielehrkräfte im Kompetenzbereich Erkenntnisgewinnung: Eine Analyse schwierigkeitserzeugender Aufgabenmerkmale bei Multiple-Choice-Tests*. [Masterarbeit, Christian-Albrechts-Universität zu Kiel].
- Godfrey-Smith, P. (2003). *Theory and reality: An introduction to the philosophy of science. Science and its conceptual foundations*. University of Chicago Press.
- Godfrey-Smith, P. (2007). The strategy of model-based science. *Biology & Philosophy*, 21(5), 725–740. <https://doi.org/10.1007/s10539-006-9054-6>
- Göhner, M. & Krell, M. (2022). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*, 52(2), 395–414. <https://doi.org/10.1007/s11165-020-09945-7>
- Gresch, H., Hasselhorn, M. & Bögeholz, S. (2013). Training in decision-making strategies: An approach to enhance students' competence to deal with socio-scientific issues. *International Journal of Science Education*, 35(15), 2587–2607. <https://doi.org/10.1080/09500693.2011.617789>
- Gropengießer, H. (2023). Erkunden und Erkennen. In H. Gropengießer & U. Harms (Hrsg.), *Fachdidaktik Biologie* (S. 246–251). Aulis.
- Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education. Professional development and practice series*. Teachers College Press.
- Großmann, L. & Krüger, D. (2022). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar*innen für die Qualität ihrer Unterrichtsentwürfe? *Zeitschrift für Didaktik der Naturwissenschaften*, 28(1), 271. <https://doi.org/10.1007/s40573-022-00141-w>
- Großmann, L. & Krüger, D. (2023). Identifying performance levels of enacted pedagogical content knowledge in trainee biology teachers' lesson plans. In G. S. Carvalho, A. S. Afonso & Z. Anastácio (Hrsg.), *Contributions from science education research: Bd. 13. Fostering scientific citizenship in an uncertain world: Selected papers from the ESERA 2021 conference* (S. 95–116). Springer. https://doi.org/10.1007/978-3-031-32225-9_7

- Großmann, L. & Krüger, D. (2024). Erkenntnisgewinnung (v)erklärt: Wie plant man hypothesengeleiteten Biologieunterricht? In B. Reinisch, D. Krüger & D. Mahler (Hrsg.), *Biologiedidaktische Nature of Science-Forschung: Zukunftsweisende Praxis* (S. 99–112). Springer Berlin/Heidelberg. https://doi.org/10.1007/978-3-662-68409-2_9
- Großschedl, J., Harms, U., Kleickmann, T. & Glowinski, I. (2015). Preservice biology teachers' professional knowledge: Structure and learning opportunities. *Journal of Science Teacher Education*, 26(3), 291–318. <https://doi.org/10.1007/s10972-015-9423-6>
- Großschedl, J., Welter, V. D. E. & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402–439. <https://doi.org/10.1002/tea.21482>
- Günther, S. L., Fleige, J., Upmeyer zu Belzen, A. & Krüger, D. (2019). Using the case method to foster preservice biology teachers' content knowledge and pedagogical content knowledge related to models and modeling. *Journal of Science Teacher Education*, 30(4), 321–343. <https://doi.org/10.1080/1046560X.2018.1560208>
- Gut-Glanzmann, C. & Mayer, J. (2018). Experimentelle Kompetenz. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 121–140). Springer. https://doi.org/10.1007/978-3-662-56320-5_8
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3. Aufl.). Routledge. <https://doi.org/10.4324/9780203825945>
- Hammann, M. (2007). Das Scientific Discovery as Dual Search-Modell. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 187–196). Springer. https://doi.org/10.1007/978-3-540-68166-3_17
- Hartmann, S., Upmeyer zu Belzen, A., Krüger, D. & Pant, H. A. (2015). Scientific reasoning in higher education: Constructing and evaluating the criterion-related validity of an assessment of preservice science teachers' competencies. *Zeitschrift für Psychologie*, 223(1), 47–53. <https://doi.org/10.1027/2151-2604/a000199>
- Hattie, J. (2023). *Visible learning: The sequel. A synthesis of over 2,100 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9781003380542>
- Hilfert-Rüppell, D., Meier, M., Horn, D. & Höner, K. (2021). Professional knowledge and self-efficacy expectations of pre-service teachers regarding scientific reasoning and diagnostics. *Education Sciences*, 11(10), 629. <https://doi.org/10.3390/educsci11100629>
- Hill, H. C., Ball, D. L. & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372–400. <https://doi.org/10.5951/jresmetheduc.39.4.0372>
- Hill, H. C., Rowan, B. & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406. <https://doi.org/10.3102/00028312042002371>
- Hogan, T., Rabinowitz, M. & Craven, J. A. (2003). Representation in teaching: Inferences from research of expert and novice teachers. *Educational Psychologist*, 38(4), 235–247. https://doi.org/10.1207/S15326985EP3804_3
- Hollingworth, L., Beard, J. J. & Proctor, T. P. (2007). *An investigation of item type in a standards-based assessment*. <https://doi.org/10.7275/6GGZ-8837>
- Hubley, A., & Zumbo, B. (2017). Response processes in the context of validity: Setting the stage. In B. Zumbo & A. Hubley (Hrsg.), *Understanding and investigating response processes in validation research: Bd. 69. Social Indicators Research Series* (S. 1–12). Springer. https://doi.org/10.1007/978-3-319-56129-5_1
- Hume, A., Cooper, R. & Borowski, A. (Hrsg.). (2019). *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*. Springer Singapore. <https://doi.org/10.1007/978-981-13-5898-2>
- Irmer, M., Traub, D., Böhm, M., Förtsch, C. & Neuhaus, B. J. (2023). Using video-based simulations to foster pPCK/ePCK—New thoughts on the refined consensus model of PCK. *Education Sciences*, 13(3), 261. <https://doi.org/10.3390/educsci13030261>

- Jin, H., Shin, H., Johnson, M. E., Kim, J. & Anderson, C. W. (2015). Developing learning progression-based teacher knowledge measures. *Journal of Research in Science Teaching*, 52(9), 1269–1295. <https://doi.org/10.1002/tea.21243>
- Justi, R. & van Driel, J. (2006). The use of the interconnected model of teacher professional growth for understanding the development of science teachers' knowledge on models and modelling. *Teaching and Teacher Education*, 22(4), 437–450. <https://doi.org/10.1016/j.tate.2005.11.011>
- Jüttner, M., Boone, 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. <https://doi.org/10.1007/s11092-013-9157-y>
- Kane, M. (2013a). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457. <https://doi.org/10.1080/02796015.2013.12087465>
- Kane, M. (2013b). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kane, M. & Mislevy, R. (2017). Validating score interpretations based on response processes. In K. Ercikan & J. W. Pellegrino (Hrsg.), *Validation of score meaning for the next generation of assessments: The use of response processes* (S. 11–24). Routledge.
- Keller, M. M., Neumann, K. & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, 54(5), 586–614. <https://doi.org/10.1002/tea.21378>
- Kelly, M. P. & Staver, J. R. (2005). A case study of one school system's adoption and implementation of an elementary science program. *Journal of Research in Science Teaching*, 42(1), 25–52. <https://doi.org/10.1002/tea.20043>
- Khan, S. & Krell, M. (2019). Scientific reasoning competencies: A case of preservice teacher education. *Canadian Journal of Science, Mathematics and Technology Education*, 19(4), 446–464. <https://doi.org/10.1007/s42330-019-00063-9>
- Khishfe, R. & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching*, 39(7), 551–578. <https://doi.org/10.1002/tea.10036>
- Kind, P. & Osborne, J. (2017). Styles of scientific reasoning: A cultural rationale for science education? *Science Education*, 101(1), 8–31. <https://doi.org/10.1002/sce.21251>
- Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential for progress. *Studies in Science Education*, 45(2), 169–204. <https://doi.org/10.1080/03057260903142285>
- Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Kirschner, S., Sczudlek, M., Tepner, O., Borowski, A., Fischer, H. E., Lenske, G., Leutner, D., Neuhaus, B. J., Sumfleth, E., Thillmann, H. & Wirth, J. (2017). Professionswissen in den Naturwissenschaften (ProwiN). In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals: Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 113–130). Springer. https://doi.org/10.1007/978-3-658-07274-2_7
- Klahr, D. (Hrsg.). (2000). *Exploring science: The cognition and development of discovery processes*. MIT Press. <https://doi.org/10.7551/mitpress/2939.001.0001?locatt=mode:legacy>
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1–48. https://doi.org/10.1207/s15516709cog1201_1
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S. & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge. *Journal of Teacher Education*, 64(1), 90–106. <https://doi.org/10.1177/0022487112460398>

- Klemm, K. (2020). *Lehrkräftemangel in den MINT-Fächern: Kein Ende in Sicht: Zur Bedarfs- und Angebotsentwicklung in den allgemeinbildenden Schulen der Sekundarstufe I und II am Beispiel Nordrhein-Westfalen*. <https://www.telekom-stiftung.de/sites/default/files/mint-lehrkraeftebedarf-2020-ergebnisbericht.pdf>
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876–903. <https://doi.org/10.25656/01:4493>
- KMK (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss: Beschluss vom 16.12.2004*. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Biologie.pdf
- KMK (2020a). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife. (Beschluss der Kultusministerkonferenz vom 18.06.2020)*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf
- KMK (2020b). *Bildungsstandards im Fach Chemie für die Allgemeine Hochschulreife. (Beschluss der Kultusministerkonferenz vom 18.06.2020)*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Chemie.pdf
- KMK (2020c). *Bildungsstandards im Fach Physik für die Allgemeine Hochschulreife. (Beschluss der Kultusministerkonferenz vom 18.06.2020)*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Physik.pdf
- KMK (2022). *Lehrkräfteeinstellungsbedarf und -angebot in der Bundesrepublik Deutschland 2021 - 2035: Zusammengefasste Modellrechnungen der Länder*. https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/Dok_233_Bericht_LEB_LEA_2021.pdf
- KMK (2024a). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. (Beschluss der Kultusministerkonferenz vom 16.10.2008 i. d. F. vom 16.05.2019)*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf
- KMK (2024b). *Weiterentwickelte Bildungsstandards in den Naturwissenschaften für das Fach Biologie (MSA). (Beschluss der Kultusministerkonferenz vom 16.12.2004 i.d.F. vom 13.06.2024)*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2024/2024_06_13-WeBiS_Biologie_MSA.pdf
- KMK (2024c). *Weiterentwickelte Bildungsstandards in den Naturwissenschaften für das Fach Chemie (MSA). (Beschluss der Kultusministerkonferenz vom 16.12.2004 i.d.F. vom 13.06.2024)*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2024/2024_06_13-WeBiS_Chemie_MSA.pdf
- KMK (2024d). *Weiterentwickelte Bildungsstandards in den Naturwissenschaften für das Fach Physik (MSA). (Beschluss der Kultusministerkonferenz vom 16.12.2004 i.d.F. vom 13.06.2024)*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2024/2024_06_13-WeBiS_Physik_MSA.pdf
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42(2), 262–271. <https://doi.org/10.1016/j.shpsa.2010.11.034>
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M. & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716–725. <https://doi.org/10.1037/0022-0663.100.3.716>

- Krell, M. (2018). Schwierigkeitserzeugende Aufgabenmerkmale bei Multiple-Choice-Aufgaben zur Experimentierkompetenz im Biologieunterricht: Eine Replikationsstudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1), 1–15. <https://doi.org/10.1007/s40573-017-0069-0>
- Krell, M., Czeskleba, A. & Krüger, D. (2012). Validierung von Forced Choice-Aufgaben durch lautes Denken. *Erkenntnisweg Biologiedidaktik*, 11, 53 - 70.
- Krell, M., Dawborn-Gundlach, M. & van Driel, J. (2020). Scientific Reasoning Competencies in Science Teaching. *Teaching Science*, 66(2), 32–42.
- Krell, M., Khan, S., Vergara, C., Cofré, H., Mathesius, S. & Krüger, D. (2022). Pre-service science teachers' scientific reasoning competencies: Analysing the impact of contributing factors. *Research in Science Education*, 53(1), 25. <https://doi.org/10.1007/s11165-022-10045-x>
- Krell, M. & Krüger, D. (2022). Erkenntnisgewinnung im Biologieunterricht: Welches Methodenwissen soll vermittelt werden? *MNU Journal*, 5, 376–382.
- Krell, M., Upmeyer zu Belzen, A. & Krüger, D. (2016). Modellkompetenz im Biologieunterricht. In A. Sandmann & P. Schmiemann (Hrsg.), *Biologie lernen und lehren: Band 1. Biologiedidaktische Forschung: Schwerpunkte und Forschungsstände* (S. 83–102). Logos Verlag Berlin.
- Krell, M., Vorholzer, A. & Nehring, A. (2022). Scientific reasoning in science education: From global measures to fine-grained descriptions of students' competencies. *Education Sciences*, 12(2), 97. <https://doi.org/10.3390/educsci12020097>
- Krepf, M., Plöger, W., Scholl, D., & Seifert, A. (2018). Pedagogical content knowledge of experts and novices—what knowledge do they activate when analyzing science lessons? *Journal of Research in Science Teaching*, 55(1), 44–67. <https://doi.org/10.1002/tea.21410>
- Krüger, D., Hartmann, S., Nordmeier, V. & Upmeyer zu Belzen, A. (2020). Measuring scientific reasoning competencies. In O. Zlatkin-Troitschanskaia, H. A. Pant & C. Lautenbach (Hrsg.), *Student learning in german higher education: Innovative modelling and measurement approaches and research results* (S. 261–280). Springer. https://doi.org/10.1007/978-3-658-27886-1_13
- Krüger, D. & Upmeyer zu Belzen, A. (2021). Kompetenzmodell der Modellierkompetenz – Die Rolle abduktiven Schließens beim Modellieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 27(1), 127–137. <https://doi.org/10.1007/s40573-021-00129-y>
- Kulgemeyer, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Reinhold, P., Riese, J., Schecker, H., Schröder, J. & Vogelsang, C. (2020). Professional knowledge affects action-related skills: The development of preservice physics teachers' explaining skills during a field experience. *Journal of Research in Science Teaching*, 57(10), 1554–1582. <https://doi.org/10.1002/tea.21632>
- Kulgemeyer, C., Kempin, M., Weißbach, A., Borowski, A., Buschhüter, D., Enkrott, P., Reinhold, P., Riese, J., Schecker, H., Schröder, J. & Vogelsang, C. (2021). Exploring the impact of pre-service science teachers' reflection skills on the development of professional knowledge during a field experience. *International Journal of Science Education*, 43(18), 3035–3057. <https://doi.org/10.1080/09500693.2021.2006820>
- Kulgemeyer, C. & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, 55(10), 1393–1418. <https://doi.org/10.1002/tea.21457>
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>
- Lane, S. (2017). Commentary II: The value of response process studies to evaluate score meaning. In K. Ercikan & J. W. Pellegrino (Hrsg.), *Validation of score meaning for the next generation of assessments: The use of response processes* (S. 138–144). Routledge.
- Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education*, 2(3), 307–338. <https://doi.org/10.1007/s10763-004-3224-2>

- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of Guidance. *Review of Educational Research*, 86(3), 681–718.
<https://doi.org/10.3102/0034654315627366>
- Lederman, J., Bartels, S., Jimenez, J., Lederman, N. G., Acosta, K., Adbo, K., Akerson, V. L., Andrade, M. a. B. S. de, Avraamidou, L., Barber, K.-A., Blanquet, E., Boujaoude, S., Cardoso, R., Cesljarev, C. D., Chaipidech, P., Connolly, C. P., Crowther, D. T., Das, P. M., Dogan, Ö. K., . . . Zhu, Q. (2023). Completing the progression establishing an international baseline of primary, middle and secondary students' views of scientific inquiry. *International Journal of Science Education*, 46(7), 715–731. <https://doi.org/10.1080/09500693.2023.2256458>
- Lederman, J., Lederman, N. G., Bartels, S., Jimenez, J., Acosta, K., Akubo, M., Aly, S., Andrade, M. A. B. S. de, Atanasova, M., Blanquet, E., Blonder, R., Brown, P., Cardoso, R., Castillo-Urueta, P., Chaipidech, P., Concannon, J., Dogan, O. K., El-Deghaidy, H., Elzorkani, A., . . . Wishart, J. (2021). International collaborative follow-up investigation of graduating high school students' understandings of the nature of scientific inquiry: is progress Being made? *International Journal of Science Education*, 43(7), 991–1016.
<https://doi.org/10.1080/09500693.2021.1894500>
- Lederman, J., Lederman, N. G., Bartels, S., Jimenez, J., Akubo, M., Aly, S., Bao, C., Blanquet, E., Blonder, R., Bologna Soares de Andrade, M., Buntting, C., Cakir, M., EL-Deghaidy, H., ELZorkani, A., Gaigher, E., Guo, S., Hakanen, A., Hamed Al-Lal, S., Han-Tosunoglu, C., . . . Zhou, Q. (2019). An international collaborative investigation of beginning seventh grade students' understandings of scientific inquiry: Establishing a baseline. *Journal of Research in Science Teaching*, 56(4), 486–515. <https://doi.org/10.1002/tea.21512>
- Lederman, J., Akerson, V., Bartels, S. & Schwartz, R. (2024). Attention science educators, we have a problem: lack of global functional scientific literacy. *International Journal of Science Education*. <https://doi.org/10.1080/09500693.2024.2361934>
- Lee, E., Brown, M. N., Luft, J. A. & Roehrig, G. H. (2007). Assessing beginning secondary science teachers' PCK: Pilot year results. *School Science and Mathematics*, 107(2), 52–60.
<https://doi.org/10.1111/j.1949-8594.2007.tb17768.x>
- Lee, S. W.-Y. (2024). Investigating the effects of explicit instructional approaches on students' understanding of scientific models and modeling. *Research in Science & Technological Education*, 42(3), 578–594. <https://doi.org/10.1080/02635143.2022.2153244>
- Lin, S. & Wang, J. (2024). Development and application of an instrument for assessing upper-secondary school biology teachers' pedagogical content knowledge of scientific thinking. *Journal of Baltic Science Education*, 23(3), 495–517. <https://doi.org/10.33225/jbse/24.23.495>
- Loughran, J., Mulhall, P. & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching*, 41(4), 370–391. <https://doi.org/10.1002/tea.20007>
- Lübeck, M. (2020). „Basiskonzepte“ der Erkenntnisgewinnung im Biologieunterricht: Ein Nachschlagewerk mit Aufgabenbeispielen. *Beiträge zur Schulentwicklung. Praxis*. Waxmann.
- Luft, J. A. & Zhang, C. (2014). The pedagogical content knowledge and beliefs of newly hired secondary science teachers: the first three years. *Educación Química*, 25(3), 325–331.
[https://doi.org/10.1016/S0187-893X\(14\)70548-8](https://doi.org/10.1016/S0187-893X(14)70548-8)
- Magnusson, S., Krajcik, J. & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Hrsg.), *Examining pedagogical content knowledge* (S. 95–132). Kluwer Academic Publishers.
https://doi.org/10.1007/0-306-47217-1_4
- Mahler, D., Bock, D., Schaubert, S. & Harms, U. (2024). Using longitudinal models to describe preservice science teachers' development of content knowledge and pedagogical content knowledge. *Teaching and Teacher Education*, 144, 104583.
<https://doi.org/10.1016/j.tate.2024.104583>

- Mahler, D., Großschedl, J. & Harms, U. (2017). Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. *International Journal of Science Education*, 39(2), 213–237. <https://doi.org/10.1080/09500693.2016.1276641>
- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34(4), 207–218. https://doi.org/10.1207/s15326985ep3404_2
- Mathesius, S., Hartmann, S., Upmeier zu Belzen, A., & Krüger, D. (2016). Scientific reasoning as an aspect of pre-service biology teacher education: Assessing competencies using a paper-pencil-test. In T. Tal & A. Yarden (Hrsg.), *The future of biology education research* (S. 93–110).
- Mathesius, S., Upmeier zu Belzen, A. & Krüger, D. (2018). Eyetracking als Methode zur Untersuchung von Multiple-Choice-Aufgaben zum wissenschaftlichen Denken. In M. Hammann & M. Lindner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik: Band 8* (S. 225–244). StudienVerlag.
- Matlen, B. J. & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science*, 41(3), 621–634. <https://doi.org/10.1007/s11251-012-9248-z>
- Mavhunga, E. & Rollnick, M. (2013). Improving PCK of chemical equilibrium in pre-service teachers. *African Journal of Research in Mathematics, Science and Technology Education*, 17(1-2), 113–125. <https://doi.org/10.1080/10288457.2013.828406>
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 177–186). Springer. https://doi.org/10.1007/978-3-540-68166-3_16
- McCoach, D. B., Gable, R. K. & Madura, J. P. (2013a). Additional evidence based on the internal structure of the instrument. In D. B. McCoach, R. K. Gable & J. P. Madura (Hrsg.), *Instrument development in the affective domain: School and corporate applications* (3. Aufl., S. 163–208). Springer. https://doi.org/10.1007/978-1-4614-7135-6_5
- McCoach, D. B., Gable, R. K. & Madura, J. P. (2013b). Evidence based on relations to other variables: Bolstering the empirical validity arguments for constructs. In D. B. McCoach, R. K. Gable & J. P. Madura (Hrsg.), *Instrument development in the affective domain: School and corporate applications* (3. Aufl., S. 209–248). Springer. https://doi.org/10.1007/978-1-4614-7135-6_6
- McCoach, D. B., Gable, R. K. & Madura, J. P. (2013c). Evidence based on test content. In D. B. McCoach, R. K. Gable & J. P. Madura (Hrsg.), *Instrument development in the affective domain: School and corporate applications* (3. Aufl., S. 91–107). Springer. https://doi.org/10.1007/978-1-4614-7135-6_3
- McCoach, D. B., Gable, R. K. & Madura, J. P. (2013d). Evidence based on the internal structure of the instrument: Factor analysis. In D. B. McCoach, R. K. Gable & J. P. Madura (Hrsg.), *Instrument development in the affective domain: School and corporate applications* (3. Aufl., S. 109–161). Springer. https://doi.org/10.1007/978-1-4614-7135-6_4
- Meinshausen, F. (2024). *Erfassung des fachdidaktischen Wissens von Biologielehrkräften im Kompetenzbereich Erkenntnisgewinnung*. [Masterarbeit, Christian-Albrechts-Universität zu Kiel].
- Messick, S. (1989). Validity. In R. L. Linn (Hrsg.), *Educational measurement* (3. Aufl., S. 13–103). American Council on Education.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Messick, S. (1996). Validity of performance assessments. In G. W. Phillips (Hrsg.), *Technical issues in large-scale performance assessment* (S. 1–18). National Center for Education Statistics.
- Mineduc (2012). *Estándares orientadores para carreras de pedagogía en Educación Media* [Richtlinien für die pädagogische Laufbahn im Sekundarbereich]. https://www.cpeip.cl/wp-content/uploads/2018/09/Est%C3%A1ndares_Media.pdf

- Mineduc (2015). *Bases Curriculares: 7° básico a 2° medio* [Lehrplan: 7. Grundstufe bis 2. Mittelstufe]. https://www.curriculumnacional.cl/614/articles-34949_Bases.pdf
- Mineduc (2019). *Bases Curriculares: 3° y 4° medio* [Lehrplan: 3. und 4. Klassenstufe]. https://www.curriculumnacional.cl/614/articles-91414_bases.pdf
- Ministry of Education (2008a). *The Ontario Curriculum: Grades 9 and 10. Science*. https://www.edu.gov.on.ca/eng/curriculum/secondary/science910_2008.pdf
- Ministry of Education (2008b). *The Ontario Curriculum: Grades 11 and 12. Science*. https://www.edu.gov.on.ca/eng/curriculum/secondary/2009science11_12.pdf
- Minner, D. D., Levy, A. J. & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496. <https://doi.org/10.1002/tea.20347>
- Morris, B. J., Croker, S., Masnick, A. M. & Zimmerman, C. (2012). The emergence of scientific reasoning. In H. Kloos, B. J. Morris & J. L. Amaral (Hrsg.), *Current topics in children's learning and cognition* (S. 61–82). IntechOpen.
- Nehm, R. H. & Ridgway, J. (2011). What do experts and novices “see” in evolutionary problems? *Evolution: Education and Outreach*, 4(4), 666–679. <https://doi.org/10.1007/s12052-011-0369-7>
- Neumann, K., Kind, V. & Harms, U. (2019). Probing the amalgam: The relationship between science teachers' content, pedagogical and pedagogical content knowledge. *International Journal of Science Education*, 41(7), 847–861. <https://doi.org/10.1080/09500693.2018.1497217>
- Newton, P. E. (2018). What is response process validation evidence and how important is it? An essay reviewing Ercikan and Pellegrino (2017) and Zumbo and Hubley (2017). *Assessment in Education: Principles, Policy & Practice*, 26(2), 245–253. <https://doi.org/10.1080/0969594X.2018.1447909>
- NGSS Lead States (2013). *Next Generation Science Standards*. National Academies Press. <https://nap.nationalacademies.org/catalog/18290/next-generation-science-standards-for-states-by-states> <https://doi.org/10.17226/18290>
- Nielsen, S. S. & Nielsen, J. A. (2021). A competence-oriented approach to models and modelling in lower secondary science education: Practices and rationales among danish teachers. *Research in Science Education*, 51(S2), 565–593. <https://doi.org/10.1007/s11165-019-09900-1>
- OECD (2019). *PISA 2018: Assessment and analytic framework*. PISA. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- OECD (2022). *PISA 2024: Strategic vision and direction for science*.
- Opfer, J. E., Nehm, R. H. & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. *Journal of Research in Science Teaching*, 49(6), 744–777. <https://doi.org/10.1002/tea.21028>
- Opitz, A., Heene, M. & Fischer, F. (2017). Measuring scientific reasoning – a review of test instruments. *Educational Research and Evaluation*, 23(3-4), 78–101. <https://doi.org/10.1080/13803611.2017.1338586>
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265–279. <https://doi.org/10.1016/j.tsc.2013.07.006>
- Osborne, J. (2014a). Scientific practices and inquiry in the science classroom. In N. G. Lederman & S. K. Abell (Hrsg.), *Handbook of research on science education: Volume II* (S. 579–599). Routledge.
- Osborne, J. (2014b). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, 25(2), 177–196. <https://doi.org/10.1007/s10972-014-9384-1>
- Osborne, J. (2023). Science, scientific literacy, and science education. In N. G. Lederman, D. L. Zeidler & J. Lederman (Hrsg.), *Handbook of research on science education* (S. 785–816). Routledge. <https://doi.org/10.4324/9780367855758-30>

- Ottogalli, M. E. & Bermudez, G. M. (2024). A PCK-mapping approach to show the integration among components of the pedagogical content knowledge of elementary education teacher educators about biodiversity. *Teaching and Teacher Education*, 151, 104746. <https://doi.org/10.1016/j.tate.2024.104746>
- Padilla, J.-L. & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136–144. <https://doi.org/10.7334/psicothema2013.259>
- Park, S. (2019). Reconciliation between the refined consensus model of PCK and extant PCK models for advancing PCK research in science. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 119–130). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_4
- 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. <https://doi.org/10.1002/tea.21022>
- Park, S. & Oliver, J. S. (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. <https://doi.org/10.1007/s11165-007-9049-6>
- Park, S. & Suh, J. K. (2019). The PCK map approach to capturing the complexity of enacted PCK (ePCK) and pedagogical reasoning in science teaching. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (S. 185–197). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_8
- Petermann, V. & Vorholzer, A. (2023). Teachers' use of explicit instruction when planning lessons to foster students' scientific inquiry competencies. In G. S. Carvalho, A. S. Afonso & Z. Anastácio (Hrsg.), *Contributions from science education research: Bd. 13. Fostering scientific citizenship in an uncertain world: Selected papers from the ESERA 2021 Conference* (S. 219–234). Springer. https://doi.org/10.1007/978-3-031-32225-9_14
- Petermann, V., Vorholzer, A. & Aufschneider, C. von (2024). Science teachers' beliefs about teaching and learning related to content and procedural goals. *Journal of Research in Science Teaching*, 22003. <https://doi.org/10.1002/tea.22003>
- Pohlmann, M. (2019). *Förderung ethischer Bewertungskompetenz: Der Einfluss ausgewählter Lerngelegenheiten auf die inhaltliche Ausdifferenzierung und die Kohärenz der Komponenten des fachdidaktischen Wissens von Biologielehrkräften*. <https://d-nb.info/1182534066/34>
- Popper, K. (1935). *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft*. Springer Vienna. <https://doi.org/10.1007/978-3-7091-4177-9>
- Priemer, B. (2011). Was ist das Offene beim offenen Experimentieren? *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 315–337.
- Riese, J. & Reinhold, P. (2012). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen. *Zeitschrift für Erziehungswissenschaft*, 15(1), 111–143. <https://doi.org/10.1007/s11618-012-0259-y>
- Rios, J. & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema* 26(1), 108–116. <https://doi.org/10.7334/psicothema2013.260>
- Rönnebeck, S., Bernholt, S. & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>
- Rosenkränzer, F., Hörsch, C., Schuler, S. & Riess, W. (2017). Student teachers' pedagogical content knowledge for teaching systems thinking: effects of different interventions. *International Journal of Science Education*, 39(14), 1932–1951. <https://doi.org/10.1080/09500693.2017.1362603>
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N. & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020–1049. <https://doi.org/10.3102/0002831213477680>

- Sannert, R., Petermann, V., Lieberei, T., Welter, V. D. E. & Krell, M. (2024). Erkenntnisgewinnungskompetenzen im naturwissenschaftlichen Unterricht fördern. In N. Graulich, J. Arnold, S. Sorge & M. Kubsch (Hrsg.), *Lehrkräftebildung von morgen* (S. 281–290). Waxmann. <https://doi.org/10.31244/9783830997962.30>
- Sannert, R., Voitle, F., Haastert, K. & Krell, M. (2024). Ein Experiment – Zwei Lernziele: Schüler/innen gezielt beim Aufbau von Sachkompetenz oder Erkenntnisgewinnungskompetenz unterstützen. *MNU Journal*, 77(6), 454–459.
- Sannert, R., van Driel, J. & Krell, M. (2024). Using content representations to measure components of teachers' PCK about scientific reasoning. [Manuskript eingereicht]
- Schiering, D., Sorge, S., Keller, M. M. & Neumann, K. (2022). A proficiency model for pre-service physics teachers' pedagogical content knowledge (PCK)—what constitutes high-level PCK? *Journal of Research in Science Teaching*, 52(11), 271. <https://doi.org/10.1002/tea.21793>
- Schiering, D., Sorge, S. & Neumann, K. (2021). Hilft viel viel? Der Einfluss von Studienstrukturen auf das Professionswissen angehender Physiklehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 24(3), 545–570. <https://doi.org/10.1007/s11618-021-01003-w>
- Schlichter, N. (2012). *Lehrerüberzeugungen zum Lehren und Lernen*. [Dissertation, Georg-August-Universität Göttingen].
- Schmelzing, S., van Driel, J., Jüttner, M., Brandenbusch, S., Sandmann, A. & Neuhaus, B. J. (2013). Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the "cardiovascular system". *International Journal of Science and Mathematics Education*, 11(6), 1369–1390. <https://doi.org/10.1007/s10763-012-9384-6>
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>
- Schunn, C. D. & Klahr, D. (2000). Multiple-space search in a more complex discovery microworld. In D. Klahr (Hrsg.), *Exploring science: The cognition and development of discovery processes* (S. 161–199). MIT Press.
- Schwartz, R., Lederman, N. G. & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education*, 88(4), 610–645. <https://doi.org/10.1002/sce.10128>
- Schwartz, R., Lederman, J. & Enderle, P. J. (2023). Scientific inquiry literacy. In N. G. Lederman, D. L. Zeidler & J. Lederman (Hrsg.), *Handbook of research on science education* (S. 749–782). Routledge. <https://doi.org/10.4324/9780367855758-28>
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Mifflin and Company.
- She, J. (2023). *Clarifying the relationship between science teachers pedagogical content knowledge and student learning outcomes*. [Dissertation, The University of Hong Kong].
- She, J., & Chan, K. K. H. (2023). Situated and dynamic versus declarative and static forms of pedagogical content knowledge: An evaluation of the differences in test reactions and performance. *Journal of Research in Science Teaching*, 60(3), 568–607. <https://doi.org/10.1002/tea.21810>
- She, J., Chan, K. K. H., Wang, J., Hu, X. & Liu, E. (2024). Effect of science teachers' pedagogical content knowledge on student achievement: Evidence from both text- and video-based pedagogical content knowledge tests. *American Educational Research Journal*. <https://doi.org/10.3102/00028312241278627>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Shulman, L. S. (2015). PCK: Its genesis and exodus. In A. Berry, P. Friedrichsen & J. Loughran (Hrsg.), *Re-examining pedagogical content knowledge in science education* (S. 3–13). Routledge.
- Sireci, S. (1998). The construct of content validity. *Social Indicators Research*, 45(1/3), 83–117. <https://doi.org/10.1023/A:1006985528729>

- Sireci, S. & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107. <https://doi.org/10.7334/psicothema2013.256>
- Sireci, S. & Parker, P. (2006). Validity on trial: Psychometric and legal conceptualizations of validity. *Educational Measurement: Issues and Practice*, 25(3), 27–34. <https://doi.org/10.1111/j.1745-3992.2006.00065.x>
- Slomp, D. H. & Fuite, J. (2004). Following phaedrus: Alternate choices in surmounting the reliability/validity dilemma. *Assessing Writing*, 9(3), 190–207. <https://doi.org/10.1016/j.asw.2004.10.001>
- Smith, P. S. & Banilower, E. R. (2015). Assessing PCK: A new application of the uncertainty principle. In A. Berry, P. Friedrichsen & J. Loughran (Hrsg.), *Re-examining pedagogical content knowledge in science education* (S. 88–103). Routledge.
- Sorge, S., Keller, M. M., Neumann, K. & Möller, J. (2019). Investigating the relationship between pre-service physics teachers' professional knowledge, self-concept, and interest. *Journal of Research in Science Teaching*, 56(7), 937–955. <https://doi.org/10.1002/tea.21534>
- Stender, A. & Brückmann, M. (2020). Processes of knowledge integration between realms of pedagogical content knowledge and how to capture them. In T. Lehmann (Hrsg.), *International perspectives on knowledge integration: Theory, research, and good practice in pre-service teacher and higher education* (S. 111–132). BRILL. https://doi.org/10.1163/9789004429499_006
- Suh, J. K. & Park, S. (2017). Exploring the relationship between pedagogical content knowledge (PCK) and sustainability of an innovative science teaching approach. *Teaching and Teacher Education*, 64, 246–259. <https://doi.org/10.1016/j.tate.2017.01.021>
- Sweller, J., Zhang, L., Ashman, G., Cobern, W. & Kirschner, P. A. (2024). Response to De Jong et al.'s (2023) paper "Let's talk evidence – The case for combining inquiry-based and direct instruction". *Educational Research Review*, 42, 100584. <https://doi.org/10.1016/j.edurev.2023.100584>
- SWK (2024). *Empfehlungen zum Umgang mit dem akuten Lehrkräftemangel: Stellungnahme der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz*. <http://dx.doi.org/10.25656/01:26372>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education*, 4(2), 99–110. [https://doi.org/10.1016/0742-051X\(88\)90011-X](https://doi.org/10.1016/0742-051X(88)90011-X)
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B. J., Sandmann, A., Sumfleth, E., Thillmann, H. & Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 7–28.
- Upmeyer zu Belzen, A., Engelschalt, P. & Krüger, D. (2021). Modeling as scientific reasoning—The role of abductive reasoning for modeling competence. *Education Sciences*, 11(9), 495. <https://doi.org/10.3390/educsci11090495>
- Upmeyer zu Belzen, A. & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57.
- Upmeyer zu Belzen, A. & Krüger, D. (2019). Ein Fall für Erkenntnisgewinnung: Biologische Beiträge zu einem Verständnis naturwissenschaftlichen Modellierens. *Unterricht Chemie* (171), 38–41.
- Upmeyer zu Belzen, A., van Driel, J., & Krüger, D. (2019). Introducing a framework for modeling competence. In A. Upmeyer zu Belzen, D. Krüger, & J. van Driel (Hrsg.), *Towards a competence-based view on models and modeling in science education: Bd. 12. Models and modeling in science education* (S. 3–19). Springer. https://doi.org/10.1007/978-3-030-30255-9_1
- van Driel, J. & Berry, A. (2010). Pedagogical content knowledge. In P. L. Peterson (Hrsg.), *International encyclopedia of education* (3. Aufl., S. 656–661). Academic Press. <https://doi.org/10.1016/B978-0-08-044894-7.00642-4>

- van Driel, J. & Berry, A. (2019). Pedagogical content knowledge in preservice teacher education. In M. A. Peters (Hrsg.), *Encyclopedia of teacher education* (S. 1–6). Springer Singapore.
https://doi.org/10.1007/978-981-13-1179-6_176-1
- van Driel, J., Hume, A. & Berry, A. (2023). Research on science teacher knowledge and its development. In N. G. Lederman, D. L. Zeidler & J. Lederman (Hrsg.), *Handbook of research on science education* (S. 1123–1161). Routledge. <https://doi.org/10.4324/9780367855758-41>
- van Joolingen, W. R. & de Jong, T. (1997). An extended dual search space model of scientific discovery learning. *Instructional Science*, 25(5), 307–346.
<https://doi.org/10.1023/A:1002993406499>
- van Uum, M. S. J., Peeters, M. & Verhoeff, R. P. (2021). Professionalising primary school teachers in guiding inquiry-based learning. *Research in Science Education*, 51(S1), 81–108.
<https://doi.org/10.1007/s11165-019-9818-z>
- Vo, T., Forbes, C. T., Zangori, L. & Schwarz, C. V. (2015). Fostering third-grade students' use of scientific models with the water cycle: Elementary teachers' conceptions and practices. *International Journal of Science Education*, 37(15), 2411–2432.
<https://doi.org/10.1080/09500693.2015.1080880>
- Vorholzer, A. & Aufschnaiter, C. von (2019). Guidance in inquiry-based instruction – an attempt to disentangle a manifold construct. *International Journal of Science Education*, 41(11), 1562–1577. <https://doi.org/10.1080/09500693.2019.1616124>
- Vorholzer, A., Aufschnaiter, C. von & Boone, W. J. (2020). Fostering upper secondary students' ability to engage in practices of scientific investigation: A comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education*, 50(1), 333–359.
<https://doi.org/10.1007/s11165-018-9691-1>
- Vorholzer, A., Petermann, V., Weber, J., Upmeyer zu Belzen, A. & Tiemann, R. (2022). Explicit instruction on procedural and epistemic knowledge – is it happening? A video-based exploration of classroom practice. *Research in Science & Technological Education*, 42(3), 513–532. <https://doi.org/10.1080/02635143.2022.2153245>
- Waßmann, S. (2014). *Innovationsfähigkeit im demografischen Wandel: Entwicklung und Validierung des Assessment-Tools NovaDemo zur Erfassung der Innovationsfähigkeit von Einzelpersonen und Arbeitsgruppen*. [Dissertation, Karlsruher Institut für Technologie]. KIT Scientific Publishing. <https://publikationen.bibliothek.kit.edu/1000043686>
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.) *Leistungsmessungen in Schulen* (S. 17–31). Beltz.
- Wellnitz, N. & Mayer, J. (2013). Erkenntnismethoden in der Biologie - Entwicklung und Evaluation eines Kompetenzmodells. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 315–345.
- Wellnitz, N. & Mayer, J. (2016). Methoden der Erkenntnisgewinnung im Biologieunterricht. In A. Sandmann & P. Schmiemann (Hrsg.), *Biologie lernen und lehren: Bd. 1. Biologiedidaktische Forschung: Schwerpunkte und Forschungsstände* (S. 61–82). Logos Verlag Berlin.
- Welter, V. D. E., Herzog, S., Harms, U., Steffensky, M. & Großschedl, J. (2022). School subjects' synergy and teacher knowledge: Do biology and chemistry teachers benefit equally from their second subject? *Journal of Research in Science Teaching*, 59(2), 285–326.
<https://doi.org/10.1002/tea.21728>
- Wise, L. L. (2017). Commentary I: Validation of score meaning in the next generation of assessments. In K. Ercikan & J. W. Pellegrino (Hrsg.), *Validation of score meaning for the next generation of assessments: The use of response processes* (S. 52–59). Routledge.
- Wolff, C. E., van den Bogert, N., Jarodzka, H. & Boshuizen, H. P. A. (2015). Keeping an eye on learning: Differences between expert and novice teachers' representations of classroom management events. *Journal of Teacher Education*, 66(1), 68–85.
<https://doi.org/10.1177/0022487114549810>

- Wu, D., Liao, T., Yang, W. & Li, H. (2021). Exploring the relationships between scientific epistemic beliefs, science teaching beliefs and science-specific PCK among pre-service kindergarten teachers in China. *Early Education and Development*, 32(1), 82–97.
<https://doi.org/10.1080/10409289.2020.1771971>
- Yang, Y., Liu, X. & Gardella, J. A. (2020). Effects of a professional development program on science teacher knowledge and practice, and student understanding of interdisciplinary science concepts. *Journal of Research in Science Teaching*, 57(7), 1028–1057.
<https://doi.org/10.1002/tea.21620>
- Zeidler, D. L., Sadler, T. D., Simmons, M. L. & Howes, E. V. (2005). Beyond STS: A research-based framework for socioscientific issues education. *Science Education*, 89(3), 357–377.
<https://doi.org/10.1002/sce.20048>
- Zhang, L., Kirschner, P. A., Cobern, W. W. & Sweller, J. (2022). There is an evidence crisis in science educational policy. *Educational Psychology Review*, 34(2), 1157–1176.
<https://doi.org/10.1007/s10648-021-09646-1>

12 Anhang

12.1 Multiple-Choice-Instrument zur Erfassung fachdidaktischen Wissens im Kompetenzbereich Erkenntnisgewinnung (PCK_{SR}-bio)

Im Folgenden befindet sich eine Übersicht über die Systematik des PCK_{SR}-bio (Tabelle 12.1), eine Übersicht über die Items sowie ein dazugehöriger Lösungsbogen (Tabelle 12.2).

Verwendet in Studie 1, 2, 3 und 5.

Tabelle 12.1: Übersicht über die Systematik des PCK_{SR}-bio

	Experimentieren	Hypothesen formulieren	Modellieren
KISR	Item #1	Item #2	Item #3
KSU	Item #4	Item #5	Item #6
KISR→KSU	Item #7	Item #8	Item #9
KSU→KISR	Item #10	Item #11	Item #12

Item#1: KISR x Experimentieren

Eine Biologielehrkraft führt im Themenfeld „Lebensräume und ihre Bewohner“ mit ihren Schüler:innen ein Experiment zur Habitatpräferenz von Asseln durch. Hierbei sollen die Schüler:innen untersuchen, inwiefern Asseln helle oder dunkle beziehungsweise feuchte oder trockene Standorte bevorzugen. Das Ziel dieser Stunde ist es, die Experimentierkompetenz der Schüler:innen zu fördern.

Bitte kreuzen Sie an, was die Lehrkraft bei der Planung der Stunde mit dieser Zielsetzung berücksichtigen sollte.

Die Schüler:innen sollten in der Stunde ...

- (1) ... unterschiedliche Experimente zur Habitatpräferenz der Asseln durchführen, um eine Aussage treffen zu können, welches Habitat Asseln präferieren.
- (2) ... mit einer Anleitung praktisch zur Habitatpräferenz von Asseln experimentieren, um möglichst aussagekräftige Ergebnisse zu erhalten.
- (3) ... selbst praktisch zur Habitatpräferenz von Asseln experimentieren, um in der Auswertung die Aussagekraft der Experimente zu besprechen.

Item #2: KISR x Hypothesen formulieren

Eine Biologielehrkraft setzt in einer Unterrichtssequenz zu „Pflanzen und Pflanzenphysiologie“ wiederholt praktische Versuche mit den Schüler:innen ein. Die Funktion von Hypothesen hat sie bislang noch nicht thematisiert. Die Lehrkraft verfolgt in einer Stunde das Ziel, mit den Schüler:innen das Formulieren von naturwissenschaftlichen Hypothesen zu erarbeiten. Dies erfolgt am Beispiel eines Experiments zu Einflussfaktoren der Samenkeimung.

Bitte kreuzen Sie an, was die Lehrkraft bei der Planung der Stunde mit dieser Zielsetzung berücksichtigen sollte.

Die Lehrkraft sollte ...

- (1) ... den Schüler:innen im Einstieg möglichst genau erklären, was eine Hypothese ist, damit sie verstehen, welchen Einfluss die in der Stunde betrachteten Faktoren auf die Samenkeimung haben.
- (2) ... die Schüler:innen auffordern, eigene Hypothesen zu Einflussfaktoren der Samenkeimung zu formulieren, damit sie zu Beginn der Stunde das Vorwissen der Schüler:innen aktivieren kann.
- (3) ... die Schüler:innen Hypothesen zu Einflussfaktoren der Samenkeimung formulieren lassen, damit in der Sicherung darüber nachgedacht werden kann, ob diese mit dem Experiment überprüft werden konnten.

Item #3: KISR x Modellieren

Eine Biologielehrkraft setzt im Themengebiet „Humanbiologie“ häufig Modelle ein, mit deren Hilfe sich die Schüler:innen Struktur und Funktion menschlicher Organe erschließen. Nun möchte sie in der Unterrichtssequenz zu „Struktur und Funktion der Atmungsorgane“ eine Stunde zur Förderung der Kompetenz naturwissenschaftlicher Modellbildung anhand von Lungenmodellen durchführen.

Bitte kreuzen Sie an, was die Lehrkraft bei der Planung der Stunde mit dieser Zielsetzung berücksichtigen sollte.

Die Schüler:innen sollten ...

- (1) ... nach der Erarbeitung von Struktur und Funktion der Lunge ein Modell mit der Lunge vergleichen und Unterschiede zwischen Modell und Original beschreiben.
- (2) ... auf der Basis des vorher erarbeiteten Wissens mit verschiedenen Materialien ein eigenes Modell zu Struktur und Funktion der Lunge entwickeln.
- (3) ... Struktur und Funktion der Lunge mit Hilfe selbst gebauter Modelle erklären und Gründe für die Unterschiedlichkeit dieser Modelle diskutieren.

Item #4: KSU x Experimentieren

Eine Biologielehrkraft möchte mit ihren Schüler:innen den Einfluss abiotischer Faktoren auf die Fotosynthese untersuchen. Sie möchte dabei die Experimentierkompetenz fördern, die bei ihren Schüler:innen bislang kaum ausgeprägt ist. Um mit einem einfachen Experiment einzusteigen, hat die Lehrkraft die Frage ausgewählt, welchen Einfluss die Menge an Kohlenstoffdioxid im Wasser auf die Fotosyntheserate von Wasserpflanzen hat. Dies lässt sich mit Wasserpflanzen in einem Glasgefäß untersuchen, bei denen die entstehenden Gasbläschen gezählt werden. Die Lehrkraft fragt ihre Schüler:innen im Einstieg der Stunde, wie sie das Experiment planen würden, und bittet darum, das Vorgehen zu begründen.

Bitte wählen Sie die Antwort aus, die die Schüler:innen am wahrscheinlichsten geben werden.

- (1) „Wir stecken eine Wasserpflanze in ein mit Leitungswasser gefülltes Glasgefäß und belichten es mit einer Lampe, damit wir die Gasbläschen sehen und zählen können.“
- (2) „Wir stecken eine Wasserpflanze in ein mit Leitungswasser gefülltes Glasgefäß, belichten es mit einer Lampe und geben schrittweise Sprudelwasser hinzu, damit wir die Veränderung der Gasbläschen sehen können.“
- (3) „Wir stecken eine Wasserpflanze in ein mit Leitungswasser und eine zweite in ein mit Sprudelwasser gefülltes Glasgefäß und belichten beide, damit wir die Anzahl der Gasbläschen vergleichen können.“

Item #5: KSU x Hypothesen formulieren

Eine Biologielehrkraft arbeitet mit ihrer Klasse aktuell im Themenfeld „Stoffwechsel des Menschen“. Sie hat bereits die Struktur und die Funktion der Verdauungsorgane erarbeitet und möchte nun die Wirkungsweise von Enzymen thematisieren. Für die nächsten Stunden hat sie erstmals praktisches Arbeiten vorgesehen, für das ihre Schüler:innen auch das Formulieren naturwissenschaftlicher Hypothesen lernen sollen. Die Lehrkraft plant für die nächste Stunde, am Beispiel der Temperaturabhängigkeit von Amylase das Entwickeln von Hypothesen zu üben. Dazu bittet sie die Schüler:innen im Einstieg, an diesem Beispiel zu beschreiben und zu begründen, zu welchem Zweck Hypothesen formuliert werden.

Bitte wählen Sie die Antwort aus, die die Schüler:innen am wahrscheinlichsten geben werden.

„Ich formuliere meine Erwartungen, welchen Einfluss die Temperatur auf die Aktivität von Amylase hat, ...

- (1) ... um in den Versuchsergebnissen zu sehen, ob ich richtig liege.“
- (2) ... um meinen Versuch Schritt für Schritt durchführen zu können.“
- (3) ... um meine Erwartungen mit dem Versuch zu bestätigen oder zu verwerfen.“

Item #6: KSU x Modellieren

Eine Biologielehrkraft plant für ihre Klasse im Themengebiet „Die Zelle – kleinste Funktionseinheit des Lebendigen“ eine Stunde zur Modellbildung. Da sie in dieser Klasse bislang kaum mit Modellen gearbeitet hat, möchte sie die Kompetenz naturwissenschaftlicher Modellbildung bei ihren Schüler:innen fördern. Hierzu entwickeln und diskutieren sie Modelle des einzelligen Organismus Euglena. Da die Lehrkraft weiß, dass es lernförderlich ist, die Vorstellungen von Schüler:innen zu erfassen und in der Stundenplanung zu berücksichtigen, macht sie vorab eine kurze schriftliche Umfrage zum Einsatz von Modellen in der Biologie.

Bitte wählen Sie die Antwort aus, die die Schüler:innen am wahrscheinlichsten geben werden.

- (1) „Biologische Modelle werden von Wissenschaftler:innen erstellt, um mit ihnen forschen zu können.“
- (2) „In der Biologie werden Modelle genutzt, um naturwissenschaftliche Dinge zu erklären.“
- (3) „Biolog:innen setzen unterschiedliche Modelle ein, um sie vergleichen zu können.“

Item #7: KISR→KSU x Experimentieren

Eine Biologielehrkraft möchte im Themenfeld „Ernährung“ die Experimentierkompetenz ihrer Schüler:innen fördern. Sie plant dazu eine Stunde, in der die Schüler:innen ein Experiment zum Einfluss der Temperatur auf die Aktivität des Enzyms Katalase durchführen. Die Lehrkraft wiederholt im Einstieg kurz die Reaktionsgeschwindigkeit-Temperatur-Regel (RGT-Regel) und leitet dann in die Erarbeitungsphase über. Die Schüler:innen führen das vorbereitete Experiment in Gruppen durch und werten es im Plenum aus, indem sie ihre Ergebnisse zur Reaktionsgeschwindigkeit bei unterschiedlichen Temperaturen für die Verdeutlichung der RGT-Regel nutzen.

Bitte kreuzen Sie an, welche Vorstellung über Experimente bei den Schüler:innen durch dieses Vorgehen am wahrscheinlichsten entwickelt wird.

Das Experimentieren ist eine Methode, ...

- (1) ... mit der die in einer Regel formulierten Zusammenhänge überprüft werden können.
- (2) ... mit der eine bereits bestehende Regel bestätigt werden kann.
- (3) ... mit der durch viele Ergebnisse eine Regel verallgemeinert werden kann.

Item #8: KISR→KSU x Hypothesen formulieren

Eine Biologielehrkraft hat eine Unterrichtssequenz zum Experimentieren geplant. Sie hat mehrere Experimente vorbereitet, die die Schüler:innen über einen längeren Zeitraum selbständig durchführen sollen. Zum Einstieg in diese Sequenz sammelt die Lehrkraft die Hypothesen ihrer Schüler:innen zur Frage: „Wodurch können enzymatische Reaktionen in Organismen beeinflusst werden?“ Die Lehrkraft weist darauf hin, dass am Ende der Sequenz noch einmal geschaut werden kann, welche der Hypothesen bestätigt werden können, und teilt dann die Anweisungen zur Durchführung der vorbereiteten Experimente aus.

Bitte kreuzen Sie an, welche Vorstellung über Hypothesen bei den Schüler:innen durch dieses Vorgehen am wahrscheinlichsten entwickelt wird.

Das Formulieren von Hypothesen ...

- (1) ... ist eine gute Möglichkeit, das eigene bisherige Wissen durch ein Experiment zu überprüfen.
- (2) ... ist im naturwissenschaftlichen Unterricht wichtig, um Experimente planen zu können.
- (3) ... ist ein Vorgehen, bei dem die in einem Experiment betrachteten Variablen berücksichtigt werden.

Item #9: KISR→KSU x Modellieren

Eine Biologielehrkraft nutzt Strukturmodelle, um mit ihren Schüler:innen den Aufbau des menschlichen Herzens zu erarbeiten. Die Schüler:innen lesen in Einzelarbeit einen Informationstext über die Struktur des Herzens und identifizieren in Partnerarbeit einzelne Bestandteile des Herzens an Modellen. Abschließend wird im Plenum der Weg des Blutes an den Modellen besprochen, um die neu erarbeiteten Inhalte zu sichern.

Bitte kreuzen Sie an, welche Vorstellung über Modelle bei den Schüler:innen durch dieses Vorgehen am wahrscheinlichsten entwickelt wird?

- (1) Modelle veranschaulichen ein reales Objekt, um das Lernen zu unterstützen.
- (2) Modelle bieten unterschiedliche Möglichkeiten, ein reales Objekt zu veranschaulichen.
- (3) Modelle zeigen, wie man sich ein reales Objekt vorstellen kann.

Item #10: KSU→KISR x Experimentieren

Eine Biologielehrkraft hat mit ihrer Klasse im Biologieunterricht bereits häufig experimentiert. Sie ist überzeugt, dass Schüler:innen die biologischen Zusammenhänge viel besser verstehen, wenn sie sich diese aktiv mit Experimenten erschließen. Im Laufe des Schuljahres möchte die Lehrkraft erfahren, was ihre Klasse bereits über Experimente in der Biologie gelernt hat. Hierzu lässt sie ihre Schüler:innen notieren, was Experimente sind und wofür sie in der Biologie genutzt werden. Dabei zeigt sich: Die vorherrschende Vorstellung in ihrer Klasse ist, dass Experimente im Biologieunterricht genutzt werden, um praktisch zu veranschaulichen, was vorher gelernt wurde.

Bitte kreuzen Sie an, auf welche Weise die Lehrkraft diese Vorstellung über Experimente aufgreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis erweitern sollte.

Sie sollte ihre Schüler:innen ...

- (1) ... selbstgeplante Experimente durchführen und in der Sicherung die Durchführung der Experimente reflektieren lassen.
- (2) ... ein vorab erarbeitetes naturwissenschaftliches Phänomen mit Alltagsbezug mit Hilfe eines Experiments untersuchen lassen.
- (3) ... zu einer noch nicht im Unterricht beantworteten Frage Hypothesen formulieren und in einem Experiment überprüfen lassen.

Item #11: KSU→KISR x Hypothesen formulieren

Eine Biologielehrkraft lässt ihre Schüler:innen zu Beginn ihrer Unterrichtsstunden regelmäßig Hypothesen formulieren. In der Klasse fragt ein Schüler, was das solle: „Wenn ich nicht weiß, wie das Phänomen zu erklären ist, dann rate ich halt irgendetwas, was richtig oder falsch sein kann.“

Bitte kreuzen Sie an, auf welche Weise die Lehrkraft diese Vorstellung über Hypothesen aufgreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis erweitern sollte.

- (1) Sie sollte ihre Schüler:innen dazu auffordern, bei der Auswertung von Experimenten die Funktion der eingangs formulierten Hypothesen zu reflektieren.
- (2) Sie sollte ihren Schüler:innen erklären, dass sich nach dem Experiment durch Verifikation oder Falsifikation der Hypothesen das Phänomen erklären lässt.
- (3) Sie sollte ihren Schüler:innen transparent machen, dass sie durch die formulierten Hypothesen an das Vorwissen ihrer Schüler:innen anknüpfen kann.

Item #12: KSU→KISR x Modellieren

Eine Biologielehrkraft hat im Themenfeld „Genetik“ mit ihren Schüler:innen den Aufbau der DNA mit Modellen erarbeitet. In einer folgenden Stunde fragt sie ihre Schüler:innen, was Modelle sind und welche Funktion diese in der Biologie haben. Eine Schülerin antwortet: „Modelle sind Abbildungen von etwas, zum Beispiel von der DNA. Damit können wir lernen, wie die DNA aufgebaut ist.“

Bitte kreuzen Sie an, auf welche Weise die Lehrkraft diese Vorstellung über Modelle aufgreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis erweitern sollte.

Sie sollte mit ihren Schüler:innen ...

- (1) ... eigene DNA-Modelle herstellen, um mit ihnen das Potenzial von Modellen als Medien herauszustellen.
- (2) ... Modellkritik üben, indem sie Gemeinsamkeiten und Unterschiede zwischen DNA-Modell und DNA erarbeiten.
- (3) ... über Modelle reflektieren, indem sie die Bedeutung der DNA-Modelle für die Entdeckung der DNA-Struktur erarbeiten.

Tabelle 12.2: Lösungsbogen des PCK_{SR}-bio

Item	Richtige Antwort	Begründung
#1	(3)	Falsche Antwortoptionen Antwortoption (1) beschreibt ein Unterrichtsvorgehen zur Förderung von Fachwissen in Bezug auf die Habitatpräferenz von Asseln. Auch in Antwortoption (2) wird durch die Durchführung des Experiments mit einer Anleitung das Ziel verfolgt, Fachwissen zu vermitteln. Vor dem Hintergrund der Zielsetzung der Unterrichtsstunde sind diese Vorgehen damit weniger geeignet, da die Experimentierkompetenz lediglich implizit adressiert wird (Matlen & Klahr, 2013; Vorholzer et al., 2020).
		Richtige Antwortoption Dahingehend eignet sich das eigenständige Experimentieren sowie insbesondere die Besprechung der Aussagekraft der in diesen Experimenten erzielten Ergebnisse in Antwortoption (3), um explizit die Experimentierkompetenz der Schüler:innen zu fördern (Baur et al., 2019; Krell & Krüger, 2022).
#2	(3)	Falsche Antwortoptionen Antwortoption (1) beschreibt ein Unterrichtsvorgehen, bei dem der Fokus auf der Förderung von Fachwissen in Bezug auf Einflussfaktoren der Samenkeimung liegt, weshalb es sich vor dem Hintergrund der beschriebenen Zielsetzung nicht eignet, da die Funktion von Hypothesen lediglich implizit adressiert wird (Matlen & Klahr, 2013; Vorholzer et al., 2020). Auch das in Antwortoption (2) beschriebene Vorgehen eignet sich nicht, da die Funktion von Hypothesen zum naturwissenschaftlichen Erkenntnisgewinn statt zur Aktivierung des Vorwissens der Schüler:innen vermittelt werden sollte (Großmann & Krüger, 2024).
		Richtige Antwortoption Um diese Vorstellung zu vermitteln eignet sich das in Antwortoption (3) beschriebene Vorgehen, da nach Durchführung der Experimente reflektiert werden kann, ob die zuvor formulierten Hypothesen durch diese überprüft werden konnten.
#3	(3)	Falsche Antwortoptionen Antwortoption (1) beschreibt ein Vorgehen, bei dem ein Modell mit dem Original verglichen wird. In dem in Antwortoption (2) beschriebenen Vorgehen entwickeln die Schüler:innen eigene Modelle, es wird jedoch nicht auf ihre Rolle für den naturwissenschaftlichen Erkenntnisgewinn eingegangen. Die Zielsetzung der Unterrichtsstunde (Förderung der Kompetenz naturwissenschaftlicher Modellbildung) kann mit diesem Vorgehen nicht erfüllt werden (Upmeyer zu Belzen et al., 2019).
		Richtige Antwortoption Nur in Antwortoption (3) setzen sich die Schüler:innen mit der Unterschiedlichkeit verschiedener Modelle auseinander und lernen auf diese Weise, dass unterschiedliche Modelle für dasselbe Objekt existieren und sich aus diesen unterschiedliche Erkenntnisse ableiten lassen (Göhner & Krell, 2022; Knuuttila, 2011).

Tabelle 12.2: Lösungsbogen des PCK_{SR}-bio (Fortsetzung)

Item	Richtige Antwort	Begründung
#4	(1)	Falsche Antwortoptionen Die in Antwortoption (2) beschriebene Vorstellung über die Beziehung von zwei Variablen zueinander sowie die in Antwortoption (3) beschriebene Vorstellung über die systematische Kontrolle von Variablen setzen eine ausgeprägte Experimentierkompetenz voraus, weshalb sie von der beschriebenen Lerngruppe eher nicht zu erwarten sind (Wellnitz & Mayer, 2016).
		Richtige Antwortoption Da die Schüler:innen bisher über kaum ausgeprägte Experimentierkompetenz verfügen, ist es am wahrscheinlichsten, dass sie wie in Antwortoption (1) beschrieben eindrucksvolle Effekte durch ein Experiment erzielen wollen und die betrachteten Variablen daher unsystematisch variieren (Kranz et al., 2023; Wellnitz & Mayer, 2016).
#5	(1)	Falsche Antwortoptionen Die in Antwortoption (2) beschriebene Vorstellung erfordert ein Verständnis über das hypothesengeleitete Experimentieren und die in Antwortoption (3) beschriebene Vorstellung ein Verständnis über die Falsifizierbarkeit von Hypothesen, welches ein elaborierteres Verständnis über die Rolle von Hypothesen für den naturwissenschaftlichen Erkenntnisgewinn voraussetzen und von der beschriebenen Lerngruppe daher eher nicht zu erwarten sind (Wellnitz & Mayer, 2016).
		Richtige Antwortoption Da die Schüler:innen bisher keine Kenntnisse über das Formulieren von Hypothesen aufweisen, ist zu erwarten, dass sie die in Antwortoption (1) beschriebene Vorstellung aufweisen und Hypothesen als Möglichkeit zum Testen des eigenen Wissens ansehen (Kattmann, 2015; Kranz et al., 2023).
#6	(2)	Falsche Antwortoptionen Da die Lehrkraft in der Klasse bisher kaum mit Modellen gearbeitet hat, ist nicht zu erwarten, dass die Schüler:innen Modelle wie in Antwortoption (1) beschrieben als Mittel zur Erkenntnisgewinnung sehen oder wie in Antwortoption (3) beschrieben Kenntnis davon haben, dass zu einem Phänomen unterschiedliche Modelle existieren können (Upmeyer zu Belzen et al., 2019).
		Richtige Antwortoption Daher ist von der beschriebenen Lerngruppe davon auszugehen, dass die Schüler:innen nur über fundamentale Modellierkompetenzen verfügen und Modelle wie in Antwortoption (2) beschrieben lediglich als Anschauungsobjekte ansehen (Upmeyer zu Belzen et al., 2019).

Tabelle 12.2: Lösungsbogen des PCK_{SR}-bio (Fortsetzung)

Item	Richtige Antwort	Begründung
#7	(2)	Falsche Antwortoptionen Damit die Schüler:innen die in Antwortoption (1) beschriebene Vorstellung über die Zusammenhänge einer Regel oder die in Antwortoption (3) beschriebene Vorstellung über die Verallgemeinerung einer Regel entwickeln können, ist es erforderlich, eine unbekannte Regel durch hypothesengestütztes Experimentieren erarbeiten zu lassen (Baur et al., 2019). Dies ist in dem beschriebenen Unterrichtsvorgehen nicht gegeben, da den Schüler:innen die in dem Experiment veranschaulichte RGT-Regel bereits bekannt ist und die durchzuführenden Experimente vorbereitet sind.
		Richtige Antwortoption Durch das beschriebene Unterrichtsvorgehen ist es wahrscheinlich, dass die in Antwortoption (2) beschriebene Vorstellung, dass Experimente der Veranschaulichung bereits bestehenden Wissens dienen, entwickelt wird (Kranz et al., 2023; Wellnitz & Mayer, 2016).
#8	(1)	Falsche Antwortoptionen Um die in Antwortoption (2) und Antwortoption (3) beschriebenen Vorstellungen hinsichtlich der Rolle von Hypothesen für die Planung von Experimenten zu entwickeln, ist es notwendig, solche Experimente auf Grundlage von zuvor formulierten Hypothesen zu planen und im Anschluss insbesondere die Passung der durchgeführten Experimente zur Überprüfung dieser zu reflektieren (Baur et al., 2019). Da die Experimente in dem beschriebenen Unterrichtsvorgehen jedoch vorbereitet sind und die Durchführung auf Grundlage einer Anleitung erfolgt, ist nicht davon auszugehen, dass diese Vorstellungen entwickelt werden.
		Richtige Antwortoption Durch das beschriebene Unterrichtsvorgehen ist hingegen davon auszugehen, dass die Schüler:innen die in Antwortoption (1) beschriebene Vorstellung entwickeln, dass Hypothesen zur Testung des eigenen Wissens verwendet werden (Kranz et al., 2023; Wellnitz & Mayer, 2016).
#9	(1)	Falsche Antwortoptionen Um die in Antwortoption (2) beschriebene Vorstellung zu entwickeln, müssten die Schüler:innen Kenntnisse über alternative Darstellungsformen und deren unterschiedliche Funktion erlangen (Upmeier zu Belzen et al., 2019). Auch um die in Antwortoption (3) beschriebene Vorstellung zu entwickeln ist es erforderlich, dass die Schüler:innen lernen, dass Modelle nur theoretische Rekonstruktionen eines Phänomens und damit nur Möglichkeiten sind, dieses Phänomen darzustellen (Upmeier zu Belzen et al., 2019). Da die Modelle in dem beschriebenen Unterrichtsvorgehen jedoch lediglich dafür genutzt werden, um das zuvor gelernte Wissen visuell zu veranschaulichen, ist es unwahrscheinlich, dass die Schüler:innen diese Vorstellungen entwickeln.
		Richtige Antwortoption Wahrscheinlicher ist durch das beschriebene Unterrichtsvorgehen auszugehen, dass die Schüler:innen die in Antwortoption (1) beschriebene Vorstellung entwickeln, dass ein Modell ein reales Objekt so abbildet, wie es ist (Upmeier zu Belzen et al., 2019).

Tabelle 12.2: Lösungsbogen des PCK_{SR}-bio (Fortsetzung)

Item	Richtige Antwort	Begründung
#10	(3)	Falsche Antwortoptionen In Antwortoption (1) planen die Schüler:innen selbstständig Experimente, jedoch wird im Anschluss nur die Durchführung und nicht die Aussagekraft dieser reflektiert. In Antwortoption (2) wird ein Experiment zu einem bereits bekannten Phänomen durchgeführt, wodurch lediglich das bereits vorhandene Wissen der Schüler:innen bestätigt wird. Da die Experimente auf diese Weisen nicht genutzt werden, um neue Erkenntnisse zu erlangen, eignen sich diese Unterrichtsvorgehen nicht, um die gezeigte Vorstellung aufzugreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis zu ergänzen (Krell & Krüger, 2022).
		Richtige Antwortoption Antwortoption (3) stellt hingegen ein Vorgehen da, bei dem die Schüler:innen Experimente zu einem ihnen unbekannten Phänomen auf Grundlage von zuvor formulierten Hypothesen entwickeln und auf diese Weise den Prozess der naturwissenschaftlichen Erkenntnisgewinnung nachvollziehen können (Baur et al., 2019).
#11	(1)	Falsche Antwortoptionen In Antwortoption (2) wird die Vorstellung vermittelt, dass Experimente dazu genutzt werden können, um Hypothesen auf ihre Richtigkeit zu überprüfen. In Antwortoption (3) werden Hypothesen als Erklärungsversuche zur Aktivierung von Vorwissen verwendet. Beide Unterrichtsvorgehen vermitteln damit keine Kenntnisse über die Rolle von Hypothesen für den naturwissenschaftlichen Erkenntnisgewinn als nicht verifizierbare, sondern lediglich falsifizierbare, empirisch überprüfbare Voraussage eines Untersuchungsergebnisses (Großmann & Krüger, 2024; Krell & Krüger, 2022).
		Richtige Antwortoption Antwortoption (1) beschreibt im Gegensatz dazu ein sinnvolles Unterrichtsvorgehen, bei dem die Vorstellung der Schüler:innen aufgegriffen werden kann und insbesondere die Rückschau dazu genutzt werden kann, die Rolle von Hypothesen als Grundlage des hypothesengeleiteten Experimentierens zu reflektieren.
#12	(3)	Falsche Antwortoptionen In Antwortoption (1) werden Modelle ausschließlich als Medien beschrieben, in Antwortoption (2) wird ausschließlich Modellkritik geübt. Beides entspricht den gängigen Vorstellungen von Schüler:innen (Upmeyer zu Belzen et al., 2019). Auch wenn beide Vorgehen im Biologieunterricht generell sinnvoll sind, beschreiben sie keinen geeigneten Weg, wie die Lehrkraft diese Vorstellung über Modelle aufgreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis erweitern sollte.
		Richtige Antwortoption Antwortoption (3) hingegen beschreibt ein Unterrichtsvorgehen, in dem die Schüler:innen lernen, wie Erkenntnisse durch die Verwendung von Modellen gewonnen werden können (Göhner & Krell, 2022; Knuuttila, 2011) und beschreibt damit ein geeignetes Vorgehen.

Literatur

- Baur, A., Emden, M. & Bewersdorff, A. (2019). Welche Unterrichtsprinzipien sollten für den Aufbau von Kompetenzen zum Experimentieren Beachtung finden? Eine Ableitung auf Basis multiperspektivisch begründeter Unterrichtsziele. *Zeitschrift für Didaktik der Biologie*, 23, 10–24. <https://doi.org/10.4119/zdb-1738>
- Göhner, M. & Krell, M. (2022). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*, 52(2), 395–414. <https://doi.org/10.1007/s11165-020-09945-7>
- Großmann, L. & Krüger, D. (2024). Erkenntnisgewinnung (v)erklärt: Wie plant man hypothesengeleiteten Biologieunterricht? In B. Reinisch, D. Krüger & D. Mahler (Hrsg.), *Biologiedidaktische Nature of Science-Forschung: Zukunftsweisende Praxis* (S. 99–112). Springer. https://doi.org/10.1007/978-3-662-68409-2_9
- Kattmann, U. (2015). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht*. Aulis Verlag.
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42(2), 262–271. <https://doi.org/10.1016/j.shpsa.2010.11.034>
- Kranz, J., Baur, A. & Möller, A. (2023). Learners' challenges in understanding and performing experiments: A systematic review of the literature. *Studies in Science Education*, 59(2), 321–367. <https://doi.org/10.1080/03057267.2022.2138151>
- Krell, M. & Krüger, D. (2022). Erkenntnisgewinnung im Biologieunterricht: Welches Methodenwissen soll vermittelt werden? *MNU Journal*, 5, 376–382.
- Matlen, B. J. & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science*, 41(3), 621–634. <https://doi.org/10.1007/s11251-012-9248-z>
- Vorholzer, A., Aufschnaiter, C. von & Boone, W. J. (2020). Fostering upper secondary students' ability to engage in practices of scientific investigation: A comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education*, 50(1), 333–359. <https://doi.org/10.1007/s11165-018-9691-1>
- Wellnitz, N. & Mayer, J. (2016). Methoden der Erkenntnisgewinnung im Biologieunterricht. In A. Sandmann & P. Schmiemann (Hrsg.), *Biologie lernen und lehren: Bd. 1. Biologiedidaktische Forschung: Schwerpunkte und Forschungsstände* (S. 61–82). Logos Verlag Berlin.
- Upmeier zu Belzen, A., van Driel, J., & Krüger, D. (2019). Introducing a framework for modeling competence. In A. Upmeier zu Belzen, D. Krüger, & J. van Driel (Hrsg.), *Towards a competence-based view on models and modeling in science education: Bd. 12. Models and modeling in science education* (S. 3–19). Springer. https://doi.org/10.1007/978-3-030-30255-9_1

12.2 Englische Version des PCK_{SR}-bio

Im Folgenden befindet sich eine Übersicht über die Items der englischsprachigen Version des PCK_{SR}-bio.

Verwendet in Studie 5.

Item 1: KISR x "Planning and Conducting Investigations"

A biology teacher conducts an investigation with her students on the habitat preference of isopods. The students are to investigate the extent to which isopods prefer light or dark or moist or dry locations. The aim of this lesson is to promote the students' ability to investigate.

Please tick what the teacher should consider when planning the lesson with this objective.

In the lesson, the students should...

- (1) ... conduct different investigations on the habitat preference of isopods to make a statement about which habitat isopods prefer.
- (2) ... investigate practically on the habitat preference of isopods with written instructions in order to obtain the most meaningful results.
- (3) ... conduct investigations themselves on the habitat preference of isopods to discuss the significance of the experiments.

Item 2: KISR x "Generating Hypotheses"

A biology teacher repeatedly uses practical experiments with his students in a lesson sequence on "Plants and Plant Physiology". He has not yet addressed the function of hypotheses. In one lesson, the teacher aims to improve the students' ability to formulate a scientific hypothesis. This is done using the example of an experiment on factors influencing seed germination.

Please tick what the teacher should consider when planning the lesson with this objective.

The teacher should...

- (1) ... explain to the students as precisely as possible in the introduction of the lesson what a hypothesis is. Through this, the students understand what influence the factors considered in the lesson have on seed germination.
- (2) ... ask the students to formulate their own hypotheses on factors influencing seed germination. Through this, the teacher can activate the students' prior knowledge at the beginning of the lesson.
- (3) ... have the students formulate hypotheses on factors influencing seed germination. Through this, the students can reflect on whether these could be tested with the experiment.

Item 3: KISR x "Using Scientific Models"

A biology teacher often uses models to help students understand the structure and function of human organs. In the lesson sequence on "Structure and Function of the Respiratory Organs", she would like to conduct a lesson to promote their students' understanding of scientific modelling using lung models.

Please tick what the teacher should consider when planning the lesson with this objective.

The students should...

- (1) ... after learning about the structure and function of the lungs, compare a model with the lungs and describe the differences between the model and the object.
- (2) ... develop their own models of the structure and function of the lungs based on the previously acquired knowledge and using various materials.
- (3) ... explain the structure and function of the lungs with the help of self-made models and discuss reasons for the differences between these models.

Item 4: KSU x "Planning and Conducting Investigations"

A biology teacher wants to investigate the influence of abiotic factors on photosynthesis with his students. In doing so, he would like to promote the students' experimental ability, which has not yet been fully. The teacher has chosen the question of how the amount of carbon dioxide in the water influences the photosynthesis rate of aquatic plants. This can be investigated with aquatic plants in a glass jar and counting the gas bubbles that are produced. At the beginning of the lesson, the teacher asks the students how they would plan the experiment and asks them to justify the procedure.

Please select the answer that the students are most likely to give.

- (1) "We put an aquatic plant in a glass jar filled with tap water and expose it to light with a lamp so we can see and count the gas bubbles."
- (2) "We put an aquatic plant in a glass jar filled with tap water, exposed it to light with a lamp, and gradually add sparkling water so we can see the change in gas bubbles."
- (3) "We put one water plant in a glass jar filled with tap water and a second one in a glass jar filled with sparkling water and expose them both to light so that we can compare the number of gas bubbles."

Item 5: KSU x "Generating Hypotheses"

A biology teacher has already worked with her students on the structure and function of the digestive organs and now wants to focus on how enzymes work. For the next lessons, she has planned practical work for the first time, for which her students should also learn how to formulate scientific hypotheses. Therefore, she plans to use the example of the temperature dependence of amylase to practise developing hypotheses. For this purpose, she asks the students to describe and justify the purpose for which hypotheses are formulated.

Please select the answer that the students are most likely to give.

"I formulate my expectations about the effect of temperature on the activity of amylase, ...

- (1) ... to see in the experimental results if I am right."
- (2) ... to be able to conduct my experiment step by step."
- (3) ... to confirm or reject my expectations with the experiment."

Item 6: KSU x "Using Scientific Models"

A biology teacher is planning a lesson on modelling for his class on the topic "The Cell - the Smallest Functional Unit of Life". As he has worked with models very little in this class so far, he wants to promote the ability of scientific modelling in his students. For this purpose, they develop and discuss models of the unicellular organism *Euglena*. Since the teacher knows that recording the students' ideas and consider them in the lesson planning is beneficial for learning, he first makes a short written survey on the use of models in biology.

Please select the answer that the students are most likely to give.

- (1) "Biological models are created by scientists to conduct research with them."
- (2) "In biology, models are used to explain scientific content."
- (3) "Biologists use different models to compare them."

Item 7: KISR→KSU x "Planning and Conducting Investigations"

In a lesson series on nutrition, a biology teacher would like to promote the experimental ability of her students. She plans a lesson in which the students conduct an experiment on the influence of temperature on the activity of the enzyme catalase. The teacher briefly repeats the rule about the effect of temperature on reaction-rate in the beginning of the lesson. Then, the students conduct the prepared experiment in groups and evaluate it in plenary by using their results on the reaction rate at different temperatures to illustrate the rule about the effect of temperature on reaction-rate.

Please tick which idea about experiments is most likely to be developed in the students through this procedure.

Experimentation is a method...

- (1) ... with which the relationships formulated in a rule can be checked.
- (2) ... with which an already existing rule can be confirmed.
- (3) ... with which a rule can be generalised through many results.

Item 8: KISR→KSU x “Generating Hypotheses”

A biology teacher has prepared several experiments that the students are to conduct independently over a longer period of time. To begin the sequence, the teacher collects hypotheses from his students on the question: "How can enzymatic reactions in organisms be influenced?" The teacher points out that at the end of the sequence it is possible to look again to see which of the hypotheses can be confirmed, and then hands out the instructions for carrying out the prepared experiments.

Please tick which idea about hypotheses is most likely to be developed in the students through this procedure.

Formulating hypotheses...

- (1) ... is a good way to test one's own prior knowledge through an experiment.
- (2) ... is important in science lessons to be able to plan experiments.
- (3) ... is a procedure in which the variables in an experiment are considered.

Item 9: KISR→KSU x “Using Scientific Models”

A biology teacher uses structural models to work with her students on the structure of the human heart. The students work individually to read an information text about the structure of the heart and work in pairs to identify individual components of the heart in models. Finally, the path of the blood is discussed with the class using the models in order to secure the newly acquired content.

Please tick which idea about models is most likely to be picked up by students through this approach.

- (1) Models illustrate a real object to support learning.
- (2) Models provide different ways of representing a real object.
- (3) Models show how to imagine a real object.

Item 10: KSU→KISR x “Planning and Conducting Investigations”

A biology teacher has experimented frequently with his class in biology lessons. He is convinced that students better understand biological relationships when they actively explore them through experiments. The teacher wants to determine what his students have learned about experiments in biology. He asks his students what experiments are and what they are used for in biology. This shows that the predominant idea in his class is that experiments are used in biology lessons to practically illustrate what has been learned before.

Please tick the ways in which the teacher should take up this idea about experiments and extend it through an understanding that is as scientifically appropriate as possible.

He should have his students...

- (1) ... conduct self-planned experiments relating to a self-chosen topic and reflect on how the experiments were conducted.
- (2) ... investigate a scientific phenomenon related to everyday life that they have worked out in advance with the help of an experiment.
- (3) ... formulate hypotheses on a question that has not yet been answered in class and have the hypotheses tested in an experiment.

Item 11: KSU→KISR x “Generating Hypotheses”

A biology teacher regularly asks her students to formulate hypotheses at the beginning of their lessons. A student asks what the point is: "If I don't know how to explain the phenomenon, then I just guess something that can be right or wrong."

Please tick the ways in which the teacher should take up this idea about hypotheses and extend it through an understanding that is as scientifically appropriate as possible.

- (1) She should ask her students to reflect on the function of the hypotheses formulated at the beginning when evaluating experiments.
- (2) She should explain to her students that after the experiment, the phenomenon can be explained by verifying or falsifying the hypotheses.
- (3) She should explain to her students that she can build on their prior knowledge through the formulated hypotheses.

Item 12: KSU→KISR x “Using Scientific Models”

A biology teacher has worked with his students on the structure of DNA using models. In a subsequent lesson, he asks his students what models are and what function they have in biology. One student answers: "Models are visualisations of something, for example DNA. We can use them to learn how DNA is built."

Please tick the ways in which the teacher should take this idea about models and extend it through an understanding that is as scientifically appropriate as possible.

He should...

- (1) ... ask his students to make their own DNA models to highlight the potential of models as representations.
- (2) ... practise “model critique” with his students by working out similarities and differences between the DNA model and DNA.
- (3) ... reflect on models by elaborating on the importance of DNA models in the discovery of DNA structure.

12.3 Erweiterung des PCK_{SR}-bio

Im Folgenden befindet sich eine Übersicht über die über den Rahmen dieser Dissertation hinaus zusätzlich entwickelten Items des PCK_{SR}-bio (Tabelle 12.3), eine Übersicht über die Items sowie ein dazugehöriger Lösungsbogen (Tabelle 12.4).

Die neu entwickelten Items wurden gemeinsam mit Jule Flachsberger entwickelt und erstmals mit Biologie-Lehramtsstudierenden im Bachelor und Master sowie praktizierender Lehrkräfte unterschiedlicher naturwissenschaftlicher Fächer im Rahmen einer Masterarbeit eingesetzt (Meinshausen, 2024).

Tabelle 12.3: Übersicht über die Systematik der neu entwickelten Items des PCK_{SR}-bio

	Experimentieren	Hypothesen formulieren	Modellieren
KISR	Item #13	Item #14	Item #15
KSU	Item #16	Item #17	Item #18
KISR→KSU	Item #19	Item #20	Item #21
KSU→KISR	Item #22	Item #23	Item #24

Item #13: KISR x Experimentieren

Um die Experimentierkompetenz ihrer Schüler:innen zu fördern, führt eine Biologielehrkraft im Themenfeld „Vererbung“ ein digitales Experiment zur Weitergabe von Erbinformationen anhand der Mendelschen Regeln durch. Dabei wird der Einfluss von dominanten und rezessiven Allelen auf die unterschiedlichen Phänotypen der Erbsensamen in den Nachfolgegenerationen untersucht werden. Am Computer haben die Schüler:innen die Möglichkeit, digitale Erbsenpflanzen mit gelben und grünen sowie glatten und runzligen Samen zu kreuzen.

Bitte kreuzen Sie an, was die Lehrkraft bei der Planung der Stunde mit dieser Zielsetzung berücksichtigen sollte.

Die Schüler:innen sollten...

- (1) ... ihre Ergebnisse am Ende der Stunde schriftlich sichern, um die Mendelschen Regeln zu erarbeiten.
- (2) ... nach Durchführung der Experimente gemeinsam die Aussagekraft dieser in der Sicherung reflektieren.
- (3) ... nach der Durchführung Schwierigkeiten im Umgang mit dem digitalen Experiment diskutieren.

Item #14: KISR x Hypothesen formulieren

In einer Unterrichtsstunde der Sequenz „Geschichte und Verwandtschaft des Menschen“ möchte eine Biologielehrkraft mit ihren Schüler:innen das erste Mal die Funktion von naturwissenschaftlichen Fragestellungen und Hypothesen in Bezug auf das Vergleichen und Ordnen von biologischen Merkmalen einführen. Dazu stellt sie die Schädel unterschiedlicher Vertreter der *Homo*-Gattung sowie einige Vergleichskriterien zur Verfügung. Mit Hilfe dieser Kriterien sollen die Schüler:innen, anhand unterschiedlicher Merkmale der Schädel, Hypothesen zu deren Verwandtschaft aufstellen.

Bitte kreuzen Sie an, was die Lehrkraft bei der Planung der Stunde mit dieser Zielsetzung berücksichtigen sollte.

Die Lehrkraft sollte...

- (1) ... in der Rückschau überprüfen, ob die Hypothesen dazu genutzt werden können, die Unterschiede der beobachteten Merkmale zu erklären.
- (2) ... zu Beginn den Aufbau von Hypothesen erklären, damit die Schüler:innen angemessene Hypothesen zur Verwandtschaft der Schädel formulieren.
- (3) ... die Schüler:innen eigenständig eine Fragestellung formulieren lassen, auf deren Basis sie im Anschluss Beobachtungs- und Vergleichshypothesen aufstellen.

Item #15: KISR x Modellieren

Eine Biologielehrkraft hat im Themengebiet „Humanbiologie“ bereits häufig Modelle verwendet, um die Struktur menschlicher Organe zu veranschaulichen. Nachdem sie im Unterricht das Thema „Blutgruppen des Menschen“ behandelt hat, möchte sie nun die Modellierkompetenz ihrer Schüler:innen in Bezug auf Modelle als Werkzeuge naturwissenschaftlicher Erkenntnisgewinnung fördern. Vor diesem Hintergrund sollen die Schüler:innen selbstständig Modelle zu den unterschiedlichen menschlichen Blutgruppen entwickeln.

Bitte kreuzen Sie an, was die Lehrkraft bei der Planung der Stunde mit dieser Zielsetzung berücksichtigen sollte.

Die Lehrkraft sollte am Ende der Unterrichtsstunde...

- (1) ... mit ihren Schüler:innen darüber diskutieren, warum es sinnvoll ist, die unterschiedlichen Modelle miteinander zu vergleichen.
- (2) ... sicherstellen, dass alle Modelle den Sachverhalt korrekt wiedergeben, damit das Lernziel von allen Schüler:innen erreicht ist.
- (3) ... deutlich machen, welche Vorteile Modelle für die Darstellung menschlicher Organe im Vergleich zur Funktion menschlicher Blutgruppen haben.

Item #16: KSU x Experimentieren

Eine Biologielehrkraft hat mit ihrer Klasse bisher noch nie experimentiert und möchte dies nun erstmals in einer Unterrichtsstunde zum Kniesehenreflex tun. Um den Unterschied zwischen einer gewollten und einer ungewollten Bewegung (Reflex) zu untersuchen, sollen sich die Schüler:innen zunächst in Zweiergruppen aufteilen. Während eine Person auf einem Stuhl sitzt, schlägt eine andere vorsichtig mit einem Holzlöffel an verschiedene Stellen des Beines der sitzenden Person. Sobald diese eine Berührung fühlt, streckt sie das Bein aus, wodurch ein vor ihr aufgebauter Stromkreis unterbrochen wird. Dabei wird die Zeit zwischen dem Schlag und dem Unterbrechen des Stromkreises gemessen. Trifft der Holzlöffel die richtige Stelle, sodass der Kniesehenreflex ausgelöst wird, führt dies zu einer schnelleren Unterbrechung als bei einer bewussten Bewegung. Vor der Durchführung dieses Experiments bittet die Lehrkraft ihre Schüler:innen zunächst, dieses Vorgehen zu begründen.

Bitte wählen Sie die Antwort aus, die die Schüler:innen am wahrscheinlichsten geben werden.

- (1) „Mit diesem Vorgehen können wir den Versuch häufig wiederholen, wenn wir das Bein versehentlich bewegt haben.“
- (2) „Durch dieses Vorgehen können wir an verschiedenen Stellen des Beins vergleichen, ob der Kniesehenreflex ausgelöst wird.“
- (3) „Dadurch können wir die Stelle am Bein finden, mit der wir es schaffen, das Bein möglichst schnell zu bewegen.“

Item #17: KSU x Hypothesen formulieren

Eine Biologielehrkraft hat mit ihren Schüler:innen im Themenfeld „Geschichte und Verwandtschaft“ über die Rolle von Stammbäumen als naturwissenschaftliche Modelle gesprochen und möchte nun das Formulieren naturwissenschaftlicher Hypothesen anhand dieser erarbeiten. Da ihre Schüler:innen die Funktion einer Hypothese bisher noch nicht kennen, fragt die Lehrkraft sie zum Einstieg, zu welchem Zweck Hypothesen bei der Arbeit mit naturwissenschaftlichen Modellen formuliert werden.

Bitte wählen Sie die Antwort aus, die die Schüler:innen am wahrscheinlichsten geben werden.

„Hypothesen können genutzt werden, um...

- (1) ... danach ein Experiment zu planen, mit dem ich meine Frage beantworten kann.“
- (2) ... im Anschluss zu überprüfen, ob meine Vermutung am Anfang richtig war.“
- (3) ... nach der Überprüfung zu sehen, ob ich den Stammbaum anpassen muss.“

Item #18: KSU x Modellieren

Eine Biologielehrkraft plant für ihre Klasse eine Stunde zur Förderung der bisher kaum vorhandenen Modellierkompetenz ihrer Schüler:innen. Am Beispiel der Hausmaus (*Mus musculus*) möchte sie den Zweck von Modellorganismen für die Erforschung von Alzheimer-Demenz in der Wissenschaft behandeln. Für ihre Stundenplanung erfragt sie dazu zunächst die Vorstellungen ihrer Schüler:innen über den Zweck von Modellorganismen.

Bitte wählen Sie die Antwort aus, die die Schüler:innen am wahrscheinlichsten geben werden.

„Modellorganismen haben den Zweck...

- (1) ... Experimente an ihnen durchzuführen, weil sie sich schnell vermehren können.“
- (2) ... zu sehen, ob die Maus stirbt und man das nicht an Menschen testen kann.“
- (3) ... die Wirkung von verschiedenen Medikamenten testen zu können.“

Item #19: KISR→KSU x Experimentieren

Eine Biologielehrkraft hat im Unterricht zum Themenfeld „Vererbung“ die Mendelschen Regeln bereits ausführlich behandelt. Nun möchte sie die Experimentierkompetenz ihrer Schüler:innen mit Hilfe eines digitalen Experiments fördern. Zu Beginn der Unterrichtsstunde wiederholt sie deshalb die Uniformitäts-, Spaltungs- und Unabhängigkeitsregel und teilt die Klasse im Anschluss in drei Gruppen auf. Jede Gruppe simuliert daraufhin mit Hilfe von Kreuzungsexperimenten eine der drei Regeln am Computer. Im Anschluss werden die Ergebnisse im Plenum gesammelt, um die Mendelschen Regeln für alle Schüler:innen noch einmal zu verdeutlichen.

Bitte kreuzen Sie an, welche Vorstellung über Experimente bei den Schüler:innen durch dieses Vorgehen am wahrscheinlichsten entwickelt wird.

- (1) Experimente werden entwickelt und durchgeführt, um meine Vermutungen zu einer Regel zu überprüfen.
- (2) Mit Experimenten können wir nachweisen, dass unser theoretisches Wissen auch in der Realität funktioniert.
- (3) Experimente dienen dazu, mein Wissen mit einem Versuch noch einmal praktisch überprüfen zu können.

Item #20: KISR→KSU x Hypothesen formulieren

Eine Biologielehrkraft hat eine Unterrichtsstunde über den Wassertransport in Pflanzen vorbereitet, in der auch ein Experiment durchgeführt werden soll. Im Vorfeld hat sie bereits den Aufbau von Pflanzen mit Wasserleitung in der Sprossachse thematisiert. Nun bringt sie zwei weiße Rosen mit, die frisch angeschnitten in zwei Bechergläser gestellt werden. Diese sind mit unterschiedlich gefärbtem Wasser gefüllt. Im Anschluss stellen die Schüler:innen Hypothesen dazu auf, was bei den Rosen zu beobachten sein wird. Die Lehrkraft kündigt an, dass am Ende der Unterrichtsstunde überprüft werden kann, welche der Schüler:innen richtig lagen.

Bitte kreuzen Sie an, welche Vorstellung über Hypothesen bei den Schüler:innen durch dieses Vorgehen am wahrscheinlichsten entwickelt wird.

Hypothesen...

- (1) ... müssen nicht immer richtig sein, denn auch durch widerlegte Hypothesen können neue Erkenntnisse gewonnen werden.
- (2) ... dienen dazu, mein Wissen über den Ausgang eines Experiments zu überprüfen.
- (3) ... können überprüft werden, indem man ein Experiment entwickelt, mit dem man sie testet.

Item #21: KISR→KSU x Modellieren

Eine Biologielehrkraft möchte im Themengebiet „Ökologie“ eine Unterrichtsstunde zur Änderung von Modellen durchführen. Dazu hat sie mehrere Modelle vorbereitet, die den Zusammenhang von abiotischen und biotischen Umweltfaktoren auf das Verhalten von Libellen verdeutlichen. Die Modelle unterscheiden sich dabei hinsichtlich ihrer Komplexität und Darstellungsweise und spiegeln die schrittweise Entwicklung eines Ursprungsmodells wider. In der Unterrichtsstunde sollen die Schüler:innen die Modelle nun in die richtige Reihenfolge bringen. Dazu steht ihnen ein Informationstext sowie eine Unterhaltung zwischen zwei Forschenden über die schrittweise Entwicklung des Modells auf Basis von Forschungsdaten zur Verfügung.

Bitte kreuzen Sie an, welche Vorstellung über Modelle bei den Schüler:innen durch dieses Vorgehen am wahrscheinlichsten entwickelt wird?

- (1) ... Wenn man einen Fehler findet, muss das Modell überarbeitet werden, damit es richtig ist.
- (2) ... Modelle werden verändert, um neue Erkenntnisse aus Untersuchungen zu integrieren.
- (3) ... Es gibt ein richtiges Modell, welches die Fehler der jeweiligen Vorgängermodelle behebt.

Item #22: KSU→KISR x Experimentieren

Eine Biologielehrkraft hat mit ihrer Klasse im Verlauf des Schuljahres bereits häufiger experimentiert. Zu Beginn einer neuen Unterrichtsreihe mit dem Thema „Signaltransduktion“ fragt sie die Schüler:innen, wie sie ein Experiment zur Kommunikation von Zellen durchführen würden. Ein Schüler antwortet: „Woher sollen wir das wissen? Experimente durchzuführen ergibt doch nur Sinn, wenn man schon etwas über das Thema gelernt hat!“. Er erhält von seinen Mitschüler:innen Zuspruch.

Bitte kreuzen Sie an, auf welche Weise die Lehrkraft diese Vorstellung über Experimente aufgreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis erweitern sollte.

Die Lehrkraft sollte diese Aussage aufgreifen...

- (1) ... und ein Alltagsphänomen zusammen mit den Schüler:innen erarbeiten und in der Folgestunde anhand eines Experiments demonstrieren.
- (2) ... und verdeutlichen, dass in der Wissenschaft unvoreingenommen an ein Thema gegangen werden sollte, um vertrauenswürdige Ergebnisse zu erzielen.
- (3) ... und den Schüler:innen in der darauffolgenden Stunde zeigen, wie Experimente dabei helfen können, neue Erkenntnisse zu erlangen.

Item #23: KSU→KISR x Hypothesen formulieren

In einer Unterrichtsreihe zu dem Thema „Reproduktion und Vererbung“ hat eine Biologielehrkraft mit ihren Schüler:innen bereits mehrfach das Formulieren von Hypothesen geübt. Um herauszufinden, welchen Einfluss dies auf die Vorstellung ihrer Schüler:innen in Bezug auf die Entwicklung von Hypothesen hatte, führt sie eine kurze schriftliche Umfrage durch. Dabei stellt sich heraus, dass viele Schüler:innen davon ausgehen, Hypothesen seien eine Formalität, um vor Beginn eines Experiments zu beschreiben, wie es ausgeht.

Bitte kreuzen Sie an, auf welche Weise die Lehrkraft diese Vorstellung über Hypothesen aufgreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis erweitern sollte.

Die Lehrkraft sollte...

- (1) ... die Schüler:innen bewusst mit falschen Hypothesen in die Irre laufen lassen, um ihnen zu zeigen, dass diese widerlegt werden können.
- (2) ... die Verifizierbarkeit von Hypothesen verdeutlichen, indem sie selbstständig formulierte Hypothesen der Schüler:innen überprüfen lässt.
- (3) ... verschiedene Hypothesen ihrer Schüler:innen zu einem unbekannten Phänomen sammeln und basierend darauf Experimente entwickeln lassen.

Item #24: KSU→KISR x Modellieren

In einer Unterrichtsstunde über Epidemien hat eine Biologielehrkraft mit ihren Schüler:innen die Relevanz der Revision von Modellen am Beispiel der Infektionszahlen und Handlungsstrategien zu unterschiedlichen Zeitpunkten während der COVID-19-Pandemie behandelt. Eine Schülerin fragt am Ende der Stunde: „Warum gab es überhaupt so viele verschiedene Modelle von unterschiedlichen Wissenschaftler:innen? Man hätte sich doch direkt am Anfang auf eins einigen können!“

Bitte kreuzen Sie an, auf welche Weise die Lehrkraft diese Vorstellung über Modelle aufgreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis erweitern sollte.

Die Lehrkraft sollte diese Aussage aufgreifen indem sie...

- (1) ... die Schüler:innen verschiedene Modelle entwickeln lässt, die die Auswirkungen unterschiedlicher Handlungsstrategien auf die Infektionszahlen darstellen.
- (2) ... die Schüler:innen Modelle, die sich im Nachhinein als falsch herausgestellt haben, korrigieren lässt, um den Prozess der Modellrevision zu verdeutlichen.
- (3) ... die Schüler:innen eigene Modelle basierend auf ihrem aktuellen Wissensstand entwickeln lässt und diese mit den Modellen während der COVID-19-Pandemie vergleicht.

Tabelle 12.4: Lösungsbogen der neu entwickelten Items des PCK_{SR}-bio

Item	Richtige Antwort	Begründung
#13	(2)	Falsche Antwortoptionen Antwortoption (1) beschreibt ein Unterrichtsvorgehen zur Förderung von Fachwissen bezüglich der Mendelschen Regeln. Auch in Antwortoption (3) liegt der Fokus nicht auf der Experimentierkompetenz, sondern in der Vermittlung des korrekten Umgangs mit der digitalen Experimentierumgebung. Vor dem Hintergrund der Zielsetzung der Unterrichtsstunde sind diese Vorgehen damit weniger geeignet, da die Experimentierkompetenz lediglich implizit adressiert wird (Matlen & Klahr, 2013; Vorholzer et al., 2020).
		Richtige Antwortoption Dahingehend eignet sich die eigenständige Durchführung der Experimente sowie insbesondere die Reflexion der Aussagekraft dieser in Antwortoption (2), um explizit die Experimentierkompetenz der Schüler:innen zu fördern (Baur et al., 2019; Krell & Krüger, 2022).
#14	(1)	Falsche Antwortoptionen Antwortoption (2) beschreibt ein Unterrichtsvorgehen, bei dem der Fokus auf die Vermittlung von Fachwissen in Bezug auf die Verwandtschaft von menschlichen Schädeln liegt und sich vor dem Hintergrund der Zielsetzung weniger eignet, da die Funktion von Hypothesen lediglich implizit adressiert wird (Matlen & Klahr, 2013; Vorholzer et al., 2020). Das in Antwortoption (3) beschriebene Vorgehen der eigenständigen Formulierung einer Fragestellung und darauf aufbauender Hypothesen kann bei Lernenden, welche über wenig Vorwissen verfügen, dazu führen, dass nicht trennscharfe Kriterien zum Vergleichen herangezogen werden (Wellnitz & Mayer, 2016), weshalb dies ebenfalls vermieden werden sollte.
		Richtige Antwortoption Das in Antwortoption (1) beschriebene Vorgehen eignet sich hingegen, da in der Rückschau überprüft wird, ob die Hypothesen für den vorgesehenen Zweck (Vergleich von Merkmalen) genutzt werden konnten, und den Schüler:innen die Funktion von Hypothesen zum naturwissenschaftlichen Erkenntnisgewinn damit deutlich gemacht wird (Großmann & Krüger, 2024).
#15	(1)	Falsche Antwortoptionen Durch das in Antwortoption (2) beschriebene Vorgehen wird das Ziel verfolgt, das Fachwissen der Schüler:innen in Bezug auf die Blutgruppen des Menschen zu fördern, was vor dem Hintergrund der Zielsetzung der Unterrichtsstunde nicht sinnvoll ist (Göhner & Krell, 2022; Knuuttila, 2011). Auch das in Antwortoption (3) beschriebene Vorgehen ist ungeeignet, da die Lehrkraft herausstellen sollte, dass Modelle nicht nur die Funktion haben, biologische Phänomene zu visualisieren, sondern diese auch zu erklären (Upmeier zu Belzen et al., 2019).
		Richtige Antwortoption Vor dem Hintergrund der Zielsetzung der Stunde eignet sich deshalb insbesondere das in Antwortoption (1) beschriebenen Vorgehen, da den Schüler:innen deutlich gemacht wird, warum verschiedene Modelle einen Sachverhalt unterschiedlich wiedergeben und wie sie auf diese Weise als Werkzeug naturwissenschaftlicher Erkenntnisgewinnung genutzt werden können (Göhner & Krell, 2022; Knuuttila, 2011).

Tabelle 12.4: Lösungsbogen der neu entwickelten Items des PCK_{SR}-bio (Fortsetzung)

Item	Richtige Antwort	Begründung
#16	(3)	Falsche Antwortoptionen Die in Antwortoption (1) beschriebene Vorstellung über das systematische Wiederholen von Experimenten sowie die in Antwortoption (2) beschriebene Vorstellung über die Variation einer Variablen setzen eine ausgeprägte Experimentierkompetenz voraus, weshalb sie eher nicht zu erwarten sind (Wellnitz & Mayer, 2016).
		Richtige Antwortoption Da die Schüler:innen bisher über kaum ausgeprägte Experimentierkompetenz verfügen, ist davon auszugehen, dass sie Experimente, wie in Antwortoption (3) beschrieben, als Möglichkeit ansehen, um ein eindrucksvolles Ergebnis zu erzielen, weshalb keine systematische Kontrolle der Variablen erfolgt (Kranz et al., 2023; Wellnitz & Mayer, 2016).
#17	(2)	Falsche Antwortoptionen Die in Antwortoption (1) beschriebene Vorstellung über die Planung eines hypothesengeleiteten Experiments erfordert ein elaborierteres Verständnis über die Funktion von Hypothesen zum naturwissenschaftlichen Erkenntnisgewinn und ist deshalb von der beschriebenen Lerngruppe eher nicht zu erwarten (Wellnitz & Mayer, 2016). Ebenfalls ist es von Schüler:innen ohne Wissen über die Funktion einer Hypothese eher nicht zu erwarten, dass sie wie in Antwortoption (3) beschrieben Kenntnisse darüber besitzen, wie Modelle basierend auf falsifizierter Hypothesen geändert werden können (Upmeier zu Belzen et al., 2019).
		Richtige Antwortoption Da die Schüler:innen die Funktion einer Hypothese noch nicht kennen, ist davon auszugehen, dass sie die in Antwortoption (2) beschriebene Vorstellung aufweisen und davon ausgehen, dass Hypothesen als Belege dienen (Kattmann, 2015; Kranz et al., 2023).
#18	(2)	Falsche Antwortoptionen Von Schüler:innen mit kaum ausgeprägter Modellierkompetenz ist nicht zu erwarten, dass sie wie in Antwortoption (1) beschrieben eine Vorstellung darüber haben, dass Modellorganismen Vorteile im Vergleich zu Menschen oder anderen Tierarten aufweisen (zum Beispiel eine kürzere Reproduktionszeit). Auch ist nicht zu erwarten, dass sie wie in Antwortoption (3) beschrieben eine Vorstellung darüber haben, dass sie dazu genutzt werden können, den systematischen Zusammenhang der Wirkung unterschiedlicher Medikamente auf die Gesundheit des Modellorganismus zu vergleichen (Upmeier zu Belzen et al., 2019).
		Richtige Antwortoption Daher ist von der beschriebenen Lerngruppe davon auszugehen, dass sie Modellorganismen wie in Antwortoption (2) beschrieben nur als alternative Darstellung für den menschlichen Organismus ansehen (Upmeier zu Belzen et al., 2019).

Tabelle 12.4: Lösungsbogen der neu entwickelten Items des PCK_{SR}-bio (Fortsetzung)

Item	Richtige Antwort	Begründung
#19	(3)	Falsche Antwortoptionen Damit die in Antwortoption (1) beschriebene Vorstellung über die hypothesengestützte systematische Entwicklung eines Experiments basierend auf den eigenen Vermutungen oder die in Antwortoption (2) beschriebene Generalisierbarkeit von Regeln entwickelt wird, ist es erforderlich, eine unbekannte Regel durch hypothesengestütztes Experimentieren erarbeiten zu lassen (Baur et al., 2019). Dies ist in dem beschriebenen Unterrichtsvorgehen nicht gegeben, da den Schüler:innen die Mendelschen Regeln bereits bekannt sind und ihr Wissen durch die Simulation lediglich bestätigt wird.
		Richtige Antwortoption Durch das beschriebene Unterrichtsvorgehen ist deshalb davon auszugehen, dass die Schüler:innen die in Antwortoption (1) beschriebene Vorstellung, dass Experimente nur zur Bestätigung des eigenen Wissens dienen, entwickeln (Kranz et al., 2023; Wellnitz & Mayer, 2016).
#20	(2)	Falsche Antwortoptionen Um die in Antwortoption (1) beschriebene Vorstellung, dass Hypothesen potenziell falsifizierbar sind oder die in Antwortoption (3) beschriebene Vorstellung, dass Experimente entwickelt werden, um zuvor formulierte Hypothesen zu überprüfen, zu entwickeln, ist es notwendig, Experimente auf Grundlage von zuvor formulierten Hypothesen zu entwickeln und die Passung der durchgeführten Experimente zur Überprüfung dieser zu reflektieren (Baur et al., 2019). Dies ist in dem beschriebenen Unterrichtsvorgehen nicht gegeben, da das Experiment bereits vorbereitet ist, die Hypothesen erst während der Durchführung aufgestellt werden und die Schüler:innen den Ausgang bereits kennen.
		Richtige Antwortoption Durch das beschriebene Unterrichtsvorgehen ist davon auszugehen, dass die Schüler:innen die in Antwortoption (2) beschriebene Vorstellung entwickeln, dass Hypothesen dazu dienen, das eigene Wissen zu überprüfen (Kranz et al., 2023; Wellnitz & Mayer, 2016).
#21	(2)	Falsche Antwortoptionen Die in Antwortoption (1) beschriebene Vorstellung, dass Modelle Fehler aufweisen, welche korrigiert werden müssen, oder die in Antwortoption (3) beschriebene Vorstellung, dass nur ein richtiges Modell existiert, wird durch das beschriebene Vorgehen bewusst vermieden, da das Unterrichtsvorgehen die sukzessive Überarbeitung eines Modells beinhaltet.
		Richtige Antwortoption Durch das beschriebene Unterrichtsvorgehen ist deshalb davon auszugehen, dass die Schüler:innen wie in Antwortoption (2) beschrieben lernen, wie Modelle durch das Erlangen neuer Erkenntnisse schrittweise revidiert und dabei hinsichtlich ihrer Komplexität und Darstellungsweise weiterentwickelt werden (Göhner & Krell, 2023; Upmeyer zu Belzen et al., 2019).

Tabelle 12.4: Lösungsbogen der neu entwickelten Items des PCK_{SR}-bio (Fortsetzung)

Item	Richtige Antwort	Begründung
#22	(3)	Falsche Antwortoptionen In dem in Antwortoption (1) beschriebenen Unterrichtsvorgehen erarbeiten die Schüler:innen zunächst Wissen zu einem Alltagsphänomen und prüfen dieses im Anschluss durch ein Experiment. Durch das in Antwortoption (2) beschriebene Vorgehen kann der Eindruck entstehen, dass es notwendig ist, Experimente ohne Vorwissen zu planen. Beide Unterrichtsvorgehen eignen sich damit nicht, um die Vorstellung aufzugreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis zu erweitern (Krell & Krüger, 2022).
		Richtige Antwortoption Im Gegensatz dazu stellt das in Antwortoption (3) beschriebene Vorgehen eine Möglichkeit dar, die Vorstellung der Lernenden um ein Verständnis über Experimente zum Erkenntnisgewinn zu erweitern (Baur et al., 2019).
#23	(3)	Falsche Antwortoptionen Durch das in Antwortoption (2) beschriebene Vorgehen kann die Vorstellung vermittelt werden, dass sich Hypothesen durch Experimente bestätigen lassen. Antwortoption (1) beschreibt hingegen ein Unterrichtsvorgehen, bei dem die Schüler:innen zwar die Vorstellung entwickeln können, dass Hypothesen potenziell falsifizierbar sind, jedoch keine Kenntnisse über die Rolle von Hypothesen als empirisch überprüfbare Voraussage eines Untersuchungsergebnisses erlangen (Großmann & Krüger, 2024; Krell & Krüger, 2022).
		Richtige Antwortoption Dies kann allerdings durch das in Antwortoption (3) beschriebene Vorgehen erreicht werden, da die Vorstellung der Schüler:innen durch die Durchführung hypothesengeleiteter Experimente zu einem unbekannten Phänomen aufgegriffen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis erweitert wird.
#24	(1)	Falsche Antwortoptionen In Antwortoption (2) wird die Behebung von Mängeln an dem betrachteten Modell beschrieben, in Antwortoption (3) werden die von den Schüler:innen entwickelten Modelle mit einem vermeintlich richtigen Modell verglichen. Beides entspricht gängigen Vorstellungen von Schüler:innen (Upmeyer zu Belzen et al., 2019), eignet sich jedoch nicht, um die genannte Vorstellung der Schülerin aufzugreifen und durch ein wissenschaftsmethodisch möglichst angemessenes Verständnis zu erweitern.
		Richtige Antwortoption Antwortoption (1) beschreibt hingegen ein Vorgehen, bei dem veranschaulicht werden kann, wie Modelle durch die systematische Überarbeitung und Ableitung entsprechender Handlungsstrategien genutzt werden konnten, um neue Erkenntnisse zu erlangen (Göhner & Krell, 2022; Knuuttila, 2011).

Literatur

- Baur, A., Emden, M. & Bewersdorff, A. (2019). Welche Unterrichtsprinzipien sollten für den Aufbau von Kompetenzen zum Experimentieren Beachtung finden? Eine Ableitung auf Basis multiperspektivisch begründeter Unterrichtsziele. *Zeitschrift für Didaktik der Biologie*, 23, 10–24. <https://doi.org/10.4119/zdb-1738>
- Göhner, M. & Krell, M. (2022). Preservice science teachers' strategies in scientific reasoning: The case of modeling. *Research in Science Education*, 52(2), 395–414. <https://doi.org/10.1007/s11165-020-09945-7>
- Göhner, M. & Krell, M. (2023). Umweltfaktoren beeinflussen das Libellenverhalten: Modelle als Werkzeuge zur Erkenntnisgewinnung nutzen. *Unterricht Biologie*, 47(481), 20–25.
- Großmann, L. & Krüger, D. (2024). Erkenntnisgewinnung (v)erklärt: Wie plant man hypothesengeleiteten Biologieunterricht? In B. Reinisch, D. Krüger & D. Mahler (Hrsg.), *Biologiedidaktische Nature of Science-Forschung: Zukunftsweisende Praxis* (S. 99–112). Springer. https://doi.org/10.1007/978-3-662-68409-2_9
- Kattmann, U. (2015). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht*. Aulis Verlag.
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42(2), 262–271. <https://doi.org/10.1016/j.shpsa.2010.11.034>
- Kranz, J., Baur, A. & Möller, A. (2023). Learners' challenges in understanding and performing experiments: A systematic review of the literature. *Studies in Science Education*, 59(2), 321–367. <https://doi.org/10.1080/03057267.2022.2138151>
- Krell, M. & Krüger, D. (2022). Erkenntnisgewinnung im Biologieunterricht: Welches Methodenwissen soll vermittelt werden? *MNU Journal*, 5, 376–382.
- Matlen, B. J. & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science*, 41(3), 621–634. <https://doi.org/10.1007/s11251-012-9248-z>
- Vorholzer, A., Aufschnaiter, C. von & Boone, W. J. (2020). Fostering upper secondary students' ability to engage in practices of scientific investigation: A comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education*, 50(1), 333–359. <https://doi.org/10.1007/s11165-018-9691-1>
- Wellnitz, N. & Mayer, J. (2016). Methoden der Erkenntnisgewinnung im Biologieunterricht. In A. Sandmann & P. Schmiemann (Hrsg.), *Biologie lernen und lehren: Bd. 1. Biologiedidaktische Forschung: Schwerpunkte und Forschungsstände* (S. 61–82). Logos Verlag Berlin.
- Upmeier zu Belzen, A., van Driel, J., & Krüger, D. (2019). Introducing a framework for modeling competence. In A. Upmeier zu Belzen, D. Krüger, & J. van Driel (Hrsg.), *Towards a competence-based view on models and modeling in science education: Bd. 12. Models and modeling in science education* (S. 3–19). Springer. https://doi.org/10.1007/978-3-030-30255-9_1

12.4 Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1³²

Im Folgenden befindet sich eine vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1.

Tabelle 12.5: Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1

Categories	Description	Empirical Evidence
Revision Note	Justifications of novices and experts that indicate a revision of the items. These can be explicit (i.e., addressed directly) or implicit (i.e., it can be seen that the item was not understood).	
Answer		
Right Answer	The item is answered correctly, only the final explanation and choice counts.	
Wrong Answer	The item is answered incorrectly.	
No Answer	No answer is given to the item.	
Explanation		
Intended	The explanation of the choice of an answer option is as intended (i.e., as intended when the item was created).	
Not Intended	The explanation of the choice of an answer option is not intended (i.e., not as intended when the item was created).	

³² Zusatzmaterial erschienen als *Supplementary Material* der zu Studie 1 zugehörigen Publikation.

Tabelle 12.5: Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1 (Fortsetzung)

Categories		Description	Empirical Evidence
Surface Thinking			
Item Stem	Student Characteristics	The (considered as inadequate) description of the learning group is used for the explanation. Only surface characteristics of the learning group are mentioned, such as the grade level, the expected competencies (not based on a subject didactic assessment, but without further explanations), ... The language used in students' statements mentioned in the item stem is used to explain the choice of answer options. Answer options in everyday language are chosen only for this reason, answer options in technical language are excluded only for this reason. Based on Kunter and Trautwein (2013).	<i>I think that the phrase 'target-oriented approach' sounds a little too high for eighth graders [and is therefore not used by them] - (Novice 10)</i>
	Topic	The topic of the classroom situation addressed in the stem is used to justify the chosen answer option without using PCK. Based on Kunter and Trautwein (2013).	<i>[The students know that they can change the CO₂ concentration by adding sparkling water, because] I simply believe that the students know the difference between tap water and sparkling water in the eighth grade [...]. Nowadays I think they know the difference, in times of SodaStream and things like that. - (Novice 3)</i>

Tabelle 12.5: Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1 (Fortsetzung)

Categories		Description	Empirical Evidence
Item Stem	Teaching Methods and Social Forms	The justification of the choice of answer options is made with reference to the teaching methods or social forms addressed in the classroom situation <u>without using PCK</u>. The answer option is selected because the teaching situation in the item is too easy/difficult for the students (<u>without using PCK</u>). Based on Kunter and Trautwein (2013).	<i>I didn't choose answer option (1) because I don't think that DNA is a good topic where you can start with students making their own models themselves. - (Novice 3)</i> <i>[I think the students are more likely to develop this idea] because they work in different groups and really have a lot of [different] results and by comparing all of them they can develop the rule themselves [...]. - (Novice 8)</i>
Personal Experience		The selection is based on personal experience. It includes both reported experiences from the perspectives of student and teacher. Only coded if own experiences are explicit and specifically mentioned.	<i>[...] I can only say that based on my personal experience, [...] I don't think that I would have understood anything about biology as a science in eighth grade, [...] that models are not only being developed for students, but also for people who really know about it. - (Novice 3)</i>

Tabelle 12.5: Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1 (Fortsetzung)

Categories	Description	Empirical Evidence
Intuition	The selection is explained intuitively. It is argued with the feeling (e.g., “I think this is right”), <u>without using PCK</u>. The participant cannot explain in more detail why the choice was made. The reason given is that an item/answer option is not understood. The rationale for selecting the response option is not from a constructivist perspective, but from a transmissive perspective on teaching and learning (Seidel & Reiss, 2014).	<i>I would take [this] answer option because [I think the students would say this because] I don't know. So somehow- I don't know, I think that's how they phrase it, I don't know - (Novice 4)</i> <i>[...] It's important that the students have an instruction, because [as a teacher] you should support them in experimenting. - (Novice 1)</i>
Wrong Interpretation	The necessary understanding exists but the item / answer option is interpreted differently than intended. As a result, the choice of answer option and / or the explanation is incorrect. Serves as an indication that the phrasing of the item is unclear, which is why it is often coded with “Revision Note”.	<i>So, they have already used models frequently. From this, you could conclude that the models they have used before were manufactured models [...]. - (Expert 1)</i>
Misconception	The reason for selection is wrong/missing knowledge (e.g., in a task on modelling skills, the person has no knowledge that these can be used to gain knowledge).	<i>[I didn't choose the answer option] "Biological models are created by scientists to conduct research with them." [...] because you are not researching models, but the actual things. Well, in science you don't do research with models, but they reflect what is being researched in science [...]. - (Novice 6)</i>

Tabelle 12.5: Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1 (Fortsetzung)

Categories	Description	Empirical Evidence
Pedagogical Knowledge	Explanations are not based on PCK, but on general pedagogical ideas or argued with pedagogical beliefs. The expected teaching time for a certain action is used to justify the choice of response options. Constructivist justifications are used without reference to PCK (e.g., “I choose this because they work there themselves and thus, they learn better”). Distinction based on Shulman (1986, 1987) and the empirical evidence by Krepf et al. (2018).	<i>I wouldn't conduct different experiments [in the lesson], because I think that would be a bit too much for one lesson. I think one experiment is enough [...]. - (Novice 1)</i> <i>Well, for example, boys somehow also often learn better if they touch something or make something themselves with their hands. I think that modelling is much more understandable if you rebuilt the models yourself - (Novice 7)</i>
Content Knowledge	Explanation of the selection is based on CK, no reference to PCK. Distinction based on Shulman (1986, 1987) and the empirical evidence by Krepf et al. (2018).	<i>I wouldn't choose this because models usually do not completely [represent the reality] [...]. A model does not show exactly how something is in reality, but only approximately so that you can imagine it better. - (Novice 8)</i>

Tabelle 12.5: Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1 (Fortsetzung)

Categories	Description	Empirical Evidence
No Relation to Scientific Reasoning	Explanations of the selection are based on other reasons for the decision (e.g., what teachers or students would answer). The choice is explained by the exclusion of other response options. The linguistic wording of the response option is used as an explanation for non-selection. Arguments are made with other competences that cannot be assigned to scientific reasoning (e.g., model criticism in the competence area modelling).	<i>[...] That's why I think that the model critique is much more conducive to promoting an appropriate scientific understanding in the students than the other answer. - (Novice 3)</i>
Deep Thinking		
Orientations to Teaching Science	Choice of an answer option is explained with purposes and goals of science teaching with reference to <i>pedagogical content knowledge</i>. Explanations regarding beliefs about (biology) teaching in general are not included. (Park & Chen, 2012; Park & Oliver, 2008) Explanations are based on competence models or other theoretical models (e.g., SDDS model).	<i>[I] would probably choose this answer option because that's most likely to encompass the meaningfulness of experiments. If you think about the SDDS model, and the competencies described there. For example, analysis of evidence. - (Expert 8)</i>

Tabelle 12.5: Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1 (Fortsetzung)

Categories	Description	Empirical Evidence
Knowledge of Students' Understanding	Choice of an answer option is based on knowledge about student perceptions (i.e., learning difficulties, motivation and differences in abilities, individual learning styles and interests) as well as knowledge about common conceptions and the prior knowledge of students (Park & Chen, 2012; Park & Oliver, 2008).	<i>[These two answer options are implausible] [...] because students tend to see models as visualizations rather than for research [...].</i> <i>[The other answer option is more suitable] because it addresses the reality of the students. This is because it is the function of models that they know and that they can reproduce. - (Expert 1)</i>
Knowledge of Science Curriculum	Choice of an answer option is explained with the use of the curriculum; argumentation on a horizontal or vertical level; explanation of the choice based on the curriculum as a whole. Core conceptions of the curriculum are applied to the item situation. Referencing directly to the curriculum (Park & Chen, 2012; Park & Oliver, 2008).	<i>[It reminds me of] the curriculum biology secondary level 1 [opens framework on the computer and searches in]. So, then we have "dealing with models". There we have the three big areas "Use", "Test" and "Change". And I think these answers could be assigned quite well. - (Expert 2)</i>

Tabelle 12.5: Vollständige Übersicht des verwendeten Kategoriensystems aus Studie 1 (Fortsetzung)

Categories		Description	Empirical Evidence
Knowledge of Instructional Strategies and Representations of Teaching Science	Topic-Specific Strategies	Choice of an answer option is explained based on teaching strategies related to the specific topic in the item situation. Thereby, the focus of the explanation is on teaching methods relating to a specific topic, not teaching strategies in general (Park & Oliver, 2008).	<i>But to think about this example with this specific subject content, I would not reflect on the goal-oriented procedure of hypothesis formation [for this topic]. - (Expert 2)</i>
	Subject-Specific Strategies	Choice of an answer option is explained based on teaching strategies relating to achieve the goals of biology teaching related to scientific methods and scientific reasoning. Includes also inquiry-based teaching strategies (Park & Chen, 2012; Park & Oliver, 2008).	<i>I would exclude this answer, because it is not about developing competencies about experimentation, but only about finding a result to the research question. - (Expert 8)</i>
Knowledge of Assessment of Science Learnings		Choice of an answer option is explained based on “knowledge of specific instruments, approaches, or activities” (p. 266) for the assessment of science learnings. (Park & Oliver, 2008)	No Coding

Literatur

- Krepf, M., Plöger, W., Scholl, D., & Seifert, A. (2018). Pedagogical content knowledge of experts and novices—what knowledge do they activate when analyzing science lessons? *Journal of Research in Science Teaching*, 55(1), 44–67. <https://doi.org/10.1002/tea.21410>
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts*. Schöningh. <http://www.utb-studi-e-book.de/9783838538952>
- 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. <https://doi.org/10.1002/tea.21022>
- Park, S. & Oliver, J. S. (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. <https://doi.org/10.1007/s11165-007-9049-6>
- Seidel, T. & Reiss, K. (2014). Lerngelegenheiten im Unterricht. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (6. Aufl., S. 253–275). Julius Beltz.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>

12.5 Publikationen im Rahmen der Dissertation

Lieberei, T., Welter, V. D. E. & Krell, M. (eingereicht). Pedagogical content knowledge (PCK) about scientific reasoning in different groups of science teachers and its relation to other variables. *Journal of Research in Science Teaching*.

2024

Lieberei, T., Großmann, L., Welter, V. D. E., Krüger, D. & Krell, M. (angenommen). Considering multiple sources of validity evidence can help to address challenges in the development of pedagogical content knowledge (PCK) multiple-choice items. *Research in Science Education*.

Lieberei, T., Welter, V. D. E. & Krell, M. (2024). Using the PCK map approach to obtain validity evidence based on response processes for a multiple-choice instrument. In G. Kaya, M. Sardag & I. Donmez (Hrsg.), *Proceedings book series-V of the ESERA 2023 conference: Connecting science education with cultural heritage* (S. 114–126). Nobel Bilimsel Eserler.

Lieberei, T., Dawborn-Gundlach, M., van Driel, J. & Krell, M. (2024). Preservice science teachers' knowledge about how to teach scientific reasoning. *Teaching Science*, 70(2), 30-43.

Sannert, R., Petermann, V., **Lieberei, T.,** Welter, V.D.E. & Krell, M. (2024). Erkenntnisgewinnungskompetenzen im naturwissenschaftlichen Unterricht fördern. Empfehlungen für die Gestaltung von Lehrkonzepten für die Aus- und Fortbildung von Lehrkräften. In Graulich, N., Arnold, J., Sorge, S., Kubsch, M., (Hrsg.), *Lehrkräftebildung von morgen. Beiträge der Naturwissenschaftsdidaktiken zur Förderung überfachlicher Kompetenzen* (S. 281-290). Waxmann.

2023

Lieberei, T., Welter, V. D. E., Großmann, L. & Krell, M. (2023). Findings from the expert-novice paradigm on differential response behavior among multiple-choice items of a pedagogical content knowledge test – implications for test development. *Frontiers in Psychology*, 14, 1240120. <https://doi.org/10.3389/fpsyg.2023.1240120>

Lieberei, T., Großmann, L., Welter, V. D. E., Krüger, D. & Krell, M. (2023). Entwicklung und Evaluation eines Instruments zur Erhebung fachdidaktischen Wissens angehender Biologielehrkräfte im Kompetenzbereich Erkenntnisgewinnung: PCKSR-bio. In M. Müller & S. Schaal (Hrsg.), *Herausforderung Zukunft: Internationale Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO*. (S. 491-493).

12.6 Vorträge und Poster im Rahmen der Dissertation

- Lieberei, T.* & Krell, M.** (2024). Fachdidaktisches Wissen im Kompetenzbereich Erkenntnisgewinnung – Vergleich unterschiedlicher Gruppen von Lehrkräften und die Beziehung zu anderen Variablen. Vortrag auf der WA-Sommertagung 2024 vom IPN, Juni 2024, Kiel.
- Lieberei, T.***, Großmann, L., Welter, V. D. E., Krüger, D., Krell, M. (2024). Considering Multiple Sources of Validity Evidence to Address Challenges in Developing PCK Multiple-Choice Items. Vortrag auf der 2024 Annual International Conference (NARST): NARST's 97th Annual International Conference, März 2024, Denver.
- Lieberei, T.***, Großmann, L., Welter, V.D.E., Krüger, D., Krell, M. (2023). Entwicklung und Evaluation eines Instruments zur Erhebung fachdidaktischen Wissens angehender Biologielehrkräfte im Kompetenzbereich Erkenntnisgewinnung: PCKSR-bio. Vortrag auf der Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, September 2023, Ludwigsburg.
- Lieberei, T.***, Welter, V.D.E., Krell, M. (2023). Using the PCK Map Approach to Obtain Validity Evidence Based on Response Processes for a Multiple-Choice Instrument. Vortrag auf der 15th Conference of the European Science Education Research Association (ESERA), August 2023, Nevşehir.
- Lieberei, T.* & Krell, M.** (2022). Development and Evaluation of a Multiple-Choice Instrument to Assess Pedagogical Content Knowledge on Scientific Reasoning. Gastvortrag an der Melbourne Graduate School of Education, University of Melbourne, November 2022, Melbourne.
- Lieberei, T.***, Großmann, L., Krüger, D., Krell, M. (2022). PCK about Scientific Reasoning: Response Processes of Novices and Experts. Vortrag auf der Swiss Doctoral School of Science Education 2022, Juni 2022, Seengen.
- Lieberei, T.***, Großmann, L., Krüger, D., Krell, M. (2022). Fachdidaktisches Wissen von angehenden Biologielehrkräften im Kompetenzbereich Erkenntnisgewinnung. Posterpräsentation auf der Internationalen Frühjahrsschule der Fachsektion Didaktik der Biologie im VBIO, März 2022, Bielefeld.
- Großmann, L.*, **Lieberei, T.**, Dawborn-Gundlach, M., van Driel, J., Krüger, D., Krell, M. (2022). Measuring Pre-service Biology Teachers' PCK about Scientific Reasoning – A case of validity. Vortrag auf der 2022 Annual International Conference (NARST): NARST's 95th Annual International Conference, März 2022, Vancouver (online).

* Präsentator:in

Danksagung

Abschließend möchte ich all denjenigen Personen danken, die mich während meiner aufregenden aber anstrengenden Promotionszeit begleitet und auf diese Weise die Fertigstellung dieser Dissertation möglich gemacht haben.

Besonderer Dank gilt zunächst Prof. Dr. Moritz Krell, welcher mir die Möglichkeit gegeben hat, diese Dissertation anzufertigen und damit mein Interesse an der Professionsforschung geweckt hat. Seine Offenheit für meine Ideen, das entgegengebrachte Vertrauen, die Wertschätzung und Unterstützung sowie das hohe Maß an Flexibilität haben maßgeblich zu meiner beruflichen und persönlichen Weiterentwicklung beigetragen.

Auch möchte ich mich bei Prof. Dr. Dirk Krüger bedanken, welcher mir durch die produktive Zusammenarbeit sowie seinen konstruktiven Rückmeldungen wertvolle Impulse gegeben und auf diese Weise neue Ansätze und Perspektiven ermöglicht hat.

Einen großen Dank möchte ich ebenfalls an meine Mentor:innen richten, die mich jederzeit unterstützt und in den richtigen Momenten aufgefangen und motiviert haben, mir bei Fragen stets zur Seite standen sowie die Realität zu den notwendigen Zeitpunkten vor Augen geführt haben. Danke an Virginia Welter, die mir unermüdlich zur Seite stand und auch nach Verlassen des IPN jederzeit ein offenes Ohr für mich hatte. Danke an Finja Grospietsch für die fachliche und emotionale Unterstützung, die mir insbesondere in den stressigen Phasen den benötigten Halt gegeben hat. Danke an Stefan Sorge, der jederzeit für mich ansprechbar war und mich durch unsere Gespräche auf neue Ideen brachte.

Danke an Kim, Mary und Richard für die angenehme Zeit in unserem gemeinsamen Büro, die Unterstützung sowie den Zusammenhalt unter uns Doktorand:innen. Ebenfalls möchte ich Frauke und Till danken, die mir bei Fragen in und außerhalb unserer Jours fixes jederzeit mit gutem Rat zur Seite standen.

Darüber hinaus danke ich allen weiteren Kolleg:innen, die mich während meiner Zeit am IPN begleitet und mich durch die fachlichen Diskussionen auf neue Ideen gebracht, aber ebenso durch den regelmäßigen informellen Austausch sowie den gemeinsamen Mensabesuchen für einen angenehmen Ausgleich gesorgt haben: Anja, Berrit, Birgit, Burkhard, Charlotte, Caro, Christina, Deni, Doro, Ernestine, Hannah, Hanno, Isi, Katharina, Kathryn, Kirsten, Kristin, Marc, Maren, Jana, Johannes, Sandra, Thomas, Uli sowie natürlich Frau Prof. Dr. Ute Harms als Leiterin der Abteilung.

My heartfelt thanks to Jan, Merryn and all colleagues from the MGSE, who welcomed me with open arms and made my stay in Melbourne the most wonderful time during my PhD.

Danke an unsere Hilfskräfte Alexa, Andreas, Jule, Kerstin, Naemi und Phyllis für die regelmäßige und verlässliche Unterstützung, der Übernahme diverser kleiner und großer Aufgaben sowie der damit einhergehenden Entlastung im Arbeitsalltag.

Danke an meine Familie und Freunde, auf die ich mich jederzeit verlassen konnte. Danke an meine Eltern und Großeltern für die bedingungslose Unterstützung. Danke an Benjamin, Celia, Nicola und Vanessa, die mich seit meiner Schulzeit begleiten. Danke an Daniel und Mario, die mich jederzeit mit offenen Armen bei sich empfangen und für einen angemessenen Ausgleich an den Wochenenden gesorgt haben.

Zuletzt gilt mein Dank allen Studierenden, Lehrkräften und Wissenschaftler:innen, die diese Dissertation durch ihre Teilnahme an den Studien erst möglich gemacht haben.

Erklärung

Hiermit erkläre ich, dass die vorliegende Dissertation – abgesehen von der Beratung durch meine Betreuer – nach Inhalt und Form meine eigene Arbeit ist und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden. Die Dissertation ist unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden. Sie wurde weder ganz noch zum Teil einer anderen Stelle im Rahmen eines Promotionsverfahrens vorgelegt. Ein Teil dieser Dissertation wurde bereits publiziert bzw. ist zur Publikation eingereicht (Kapitel 4 bis 8, siehe auch Tabelle 3.1). Ein akademischer Grad ist mir nicht entzogen worden.

Köln, den

Tobias Lieberei