

Sebastian Nickel

Repräsentationskompetenzen im Fach Chemie

Eine empirische Untersuchung
zur Dimensionalität des Konstrukts
und zu dessen Zusammenhang mit
Fachwissen und räumlichen Fähigkeiten



λογος

Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf und Mathias Ropohl

Diese Reihe im Logos Verlag Berlin lädt Forscherinnen und Forscher ein, ihre neuen wissenschaftlichen Studien zum Physik- und Chemielernen im Kontext einer Vielzahl von bereits erschienenen Arbeiten zu quantitativen und qualitativen empirischen Untersuchungen sowie evaluativ begleiteten Konzeptionsentwicklungen zu veröffentlichen. Die in den bisherigen Studien erfassten Themen und Inhalte spiegeln das breite Spektrum der Einflussfaktoren wider, die in den Lehr- und Lernprozessen in Schule und Hochschule wirksam sind.

Die Herausgeber hoffen, mit der Förderung von Publikationen, die sich mit dem Physik- und Chemielernen befassen, einen Beitrag zur weiteren Stabilisierung der physik- und chemiedidaktischen Forschung und zur Verbesserung eines an den Ergebnissen fachdidaktischer Forschung orientierten Unterrichts in den beiden Fächern zu leisten.

Martin Hopf und Mathias Ropohl

Studien zum Physik- und Chemielernen

Band 396

Sebastian Nickel

Repräsentationskompetenzen im Fach Chemie

Eine empirische Untersuchung zur Dimensionalität
des Konstrukts und zu dessen Zusammenhang mit
Fachwissen und räumlichen Fähigkeiten

Logos Verlag Berlin



Studien zum Physik- und Chemielernen

Martin Hopf und Mathias Ropohl [Hrsg.]

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.



Dieses Werk ist lizenziert unter der Creative Commons Lizenz CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Die Bedingungen der Creative-Commons-Lizenz gelten nur für Originalmaterial. Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit Quellenangabe) wie z. B. Schaubilder, Abbildungen, Fotos und Textauszüge erfordert ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.

Logos Verlag Berlin GmbH 2026

ISBN 978-3-8325-6074-4

ISSN 1614-8967

DOI 10.30819/6074

Logos Verlag Berlin GmbH
Georg-Knorr-Str. 4, Geb. 10
D-12681 Berlin

Tel.: +49 (0)30 / 42 85 10 90

Fax: +49 (0)30 / 42 85 10 92

<https://www.logos-verlag.de>

Repräsentationskompetenzen im Fach Chemie

Eine empirische Untersuchung zur Dimensionalität des
Konstrukts und zu dessen Zusammenhang mit Fachwissen
und räumlichen Fähigkeiten

Dissertation

zur Erlangung des akademischen Grades
Doktor der Philosophie
- *Dr. phil.* -

Philosophische Fakultät und Fachbereich Theologie,
Lehrstuhl für Didaktik der Chemie
der
Friedrich-Alexander-Universität Erlangen-Nürnberg

vorgelegt von
Sebastian Nickel,
geboren am 05.07.1996 in Siegen.

Nürnberg, 2025

Die vorliegende Arbeit wurde im Zeitraum von Mai 2022 bis August 2025 im Arbeitskreis von Prof. Dr. Sebastian Habig an der Philosophischen Fakultät (Didaktik der Chemie) der Friedrich-Alexander-Universität Erlangen-Nürnberg zum Erlangen eines Doktors der Philosophie verfasst.

Als Dissertation genehmigt
von der Philosophischen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 14. November 2025

Prüfungskommission: Prof. Dr. Sebastian Habig
Prof. Dr. Claudia von Aufschnaiter
Prof. Dr. Johannes Hasselhorn

*Für meine Schwester
Ann-Christin*

Danksagung

Die vergangenen drei Jahre, geprägt durch die Anfertigung dieser Dissertation, waren eine spannende und interessante Zeit mit großartigen Erlebnissen und noch großartigeren Menschen. Ihnen allen spreche ich an dieser Stelle ein herzliches „Dankeschön“ aus, denn ohne sie hätte ich das nicht geschafft.

Zunächst wende ich mich an meinen Doktorvater Prof. Dr. Sebastian Habig. 2016 durfte ich als studentische Hilfskraft bei dir anfangen, 2020 meine Masterarbeit unter deiner Betreuung schreiben und 2022 schließlich meine Promotion bei dir beginnen. Für dein Vertrauen in mich, die stets konstruktiven und inspirierenden Gespräche sowie dein offenes Ohr zu jeder Zeit danke ich dir von Herzen. Auch wenn ich zunächst gezögert habe, meinen Lebensmittelpunkt zumindest zeitweise nach Nürnberg zu verlegen, kann ich heute mit voller Überzeugung sagen: Es war die richtige Entscheidung, und ich bereue keinen Moment. Vielen Dank für alles – und für die Zusammenarbeit mit einem wirklich großartigen Chef!

Ein großes Dankeschön gebührt auch Prof. Dr. Claudia von Aufschnaiter: Deine unterstützende und offene Art während der ESERA Summer School 2023 ist mir sehr positiv in Erinnerung. Daher freue ich mich außerordentlich über deine Bereitschaft, die Zweitbegutachtung meiner Arbeit zu übernehmen. Vielen Dank, Claudia!

Prof. Dr. Johannes Hasselhorn danke ich herzlich für die inspirierenden Anregungen und Diskussionen im Rahmen des Forschungskolloquiums. Ich freue mich sehr, dass du dich auch als Prüfer in meiner mündlichen Prüfung engagierst. Deine Unterstützung war stets besonders wertvoll!

Meinen großartigen Kolleg:innen Xenia Schäfer, Markus Stocker, David Johannes Hauck, Tobias Petermann, Hella Rieß, Lisa Eggers und Christian Andre danke ich für die liebevolle und humorvolle Atmosphäre, die moralische Unterstützung und die anregenden Diskussionen – ich bin sehr froh, dass ihr mich alle auf diesem Weg begleitet habt. Besonders hebe ich Steffen Brockmüller hervor: Steffen, du hast meine Arbeit von Beginn an mitbetreut, mich in allen Phasen mit wertvollen Anregungen unterstützt und so maßgeblich zum Erfolg beigetragen. Vielen Dank!

Auch den wissenschaftlichen Hilfskräften, die mich in dieser Zeit bei verschiedenen Projekten und natürlich bei dieser Arbeit unterstützt haben, danke ich sehr. Johannes, Pauline, Maria, Mara, Lily, Nico, Marianna, Sophie, Gertrud, Hannah Luna, Hannah, Nessi: Ohne euch würde ich wahrscheinlich immer noch die ca. 2.500 Testhefte digitalisieren und die 24 Stunden Videomaterial transkribieren. Eure Unterstützung war fantastisch!

Ein großes Dankeschön gilt Dr. Tanja Retzer, Dr. Olaf Brummel, Dr. Victoria Teller, Dr. Vanessa Fischer sowie Dr. Carolin Eitemüller, die mich in ihre Vorkurse eingeplant und damit die Durchführung meiner Datenerhebung ermöglicht haben. In diesem Zusammenhang gilt mein Dank auch allen Studierenden, die an meinen Datenerhebungen teilgenommen haben – ich weiß die Mühe und die investierte Zeit sehr zu schätzen.

Letztlich habe ich meiner Familie von Herzen zu danken, die mir in den vergangenen Jahren stets Rückhalt gegeben hat. Dies gilt besonders für meine Eltern, Ralf und Petra! Ihr seid immer an meiner Seite gewesen und habt mich in all meinen Entscheidungen unterstützt. Dafür bin ich euch unendlich dankbar – bessere Eltern kann man sich nicht wünschen.

Ebenso danke ich meinem Verlobten Simon: Du hast mich von Anfang an in meinem Wunsch zu promovieren bestärkt und dafür die Herausforderungen einer Fernbeziehung mitgetragen. In den vergangenen acht Jahren konnten wir viele Erinnerungen – schöne wie schwierige – miteinander teilen, und ich bin unendlich froh, dass wir das gemeinsam durchgestanden haben.

Herzlichen Dank euch allen!

Inhaltsverzeichnis

Zusammenfassung	1
Abstract	5
1 Einleitung und Problemorientierung	9
2 Repräsentationen und ihre Bedeutung für das Lernen	11
2.1 Klassifikation von Repräsentationen.....	11
2.2 Repräsentationen in der Chemie und anderen Naturwissenschaften	18
2.2.1 Repräsentation und Modell – Was unterscheidet sie?	21
2.2.2 Lernen mit multiplen externen Repräsentationen.....	22
2.2.3 Lernen mit und über Repräsentationen – Ein Dilemma.....	27
2.2.4 Repräsentationskompetenzen	30
3 Chemie und das räumliche Denken	37
3.1 Räumliche Fähigkeiten: Eine definitorische Annäherung	38
3.2 Evolution der Theorien zur menschlichen Intelligenz und Einordnung der räumlichen Fähigkeiten.....	39
3.3 Die Cattell-Horn-Carroll Theorie mit Verortung und Beschreibung der räumlichen Fähigkeiten	41
3.3.1 Aktuelle Forschungen und Entwicklung des Cattell-Horn-Carroll Modells	44
3.3.2 Statische und dynamische räumliche Fähigkeiten	45
3.3.3 Kritische Betrachtung des Cattell-Horn-Carroll Modells	46
3.3.4 Geschlechtsunterschiede bei räumlichen Fähigkeiten.....	47
3.4 Die Rolle der räumlichen Fähigkeiten in der Chemie und anderen Naturwissenschaften – Ein Einblick in den aktuellen Forschungsstand	48
3.5 Trainierbarkeit räumlicher Fähigkeiten.....	50
4 Zielsetzungen und Fragestellungen	53

5	Forschungsanliegen I – Entwicklung und Evaluation des CRI:TIC: Untersuchung der Struktur der Repräsentationskompetenzen und des Zusammenhangs mit dem Fachwissen in der Chemie	61
5.1	Design.....	61
5.2	Messinstrumente	64
5.2.1	Gütekriterien zur Entwicklung und Evaluation von Messinstrumenten	64
5.2.2	Messung der Repräsentationskompetenzen	68
5.2.3	Messung des chemischen Fachwissens	81
5.3	Testtheorien	83
5.3.1	Die Klassische Testtheorie (KTT)	83
5.3.2	Die Item-Response-Theorie (IRT).....	86
5.4	Psychometrische und statistische Methoden	90
5.5	Datenerhebung und Stichprobe	93
5.6	Ergebnisse	96
5.6.1	Evaluation des CRI:TIC	96
5.6.2	Untersuchung der Trennbarkeit der lower-level skills der Repräsentationskompetenzen	103
5.6.3	Evaluation der Fachwissensitems	105
5.6.4	Untersuchung der Trennbarkeit von Fachwissen und Repräsentationskompetenzen	109
5.6.5	Bewertung der Gütekriterien	110
5.7	Zusammenfassung und Diskussion der Ergebnisse	114
6	Forschungsanliegen II – Einfluss der räumlichen Fähigkeiten beim Bearbeiten repräsentationsbasierter Chemieaufgaben	121
6.1	Design.....	121
6.2	Messinstrumente	123
6.2.1	Messung der Repräsentationskompetenzen und des Fachwissens	123
6.2.2	Messung der räumlichen Fähigkeiten.....	124
6.3	Psychometrische und statistische Methoden	134
6.4	Datenerhebung und Stichprobe	138

6.5	Ergebnisse.....	141
6.5.1	Forschungsanliegen I – Wiederholung der Analysen zur Prüfung der Replizierbarkeit	141
6.5.2	Forschungsanliegen II – Untersuchung des Zusammenhangs zwischen Repräsentationskompetenzen, Fachwissen und räumlichen Faktoren	150
6.6	Zusammenfassung und Diskussion der Ergebnisse	158
7	Forschungsanliegen III – Hindernisse beim Bearbeiten repräsentationsbasierter Chemieaufgaben	169
7.1	Design.....	169
7.2	Methodisches Vorgehen und Datenauswertung	170
7.2.1	Hindernisse im Bearbeitungsprozess	170
7.2.2	Entwicklung des Materials	172
7.2.3	Lautes Denken.....	178
7.2.4	Qualitative Inhaltsanalyse nach Mayring.....	180
7.2.5	Umsetzung der qualitativ orientierten kategoriengeleiteten Textanalyse in dieser Studie (qualitative Inhaltsanalyse).....	182
7.3	Datenerhebung und Stichprobe	187
7.4	Ergebnisse.....	189
7.4.1	Hindernisse bei der <i>Interpretation</i>	198
7.4.2	Hindernisse bei der <i>Translation</i>	202
7.4.3	Hindernisse bei der <i>Konstruktion</i>	209
7.5	Zusammenfassung und Diskussion	213
8	Integration der Forschungsanliegen.....	221
8.1	Integrative Gesamtdiskussion und Limitationen	226
8.2	Implikationen für Praxis und Forschung	232
9	Verzeichnisse	237
9.1	Literaturverzeichnis	237
9.2	Abbildungsverzeichnis	267
9.3	Tabellenverzeichnis.....	272
9.4	Abkürzungsverzeichnis	276
10	Anhang.....	279

Zusammenfassung

Repräsentationen spielen in den Naturwissenschaften, besonders in der Chemie, eine zentrale Rolle, da sie eine Kommunikation über und eine Erklärung von Phänomenen auf einer nicht erfassbaren Ebene ermöglichen. Sie umfassen beispielsweise Graphen und Diagramme sowie, kennzeichnend für die Chemie, Repräsentationen von Atomen, Molekülen und Ionen, Reaktionsgleichungen und -mechanismen. In Lehr-Lern-Prozessen zeigt sich, dass Repräsentationen eine Herausforderung für Lernende darstellen. Zum einen zeigt sich, dass das Lernen *von* und das Lernen *über* Repräsentationen gleichzeitig stattfindet. Lernende müssen also unbekannte fachliche Inhalte von Repräsentationen lernen, ohne zu verstehen, wie diese Repräsentationen die Inhalte vermitteln. Dieses Problem wird als Repräsentationsdilemma bezeichnet. Um dieses zu überwinden, müssen Lernende Repräsentationskompetenzen entwickeln, wobei die *Interpretation*, *Translation* und *Konstruktion* (Kozma & Russell, 1997, 2005) grundlegende Fähigkeiten zum erfolgreichen Umgang mit Repräsentationen darstellen. Zum anderen sind Repräsentationen in der Chemie häufig sehr abstrakt und räumlich anspruchsvoll. Daher benötigen Lernende zusätzlich ausgeprägte räumliche Fähigkeiten, beispielsweise zwei- und dreidimensionale mentale Rotationsfähigkeiten.

Forschungsanliegen I – In Bezug auf die Repräsentationskompetenzen *Interpretation*, *Translation* und *Konstruktion* ist bisher nicht abschließend geklärt, inwieweit sich diese empirisch trennen lassen und sie ein zum Fachwissen distinktes Konstrukt darstellen. Die Beantwortung dieser Fragen ist für die Optimierung repräsentationsbasierter Lehr-Lern-Prozesse von essenzieller Bedeutung. Zur Untersuchung der Forschungsfragen wurde das *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC) entwickelt und mittels multidimensionaler Raschanalysen evaluiert. Dieses Instrument wurde von $N = 185$ Studierenden eines Chemievorkurses im Wintersemester 2023/24 bearbeitet. Die Ergebnisse deuten darauf hin, dass die *Interpretation*, *Translation* und *Konstruktion* stark interkorrelierte und, empirisch nicht trennbare Fähigkeiten darstellen, obwohl diese unterschiedliche Anforderungen mit sich bringen. Zusätzlich zum CRI:TIC wurde das deklarative Fachwissen erhoben und durch vergleichbare Methoden die Abgrenzung zu den Repräsentationskompetenzen geprüft. Hier zeigt sich eindeutig, dass die beiden Konstrukte im Rahmen dieser Studie empirisch trennbar sind.

Forschungsanliegen II – Zwar liegen bereits Evidenzen zum Zusammenhang einzelner räumlicher Faktoren mit der Leistung in der Chemie und der Bearbeitung repräsentationsbasierter Chemieaufgaben vor, diese Studien berücksichtigen jedoch nicht die Interkorrelation zwischen den räumlichen Faktoren. Neben den Repräsentationskompetenzen (CRI:TIC) und dem deklarativen Fachwissen wurden acht räumliche Faktoren erhoben und der Zusammenhang der Konstrukte mittels einer Pfadanalyse untersucht. Insgesamt nahmen an der Studie $N = 494$ Studierende dreier Chemie-Vorkurse im Wintersemester 2024/25 teil. Die Ergebnisse verdeutlichen, dass das Fachwissen den deutlichsten und höchsten Zusammenhang mit den repräsentationsbasierten Chemieaufgaben aufweist. Daneben zeigen sich auch die zwei- und dreidimensionale mentale Rotationsfähigkeit sowie die mentale Manipulationsfähigkeit als relevant für die Bearbeitung dieser Aufgaben. In einer differenzierten Betrachtung für die *Interpretation*, *Translation* und *Konstruktion* zeigt sich, dass sich die Ausprägung der relevanten Faktoren unterscheidet. Somit lässt sich darauf schließen, dass für die *Interpretation*, *Translation* und *Konstruktion* unterschiedliche kognitive Prozesse und Fähigkeiten relevant sind.

Forschungsanliegen III – Um die Ergebnisse der ersten beiden Forschungsanliegen zu vertiefen, wurden Hindernisse bei der Bearbeitung repräsentationsbasierter Chemieaufgaben untersucht. Dadurch sollten Rückschlüsse auf die kognitiven Prozesse und Fähigkeiten bei der *Interpretation*, *Translation* und *Konstruktion* möglich werden. Hierzu wurde eine qualitative Studie unter Anwendung der Methode des lauten Denkens mit $N = 25$ Lehramtsstudierenden des Fachs Chemie durchgeführt. Es zeigt sich, dass das Fachwissen für alle drei Fähigkeiten eine zentrale Rolle spielt, räumliche Fähigkeiten jedoch eher für die *Translation* und *Konstruktion* chemischer Repräsentationen relevant sind.

Das Projekt knüpft an bestehende Forschung zum Lernen und Arbeiten mit Repräsentationen in der Chemie an und trägt zur empirischen Fundierung der Struktur von Repräsentationskompetenzen bei. Zudem ermöglicht es eine differenzierte Abgrenzung dieses Konstrukts vom chemischen Fachwissen, indem die zugrunde liegenden Konstrukte klar und theoriegeleitet operationalisiert werden – ein Aspekt, der in bisherigen Studien häufig unzureichend berücksichtigt worden ist. Darüber hinaus stellt der CRI:TIC ein objektives, valides und reliables Instrument zur Erfassung von Repräsentationskompetenzen bereit. Durch die Identifikation zentraler räumlicher Faktoren und Hindernisse bei der Bearbeitung repräsentationsbasierter Chemieaufgaben leistet das

Projekt einen wichtigen Beitrag zur Erweiterung des aktuellen Forschungsstandards.

Daraus ergeben sich sowohl praxisrelevante Implikationen für die Gestaltung von Lehr-Lern-Prozessen als auch richtungsweisende Perspektiven für künftige Forschungsanliegen. Hinsichtlich der Praxis wird deutlich, dass Repräsentationskompetenzen neben dem chemischen Fachwissen als zentrales Element in den Lehr-Lern-Prozess integriert werden müssen. Räumliche Fähigkeiten spielen dabei eine ebenso wichtige Rolle. Insbesondere räumlich anspruchsvolle Aufgaben sollten gezielt durch das Training räumlicher Faktoren sowie durch zusätzliches Material (z. B. haptische oder virtuelle Modelle) ergänzt werden, um Lernende im Bearbeitungsprozess optimal zu unterstützen. Aus forschungsbezogener Perspektive stellen sich weiterführende Fragen nach der hierarchischen Struktur der Repräsentationskompetenzen, insbesondere im Hinblick auf die Fähigkeiten *Interpretation*, *Translation* und *Konstruktion*. Ebenso gilt es zu untersuchen, inwiefern ein gezieltes Training zwei- und dreidimensionaler mentaler Rotations- sowie Manipulationsfähigkeiten zu einer signifikanten Leistungssteigerung bei der Bearbeitung repräsentationsbasierter Chemieaufgaben führen kann.

Abstract

Representations play a key role in sciences, particularly in chemistry, as they enable communication about and explanation of phenomena on a non-observable level. These include, for example, graphs, diagrams, and more specifically in chemistry, representations of atoms, molecules, and ions, as well as chemical equations and reaction mechanisms. However, in teaching and learning processes, representations often pose a challenge for learners. One key issue is that learning *with* and *about* representations occurs simultaneously – learners are expected to learn content knowledge they do not understand from representations without understanding how these representations convey that content. This problem is referred to as the representation dilemma. To overcome this dilemma, learners must develop representational competence, with *interpretation*, *translation*, and *construction* (Kozma & Russell, 1997, 2005) being basic skills in dealing with scientific representations. Moreover, representations in chemistry are often highly abstract and spatially demanding. As such, learners also require well-developed spatial abilities, for instance the ability for two- and three-dimensional mental rotation.

Research Objective I – With regard to the representational skills *interpretation*, *translation*, and *construction*, it remains unclear to what extent these can be empirically distinguished from one another and whether they represent a construct that is distinct from content knowledge. Answering these questions is essential for enhancing teaching and learning processes involving chemical representations. To investigate these questions, a test instrument – the *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC) – was developed and evaluated using multidimensional Rasch analyses. The test was administered to $N = 185$ students enrolled in a preparatory chemistry course during the wintersemester 2023/24. The results suggest that *interpretation*, *translation*, and *construction* are highly interrelated and cannot be clearly distinguished from one another empirically, although they require different abilities. In addition to the CRI:TIC, content knowledge was assessed using comparable methods, in order to examine its distinction from representational competence. The findings clearly indicate that the two constructs – representational competence and content knowledge – are empirically separable within the context of this study. The development, validation, and the English version of the CRI:TIC are detailed in Nickel et al. (2025).

Research Objective II – Although previous studies have found evidence for relation between spatial factors and performance on representation-based chemistry tasks, they often fail to account the intercorrelation between different spatial factors. In this study, in addition to assessing representational competence (via CRI:TIC) and content knowledge, eight spatial factors are measured and examined using path analysis. A total of $N = 494$ students participated in the study during preparatory chemistry courses in winter 2024/25. The results show that content knowledge has the strongest impact on the representation-based tasks. In addition, two- and three-dimensional mental rotation abilities, as well as mental manipulation ability, emerged as relevant for working with representation-based chemistry tasks. A differentiated analysis of the individual skills – *interpretation*, *translation*, and *construction* – reveals variations in the contribution of spatial factors. These findings suggest that different cognitive processes and abilities are involved in each of the three skills.

Research Objective III – To deepen the insights gained from the first two research objectives, this study investigates obstacles encountered by learners when solving representation-based chemistry tasks. This allows for further conclusions about the cognitive processes and abilities involved in *interpretation*, *translation*, and *construction*. For this purpose, a qualitative think-aloud study was conducted with $N = 25$ preservice chemistry teachers. The findings indicate that content knowledge plays a central role in all three skills. However, spatial abilities are particularly relevant for the *translation* and *construction* of chemical representations.

The project builds upon existing research on learning and working with representations in chemistry and contributes to the empirical foundation of the structure of representational competence. Furthermore, it enables a nuanced delineation of this construct from chemical content knowledge by operationalising the underlying constructs in a clear and theory-based manner – an aspect often insufficiently addressed in previous studies. In addition, the project provides an objective, valid, and reliable instrument for assessing representational competence in the form of the CRI:TIC. By identifying key spatial factors and difficulties involved in solving representation-based chemistry tasks, this project makes a significant contribution to advancing current research. This yields both practice-oriented implications for the design of teaching and

learning processes and perspectives for future research. With regard to educational practice, it becomes evident that representational competence must be integrated as a central element alongside chemical content knowledge in teaching and learning. Spatial abilities also play a crucial role in this context. Therefore, tasks that are spatially demanding should be supplemented through targeted training of spatial factors as well as with additional resources (e.g., concrete or virtual models) in order to support learners in the problem-solving process. From a research perspective, further questions arise concerning the hierarchical structure of representational competence, particularly regarding the skills of *interpretation*, *translation*, and *construction*. It is also essential to investigate the extent to which targeted training of two- and three-dimensional mental rotation and manipulation abilities can lead to a significant improvement in performance when working on representation-based chemistry tasks.

1

Einleitung und Problemorientierung

Das Unterrichtsfach Chemie stößt nach wie vor bei vielen Lernenden auf wenig Beliebtheit, ähnlich wie die anderen „harten“ MINT-Fächer Mathematik und Physik – und das sowohl im nationalen (z. B. Jansen et al., 2019; Maaß et al., 2022; SINUS, 2024) als auch im internationalen Kontext (z. B. Mansour, 2024). In verschiedenen Studien wurden von Lernenden vielfältige Gründe hierfür genannt. So wird Chemie häufig als ein komplexes und herausforderndes Fach beschrieben, dessen Inhalte als abstrakt, wenig lebensweltbezogen und mitunter als langweilig gelten. Zudem wurde ein Mangel an Vorstellungsvermögen in Bezug auf chemische Konzepte genannt (Cardellini, 2012; Mansour, 2024). Chemie befasst sich mit der Erklärung beobachtbarer Phänomene auf einer submikroskopischen Ebene, die der direkten Wahrnehmung nicht zugänglich ist und daher durch externe Repräsentationen vermittelt werden muss. Wu und Shah (2004) bezeichnen Chemie daher als „visual science“ (S. 465), was Bucat und Mocerino (2009) treffend ergänzen mit der Charakterisierung der submikroskopischen Ebene als „[...] unobservable world, accessible only by imagination“ (S. 12). Aus dieser Perspektive erscheinen die von Lernenden genannten Schwierigkeiten nachvollziehbar und erklärbar.

Die submikroskopische Ebene wird durch Repräsentationen epistemologisch zugänglich gemacht, weshalb die Chemie durch eine hohe Dichte und Vielfalt unterschiedlicher Repräsentationen geprägt ist. Lernende sind daher darauf angewiesen, spezifische Fähigkeiten im Umgang mit diesen zu entwickeln – sogenannte Repräsentationskompetenzen. Diese befähigen sie, Informationen aus chemischen Repräsentationen zu extrahieren und für die Entwicklung chemischer Konzepte zu nutzen (z. B. Kozma & Russell, 1997, 2005; Rau, 2017c). Repräsentationskompetenzen und chemisches Fachwissen stehen in einer engen, wechselseitigen Beziehung und sind in der Praxis kaum voneinander zu trennen. Diese enge Verbindung äußert sich unter anderem im sogenannten Repräsentationsdilemma (Rau, 2018). Die empirische Trennbarkeit der beiden Konstrukte ist bislang nicht abschließend geklärt. Ebenso besteht

Forschungsbedarf bezüglich der Differenzierung spezifischer Fähigkeiten innerhalb der Repräsentationskompetenz. Diese offenen Fragen bilden den Schwerpunkt des ersten Forschungsanliegens.

Neben Repräsentationskompetenzen und Fachwissen haben sich räumliche Fähigkeiten als wichtige Prädiktoren für die Leistung in der Chemie und beim Arbeiten mit chemischen Repräsentationen herausgestellt (z. B. Buckley et al., 2018; Harle & Towns, 2011; Stieff et al., 2012). Einzelne räumliche Faktoren erwiesen sich in diesem Kontext als besonders relevant, beispielsweise die zwei- und dreidimensionale mentale Rotationsfähigkeit sowie die Fähigkeit zur mentalen Manipulation. Es mangelt jedoch an Studien, die räumliche Faktoren systematisch und umfassend erheben und deren Einfluss auf die Bearbeitung repräsentationsbasierter Chemieaufgaben unter Berücksichtigung des Fachwissens sowie möglicher Interkorrelationen untersuchen – dies stellt den Fokus in Forschungsanliegen II dar.

Forschungsanliegen III zielt darauf ab, vertiefte Einblicke in die Bearbeitung repräsentationsbasierter Chemieaufgaben zu gewinnen. Dies soll durch die Identifikation und Analyse der dabei auftretenden Hindernisse geschehen. Ziel ist es, zentrale kognitive Prozesse und Fähigkeiten herauszuarbeiten, die für eine erfolgreiche Bearbeitung solcher Aufgaben erforderlich sind.

2

Repräsentationen und ihre Bedeutung für das Lernen

Repräsentationen spielen in unserem Alltag eine maßgebliche Rolle. Sie ermöglichen es, Gedanken in Form von Sprache, Symbolen, Bildern oder Gesten zu externalisieren und eröffnen uns damit die Kommunikation mit anderen Menschen (Ainsworth, 2006; Krey & Schwanewedel, 2018). Besonders in den Naturwissenschaften sind Repräsentationen entscheidend. Sie dienen der Darstellung einer visuell nicht erfassbaren oder beobachtbaren Ebene, beispielsweise der Atome und Moleküle in der Chemie (Bucat & Mocerino, 2009; Chiu & Wu, 2009). Repräsentationen erlauben es uns somit, unsichtbare Phänomene anschaulich zu erklären (Chiu & Wu, 2009; Rapp, 2005).

In diesem Kapitel wird die Bedeutung des Lernens mit Repräsentationen in der Chemie dargestellt, damit einhergehende Anforderungen und Herausforderungen herausgearbeitet und ein Überblick über aktuelle Forschung in diesem Kontext gegeben. Dazu werden zunächst die Konzeptualisierung und Kategorisierung von Repräsentationen näher beleuchtet und grundlegende Theorien zum diesbezüglichen Lernen erläutert.

2.1 Klassifikation von Repräsentationen

Mit dem Gemälde *La trahison des images* (Der Verrat der Bilder, Abbildung 1), auf dem eine Pfeife mit dem Schriftzug „*Ceci n'est pas une pipe*“ („Dies ist keine Pfeife“) zu sehen ist, verdeutlicht der belgische Künstler René Magritte (1929) die Beziehung zwischen einem Objekt (der Pfeife), seiner ikonischen Darstellung und dem Wort „Pfeife“. Weder die abgebildete Pfeife noch das Wort „Pfeife“ lassen sich stopfen oder rauchen. Sie sind lediglich Repräsentationen des realen Gegenstands.



Abbildung 1: La trahison des images (Der Verrat der Bilder) mit dem Schriftzug „Ceci n’est pas une pipe“ („Dies ist keine Pfeife“) von René Magritte (1929).

Magritte (1929) veranschaulicht hier die Perspektive der Semiotik: Repräsentationen sind Zeichen, die für einen Gegenstand oder einen Prozess stehen (Schnotz et al., 2010). Sie sind durch einen spezifischen Code (verbal, ikonisch oder numerisch) sowie eine spezifische Modalität (sinnliche Wahrnehmung, vor allem auditiv, visuell oder haptisch) gekennzeichnet (Schanze & Girwidz, 2018). Absichtlich erzeugte Zeichen dienen der Kommunikation mit anderen Personen und der Wissensvermittlung: Personen externalisieren ihre Vorstellungen von einem Gegenstand oder Prozess, indem sie Zeichen beziehungsweise Repräsentationen produzieren (Schnotz et al., 2010).

Hier lässt sich eine grundlegende Unterscheidung zwischen *internen* und *externen* Repräsentationen treffen. Aus kognitionspsychologischer Sicht werden *interne* Repräsentationen als mentale Abbilder der Umwelt verstanden, die erzeugt werden, um diese gedanklich zu verarbeiten. Dabei beeinflussen bereits bestehende *interne* Repräsentationen diesen Prozess. Durch die Erzeugung von Zeichen können *interne* Repräsentationen in *externe* überführt werden; diese externen Repräsentationen wiederum dienen als Grundlage für die Konstruktion neuer *interner* Repräsentationen. Somit stehen *interne* und *externe* Repräsentationen in einer fortwährenden Wechselbeziehung zueinander (Krey & Schwanewedel, 2018; Schnotz et al., 2010; Zhang, 2001).

Externe Repräsentationen können weiter anhand ihrer Kodalität differenziert werden. Schnotz und Bannert (2003) teilen sie grob in *depiktionale* und *deskriptionale* Repräsentationen ein. Gemälde, Bilder und physische Modelle beispielsweise basieren auf ikonischen Zeichen und werden daher als *depiktionale* Repräsentationen bezeichnet. Geschriebene und gesprochene Texte sowie mathematische Formeln setzen sich hingegen aus Symbolen zusammen, die einen Gegenstand beschreiben und werden daher als *deskriptionale* Repräsentationen bezeichnet. Die verwendeten Symbole besitzen eine willkürliche Struktur, wodurch sie nicht direkt, sondern durch spezifische Konventionen mit dem repräsentierten Gegenstand verbunden sind (z. B. Wörter). Solche Symbole finden sich bei *depiktionalen* Repräsentationen nicht wieder (Krey & Schwanewedel, 2018; Schnotz & Bannert, 2003; Talanquer, 2022).

Auf dieser Grundlage können Repräsentationen weiter und feiner differenziert werden. Die *depiktionalen* Repräsentationen lassen sich in *visuell-graphische* Repräsentationen (realistische und logische Bilder) sowie *handelnd-operative* Repräsentationen (etwa Gesten, physische Modelle, Experimente) unterscheiden. Die *deskriptionalen* Repräsentationen können in *verbal-textliche* Repräsentationen (gesprochener und geschriebener Text) und *symbolisch-mathematische* Repräsentationen (etwa Gleichungen, Formeln) gegliedert werden (Tsui, 2003; Wu & Puntambekar, 2012).

Eine vergleichbare Klassifikation beschreibt Gilbert (2005) mit den fünf *Modes of Representations*: (1) *concrete* bzw. *material mode* (haptische, materielle Repräsentationen), (2) *verbal mode* (gesprochener und geschriebener Text), (3) *symbolic mode* (chemische Symbole, Formeln, Gleichungen usw.), (4) *visual mode* (Graphen, Diagramme, Animationen usw.) und (5) *gestural mode* (Gesten, Körperbewegungen). Diese Kategorisierung ist nicht gänzlich neu und wurde im Laufe der Jahre immer wieder adaptiert und weiter ausdifferenziert. Bereits Bruner (1966) unterschied zum Beispiel zwischen enaktiven, ikonischen, symbolischen Repräsentationen, wobei die enaktive Repräsentation vollzogene Handlungen beschreiben.

Die vier Kategorien nach Tsui (2003) bzw. Wu und Puntambekar (2012) lassen sich anhand des Abstraktionsgrades weiter charakterisieren. Repräsentationen, die über einen hohen symbolischen Anteil mit willkürlicher Struktur verfügen und daher nur über Konventionen mit einem repräsentierten Gegenstand oder Prozess verknüpft sind, weisen einen hohen Abstraktionsgrad auf.

Dies gilt besonders für *mathematisch-symbolische* Repräsentationen (Müller et al., 2017; Talanquer, 2022). *Verbal-textliche* Repräsentationen können als eine Art Intermediat verstanden werden, das die Interpretation des Zusammenhangs der Symbole in einer *mathematisch-symbolischen* Repräsentation unterstützt (Müller et al., 2017). Verwenden Repräsentationen hingegen ikonische Zeichen, die in direkter Beziehung zum dargestellten Gegenstand oder Prozess stehen, ist ihr Abstraktionsgrad gering. Bei *visuell-graphischen* Repräsentationen, im Sinne von logischen Bildern, beruhen diese ikonischen Zeichen häufig auf Analogien, z. B. die Darstellung eines Atoms als Kugel (Talanquer, 2022). Dadurch stellen sie einen direkten Bezug zum repräsentierten Gegenstand her und sind folglich konkreter und weniger abstrakt. Demnach weisen realistische Bilder wie Fotografien eine höhere Konkretheit auf als logische Bilder; reale Experimente im Sinne *handelnd-operativer* Repräsentationen sind noch konkreter (Müller et al., 2017; Nitz, Ainsworth, et al., 2014; Taber, 2009; Talanquer, 2022). Entsprechend kann statt des Begriffs „Abstraktionsgrad“ auch „Grad der Ikonizität“ verwendet werden (Talanquer, 2022), wobei dieser invers dazu eingesetzt werden muss.

Häufig lassen sich Repräsentationen nicht eindeutig einer Kategorie zuordnen, da sie die Kodalität und Charakteristika verschiedener Kategorien vereinen. Dies gilt besonders in der Chemie für *mathematisch-symbolische* und auch *visuell-graphische* Repräsentationen (Dickmann, 2019; Talanquer, 2022). Daher wird diese Kategorisierung um eine „hybride Form“ erweitert, um zu verdeutlichen, dass Repräsentationen hinsichtlich ihrer Zuordnung häufig nicht trennscharf sind bzw. nicht exakt einer Kategorie zugeordnet werden können. Eine Synthese der bisher beschriebenen Kategorisierung ist in Abbildung 2 dargestellt.

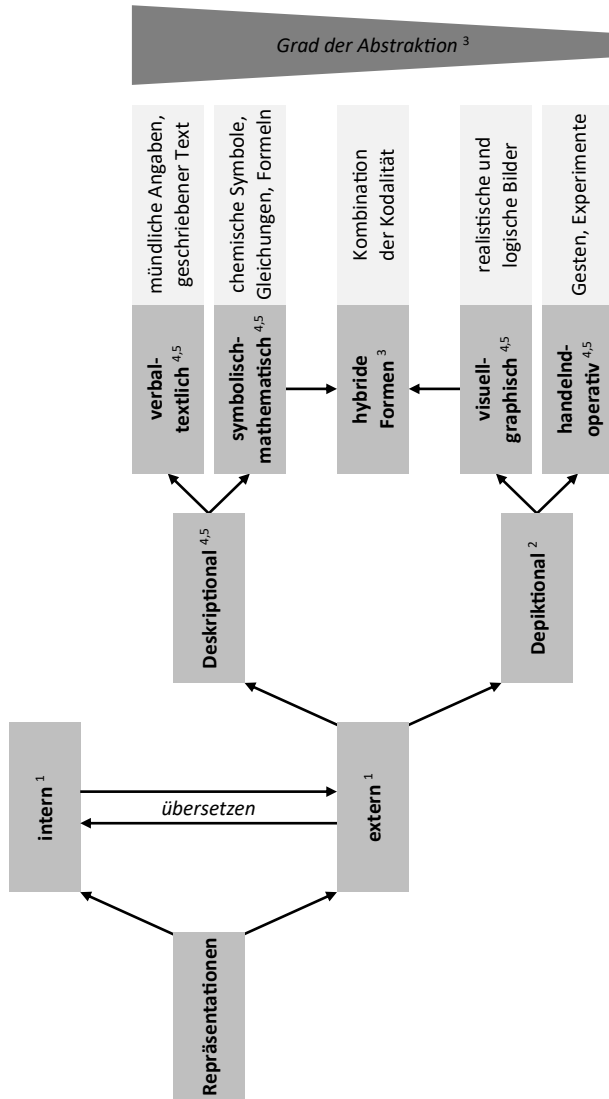


Abbildung 2: Synthese der Kategorisierung von Repräsentationen und Charakterisierung durch den Abstraktionsgrad (¹Krey & Schwanewedel, 2018; ²Schnotz & Bannert, 2003; ³Talanquer, 2022; ⁴Tsui, 2003; ⁵Wu & Puntambekar, 2012), adaptiert nach Nickel et al. (2025).

Eine weitere Klassifikation bzw. Konzeptualisierung von Repräsentationen, die besonders in der Chemie Anwendung findet, basiert nicht auf Charakteristika der Repräsentationen selbst, sondern auf der Art des dargestellten Wissens (Krey & Schwanewedel, 2018). Repräsentationen werden dabei auf drei Levels betrachtet: *macro*, *submicro* und *symbolic* (Gilbert & Treagust, 2009; Russell et al., 1997; Talanquer, 2011; Treagust et al., 2003). Diese werden im *Johnstone-Dreieck* (Johnstone, 1991) zueinander in Beziehung gesetzt, wobei jedes Level gleichermaßen relevant ist und diese sich gegenseitig ergänzen. Unter dem *macroscopic level* (häufig auch *phenomenological level*) werden beobachtbare Phänomene verstanden, die im Alltag oder im Labor wahrnehmbar und somit messbar sind. Dazu zählen Eigenschaften wie die Masse, Temperatur oder der pH-Wert (Gilbert & Treagust, 2009). Um diese Phänomene zu erklären, müssen Chemiker:innen auf eine Ebene zurückgreifen, die mit den Sinnen nicht direkt wahrnehmbar ist: die submikroskopische Ebene der Atome, Ionen und Moleküle. Diese ist nur indirekt zugänglich und kann ausschließlich durch Repräsentationen dargestellt und kommuniziert werden (Bucat & Mocerino, 2009; Chiu & Wu, 2009). Diesen Zusammenhang zwischen dem *macroscopic level* und dem *submicroscopic level* beschreibt Bent (1984) treffend: „Evidently there’s more to seeing than meets the eye. To see what a chemist sees one needs to know what a chemist knows“ (Bent, 1984, S. 774).

Letztlich lassen sich sowohl das *macroscopic* als auch das *submicroscopic level* auf ein *symbolic level* abstrahieren, indem man Symbole verwendet und diesen bestimmte Bedeutungen zuordnet: Symbole zur Kennzeichnung von Atomen oder Verbindungen (H für ein Wasserstoffatom), der elektrischen Ladungen (H^+ , wobei das hochgestellte + eine einfache positive Ladung des Ions angibt), der Anzahl von Atomen, Ionen oder Molekülen durch Koeffizienten und Indizes (zum Beispiel $2 H_2O$; der vorangestellte Koeffizient gibt die Anzahl der Wassermoleküle an, der Index die Anzahl der Atome des vorangehenden Elements in der Verbindung), Buchstaben zur Kennzeichnung des Aggregatzustandes ($H_2O_{(l)}$, l für flüssig), sowie spezielle Konventionen für komplette chemische Reaktionsgleichungen ($2 H_{2(g)} + O_{2(g)} \rightarrow 2 H_2O_{(l)}$). Die drei Level ergänzen sich gegenseitig und können ineinander übersetzt werden (Gilbert & Treagust, 2009; Johnstone, 2000). Beispielsweise lässt sich das Sieden von Wasser auf dem *macroscopic level* beobachten und die Siedetemperatur messen. Auf dem *submicroscopic level* kann der Prozess durch ein Teilchenmodell veranschaulicht oder mittels einer Siedekurve graphisch dargestellt werden.

Letztlich lässt sich der Prozess mit einer Reaktionsgleichung auf dem *symbolic level* ausdrücken, wobei diese Änderung des Aggregatzustands einen physikalischen Vorgang und keine chemische Reaktion darstellt.

Repräsentationen werden auf vielfältige Weise konzeptualisiert, wobei diese Konzeptualisierungen nicht immer klar voneinander abgegrenzt sind. Häufig weisen sie Überschneidungen auf, beispielsweise zwischen dem *Johnstone Dreieck* und der Kategorisierung von Tsui (2003) und Wu und Puntambekar (2012). Auf dem *symbolic level* finden sich vorwiegend symbolisch-mathematische Repräsentationen, auf dem *submicroscopic level* eher visuell-graphische Repräsentationen und auf dem *macroscopic level* am ehesten handelnd-operative Repräsentationen. Zu betonen ist jedoch, dass diese Zuordnungen nicht immer eindeutig sind, da die jeweiligen Kategorisierungen auf unterschiedlichen Grundlagen beruhen. Ainsworth (2006) fasst zusammen: “although there is some overlap between the taxonomies, no one classification is universally accepted” (S. 190). Daher wird in der vorliegenden Arbeit die Kategorisierung von Tsui (2003) und Wu und Puntambekar (2012) zugrunde gelegt (Abbildung 2), die Repräsentationen in verbal-textliche, symbolisch-mathematische, visuell-graphische und handelnd-operative Repräsentationen kategorisiert, was eine adäquate und geeignete Grundlage für das Projekt bildet.

Neben den bisher beschriebenen Kategorisierungen lassen sich externe Repräsentationen zusätzlich in statische und dynamische Repräsentationen einordnen. Dynamische Repräsentationen veranschaulichen Prozesse, die sich in Abhängigkeit der Zeit verändern (Ainsworth & VanLabeke, 2004). Das bekannteste Beispiel hierfür sind Animationen. Die Meta-Studien von Berney und Bétrancourt (2016) sowie Höffler und Leutner (2007) zeigen, dass Animationen bzw. dynamische Repräsentationen im Vergleich zu statischen Repräsentationen lernförderlich sind, wenn es um die Darstellung dynamischer Prozesse geht. Castro-Alonso et al. (2019) betonen aber, dass diese Ergebnisse mit Vorsicht zu interpretieren sind und keine endgültige Schlussfolgerung zulassen, da einige der in den Meta-Studien einbezogenen Studien methodische Mängel aufweisen. Dennoch liefern Castro-Alonso et al. (2019) in ihrer Meta-Studie ebenfalls Evidenz für einen positiven Effekt dynamischer Repräsentation auf das Lernen. Dabei konnten sie Moderatorvariablen identifizieren, etwa Personenvariablen sowie die Art und Weise der Intervention. Besonders das

Geschlecht spielt der Meta-Studie zufolge eine entscheidende Rolle. Das Lernen mit dynamischen Repräsentationen ist bei Männern effektiver als bei Frauen (Castro-Alonso et al., 2019). Im vorliegenden Projekt stehen jedoch statische Repräsentationen im Fokus, da sie eine wesentliche Rolle im Lehr-Lern-Prozess in naturwissenschaftlichen Fächern der Schule und Hochschule einnehmen, z. B. im Hinblick auf Schul- und Fachbücher (Gurung et al., 2022).

2.2 Repräsentationen in der Chemie und anderen Naturwissenschaften

Die Chemie und andere Naturwissenschaften sind durch einen ausgeprägten Gebrauch von Repräsentationen charakterisiert (Eilam & Gilbert, 2014; Gurung et al., 2022; Rau, 2017b). Dies lässt sich durch die Natur des Faches Chemie erklären: „Chemie ist die Wissenschaft, die sich mit der Charakterisierung, Zusammensetzung und Umwandlung von Stoffen befasst“ (Mortimer & Müller, 2020, S. 19). Wie bereits betont, lassen sich chemische Phänomene und Stoffeigenschaften zwar makroskopisch beobachten bzw. beschreiben, die zugrunde liegenden Prozesse und ursächliche Strukturen entziehen sich jedoch unserer Wahrnehmung. Daher müssen wir auf Repräsentationen zurückgreifen, um diese veranschaulichen und erklären zu können (Bucat & Mocerino, 2009; Chiu & Wu, 2009). Hierin liegt eine der Ursachen, weshalb Chemie von Lernenden häufig als komplexes, abstraktes und schwieriges Fach wahrgenommen wird (Cardellini, 2012).

Typisch für die Chemie sind unterschiedliche Repräsentationen von Atomen, Verbindungen, Ionen und Reaktionen. Dabei existiert eine große Bandbreite an Repräsentationen, die es ermöglicht, einen Sachverhalt auf unterschiedliche Weisen darzustellen. Wie in Abbildung 2 dargestellt, lassen sich diese Repräsentationen in verbal-textliche, symbolisch-mathematische oder visuell-graphische sowie hybride Formen einteilen. Sie können zudem als eher abstrakt oder eher konkret charakterisiert werden (Talanquer, 2022). Abbildung 3 zeigt exemplarisch neun unterschiedliche Repräsentationen eines (S)-Milchsäuremoleküls, die nach ihrem Abstraktionsgrad geordnet sind: von eher abstrakt zu eher konkret (Nickel et al., 2025; Talanquer, 2022). Selbstverständlich existieren noch zahlreiche weitere Darstellungsmöglichkeiten, die Übersicht

beschränkt sich bewusst auf die in Lehre und Forschung am häufigsten verwendeten Repräsentationen.

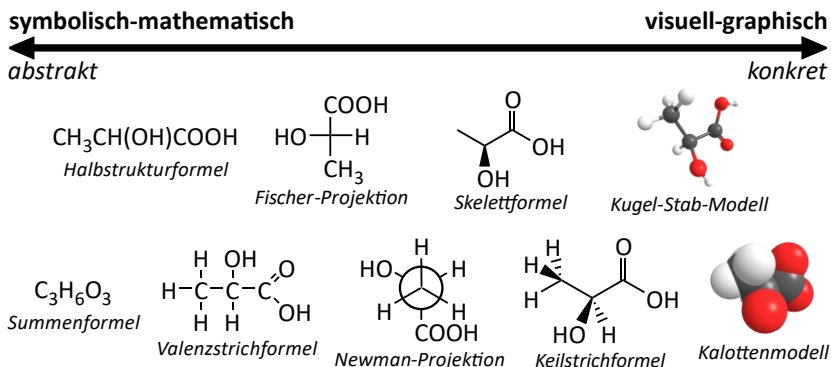


Abbildung 3: Unterschiedliche Repräsentation eines (S)-Milchsäuremoleküls ((S)-2-Hydroxypropansäure), charakterisiert und geordnet nach dem Abstraktionsgrad (adaptiert nach Nickel et al., 2025).

Damit stellt sich die Frage: Warum existieren so viele verschiedene Repräsentationen, um ein und dasselbe darzustellen? Auch wenn die Repräsentationen vermeintlich das Gleiche darstellen, beispielsweise das (S)-Milchsäuremolekül in Abbildung 3, können aus den Repräsentationen unterschiedliche Informationen gewonnen werden. Aus der Summenformel lässt sich die Anzahl der Atome jeder Atomsorte leicht ablesen, sie gibt jedoch keine Informationen über die Art der Verbindung, bzw. die Bindungen zwischen den Atomen oder die räumliche Anordnung. So besitzen beispielsweise Milchsäure, 1,3-Dihydroxacetone und Kohlendimethylester allesamt die Formel $\text{C}_3\text{H}_6\text{O}_3$. Auf die Struktur der Verbindung lässt sich erst aus der Halbstrukturformel schließen, sofern die entsprechenden Konventionen beherrscht werden. Um die absolute Konfiguration des Milchsäuremoleküls (*R* bzw. *S*) zu bestimmen, d. h. die optische Aktivität des Moleküls, müssen räumliche Informationen in der Struktur dargestellt werden. Die *R*- und *S*-Konfiguration eines Moleküls verhalten sich zueinander wie Bild und Spiegelbild: Sie sind nicht durch bloße Rotation zur Deckung bringbar. Dieser vermeintlich kleine Unterschied kann enorme Auswirkungen auf chemische und physikalische Eigenschaften der Verbindung haben, weshalb die Unterscheidung in vielen Kontexten höchst

relevant ist. Ein Beispiel: Das *R*-Carvon riecht intensiv nach Kümmel, während das *S*-Carvon nach Minze duftet, obwohl die beiden Moleküle lediglich spiegelbildlich zueinander sind (Schmuck, 2013).

Jede Repräsentation kodiert somit spezifische Informationen (Rau et al., 2015). Einige davon, beispielsweise die Anzahl der Atome jeder Atomsorte in einer Verbindung, lassen sich aus mehreren Repräsentationen gewinnen. Andere Informationen, wie etwa die maßstabsgetreuen Bindungslängen, sind in Abbildung 3 ausschließlich dem Kugel-Stab-Modell und dem Kalottenmodell zu entnehmen. Um fachliche Probleme zu lösen, muss somit eine zielführende Repräsentation gewählt werden, die dem geplanten Zweck gerecht wird. Sollen beispielsweise die Reaktionsprodukte der monomolekularen nucleophilen Substitution erster Ordnung (S_N1 -Mechanismus) von (*S*)-3-Chlor-3-methylhexan und Hydroxidionen vorhergesagt werden, ist eine Strukturformel mit räumlicher Information erforderlich, zum Beispiel die Keilstrichformel. Bei der Reaktion findet ein Übergang von einem vierbindigen, tetraedrischen Edukt zu einem dreibindigen, planaren Carbeniumion als Zwischenprodukt statt. Dieses planare Carbeniumion reagiert mit einem nucleophilen Hydroxidion (OH^-), das von beiden Seiten angreifen kann, sodass ein racemisches Gemisch aus (*R*)- und (*S*)-3-Methylhexan-3-ol entsteht, die sich wie Bild und Spiegelbild gleichen (Abbildung 4). Werden hier Repräsentationen wie die Summenformel oder die Valenzstrichformel verwendet, kann diese Aufgabe nicht ausreichend gelöst werden, da die (*R*)- beziehungsweise (*S*)-Konfiguration des Produkts nicht erkennbar ist (Latscha et al., 2016; Schmuck, 2013).

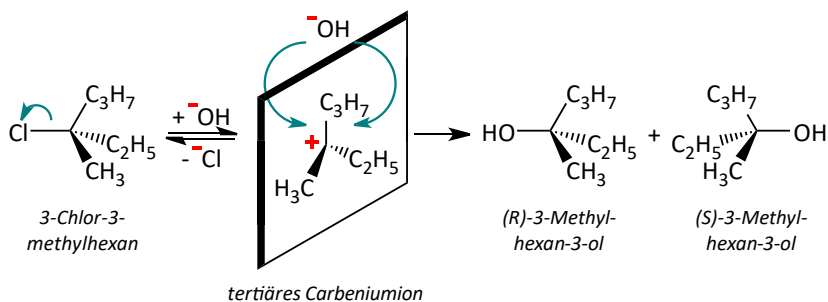


Abbildung 4: S_N1 -Mechanismus der Reaktion von (*S*)-3-Chlor-3-methylhexan mit Hydroxidionen (OH^-) über ein tertiäres Carbeniumion als Zwischenstufe zu (*R*)- und (*S*)-3-Methylhexan-3-ol (adaptiert nach Latscha et al., 2016).

2.2.1 Repräsentation und Modell – Was unterscheidet sie?

Eng verbunden mit dem Begriff der „Repräsentation“ ist in den Naturwissenschaften der Begriff „Modell“. So zeigt Abbildung 3 beispielsweise ein Kugel-Stab-Modell sowie das Kalottenmodell desselben Moleküls. Worin unterscheiden sich jedoch Repräsentation und Modell, und was haben sie gemeinsam?

Modelle lassen sich als „Repräsentationen von Originalen bzw. Phänomenen oder Systemen einer Erfahrungswelt auffassen, die diese Phänomene für einen bestimmten Zweck und zeitlich befristet beschreiben oder erlauben, über die Erfahrungswelt Hypothesen aufzustellen und diese zu untersuchen“ (Krüger et al., 2018, S. 142). Ähnlich definiert auch Kircher (2015a) Modelle, wobei er das Subjekt noch stärker einbezieht: Modelle (M) werden von Subjekten (S) für einen bestimmten Zweck genutzt oder geschaffen, können theoretisch oder gegenständlich sein und weisen zu bestimmten Elementen des modellierten Objekts (O) Analogien auf. Aus diesen Definitionen lässt sich bereits ableiten, dass Modelle ebenfalls Repräsentationen (Kapitel 2.1) sind und beide als Werkzeuge zur Kommunikation dienen. Dazu werden aus beobachtbaren Phänomenen Information abstrahiert und häufig in vereinfachter Weise dargestellt (Krey & Schwanewedel, 2018; Krüger et al., 2018).

Repräsentationen dienen in der Regel der klaren Darstellung und Kommunikation von Informationen. Im Modellierungsprozess können sie als Werkzeuge genutzt werden, um Modelle (mental) zu entwickeln, zu überprüfen und zu erweitern; eine zentrale Kompetenz in den Naturwissenschaften (Justi et al., 2009; Krey & Schwanewedel, 2018). Modelle sind Projektionen eines Objekts, Systems, Phänomens oder Prozesses (Abbildungsmerkmal) und können ihr Zielobjekt nicht eindeutig und endgültig darstellen. Somit gibt es in der Realität immer Elemente, die im Modell keine Entsprechung finden (Verkürzungsmerkmal) (Kircher, 2015a).

In diesen Aussagen zeigt sich eine epistemologische Perspektive, die für die Erkenntnisgewinnung zentral ist. Modelle dienen nicht nur der Entwicklung von Fachwissen, sondern fördern auch das Bewusstsein dafür, dass sie lediglich Werkzeuge darstellen und sich grundsätzlich vom Original unterscheiden (Krey & Schwanewedel, 2018; Krüger et al., 2018; KMK, 2020).

Modelle und Repräsentationen weisen somit eine erhebliche Überschneidung auf. Eine klare Abgrenzung ist häufig schwierig, und auch in der einschlägigen Fachliteratur werden die Begriffe oft synonym oder nicht trennscharf verwendet. In dieser Arbeit wird daher durchgehend der Begriff der Repräsentation verwendet, da dieser Modelle mit umfasst und ein breiteres Feld als diese abdeckt. So werden beispielsweise realistische Bilder (z. B. Fotos) und Diagramme mit einbezogen, die unter dem Modellbegriff auf Basis der Definition in der Regel nicht eingeordnet werden können (Kircher, 2015a).

2.2.2 Lernen mit multiplen externen Repräsentationen

Im Lehr-Lern-Prozess werden häufig nicht einzelne, sondern multiple externe Repräsentationen eingesetzt, die miteinander kombiniert werden müssen (Ainsworth, 2006; Rau et al., 2015; Rexigel et al., 2024). So können beispielsweise bei der Auswertung einer Titration (*handelnd-operativ*) einer Salzsäure-Lösung mit Natronlauge zusätzlich Diagramme (hybride Repräsentationsform) genutzt werden, die den Titrationsverlauf bzw. den pH-Wert in Abhängigkeit vom Volumen der zugegebenen Natronlauge darstellen. Ergänzend werden auch Reaktionsgleichungen (*symbolisch-mathematisch*) eingesetzt, um die Säure-Base-Reaktion zu veranschaulichen.

Die Forschung zum Lernen mit multiplen Repräsentationen hat in den letzten Jahren verstärkt den kombinierten Einsatz multipler externer Repräsentationen in den Vordergrund gerückt und Evidenz dafür geliefert, dass dieser im Vergleich zum Einsatz einzelner externer Repräsentationen das Lernen verbessert (Kozma & Russell, 2005; Rexigel et al., 2024). Dabei wurde vornehmlich das Lernen mit textlichen Repräsentationen und ergänzenden visuell-graphischen Repräsentationen untersucht. Die Ergebnisse der Studien zeigen, dass Informationen aus Texten besser verstanden und behalten werden, wenn sie mit visuellen Repräsentationen verknüpft sind (Rau et al., 2015; Schnotz & Bannert, 2003).

Im Lehr-Lern-Prozess werden nicht nur textliche mit visuell-graphischen Repräsentationen kombiniert, sondern Repräsentationen aller Kategorien (Abbildung 2). Durch die Kombination verschiedener Repräsentationsformen, die jeweils unterschiedliche Aspekte eines zugrundeliegenden Konzepts in den

Vordergrund rücken, wird Lernenden ermöglicht, Konzepte ganzheitlich zu erfassen. Das ermöglichen einzelne Repräsentationen in der Regel nicht (Lichtenberger et al., 2024; Tsui & Treagust, 2013). Zur Interpretation dieser Ergebnisse ist die *Cognitive Theory of Multimedia Learning* (Mayer, 2021) heranzuziehen, die die Verarbeitung multimodaler und -kodaler Medien beschreibt.

Die Cognitive Theory of Multimedia Learning – Die Theorie ist eine evidenzbasierte Theorie zum Lernen mit deskriptionalen und depiktionalen Repräsentationen (Abbildung 2). Sie basiert auf drei Grundannahmen: (1) Menschen verfügen über zwei separate Kanäle zur Verarbeitung von Informationen. (2) Jeder dieser Kanäle besitzt nur eine begrenzte Kapazität für die Informationsverarbeitung im Arbeitsgedächtnis. (3) Zudem wird Lernen als aktiver Prozess gesehen, bei dem relevante Informationen in eine kohärente mentale Repräsentation organisiert und in bestehende Wissensstrukturen integriert werden (Mayer, 2021).

Zwei Kanäle – Menschen verfügen über zwei Kanäle zur Verarbeitung von Informationen: einen verbalen Kanal für sprachliche und einen bildlichen Kanal für visuelle Inhalte. Werden beispielsweise Bilder oder Videos präsentiert, so wird über die Augen der visuelle Kanal aktiviert, während gesprochene Wörter oder geschriebener Text über das Gehör primär den verbalen Kanal ansprechen. Obwohl Informationen über einen Sinneskanal aufgenommen werden, können sie im Arbeitsgedächtnis konvertiert werden. Es ist möglich, Bilder mental in Worte zu übersetzen und umgekehrt (Baddeley, 2001; Mayer, 2021; Schnotz, 2021).

Begrenzte Kapazitäten – Wird einer Person beispielsweise ein Bild gezeigt, so werden die Informationen über den visuellen Kanal aufgenommen. Die Kapazität dieses Kanals ist jedoch begrenzt: Es können jeweils nur wenige Elemente gleichzeitig im Arbeitsgedächtnis verarbeitet werden, sodass keine vollständige mentale „Kopie“ des Originals entsteht. Dasselbe gilt für Informationen, die über den verbalen Kanal aufgenommen werden (Baddeley, 2001; Mayer, 2021; Sweller et al., 1998, 2019).

Aktiver Prozess – Das Erstellen einer kohärenten mentalen Repräsentation wird als aktiver kognitiver Prozess verstanden. Menschen nehmen aktiv relevante Informationen in das Arbeitsgedächtnis auf, organisieren diese im Ar-

beitsgedächtnis in mentale Repräsentationen und integrieren diese mit bereits vorhandenen mentalen Repräsentationen, dem Vorwissen aus dem Langzeitgedächtnis (Mayer, 2021).

Diese drei Grundannahmen lassen sich in ein Modell integrieren, das in Abbildung 5 dargestellt ist.

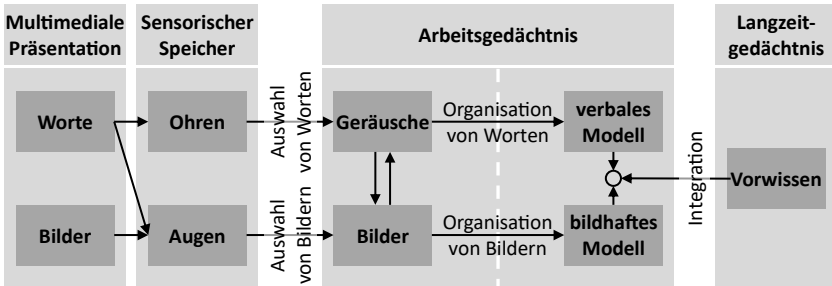


Abbildung 5: Modell der *Cognitive Theory of Multimedia Learning* nach Mayer (2021).

Vor dem Hintergrund der *Cognitive Theory of Multimedia Learning* (Abbildung 5) lassen sich die empirischen Befunde zur Förderung des Lernens mit multiplen externen Repräsentationen interpretieren und erklären. Lernende wählen zunächst relevante visuelle und verbale Informationen aus den Repräsentationen. Dabei werden durch unterschiedliche sensorische Informationen verschiedene Kanäle angesprochen, wodurch die Kapazitäten der Speicher optimal genutzt und in Bezug auf das einzelne Konzept erweitert werden können. Das Arbeitsgedächtnis erzeugt in der Folge interne verbal-textliche und visuell-graphische Repräsentationen, die mit dem bereits vorhandenen relevanten fachlichen Vorwissen zu einem kohärenten Konzept integriert werden (Mayer, 2021; Rau, 2017b).

Das Lernen mit multiplen externen Repräsentationen stellt jedoch zeitgleich eine Herausforderung dar, da die Lernenden in der Lage sein müssen, die Aspekte der Repräsentationen miteinander zu verknüpfen (Rau, 2017b). Ainsworth (2008) betont, dass Lernende häufig daran scheitern, die Vorteile multipler externer Repräsentationen zu nutzen. Im schlimmsten Fall kann es zu einer Hemmung des Lernens kommen, wenn multiple Repräsentationen nicht

adäquat kombiniert werden (Ainsworth, 2008). Diese Beobachtung wird häufig mit der *Cognitive Load Theory* begründet (Tsui & Treagust, 2013). Ainsworth (2008) fasst zusammen: „[...] multiple representations are powerful tools but like all powerful tools they need careful handling if learners are to use them successfully“ (Ainsworth, 2008, S. 191).

Die Cognitive Load Theory – Die Theorie zielt darauf ab, zu erklären, wie die von Lernaufgaben erzeugte kognitive Belastung die Verarbeitung von neuen Informationen und den Aufbau von Wissen im Langzeitgedächtnis beeinflusst. Die Grundannahme der Theorie ist, wie auch bei der *Cognitive Theory of Multimedia Learning*, dass das Arbeitsgedächtnis über eine begrenzte Kapazität verfügt und somit nur ein eingeschränkter Umfang an Informationen zeitgleich verarbeitet werden kann. Ursprünglich wurden drei Arten von kognitiver Belastung unterschieden:

(1) Intrinsische kognitive Belastung (*intrinsic cognitive load*)

Fachliche Konzepte unterscheiden sich in der Belastung des Arbeitsgedächtnisses dadurch, wie viele Elemente parallel verarbeitet werden müssen und wie stark diese miteinander interagieren. Je stärker die Interaktivität der Elemente ist, desto komplexer sind die Informationen und desto höher ist die intrinsische kognitive Belastung. Dies ist beim Lernen fachlicher Konzepte unvermeidbar (Kirschner et al., 2018; Paas & Sweller, 2021; Sweller, 2011; Sweller et al., 1998, 2019).

(2) Extrinsische kognitive Belastung (*extraneous cognitive load*)

Die extrinsische kognitive Belastung wird im Gegensatz zur intrinsischen kognitiven Belastung nicht durch die Komplexität der Informationen an sich, sondern durch die Art der Präsentation der Informationen beeinflusst. Diese Art der Belastung wird erhöht, wenn beispielsweise viele irrelevante Informationen unübersichtlich präsentiert werden und so eine vermeidbare hohe Interaktion zwischen den Elementen erzeugt wird. Das Ausmaß dieser Belastung kann durch das Design der Aufgabe und die instruktionale Gestaltung gesteuert und möglichst gering gehalten werden (Kirschner et al., 2018; Paas & Sweller, 2021; Sweller, 2011; Sweller et al., 1998, 2019).

(3) Lernbezogene kognitive Belastung (*germane cognitive load*)

Ursprünglich wurden unter dieser Art der Belastung Ressourcen verstanden, die hauptsächlich zur Bewältigung der intrinsischen kognitiven und

nebensächlich der extrinsischen kognitiven Belastung aufgewendet werden müssen und somit für das Lernen notwendig sind. Je mehr Ressourcen zur Bewältigung der extrinsischen kognitiven Belastung benötigt werden, desto weniger Ressourcen stehen für die Bewältigung der intrinsischen kognitiven Belastung zur Verfügung und desto weniger wird folglich gelernt (Paas & Sweller, 2021; Sweller et al., 1998, 2019).

Nach dieser Definition stellt die lernbezogene kognitive Belastung eine eigene Belastungsart dar, die zur kognitiven Gesamtbelastung beiträgt, indem sie die extrinsische kognitive Belastung ersetzt. Unter dieser Annahme dürfte eine Reduktion der extrinsischen kognitiven Belastung die Gesamtbelastung nicht verringern. Zahlreiche Studien zeigen jedoch genau das Gegenteil. Durch die Verringerung der extrinsischen kognitiven Belastung sinkt auch die kognitive Gesamtbelastung. Daher wird die lernförderliche kognitive Belastung heute nicht mehr als eigenständige Belastungsart verstanden, sondern vielmehr als Umverteilungsfunktion für Ressourcen von extrinsischer zu intrinsischen Belastung (Paas & Sweller, 2021; Sweller et al., 2019).

Kurz zusammengefasst lässt sich festhalten, dass die extrinsische kognitive Belastung möglichst gering gehalten werden sollte, sodass ein Maximum an Ressourcen für die Bewältigung der intrinsischen kognitiven Belastung zur Verfügung steht, um das Lernen zu optimieren (Paas & Sweller, 2021; Sweller et al., 2019).

Bezieht man die *Cognitive Load Theory* auf die Befunde, dass Lernende multiple externe Repräsentationen häufig nicht adäquat kombinieren können und das Lernen damit folglich nicht förderlich ist, so kann dies auch auf die extrinsische kognitive Belastung zurückgeführt werden. Die Lernenden müssen viele Ressourcen aufwenden, um jede einzelne Repräsentation zu analysieren und anschließend die einzelnen Repräsentationen aufeinander zu beziehen und miteinander zu verknüpfen. So entsteht eine hohe extrinsische kognitive Belastung, wodurch die Ressourcen zur Bewältigung der intrinsischen kognitiven Belastung fehlen (Kenneally & Bentley, 2024; Tsui & Treagust, 2013). Dies gilt besonders, wenn es sich um neuartige Repräsentationen handelt, die den Lernenden unbekannt sind oder redundante Informationen darstellen (Nyachwaya & Gillaspie, 2016; Rau, 2018).

Aus der *Cognitive Load Theory* und der *Cognitive Theory of Multimedia Learning* lassen sich somit Prinzipien ableiten, die bei der Gestaltung von multimedialen Lerngelegenheiten berücksichtigt werden sollten, um das Lernen zu optimieren. Hier lässt sich exemplarisch das *Modalitätsprinzip* nennen: Visuell-graphische Repräsentationen sollten mit gesprochenen Texten ergänzt werden, sodass der visuelle und der verbale Kanal angesprochen und die Kapazitäten beider Kanäle optimal genutzt werden können (Castro-Alonso & Sweller, 2021). Weitere Beispiele sind das *Kohärenzprinzip*, *Signalisierungsprinzip*, *Redundanzprinzip*, *Segmentierungsprinzip* und der *Split-Attention-Effekt*. Daneben existieren weitere Prinzipien, die hier jedoch nicht näher behandelt werden sollen. Um diese nachzuvollziehen und zu vertiefen, empfehlen sich die Publikationen von Fiorella und Mayer (2021), Castro-Alonso und Sweller (2021) sowie Sweller und Kolleg:innen (2019).

2.2.3 Lernen mit und über Repräsentationen – Ein Dilemma

Im Fach Chemie werden Repräsentationen genutzt, um Konzepte zu kommunizieren und zu vermitteln. Somit stehen sie immer in einem direkten Zusammenhang mit chemischem Fachwissen, was durch das Zitat von Bent (1984) noch einmal akzentuiert wird: „To see what a chemist sees one needs to know what a chemist knows“ (Bent, 1984, S. 774). An dieser Stelle muss die Frage beantwortet werden, was *Fachwissen* eigentlich umfasst.

Grundsätzlich wird unter Fachwissen das Wissen über fachliche Konzepte und Prozesse zum Erklären naturwissenschaftlicher Phänomene sowie methodisches Wissen verstanden (Kircher, 2015b). Es ist domänenspezifisch, d. h., auch innerhalb eines Fachs kann es sich das Fachwissen intraindividuell unterscheiden. Eine Person mit ausgeprägtem Fachwissen in der Organischen Chemie muss nicht zwangsläufig auch in der Physikalischen Chemie bewandert sein (Tepner et al., 2012). Hier wird die Konzeptualisierung von Fachwissen nach Anderson und Krathwohl (2001) als Rahmen verwendet, die zwischen vier Wissensaspekten unterscheidet: Faktenwissen, konzeptuelles Wissen, prozedurales Wissen und metakognitives Wissen.

Faktenwissen umfasst Wissen, das sich weitgehend auf fachinhaltliche Fakten und Fachbegriffe beschränkt. *Konzeptuelles Wissen* bezeichnet hingegen ver-

netztes Wissen über Kategorien, Strukturen, Modelle sowie Theorien. Zusammen bilden Faktenwissen und konzeptuelles Wissen das *deklarative Wissen*. Durch das *prozedurale Wissen* wird Wissen über Abläufe, Prozesse und Vorgänge abgebildet, das von basalen Verhaltensweisen bis hin zu komplexen Handlungsmustern reicht. Das *metakognitive Wissen* stellt das Wissen über die eigene Kognition dar und umfasst die Fähigkeiten, Lernprozesse selbst zu steuern, Informationen gezielt zu verarbeiten und Strategien zur Problemlösung anzuwenden (Anderson & Krathwohl, 2001; Gebhard et al., 2017; Maier et al., 2010; Vukić et al., 2020).

Betrachtet man die Chemie (und andere Naturwissenschaften), so kristallisiert sich hier eine enorme Herausforderung für die Lernenden heraus. Chemisches Fachwissen kann kaum ohne Repräsentationen erlernt werden. Gleichzeitig muss aber auch über die Repräsentationen selbst gelernt werden, also darüber, wie diese chemische Konzepte vermitteln, was wiederum nicht ohne chemisches Fachwissen gelingen kann (Dreher & Kuntze, 2015; Rau et al., 2021; Steinbach et al., 2024). Somit werden Lernende in Lernsituationen häufig mit dem Problem konfrontiert, dass sie neue, zunächst unverständliche chemische Konzepte mit unbekannten Repräsentationen lernen müssen. Dieses Problem wird als *Repräsentationsdilemma* bezeichnet (Rau, 2017b, 2018) und macht das Lernen mit Repräsentationen zu einem zweiseitigen Schwert: Auf der einen Seite müssen Lernende chemische Konzepte *von* chemischen Repräsentationen erlernen, auf der anderen Seite müssen sie *über* chemische Repräsentationen lernen (Rau, 2017a, 2017c; Rau et al., 2021).

Dies lässt sich beispielsweise an der Titrationskurve einer schwachen Säure (hier Essigsäure, $\text{CH}_3\text{COOH}_{(\text{aq})}$) mit einer starken Base (hier Natronlauge, $\text{NaOH}_{(\text{aq})}$) veranschaulichen (Abbildung 6). Dabei sind insbesondere der Pufferbereich (mit dem Pufferpunkt) sowie der Äquivalenzpunkt als charakteristische Punkte beziehungsweise Bereiche zu identifizieren, die sich in der Kurve als Wendepunkte erkennen lassen. Außerdem ist das Diagramm als solches zu analysieren und zu beschreiben, beispielsweise dass es den pH-Wert der Probelösungen in Abhängigkeit von der zugegebenen Natronlauge darstellt. Diese Aspekte lassen sich dem Wissen *über* Repräsentationen, die dahinterstehenden fachlichen Aussagen und Erklärungen dem Fachwissen zuordnen. Dazu gehört das Verständnis, dass in einem Essigsäure-Acetat-Puffer am Pufferpunkt die Stoffmenge an Essigsäure (CH_3COOH) gleich der Stoffmenge an Acetationen (CH_3COO^-) ist. Dadurch bleibt der pH-Wert bei Zugabe kleiner

Mengen Säure oder Basen weitgehend konstant. Im Äquivalenzpunkt ist die Stoffmenge der Acetationen gleich der Stoffmenge der Natriumionen (Na^+), die durch Dissoziation des Natriumhydroxids (NaOH) neben den Hydroxidionen (OH^-) in Wasser entstehen (entspricht der Natronlauge, $\text{NaOH}_{(\text{aq})}$). Da Essigsäure eine schwache Säure ist, reagiert ein Teil der Acetationen mit Wasser zu Essigsäuremolekülen und Hydroxidionen, wodurch der pH-Wert im Äquivalenzpunkt nicht neutral ist ($\text{pH} = 7$), sondern im basischen Bereich. Bei der Titration einer starken Säure (z. B. Salzsäure, $\text{HCl}_{(\text{aq})}$) mit einer starken Base fallen Äquivalenz- und Neutralpunkt dagegen zusammen (Binnewies et al., 2016; Mortimer & Müller, 2020).

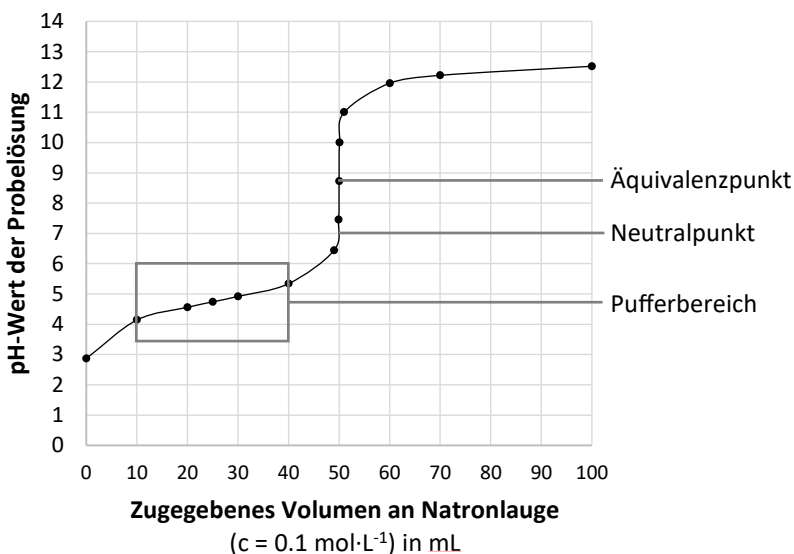


Abbildung 6: Titrationskurve der Titration von 50 mL Essigsäurelösung ($c = 0.1 \text{ mol} \cdot \text{L}^{-1}$) mit Natronlauge ($c = 0.1 \text{ mol} \cdot \text{L}^{-1}$) mit markierten charakteristischen Punkten und Bereichen (adaptiert nach Mortimer & Müller, 2020, S. 310 f.).

Im Lerngeschehen, vor allem in der Schule, verschärft sich dieses Dilemma: Repräsentationen werden oft lediglich als Mittel zur Vermittlung chemischer Konzepte eingesetzt. Dass sie selbst eine Herausforderung darstellen, wird

häufig nicht beachtet, und die Frage, wie Repräsentationen chemische Konzepte eigentlich kommunizieren, rückt oft in den Hintergrund (Popova & Jones, 2021; Steinbach et al., 2024). Lernenden müssen Fähigkeiten und Kompetenzen vermittelt werden, mit denen sie Informationen aus Repräsentationen entnehmen und somit erfolgreich und zielführend mit diesen arbeiten können (z. B. KMK, 2020; Kapitel 2.2.4). Diese Fähigkeiten werden unter dem Begriff *Repräsentationskompetenzen* zusammengefasst (Rau, 2017c, 2018; Rau et al., 2015, 2021; Steinbach et al., 2024).

In diesem Zusammenhang stellt sich folgende Frage: Wenn Lernen *mit* und Lernen *über* Repräsentationen Hand in Hand gehen, inwiefern können Repräsentationskompetenzen und Fachwissen dann als empirisch differenzierbare Konstrukte betrachtet werden, oder sollten sie vielmehr als zwei Facetten eines übergeordneten Konstrukts gefasst werden? Diese Frage beschäftigt viele Forscher:innen und gibt auch aktuell Anlass zu Diskussionen. Aus der naturwissenschaftsdidaktischen Forschung liegen jedoch einzelne Befunde, die diese Frage aufgreifen. Die Ergebnisse faktoranalytischer Untersuchungen von Nitz, Ainsworth und Kolleg:innen (2014) sowie von Edelsbrunner und Kolleg:innen (2023) deuten darauf hin, dass sich Repräsentationskompetenzen und Fachwissen zwar empirisch trennen lassen, aber dennoch stark zusammenhängen. Des Weiteren konnten sie erkennen, dass Lernende mit gut ausgeprägten Repräsentationskompetenzen über ein mangelhaftes Fachwissen verfügen können, während die umgekehrte Konstellation nicht beobachtet werden konnte. Daraus leiten sie die Hypothese ab, dass Repräsentationskompetenzen eine Grundlage zur Entwicklung von Fachwissen darstellen (Edelsbrunner et al., 2023; Nitz, Ainsworth, et al., 2014) und untermauern damit die Aussagen zur Relevanz von Repräsentationskompetenzen von Rau (2017c, 2018).

2.2.4 Repräsentationskompetenzen

Aus dem vorangegangenen Kapitel geht hervor, dass Repräsentationskompetenzen eine zentrale Rolle im Lehr-Lern-Prozess spielen. Es stellt sich die Frage: Was genau sind Repräsentationskompetenzen? Zur Beantwortung dieser Frage nachfolgend einige zentrale Punkte zum Lernen mit Repräsentationen:

- Repräsentationen nutzen Zeichen und Symbole, um Informationen auf unterschiedliche Art und Weise zu kommunizieren (Kapitel 2.1).
- In der Chemie (und anderen Naturwissenschaften) sind Repräsentationen unverzichtbar: Sie ermöglichen die Kommunikation über nicht beobachtbare chemische Phänomene (Kapitel 2.1) und sind somit zentrale Voraussetzung für chemische Problemlöseprozesse (Kapitel 2.2).
- Es existiert eine große Bandbreite an chemischen Repräsentationen, wobei jede einzelne einem spezifischen Zweck dient. Erst durch die Kombination und Verknüpfung multipler externen Repräsentationen entsteht ein kohärentes Gesamtbild des Sachverhalts (Kapitel 2.2).
- Beim Lernen mit multiplen externen Repräsentationen sollte die extrinsische kognitive Belastung minimiert werden, sodass die Ressourcen des Arbeitsgedächtnisses optimal für die Bewältigung der intrinsischen kognitiven Belastung zur Verfügung stehen (Kapitel 2.2.2).
- Lernen mit Repräsentationen impliziert ein Dilemma: Das Lernen chemischer Konzepte erfolgt häufig parallel zum Lernen über die Repräsentationen, also dazu, wie diese das chemische Konzept vermitteln (Kapitel 2.2.3).

Diese Punkte verdeutlichen, dass das Lernen mit chemischen Repräsentationen eine Herausforderung darstellt und Lernende daher über spezifische Fähigkeiten und Kompetenzen verfügen müssen, um diese Herausforderungen zu bewältigen. Sie müssen in der Lage sein, Repräsentationen zu dekodieren, um die darin enthaltenen Informationen zu erschließen (Steinbach et al., 2024; Tsui & Treagust, 2013). Sie müssen Aspekte multipler Repräsentationen aufeinander übertragen und diese miteinander verknüpfen können (Ainsworth, 2008; Kozma & Russell, 2005; Rau, 2017c), um so etwa die extrinsische kognitive Belastung zu reduzieren (Rau, 2018; Tsui & Treagust, 2013). Es geht zudem darum, Repräsentationen flexibel und zielführend in Problemsituationen anwenden zu können (Kozma & Russell, 1997, 2005; Rau et al., 2021). Das Repräsentationsdilemma muss überwunden werden, um den Lernenden den Erwerb chemischer Konzepte anhand von Repräsentationen zu ermöglichen (Rau, 2017c, 2018; Rau et al., 2015, 2021; Steinbach et al., 2024). Diese Fähigkeiten werden als *Repräsentationskompetenz* bezeichnet.

Die Relevanz von Repräsentationskompetenzen zeigt sich auch in den Bildungsstandards des Fachs Chemie zur allgemeinen Hochschulreife. Sie leiten sich aus den Charakterisierungen der vier Kompetenzbereiche ab und finden

in den Kompetenzerwartungen konkrete Umsetzung. Chemische Konzepte müssen u. a. mit Modellen erarbeitet, auf Teilchenebene ausgewertet, mit fachtypischen Repräsentationen dargestellt und kritisch beurteilt werden (KMK, 2020):

- **Sachkompetenz:** „Die Sachkompetenz der Lernenden zeigt sich in der Kenntnis naturwissenschaftlicher Konzepte, Theorien [...]. Im Mittelpunkt steht hierbei die modellhafte Deutung beobachtbarer Phänomene auf Teilchenebene“ (KMK, 2020, S. 13 f.).
- **Erkenntnisgewinnungskompetenz:** „Im Bereich der Erkenntnisgewinnungskompetenz ist es wichtig, [...] Modelle sachgerecht zur Beschreibung eines Phänomens oder zur Gewinnung von Erkenntnissen einsetzen zu können“ (KMK, 2020, S. 15).
- **Kommunikationskompetenz:** „Die Kommunikationskompetenz der Lernenden zeigt sich in der Kenntnis von [...] fachtypischen Darstellungen [...] und in der Fähigkeit, diese zu nutzen, um fachbezogene Informationen zu erschließen, adressaten- und situationsgerecht darzustellen und auszutauschen“ (KMK, 2020, S. 16).
- **Bewertungskompetenz:** „Die Bewertungskompetenz der Lernenden zeigt sich in der Kenntnis von fachlichen und überfachlichen Perspektiven und Bewertungsverfahren und in der Fähigkeit, diese zu nutzen, um Aussagen bzw. Daten anhand verschiedener Kriterien zu beurteilen [...]“ (KMK, 2020, S. 17).

Repräsentationskompetenzen sind nicht nur in der Chemie von zentraler Bedeutung, sondern auch in anderen Naturwissenschaften, wie etwa der Biologie (z. B. Nitz et al., 2014), der Physik (z. B. Kohl & Finkelstein, 2017) und der Mathematik (z. B. Cheng, 1999). Kozma und Russell (1997, 2005) haben sich explizit mit Repräsentationskompetenzen in der Chemie beschäftigt und definieren diese als „set of skills and practices that allow a person to reflectively use a variety of representations or visualizations, singly and together, to think about, communicate, and act on chemical phenomena in terms of underlying, a perceptual physical entities and processes“ (Kozma & Russell, 2005, S. 131). Rau (2017c, 2018) differenziert außerdem zwischen *sense-making skills* und *perceptual fluency*. *Sense-making skills* fassen Fähigkeiten zusammen, die es ermöglichen, zu verstehen, wie einzelne Repräsentationen Informationen darstellen, Gemeinsamkeiten und Unterschiede zwischen multiplen Repräsentationen zu beschreiben sowie diese mit relevanten chemischen Konzepten zu

verknüpfen. *Perceptual fluency* bezeichnet hingegen die Fähigkeit, solche Unterschiede und Gemeinsamkeiten mit geringem kognitivem Aufwand und schnell zu identifizieren, wie es beispielsweise bei Expert:innen ausgeprägt ist (Rau, 2017b, 2017c, 2018).

In ihren Untersuchungen haben Kozma und Russell (1997, 2005) Noviz:innen und Expert:innen in Bezug auf das Verständnis, die Nutzung und die Verknüpfen chemischer Repräsentationen miteinander verglichen. Auf der Grundlage ihrer Ergebnisse haben sie Fähigkeiten abgeleitet, die Expert:innen im Vergleich zu Noviz:innen besitzen und die es ihnen ermöglichen, erfolgreich mit chemischen Repräsentationen umzugehen.

Nitz (2012) hat diese Fähigkeiten kategorisiert und zusammengefasst und Gurung et al. (2022) haben sie in *lower-level skills* und *higher-level skills* eingeordnet. Unter *lower-level skills* werden Fähigkeiten verstanden, chemische Repräsentationen zu interpretieren, zu übersetzen und zu konstruieren. *Higher-level skills* umfassen dagegen die Fähigkeiten, chemische Repräsentationen zur Argumentation zu nutzen, diese kritisch zu vergleichen sowie eine epistemologische Perspektive darauf einzunehmen, weshalb sie häufig auch als *metarepresentational skills* bezeichnet werden (Gurung et al., 2022; Kozma & Russell, 1997, 2005; Nitz, 2012; Nitz et al., 2014). Diese Kategorisierung mit Definitionen kann Tabelle 1 entnommen werden.

Tabelle 1: Übersicht der Repräsentationskompetenzen nach ¹Kozma und Russell (1997; 2005), kategorisiert nach ²Nitz (2012) und ³Gurung et al. (2022).

Kategorisierung ²		Definitionen ¹
lower-level skills ³	Interpretation	Die Fähigkeit, Merkmale (z. B. einen Peak in einem Graphen) und Muster (z. B. den Kurvenverlauf in einem Graphen) in chemischen Repräsentationen zu identifizieren und zu analysieren, sowie Repräsentationen zum Beschreiben chemischer Phänomene in Bezug auf zugrundeliegende molekulare Zusammenhänge und Vorgänge zu nutzen.
	Translation	Die Fähigkeit, Zusammenhänge zwischen multiplen Repräsentationen zu erkennen, Aspekte einer Repräsentation in einer anderen Repräsentation zu identifizieren (z. B. den Peak eines Graphen auf ein Molekül übertragen) und Repräsentationen ineinander zu übersetzen.
	Konstruktion	Die Fähigkeit, eine Repräsentation zu konstruieren oder auszuwählen und zu erklären, weshalb die Repräsentation für den geplanten Einsatz geeignet ist.

higher-level skills ³	Vergleich & Kritik	Die Fähigkeit, zu beschreiben, wie unterschiedliche Repräsentationen die gleichen Informationen auf unterschiedliche Art und Weise kommunizieren und zu erklären, wie eine Repräsentation Aussagen zulässt, die aus anderen Repräsentationen nicht abgeleitet werden können.
	Epistemologie & Funktionsweise	Die Fähigkeit, den erkenntnistheoretischen Standpunkt einzunehmen und die wechselseitige Beziehung zwischen chemischem Phänomen und Repräsentation zu erläutern.
	Argumentation	Die Fähigkeit, Repräsentationen als Evidenz zur Stützung von Aussagen zu nutzen, Rückschlüsse aus ihnen zu ziehen und Hypothesen über beobachtbare chemische Phänomene daraus abzuleiten.

Die lower-level skills der Repräsentationskompetenz umfassen basale Fähigkeiten im Umgang mit Repräsentationen und bilden die Grundlage für die Entwicklung der higher-level skills (Gurung et al., 2022). Ähnliche Anforderungen formuliert Gilbert (2005, 2008) mit seinen Kriterien der Metavisualisierung, also der Fähigkeit, flexibel und flüssig mit Repräsentationen umzugehen – das heißt, aus ihnen zu lernen, sie zu entschlüsseln und in bestehende Wissensstrukturen zu integrieren. Um dies zu erreichen, müssen Lernende in der Lage sein, die Konventionen von Repräsentationen zu verstehen (*Interpretation*), zwischen verschiedenen Repräsentationen zu übersetzen (*Translation*) und Repräsentationen zu erzeugen (*Konstruktion*) (Gilbert, 2005, 2008). Somit werden auch hier die Kriterien zur Metavisualisierung, wie die lower-level kills von Kozma und Russell (1997, 2005) als grundlegende Fähigkeiten für das Lernen mit Repräsentationen dargestellt.

Um zu prüfen, inwieweit die Repräsentationskompetenzen nach Kozma und Russell (Tabelle 1) im schulischen Kontext in den USA gefördert und thematisiert werden, haben Gurung et al. (2022) dortige Schulbücher für die Organische Chemie analysiert. Die Ergebnisse zeigen, dass vornehmlich die lower-level skills *Interpretation*, *Translation* und *Konstruktion* gefördert und die higher-level skills weitgehend vernachlässigt werden. Des Weiteren wird betont, dass viele Schul- und Sachbücher (besonders im Bereich der Organischen Chemie) Repräsentationen nicht effektiv in den Lehr-Lern-Prozess einbinden und beispielsweise Gestaltungsprinzipien häufig nicht optimal umsetzen (z. B. werden redundante und dekorative Repräsentationen verwendet). Dadurch wird die extrinsische kognitive Belastung des Arbeitsgedächtnisses unnötig erhöht (Gurung et al., 2022; Nyachwaya & Gillaspie, 2016). Außerdem gelingt es

vielen Lehrbüchern nur unzureichend, den Zusammenhang zwischen einzelnen Repräsentationen und der makroskopischen, submikroskopischen und symbolischen Ebene adäquat darzustellen (Aydin et al., 2014; Nyachwaya & Gillaspie, 2016).

Die Ergebnisse dieser Studien verdeutlichen, dass die Förderung der Repräsentationskompetenzen im Lehr-Lern-Prozess, besonders mit Blick auf die higher-level skills, zu kurz kommt, obwohl diese unerlässlich für das Erlernen chemischer Konzepte sind. Vielmehr werden Repräsentationen oft als Bestandteil des Fachwissens gesehen, wodurch die Entwicklung der Repräsentationskompetenzen häufig nur nebensächlich und zweitrangig abläuft. Die Förderung der Repräsentationskompetenzen sollte jedoch einen zentralen Aspekt im Unterricht einnehmen, sodass chemische Konzepte optimal entwickelt werden können (Aydin et al., 2014; Gurung et al., 2022; Nyachwaya & Gillaspie, 2016; Popova & Jones, 2021; Rau et al., 2021; Sim & Daniel, 2014; Wu & Rau, 2018).

Damit stellt sich die Frage: Wie können Repräsentationskompetenzen gefördert werden und welche Effekte lassen sich damit erzielen? Dies wurde bereits in einigen Studien untersucht, die Evidenzen dafür liefern konnten, dass Fördermaßnahmen zum einen den Einsatz und Gebrauch von Repräsentationen durch Lernende in Problemlösesituationen erhöhen und zum anderen das Lernen von Fachwissen mithilfe von Repräsentationen fördern (Stieff et al., 2016; Stull et al., 2012, 2016; Wu & Rau, 2018). Dabei wurden unterschiedliche Fördermaßnahmen untersucht. Rau und Kolleg:innen (2021) konnten aufzeigen, dass ein adaptives im Vergleich zu einem nicht-adaptiven Unterstützungsangebot dabei hilft, Fehlern und Verwirrungen bei der Arbeit mit Repräsentationen im Problemlöseprozess entgegenzuwirken und gleichzeitig das Lernen von Fachwissen zu verbessern. Mithilfe von Skizzen-Prompts haben Wu und Rau (2018) gezeigt, dass sich einerseits die Qualität der Skizzen von Lernenden verbessert und andererseits, dass dadurch das Verständnis darüber intensiviert wird, wie die Repräsentationen Inhalte darstellen. Stull und Kolleg:innen (2016) sowie die Arbeitsgruppe um Stieff (2016) liefern Evidenz dafür, dass der Einsatz haptischer und virtueller Modelle in der (Organischen) Chemie die Fähigkeit fördert, zwischen verschiedenen Repräsentationen zu übersetzen. Diese Befunde unterstützen damit die Forderung nach einer stärkeren Förderung der Repräsentationskompetenzen im Lehr-Lern-Prozess.

Die in Tabelle 1 dargestellten Repräsentationskompetenzen von Kozma und Russell (1997, 2005) finden im naturwissenschaftsdidaktischen Kontext auf breite Anerkennung und werden in zahlreichen Studien als Fundament für die Untersuchungen verwendet (z. B. Gurung et al., 2022; Nitz, Prechtel, et al., 2014; Popova & Jones, 2021; Sim & Daniel, 2014; Stieff & DeSutter, 2021). Bisher gibt es jedoch nur wenige Studien, die die Struktur des Modells empirisch überprüfen. Faktoranalytische Untersuchungen von Scheidt et al. (2018) sowie des Teams um Taskin (2015) liefern jedoch Evidenz dafür, dass sich die Repräsentationskompetenz nicht in separate Teilfähigkeiten trennen lässt. Auch Ward und Kolleg:innen (2025) liefern hier einen wichtigen Beitrag: Mithilfe eines entwickelten Messinstruments zur Erfassung der Fähigkeiten Interpretation, Translation und Argumentation (Tabelle 1) konnten sie zeigen, dass sich die drei Fähigkeiten empirisch nicht trennen lassen. Dadurch wird betont, dass diese Fähigkeiten eng miteinander verknüpft sind. Repräsentationen müssen zunächst interpretiert, die zugrunde liegenden chemischen Konzepte verstanden, relevante von irrelevanten Aspekten unterschieden und die relevanten Informationen mit den entsprechenden Konzepten verknüpft werden, bevor sie beispielsweise zur Hypothesenbildung oder als Evidenz in fachlichen Diskussionen, in argumentativen Prozessen genutzt werden können (Ward et al., 2025).

Interpretation, Translation und Konstruktion sind fundamentale Fähigkeiten bei der Arbeit mit Repräsentationen. Sie stellen die Grundlage für die Entwicklung der higher-level skills wie *Vergleich & Kritik, Epistemologie & Funktionsweise* und *Argumentation* dar (Gurung et al., 2022). Zwar lässt sich zusammenfassen, dass die Struktur der Repräsentationskompetenzen bereits in verschiedenen Studien behandelt wurde, diese Studien sind jedoch kritisch zu betrachten, da sich inhaltliche Schwächen zeigen, insbesondere in Bezug auf die zugrunde gelegte Definition von Repräsentationskompetenzen, deren Abgrenzung vom Fachwissen sowie in den eingesetzten Erhebungsinstrumenten. Vor diesem Hintergrund sind Aussagekraft und Validität der Befunde dieser Studien als eingeschränkt zu betrachten und erfordern eine kritische Prüfung, welche in Kapitel 5.2.2 detailliert erfolgt.

3

Chemie und das räumliche Denken

Neben Repräsentationskompetenzen und Fachwissen spielt ein weiterer Faktor eine essenzielle Rolle beim Lernen mit Repräsentationen in der Chemie: Räumliche Fähigkeiten. Zur Darstellung chemischer Verbindungen wird eine große Bandbreite an Repräsentationen verwendet (Abbildung 3). Dabei werden dreidimensionale Strukturen häufig zweidimensional dargestellt, wobei die Dreidimensionalität teilweise durch spezifische Konventionen angedeutet wird, beispielsweise in der Keilstrichformel (Abbildung 7). Einfache Linien liegen in der Ebene, gestrichelte Keile zeigen von der Ebene nach hinten und gefüllte von der Ebene nach vorne (Mortimer & Müller, 2020). Externe zweidimensionale Repräsentationen müssen somit in dreidimensionale mentale Repräsentationen übersetzt und anschließend manipuliert werden, beispielsweise bei der Darstellung von Reaktionsmechanismen wie in Abbildung 4 (Carlisle et al., 2015; Harle & Towns, 2011; Kenneally & Bentley, 2024; Merchant et al., 2013; Salame & Kabir, 2022; Stieff & Uttal, 2015).

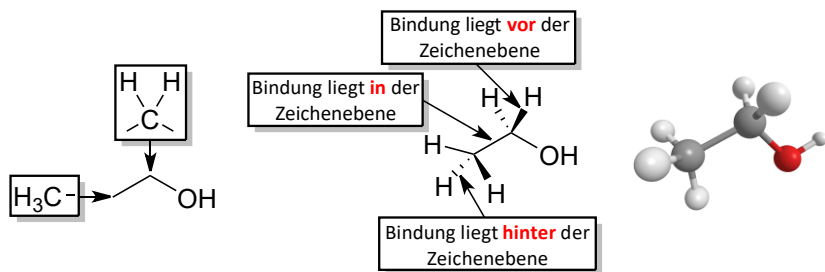


Abbildung 7: Skelettformel, Keilstrichformel und Kugel-Stab-Modell eines Ethanolmoleküls (C_2H_5OH) mit Informationen zur Interpretation spezifischer Repräsentationsmerkmale (verändert nach Schmuck, 2013).

Auch mit Blick auf die Repräsentationskompetenzen lässt sich ableiten, dass räumliche Fähigkeiten zum Arbeiten mit Repräsentationen erforderlich sind. Um Repräsentationen zum Beschreiben chemischer Phänomene einzusetzen

und dabei die zugrundeliegenden molekularen Zusammenhänge und Vorgänge zu beschreiben, Muster in Repräsentationen zu erkennen oder die Perspektive auf ein Objekt zu wechseln (Tabelle 1), müssen räumliche Aspekte dargestellt und verstanden werden (Gilbert, 2008; Harle & Towns, 2011).

Räumliche Fähigkeiten (*spatial abilities* oder *visual processing*) nehmen somit in der Chemie, aber auch in anderen Naturwissenschaften, eine zentrale Rolle ein (Buckley et al., 2018; Stieff & Uttal, 2015; Stull et al., 2012; Wu & Shah, 2004). Bevor dies näher beleuchtet werden kann, müssen räumliche Fähigkeiten beschrieben und definiert werden.

3.1 Räumliche Fähigkeiten: Eine definitorische Annäherung

Eine frühe Beschreibung räumlicher Fähigkeiten liefert Galton (1879) mit dem Begriff der „Visualisierungsfähigkeit“ (*visualising faculty*), also der Fähigkeit, mentale Objekte nach Belieben abzuspeichern, zu verwerfen, zu manipulieren und aus verschiedenen Perspektiven zu betrachten (Galton, 1879). Diese Beschreibung präzisiert und erweitert Galton später (1880): Die Visualisierungsfähigkeit umfasst für ihn, klare und vollständige Bilder von Objekten mental abzurufen, diese aus verschiedenen Perspektiven zu betrachten, die Umrisse des Objekts zu skizzieren, mentale Bilder aus Beschreibungen zu konstruieren und zu manipulieren, sowie verschiedene Bilder eines Objekts zu einem generischen Bild zusammenzuführen (Galton, 1880). Etwa 100 Jahre später formuliert Lohman (1979) eine Definition für räumliche Fähigkeiten, die bis heute regelmäßig zitiert wird. Er beschreibt sie als „the ability to generate, retain, and manipulate abstract visual images“ (Lohman, 1979, S. 126 f.). Eine aktuellere Beschreibung von räumlichen Fähigkeiten gibt Newcombe (2010): „Spatial thinking concerns the locations of objects, their shapes, their relations to each other, and the paths they take as they move“ (Newcombe, 2010, S. 30).

Ähnlich wie bei der Klassifikation von Repräsentationen (Abbildung 2, Kapitel 2.1) gibt es keinen gemeinsamen Konsens zur Definition von räumlichen Fähigkeiten (Meehl, 2013). Die Definitionen und Beschreibungen weisen zwar Überschneidungen aus, unterscheiden sich aber in wesentlichen Punkten. Eine Gemeinsamkeit aller Definitionen ist, dass räumliche Fähigkeiten aus

mehreren unterschiedlichen Faktoren bestehen und als eigenständiger kognitiver Faktor verstanden werden. Daher erscheint es sinnvoll und angemessen, räumliche Fähigkeiten über ihre Faktoren zu definieren.

Hierfür liefert das Cattell-Horn-Carroll Modell (Schneider & McGrew, 2012) einen geeigneten theoretischen Rahmen. Um dieses Modell sowie die Einordnung der räumlichen Fähigkeiten darin nachvollziehen zu können, ist ein Blick auf die historische Evolution des Konstrukts und des Verständnisses menschlicher Intelligenz notwendig (Buckley et al., 2018).

3.2 Evolution der Theorien zur menschlichen Intelligenz und Einordnung der räumlichen Fähigkeiten

Zu Beginn des 20. Jahrhunderts führte Charles Spearman Studien zur Struktur der menschlichen Intelligenz durch. 1904 postulierte er einen allgemeinen Intelligenzfaktor (*general intelligence*, *g*-Faktor) und entwickelte daraus die Zwei-Faktoren-Theorie der menschlichen Intelligenz (Spearman, 1927). Basierend auf faktoranalytischen Analysen verschiedener Intelligenztests identifizierte er zwei Faktoren: die allgemeine Intelligenz (*general intelligence*, *g*-Faktor) und einen spezifischen Faktor (*specific ability*, *s*-Faktor). Die Analysen ergaben eine hohe Korrelation. Wer in einem Test gut abschnitt, erzielte auch in den anderen gute Ergebnisse - und umgekehrt. Aus diesem positiven Zusammenhang zwischen den Ergebnissen der einzelnen Testinstrumente schloss Spearman, dass ein allgemeiner Intelligenzfaktor, der *g*-Faktor, existieren müsse. Darüber hinaus postulierte er testspezifische Faktoren (*s*-Faktoren), die sich nicht auf andere Tests verallgemeinern lassen (Brody, 2000; Buckley et al., 2018; Williams et al., 2003; Willis et al., 2011).

Dieses Modell wurde von anderen Forscher:innen kritisch betrachtet, da der *g*-Faktor allein nur einen Teil der Varianz in den Analysen erklärt und somit für die Erklärung der Ergebnisse nicht ausreicht bzw. darauf hindeutet, dass weitere Faktoren notwendig sind und existieren müssen (Brody, 2000). Des Weiteren wird das Argument angeführt, dass die positive Korrelation aus verschiedenen Gründen auftreten kann und nicht zwingend einen gemeinsamen Faktor (*g*) impliziert (Willis et al., 2011). Diese Kritiker von Spearmans *g*-Theorie, etwa Thorndike (1927) und Thurstone (1938), vertraten multifaktorielle Theorien der Intelligenz. Thurstone (1938) hat zur Untersuchung der Struktur der

Intelligenz neue faktoranalytische Verfahren entwickelt, um die Anzahl und die Natur latenter Konstrukte der beobachtbaren Variablen zu identifizieren. Seine Ergebnisse deuten darauf hin, dass Intelligenz nicht als ein einheitlicher Faktor beschrieben werden sollte, wie Spearman (1927) es postulierte. Er konnte aus seinen Ergebnissen sieben primäre mentale Fähigkeiten (*primary mental abilities*) zur Beschreibung der Intelligenz ableiten. Darunter befindet sich u. a. ein Faktor zum räumlichen Vorstellungsvermögen (*spatial relation*) (Brody, 2000; Buckley et al., 2018; Thurstone, 1938; Willis et al., 2011).

Die Studien von Spearman (1904, 1927) und Thurstone (1938) bilden mitunter die Grundlage für Cattells *Gf-Gc*-Theorie (1943, 1963). Er beschreibt die Existenz zweier gleich relevanter Faktoren, die auf den *g*-Faktor laden: Er erläutert eine fluide Intelligenz (*fluid intelligence, g_f*), die eine biologische Komponente innehat und Prozesse und Fähigkeiten umfasst, die u. a. schnelles Denken und Problemlösen ermöglichen und mit dem Verlauf des Erwachsenenalters abnehmen. Darüber hinaus stellt er die kristalline Intelligenz (*crystallized Intelligence, g_c*) dar, die von durch Bildung und Kultur erworbenem Wissen und Fähigkeiten beeinflusst wird und mit dem Erwachsenenalter nicht abnimmt (Brody, 2000; Buckley et al., 2018; Cattell, 1943, 1963; Schneider & McGrew, 2012; Willis et al., 2011). Horn (1965), der Doktorand von Cattell, konnte mit seinen anknüpfenden Studien und Analysen zum einen Cattells *Gf-Gc*-Theorie bestätigen und zum anderen um weitere Faktoren ergänzen. Die zusätzlichen Faktoren korrespondieren mit jenen aus Thurstones (1938) Theorie der primären mentalen Fähigkeiten und laden gleichermaßen auf den *g*-Faktor wie die fluide (*G_f*) und die kristalline Intelligenz (*G_c*). Einer dieser erweiterten Faktoren ist die visuelle Intelligenz (*visual intelligence, G_v*), die eine Adaption des Faktors *spatial relation* (Thurstone, 1938) darstellt. Im Rahmen dieser Erweiterung der *Gf-Gc*-Theorie wurde die Notation *g_f* und *g_c* von Cattell (1943, 1963) zu *G_f* und *G_c* angepasst (Buckley et al., 2018; Horn, 1965; Horn & Cattell, 1966; Schneider & McGrew, 2012).

Einige Jahre später hat Carroll (1993) durch die faktoranalytische Untersuchung umfassender Studien seine *Three-Stratum Theory* postuliert, in der er die kognitiven Fähigkeiten in drei Ebenen unterteilt: die *general ability (g)* als *Stratum III*, die *broad abilities* als *Stratum II* und die *narrow abilities* als *Stratum I*. Die *narrow abilities* stellen hochspezifische Fähigkeiten dar, die sich jeweils zu einer der *broad abilities* zusammenfassen lassen. Die *broad abilities* wiederum bilden in der Summe die allgemeine Intelligenz (*g*) ab (Brody, 2000;

Carroll, 1993; Davidson & Downing, 2000; Willis et al., 2011). Hier lässt sich bereits eine sehr hohe Überschneidung zwischen dem *Gf-Gc-Modell* von Cattell und Horn (1966) und der zweiten und dritten Ebene von Carrolls *Three-Stratum Modell* (1993) erkennen.

3.3 Die Cattell-Horn-Carroll Theorie mit Verortung und Beschreibung der räumlichen Fähigkeiten

Bei einem Treffen namhafter Testtheoretiker:innen und -entwickler:innen wurden die Überschneidungen der erweiterten *Gf-Gc*-Theorie und der *Three-Stratum Theorie* sehr deutlich und führten zur Integration beider Ansätze hin zum Cattell-Horn-Carroll Modell (McGrew, 2009; Schneider & McGrew, 2012).

Tatsächlich hatte diese Synthese pragmatische Gründe. Das Modell sollte eine Brücke zwischen Theorie und Praxis ermöglichen, um einen gemeinsamen Rahmen für die Interpretation, Entwicklung und Anpassung von Intelligenztests zu schaffen, insbesondere für die hochspezialisierten Fähigkeiten des *Stratum I* (Davidson & Kemp, 2011). Das daraus resultierende Modell beinhaltet einen Faktor dritter Ordnung (*third-order factor*) – die allgemeine Intelligenz (den *g*-Faktor). Darauf laden 16 gleichwertige Faktoren zweiter Ordnung (*second-order factors*), auf die wiederum insgesamt 84 hochspezifische Faktoren erster Ordnung (*first-order factors*) laden (Buckley et al., 2018; McGrew, 2009; Schneider & McGrew, 2012). Die Struktur dieses hierarchischen Modells ist in Abbildung 8 schematisch veranschaulicht.

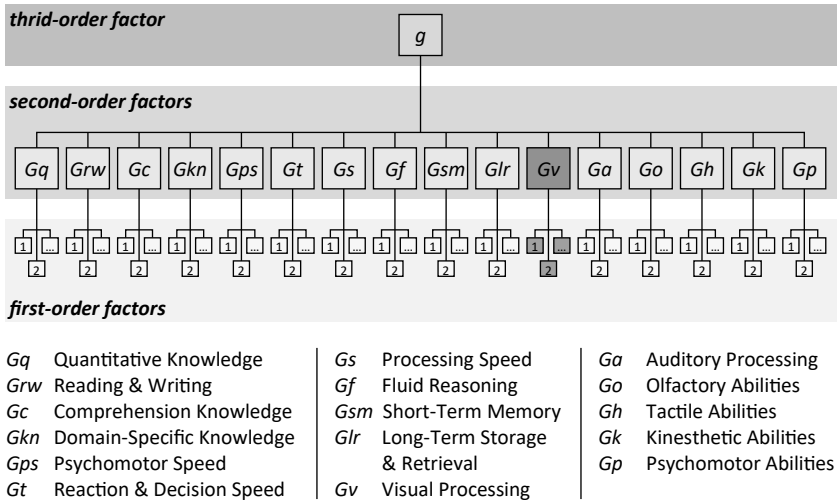


Abbildung 8: Hierarchische Struktur des Cattell-Horn-Carroll Modells mit Hervorhebung der räumlichen Fähigkeiten (*Visual Processing, Gv*) nach Schneider und McGrew (2012) sowie Buckley et al. (2018).

Das Cattell-Horn-Carroll Modell ist keineswegs statisch. Vielmehr ist es als dynamisches, sich weiterentwickelndes Modell zu betrachten (Flanagan & Dixon, 2014), das aktuell der bestmögliche Erklärungsansatz für die Struktur und Entwicklung von Fähigkeiten ist, die als Indikatoren menschlicher Intelligenz gefasst werden können (Horn & Blankson, 2012). Das Modell wird stetig vor dem Hintergrund neuer Theorien untersucht und auf Fehler geprüft. Darauf basierend werden Faktoren erster und zweiter Ordnung ergänzt, verworfen oder angepasst (Flanagan & Dixon, 2014). Beispiele dafür sind, mit Blick auf räumliche Fähigkeiten, die Arbeiten von Schneider und McGrew (2012) sowie von Buckley und Kolleg:innen (2018).

Die räumlichen Fähigkeiten stellen in dem Modell einen von 16 Faktoren zweiter Ordnung dar (Abbildung 8). Darauf laden elf Faktoren erster Ordnung, die im Folgenden als räumliche Faktoren bezeichnet werden. Schneider und McGrew (2012) definieren diese Fähigkeiten als “ability to make use of simulated mental imagery (often in conjunction with currently perceived images) to solve problems” (Schneider & McGrew, 2012, S. 129). Diese Definition dient jedoch lediglich als grobe Beschreibung, die alle entsprechenden Faktoren

miteinander verbindet. Eine präzisere Definition liefern die einzelnen räumlichen Faktoren selbst (Buckley et al., 2018). Die elf räumlichen Faktoren des Cattell-Horn-Carroll Modells sind in Tabelle 2 aufgelistet und definiert.

Tabelle 2: Definition der elf räumlichen Faktoren, die auf die räumlichen Fähigkeiten im Cattell-Horn-Carroll Modells laden (Flanagan & Dixon, 2014; McGrew, 2009; Schneider & McGrew, 2012).

Räumlicher Faktor	Definition
Closure Speed (CS)	Die Fähigkeit, unvollständige oder teilweise verdeckte visuelle Stimuli oder Muster schnell mental zu vervollständigen, ohne im Voraus zu wissen, worum es sich dabei handelt.
Flexibility of Closure (CF)	Die Fähigkeit, eine im Voraus bekannte Figur in einem komplexen, ablenkenden Muster zu erkennen und zu identifizieren.
Imagery (IM)	Die Fähigkeit, mental sehr lebendige Bilder zu erzeugen und diese zu manipulieren.
Length Estimation (LE)	Die Fähigkeit, Abstände oder Längen präzise ohne Messinstrumente zu schätzen und zu vergleichen.
Perceptual Alternations (PN)	Die Fähigkeit, flüssig und stetig zwischen unterschiedlichen visuellen Wahrnehmungen zu wechseln. <u>Beispiel:</u> Die Vorderseite eines Necker Würfels (Würfel, bei dem sämtliche Kanten als durchgezogene Linien dargestellt sind) zu wechseln.
Perceptual Illusions (IL)	Die Fähigkeit, sich von optischen Täuschungen mit geometrischen Figuren nicht täuschen zu lassen.
Serial Perceptual Integration (PI)	Die Fähigkeit, ein Muster zu erkennen, wenn Teile des Musters schnell hintereinander präsentiert werden. <u>Beispiel:</u> Einen Hirsch erkennen, wenn dieser hinter einer Reihe von Bäumen herläuft und er immer nur teilweise sichtbar ist.
Spatial Scanning (SS)	Die Fähigkeit, ein visuelles Feld oder Muster schnell zu überblicken und einen Weg dadurch zu finden.
Speeded Rotation (SR)	Die Fähigkeit, Probleme schnell zu lösen, indem einfache Bilder mental (zweidimensional) rotiert werden.
Visual Memory (MV)	Die Fähigkeit, eine mentale Repräsentation eines visuellen externen Stimulus zu erstellen, diese abzuspeichern und sie zu einem späteren Zeitpunkt wieder abzurufen.
Visualization (VZ)	Die Fähigkeit, komplexe Muster wahrzunehmen und mental zu simulieren, wie sie aussehen könnten, wenn sie verändert werden (z. B. durch Rotation, Größenänderung oder teilweise Abdeckung).

3.3.1 Aktuelle Forschungen und Entwicklung des Cattell-Horn-Carroll Modells

Faktoranalytische Untersuchungen mit dem Ziel, das Cattell-Horn-Carroll Modell bzw. die räumlichen Fähigkeiten weiter zu untersuchen und zu optimieren, bringen immer neue Ergebnisse zum Vorschein. Diese werden jedoch teilweise stark diskutiert und benötigen weitere Evidenz, um auf konsensualer Basis in das Modell integriert zu werden. Buckley und Kolleg:innen (2018) stellen als zentrale Kriterien zur Identifikation räumlicher Faktoren heraus, dass es sich dabei um spezifische mentale Fähigkeiten handelt, die visuelle Stimuli verarbeiten und losgelöst von semantischem Wissen sind. Des Weiteren müssen diese Faktoren klar von bereits bestehenden räumlichen Faktoren abgrenzbar sein.

Einige Faktoren erster Ordnung werden zwar nicht den räumlichen Fähigkeiten zugeordnet, stehen mit diesen jedoch in engem Zusammenhang. Beispielsweise wird der Faktor *Perceptual Speed (P)*, der die Fähigkeit beschreibt, räumlich nebeneinander oder separiert angeordnete visuelle Symbole oder Muster schnell miteinander zu vergleichen (Flanagan & Dixon, 2014), dem *Processing Speed (Gs)* zugeordnet (Abbildung 8). Es liegt jedoch eindeutige Evidenz vor, dass es sich bei Verwendung visuell-grafischer Stimuli um einen räumlichen Faktor handelt (Buckley et al., 2018; Pellegrino et al., 1987).

Weitere Untersuchungen, die besonders mit Blick auf die Chemie als interessant und relevant erscheinen, sind die Diskussion zur Differenzierbarkeit des räumlichen Faktors *Visualization (VZ)* und der zwei- sowie dreidimensionalen Rotation. Chemiker:innen müssen beispielsweise bei der Betrachtung von Reaktionsmechanismen, beim Vergleichen von Molekülen, der Bestimmung der absoluten Konfiguration von Molekülen und der Bestimmung der Isomerie in der Lage sein, Moleküle mental zu manipulieren und zu rotieren, sowohl zweis- als auch dreidimensional. Die zweidimensionale Rotation wird durch den räumlichen Faktor *Speeded Rotation (SR)* abgebildet (Tabelle 2); dieser beschränkt sich auf einfache Bilder und hat einen zeitlichen Faktor inne. Ein räumlicher Faktor zur dreidimensionalen Rotation existiert im Cattell-Horn-Carroll Modell nicht, sondern ist in den räumlichen Faktor *Visualization (VZ)* integriert, der die Manipulation, mitunter die Rotation, komplexer Bilder ohne zeitliche Komponente beschreibt (Tabelle 2). Evidenz zur Trennbarkeit der

zwei- und dreidimensionalen Rotation als mentale Faktoren liefert beispielsweise Moreau (2013). Durch Trainings zu beiden Rotationstypen konnte er zeigen, dass ein Training der dreidimensionalen Rotation auch die zweidimensionale Rotation fördert, jedoch nicht umgekehrt. Dieses Erkenntnis legt die Existenz eines räumlichen Faktors *Spatial Relation* nahe, der als Fähigkeit definiert wird, Probleme zeitunabhängig durch die (dreidimensionale) Rotation komplexer Bilder zu lösen. Es stellt sich nun die Frage, ob dieser Faktor als Indikator für *Visualization* (VZ) dienen kann oder als separater Faktor in das Modell integriert werden sollte (Buckley et al., 2018).

3.3.2 Statische und dynamische räumliche Fähigkeiten

Bei der Betrachtung der räumlichen Faktoren und der Messinstrumente zur Erfassung dieser Faktoren fällt auf, dass es sich um zweidimensionale, statische Messinstrumente handelt, die sich nur schwer mit einer dreidimensionalen, dynamischen Umwelt in Einklang bringen lassen (Larson, 1996). Daher erscheint es sinnvoll, räumliche Faktoren in statische und dynamische Faktoren zu differenzieren bzw. die räumlichen Faktoren zu erweitern. Hier können zunächst zwei Ansätze zum Verständnis von statischen und dynamischen räumlichen Faktoren unterschieden werden. Einerseits kann sich „statisch“ und „dynamisch“ auf mentale Prozesse beziehen, beispielsweise ist die mentale Rotation eines Objekts als dynamischer Prozess zu verstehen (Buckley et al., 2018; Newcombe & Shipley, 2015). Andererseits kann sich „statisch“ und „dynamisch“ auf die Art des externen Stimulus beziehen. Statische Stimuli meinen hierbei unbewegte Objekte während dynamische Stimuli bewegliche Stimuli intendieren (Buckley et al., 2018; Pellegrino et al., 1987). Letzteres ist dabei das mehrheitlich angenommene Verständnis von statischen und dynamischen Fähigkeiten, zu dem bereits Evidenz vorliegt. Studien von Roff (1952), Contreras und Kolleg:innen (2003) sowie D'Oliveira (2004) beispielsweise konnten statische und dynamische räumliche Faktoren faktoranalytisch trennen und schlagen u. a. Faktoren wie *Movement Detection*, *Speed Judgement* und *Directional Judgement* als dynamische Faktoren vor.

3.3.3 Kritische Betrachtung des Cattell-Horn-Carroll Modells

Das Cattell-Horn-Carroll Modell wird zwar von einer Großzahl der Forscher:innen als aktuell bester Erklärungsansatz zur Struktur und Entwicklung von Intelligenz gesehen (Flanagan & Dixon, 2014), doch es gibt auch kritische Stimmen. So wird die uneinheitliche Benennung, Definition und Zuordnung von räumlichen Faktoren und Messinstrumenten kritisiert, die zu Unklarheiten bezüglich der Zugehörigkeit von Messinstrumenten führt (D'Oliveira, 2004). Ferner werden die Messinstrumente selbst kritisiert. Die Struktur des Modells basiert auf der faktoranalytischen Auswertung standardisierter psychometrischer Messinstrumente, zu deren Lösung beispielsweise ein räumlicher Faktor (Tabelle 2) als Voraussetzung erachtet wird. Der *Revised Purdue Spatial Visualization Test: Visualization of Rotation (Revised PSVT:R; Yoon, 2011b)* wird beispielsweise zur Messung der Fähigkeit *Spatial Relation* eingesetzt, um Probleme mithilfe dreidimensionaler Rotation komplexer Objekte zu lösen (Abbildung 9).

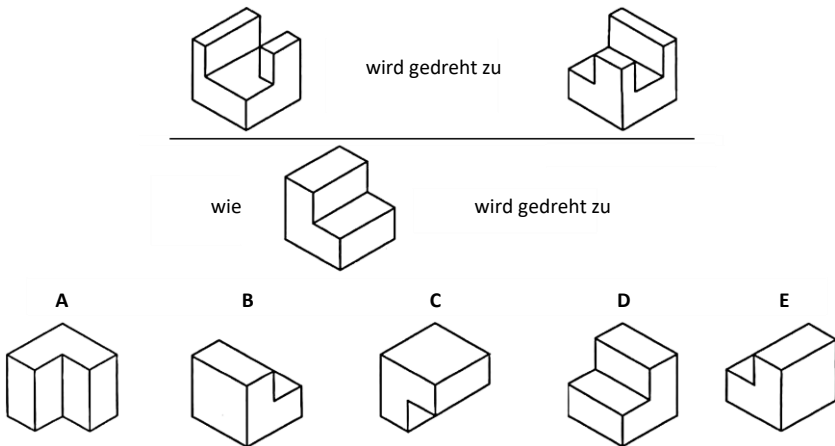


Abbildung 9: Beispiel-Item aus dem *Revised Purdue Spatial Visualization Test: Visualization of Rotation* (Yoon, 2011b) zur Messung der dreidimensionalen Rotationsfähigkeit.

Aktuelle Studien zum *Mental Rotation Test* von Vandenberg und Kuse (1978), einem eng verwandten Messinstrument des *Revised PSVT:R*, zeigen, dass sich

die Lösungsstrategien bei Proband:innen durchaus unterscheiden und nicht immer eine Rotation genutzt werden muss, um ein vermeintliches Rotationsproblem zu lösen. Strategien, wie den Blickwinkel auf das Objekt und somit die Position der eigenen Person zum Objekt zu verändern, anstatt das Objekt an sich zu rotieren, ist eine weitere Möglichkeit, die Aufgaben zu lösen.

Auch analytische Strategien haben sich als effektiv erwiesen (Hegarty, 2018). Bei der Lösung von Aufgaben mit räumlichen Anforderungen können somit holistische und analytische Strategien differenziert werden. Zu den holistischen Strategien zählen zum einen die *move-object* Strategie, bei der ein Objekt mental rotiert wird, was der *Revised PSVT:R* und der *Mental Rotation Test* erheben sollen; zum anderen wird die *move-self* Strategie inkludiert, bei der die eigene Position zum Objekt verändert wird. Bei der analytischen Strategie, auch *key-feature* Strategie genannt, werden Vergleiche zwischen den Objekten herangezogen, zentrale diesbezügliche Aspekte betrachtet und logisch-schlussfolgernde Denkprozesse genutzt (Maresch, 2014; Wang & Carr, 2014). Somit können verschiedene zielführende und erfolgreiche Lösungsstrategien zur Bewältigung von Aufgaben in Messinstrumenten wie dem *Revised PSVT:R* herangezogen werden.

Trotz der berechtigten Kritik am Cattell-Horn-Carroll Modell und den psychometrischen Messinstrumenten bilden diese die adäquateste Ausgangslage zur Untersuchung der räumlichen Fähigkeiten. Das zentrale Argument ist hier die Vergleichbarkeit der Ergebnisse mit der bisherigen Forschung, da diese mehrheitlich auf die standardisierten Messinstrumente zurückgreift, die im Kontext des Cattell-Horn-Carroll Modells zur Erfassung räumlicher Faktoren Erwähnung finden.

3.3.4 Geschlechtsunterschiede bei räumlichen Fähigkeiten

Unter den kognitiven Faktoren sind die räumlichen Fähigkeiten der Faktor, der die stabilsten und beständigsten Unterschiede mit Blick auf das Geschlecht aufweist. Dabei schneiden Männer in Tests zur Messung räumlicher Faktoren durchschnittlich besser ab als Frauen (Reilly et al., 2017). Besonders mit Blick auf die Fähigkeit *Spatial Relation*, die dreidimensionale Rotationsfähigkeit, werden diese Ergebnisse häufig berichtet (z. B. Estes & Felker, 2012; Hegarty, 2018; Linn & Petersen, 1985; Lippa et al., 2010). Interessant ist die Metastudie

von Voyer und Kolleg:innen (1995), da diese über verschiedene räumliche Faktoren hinweg eine Zunahme des Effekts mit zunehmendem Alter zeigen konnte. Dieser Effekt ist bei dem Faktor *Spatial Relation* am größten: $d = 0.33$ für unter 13-Jährige, $d = 0.45$ für 13- bis 18-Jährige und $d = 0.66$ für über 18-Jährige. Ein möglicher Erklärungsansatz für diese Entwicklung ist der Einfluss von Sexualhormonen (Androgene und Östrogene) auf die Entwicklung des Gehirns und das Verhalten während der Pubertät. Aktuell gibt es verschiedene Indikatoren, dass Androgene, die vornehmlich von Männern produziert werden, die Entwicklung der räumlichen Fähigkeiten positiv beeinflussen. Weitere Faktoren die diese Entwicklung beeinflussen könnten, sind stereotypisch behaftete Spielzeuge, z. B. Bauklötze und Computerspiele, mit denen Jungen vermeintlich eher als Mädchen spielen (Reilly et al., 2017). Daneben gibt es weitere Erklärungsansätze, die hier nicht weiter aufgeführt werden. Eine detaillierte Zusammenfassung bieten beispielsweise Reilly et al. (2017).

3.4 Die Rolle der räumlichen Fähigkeiten in der Chemie und anderen Naturwissenschaften – Ein Einblick in den aktuellen Forschungsstand

Die Annahme, dass räumliche Fähigkeiten neben Repräsentationskompetenzen und Fachwissen zum Arbeiten mit Repräsentationen benötigt werden, lässt sich durchaus mit umfassenden Studien belegen. In der Chemie (besonders der Organischen Chemie) und weiteren Naturwissenschaften haben sich gut ausgeprägte räumliche Fähigkeiten mit einem positiven Effekt auf die Leistung von Lernenden erwiesen (Buckley et al., 2018; Stieff & Uttal, 2015; Uttal et al., 2013; Uttal & Cohen, 2012; Yang et al., 2020). Auch mit Blick auf Problemlösefähigkeiten zeigt sich ein hoher signifikanter Zusammenhang mit räumlichen Fähigkeiten (Carter et al., 1987; Pribyl & Bodner, 1987). An dieser Stelle muss jedoch betont werden, dass die Aussage „gute räumliche Fähigkeiten“ sehr vage ist. Die räumlichen Fähigkeiten setzen sich, wie im Cattell-Horn-Carroll Modell (Abbildung 8) dargestellt, aus einer Vielzahl räumlicher Faktoren zusammen (Tabelle 2). Darauf basierend sollte eine differenzierte und keine verallgemeinerte Betrachtung vorgenommen werden. Ein Großteil der Studien nutzt ein bis drei psychometrische Messinstrumente zur Messung der räumlichen Faktoren und trifft damit eine verallgemeinerte Aussage über

die räumlichen Fähigkeiten. Einige Studien berichten jedoch auch über spezifische räumliche Faktoren im Zusammenhang mit der Chemie:

- ***Spatial Relation***: Die Fähigkeit zur dreidimensionalen mentalen Rotation ist in diesem Kontext der am meisten untersuchte Faktor (Kenneally & Bentley, 2024; Stieff et al., 2018) und wird häufig als wichtigster Prädiktor für die Leistung in der Chemie und anderen Naturwissenschaften benannt. Besonders mit Blick auf fachliche Problemlöseprozesse und die Bewältigung komplexer Aufgaben stellt sich der Faktor *Spatial Relation* als wichtiger Prädiktor heraus (Bodner & Guay, 1997; Carter et al., 1987; Harle & Towns, 2011; Harris et al., 2013; Sorby, 2009).
- ***Speeded Rotation***: Neben dem Zusammenhang zwischen der Chemieleistung und der dreidimensionalen mentalen Rotation wurde auch ein (schwacher) Zusammenhang mit der zweidimensionalen mentalen Rotationsfähigkeit gefunden (Harris et al., 2013; Wu & Shah, 2004).
- ***Visualization***: Die Fähigkeit, Repräsentationen mental zu manipulieren (hiermit ist nicht die mentale Rotation gemeint, diese wird differenziert mit den Faktoren *Speeded Rotation* und *Spatial Relation* betrachtet), zeigt ebenfalls einen signifikanten Zusammenhang mit der Leistung in den Naturwissenschaften (Harris et al., 2013; Wai et al., 2009).
- ***Flexibility of Closure***: Bekannte Figuren in ablenkenden Mustern zu erkennen, zeigt einen (schwachen) positiven Zusammenhang mit der Leistung in (Organischer) Chemie, aber auch bei fachlichen Problemlöseprozessen und bei der Bewältigung komplexer Aufgaben, ähnlich wie der Faktor *Spatial Relation* (Bodner & McMillen, 1986; Carter et al., 1987; Hodgkiss et al., 2018; Wu & Shah, 2004).
- ***Spatial Scaling***: Hodgkiss und Kolleg:innen (2018) konnten zeigen, dass die Fähigkeit, über Relationen zwischen einer Repräsentation und einer Referenzrepräsentation nachzudenken (Newcombe et al., 2019), ebenfalls mit der Leistung in der Chemie zusammenhängt. Dieser Faktor wird jedoch erst seit Kurzem in diesem Zusammenhang untersucht, weshalb die Verortung des Faktors aktuell im Cattell-Horn-Carroll Modell empirisch nicht bestätigt ist.

Mentale Rotationsprozesse, die Darstellung von dreidimensionalen Strukturen, die Identifikation funktioneller Gruppen und die Bestimmung der absoluten Konfiguration von Molekülen sind nur einige Beispiele, die typisch für die

Chemie sind und Herausforderungen für Lernende darstellen (Salame & Kabir, 2022). Barke und Engida (2001) empfehlen daher, basierend auf den Ergebnissen ihrer und anderer Studien, räumliche Strukturen so früh wie möglich in den Unterricht zu integrieren, um so einerseits den Umgang mit Repräsentationen und das Chemieverständnis zu fördern und andererseits, um die Entwicklung der räumlichen Fähigkeiten frühestmöglich zu unterstützen.

Basierend auf den Vergleichen von Noviz:innen und Expert:innen konnte Stieff (2007) zeigen, dass beide Gruppen über vergleichbare Fähigkeiten bei der mentalen Rotation (*Spatial Relation*) verfügen, sich aber signifikant beim Umgang und der Bestimmung von Symmetrien dreidimensionaler Moleküle unterscheiden. Er konnte Evidenz liefern, dass Expert:innen die mentale Rotation mithilfe von semantischem Wissen über Moleküle substituieren können (Stieff, 2007; Uttal & Cohen, 2012).

3.5 Trainierbarkeit räumlicher Fähigkeiten

In den vorangegangenen Kapiteln wurde bereits an einigen Stellen angesprochen, dass räumliche Fähigkeiten in verschiedenen Studien erfolgreich gefördert werden konnten. Da sich räumliche Fähigkeiten als wichtiger Prädiktor in den Naturwissenschaften herausgestellt haben, erscheint es sinnvoll, diese Fähigkeiten zu trainieren und zu fördern. Zur Förderung von räumlichen Fähigkeiten liegen bereits zahlreiche Studien und Meta-Analysen vor. Die Interventionen und Trainings können dabei in drei Kategorien eingeordnet werden: Das Nutzen von Computerspielen, speziell entwickelte Kurse, die räumliche Fähigkeiten indirekt adressieren, und Trainings mithilfe spezifischer räumlicher Aufgaben (Uttal et al., 2013). Die Ergebnisse der Studien unterstützen die Hypothesen, dass durch Interventionen und Trainings zum einen räumliche Fähigkeiten und zum anderen die Leistung in naturwissenschaftlichen Fächern verbessert werden können (Carlisle et al., 2015; Hawes et al., 2022; Merchant et al., 2013; Sorby, 2009; Stieff & Uttal, 2015; Uttal et al., 2013; Yang et al., 2020). Stieff und Uttal (2015) betonen in diesem Zusammenhang, dass selbst bei geringen Effekten solcher Interventionen und Trainings besonders Lernenden mit schwach ausgeprägten räumlichen Fähigkeiten der Weg für eine Karriere in den Naturwissenschaften eröffnet werden kann, der ihnen ansonsten verwehrt bliebe.

Ein Beispiel für ein solches Training beschreiben Sorby et al. (2013). Ihr Ziel war die Förderung der räumlichen Fähigkeiten von Studieneinsteiger:innen der Ingenieurwissenschaften mit schwach ausgeprägten räumlichen Fähigkeiten. Dazu haben sie mithilfe eines Arbeitsbuchs und einer Multimedia-Software in zehn Modulen von je 90 Minuten unterschiedliche Aspekte räumlicher Fähigkeiten abgedeckt. Mit einem Prä-Post-Design haben sie die Effektivität des Trainings u. a. anhand des *Purdue Spatial Visualization Test: Rotations* (PSVT:R, Kapitel 6.2.2.8) gemessen. Die Ergebnisse des PSVT:R im Prätest haben sie zur Einteilung in Interventions- und Kontrollgruppe genutzt: Studierende mit Testwerten < 60 % wurden der Interventionsgruppe, Studierende mit Testwerten ≥ 60 % werden der Kontrollgruppe zugeordnet. Die Module für die Interventionsgruppe behandeln nach Sorby et al. (2013) folgende Trainings:

1. **Flächen und Körper in Rotation** – Durch das Drehen zweidimensionaler Objekte um eine bestimmte Achse sollen dreidimensionale Flächen und Körper erzeugt werden.
2. **Kombinieren von Objekten** – Neue Objekte werden durch das Schneiden, Verbinden oder Überschneiden zweier Objekte gestaltet.
3. **Isometrisches Skizzieren** – Kodierte Pläne werden verwendet, um Objekte zu definieren, die aus Würfeln und Blöcken aufgebaut sind. Anschließend werden diese aus verschiedenen Eckansichten skizziert.
4. **Orthographische Projektion mit normalen Flächen** – Die Konstruktion der Ober-, Vorder- und Seitenansicht von Objekten wird vermittelt, wobei auch verdeckte Merkmale von Objekten einbezogen werden.
5. **Orthographische Projektion mit schrägen und gekrümmten Flächen** – Die Konstruktion der Ober-, Vorder- und Seitenansicht von Objekten mit schrägen oder gekrümmten Flächen wird vermittelt.
6. **Entwicklung von flachen Mustern** – Die Umwandlung von zwei- in dreidimensionale flache Muster, die zu dreidimensionalen Objekten gefaltet werden können, wird behandelt.
7. **Rotation von Objekten um eine Achse** – Die Rotation von Objekten um eine bestimmte Achse wird durch das Skizzieren der rotierten Objekte gefördert.
8. **Drehen von Objekten um mehrere Achsen** – In Fortsetzung des vorangehenden Moduls werden Objekte um mindestens zwei Achsen rotiert und anschließend skizziert.

9. **Objektspiegelung und Symmetrie** – Das Konzept der Spiegelung eines Objekts in einer Ebene und das Konzept der Objektsymmetrie werden hier behandelt.
10. **Querschnitte** – Die Querschnitte durch ein Objekt werden für bestimmte Schnittebenen durch das Objekt vorhergesagt.

Ein zentrales Ergebnis der Analysen ist, dass die Intervention ein signifikanter Prädiktor für die Leistung im PSVT:R im Posttest mit einer kleinen, aber dennoch bedeutsamen Effektstärke ($d = 0.22$) ist. Die Verzerrung durch den Wert des Prätests als Zuweisungsvariable wurde in den Analysen berücksichtigt. Basierend darauf sprechen sie eine Empfehlung aus, dass Studieneinsteiger:innen mit schwach ausgeprägten räumlichen Fähigkeiten im ersten Semester identifiziert und gezielt beim Aufbau der räumlichen Fähigkeiten unterstützt werden sollten (Sorby et al., 2013), ähnlich wie es auch Uttal et al. (2013) basierend auf den Befunden ihrer Metastudie zur Förderbarkeit räumlicher Fähigkeiten empfehlen.

4

Zielsetzungen und Fragestellungen

Repräsentationskompetenzen spielen in den Naturwissenschaften und besonders in der Chemie eine zentrale Rolle. Naturwissenschaftler:innen sind auf Repräsentationen angewiesen, um Phänomene zu erklären und über sie zu kommunizieren. Dafür müssen sie in der Lage sein, diese zu *interpretieren*, zu *übersetzen*, zu *konstruieren* und letztlich mit ihnen zu *argumentieren*, die *epistemologische Perspektive* zu verstehen und sie *kritisch zu vergleichen* (Tabelle 1) – kurz gesagt: Sie müssen Repräsentationskompetenzen entwickeln, um erfolgreich mit Repräsentationen zu arbeiten (Kozma & Russell, 1997, 2005). Diese Fähigkeiten spielen auch bei der Überwindung des Repräsentationsdilemmas eine wesentliche Rolle, indem sie die Möglichkeit eröffnen, chemische Konzepte mithilfe von Repräsentationen zu erlernen (Rau, 2018).

Die von Kozma und Russell (1997, 2005) postulierten Repräsentationskompetenzen stellen eine gute Basis für die Forschung im Kontext von Repräsentationen dar, aber auch, um diese im Lehr-Lern-Prozess zu adressieren und Lerngelegenheiten zu ihrer Förderung zu entwickeln. Sowohl mit Blick auf die Forschung als auch auf die Schaffung von Lerngelegenheiten muss jedoch die Struktur der Repräsentationskompetenzen geklärt werden, um den Zusammenhang der einzelnen Fähigkeiten zu verstehen und diese adäquat adressieren und umsetzen zu können.

Aktuelle Studien liefern zwar Evidenz dafür, dass es sich bei den einzelnen Fähigkeiten der Repräsentationskompetenzen um untrennbare, stark miteinander verknüpfte Fähigkeiten handelt (Nitz, Ainsworth, et al., 2014; Scheid et al., 2018; Taskin et al., 2015; Ward et al., 2025), sie weisen jedoch Schwächen mit Blick auf die eingesetzten Messinstrumente auf. Daher müssen Aussagekraft und Validität der Ergebnisse kritisch hinterfragt werden. Das *Chemical Representation Inventory* (Taskin et al., 2015) und der *Test of Representational Competence* (Sim & Daniel, 2014) beispielsweise berücksichtigen die Repräsentationskompetenzen nach Kozma und Russell (1997, 2005), differenzieren jedoch nicht gezielt mit Blick auf die einzelnen Fähigkeiten (Tabelle 1). Zusätzlich

messen einzelne Items der Instrumente potenziell eher Fachwissen, beispielsweise wenn nach der Strukturformel einer speziellen Verbindung gefragt wird. Ist die Verbindung nicht bekannt, kann auch die Struktur nicht skizziert werden, obwohl die Fähigkeiten gegebenenfalls vorhanden sind. Letzteres gilt auch für das *Organic Chemistry Representational Competence Assessment* (Ward et al., 2025). Dadurch wird deutlich, dass Repräsentationskompetenzen und Fachwissen stark zusammenhängen, auch wenn einige Studien Evidenz zur Trennbarkeit der Konstrukte liefern (Edelsbrunner et al., 2023; Nitz, Ainsworth, et al., 2014).

„Chemistry is a visual science“ (Wu & Shah, 2004, S. 465). Mit diesen Worten fassen Wu und Shah (2004) das Fach Chemie gut zusammen. Neben Repräsentationskompetenzen und Fachwissen spielt auch das räumliche Denken eine wichtige Rolle beim Arbeiten mit Repräsentationen. Diese Fähigkeiten werden als wichtige Prädiktoren für Leistungen im Fach Chemie angesehen. Die Aussage ist jedoch vage, da sich die räumlichen Fähigkeiten aus vielen unterschiedlichen Faktoren zusammensetzen, wenn man das Cattell-Horn-Carroll Modell betrachtet. Beispielsweise haben sich bereits die Faktoren *Flexibility of Closure*, *Spatial Relation*, *Visualization* und *Speeded Rotation* (Kapitel 3.3) als wichtige räumliche Faktoren und Prädiktoren für die Naturwissenschaften und vor allem die Chemie erwiesen. Bei genauer Betrachtung der Studien, die räumliche Fähigkeiten im naturwissenschaftlichen Kontext untersuchen, fällt jedoch auf, dass diese häufig nur ein bis drei räumliche Faktoren erheben und auf dieser Basis verallgemeinerte Aussagen über räumliche Fähigkeiten treffen.

Forschungsanliegen I – Entwicklung und Evaluation des CRI:TIC: Untersuchung der Struktur der Repräsentationskompetenzen und Zusammenhang mit dem Fachwissen in der Chemie

Da die Struktur der Repräsentationskompetenzen nicht abschließend geklärt ist und bisherige Ergebnisse nur mäßig belastbar sind, besteht das Ziel des ersten Forschungsanliegens darin, entsprechende Untersuchungen zu vertiefen. Dabei soll jedoch nicht die Struktur der gesamten Fähigkeiten der Repräsentationskompetenzen geprüft, sondern ein Fokus auf die lower-level skills *Interpretation*, *Translation* und *Konstruktion* gelegt werden. Diese drei Fähigkeiten stellen basale Fähigkeiten dar, die das Arbeiten mit Repräsentationen

erst ermöglichen und somit auch die Grundlage für die Entwicklung der higher-level skills *Vergleich und Kritik*, *Epistemologie* und *Argumentation* bilden (Gilbert, 2008; Gurung et al., 2022; Ward et al., 2025). Zudem soll hier die Trennbarkeit bzw. der Zusammenhang von Repräsentationskompetenzen und Fachwissen in der Chemie untersucht werden, da diese Differenzierung im Kontext des Lernens mit Repräsentationen eine zentrale Rolle spielt.

Um dieses Ziel zu erreichen, ist zunächst die Entwicklung eines adäquaten Instruments nötig, da, wie zuvor angemerkt, andere Messinstrumente nicht zwischen den einzelnen Fähigkeiten differenzieren und das Repräsentationsdilemma nicht adäquat berücksichtigen. Einzelne Items messen in diesen Instrumenten potenziell eher Fachwissen als Repräsentationskompetenzen. Daher wurde das *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC) (Nickel et al., 2025) für dieses Projekt entwickelt und evaluiert. So wird die Messung der Repräsentationskompetenzen klar von der jener des chemischen Fachwissens abgegrenzt und operationalisiert. Das erste Forschungsanliegen wird von folgenden Forschungsfragen geleitet:

FF1.1 Inwieweit lassen sich die drei lower-level skills der Repräsentationskompetenzen *Interpretation*, *Translation* und *Konstruktion* empirisch trennen?

FF1.2 Inwieweit lassen sich Repräsentationskompetenzen und Fachwissen empirisch trennen?

Der aktuelle Forschungsstand deutet eher darauf hin, dass Repräsentationskompetenzen als Konstrukt aus eng miteinander verknüpften Fähigkeiten zu verstehen ist (z. B. Taskin et al., 2015; Ward et al., 2025). Diese Befunde sind aufgrund von Mängeln der eingesetzten Messinstrumente jedoch nur mäßig belastbar, vor allem mit Blick auf das Repräsentationsdilemma, also den Umgang mit chemischem Fachwissen in den Erhebungen. Ausgehend von diesen Befunden und Erkenntnissen lassen sich folgende Hypothesen formulieren:

Bei einer adäquaten Berücksichtigung des Repräsentationsdilemmas, d. h. einer klaren Abgrenzung und Operationalisierung der Repräsentationskompetenzen (nach Kozma & Russell, 1997, 2005) und des Fachwissens in der Chemie, können ...

- H1.1** die drei lower-level skills *Interpretation*, *Translation* und *Konstruktion* empirisch voneinander getrennt werden.
- H1.2** Fachwissen und Repräsentationskompetenzen empirisch voneinander getrennt werden.

Forschungsanliegen II – Einfluss der räumlichen Fähigkeiten beim Bearbeiten repräsentationsbasierter Chemieaufgaben

Aktuell fehlen Studien, die eine hohe Bandbreite an räumlichen Faktoren erheben und deren Zusammenhang mit der Leistung beim Bearbeiten repräsentationsbasierter Chemieaufgaben untersuchen, aber auch den Zusammenhang mit dem Fachwissen betrachten. Das Cattell-Horn-Carroll Modell bietet aufgrund seiner hierarchischen Struktur einen guten Rahmen für die geplante Untersuchung, da es einerseits durch die deutliche Differenzierung einzelner räumlicher Faktoren eine klare Struktur vorgibt und andererseits Messinstrumente zur Erfassung der räumlichen Faktoren empfiehlt, da das Modell durch faktoranalytische Untersuchung dieser entstanden ist. Das zweite Forschungsanliegen wird daher durch folgende Forschungsfragen angeleitet:

- FF2.1** Welcher korrelative Zusammenhang lässt sich zwischen Repräsentationskompetenzen bzw. Fachwissen und räumlichen Faktoren in der Chemie identifizieren?
- FF2.2** Welche räumlichen Faktoren sind beim Bearbeiten repräsentationsbasierter Aufgaben prädiktiv für die Leistung?

Aus der bisheriger Forschung lässt sich ableiten, dass neben dem Fachwissen (Cook et al., 2008; Rau, 2018) alle räumlichen Faktoren, die im Rahmen der Cattell-Horn-Carroll Theorie definiert werden, in einem bestimmten Ausmaß mit den Repräsentationskompetenzen im Zusammenhang stehen (Buckley et al., 2018; Uttal et al., 2013). Die Faktoren *Spatial Relation*, *Speeded Rotation*, *Visualization* und *Flexibility of Closure* sind in diesem Kontext am besten untersucht und haben sich in verschiedenen Arbeiten als zentrale Faktoren im Zusammenhang mit der Leistung und vor allem beim Arbeiten mit Repräsentationen in der Chemie herausgestellt (Bodner & Guay, 1997; Bodner & McMillen, 1986; Carter et al., 1987; Harle & Towns, 2011; Harris et al., 2013;

Hodgkiss et al., 2018; Kenneally & Bentley, 2024; Sorby, 2009; Stieff et al., 2018; Wai et al., 2009; Wu & Shah, 2004). Daraus lassen sich folgende Hypothesen ableiten:

- H2.1** Bei der zugrunde gelegten Operationalisierung lassen sich Korrelationen zwischen sämtlichen erhobenen räumlichen Faktoren und den Repräsentationskompetenzen identifizieren.
- H2.2** Die räumlichen Faktoren *Spatial Relation*, *Speeded Rotation*, *Visualization* und *Flexibility of Closure* nehmen neben dem Fachwissen eine zentrale Rolle beim Bearbeiten repräsentationsbasierter Chemieaufgaben ein und sind signifikante Prädiktoren für die Repräsentationskompetenzen.

Mit Blick auf den Zusammenhang zwischen Fachwissen und räumlichen Faktoren lassen sich weniger klare Hypothesen ableiten. In bisherigen Untersuchungen wird zwar der Zusammenhang zwischen räumlichen Fähigkeiten und der Leistung in der Chemie betont, daraus lässt sich jedoch nicht ableiten, inwieweit diese mit dem Fachwissen, wie es in dieser Arbeit operationalisiert und gemessen wird, zusammenhängen. Daher wird der korrelative Zusammenhang zwischen Fachwissen und räumlichen Faktoren hier eher explorativ untersucht.

Auf Basis der Theorie und den Hypothesen lässt sich ein Modell ableiten, dass im Rahmen dieses Forschungsanliegens (FF2.2) untersucht und geprüft werden soll (Abbildung 10). Dieses Modell berücksichtigt neben dem Einfluss des Fachwissens und der räumlichen Faktoren auf die Repräsentationskompetenzen auch die Interkorrelationen der räumlichen Faktoren und eignet sich daher, um Forschungsfragen 2.2 zu beantworten.

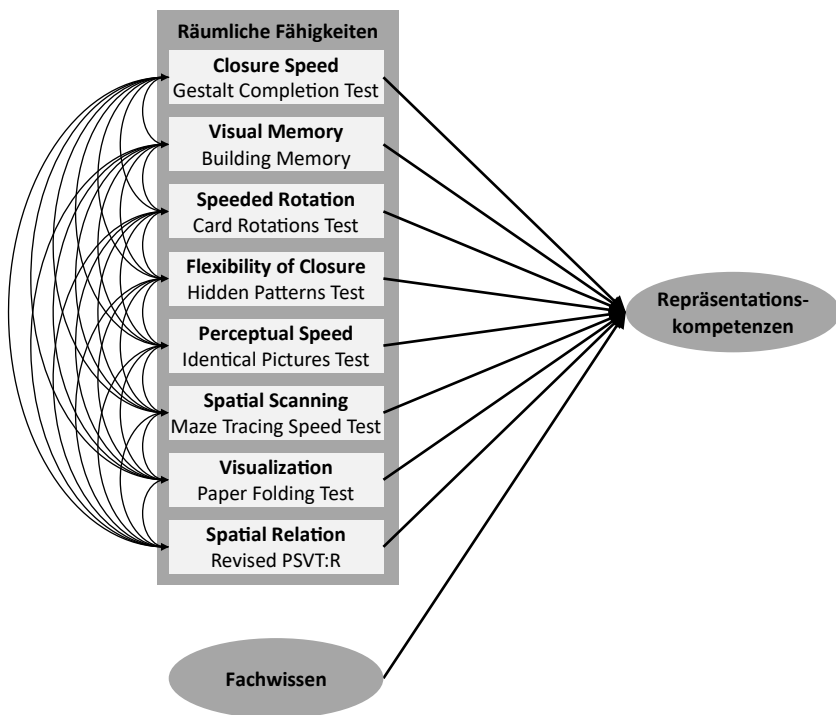


Abbildung 10: Theoretisch abgeleitetes Modell, dass im Rahmen von Forschungsanliegen II untersucht und geprüft wird.

Forschungsanliegen III - Hindernisse beim Bearbeiten repräsentationsbasierter Chemieaufgaben

Die ersten beiden Forschungsanliegen liefern einen wichtigen Einblick in die Struktur der lower-level skills der Repräsentationskompetenzen und vertiefen die Fragen, inwieweit sich die Repräsentationskompetenzen vom Fachwissen abgrenzen lassen und ob und in welchem Ausmaß räumliche Faktoren Prädiktoren für die Leistung bei der Bearbeitung repräsentationsbasierter Chemieaufgaben darstellen. Durch diese Untersuchungen kann jedoch kein Einblick in die kognitiven Prozesse beim Bearbeiten repräsentationsbasierter Chemieaufgaben gewonnen werden. Die Definitionen der lower-level skills

lassen bereits darauf schließen, dass die Anforderungen bei der *Interpretation*, *Translation* und *Konstruktion* unterschiedliche kognitive Prozesse und Fähigkeiten erfordern. Dies soll das Ziel des dritten Forschungsanliegens sein. Um ein besseres Verständnis für die kognitiven Prozesse und Fähigkeiten beim *Interpretieren*, *Übersetzen* und *Konstruieren* von Repräsentationen zu erlangen, sollen Hindernisse identifiziert werden, über die auf zentrale und notwendige Fähigkeiten und Kompetenzen im Bearbeitungsprozess geschlossen werden kann. Es ist das Ziel, damit die Ergebnisse der Forschungsanliegen I und II anzureichern. Einerseits soll die Rolle des Fachwissens und der räumlichen Fähigkeiten im Kontext des Lernens mit Repräsentationen vertieft untersucht werden, andererseits gilt es zu ermitteln, inwieweit sich weitere notwendige Faktoren für die Bearbeitung repräsentationsbasierter Chemieaufgaben identifizieren lassen. Folgende Forschungsfragen leiten sich daraus ab:

FF3.1 Welche zentralen Hindernisse lassen sich beim *Interpretieren*, *Übersetzen* und *Konstruieren* von chemischen Repräsentationen identifizieren?

FF3.2 Inwieweit unterscheiden sich die Hindernisse bzw. die kognitiven Prozesse und Fähigkeiten bei der *Interpretation*, *Translation* und *Konstruktion* von chemischen Repräsentationen?

5

Forschungsanliegen I – Entwicklung und Evaluation des CRI:TIC: Untersuchung der Struktur der Repräsentationskompetenzen und des Zusammenhangs mit dem Fachwissen in der Chemie

In diesem Kapitel geht es um die Vorstellung des ersten Forschungsanliegens. Dazu werden zunächst das Design und die Methoden zur Untersuchung der Forschungsfragen sowie die Entwicklung des verwendeten Messinstruments, des CRI:TIC, beschrieben. Es folgen Darstellung und Diskussion der Ergebnisse.

5.1 Design

Um die Forschungsfrage zu adressieren, wurde eine *quantitative Querschnittsstudie* geplant, in deren Rahmen sich Leistungstests besonders eignen, um mithilfe manifester Variablen (Items) ein latentes Konstrukt wie die Repräsentationskompetenzen oder das Fachwissen zu erfassen. Da die Studie theorie- und hypothesengeleitet ist und mit einer nicht-randomisierten Stichprobe durchgeführt wurde, handelt es sich um ein *explanatives, nicht-experimentelles* Studiendesign (Döring, 2023f).

Forschungsfrage 1.1

Zur Beantwortung der Forschungsfrage 1.1, inwieweit die drei lower-level skills *Interpretation Translation* und *Konstruktion* empirisch trennbar sind, wurden theoriegeleitet drei Modelle abgeleitet. Anschließend folgten Prüfung und Vergleich. Dabei stellt das dreidimensionale Modell die Gegenhypothese dar, während das ein- und zweidimensionale Modell Nullhypothesen abbilden:

1. **Eindimensionales Modell (1D)** – In diesem Modell werden *Interpretation*, *Translation* und *Konstruktion* als nicht trennbare, zusammengefasste latente Variable abgebildet.
2. **Zweidimensionales Modell (2D)** – Dieses Modell kann als eine Art Intermediat zwischen dem ein- und dreidimensionalen Modell verstanden werden, das die *Translation* und *Konstruktion* aufgrund der Überschneidungen in den Definitionen in einer latenten Variable zusammenfasst, was in Kapitel 5.2.2.2 näher beleuchtet wird. Demgegenüber steht die *Interpretation* als latente Variable.
3. **Dreidimensionales Modell (3D)** – Das dreidimensionale Modell betrachtet die drei lower-level skills getrennt als separate latente Variablen.

Um diese Modelle zu prüfen und zu vergleichen, musste jedoch zunächst ein geeignetes Messinstrument zur Erfassung der drei lower-level skills der Repräsentationskompetenzen entwickelt werden, das *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC). Dieser Schritt war notwendig, um einerseits die Struktur der lower-level skills zu untersuchen und andererseits sicherzustellen, dass eine valide Messung der Repräsentationskompetenzen ohne Verzerrung durch das Fachwissen ermöglicht wird (FF1.2). Dazu wurden bestehende Messinstrumente evaluiert und Kriterien abgeleitet, die die Konstruktion des CRI:TIC leiteten. Anschließend ging es darum, Items aus verschiedenen bestehenden Messinstrumenten zu adaptieren und neue Items auf Basis der Kriterien zu entwickeln. Das Vorgehen zur Entwicklung des CRI:TIC wird in Kapitel 5.2.2 ausführlich beschrieben.

Bevor die Forschungsfragen untersucht werden konnten, musste das CRI:TIC evaluiert werden. Dazu wurden (*multidimensionale*) *Partial Credit Modelle* (PCM) aus der Familie der Rasch-Modelle (Item-Response-Theorie) gewählt, da dieses unterschiedlich skalierte Items (dichotom und polytom) in ein einheitliches Messmodell integriert (Masters, 1982). Personenfähigkeiten und Itemschwierigkeiten werden metrisch skaliert und auf der gleichen linearen Skala abgebildet. Des Weiteren ist der Umgang mit fehlenden Daten im PCM für die Auswertung von Vorteil, da diese bei der Schätzung der Personenfähigkeiten nicht einbezogen werden (Boone et al., 2014). Zur Evaluation wurde das dreidimensionale Modell als Grundmodell verwendet, da von der Trennbarkeit der drei lower-level skills ausgegangen wird (H1.1). Mithilfe geeigneter Parameter wurden zunächst die Grundannahmen des PCM geprüft. Um die

Passung der Items zum PCM und die globale Güte des Modells zu prüfen, wurden Modell- und Itemfit Indizes evaluiert und auf dieser Basis kritische Items geprüft und aussortiert. Mit dem aufgearbeiteten Datensatz wurden anschließend alle drei theoretisch abgeleiteten Modelle hinsichtlich ihrer Güte geprüft.

Anschließend konnten die drei Modelle mithilfe verschiedener statistischer und graphischer Methoden untersucht und verglichen werden. Dadurch kann das Modell mit der besten Passung identifiziert und eine Aussage zur empirischen Trennbarkeit der drei lower-level skills getroffen werden.

Forschungsfrage 1.2

Zur Beantwortung von Forschungsfrage 1.2 wurde ein analoges Verfahren verwendet. Basierend auf den Ergebnissen von Forschungsfrage 1.1 wird das Modell mit der besten Passung für die Repräsentationskompetenzen dem chemischen Fachwissen gegenübergestellt. Dafür lassen sich zwei Modelle für die Untersuchung ableiten:

- 1. Eindimensionales Modell (1D)** – Das eindimensionale Modell fasst das Fachwissen und die Repräsentationskompetenzen in einer latenten Variable zusammen und betrachtet diese als nicht trennbar (Nullhypothese).
- 2. Mehrdimensionales Modell** – Fachwissen und die Repräsentationskompetenzen (ein- bis dreidimensional) werden als separate latente Konstrukte behandelt (Gegenhypothese).

Neben dem CRI:TIC zur Messung der Repräsentationskompetenzen wurden Items zur Messung des Fachwissens aus der ALSTER-Studie (Averbeck, 2021) eingesetzt. Dabei wurde eine inhaltliche Passung zu den Items des CRI:TIC beachtet (Kapitel 5.2.3).

Nach einer Prüfung der Passung der Fachwissensitems zum Modell und der Prüfung der Modell- und Itemfit Indizes konnten kritische Items identifiziert und aussortiert werden. Der aufgearbeitete Datensatz wurde im Anschluss genutzt, um die zweite Forschungsfrage, die Trennbarkeit der Repräsentationskompetenzen und des Fachwissens, zu untersuchen. Dafür wurden verschiedene statistische und graphische Methoden äquivalent zum Vorgehen bei Forschungsfrage 1.1 verwendet.

5.2 Messinstrumente

Leistungstests eignen sich im Rahmen quantitativer Studien besonders, um latente Merkmale wie die Repräsentationskompetenzen oder das Fachwissen mithilfe von manifesten Variablen, den Items, zu erfassen. In den folgenden Kapiteln werden zunächst theoretische Anforderungen an die Qualität und Güte von Messinstrumenten beschrieben. Anschließend wird die Notwendigkeit der Konstruktion des CRI:TIC ausführlich dargestellt und das Messinstrument erläutert, ebenso die Messung des Fachwissens und der Kontrollvariablen.

5.2.1 Gütekriterien zur Entwicklung und Evaluation von Messinstrumenten

Messinstrumente zur Erfassung latenter Merkmale müssen bestimmte Standards und Kriterien erfüllen. Dies ist notwendig, um die Qualität einer Testung zu gewährleisten, also sicherzustellen, dass die aus dem Test gewonnenen Ergebnisse geeignet sind, um Schlussfolgerungen aus untersuchtem Kontext zu ziehen. Dazu hat sich eine Reihe von Testgütekriterien zur Beurteilung etabliert (Döring, 2023e; Moosbrugger & Kelava, 2020; Newton & Shaw, 2013).

Hauptgütekriterien – Bei der Entwicklung und Evaluation eines (psychometrischen) Messinstruments spielen die drei Hauptgütekriterien *Objektivität*, *Reliabilität* und *Validität* eine zentrale Rolle.

(1) Objektivität – Die Ergebnisse eines Tests müssen unabhängig von der Durchführung (*Durchführungsobjektivität*) und der auswertenden Person (*Auswertungsobjektivität*) sein, d. h. unterschiedliche Testanwender müssen bei gleichen Testpersonen zum selben Ergebnis kommen – es darf also keinen intersubjektiven Spielraum bei der Auswertung des Tests geben. Auch die Interpretation der Ergebnisse (*Interpretationsobjektivität*) muss unabhängig von der Person, also standardisiert sein (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020; Rost, 2004; Witte, 2019a).

Das Aufgabenformat bzw. das Testmanual spielt dabei eine entscheidende Rolle: *Geschlossene Items* geben Alternativ- oder Auswahlantworten vor (z. B. richtig oder falsch, Multiple Choice), wodurch genau bekannt ist, welche Ant-

wort(en) richtig oder falsch ist (sind). Bei *halboffenen Items* wird die Formulierung der Antwort der Testperson überlassen. Die Aufgabe sollte jedoch so präzise sein, dass nur eine Antwort richtig ist (z. B. Einfach- und Mehrfachantworten). *Offene Items* überlassen die Antwort ebenfalls der Testperson, wobei die Aufgabenstellungen weniger präzise sind und dadurch vermehrt Handlungsspielraum bei der Auswertung und der Interpretation entsteht (etwa Einschätzungen oder Begründungen zu einem Sachverhalt). Sie widersprechen damit der Logik psychometrischer Tests und finden daher keine bzw. selten Anwendung (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Brandt, 2020; Witte, 2019a).

(2) Reliabilität – Ein guter Test muss zuverlässig und exakt messen und gilt als reliabel, wenn ein Testwert einer bestimmten Person nach einer bestimmten Zeit unter identischen Bedingungen repliziert werden kann. Der Wert sollte allenfalls durch einen Messfehler belastet werden. Ein häufig genutztes Maß zur Bewertung der Reliabilität ist die *interne Konsistenz*. Der wohl bekannteste *Reliabilitätskoeffizient* zu deren Beurteilung ist *Cronbachs Alpha* (α), welcher ein Maß für die Homogenität der Items ist und diese als durchschnittliche Korrelation zwischen den Items definiert. Cronbachs Alpha und andere Reliabilitätskoeffizienten nehmen in der Regel Werte zwischen 0 (völlig unzuverlässige Messung) bis 1 (perfekt zuverlässige Messung) an (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020; Rost, 2004; Witte, 2019a).

(3) Validität – Ein Test gilt dann als valide, wenn er das Merkmal misst, das er zu messen beabsichtigt (Moosbrugger & Kelava, 2020; Rost, 2004). Auch hier sind verschiedene Aspekte zu differenzieren. Unter der *Inhaltsvalidität* versteht man, wie repräsentativ ein Test(item) bei der Merkmalmessung ist. Maßgeblich wird dies durch Expert:innen beurteilt und nicht durch Kennwerte. Davon abzugrenzen ist die *Augenscheinvalidität*, die angibt, inwieweit der Validitätsanspruch von Laien (z. B. von den Testpersonen) als gerechtfertigt angesehen wird. Die *Konstruktvalidität* gibt Auskunft darüber, inwieweit durch die Testitems wirklich auf das beabsichtigte und auf kein anderes Merkmal geschlossen werden kann, und steht daher in enger Beziehung mit den zugrunde gelegten theoretischen Annahmen und Modellen. Hier stellen explorative bzw. konfirmatorische Faktorenanalysen Möglichkeiten zur Gewinnung bzw. Prüfung von Hypothesen dar. Des Weiteren kann hier zwischen der *Konvergenzvalidität*, einem hohen (korrelativen) Zusammenhang mit Mes-

sungen anderer Tests zum gleichen oder einem sehr nah verwandten Merkmal, und der *Diskriminanzvalidität*, einem niedrigen (korrelativen) Zusammenhang mit Messungen anderer Tests zu verwandten, aber dennoch distinkten Merkmalen, unterschieden werden. Wie gut von dem Verhalten einer Testperson innerhalb der Testsituation auf das Verhalten der Person außerhalb der Testsituation geschlossen werden kann, wird durch die *Kriteriumsvalidität* beschrieben. Es ist zu erwarten, dass Personen, die im Test gut abschneiden, auch außerhalb des Tests gut abschneiden und leistungsschwache Personen im Test auch außerhalb der Testsituation leistungsschwach sind (Korrelationsschluss). Die Validität wird als das wichtigste Gütekriterium angesehen, da Messinstrumente, die nicht das Konstrukt messen, das sie erfassen sollen, sowohl für die Diagnostik als auch für die Forschung unbrauchbar sind (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020; Pospeschill, 2022; Witte, 2019a).

Die drei Gütekriterien stehen insofern in einem Zusammenhang, dass eine mangelhafte Objektivität eine schlechte Reliabilität herbeiführt und dies zwangsläufig zu einer schlechten Validität führt. Umgekehrt impliziert ein Test mit hoher Reliabilität jedoch keinen validen Test (Döring, 2023c, 2023e; Witte, 2019a).

Nebengütekriterien – Neben den drei Hauptgütekriterien existieren noch sieben weitere Nebengütekriterien, die zwar ebenfalls eine Rolle bei der Testentwicklung und -evaluation spielen, jedoch von weniger zentraler Bedeutung sind. Sie werden im Folgenden kurz dargestellt:

(1) Skalierung – Dieses Kriterium gilt als erfüllt, wenn leistungsfähige Testpersonen bessere Testwerte als weniger leistungsfähige Personen erreichen. Dabei reicht eine *nominale* Skalierung der Daten nicht aus, diese müssen mindestens *ordinal* skaliert sein. Ein Intervallskalenniveau ermöglicht darüber hinaus eine Beurteilung der Größendifferenz der Testwerte auf inter- und intra-individueller Ebene. Während in der *klassische Testtheorie* (KTT; Kapitel 5.3.1) üblicherweise Summenscores gebildet werden, ohne die messtheoretischen Eigenschaften der Items zu prüfen, wird die Skalierung aller Items bei der *Item-Response-Theorie* (IRT; Kapitel 5.3.2) empirisch auf die Modellkonformität geprüft (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020).

(2) Normierung – Damit ein Test als normiert bezeichnet werden kann, müssen repräsentative Vergleichsstichproben (unterschiedlicher Bevölkerungsgruppen) vorliegen. Basierend auf einem Testrohwert muss dafür ein standardisierter Testwert (z. B. z-Wert) generiert werden (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020).

(3) Testökonomie – Die Testökonomie bewertet Testdauer, den materiellen Aufwand, die Handhabbarkeit und die Dauer für die Auswertung eines Tests in Relation zum Erkenntnisgewinn. Computergestützte Tests sind aus dieser Perspektive besonders ökonomisch. Allgemein gilt, dass die Aussagekraft eines Tests höher als die Aufwandsminimierung gewichtet werden sollte (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020).

(4) Nützlichkeit – Wenn es an einem Messinstrument zur Erfassung eines Merkmals mangelt oder bestehende Messinstrumente eine fragwürdige Testgüte aufweisen, kann die Nützlichkeit eines Tests als hoch angesehen werden. Dementsprechend muss die Nützlichkeit mit der Entwicklung neuer Instrumente erneut kritisch geprüft werden (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020).

(5) Zumutbarkeit – Im Vergleich zum Erkenntnisgewinn sollten Personen nicht übermäßig zeitlich, physisch oder psychisch belastet werden. Dieses Gütekriterium vertritt somit einen forschungsethischen Grundsatz, dass Testpersonen nicht übermäßig belastet oder beeinträchtigt werden dürfen (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020).

(6) Unverfälschbarkeit – Damit ist die Durchschaubarkeit des Messprinzips durch die Testperson gemeint und inwieweit diese die eigenen Testwerte steuern bzw. verzerren kann. Dieser Aspekt spielt besonders bei Persönlichkeitsfragebögen eine entscheidende Rolle, da einem Antwortverhalten gemäß einer „sozialen Erwünschtheit“ vorzubeugen ist. Bei Leistungstests ist dies in der Regel unproblematisch, selbst wenn das Messprinzip durchschaut wird (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020).

(7) Fairness – Ein Test ist als fair zu bewerten, wenn Ergebnisse des Tests zu keiner Benachteiligung bestimmter Personengruppen führen. Damit ist die Zugehörigkeit zu ethnischen, sozokulturellen und geschlechtsspezifischen Per-

sonengruppen gemeint. Beispielsweise betrifft dies auch die sprachliche Gestaltung der Aufgaben, sodass Items auch ohne hohe sprachliche Fähigkeiten und unabhängig vom soziokulturellen Status gelöst werden können (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020).

5.2.2 Messung der Repräsentationskompetenzen

Um ein latentes Merkmal wie die Repräsentationskompetenzen einer Person zu untersuchen, bedarf es eines geeigneten psychometrischen Messinstruments (Tests), das dieses Merkmal mithilfe manifester Variablen (Items) modelliert. Dazu eignen sich Leistungstest mit begrenzter Bearbeitungszeit (Speed-Tests) und/oder steigendem Aufgabenniveau (Power-Tests). Dazu muss das Messinstrument die Hauptgütekriterien *Objektivität*, *Validität* und *Reliabilität* erfüllen, sodass die Legitimität und Gültigkeit des Instruments und der daraus gewonnen Daten beansprucht werden kann. Ebenso sollten die Nebengütekriterien berücksichtigt werden (Döring, 2023e, 2023c; Krumm et al., 2021; Moosbrugger & Kelava, 2020).

Basierend darauf wurden bereits bestehende Messinstrumente zur Erhebung von Repräsentationskompetenzen gesichtet und die Eignung für die Forschungsanliegen kritisch geprüft. Diese Evaluation hat ergeben, dass die Konstruktion eines neuen Instruments, des *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC) notwendig war, um die Forschungsanliegen adäquat zu adressieren.

Um die Notwendigkeit der Entwicklung des CRI:TIC transparent und nachvollziehbar darzustellen, werden im Folgenden die Ergebnisse der Evaluation bestehender Messinstrumente beschrieben. Daraus werden positive und negative Aspekte herausgearbeitet, um Kriterien zu Gestaltung des CRI:TIC abzuleiten. Auf dieser Basis wird die Entwicklung und Konstruktion des CRI:TIC erläutert.

5.2.2.1 Kritische Evaluation bereits existierender psychometrischer Messinstrumente zur Erfassung von Repräsentationskompetenzen in der Chemie

Um die ersten zwei Forschungsanliegen zu adressieren, wurden bereits existierende Messinstrumente zur Erfassung von Repräsentationskompetenzen im Fach Chemie betrachtet und die Eignung für die Untersuchungen geprüft. Daraus hat sich ergeben, dass sich keines der Instrumente in seiner aktuellen Form für die Untersuchungen eignet.

Struktur des Messinstruments – Mit Blick auf die Untersuchung der Struktur der lower-level skills der Repräsentationskompetenzen ist ein Messinstrument notwendig, das die Items eindeutig einer der drei Fähigkeiten (*Interpretation*, *Translation* und *Konstruktion*) zuordnet. Das *Chemical Representation Inventory* (Taskin et al., 2015) und der *Test of Representational Competence* (Sim & Daniel, 2014) berücksichtigen diese Fähigkeiten zwar, ordnen die Items den Fähigkeiten jedoch nicht zu. Die Tests beabsichtigen keine differenzierte Betrachtung der einzelnen Fähigkeiten, sondern zielen darauf ab, die Repräsentationskompetenzen als holistisches Konstrukt zu erheben. Das *Organic Chemistry Representational Competence Assessment* (ORCA) von Ward et al. (2025) erfüllt diese Anforderungen zwar, nimmt aber einerseits statt der *Konstruktion* die *Argumentation* mit in das Messinstrument auf und stand andererseits zu Beginn des Projekts nicht zur Verfügung und konnte daher nicht berücksichtigt werden.

Validität des Messinstruments – Ein Instrument zur Messung der Repräsentationskompetenzen sollte diese unabhängig von anderen Konstrukten messen, damit es eine valide Messung und belastbare Aussagen auf Basis der Ergebnisse zulässt. Hier zeigen das *Chemical Representation Inventory* (Taskin et al., 2015) und der *Test of Representational Competence* (Sim & Daniel, 2014) jedoch einige Mängel mit Blick auf das geplante Anliegen. Einige Items in den Messinstrumenten zielen darauf ab, Repräsentationskompetenzen zu erheben, setzen dabei jedoch Fachwissen voraus, sodass diese nicht gelöst werden können, wenn das Fachwissen nicht vorhanden ist. Beispielsweise wird nach der Strukturformel für Benzoesäure gefragt. Die Struktur kann an dieser Stelle nur skizziert werden, wenn diese bereits bekannt ist. Das Repräsentationsdilemma wird hier also nicht adäquat berücksichtigt und verzerrt die Messung der Repräsentationskompetenzen. Für die ersten beiden Forschungsanliegen,

besonders aber für FF1.2, ist es wichtig, Repräsentationskompetenzen und Fachwissen eindeutig zu operationalisieren und zu definieren, sodass diese unabhängig voneinander gemessen werden können.

Bei der Betrachtung des ORCA von Ward et al. (2025) zeigen sich ähnliche Schwächen im Messinstrument. Einige Items setzen ein hohes Maß an Fachwissen voraus, beispielsweise wenn nach der Verbindung (als Strukturformel dargestellt) mit dem höchsten Siedepunkt gefragt wird. Um dieses Item zu beantworten, wird Fachwissen über intermolekulare Wechselwirkungen vorausgesetzt (Schmuck, 2013). Mit Blick auf die Items zur Erfassung der Argumentationsfähigkeit lässt sich in Frage stellen, inwieweit das Multiple Choice Format geeignet ist, eine Fähigkeit zu messen, die komplexe Denkvorgänge umfasst (Kozma & Russell, 2005). Auch in Messinstrumenten für Repräsentationskompetenzen in anderen Naturwissenschaften zeigen sich hinsichtlich der Berücksichtigung des Repräsentationsdilemmas Schwächen (z. B. Scheid et al., 2018).

Auch wenn die beschriebenen Messinstrumente bzw. einzelne Items Schwächen aufweisen, dienen sie als gute Orientierung zur Konstruktion eines für die Forschungsanliegen geeigneten Instruments. Sowohl im *Chemical Representation Inventory* (Taskin et al., 2015) als auch im *Test of Representational Competence* (Sim & Daniel, 2014) wird eine Mischung aus geschlossenen und halboffenen Items genutzt, wodurch die Objektivität der Instrumente mithilfe eines Testmanuals gewährleistet werden kann. Zudem eignet sich ein halboffenes Aufgabenformat, um die *Translation* und *Konstruktion* in einem angemessenen Rahmen zu erfassen, da so Repräsentationen wie Keilstrichformeln von den Testpersonen selbst skizziert werden können. Durch die kritische Evaluation der bestehenden Instrumente können außerdem die bestehenden Schwächen erst identifiziert und adäquat berücksichtigt werden.

5.2.2.2 Operationalisierung der Repräsentationskompetenzen

Wie bereits beschrieben, differenzieren das *Chemical Representation Inventory* (Taskin et al., 2015) und der *Test of Representational Competence* (Sim & Daniel, 2014) nicht zwischen einzelnen Fähigkeiten, sondern erheben Repräsentationskompetenzen als holistisches Konstrukt. Daher bestand eine Überlegung darin, Items aus den Instrumenten, die das Repräsentationsdilemma

adäquat berücksichtigen, zu adaptieren und den lower-level skills von Kozma und Russell (Tabelle 1) zuzuordnen. Hier zeigte sich jedoch schnell, dass eine eindeutige Zuordnung auf Basis der Definitionen von Kozma und Russell (1997, 2005) nicht möglich ist. Besonders mit Blick auf die Fähigkeiten *Translation* und *Konstruktion* offenbarten sich hier große Überschneidungen hinsichtlich der Zuordnung. Wenn beispielsweise die Halbstrukturformel einer Verbindung vorgegeben wird und daraus die Skelettformel (Abbildung 3) skizziert werden soll, ist dies eine *Translationsleistung*, da eine Repräsentation in eine andere übersetzt wird. Andererseits wird eine neue Repräsentation erstellt, weshalb das Item auch der *Konstruktion* zugeordnet werden kann. Dadurch zeigt sich, dass die Repräsentationskompetenzen stärker operationalisiert werden müssen, um Items zu den einzelnen Fähigkeiten zu adaptieren und zu erstellen. Dieses Vorgehen ist besonders mit Blick auf die Konstruktvalidität des Messinstruments notwendig.

In dieser Hinsicht hat sich die Klassifikation der Repräsentationen bzw. die Charakterisierung von Repräsentationen als eher abstrakt und eher konkret als geeigneter Ansatz angeboten. Da sowohl die *Translation* als auch die *Konstruktion* die Gestaltung neuer Repräsentationen umfassen, muss besonders in dieser Hinsicht eine Abgrenzung der Prozesse erreicht werden. Dies wurde auf der Basis zweier Aspekte ermöglicht:

Umgang mit der Referenzrepräsentation – Wird der Inhalt der Referenzrepräsentation während der Gestaltung der Zielrepräsentation selbst nicht verändert und lediglich auf eine andere Art und Weise dargestellt, handelt es sich um eine *Translation*. Dies umfasst beispielsweise die Änderung der Repräsentationsform (z. B. eine Übersetzung vom Kugel-Stab-Modell zur Keilstrichformel; Abbildung 3). Wird die Referenzrepräsentation inhaltlich jedoch verändert und eine neue, inhaltlich zur Referenzrepräsentation distinkte Repräsentation geschaffen, wird dieser Prozess der *Konstruktion* zugeordnet. Dafür muss die Repräsentationsform an sich nicht verändert werden, wenn z. B. auf Basis der Keilstrichformel des (R)-Milchsäuremoleküls die Keilstrichformel des (S)-Milchsäuremoleküls skizziert werden soll.

Änderung des Abstraktionsgrads – Vergleicht man den Abstraktionsgrad zwischen Referenz- und Zielrepräsentation, fällt auf, dass sich die explizit dargestellten Inhalte bei geringem Unterschied sehr ähnlich sind, z. B. bei Keilstrich-

formel und Kugel-Stab-Modell. Erhöht man jedoch den Abstand im Abstraktionsgrad, so unterscheiden sich auch die explizit dargestellten Inhalte sehr stark und ähneln sich teilweise kaum noch, etwa wenn Summenformel und Keilstrichformel verglichen werden. Basierend darauf sind *Translationsleistungen* im Rahmen dieser Arbeit als Operationen zu verstehen, bei denen der Abstraktionsgrad zwischen Referenz- und Zielrepräsentation wenig verändert wird und die explizit dargestellten Merkmale lediglich übersetzt werden müssen. Bei *Konstruktionsleistungen* hingegen wird der Abstraktionsgrad zwischen Referenz- und Zielrepräsentation stark verändert, sodass die Zielrepräsentation nicht einfach durch Übersetzung erreicht werden kann, sondern auf Basis der Referenzrepräsentation konstruiert werden muss (s. Kapitel 5.2.2.3).

In Bezug auf die Fähigkeit *Konstruktion* war eine weitere Anpassung notwendig. Die Definition von Kozma und Russell (1997, 2005) umfasst neben der Gestaltung einer Repräsentation auch die Erklärung, weshalb diese Repräsentation für den geplanten Einsatz angemessen ist (Tabelle 1). Es zeigt sich eine Überschneidung mit der Fähigkeit *Vergleich und Kritik* aus den higher-level skills, weshalb dieser Aspekt in der hier zugrunde gelegten Definition für die *Konstruktion* ausgelassen wird. Des Weiteren kann das Erklären eines Sachverhalts nicht adäquat durch geschlossene bzw. halboffene Aufgabenformate erhoben werden. Ein offenes Aufgabenformat ist erforderlich. Dieses Format widerspricht jedoch der Logik eines Leistungstests, da es mit Blick auf die Auswertung und Interpretation der Antworten ein eher qualitatives Auswertungsverfahren benötigt (Döring, 2023c; Moosbrugger & Brandt, 2020).

Ausgehend von diesen zwei Aspekten lassen sich die Repräsentationskompetenzen bzw. die Fähigkeiten *Translation* und *Konstruktion* besser operationalisieren und voneinander abgrenzen. In Tabelle 3 sind die drei in dieser Arbeit zugrunde gelegten lower-level skills definiert.

Tabelle 3: Operationalisierung der lower-level skills (¹Gurung et al., 2022; ²Kozma & Russell, 1997, 2005; ³Nitz, 2012) mithilfe der Charakterisierung von Repräsentationen (Abbildung 2).

Kategorisierung ³		Definitionen ²
lower-level skills ¹	Interpretation	Die Fähigkeit, Merkmale und Muster in einer Repräsentation zu identifizieren und analysieren, sowie Repräsentationen zum Beschreiben beobachtbarer chemischer Phänomene in Bezug auf zugrundeliegende molekulare Zusammenhänge und Vorgänge zu nutzen.
	Translation	Die Fähigkeit, zwischen Repräsentationen <i>mit ähnlichem Abstraktionsgrad und ähnlichen visuellen Informationen</i> zu übersetzen und die Betrachtungsperspektive zu wechseln, <i>ohne das repräsentierte Objekt zu ändern</i> .
	Konstruktion	Die Fähigkeit, neue Repräsentationen zu konstruieren oder auszuwählen, <i>bei denen sich der Abstraktionsgrad von der Referenzrepräsentation stark unterscheidet oder die distinkt zu dieser sind</i> .

Anmerkung: Die Anpassungen der Fähigkeiten auf Basis der Charakterisierung von Repräsentationen sind *kursiv* dargestellt.

Bevor die Entwicklung des CRI:TIC beschrieben und erläutert wird, müssen zwei weitere wichtige Fragen geklärt werden: (1) Wie werden Repräsentationen in Bezug auf die Messung der Repräsentationskompetenzen operationalisiert? Und (2) wie können Items zur Messung der Repräsentationskompetenzen dahingehend von Items zur Messung des chemischen Fachwissens abgegrenzt werden?

Das Ziel des CRI:TIC ist die Messung der Repräsentationskompetenzen über verschiedene Fachbereiche der Chemie (Anorganische Chemie, Organische Chemie, Allgemeine Chemie, Physikalische Chemie usw.) hinweg und dabei eine hohe Bandbreite an typischen statischen Repräsentationen in der Chemie zu integrieren, beispielsweise Diagramme, Graphen, Reaktionsgleichungen und -mechanismen sowie verschiedene Repräsentationen zur Darstellung von chemischen Verbindungen (Abbildung 3) (Gurung et al., 2022; Harle &

Towns, 2011; Kozma & Russell, 2005; Rau et al., 2021). Dadurch soll eine generalisierbare Aussage über die Fähigkeiten der Repräsentationskompetenzen getroffen werden können, als läge der Fokus auf einer bestimmten Fachrichtung oder bestimmten Repräsentationen wie beispielsweise im ORCA (Ward et al., 2025). Dies könnte gleichzeitig als Limitation betrachtet werden, da somit für einzelne Fachbereiche spezifische Repräsentationen und damit einhergehende Besonderheiten nur eingeschränkt berücksichtigt werden. Die oben genannten statischen Repräsentationen lassen sich mithilfe der Kategorisierung von Repräsentationen (Abbildung 2) als symbolisch-mathematische und visuell-graphische Repräsentationen sowie hybride Formen daraus einordnen. Diese Kategorien von Repräsentationen sind daher von zentraler Bedeutung für die Entwicklung des CRI:TIC sowie für die Messung der Repräsentationskompetenzen.

Um das Fachwissen unabhängig davon zu messen, muss das Instrument zu dessen Messung somit frei von symbolisch-mathematischen, visuell-graphischen und hybriden Repräsentationen sein und beschränkt sich daher auf rein verbal-textliche Repräsentationen (Kapitel 5.2.3). Handelnd-operative Repräsentationen werden bei der Messung beider Konstrukte ausgeschlossen, da hier der Fokus auf symbolisch-mathematischen und visuell-graphischen Repräsentationen liegt und sie andere Anforderungen an die Messung selbst sowie die Durchführung der Datenerhebung stellen.

5.2.2.3 Entwicklung des *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC)

Auf dieser Grundlage lassen sich Kriterien ableiten, die die Items zur Messung der Repräsentationskompetenzen erfüllen müssen, um die empirische Trennbarkeit der drei lower-level skills (FF1.1) und der Repräsentationskompetenzen vom Fachwissen zu untersuchen (FF1.2):

- 1. Repräsentationsbezug** – Alle Items zur Messung der lower-level skills der Repräsentationskompetenzen müssen mindestens eine symbolisch-mathematische, visuell-graphische oder eine hybride Repräsentation in die Aufgabe integrieren.

2. **Bezug zu den lower-level skills** – Die Items müssen auf Basis der drei lower-level skills (Tabelle 3) konzipiert werden, bzw. diesen zugeordnet werden können, wenn sie aus bestehenden Instrumenten adaptiert werden.
3. **Kontrolle des Fachwissens** – Um sicherzustellen, dass die Items ausschließlich die Repräsentationskompetenzen und kein Fachwissen erheben, muss in der Aufgabenstellung das Fachwissen, das zur Lösung notwendig ist, kurz angegeben werden, sofern es sich um spezialisiertes chemisches Fachwissen handelt. Zudem muss gewährleistet sein, dass die Testpersonen die Konventionen zu den aufgegriffenen Repräsentationen kennen, weshalb zu allen eingesetzten Repräsentationen eine Übersicht mit deren Konventionen zur Verfügung gestellt werden muss (Anhang A.2).
4. **Aufgabenformat** – Hinsichtlich der Objektivität (und der Testökonomie) sollten die Items im geschlossenen und halboffenen Format gestaltet sein, sodass ein Testmanual entwickelt werden kann und für jedes Item eine eindeutige Lösung vorhanden ist.

Basierend auf diesen Kriterien erfolgte zunächst die Evaluierung des *Chemical Representation Inventory* (Taskin et al., 2015), des *Tests of Representational Competence* (Sim & Daniel, 2014) und der ALSTER Studie (Averbeck, 2021), aus der insgesamt 15 geeignete Items hervorgingen. Hinzu kam eine Sichtung von Lehrbüchern mit Übungsaufgaben, woraus ein weiteres Item adaptiert werden konnte. Die 16 Items wurden teilweise vom Englischen ins Deutsche übersetzt und entsprechend den Kriterien angepasst. Zusätzlich entstanden 22 Items eigenständig auf Basis der vier Kriterien, wofür die drei genannten Messinstrumente sowie die Studien von Ward und Kolleg:innen (2022) sowie Stieff und Kolleg:innen (2016, 2018) als Orientierung dienten. Dadurch bildete sich ein Itempool von 38 Items, die den drei lower-level skills zugeordnet wurden: 14 Items der *Interpretation*, elf Items der *Translation* und 13 Items der *Konstruktion*. 22 Items liegen im halboffenen und 16 im geschlossenen Format vor. Die Herkunft und Zuordnung der Items zeigt Tabelle 4.

Tabelle 4: Herkunft der 38 Items des CRI:TIC und Zuordnung zu den drei lower-level skills.

Herkunft	Interpretation	Translation	Konstruktion	Σ
Taskin et al. (2015) ^e	Int_03	Trans_01*	Konst_03	10
	Int_04*	Trans_04*	Konst_04	
	Int_05*	Trans_05*		
	Int_06*	Trans_06		
Sim und Daniel (2014) ^e	Int_14*		Konst_07*	4
			Konst_10*	
			Konst_13*	
Averbeck (2021)	Int_08*			1
Vollhardt und Schore (2011)		Trans_08*		1
Selbst gestaltet	Int_01	Trans_02	Konst_01	22
	Int_02	Trans_03	Konst_02	
	Int_07	Trans_07	Konst_05	
	Int_09	Trans_09	Konst_06	
	Int_10	Trans_10	Konst_08	
	Int_11	Trans_11	Konst_09	
	Int_12		Konst_11	
	Int_13		Konst_12	
Σ	14	11	13	38

e = aus dem Englischen übersetzte Items; * = überarbeitete/angepasste Items

Die große Vielfalt an unterschiedlichen Repräsentationen (Diagramme, Graphen, Moleküldarstellungen usw.) sowie an Fachrichtungen in der Chemie (Organische Chemie, Anorganische Chemie, Physikalische Chemie, Allgemeine Chemie usw.) erschwerte eine systematische Konstruktion von Testitems, wie sie etwa bei Ward et al. (2025) umgesetzt wurde. Für jede der drei lower-level skills hätte für jeden Fachbereich ein Item mit der gleichen Repräsentation erstellt werden müssen. Dadurch wäre eine unrealistische Anzahl an Items entstanden, die aus testökonomischen Gründen und mit Blick auf die Zumutbarkeit für die Testpersonen nicht haltbar gewesen wäre. Hier hätten sich bei-

spielsweise aus den drei lower-level skills 15 unterschiedlichen Repräsentationen und vier Fachbereichen in der Summe 180 Items ergeben. Auch mit Blick auf die Schwierigkeit der Items hat sich kein systematisches Vorgehen als zielführend erwiesen, da die Anzahl an schwierigkeiterzeugenden Faktoren sehr hoch ist (u. a. Inhalt, Repräsentation, Sprache und Vorwissen). Daher wurden zur Gestaltung des CRI:TIC nur die vier oben herausgearbeiteten Kriterien zugrunde gelegt.

Im Folgenden wird exemplarisch an drei Items dargestellt und erläutert, inwieweit diese die genannten Kriterien erfüllen und begründet, weshalb sie der Fähigkeit *Interpretation*, *Translation* oder *Konstruktion* zugeordnet werden. Das vollständige CRI:TIC kann dem Anhang entnommen werden (Anhang A.1).

Interpretation – In Item *Int_01* (Abbildung 11) ist ein Energiediagramm dargestellt, d. h. eine eher visuell-graphische Repräsentation mit symbolischen Anteilen, das den Verlauf einer Reaktion verdeutlicht, bei dem ein Zwischenprodukt gebildet wird. Dieser Verlauf ist beispielsweise bei der S_N1 -Reaktion typisch (Abbildung 4). Aufgabe ist es, das Zwischenprodukt im Energiediagramm zu verorten. Der Verlauf des Graphen muss somit analysiert und das lokale Minimum als Zwischenprodukt identifiziert werden, weshalb dieses Item eindeutig der *Interpretation* zugeordnet werden kann.

Mit Blick auf das Fachwissen wird im Aufgabenstamm lediglich angegeben, dass bei Reaktionen Zwischenprodukte entstehen können und dass diese als lokale Minima im Verlauf identifiziert werden können. In der Schullaufbahn kommen die Lernenden mit vergleichbaren Diagrammen, z. B. bei endothermen und exothermen Reaktionen, in Kontakt (KMK, 2020). Die Studierenden wissen daher in der Studieneingangsphase, dass der Reaktionsverlauf von links nach rechts auf der Abszisse (Reaktionskoordinate) abgebildet wird, also links die Edukte und rechts die Produkte dargestellt sind. Selbst ohne dieses Wissen sollte das Item durch die Angabe, dass Zwischenprodukte als lokale Minima erkenntlich sind, lösbar sein.

Das Item ist halboffen, da die Lernenden selbstständig eine Antwort formulieren müssen. In der Aufgabe wird bewusst nicht vorgegeben, wie viele richtige Antworten existieren. Dadurch soll vermieden werden, dass die Aufgabe impliziert, nur das lokale Minimum könne die richtige Antwort sein; stattdessen

sollen auch die Edukte und Produkte für die Studierenden als richtige Antworten in Frage kommen.

Während vieler chemischer Reaktionen werden Zwischenprodukte gebildet, die in Energiediagrammen als lokale Energieminima erkenntlich sind. **Markieren** Sie in dem abgebildeten Energiediagramm das Zwischenprodukt/die Zwischenprodukte.

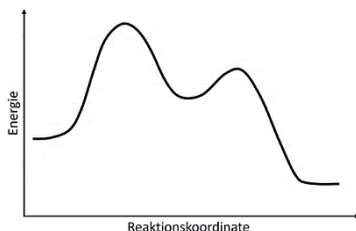


Abbildung 11: Item *Int_01* als Beispielitem für die *Interpretation* (selbst gestaltet).

Translation – Das Items *Trans_07* (Abbildung 12) gibt die Keilstrichformel des (2*S*,3*S*)-Weinsäuremoleküls vor sowie vier Kugel-Stab-Modelle: a) (2*S*,3*S*)-Weinsäure, b) (2*R*,3*R*)-Weinsäure, c) und d) (2*R*,3*S*)-Weinsäure (durch Drehung ineinander überführbar). Die Lernenden sollen die identischen Strukturen identifizieren. Zur Lösung der Aufgaben müssen die Testpersonen die Keilstrichformel in ein Kugel-Stab-Modell übersetzen. Die Referenz- und die Zielrepräsentation bilden das gleiche Molekül ab, ein (2*S*,3*S*)-Weinsäuremolekül, und der Abstraktionsgrad zwischen beiden Repräsentationen ändert sich kaum, weshalb die Aufgabe als Translationsaufgabe eingeordnet wird.

Für die Lösung des Items ist kein zusätzliches Fachwissen notwendig, da alle notwendigen Informationen in der Aufgabenstellung gegeben beziehungsweise den Repräsentationen zu entnehmen sind. Da in beiden Repräsentationsformen Einfachbindungen durch Striche und Doppelbindungen durch Doppelstriche dargestellt sind, kann geschlussfolgert werden, dass graue Kugeln für Kohlenstoffatome, rote Kugeln für Sauerstoffatome und weiße Kugeln für Wasserstoffatome stehen. Mithilfe der beigelegten Übersicht konnten die Lernenden nachschlagen, dass gefüllte Keile von der Blattebene nach vorne und gestrichelte von der Blattebene nach hinten zeigen.

Da die Antworten vorgegeben sind, handelt es sich hier um ein geschlossenes Aufgabenformat, ein Multiple Choice: Single Select Item. Die Lernenden müssen ihre Antwort nicht selbst formulieren, sondern eine Antwort wählen.

Wählen Sie das Kugel-Stab-Modell, das der abgebildeten Keilstrichformel des Weinsäure-Moleküls entspricht.

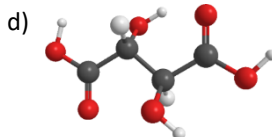
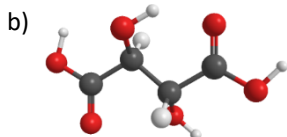
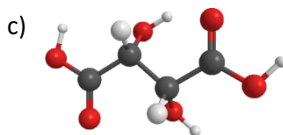
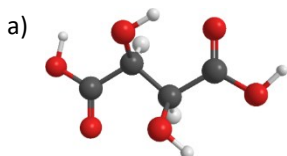
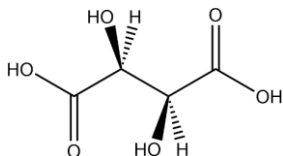


Abbildung 12: Item *Trans_07* als Beispielitem für die *Translation* (selbst gestaltet).

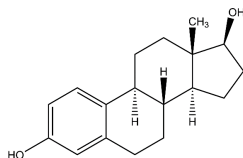
Konstruktion – In Item *Konst_05* (Abbildung 13) ist das Estradiolmolekül als Skelettformel angegeben, aus der die Studierenden die Summenformel konstruieren sollen. Der Abstraktionsgrad ändert sich hier von der Skelett- zur Summenformal stark, ohne dabei die Referenzrepräsentation selbst zu verändern, daher kann dieses Item eindeutig der *Konstruktion* zugeordnet werden.

Die Skelettformel ist eine stark verkürzte und kondensierte Strukturformel. Jede Ecke steht für ein vierbindiges Kohlenstoffatom, das mit einer maximalen Anzahl an Wasserstoffatomen gesättigt ist, sofern keine anderen Substituenten explizit angegeben sind. Ward und Kolleg:innen (2022) konnten in ihrer Studie unter Einsatz des Lauten Denkens beispielsweise zeigen, dass Lernende

teilweise große Schwierigkeiten haben, die implizit dargestellten Wasserstoffatome zu zählen. Diese Strukturformel findet auch im schulischen Kontext häufig Anwendung, sodass die Studierenden mit dieser vertraut sein sollten. Sie können aber wie bei den anderen Items auch die Übersicht zu den Strukturformeln zur Hilfe nehmen.

Hier handelt es sich um halboffenes Item, das eine eindeutige richtige Antwort aufweist und daher einfach und objektiv auswertbar ist. Es existiert genau eine richtige Anzahl an Kohlenstoff-, Wasserstoff- und Sauerstoffatomen in der Verbindung.

Geben Sie die Summenformel der abgebildeten Verbindung (Estradiol) in der Form $C_xH_yO_z$ an.



Summenformel: _____

Abbildung 13: Item *Konst_05* als Beispielitem für die *Konstruktion* (selbst gestaltet).

Auswertung des CRI:TIC und Dateneingabe

Zur Auswertung des CRI:TIC wurde ein Testmanual erstellt. Besonders für die Bewertung der halboffenen Items war dies ein notwendiger Schritt. Items wie *Konst_05* stellten dabei weniger problematische Items dar, die eindeutig und schnell bewertet werden konnten. Items wie beispielsweise *Konst_08* (Anhang A.1), bei dem eine Skelettformel auf Basis einer Summenformel skizziert werden muss, stellten eine größere Herausforderung dar, da hier mehrere korrekte Antworten möglich waren. Bei solchen Items musste jede Antwort individuell geprüft und das Testmanual dementsprechend erweitert werden, sofern neue korrekte Antworten zu identifizieren waren. Das Testmanual kann dem Anhang entnommen werden (Anhang A.3).

Die Daten wurden in IBM SPSS® Statistics-Software (Version 29) erfasst. Für die geschlossenen Items wurden die Antworten jeder Person zunächst ohne Prüfung auf Richtigkeit eingegeben. Anschließend wurden diese mithilfe einer Syntax in Punkte umkodiert, um so Fehler in der Bewertung zu vermeiden. Die halboffenen Items wurden dazu teilweise in Teilitems, d. h. in mehrere Variablen gegliedert und zum Großteil vor der Eingabe bewertet (Anhang A.3).

5.2.3 Messung des chemischen Fachwissens

Wie im Kapitel 5.2.2 beschrieben, müssen für eine valide Messung der Repräsentationskompetenzen und des chemischen Fachwissens beide Konstrukte klar operationalisiert und abgegrenzt werden. Nur dadurch kann das Repräsentationsdilemma adäquat berücksichtigt und die Forschungsfragen belastbar beantwortet werden.

Die Items zur Messung des Fachwissens müssen daher so konzipiert sein, dass sie sich inhaltlich, rein vom gefragten Gegenstand her, mit den Items zur Messung der Repräsentationskompetenzen überschneiden, aber lediglich das Fachwissen erheben. Daher dürfen die Fachwissenitems keine symbolisch-mathematischen, visuell-graphischen oder hybriden Repräsentationen davon nutzen. Hier werden rein textliche Items verwendet, da so eine klare Abgrenzung von den Items des CRI:TIC erreicht werden kann.

Mit Blick auf die vier Wissensarten nach Anderson und Krathwohl (2001) eignen sich das Faktenwissen und das konzeptuelle Wissen (deklaratives Fachwissen), um das Fachwissen zu messen. Dieses spielt eine entscheidende Rolle beim Lernen mit Repräsentationen (Rau, 2017b) sowie bei anderen Tätigkeiten und Prozessen in der Chemie, beispielsweise beim Experimentieren (Nehring & Schwichow, 2020). Das deklarative Fachwissen kann verbalisiert und somit ideal mithilfe eines Leistungstest mit geschlossenen Items gemessen werden. Das prozedurale Wissen und das metakognitive Wissen stellen implizites, nicht verbalisiertes Wissen dar, das andere Anforderungen an eine Messung stellen und nicht optimal mit einem Fachwissenstest ermittelt werden können.

Um die Messung der Repräsentationskompetenzen und des chemischen deklarativen Fachwissens vergleichen zu können, müssen sich diese hinsichtlich der gefragten Inhalte überlappen. Daher wurden die zentralen Inhalte des

CRI:TIC betrachtet und die ALSTER-Studie (Averbeck, 2021) mit Blick auf textliche, geschlossene Items mit einer inhaltlichen Passung zum CRI:TIC herangezogen. Messinstrumente der ALSTER-Studie stellen einen geeigneten Item-Pool dar, da sie wie das CRI:TIC verschiedene Fachbereiche der Chemie abdecken und für Studierende in der Studieneingangsphase konzipiert wurden, die auch in dieser Datenerhebung fokussiert wurden (Kapitel 5.5).

Insgesamt wurden aus den Messinstrumenten der ALSTER-Studie 18 Items übernommen. Diese sind im Multiple Choice: Single Select Format, das vier Antwortmöglichkeiten vorgibt, von denen immer genau eine richtig ist. In Abbildung 14 ist exemplarisch das Item *Fach_07* dargestellt.

Was versteht man unter einem Isomer?

- a) Moleküle, die dieselbe Summenformel besitzen aber unterschiedliche Strukturformeln aufweisen.
- b) Moleküle, die dieselbe Strukturformel besitzen aber unterschiedliche Summenformeln aufweisen.
- c) Atome, die gleich viele Protonen aber unterschiedlich viele Neutronen besitzen.
- d) Atome, die gleich viele Neutronen aber unterschiedlich viele Protonen besitzen.

Abbildung 14: Item *Fach_07* als Beispielitem für die ein *Fachwissensitem* (Averbeck, 2021).

Das Item kann basierend auf der Operationalisierung von Repräsentationskompetenzen und Fachwissen eindeutig dem Fachwissen zugeordnet werden, da es rein textliche Repräsentationen verwendet und Faktenwissen in Form einer Definition abfragt. Hier ist Antwort a) die korrekte Antwort und Antwort b) ein Distraktor, bei welchem die zwei zentralen Aspekte eines Isomers vertauscht wurden, wodurch eine falsche Aussage entstanden ist, Antwort c) entspricht der Definition eines Isotops, während in Antwort d) wieder eine falsche Aussage zu Isotopen getroffen wird, indem die zentralen Aspekte abermals vertauscht werden.

5.3 Testtheorien

Um die Frage zu beantworten, inwieweit die erhobenen Daten, also das Testverhalten einer Person, mit dem fokussierten latenten Konstrukt im Zusammenhang steht, ist eine Testtheorie notwendig. Diese trifft bestimmte Annahmen darüber, wie aus empirisch ermittelten Ergebnissen auf das latente Konstrukt geschlossen werden kann (Döring, 2023c). Grundlegend können zwei Testtheorien unterschieden werden: Die Klassische Testtheorie (KTT) und die Item-Response-Theorie (IRT). Beide gehen davon aus, dass Items in einem Test (manifeste Variablen) als Indikatoren für die Ausprägung einer zugrundeliegenden latenten, nicht beobachtbaren Variable dienen. Lange wurde die Klassische Testtheorie aufgrund einer unkritischen Nutzung von Methoden zur Reliabilitätsschätzung kritisiert, „da modelltheoretische Implikationen verschiedener Reliabilitätskoeffizienten keine Berücksichtigung fanden“ (Moosbrugger, Schermelleh-Engel, et al., 2020, S. 253). Diese Kritik ist heute jedoch nicht mehr gerechtfertigt, da heute sowohl die KTT als auch die IRT Modelltests zur empirischen Prüfung modellimplizierter Annahmen treffen. Beide Testtheorien stellen somit legitime Vorgehen dar, die eine hohe Übereinstimmung zeigen (Moosbrugger, Schermelleh-Engel, et al., 2020).

Zur Evaluation des CRI:TIC und zum Prüfen der Forschungsfragen des ersten Forschungsanliegens wurde in dieser Arbeit die IRT gewählt. Um diese Entscheidung zu begründen, werden im Folgenden beide Theorien vereinfacht und grundlegend erläutert.

5.3.1 Die Klassische Testtheorie (KTT)

Die KTT stellt einen seit etwa 70 Jahren bewährten Ansatz zur Bewertung der Reliabilität und Validität von Tests dar und ist aufgrund der ökonomischen und einfachen Umsetzung theoretischer Anforderungen sehr beliebt. Sie kann als „Messfehlertheorie“ bezeichnet werden, da Messungen üblicherweise mit Messfehlern einhergehen und die KTT darauf abzielt, den wahren Wert τ_i einer Messung vom Messfehler ε_i eines Items y_i zu trennen (Döring, 2023c; Moosbrugger, Gade, et al., 2020; Moosbrugger, Schermelleh-Engel, et al., 2020). Die KTT geht dabei von einer Reihe von Annahmen aus, die theoretisch gesetzt wurden und empirisch nicht prüfbar sind. Für diese gibt es bis heute keine Einigung zu einer stringenten Gliederung und Notation. Im Folgenden

werden drei Grundannahmen und zwei Zusatzannahmen über die Unabhängigkeit der Messfehler nach Döring (2023c) sowie Moosbrugger, Gädde und Kolleg:innen (2020) dargestellt:

Grundannahme I: Verknüpfungsannahme – Der Testwert y einer Person u für einen bestimmten Test oder ein bestimmtes Item i lässt sich als Summe des wahren Werts des gemessenen Merkmals τ_{ui} und eines unsystematischen Messfehlers ε_{ui} ermitteln:

$$y_{ui} = \tau_{ui} + \varepsilon_{ui} \quad (4.01)$$

Grundannahme II: Existenzannahme – Der wahre Wert eines latenten Merkmals τ_{ui} existiert als Mittelwert (Erfahrungswert E) des empirischen Testwerts. Würde also eine Person u den Test häufig wiederholen, entspräche der Mittelwert der Testwerte dem wahren Wert, da sich der Messfehler im Mittel zu 0 ausgleicht. Dies lässt sich beispielsweise durch zufällige Schwankungen in den Testungen erklären: besonders gute Konzentration beim ersten, Müdigkeit beim zweiten, Lärm beim dritten Durchgang usw.

$$E(y_{ui}) = E(\tau_{ui}) + E(\varepsilon_{ui}) \quad | \quad E(\varepsilon_{ui}) = 0 \quad (4.02)$$

$$E(y_{ui}) = E(\tau_{ui}) \quad | \quad E(\tau_{ui}) = \tau_{ui} = \text{konstant} \quad (4.03)$$

$$E(y_{ui}) = \tau_{ui} \quad (4.04)$$

Grundannahme III: Unabhängigkeitsannahme – Unsystematische Fehlereinflüsse sind unabhängig von der wahren Merkmalsausprägung, d. h. sie treten unabhängig davon auf, ob eine Person eine hohe oder niedrige Merkmalsausprägung aufweist. Daher besteht keine Korrelation (Corr) zwischen wahren Wert und Messfehlern:

$$\text{Corr}(\tau_{ui}, \varepsilon_{ui}) = 0 \quad (4.05)$$

Zusatzannahme I zur Unabhängigkeitsannahme – Wenn eine Person u die Items i und j wiederholt bearbeitet, sind die Messfehler der beiden Items unkorreliert. Beispielsweise kann das Item i missverstanden werden, während bei Item j die Konzentration nachgelassen hat:

$$\text{Corr}(\varepsilon_{ui}, \varepsilon_{uj}) = 0 \quad (4.06)$$

Zusatzannahme II zur Unabhängigkeitsannahme – Die Messfehler zweier beliebiger Personen u und m sind unkorreliert, wenn sie dasselbe Item i bearbeiten. Beispielsweise beeinflussen Konzentrationsschwankungen der Person u nicht die Befindlichkeit der Person m :

$$\text{Corr}(\varepsilon_{ui}, \varepsilon_{mi}) = 0 \quad (4.07)$$

Typischerweise wird davon ausgegangen, dass sämtliche Items dasselbe Konstrukt messen (Eindimensionalität). Zusätzlich wird von einem linearen Zusammenhang zwischen manifesten Variablen und der latenten Variable ausgegangen. Darauf basierend lässt sich eine gemeinsame latente Variable η definieren. Die Grundgleichung (4.01) wird dabei erst durch die Verwendung mehrerer Items, die dasselbe Konstrukt messen, zu einem Messmodell, mit dem sich Kennwerte der KTT bestimmen lassen. Dabei wird berücksichtigt, dass neben der Ausprägung der latenten Variable η auch die Eigenschaften des Items das Antwortverhalten einer Person beeinflussen. Dazu werden der Leichtigkeitsparameter α_i und der Trennschärfeparameter λ_i als Modellparameter zur Definition des wahren Werts τ_{ui} eingeführt (Döring, 2023c; Moosbrugger, Gädde, et al., 2020; Moosbrugger, Schermelleh-Engel, et al., 2020).

$$\tau_i = \alpha_i + \lambda_i \cdot \eta \quad (4.08)$$

Durch Einsetzen in die Grundgleichung (4.01) erhält man folgende Gleichung:

$$y_{ui} = \alpha_i + \lambda_i \cdot \eta + \varepsilon_{ui} \quad (4.09)$$

Um die Modellkonformität, also die Güte des Modelltests zu bewerten, wird der χ^2 -Test genutzt. Deuten die Ergebnisse auf einen guten Modellfit hin, können einzelne Modellparameter interpretiert und Hypothesen geprüft werden. Dies legitimiert auch die Bildung von Summenscores, um beispielsweise Testpersonen miteinander zu vergleichen. Dafür werden die Werte aller Items, die das gleiche Konstrukt messen, für eine Person aufsummiert (Döring, 2023c; Moosbrugger, Gädde, et al., 2020; Moosbrugger, Schermelleh-Engel, et al., 2020).

Da in der Messung der Testwerte Fehler enthalten sind, ist die Schätzung des Messfehleranteils anhand der Reliabilität notwendig. Diese wird als Verhältnis der Varianz des wahren Werts zur Varianz des Testwerts definiert. Hierzu wird klassischerweise Cronbachs Alpha (α) als Reliabilitätskoeffizient verwendet,

das jedoch sehr restriktive Modellannahmen mit Blick auf die Messäquivalenz der Items hat. Daher ist eine Prüfung der Güte im Vorhinein obligatorisch. Alternativ bietet sich das weniger restriktive McDonalds Omega (ω) an, das lediglich die Eindimensionalität und die Annahme unkorrelierter Messfehler voraussetzt (Döring, 2023c; Moosbrugger, Gädde, et al., 2020; Moosbrugger, Schermelleh-Engel, et al., 2020; Schermelleh-Engel & Gädde, 2020).

5.3.2 Die Item-Response-Theorie (IRT)

Wie auch in der KTT wird in der IRT davon ausgegangen, dass manifeste Variablen als Indikatoren genutzt werden können, um eine Aussage über die Ausprägung eines latenten, nicht direkt beobachtbaren Merkmals zu treffen. Die Grundidee ist, dass die Lösung einer Aufgabe zum einen von der Ausprägung des latenten Merkmals einer Person und zum anderen von der Aufgabenschwierigkeit abhängt. In der IRT gibt es eine Vielzahl an Modellen, die sich mit Blick auf die Kategorienanzahl (z. B. dichotome und polytome Antwortkategorien) und die zu schätzenden Itemparameter (z. B. Trennschärfe-, Schwierigkeits- und Schwellenparameter) unterscheiden (Döring, 2023c; Moosbrugger, Schermelleh-Engel, et al., 2020). Alle Modelle verbindet aber die gleiche Grundannahme bzw. -voraussetzung: Lokale stochastische Unabhängigkeit. Es wird von der Messung eines eindimensionalen Konstrukts ausgegangen, d. h. die Items eines Tests messen ein und dieselbe latente Variable. Dabei darf die Beantwortung eines Items – neben der Aufgabenschwierigkeit – nur durch die Ausprägung dieser latenten Variable beeinflusst werden, d. h. das Lösen bzw. Nichtlösen eines vorangegangenen Items darf die Beantwortung eines Items nicht beeinflussen. Die Items müssen unabhängig voneinander sein (Moosbrugger, Schermelleh-Engel, et al., 2020).

Die IRT stellt einen Zusammenhang zwischen manifester Variable (Item) und latenter Variable her, die in einer kurvilinearen Beziehung stehen und durch die itemcharakteristische Funktion (IC-Funktion) beschrieben wird. Die Wahrscheinlichkeit, dass eine Person ein Item löst, hängt von der Merkmalsausprägung der Person und der Aufgabenschwierigkeit ab. Für jede Person kann eine Wahrscheinlichkeit bestimmt werden, mit der sie ein Item löst, und für jedes Item eine Wahrscheinlichkeit, dass es im Sinne des Konstrukts beantwortet wird. Die logistische IC-Funktion beschreibt, dass für Personen mit höherer Personenfähigkeit θ die Lösungswahrscheinlichkeit eines Items zunimmt und

asymptotisch gegen eins geht, während für Personen mit geringer Personenfähigkeit θ eine niedrigere Lösungswahrscheinlichkeit einhergeht, die asymptotisch gegen null geht. Die logistische Beziehung wird mit folgender Gleichung ausgedrückt (Döring, 2023c; Moosbrugger, Schermelleh-Engel, et al., 2020):

$$P(y_i = 1 | \eta) = \frac{e^{\lambda_i \cdot (\theta - \xi_i)}}{1 + e^{\lambda_i \cdot (\theta - \xi_i)}} \quad (4.10)$$

Dabei stellt y_i die manifeste Variable des Items i , e die Euler'sche Zahl ($e \approx 2.718$), λ_i den Trennschärfeparameter, θ die Personenfähigkeit und ξ_i den Schwierigkeitsparameter des Items i dar (Moosbrugger, Schermelleh-Engel, et al., 2020).

Für den Fall, dass alle Trennschärfeparameter der Items den Wert eins aufweisen, vereinfacht sich die IC-Funktion zur Funktion des Rasch-Modells (Rasch, 1960), was bedeutet, dass alle Items gleich gut zwischen Personen mit niedrigen und hohen Ausprägungen des latenten Merkmals unterscheiden. In der IC-Funktion des Rasch-Modells wird somit nur ein Parameter variiert, die Itemschwierigkeit, weshalb es auch als Ein-Parameter-Logistisches-Modell (1PL-Modell) bezeichnet wird. Zusätzlich könnten auch der Trennschärfeparameter λ_i (2PL-Modell) und der Rateparameter γ_i (3PL-Modell) berücksichtigt werden (Moosbrugger, Schermelleh-Engel, et al., 2020). Diese werden hier jedoch nicht weiter behandelt, sondern der Fokus wird auf die 1PL-Modelle gelegt, die für die Studien von Relevanz sind.

Im Falle des Rasch-Modells sind spezifisch objektive Vergleiche der Personenfähigkeit und der Itemschwierigkeit möglich, da diese differenzskaliert sind und sich auf einer gemeinsamen Skala abbilden lassen. Dadurch sind auch die Abstände beider Parameter interpretierbar. Das Rasch-Modell beschränkt sich dabei auf dichotome Items und wird daher auch als *dichotomes logistisches Latent-Trait-Modell* bezeichnet (Moosbrugger, Schermelleh-Engel, et al., 2020).

Im Rasch-Modell und anderen Modellen dieser Familie müssen die Parameter der Personenfähigkeit und Itemschwierigkeit auf Basis einer Datenmatrix und der darin enthaltenen Muster geschätzt werden, sodass die beobachteten Werte der manifesten Variable mit maximaler Wahrscheinlichkeit auf Basis von Gleichung (4.10) durch die Personenfähigkeit und die Itemschwierigkeit

vorausgesagt wird. Dafür werden in der Regel Maximum-Likelihood-Verfahren (ML-Verfahren) verwendet, die eine lokale stochastische Unabhängigkeit voraussetzen (Kelava & Moosbrugger, 2020). Dabei sollen hier vier unterschiedliche Schätzer kurz dargestellt werden:

- **Joint Maximum Likelihood Estimator (JMLE)** – Hier werden Personenfähigkeit und Itemschwierigkeit zusammen (*joint*) geschätzt. Die Personenfähigkeit wird dabei als fixe Größe betrachtet und nicht als zufällige Realisierung einer latenten Variable (Kelava & Moosbrugger, 2020; Nicklin & Vitta, 2022; Robitzsch, 2021).
- **Conditional Maximum Likelihood Estimator (CMLE)** – Dieses Verfahren schätzt zunächst die Itemparameter, wobei der Summenscore, also die Summe beantworteter Items, in die Schätzung einbezogen wird. Die Personenfähigkeiten müssen in einem zweiten, separaten Schritt geschätzt werden. Dadurch wird die Problematik einer verzerrten Schätzung bei zeitgleicher Schätzung der Personenfähigkeit und Itemschwierigkeit wie beim JMLE umgangen (Kelava & Moosbrugger, 2020; Nicklin & Vitta, 2022; Robitzsch, 2021).
- **Marginal Maximum Likelihood Estimator (MMLE)** – Die Schätzung geht davon aus, dass die Personenfähigkeit eine zufällige Realisierung der latenten Variable ist und einer bestimmten Verteilung folgt (z. B. Normalverteilung). Die Verteilung der Personenfähigkeiten wird bei der Schätzung der Itemschwierigkeiten zugrunde gelegt (Kelava & Moosbrugger, 2020; Nicklin & Vitta, 2022; Robitzsch, 2021). MML ist eine weit verbreitete Schätzmethode, da sie durch die separate Schätzung der Parameter und durch die Annahme einer bestimmten Verteilung der Personenfähigkeit sehr stabil und robust ist (Trendtel et al., 2016).
- **Weighted Likelihood Estimator (WLE)** – Dieser Schätzer stellt eine Alternative zu verschiedenen ML-Schätzern dar (Warm, 1989). Im Vergleich zu ML-Schätzern bietet der WL-Schätzer den Vorteil, dass dieser robuster gegenüber bestimmten Modellannahmen ist und systematische Verzerrungen reduziert (Trendtel et al., 2016; Warm, 1989).

Eine Erweiterung des Rasch-Modells stellt das *Partial-Credit-Modell* (PCM) dar (Masters, 1982), das die Annahmen des Rasch-Modells auf polytome Items überträgt, d. h. kategoriale Items mit geordneten Antwortkategorien (Döring, 2023c; Kelava & Moosbrugger, 2020). Dieses Modell lässt sich auch auf Tests

mit unterschiedlichen Antwortkategorien übertragen (Hamel et al., 2016). Daher wird das Modell auch als *Ordinales Latent-Trait-Modell* bezeichnet. Das PCM basiert ebenso wie das Rasch-Modell auf logistischen Funktionen, allerdings wird für jede Antwortkategorie eine Kategorienfunktion (CC-Funktion) bestimmt. Stellt man die Kategorienfunktionen eines Items graphisch dar, so lassen sich bei k Kategorien insgesamt $k-1$ Schnittpunkte zwischen den Funktionen erkennen, die sogenannten *Schwellenparameter (thresholds)*. Entspricht dieser Schwellenparameter der Personenfähigkeit, so ist die Wahrscheinlichkeit, dass sich die Person für Kategorien A oder B entscheidet, identisch. Dabei verlangt das Ordinale Latent-Trait-Modell nach geordneten Schwellenparametern. Es ist davon auszugehen, dass fähigere Personen mehr Punkte erreichen als weniger fähigere Personen, d. h. die Schwellenparameter müssen von der kleinsten zur höchsten Kategorie (Punktzahl) ebenfalls zunehmen, wobei die Abstände nicht äquidistant sein müssen (Döring, 2023c).

Ein Vorteil des Rasch-Modells und des PCM ist der Umgang mit fehlenden Daten. Solche können aus mannigfaltigen Gründen auftreten: zu schwierige Items, befristete Zeit, langsames Lesen usw. Der Vorteil der Modelle liegt darin, dass bei einzelnen fehlenden Werten einer Person die Personenfähigkeit auf Basis der beantworteten Items bestimmt wird und nicht beantwortete Items nicht einbezogen werden. Daher müssen fehlende Werte nicht etwa als falsch gelöst bewertet oder Personen mit Fehlwerten ausgeschlossen werden. So können die Personenfähigkeiten von Personen mit vollständigen und von Personen mit unvollständigen Datensätzen miteinander verglichen werden. Dennoch ist zu berücksichtigen, dass eine hohe Anzahl an fehlenden Werten für eine Person oder ein einzelnes Item die Schätzung des Modells beeinflussen kann (Boone et al., 2014).

Eine Erweiterung des Partial-Credit-Modells stellt das *multidimensionale Partial-Credit-Modell* dar. In vielen Messungen erfassen die Items nicht ein, sondern mehrere latente Merkmale. Mithilfe einer Q-Matrix werden die Items eines Messinstruments einer Dimension (*between-item-dimensionality*) oder mit entsprechender Gewichtung mehreren Dimensionen zugeordnet (*within-item-dimensionality*). Dies ermöglicht die Untersuchung von Konstrukten hinsichtlich ihrer Dimensionalität, indem die berechneten Modelle verglichen werden (Feuerstahler & Wilson, 2019; Lichti, 2019; Rost, 2004).

5.4 Psychometrische und statistische Methoden

Sowohl KTT als auch IRT stellen heute legitime und vergleichbare Testtheorien dar, um empirische Modellannahmen zu prüfen (Döring, 2023c; Moosbrugger, Schermelleh-Engel, et al., 2020). In dieser Arbeit wird die IRT verwendet, da hier der Vorteil gesehen wird, dass das (multidimensionale) Partial-Credit-Modelle aus der Rasch-Familie metrisch skalierte Personenfähigkeiten und Itemschwierigkeiten für unterschiedlich skalierte Items (dichotom und polytom) auf der gleichen linearen Skala abbildet und dadurch vergleichbar macht. Zudem stellt der Umgang mit fehlenden Daten im PCM einen Vorteil dar (Boone et al., 2014).

Bei der Frage zur Struktur der Repräsentationskompetenzen und zur empirischen Trennbarkeit vom Fachwissen ist ein übliches Vorgehen, die Annahme mithilfe konfirmatorischer Faktorenanalysen (KTT) zu prüfen. Hier werden jedoch multidimensionale PCM (*between-item-dimensionality*) verwendet und verglichen, da so eine methodische Inkonsistenz vermieden wird. Beide Vorgehen erzielen ähnliche und vergleichbare Ergebnisse (Demars, 2013; Peoples & O'Dwyer, 2014).

In dieser Arbeit wird das R[®] Package TAM (Robitzsch et al., 2024) zur Datenverarbeitung und -analyse verwendet. Dieses gibt verschiedene Parameter aus, mit deren Hilfe die Güte von Messinstrumenten beurteilt werden kann (Boone et al., 2014). In einem ersten Schritt wird das CRI:TIC (und die Fachwissensitems) in Bezug auf die theoretischen Grundannahmen des PCM geprüft, die lokale stochastische Unabhängigkeit. Anschließend wird die Güte der geprüften Modelle mithilfe verschiedener Item- und Modellfit Indizes untersucht. Dies legitimiert den zweiten Schritt, den Vergleich der (multidimensionalen) Partial-Credit-Modelle mit verschiedenen statistischen und graphischen Methoden zur Beantwortung der Forschungsfragen.

Schritt I – Evaluation der Messinstrumente

Q3-Statistik – Die Grundannahme der IRT ist, dass die Beantwortung von Items lediglich auf die Personenfähigkeit und Itemschwierigkeit zurückzuführen ist und nicht durch das Lösen bzw. Nichtlösen eines vorangegangenen Items beeinflusst wird. Diese lokale stochastische Unabhängigkeit lässt sich

mithilfe der *Q3*-Statistik (Yen, 1984) untersuchen. Hierfür werden die Residuen, also die Differenz des beobachteten und des vom Modell erwarteten Antwortverhaltens, aller Items eines Messinstruments korreliert. Da der Mittelwert der korrelierten Residuen aller Itempaare leicht unter null liegt, wird die korrigierte *aQ3*-Statistik in TAM ausgegeben. Dafür wird der Mittelwert von allen Korrelationskoeffizienten subtrahiert, sodass er null-zentriert ist. Von lokaler stochastischer Unabhängigkeit kann gesprochen werden, wenn $|aQ3| \leq 0.2$ ist. Darauf basierend müssen die Korrelationen aller Itempaare betrachtet und bei einer Überschreitung des kritischen Werts genau untersucht werden (Christensen et al., 2017; Robitzsch et al., 2024). Als globales Maß wird zusätzlich die *Mean Absolute Deviation of the adjusted Q3 statistic* (*MADaQ3*) verwendet. Dafür ist der Mittelwert der Beträge der korrelierten Residuen aller Itempaare zu bilden. Für $MADaQ3 \leq 0.1$ kann eine globale stochastische Unabhängigkeit angenommen werden (Boluarte-Carbajal et al., 2023; Hotchkiss et al., 2023; Romaní-Romaní, 2024).

Itemfit – Zur Evaluation des Itemfits werden *Mean Squares (MNSQ)* als Fit-Statistik verwendet, die prüfen, inwieweit die empirisch gewonnenen Daten mit den vom Modell angenommenen Daten übereinstimmen. Hier kann grundsätzlich zwischen *infit* und *outfit* *MNSQ* unterschieden werden. Der *infit* wird als *inlier-sensitives* Maß bezeichnet, d. h. es werden Antwortmuster berücksichtigt, bei denen Itemschwierigkeiten zur Personenfähigkeit passen. Der *outfit* ist dagegen ein *outlier-sensitives* Maß und anfällig für extreme Antwortmuster, bei denen Itemschwierigkeit und Personenfähigkeit nicht zueinander passen (Boone et al., 2014; Fulmer et al., 2021; Liu & McKeough, 2005). Die Werte für *infit* und *outfit* *MNSQ* nehmen Werte im Bereich um eins an. Werte > 1 weisen auf einen *underfit*, Werte < 1 auf einen *overfit* hin. Je nach Literatur finden sich unterschiedliche Intervalle zur Bewertung des *MNSQ*. Boone et al. (2014) empfehlen ein Intervall von $0.5 \leq MNSQ \leq 1.5$ zur adäquaten Bewertung des *infit* und *outfit* *MNSQ*. Zur Evaluation wird hier jedoch ein strikteres Maß im Intervall von $0.7 \leq MNSQ \leq 1.3$ herangezogen. Zusätzlich können die Z-standardisierten Werte (*ZSTD*) des *infits* und *outfits* einbezogen werden, wenn die *MNSQ*-Werte außerhalb des Intervalls liegen. Hierbei handelt es sich um einen *t*-Test, der prüft, ob die *MNSQ*-Berechnung zufällig ist. Werte im Intervall von $-1.96 \leq ZSTD \leq 1.96$ werden als unbedenklich angesehen ($p \geq .05$). Items mit auffälligen *MNSQ*- und *ZSTD*-Werten erfordern eine kritische Betrachtung (Boone et al., 2014).

Reliabilität – TAM gibt standardmäßig die EAP-Reliabilität (expected a-posteriori) aus. Hier wird jedoch die WLE-Reliabilität genutzt, da der WLE-Schätzer die Personenfähigkeiten auf Basis der Testwerte der jeweiligen Person bestimmt, während der EAP-Schätzer die gesamte Stichprobe in die Schätzung einbezieht. Daher ist die WLE-Reliabilität hier der geeignetere Koeffizient zur Beurteilung der internen Konsistenz (Trendtel et al., 2016). Die Werte können analog zu Cronbachs Alpha (α) interpretiert werden (Rost, 2004).

Modellfit – Um die Passung des gesamten Modells zu beurteilen, wird der *Standardized Root Mean Square Residual (SRMR)* verwendet. Dieser ist ein globales Maß und sollte nahe null sein. Hu und Bentler (1999) empfehlen als Grenze einen $SRMR \leq 0.08$, wobei auch ein Wert von ≤ 0.10 als akzeptabel angesehen wird (Gäde et al., 2020; Kline, 2016; Lichti, 2019).

Schritt II - Modellvergleiche

Informationskriterien (IC) – Der Vergleich der (multidimensionalen) Partial-Credit-Modelle wird hier mithilfe von Informationskriterien (*information criteria, IC*) vorgenommen. Dabei stellen das *Akaike Information Criterion (AIC)* und das *Bayes Information Criterion (BIC)* die bekanntesten Informationskriterien dar. Diese Parameter basieren auf der Log-Likelihood (LL) eines Modells (basierend auf der Likelihood-Funktion) und beziehen die Anzahl der geschätzten Parameter (n_P) mit ein, sodass Modelle mit vielen (unnötigen) Parametern bestraft werden. Dies wird vor dem Hintergrund, dass Modelle mit umfangreichen latenten Klassen in der Regel besser zu den Daten passen als ein sparsames Modell, einbezogen. Bei der Berechnung des BIC geht zusätzlich die Stichprobengröße (N) mit ein (Döring, 2023c; Lichti, 2019; Rost, 2004):

$$AIC = -2 \cdot LL + 2 \cdot n_P \quad (4.11)$$

$$BIC = -2 \cdot LL + n_P \cdot \ln(N) \quad (4.12)$$

Beide Informationskriterien neigen bei einer kleinen Stichprobe (N) zu einem *Overfit*. Anstelle des AIC wird bei einem Verhältnis $N \cdot n_P^{-1} < 40$ daher das *bias corrected Akaike Information Criterion (AICc)* empfohlen (Brewer et al., 2016; Burnham & Anderson, 2004; Sugiyara, 1978). Gleiches gilt für das BIC : Bei kleinen Stichproben (N) und einer hohen Anzahl an geschätzten Parametern (n_P)

neigt das *BIC* zum *Overfit*, weshalb dies in die Berechnung zur Anpassung einbezogen werden muss. Dadurch erhält man das *sample size adjusted Bayes Information Criterion (saBIC)* (Slove, 1987; Yang, 2006). Die Anpassung an kleine Stichproben erfolgt dabei auf unterschiedliche Art und Weise (in Rot hervorgehoben):

$$AICc = -2 \cdot LL + 2 \cdot n_P + \frac{2 \cdot n_P (n_P + 1)}{N - n_P - 1} \quad (4.13)$$

$$saBIC = -2 \cdot LL + n_P \cdot \ln\left(\frac{N+2}{24}\right) \quad (4.14)$$

Aufgrund der im Verhältnis kleinen Stichprobe und hohen Anzahl an geschätzten Parametern werden in dieser Arbeit das *AICc* und das *saBIC* zum Modellvergleich herangezogen.

Likelihood Ratio Tests – Dieser Test vergleicht zwei Modelle basierend auf der Log-Likelihood der geschätzten Modelle. Neben den Informationskriterien kann dadurch eine Aussage getroffen werden, ob der Unterschied zwischen den Modellen signifikant ist (Robitzsch, 2021).

Cross Plots – Neben den beiden vorgestellten statistischen Verfahren wird zusätzlich ein graphisches Verfahren genutzt, um einen Einblick in die Dimensionalität zu gewinnen. Dazu werden die Personenfähigkeiten aus zwei Modellen gegeneinander aufgetragen (CI = 95 %). Bei einem mehrdimensionalen Konstrukt wird davon ausgegangen, dass das Antwortverhalten unterschiedlich ist. Dementsprechend müssten viele Datenpunkte außerhalb der Messfehlerlinien liegen. Liegen sie innerhalb, kann dies als Indikator angesehen werden, dass beide Messungen das gleiche Konstrukt erfassen (Boone & Staver, 2020).

5.5 Datenerhebung und Stichprobe

Für die Datenerhebung wurde aus dem CRI:TIC und den Fachwissensitems ein Testheft gestaltet, in welchem die Items durchmischt und in einer randomisierten Reihenfolge abgebildet wurden. Dadurch sollte vermieden werden, dass durch externe Faktoren wie Zeitmangel kategorisch Informationen verloren gehen, wenn beispielsweise die Fachwissensitems zum Schluss nicht beantwortet werden. Das Testheft bestand aus drei Teilen. Zunächst erfolgte

eine kurze Projektvorstellung, die einerseits die angestrebten Ziele umriss, andererseits die Freiwilligkeit und Anonymität der Teilnahme unterstrich und schließlich den Umgang mit den erhobenen Daten beleuchtete. Zusätzlich wurden das Format der Items und die Vorgehensweise bei Korrekturen erläutert. Diese Informationen wurden zusätzlich mündlich vermittelt. Im zweiten Teil ging es um personenbezogene Variablen: Geschlecht, Alter, (geplanter) Studiengang und die letzte Chemienote auf einem Zeugnis. Die Items des CRI:TIC und für das Fachwissen stellten den dritten Teil, den Testteil, dar.

Für die Untersuchung des ersten Forschungsanliegens wurde als Stichprobe ein Chemie-Vorkurs für Studieneinsteiger:innen gewählt. Dieser fand in den ersten beiden Oktoberwochen 2023 für das Wintersemester 2023/24 statt. Zielgruppe des Vorkurses waren Studierende, die ein Fach mit Bezug zur Chemie studieren wollten. Dies hat den Vorteil, dass auch Personen aus Studiengängen vertreten sind, bei denen Chemie nicht direkt im Vordergrund steht, sodass davon ausgegangen werden kann, dass die Spannweite der Ausprägung der Repräsentationskompetenzen und des Fachwissens groß ist. Somit war nicht zu erwarten, dass die Ergebnisse durch mangelnde Varianz in eine Richtung verzerrt werden (Brandt & Moosbrugger, 2020).

Das Testheft wurde am ersten Tag des Vorkurses als Papier-und-Bleistift-Test durchgeführt, als Online-Test war es aus zweierlei Gründen problematisch: Zum einen hätte ein Großteil der 22 halboffenen Items trotzdem händisch ausgewertet werden müssen (beispielsweise skizzierte Strukturformeln), und zum anderen wäre zu gewährleisten gewesen, dass alle Teilnehmer:innen auf vergleichbare Endgeräte zugreifen können, um die Leistung nicht zu beeinflussen. Eine Durchführung am Handy wäre beispielsweise zum Skizzieren suboptimal gewesen. Aufgrund der Organisation des Vorkurses standen nach der Einführung in die Testung 75 Minuten zur Bearbeitung der insgesamt 56 Items zur Verfügung.

Um mit fehlenden Werten in den Datensätzen (*Missings*) optimal umzugehen, muss nach Rubin (1976) zwischen drei Arten von *Missings* unterschieden werden. Hängt das Fehlen eines Werts weder von der Ausprägung der Variable noch von der Ausprägung anderer Variablen ab, werden die fehlenden Werte als *Missing Completely At Random* (MCAR) bezeichnet. Beeinflusst die Ausprägung dritter Variablen das Fehlen eines Werts, nicht aber die Ausprägung der Variable selbst, werden diese als *Missing At Random* (MAR) bezeichnet.

Ist die Ausprägung der Variable selbst verantwortlich für das Fehlen des Werts, werden diese als *Missing Not At Random* (MNAR) bezeichnet (Lüdtke et al., 2007; Rubin, 1976).

Für die durchgeführte Datenerhebung werden zwei Annahmen getroffen: Fehlende Werte innerhalb des Testheftes wurden aufgrund der Ausprägung des latenten Merkmals von der Testperson bewusst nicht beantwortet oder übersprungen und stellen damit *Missing Not At Random* dar. Diese fehlenden Werte werden dementsprechend als falsch beantwortet gewertet. Aufgrund der begrenzten Zeit von 75 Minuten wurden fehlende Werte am Ende des Testhefts, d. h. auf die kein weiteres beantwortetes Item folgt, als *Missing At Random* gewertet. Für diese fehlenden Werte kann nicht ausgeschlossen werden, dass aufgrund der Lesegeschwindigkeit oder anderer Einflussfaktoren keine Zeit mehr zur Verfügung stand. Daher sind diese Werte als *Missing At Random* in die Analyse eingegangen. Die IRT stellt im Vergleich zur KTT in diesem Fall die bessere Alternative zur Auswertung der Daten dar, da diese die Fehlwerte adäquat berücksichtigt (Boone et al., 2014).

An der Datenerhebung haben $N = 185$ Studierende mit einem deutlich höheren Anteil an Frauen teilgenommen ($n_w = 130$ (70.3 %), $n_m = 53$ (28.6 %), $n_d = 2$ (1.1 %)). Die Studierenden waren zwischen 17 und 33 Jahren alt ($M_{\text{Alter}} = 19.3$ Jahre, $SD_{\text{Alter}} = 2.0$ Jahre) und haben Noten zwischen 3 Punkten (5+) und 15 Punkten (1+) als letzte Zeugnisnote in der Chemie angegeben ($M_{\text{Note}} = 11.7$ Punkte (2+), $SD_{\text{Note}} = 2.4$ Punkte). Abbildung 15 kann die Verteilung der Testpersonen auf die geplanten Studiengänge entnommen werden.

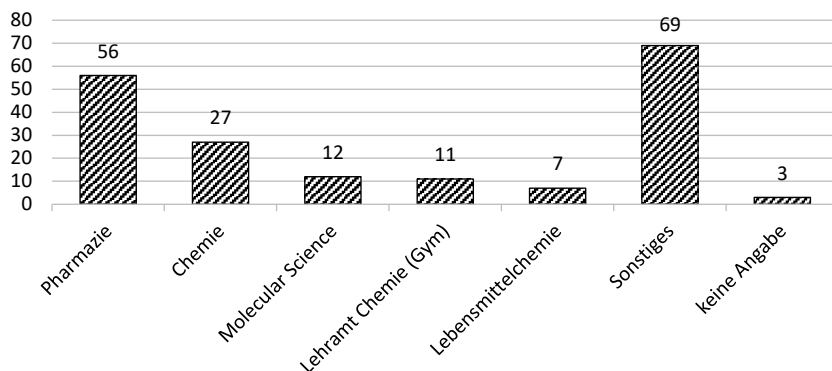


Abbildung 15: Verteilung der Studierenden auf die geplanten Studiengänge der Datenerhebung für das Forschungsanliegen I im Oktober 2023. „Sonstiges“ umfasst unter anderem die Studienfächer Biologie und Physik, weitere lehramtsbezogene Studiengänge sowie zusätzliche, nicht näher spezifizierte Studiengänge.

5.6 Ergebnisse

Die erhobenen Daten aus dem Chemie-Vorkurs werden, wie in Kapitel 5.4 beschrieben, analysiert, um die Forschungsfragen des ersten Forschungsanliegens zu beantworten. Zunächst erfolgen die Präsentation der Ergebnisse für FF1.1 und ein kurzes Fazit, da diese Ergebnisse das weitere Vorgehen für FF1.2 bestimmen. Anschließend werden die Ergebnisse der Analysen für FF1.2 vorgestellt und die Datenerhebung sowie die Messinstrumente hinsichtlich ihrer Güte mithilfe der Testgütekriterien bewertet. Abschließend kommt es zur Beantwortung der Forschungsfragen und -hypothesen anhand der dargestellten Ergebnisse.

5.6.1 Evaluation des CRI:TIC

In einem ersten Schritt wurde das dreidimensionale Grundmodell, das *Interpretation*, *Translation* und *Konstruktion* als separate latente Variablen behandelt, evaluiert. Dafür wurden zunächst die Grundannahmen des Modells geprüft. Mit Blick auf die lokale stochastische Unabhängigkeit ergibt sich, dass

die Residuen von zwei Drittel der Items mit mindestens einem anderen korrelieren. Insgesamt zeigen 22 der 703 Itempaare (3,12 %) korrelierte Residuen (Tabelle 6). Das globale Maß liegt bei $MADaQ3 = 0.071$ und kann als Indikator für die lokale stochastische Unabhängigkeit angesehen werden.

Mit Blick auf die Modellgüte lässt sich die Passung des Modells zu den Daten mit einem $SRMR = 0.100$ als noch akzeptabel beschreiben. Auf der Item-Ebene zeigen die Itemfit Statistiken einige Auffälligkeiten: Bei zwei Items sind die *infit* Werte kritisch, bei den gleichen Items und zwei weiteren liegen die *outfit* Werte außerhalb des akzeptablen Bereichs (Tabelle 6).

Es wird deutlich, dass der Test für die Stichprobe insgesamt zu schwierig war. Die Personenfähigkeiten (θ) für die drei Dimensionen liegen jeweils knapp unter 0 logits, während die Itemschwierigkeiten (ξ) für die drei Dimensionen über 0.3 logits liegen (Tabelle 5). Dabei wurden zwei Items (*Trans_04* und *Konst_04*) von weniger als fünf Prozent der Personen richtig beantwortet (Tabelle 6).

Tabelle 5: Vergleich der durchschnittlichen Personenfähigkeiten (θ) und Itemschwierigkeiten (ξ) für die *Interpretation*, *Translation* und *Konstruktion* des dreidimensionalen Grundmodells.

	Personenfähigkeiten (θ) in logits		Itemschwierigkeiten (ξ) in logits	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Interpretation</i>	-0.02	0.63	0.34	1.26
<i>Translation</i>	-0.02	0.72	0.57	1.33
<i>Konstruktion</i>	-0.08	0.91	0.81	2.07

Die WLE-Reliabilität kann für die drei Dimensionen als schlecht bis akzeptabel beschrieben werden, wobei die *Translation* die geringste ($Rel_{Trans} = 0.53$) und die *Konstruktion* die höchste ($Rel_{Konst} = 0.71$) WLE-Reliabilität aufweist. Die *Interpretation* liegt in einem noch akzeptablen Bereich ($Rel_{Int} = 0.68$).

Zwar zeigen der $MADaQ3$ und $SRMR$ auf globaler Ebene, dass lokale stochastische Unabhängigkeit und eine akzeptable Passung der Daten zum Modell angenommen werden können, auf Item-Ebene ergeben sich jedoch einige

Items mit kritischen Werten für die lokale stochastische Unabhängigkeit ($\alpha Q3$), sowie *infit* und *outfit* MNSQ und ZSTD (Tabelle 6). Auch die WLE-Reliabilität ist insgesamt als eher fragwürdig zu bewerten. Zudem ist der Test insgesamt zu schwierig. Daher müssen die Items, die kritische Werte zeigen, noch einmal betrachtet werden, um das CRI:TIC zu optimieren.

Tabelle 6: Evaluation der Items des CRI:TIC hinsichtlich der lokalen stochastischen Unabhängigkeit ($\alpha Q3$), Itemschwierigkeit (ξ) und Itemfit (*infit* und *outfit*).

Item	$\alpha Q3$ nur kritische Werte $ \alpha Q3 > .20$	ξ in logits	<i>infit</i>		<i>outfit</i>	
			MNSQ	ZSTD	MNSQ	ZSTD
Int_01	.25	-0.74	0.924	-1.278	0.896	-1.340
Int_02	-.22, -.22	-0.99	1.394	2.725	1.659	2.666
Int_03	-.24, .22, .26	0.31	0.893	-1.462	0.859	-1.330
Int_04		0.90	1.069	0.774	1.087	0.972
Int_05		0.71	0.958	-0.430	0.899	-0.709
Int_06	.21	1.94	0.951	-0.170	0.807	-0.639
Int_07		-1.04	1.070	0.924	1.126	1.228
Int_08	.26, .34, .48	2.12	0.873	-0.315	0.479	-1.149
Int_09		2.31	0.967	-0.117	0.879	-0.445
Int_10	.25	-1.16	0.941	-0.701	0.894	-0.951
Int_11		0.54	0.960	-0.734	0.941	-0.858
Int_12	-.24, -.22	-0.77	1.095	0.742	1.223	1.179
Int_13		1.67	0.966	-0.011	0.852	-0.284
Int_14	-.22	-1.05	0.992	-0.071	0.957	-0.379
Trans_01	-.22	1.12	0.991	-0.082	1.021	0.229
Trans_02		0.28	0.969	-0.683	0.978	-0.386
Trans_03	-.21, .21	-0.41	0.901	-2.146	0.882	-2.086
Trans_04	.48, .52	3.56	0.935	-0.048	0.588	-0.868
Trans_05		0.73	1.035	0.321	0.997	0.068
Trans_06	-.22, .34	1.05	0.955	-0.442	0.951	-0.426
Trans_07		-1.96	0.922	-0.501	0.815	-0.975

Trans_08	-.25, .23	0.05	1.415	3.965	1.512	4.630
Trans_09	.23	-0.33	1.018	0.403	1.021	0.385
Trans_10	-.21, .25, .34	1.73	0.906	-0.259	0.676	-0.574
Trans_11	.23	0.50	0.976	-0.372	0.964	-0.465
Konst_01		-0.46	1.060	0.559	1.103	0.608
Konst_02		-2.25	0.994	0.022	1.027	0.194
Konst_03	-.21, .23	1.08	1.144	1.228	1.158	0.865
Konst_04		4.32	1.016	0.214	1.034	0.286
Konst_05	.22, .34	0.22	0.896	-1.223	0.882	-1.262
Konst_06		2.95	1.016	0.149	0.972	0.037
Konst_07	.21, .23	0.71	0.957	-0.687	0.917	-1.011
Konst_08	-.27, -.22, -.25, .26	0.42	0.925	-0.686	0.805	-0.920
Konst_09	.34, .52	2.04	0.984	0.128	1.269	1.319
Konst_10	.21	-2.07	0.918	-0.490	0.833	-0.773
Konst_11		1.31	0.907	-0.776	0.926	-0.254
Konst_12	.23, .34	1.79	1.005	0.083	1.061	0.384
Konst_13	-.27	0.51	1.084	0.915	1.102	0.651

fett gedruckte Werte = kritische Werte; grau hinterlegte Items = ausgeschlossene Items

Alle Items mit kritischen Werten wurden erneut geprüft, besonders jene mit mehreren kritischen Werten. Auf dieser Basis wurden insgesamt fünf Items aus dem CRI:TIC ausgeschlossen:

Int_02 – Die Aufgabe bestand darin, das Zentralatom bzw. -ion in vier unterschiedlichen Komplexen (zwei als Kugel-Stab-Modell, zwei als Keilstrichformel dargestellt) zu markieren. Die *infit* und *outfit* Statistik deuteten mit einem deutlichen *Underfit* darauf hin, dass das Item unzureichend misst. Das könnte daran liegen, dass die Testpersonen aus der Aufgabenstellung die Antwort ableiten, nämlich dass das Zentralteilchen in der Mitte (zentral) liegen muss. Daher wird das Item entfernt.

Item_08 – In dem Item war das Molekülorbitalschema für ein Kohlenstoffmonoxidmolekül (CO) dargestellt, in welchem die Testpersonen das HOMO

(höchstes, mit Elektronen besetztes Molekülorbital) und das LUMO (niedrigstes unbesetztes Molekülorbital) markieren und entsprechend beschriften sollten. Hier könnte eine missverständliche Aufgabenstellung ein Hindernis dargestellt haben, was aus den Antworten der Testpersonen auch hervorgeht. Beispielsweise haben viele das insgesamt energieärmste Molekülorbital als LUMO und das energetisch höchste besetzte Orbital, das p_x -Atomorbital des Sauerstoffatoms, als HOMO markiert und beschriftet. Dies könnte auch darauf schließen lassen, dass wenig bis kein Vorwissen zu Molekülorbitalschemata besteht, da diese in den alten Lehrplänen, nach denen die Testpersonen noch unterrichtet wurden, nicht aufgegriffen werden. Aufgrund des starken, nicht erklärbaren *Overfits* und der hoch korrelierten Residuen mit den Items *Trans_04*, *Trans_10* und *Konst_09*, wurde das Item aus dem CRI:TIC ausgeschlossen.

Trans_04 – Dargestellt war ein *cis*-1,4-Dimethylcyclohexanmolekül in Sesselkonformation. Die Testpersonen sollten dieses in die Newman-Projektion übersetzen, wobei das Grundgerüst, das Cyclohexanmolekül, bereits vorgegeben war, sodass nur noch die zwei Methylgruppen ergänzt werden mussten. Dabei gab es mehrere Lösungsmöglichkeiten. Das Item zeigte einen hohen *Overfit* und stark korrelierte Residuen mit den Items *Int_08* und *Konst_09*. Zudem war das Item zu schwierig und wurde nur von sechs der 185 Testpersonen korrekt beantwortet. Dies könnte ein Hinweis darauf sein, dass dies keine klassische Translationsleistung war oder die Testpersonen eine der Repräsentationen nicht kannten. Daher wurde das Item aus dem CRI:TIC ausgeschlossen.

Trans_08 – In dem Item wurden fünf Molekülpaaire dargestellt. Die Testpersonen sollten für jedes Paar angeben, ob die Moleküle identisch oder ob sie Enantiomere bzw. Diastereomere sind, wobei eine Definition zu den beiden Begriffen gegeben wurde. Dieses Item erfragte zwei unterschiedliche Aspekte: ob die Moleküle identisch oder Isomere sind und um welche Art von Isomeren es sich handelt, wenn sie nicht identisch sind. Aus dieser Perspektive ist das Item kritisch zu betrachten und erklärt gleichzeitig die schlechten Itemfit Statistiken. Daher wurde das Item entfernt.

Konst_04 – Das Item zeigte ein Molekül ((2*R*,2'*R*)-Methylphenidat) als Skelettformel und vier an der vertikalen Achse „gespiegelte“ Stereoisomere des Me-

thylphenidats als Skelettformel, von denen das wirkliche Spiegelbild (Enantiomer) markiert werden sollte ((2S,2'S)-Methylphenidat). Das Item wurde von drei der 185 Personen korrekt beantwortet. Das Item war viel zu schwierig und wurde aus dem CRI:TIC ausgeschlossen.

Neben diesen fünf ausgeschlossenen Items wiesen noch einige andere jeweils einen einzelnen kritischen Wert auf. Dieser Umstand rechtfertigt allerdings keinen Ausschluss aus dem CRI:TIC. Die Modell- und Itemfit Statistiken sollten stets im Zusammenhang betrachtet werden. Außerdem zeigten sie allesamt keine inhaltlichen Auffälligkeiten, die die kritischen Werte erklären konnten. Zudem wird der Itempool durch den Ausschluss der fünf genannten Items bereits auf 33 Items dezimiert.

Der bereinigte Itempool mit 33 Items – zwölf Items für die *Interpretation*, neun für die *Translation* und zwölf für die *Konstruktion* – wurde im nächsten Schritt verwendet, um die Analysen zu wiederholen. Diese wurden jetzt auf das ein- und das zweidimensionale Modell ausgeweitet, um die drei Modelle zunächst hinsichtlich ihrer Güte zu vergleichen. Die Ergebnisse sind in Tabelle 7 dargestellt.

Tabelle 7: Vergleich der lokalen stochastischen Unabhängigkeit (*MADaQ3*, *aQ3*), der WLE-Reliabilität und des Modells (*SRMR*) sowie Itemfits (*outfit* und *infit MNSQ*) für die drei Modelle (FF1.1).

	1D Modell Int + Trans + Konst	2D Modell Int Trans + Konst	3D Modell Int Trans Konst
<i>MADaQ3</i>	0.071	0.071	0.071
<i>aQ3</i>	$-0.24 \leq aQ3 \leq 0.34$	$-0.28 \leq aQ3 \leq 0.33$	$-0.27 \leq aQ3 \leq 0.35$
<i>SRMR</i>	0.087	0.087	0.087
WLE-Reliabilität	$Rel = 0.86$	$Rel_{Int} = 0.66$ $Rel_{Trans+Konst} = 0.81$	$Rel_{Int} = 0.66$ $Rel_{Trans} = 0.52$ $Rel_{Konst} = 0.72$
<i>outfit MNSQ</i>	$0.787 \leq outfit \leq 1.471$	$0.790 \leq outfit \leq 1.363$	$0.800 \leq outfit \leq 1.394$
<i>infit MNSQ</i>	$0.879 \leq infit \leq 1.229$	$0.886 \leq infit \leq 1.210$	$0.890 \leq infit \leq 1.194$

Int = *Interpretation*, Trans = *Translation*, Konst = *Konstruktion*

Ein genauerer Blick auf die $\alpha Q3$ -Statistik zeigt, dass der Betrag der korrelierten Residuen von elf der insgesamt 528 (2,08 %) Itempaare $|\alpha Q3| > 0.20$ ist. Diese Itempaare sind bei den drei Modellen (1D, 2D und 3D) identisch und wurden noch einmal geprüft, um auszuschließen, dass die Antwort eines Items die Antwort eines anderen Items beeinflusst.

Die Betrachtung hat aufgezeigt, dass sich ein Teil der Itempaare (Int_1 ~ Int_10, Int_03 ~ Konst_08, Trans_06 ~ Konst_05, Trans_09 ~ Konst_12, Trans_10 ~ Konst_12, Konst_08 ~ Konst_13) inhaltlich überschneidet (etwa Bezug zu Energiediagrammen oder Isomerie), mit der gleichen Art von Repräsentation arbeitet (z. B. Skelettformel) oder Strukturen skizziert werden mussten. Bei den restlichen Itempaaren (Int_03 ~ Int_12, Int_06 ~ Konst_07, Int_14 ~ Konst_08, Trans_01 ~ Trans_06, Trans_11 ~ Konst_07) konnten in dieser Hinsicht keine Überschneidungen identifiziert werden. Es kann aus inhaltlicher Perspektive bestätigt werden, dass die Beantwortung eines Items nicht von der Beantwortung eines anderen Items abhängt. Gemeinsam mit dem *MADaQ3* (Tabelle 7) liegt für alle Modelle hohe Evidenz zur Annahme der lokalen stochastischen Unabhängigkeit vor.

Die globale Modellgüte hat sich durch den Ausschluss der fünf Items verbessert. Für alle drei Modelle kann die Güte auf Basis des *SRMR* als gut bezeichnet werden, die empirischen Daten entsprechen somit den mathematischen Modellen. Die WLE-Reliabilitäten für die *Interpretation* im zwei- und dreidimensionalen Modell liegen in einem noch akzeptablen Bereich, die der *Translation* im dreidimensionalen Modell ist jedoch als schlecht zu bewerten.

Das Item *Int_12* zeigt in allen Modellen einen *Underfit* in Bezug auf den *outfit MNSQ* (> 1.3). Da dies jedoch der einzige kritische Wert des Items ist und auch der *ZSTD* für das zwei- und dreidimensionale Modell im Rahmen liegt (< 1.96 , für das eindimensionale Modell > 1.96), wird das Item nicht ausgeschlossen.

Zusammenfassend lässt sich aus den Analysen schließen, dass die lokale stochastische Unabhängigkeit für alle drei Modelle angenommen werden kann und dass sie mit Blick auf Modell- und Itemfit in einem guten Wertebereich liegt. Lediglich die WLE-Reliabilität ist kritisch zu betrachten. Insgesamt rechtfertigen die vorliegenden Ergebnisse einen nachfolgenden Modellvergleich zur Untersuchung der Dimensionalität der Repräsentationskompetenzen.

5.6.2 Untersuchung der Trennbarkeit der lower-level skills der Repräsentationskompetenzen

Basierend auf der Log-Likelihood (LL), den geschätzten Modellparametern (n_P) und der Stichprobengröße (N) wurden die stichprobenadjustierten Informationskriterien $AICc$ und $saBIC$ berechnet. Der Vergleich zwischen den drei Modellen ergibt unterschiedliche Ergebnisse. Während das $AICc$ das ein- und zweidimensionale Modell gleichermaßen favorisiert, spricht das $saBIC$ für das zwei- sowie für das dreidimensionale Modell (Tabelle 8). Der Likelihood Ratio Test zeigt, dass sowohl das zweidimensionale Modell ($\chi^2(2) = 8.496, p = 0.014$) als auch das dreidimensionale Modell ($\chi^2(5) = 13.021, p = 0.023$) eine signifikant bessere Passung haben als das eindimensionale Modell. Zwischen dem zwei- und dreidimensionalen Modell ist kein Unterschied festzustellen ($\chi^2(3) = 4.524, p = 0.210$).

Tabelle 8: Vergleich der drei Modelle anhand des $AICc$ und des $saBIC$, basierend auf der Log-Likelihood (LL), der geschätzten Modellparameter (n_P) und der Stichprobengröße (N).

	LL	n_P	N	$AICc$	$saBIC$
1D Modell	-3747.93	68	185	7712.75	7635.46
2D Modell	-3743.68	70	185	7714.55	7631.07
3D Modell	-3741.42	73	185	7726.17	7632.71

fett gedruckte Werte = beste Werte

Neben den statistischen Werten wurden Cross Plots als graphische Methode zur Untersuchung der Dimensionalität der Repräsentationskompetenzen genutzt. Diese sind in Abbildung 16 dargestellt. Auf der Ordinate wurden dazu die Personenfähigkeiten aus dem eindimensionalen Modell gegen die Personenfähigkeiten aus dem dreidimensionalen Modell (obere drei Graphen in Abbildung 16) bzw. die Personenfähigkeiten aus dem zweidimensionalen Modell (unteren zwei Graphen in Abbildung 16) auf der Abszisse aufgetragen. In allen Diagrammen lässt sich erkennen, dass die Datenpunkte innerhalb der Messfehlerlinien liegen und hoch miteinander korreliert sind ($0.79 \leq r \leq 0.95$). Diese Befunde lassen sich so interpretieren, dass es keine Rolle spielt, welcher Pool an Items (z. B. *Interpretation*, *Translation* oder *Konstruktion* aus dem 3D

Modell) zur Messung der Repräsentationskompetenzen genutzt wird – sie liefern alle das gleiche Ergebnis. Die Cross Plots liefern somit Evidenz dafür, dass es sich bei den Repräsentationskompetenzen um ein eindimensionales Konstrukt handelt, in dem die Fähigkeiten nicht trennbar und hoch korreliert sind. Letzteres wird auch durch die Korrelation der Personenfähigkeiten der drei Fähigkeiten aus dem dreidimensionalen Modell bestätigt. Dabei korreliert die *Konstruktion* vergleichbar hoch mit der *Interpretation* ($r = 0.65$) und der *Translation* ($r = 0.62$), während die Korrelation zwischen der *Interpretation* und *Translation* etwas niedriger ist ($r = 0.56$).

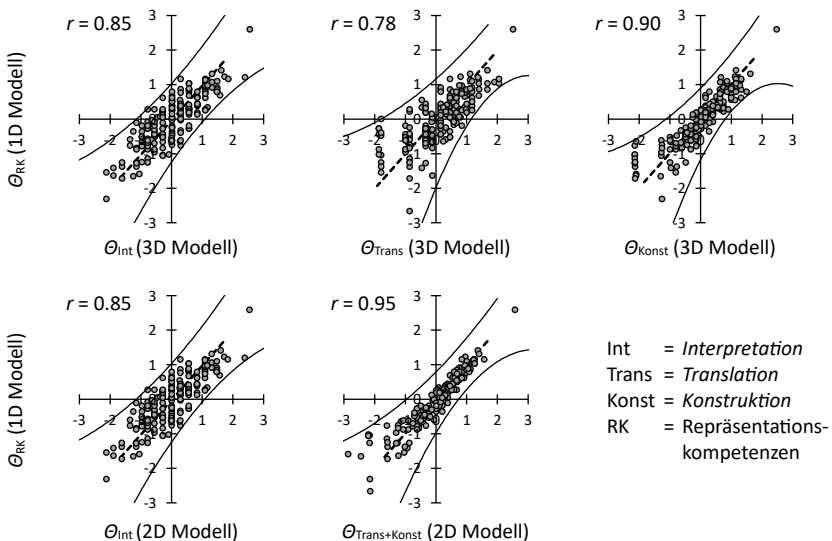


Abbildung 16: Cross Plots mit Messfehlerlinien (CI = 95 %) und Pearson-Korrelation (r) der Personenfähigkeiten (θ) aus dem eindimensionalen Modell (1D) auf der Ordinate und der Personenfähigkeiten aus dem dreidimensionalen Modell (3D, oben) bzw. dem zweidimensionalen Modell (2D, unten) auf der Abszisse.

5.6.3 Evaluation der Fachwissensitems

Zur Evaluation der 18 Fachwissensitems wurde analog zum CRI:TIC vorgegangen. Da von der Trennbarkeit der Repräsentationskompetenzen und des Fachwissens ausgegangen wird, erfolgte die Analyse des Fachwissensitems zunächst isoliert.

Zunächst wurde wieder die lokale stochastische Unabhängigkeit der Items untersucht. Von den insgesamt 153 Itempaaren zeigten sechs (3.92 %) korrelierte Residuen (Tabelle 9). Das globale Maß lässt mit $MADaQ3 = 0.073$ eine globale stochastische Unabhängigkeit annehmen.

Die Modellgüte kann bei einem $SRMR = 0.102$ als noch akzeptabel bewertet werden. Auf der Item-Ebene zeigt sich lediglich ein Item mit einem kritischen *Underfit* im *outfit MNSQ*, wobei der *outfit ZSTD* im Rahmen liegt (Tabelle 9).

Ähnlich wie bei der Evaluation des CRI:TIC zeigt sich auch hier, dass die Fachwissensitems zu schwierig für die Stichprobe waren. Die Personenfähigkeiten (θ) für das Fachwissen sind im Durchschnitt ($M_{\theta}(FW) = -0.03$ logits, $SD_{\theta}(FW) = 1.12$ logits) deutlich unter der durchschnittlichen Itemschwierigkeit (ξ) der Fachwissensitems ($M_{\xi}(FW) = 1.13$ logits, $SD_{\xi}(FW) = 1.20$ logits). Die schwierigsten Items ($\xi > 3$ logits) waren *Fach_01* und *Fach_11* (Tabelle 9). Die WLE-Reliabilität der Fachwissensitems lässt sich mit $Rel_{FW} = 0.62$ als fragwürdig bewerten.

Zusammenfassend lässt sich aus globaler Perspektive sagen, dass die Kennwerte der Fachwissensitems für eine globale stochastischen Unabhängigkeit sprechen und eine akzeptable Modellgüte besitzen. Dennoch sollten die Items mit kritischen Werten (Tabelle 9) erneut untersucht werden.

Tabelle 9: Evaluation der Fachwissensitems hinsichtlich der lokalen stochastischen Unabhängigkeit ($\alpha Q3$), Itemschwierigkeit (ξ) und Itemfit ($infit$ und $outfit$).

Item	$\alpha Q3$ nur kritische Werte $ \alpha Q3 > .20$	ξ in logits	$infit$		$outfit$	
			MNSQ	ZSTD	MNSQ	ZSTD
Fach_01	-.21, .26	3.01	0.961	-0.078	0.869	-0.218
Fach_02		0.15	0.957	-0.763	0.942	-0.751
Fach_03	-.21	0.24	0.953	-0.842	0.938	-0.789
Fach_04		1.53	1.107	1.017	1.147	0.882
Fach_05	-.23	-0.01	1.054	1.001	1.061	0.825
Fach_06		0.98	0.960	-0.502	0.912	-0.729
Fach_07		-0.80	0.951	-0.688	0.910	-0.851
Fach_08		1.64	0.946	-0.457	0.954	-0.197
Fach_09	-.23	0.77	1.118	1.689	1.168	1.579
Fach_10	-.21	-0.77	1.040	0.595	1.030	0.326
Fach_11		3.01	1.015	0.142	0.900	-0.132
Fach_12		0.59	1.023	0.380	1.072	0.791
Fach_13	.26	1.97	0.956	-0.284	0.874	-0.522
Fach_14		2.97	1.058	0.315	1.100	0.378
Fach_15		2.20	0.990	-0.011	0.908	-0.276
Fach_16	-.43	0.55	0.883	-1.902	0.848	-1.730
Fach_17	-.21	0.12	1.023	0.422	1.024	0.326
Fach_18	-.43, -.21, -.21	2.10	1.201	0.765	1.338	0.892

fett gedruckte Werte = kritische Werte; grau hinterlegte Items = ausgeschlossene Items

Mit Blick auf die Itempaare, die einen kritischen $\alpha Q3$ -Wert aufweisen, lässt sich kein Itempaar identifizieren, das aus inhaltlicher Sicht zusammenhängt. Item *Fach_18* weist einen leicht kritischen $outfit$ MNSQ auf. Da der $outfit$ ZSTD jedoch gut ist, wird das Item beibehalten. Das Item ist das letzte Item im Testheft und wurde von vielen Testpersonen nicht beantwortet (lediglich 32 Antworten). Da hier aufgrund des Zeitdrucks nicht ausgeschlossen werden kann, dass die Personen die Items aufgrund mangelnder Zeit nicht beantworten

konnten, wird das Item beibehalten und im Rahmen von Forschungsanliegen II erneut kritisch geprüft. Die unzureichende Datenlage für das Item kann Ursache für den *Underfit* erklären. Da die Fachwissensitems insgesamt viel zu schwierig für die Stichprobe sind, werden die zwei schwierigsten Items (*Fach_01* und *Fach_11*) entfernt, was auch durch die inhaltliche Überschneidung mit anderen Items gerechtfertigt ist.

Für die weiteren Analysen verblieben somit 16 Fachwissensitems. Diese wurden im nächsten Schritt mit den 33 Items des CRI:TIC in zwei Modellen verglichen, um die empirische Trennbarkeit der beiden Konstrukte zu prüfen. Dafür wurden die beiden in Kapitel 5.1 beschriebenen Modelle zur Untersuchung der Forschungsfrage (FF1.2) erneut hinsichtlich ihrer Güte überprüft.

Tabelle 10: Vergleich der lokalen stochastischen Unabhängigkeit (*MADaQ3*, *aQ3*), der WLE-Reliabilität und des Modells (*SRMR*) und Itemfits (*outfit* und *infit MNSQ*) für die zwei Modelle (FF1.2).

	1D Modell RK + FW	2D Modell RK FW
<i>MADaQ3</i>	0.072	0.070
<i>aQ3</i>	$-0.56 \leq aQ3 \leq 0.34$	$-0.53 \leq aQ3 \leq 0.35$
<i>SRMR</i>	0.100	0.098
WLE-Reliabilität	$Rel = 0.88$	$Rel_{RK} = 0.86$ $Rel_{FW} = 0.61$
<i>outfit MNSQ</i>	$0.734 \leq outfit \leq 1.534$	$0.784 \leq outfit \leq 1.481$
<i>infit MNSQ</i>	$0.870 \leq infit \leq 1.255$	$0.859 \leq infit \leq 1.247$

Mit Blick auf die *aQ3*-Statistik in Tabelle 10 lässt sich erkennen, dass teilweise korrelierte Residuen für Itempaare vorliegen. Die Matrix der 33 Items des CRI:TIC wurde bereits in Kapitel 5.6.1 betrachtet und wird hier nicht abermals erörtert. In der *aQ3*-Matrix der 16 Fachwissensitems lassen sich für beide Modelle insgesamt fünf Itempaare identifizieren, die bei erneuter Betrachtung in keinem logischen Zusammenhang stehen, wodurch ausgeschlossen werden kann, dass sich die Beantwortungen der Items gegenseitig beeinflussen. Schließlich wurde die *aQ3*-Matrix für die 33 Items des CRI:TIC mit den 16 Fachwissensitems betrachtet. Insgesamt zeigen sich hier 22 Itempaare mit kritischen Werten. Für einige Itempaare konnten entfernt inhaltliche Zusammen-

hänge erkannt werden, beispielsweise, dass beide Items chemische Reaktionen, Inhalte aus der Organischen Chemie oder aus der Komplexchemie behandeln, was jedoch auch als zufälliger Zusammenhang gewertet werden könnte und nicht zwangsläufig zu korrelierten Residuen führen muss (Fach_13 ~ Trans_05, Fach_03 ~ Trans_11, Fach_06 ~ Trans_06, Fach_15 ~ Konst_09, Fach_18 ~ Int_01 / Int_09 / Trans_10 / Konst_02 / Konst_06 / Konst_08). Für die anderen Itempaare konnte kein Zusammenhang erkannt werden (Fach_02 ~ Trans_11, Fach_03 ~ Trans_03 / Konst_03, Fach_10 ~ Trans_03, Fach_08 ~ Trans_11 / Konst_13, Fach_13 ~ Trans_11, Fach_14 ~ Trans_04, Fach_16 ~ Konst_11, Fach_17 ~ Konst_13, Fach_18 ~ Int_04 / Konst_12).

Zusammenfassend lässt sich somit sagen, dass inhaltlich keine eindeutig erkennbaren Zusammenhänge zwischen den Itempaaren vorliegen und die Korrelation der Residuen mit hoher Wahrscheinlichkeit eher zufällig ist. Besonders das Fachwissensitem *Fach_18* zeigt viele Korrelationen in der *aQ3*-Matrix, was sich wahrscheinlich wieder auf die schwache Datenlage für das Item zurückführen lässt. Zusammen mit der *MADaQ3*-Statistik für beide Modelle kann somit mit hoher Evidenz von einer lokalen stochastischen Unabhängigkeit der Items ausgegangen werden.

Mit Blick auf die Werte des *outfit MNSQ* zeigen sich auch hier zwei Items, die kritische Werte annehmen. Zum einen das Item *Int_12*, das bereits in Kapitel 5.6.1 betrachtet wurde, und Item *Fach_18*, das schon bei der ersten Evaluation der 18 Fachwissensitems auffällig war, jedoch bewusst nicht verworfen wurde.

Zusammenfassend kann aus Tabelle 10 abgeleitet werden, dass beide gerechneten Modelle einen akzeptablen Modell- und Itemfit aufweisen. Die Reliabilität ist ebenfalls gut, mit Ausnahme für das Fachwissen im zweidimensionalen Modell. Insgesamt legitimieren diese Ergebnisse somit weitere Analysen und den Vergleich der beiden Modelle.

5.6.4 Untersuchung der Trennbarkeit von Fachwissen und Repräsentationskompetenzen

In Tabelle 11 sind die Informationskriterien *AICc* und *saBIC* für die beiden untersuchten Modelle dargestellt. Im Gegensatz zur Untersuchung der Trennbarkeit der lower-level skills der Repräsentationskompetenzen, zeigt sich hier ein eindeutiges Bild. Sowohl *AICc* als auch *saBIC* favorisieren deutlich das zweidimensionale Modell, was durch den Likelihood Ratio Test bestätigt wird ($\chi^2(2) = 39.097, p < 0.001$). Die statistischen Analysen sprechen somit für die empirische Trennbarkeit der Repräsentationskompetenzen und des Fachwissens in dieser Studie.

Tabelle 11: Vergleich der zwei Modelle anhand des *AICc* und des *saBIC*, basierend auf der Log-Likelihood (*LL*), der geschätzten Modellparameter (n_p) und der Stichprobengröße (*N*).

	<i>LL</i>	n_p	<i>N</i>	<i>AICc</i>	<i>saBIC</i>
1D Modell	-5252.67	84	185	10816.14	10677.80
2D Modell	-5233.12	86	185	10790.94	10642.80

fett gedruckte Werte = beste Werte

Zusätzlich sind auch hier noch einmal die Cross Plots erstellt und betrachtet worden. Dafür wurden auf der Ordinate die Personenfähigkeiten (θ) aus dem eindimensionalen Modell gegen die Personenfähigkeiten (θ) für die Repräsentationskompetenzen und das Fachwissen aus dem zweidimensionalen Modell auf der Abszisse aufgetragen. Die Korrelation zwischen den Personenfähigkeiten des eindimensionalen Modells und der Repräsentationskompetenzen ist erwartbar hoch (Abbildung 17, links). Interessant ist das Diagramm mit dem Fachwissen (Abbildung 17, rechts): Die Datenpunkte sind viel stärker gestreut und zeigen im Vergleich zu den Repräsentationskompetenzen eine deutlich geringere Korrelation (um $\Delta r \approx 0.30$ kleiner). Zudem lässt sich erkennen, dass einige Datenpunkte außerhalb der Messfehlerlinien ($CI = 95\%$) liegen. Diese Ergebnisse können als zusätzlicher Indikator gesehen werden, dass die Fachwissensitems ein von den Items zur Erhebung der Repräsentationskompetenz distinktes Konstrukt messen, beide Konstrukte aber trotzdem zusammenhängen, was durch die Korrelation der Personenfähigkeiten für die Repräsentationskompetenzen und das Fachwissen aus dem zweidimensionalen Modell gezeigt wird ($r = 0.51$).

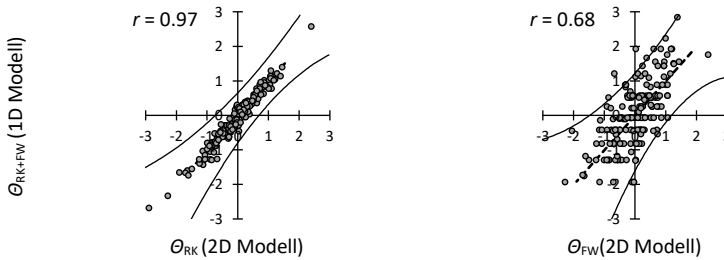


Abbildung 17: Cross Plots mit Messfehlerlinien (CI = 95 %) und Pearson-Korrelation (r) der Personenfähigkeiten (Θ) aus dem eindimensionalen Modell (1D) auf der Ordinate und der Personenfähigkeiten für die Repräsentationskompetenzen (RK) und das Fachwissen (FW) aus dem zweidimensionalen Modell (2D) auf der Abszisse.

5.6.5 Bewertung der Gütekriterien

Neben den bereits vorgestellten Ergebnissen wurden weitere Analysen durchgeführt, um die Güte der Testung mit Blick auf die Haupt- und Nebengütekriterien zu bewerten. Diese werden im Folgenden kurz dargelegt.

Objektivität – Die Objektivität der Messung lässt sich in dieser Studie als hoch einstufen. Dies wurde durch die Verwendung von geschlossenen und halboffenen Items sowie der Verwendung eines präzisen Testmanuals gewährleistet. Von den 185 Datensätzen wurden 33 (18 %) zufällig gewählte Testhefte von einer zweiten Person bewertet und ein Übereinstimmungsmaß zwischen den zwei Ratern mit Hilfe von *Cohen's Kappa* (κ) ermittelt (Cohen, 1960). Für den CRI:TIC hat sich eine Übereinstimmung von $\kappa_{RK} = 0.91$ und für die Fachwissensitems eine Übereinstimmung von $\kappa_{FW} = 1.00$ zwischen den Ratern ergeben, was als eine nahezu perfekt bzw. perfekte Übereinstimmung interpretiert (Landis & Koch, 1977) und als Maß für die *Auswertungsobjektivität* betrachtet werden kann. Durch die Verwendung von (multidimensionalen) Partial-Credit-Modellen wurden Personenfähigkeiten und Itemschwierigkeiten differenzskaliert auf der gleichen linearen Skala abgebildet (Moosbrugger, Schermelleh-Engel, et al., 2020), was für eine hohe *Interpretationsobjektivität* spricht. Mit Blick auf die *Durchführungsobjektivität* lässt sich diese hier nur theoretisch begründen, da das Testheft alle relevanten Informationen zur Studie allgemein und zum Vorgehen bei der Bearbeitung umfasst. Daher sollte die Testung auch unabhängig von der durchführenden Person sein.

Reliabilität – Zur Bewertung der Reliabilität wurde ein Reliabilitätskoeffizient mit Hilfe des WLE-Schätzers ermittelt. Er hat den Vorteil, robuster gegenüber bestimmten Modellannahmen und systematischen Verzerrungen zu sein (Trendtel et al., 2016; Warm, 1989). Dieses Maß wurde in Kapitel 5.6.1 für die überarbeitete Version des CRI:TIC (mit 33 Items) präsentiert und liegt für die unterschiedlichen Modelle und Dimensionen in einem schlechten bis guten Bereich.

Validität – Die Validität stellt das zentrale Gütekriterium dar. Durch ausführliche Literaturrecherche, kritische Evaluation bestehender Messinstrumente und Orientierung an anderen Studien in diesem Kontext konnten Kriterien zur Adaption und Konstruktion von repräsentationsbasierten Items für den CRI:TIC und für die Fachwissensitems herausgearbeitet werden. Dadurch wird eine hohe *Inhaltsvalidität* erreicht. Mit Blick auf den CRI:TIC und zur Frage nach der Trennbarkeit der *Interpretation*, *Translation* und *Konstruktion*, wurden die 38 repräsentationsbasierten Items des CRI:TIC durch drei Rater nach einer kurzen mündlichen Instruktion diesen Fähigkeiten zugeordnet. Eine Bestimmung der Übereinstimmung der Zuordnung durch die drei Rater mithilfe von *Fleiss Kappa* ergab eine mit $\kappa_{\text{Fleiss}} = 0.86$ nahezu perfekte Übereinstimmung, die hier als Maß für die *Konstruktvalidität* gesehen werden kann, da die Items mit hoher Übereinstimmung einer der drei lower-level skills zugeordnet werden konnten. Des Weiteren können die Ergebnisse zur Trennbarkeit der Repräsentationskompetenzen und des Fachwissens als Maß für die Diskriminanzvalidität gewertet werden. Die Untersuchungen zeigen, dass ein zweidimensionales Modell im Vergleich zum eindimensionalen Modell eine deutlich bessere Passung hat und die Konstrukte empirisch trennbar sind, aber dennoch hoch miteinander korrelieren ($r = 0.51$). Zur Konvergenzvalidität können hier keine Aussagen getroffen werden, da keine entsprechenden zusätzlichen Messinstrumente eingesetzt wurden.

Skalierung – Durch die Verwendung der IRT wurden metrisch skalierte Personenfähigkeiten und Itemschwierigkeiten berechnet und auf derselben Skala abgebildet, sodass die Werte direkt miteinander vergleichbar waren. Wenn die Passung der empirisch gewonnenen Daten zum Modell bestätigt wird, ist das Kriterium der Skalierung als erfüllt anzusehen, was hier der Fall ist (Krumm et al., 2021; Moosbrugger & Kelava, 2020).

Normierung – Das Güterkriterium der Normierung hat in dieser Untersuchung aus verschiedenen Gründen wenig Beachtung gefunden. Die Stichprobe ist zwar als repräsentativ zu sehen, jedoch lediglich an einem Standort erhoben worden und daher für eine Eichstichprobe zu klein. Zudem ist das Ziel dieser Arbeit die Entwicklung eines Messinstruments, mit dem sich die Struktur der Repräsentationskompetenzen und der Zusammenhang mit dem Fachwissen untersuchen lässt (Querschnittsstudie) und kein Instrument zur Individualdiagnostik. Daher ist die Normierung hier ein untergeordnetes Kriterium (Krumm et al., 2021; Moosbrugger & Kelava, 2020).

Testökonomie – Optimal wäre eine computergestützte Datenerhebung gewesen. Da die halboffenen Items teilweise Skizzen erfordern, war dies nicht möglich. Es musste sichergestellt sein, dass alle Teilnehmer:innen vergleichbare Endgeräte für die Durchführung nutzen konnten. Optimal wäre die Bereitstellung von Endgeräten gewesen, beispielsweise in Form von Tablets, um Störvariablen zu eliminieren. Bei einem Klassensatz von 30 Tablets hätte die Datenerhebung ungefähr sechs Durchgänge benötigt. Die Auswertung der 22 halboffenen Items hätte dennoch durch eine Person vorgenommen werden müssen (etwa die Bewertung der Richtigkeit von Skizzen). Daher wurde der Test als Papier-und-Bleistift-Test durchgeführt, wodurch unvermeidbar Druckkosten in überschaubarer Höhe entstanden sind. Mit Unterstützung von zwei wissenschaftlichen Hilfskräften wurden die Testhefte bewertet und die Daten in SPSS überführt.

Nützlichkeit – Mit Blick auf die Nützlichkeit lässt sich aus der aktuellen Perspektive sagen, dass der CRI:TIC ein valides Instrument zur Messung von Repräsentationskompetenzen ist, das Mängel aus anderen Messinstrumenten in diesem Kontext aufgreift und berücksichtigt. Auch mit Blick auf den ORCA (Ward et al., 2025), der ein vergleichbares Instrument zum CRI:TIC darstellt, wird beispielsweise das Repräsentationsdilemma in dieser Arbeit stärker berücksichtigt und lässt durch ein fachübergreifendes Design und die Verwendung verschiedener Repräsentationen eine allgemeingültigere Aussage zur Trennbarkeit der Repräsentationskompetenzen zu. Somit ist die Nützlichkeit des Instruments aktuell als hoch zu bewerten, da es eine valide Messung und Adressierung aktueller Forschungslücken und -interessen zulässt.

Zumutbarkeit – Vor der Testung wurden sämtliche Anforderungen transparent dargestellt und betont, dass die Teilnahme freiwillig und anonym ist. Die

anschließende Testbearbeitung hat 75 Minuten in Anspruch genommen. Daher wird die Testung hier als zumutbar bewertet.

Unverfälschbarkeit – Prinzipiell können Testpersonen ihre Testwerte auf unterschiedliche Weise verfälschen: Sie können antworten raten (besonders bei geschlossenen Formaten), von anderen abschreiben, sich „dumm stellen“ (*faking bad*) oder nach sozialer Erwünschtheit antworten (*faking good*). Letzteres kann hier ausgeschlossen werden, da dies hauptsächlich bei Persönlichkeitsfragebögen eine Rolle spielt (Krumm et al., 2021; Moosbrugger & Kelava, 2020). Den Testpersonen wurde mitgeteilt, dass die individuellen Ergebnisse wichtig sind und sie nicht von anderen abschauen und Antworten nicht raten sollten. Einzelne Personen wurden während der Testung erneut darauf hingewiesen. Außerdem wurde betont, dass die Testung anonym ist und der Test so konzipiert wurde, dass sie bei einigen Items an ihre Grenzen kommen könnten. So wurde die Unverfälschbarkeit nach besten Möglichkeiten adressiert, indem den Studierenden der Leistungsdruck genommen und die Anonymität versichert wurde, dennoch kann sie nicht gänzlich gesichert werden.

Fairness – Zur Prüfung der Fairness sind *Differential Item Functioning (DIF)* Analysen im Rahmen der IRT etablierte Verfahren. Dadurch kann geprüft werden, inwieweit Items unterschiedlich für verschiedene Subgruppen funktionieren (z. B. Geschlecht oder Alter). Dieses Verfahren war hier jedoch nicht sinnvoll, da durch die polytomen Items die Modellkomplexität bei einem Gruppenvergleich erhöht wird und die gebildeten Subgruppen viel zu klein waren (beispielsweise haben nur 53 Männer teilgenommen). Die Ergebnisse der *DIF*-Analyse wären dadurch schwierig zu interpretieren und nicht belastbar. Hier ist jedoch nicht davon auszugehen, dass die Items des CRI:TIC oder die Fachwissensitems für Subgruppen wie das Geschlecht oder das Alter unterschiedlich messen, da die Items aus bereits bestehenden Messinstrumenten adaptiert bzw. mithilfe dieser konzipiert wurden und lediglich chemische Inhalte und chemische Repräsentationen behandelt wurden. Daher spielt dieses Gütekriterium hier eine untergeordnete Rolle.

5.7 Zusammenfassung und Diskussion der Ergebnisse

Im ersten Forschungsanliegen steht die Untersuchung der Struktur der Repräsentationskompetenzen, d. h. die Untersuchung der empirischen Trennbarkeit der drei lower-level skills *Interpretation*, *Translation* und *Konstruktion* im Vordergrund (FF1.1). Dafür wurde zunächst das *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC) entwickelt und evaluiert. Zusätzlich wird mit Blick auf das Repräsentationsdilemma geprüft, inwieweit sich Repräsentationskompetenzen und chemisches Fachwissen empirisch trennen lassen (FF1.2). Dafür wurden zusätzlich zum CRI:TIC Fachwissenstests eingesetzt.

FF1.1 Inwieweit lassen sich die drei lower-level skills der Repräsentationskompetenzen *Interpretation*, *Translation* und *Konstruktion* empirisch trennen?

Im ersten Schritt wurde, basierend auf dem dreidimensionalen Modell, das die drei lower-level skills als separate latente Variablen betrachtet, das CRI:TIC evaluiert. Die Evaluation zeigt, dass etwa zwei Drittel der Items kritische Werte (Tabelle 6) in Bezug auf die lokale stochastische Unabhängigkeit ($\alpha Q3$) und bzw. oder den Itemfit (*infit* und *outfit* MNSQ und ZSTD) aufweisen. Der globale Modellfit (SRMR) ist dabei als akzeptabel zu bezeichnen. Zusätzlich hat sich der Test als zu schwierig für die Stichprobe erwiesen, da die mittleren Personenfähigkeiten für die drei Dimensionen (*Interpretation*, *Translation* und *Konstruktion*) des geprüften Modells kleiner als die durchschnittliche Itemschwierigkeit der Dimensionen sind. Die Items mit den kritischen Werten wurden genau geprüft und schlussendlich fünf Items identifiziert, die aufgrund einer Kombination aus schlechtem Itemfit (*infit* und *otufit* MNSQ und ZSTD), hoher Itemschwierigkeit, inhaltlicher Diskrepanzen und/oder missverständlicher Aufgabenstellungen aus dem CRI:TIC entfernt wurden.

Mit den 33 verbleibenden Items (je zwölf Items für die *Interpretation* und *Konstruktion*, neun Items für die *Translation*) wurden die Analysen für das dreidimensionale Modell wiederholt und für das eindimensionale Modell (Zusammenfassung der Items zu einer latenten Variable) sowie für das zweidimensionale Modell (*Interpretation* als eine latente Variable und *Translation* und

Konstruktion als zusammengefasste latente Variable) analog durchgeführt. Sowohl auf globaler Ebene (*MADaQ3*) als auch auf Item-Ebene (*aQ3*) lässt sich mit hoher Evidenz sagen, dass von lokaler stochastischer Unabhängigkeit für alle drei Modelle ausgegangen werden kann, wodurch die Grundannahme der gerechneten Modelle erfüllt ist. Ebenso zeigen die drei Modelle einen guten Modellfit (*SRMR*) und gute Itemfit Statistiken (*infit* und *outfit* *MNSQ* und *ZSTD*). Zwar weist das Item *Int_12* noch einen leichten *Underfit* auf (*outfit* *MNSQ*), dieser kann jedoch aufgrund des *outfit* *ZSTD* als zufällig beschrieben werden, weshalb das Item für die weiteren Analysen verwendet wird (Boone et al., 2014). Dadurch werden die folgenden Modellvergleiche legitimiert.

Um das Modell mit der besten Passung zu identifizieren, wurden mithilfe von Informationskriterien (*AICc*, *saBIC*) und Likelihood Ratio Tests Vergleiche durchgeführt und zusätzlich Cross Plots als graphische Methode zur Untersuchung der Dimensionalität eingesetzt. Die Ergebnisse zeigen ein ambivalentes, uneindeutiges Bild. Das eindimensionale Modell wird durch das *AICc* und die Cross Plots, das zweidimensionale Modell durch das *saBIC* und den Likelihood Ratio Test sowie das dreidimensionale Modell ebenfalls durch den Likelihood Ratio Test favorisiert. Mit Blick auf die Informationskriterien und den Likelihood Ratio Test lässt sich anmerken, dass diese auf der Log-Likelihood der Modelle basieren, wobei die adjustierten Informationskriterien *AICc* und *saBIC* zusätzlich noch die Stichprobengröße (*N*) und die Anzahl der geschätzten Parameter (*n_p*) einbeziehen, was bei dem Verhältnis (*N* zu *n_p*) in dieser Untersuchung sinnvoll ist (Brewer et al., 2016; Yang, 2006). Daher sollten diese hier stärker gewichtet werden als die Likelihood Ratio Tests. Da aber auch diese ein ambivalentes Bild zeigen, muss in diesem Fall das einfachere Modell angenommen werden, das auch von den Cross Plots favorisiert wird. Daher wird für diese Datenerhebung das eindimensionale Modell angenommen, das die drei lower-level skills als eine latente Variable zusammenfasst. Damit muss die Hypothese H1.1 verworfen werden. Die -level skills lassen sich in dieser Studie empirisch nicht trennen.

Insgesamt können die Datenerhebung und die eingesetzten Messinstrumente als objektiv und valide bezeichnet werden. Diese Kriterien werden vor allem durch die theorie- und kriteriengeleitete Konstruktion bzw. Adaption geschlossener und halboffener Items gewährleistet sowie durch die Auswertung mit entsprechenden Manualen. Die Reliabilität – hier mithilfe der WLE-Reliabilität gemessen – ist mit Blick auf die mehrdimensionalen Modelle teilweise

jedoch als kritisch zu bewerten (Tabelle 7). Ein Erklärungsansatz ist die geringe Itemanzahl von neun bzw. zwölf Items für die einzelnen Dimensionen, die die latente Variable nicht ausreichend abdecken und so zu einer unpräzisen Schätzung führen (Schermelleh-Engel & Gädde, 2020). Faktoren wie die lokale stochastische Unabhängigkeit oder die Itemqualität können hier aufgrund einer gründlichen Prüfung nahezu ausgeschlossen werden. Für das hier angenommene eindimensionale Modell kann die WLE-Reliabilität jedoch als gut bewertet werden. Zusammengefasst lässt sich das CRI:TIC nach Ausschluss der fünf Items als objektives, reliables und valides Messinstrument zur Erfassung der Repräsentationskompetenzen mit guten Modell- und Itemfit Statistiken charakterisieren.

Die Ergebnisse der Modellvergleiche zur Untersuchung der Struktur der Repräsentationskompetenzen decken sich mit den Befunden von Scheid und Kolleg:innen (2018), Taskin et al. (2015) und der Arbeitsgruppe um Ward (2025), die in ihren Studien verschiedene Facetten der Repräsentationskompetenzen empirisch ebenfalls nicht trennen konnten. Die in diesen Studien eingesetzten Messinstrumente werden hier jedoch kritisch betrachtet, da sie das Repräsentationsdilemma nicht adäquat berücksichtigen und einige Items potenziell eher Fachwissen als Repräsentationskompetenzen messen, was die Validität der Instrumente in Frage stellt und somit auch die Belastbarkeit der Ergebnisse dieser Untersuchungen. Mit den Ergebnissen des CRI:TIC, das dieses Problem stärker in den Fokus rückt, um eine valide, unverzerrte Messung der Repräsentationskompetenzen (und des Fachwissens) zu ermöglichen, wird hier jedoch weitere Evidenz geliefert, die die Ergebnisse untermauert, dass sich die Repräsentationskompetenzen empirisch nicht separieren lassen. Ward und Kolleg:innen (2025) betonen dabei, dass es sich bei den Kompetenzen um stark zusammenhängende, interkorrelierte Fähigkeiten handelt, was auch in dieser Studie bestätigt werden kann. Betrachtet man die Korrelation der *Interpretation*, *Translation* und *Konstruktion* aus dem dreidimensionalen Modell, lässt sich ein hoher Zusammenhang zwischen den drei lower-level skills erkennen ($0.56 \leq r \leq 0.65$). Dieser Zusammenhang lässt sich beispielsweise dadurch erklären, dass eine Referenzrepräsentation für die Übersetzung in eine Zielrepräsentation zunächst interpretiert werden muss. Gleiches gilt für die *Konstruktion* einer Repräsentation auf Basis einer Referenzrepräsentation (Talanquer, 2022; Ward et al., 2025). Diese Hierarchisierung ist jedoch zu-

nächst hypothetisch und bedarf weiterer Forschung. Das Design und die Ergebnisse dieser Studie lassen dazu keine belastbare Aussage zu. Daher kann zunächst nur bestätigt werden, dass die lower-level skills der Repräsentationskompetenzen stark zusammenhängende, empirisch nicht trennbare Fähigkeiten sind.

An dieser Stelle müssen die empirisch gewonnenen Ergebnisse jedoch zu der inhaltlich-theoretischen chemiedidaktischen Perspektive kontrastiert werden. Auch wenn sich hier das Bild festigt, dass Repräsentationskompetenzen ein Bündel stark vernetzter, untrennbarer Fähigkeiten darstellen, ist eine getrennte Betrachtung für den Lehr-Lern-Prozess durchaus sinnvoll. Die Übereinstimmung der Rater bei der Zuordnung der Items des CRI:TIC zu den drei lower-level skills unterstreicht, dass hinter den Fähigkeiten unterschiedliche Anforderungen stecken, die sich auch in den Aufgabenanforderungen widerspiegeln. Daher ist aus chemiedidaktischer Perspektive eine getrennte Betrachtung und Thematisierung der einzelnen Fähigkeiten anhand verschiedener mathematisch-symbolischer und visuell-graphischer Repräsentationen sowie die Reflexion der Bedeutung dieser Repräsentationen sinnvoll. Die Ergebnisse dieser Studie machen jedoch deutlich, dass die Förderung einer einzelnen Fähigkeit allein nicht sinnvoll und ausreichend ist, sondern dass die Fähigkeiten gleichermaßen Beachtung finden müssen, was sich auch aus anderen Studien in diesem Kontext ableiten lässt (Kozma & Russell, 1997, 2005; Rau, 2018).

FF1.2 Inwieweit lassen sich Repräsentationskompetenzen und Fachwissen empirisch trennen?

Die neben dem CRI:TIC eingesetzten, inhaltlich dazu abgestimmten Items zur Messung des deklarativen Fachwissens wurden einer äquivalenten Analyse unterzogen wie das CRI:TIC. Insgesamt konnten zwei Items identifiziert werden, die aufgrund einer zu hohen Schwierigkeit ausgeschlossen wurden (Tabelle 9). Der bereinigte Itempool von 16 Items zeigt insgesamt eine zufriedenstellende Modell- und Itemfit Statistik. Zudem liegt nach gründlicher Prüfung hohe Evidenz für die Annahme der lokalen stochastischen Unabhängigkeit für die Fachwissensitems vor. Zwar zeigt das Item *Fach_18* einen leichten *Underfit* in Bezug auf den *outfit* *MNSQ*, dieser kann jedoch aufgrund des passenden

outfit ZSTD als zufällig gewertet werden, weshalb das Item beibehalten wird (Boone et al., 2014).

Um die Forschungsfrage FF1.2 zu beantworten, wurden basierend auf den Ergebnissen von FF1.1 zwei Modelle gegenübergestellt: Ein eindimensionales Modell, in dem die 33 Items des CRI:TIC und die 16 Fachwissensitems als eine latente Variable operationalisiert werden und ein zweidimensionales Modell, in dem die Repräsentationskompetenzen und das Fachwissen als separate latente Variablen operationalisiert werden. Das zweidimensionale Modell basiert auf den Befunden, dass die Repräsentationskompetenzen in dieser Untersuchung empirisch nicht trennbar sind. Die zwei Modelle zeigen zufriedenstellende Modell- und Itemfit Statistiken (Tabelle 10), und die lokale stochastische Unabhängigkeit kann für beide Modelle nach sorgfältiger Prüfung angenommen werden. Dadurch wird auch hier der Modellvergleich legitimiert, um die Trennbarkeit der beiden Konstrukte zu prüfen.

Im Gegensatz zu den Befunden bei der Klärung der Struktur der Repräsentationskompetenzen zeichnet sich hier ein einheitliches und eindeutiges Bild ab, das das zweidimensionale Modell favorisiert. Sowohl die Informationskriterien (*AICc*, *saBIC*), der Likelihood Ratio Test und die Cross Plots zeigen, dass das zweidimensionale Modell im Vergleich zum eindimensionalen die bessere Passung hat und liefert somit Evidenz, dass Repräsentationskompetenzen und deklaratives Fachwissen in dieser Studie empirisch trennbar sind. Dennoch lässt sich ein deutlicher Zusammenhang zwischen dem Fachwissen und den Repräsentationskompetenzen erkennen ($r = 0.51$). Damit kann die Hypothese H1.2 angenommen werden.

Mit Blick auf die Güte der Datenerhebung und Messung lässt sich auch hier sagen, dass diese eine hohe Objektivität aufweisen. Besonders ist die Validität hervorzuheben, die bei anderen Messinstrumenten teilweise als fragwürdig einzuordnen ist. Durch die kritische Evaluation dieser Messinstrumente und die theoriegeleitete Ableitung von Kriterien zur Operationalisierung von Repräsentationskompetenzen und Fachwissen kann die Konstruktvalidität als hoch bewertet werden. Die Reliabilität lässt sich für die Repräsentationskompetenzen, wie oben bereits beschrieben, als gut bewerten, in Bezug auf das Fachwissen jedoch nur als akzeptabel (Tabelle 10). Die Itemanzahl kann dafür nicht als Erklärungsansatz herangezogen werden, ebenso wenig die lokale stochastische Unabhängigkeit, da diese ausreichend geprüft wurde. Die Items

zur Messung des Fachwissens stammen aus unterschiedlichen Instrumenten zur Messung des Fachwissens in der Allgemeinen Chemie, Analytischen Chemie, Anorganischen Chemie, Physikalischen Chemie und Organischen Chemie (Averbeck, 2021). Daher liegt hier wahrscheinlich kein eindimensionales, sondern ein multidimensionales Konstrukt vor, wodurch die Messung der Reliabilität verzerrt wird. Da die sonstigen Modell- und Itemfit Statistiken der Fachwissensitems in einem guten Bereich liegen, kann die noch akzeptable Reliabilität der Skala als weniger kritisch angesehen werden.

Nitz, Ainsworth und Koleg:innen (2014) sowie Edelsbrunner et al. (2023) haben durch faktoranalytische Untersuchungen vergleichbare Ergebnisse zur empirischen Trennbarkeit der beiden latenten Konstrukte erzielt. Dadurch liegt eine hohe Evidenz zur Annahme der Ergebnisse vor, Repräsentationskompetenzen und deklaratives Fachwissen als distinkte Konstrukte zu behandeln. Die Ergebnisse dieser Untersuchungen sind besonders mit Blick auf Lehr-Lern-Prozesse von hoher Relevanz. Sie machen deutlich, dass Repräsentationskompetenzen neben reinem Fachwissen einen zentralen Aspekt im Lehr-Lern-Prozess einnehmen sollten (Kozma & Russell, 1997, 2005; Rau, 2018; Sim & Daniel, 2014). Aktuelle Befunde zum Umgang mit Repräsentationen und Repräsentationskompetenzen zeigen, dass diese zwar thematisiert, aber nicht adäquat umgesetzt und verknüpft werden. Sie spielen lediglich eine zweitrangige, dem Fachwissen untergeordnete Rolle im Lehr-Lern-Prozess (Aydin et al., 2014; Gurung et al., 2022; Nyachwaya & Gillaspie, 2016; Popova & Jones, 2021; Rau et al., 2021; Sim & Daniel, 2014; Wu & Rau, 2018). Die vorliegenden Ergebnisse verdeutlichen, dass hier ein Umdenken notwendig ist und Repräsentationskompetenzen prominenter thematisiert werden müssen, da diese in den Naturwissenschaften und besonders in der Chemie nötig sind, um fachliche Konzepte mithilfe von Repräsentationen zu entwickeln (Edelsbrunner et al., 2023; Nitz, Ainsworth, et al., 2014).

Unter diesem Aspekt muss auch das Repräsentationsdilemma noch einmal reflektiert werden. Der zentrale Aspekt darin sagt aus, dass Repräsentationskompetenzen nötig sind, um fachliche Konzepte von Repräsentationen zu erlernen und gleichzeitig Fachwissen notwendig ist, um mit Repräsentationen diese Repräsentationskompetenzen zu entwickeln (Rau, 2017c, 2017a; Rau et al., 2021). Dieser Zusammenhang kann in der vorliegenden Studie durch die hohe Korrelation zwischen den Repräsentationskompetenzen und dem Fach-

wissen gezeigt werden – die beiden Konstrukte stehen in engem Zusammenhang. Es muss jedoch betont werden, dass im Rahmen dieser Studie eine Momentaufnahme der Repräsentationskompetenzen und des Fachwissens erhoben wurde, wodurch keine Aussagen zu kausalen Zusammenhängen und wechselseitigen Beeinflussungen getroffen werden können. Hier kann lediglich ein hoher positiver Zusammenhang der beiden Konstrukte berichtet werden. Die Hypothese von Edelsbrunner et al. (2023) und Nitz, Ainsworth und Kolleg:innen (2014), dass Repräsentationskompetenzen der Schlüssel zum Lernen von fachlichen Konzepten darstellen, kann hier nicht beantwortet und auch nicht unterstützt werden. Die Hypothese wurde auf Basis von Mustern in den Daten abgeleitet, dass Personen hohe Repräsentationskompetenzen bei niedrigem Fachwissen aufweisen können, aber nicht vice versa. Dieses Muster kann in den Daten dieser Studie nicht erkannt und bestätigt werden.

Zusammenfassend lässt sich sagen, dass mit dem CRI:TIC ein objektives, valides und reliables Instrument zur Messung der Repräsentationskompetenzen bereitgestellt wird. Basierend auf den Daten des CRI:TIC wurde die Struktur der lower-level skills der Repräsentationskompetenzen untersucht, mit dem Ergebnis, dass diese nicht separierbare, stark zusammenhängende Fähigkeiten innerhalb der Repräsentationskompetenzen darstellen (FF1.1), was mit den Befunden vergleichbarer Studien konform ist (Taskin et al., 2015; Ward et al., 2025). In Verbindung mit einer Skala zum Fachwissen zeigt sich, dass sich die Repräsentationskompetenzen jedoch davon empirisch trennen lassen, obwohl sie in engem Zusammenhang stehen (FF1.2). Auch diese Ergebnisse können durch andere Studien erhärtet werden (Edelsbrunner et al., 2023; Nitz, Ainsworth, et al., 2014).

Zur Erhöhung der Aussagekraft und der Belastbarkeit der Ergebnisse werden die Analysen mit einer neuen und größeren Stichprobe repliziert. Dies geschieht im Rahmen von Forschungsanliegen II. Die Daten werden im Anschluss gemeinsam mit Daten zu räumlichen Faktoren analysiert, um den Zusammenhang dieser Konstrukte näher zu untersuchen.

6

Forschungsanliegen II – Einfluss der räumlichen Fähigkeiten beim Bearbeiten repräsentationsbasierter Chemieaufgaben

In diesem Kapitel wird der Zusammenhang zwischen räumlichen Fähigkeiten und Repräsentationskompetenzen sowie dem Fachwissen in der Chemie untersucht und näher beleuchtet. Dazu wird das Design zunächst beschrieben und das methodische Vorgehen mit den eingesetzten Messinstrumenten dargestellt. Anschließend werden die Ergebnisse des zweiten Forschungsanliegens präsentiert und kritisch diskutiert.

6.1 Design

Für Forschungsanliegen II wurde ebenfalls eine *quantitative Querschnittsstudie mit nicht-experimentellem Design* gewählt (Döring, 2023f), um die Forschungsfragen zu untersuchen. Zwar ist das Vorgehen theoriegeleitet, jedoch konnten nur teilweise Hypothesen abgeleitet werden, weshalb sowohl explanative als auch explorative Vorgehen gewählt wurden. Die Daten wurden erneut mit Studieneinsteiger:innen in Chemie-Vorkursen erhoben.

Um die Forschungsfragen zu adressieren, wurden das evaluierte und überarbeitete CRI:TIC aus dem ersten Forschungsanliegen zur Messung der Repräsentationskompetenzen und die Fachwissensitems zur Erfassung des deklarativen Fachwissens eingesetzt. Zur Messung der räumlichen Faktoren wurde das Cattell-Horn-Carroll Modell als Grundlage verwendet und die bereits aus Untersuchungen hervorgegangenen relevanten räumlichen Faktoren *Visualization* (Harris et al., 2013; Wai et al., 2009), *Spatial Relation* (Bodner & Guay, 1997; Carter et al., 1987; Harle & Towns, 2011; Harris et al., 2013; Sorby, 2009), *Speeded Rotation* (Harris et al., 2013; Wu & Shah, 2004) und *Flexibility of Closure* (Bodner & McMillen, 1986; Carter et al., 1987; Hodgkiss et al., 2018; Wu & Shah, 2004) in die Studie einbezogen (Kapitel 3.3). Zusätzlich wurden

vier weitere räumliche Faktoren aus dem Cattell-Horn-Carroll Modell erhoben, deren Definitionen im Kontext der Naturwissenschaften, vornehmlich der Chemie, eine Rolle beim Bearbeiten repräsentationsbasierter Chemieaufgaben eine Rolle spielen könnten: *Visual Memory*, *Closure Speed*, *Perceptual Speed* und *Spatial Scanning* (Tabelle 2). Insgesamt wurden in dieser Studie somit acht unterschiedliche räumliche Faktoren erfasst und im Zusammenhang mit dem Fachwissen und den Repräsentationskompetenzen untersucht. Zur Messung der räumlichen Faktoren wurden validierte, psychometrische Messinstrumente eingesetzt. Diese wurden unter Berücksichtigung der Eignung und aus testökonomischer Perspektive verglichen (Kapitel 6.2.2).

In drei unterschiedlichen Chemie-Vorkursen wurden das CRI:TIC, die Fachwissensitems und die acht Messinstrumente zur Erfassung der entsprechenden räumlichen Faktoren eingesetzt. Für die Auswertung wurden in einem ersten Schritt die IRT-basierten Analysen zur Evaluation der Modell- und Itemfit Statistiken des CRI:TIC und der Fachwissensitems wiederholt. Daraus wurden die Personenfähigkeiten der Repräsentationskompetenzen und des deklarativen Fachwissens für die weiterführenden Analysen extrahiert. Im Gegensatz dazu wurden die psychometrischen Instrumente zur Messung der räumlichen Faktoren anhand der KTT ausgewertet und Summenscores gebildet, wie in den entsprechenden Manualen beschrieben.

Zur Beantwortung der Forschungsfrage 2.1 wurden bivariate Korrelationsanalysen mit Bonferroni-Korrektur zwischen den Personenfähigkeiten der Repräsentationskompetenzen, dem deklarativen Fachwissen und den Summenscores für die räumlichen Faktoren gerechnet. Dadurch sollte zunächst geprüft werden, inwieweit ein Zusammenhang zwischen den Konstrukten festgestellt werden kann. Anschließend wurden zur Beantwortung von Forschungsfrage 2.1 Pfadanalysen durchgeführt, um zu überprüfen, inwieweit das deklarative Fachwissen und die räumlichen Faktoren signifikante Prädiktoren für die Bearbeitung der repräsentationsbasierten Items des CRI:TIC sind und welche Variablen einen maßgeblichen und einzigartigen Beitrag zur Erklärung der Varianz leisten. Durch dieses Vorgehen konnten die Interkorrelationen zwischen den räumlichen Faktoren berücksichtigt werden, was in anderen Studien meist nicht berücksichtigt wurde.

6.2 Messinstrumente

Insgesamt wurden im Rahmen dieses Forschungsanliegens zehn unterschiedliche Variablen erhoben: Die Repräsentationskompetenzen, das deklarative Fachwissen und die acht unterschiedlichen räumlichen Faktoren. Im Folgenden werden die entsprechenden Messinstrumente beschrieben und erläutert.

6.2.1 Messung der Repräsentationskompetenzen und des Fachwissens

Die Messung der Repräsentationskompetenzen wurde mithilfe des im Rahmen von Forschungsanliegen I entwickelten und evaluierten CRI:TIC vorgenommen. Das Instrument besteht aus insgesamt 33 Items im geschlossenen und halboffenen Format und deckt verschiedene Fachbereiche aus der Chemie und eine Vielzahl an Repräsentationen ab (Kapitel 5.2.2). Die Auswertung wurde IRT-basiert mit einem Partial-Credit-Modell vorgenommen (Kapitel 5.3.2), wodurch Personenfähigkeiten und Itemschwierigkeiten differenzskaliert auf derselben linearen Skala abgebildet wurden. Basierend auf den Ergebnissen von Forschungsfrage 1.1 wurde ein eindimensionales Modell für die Analysen verwendet, da sich die lower-level skills *Interpretation*, *Translation* und *Konstruktion* im ersten Forschungsanliegen empirisch nicht eindeutig trennen ließen. Die Personenfähigkeiten wurden hier für die weiteren Analysen verwendet. Ebenso wurden die 16 Fachwissensitems aus dem ersten Forschungsanliegen zur Messung des deklarativen Fachwissens verwendet (Kapitel 5.2.3) und mithilfe der IRT ausgewertet. Die Personenfähigkeiten wurden auch hier für die weiteren Analysen extrahiert.

Wie im ersten Forschungsanliegen wurde für die Messung der beiden Konstrukte ein Testheft zusammengestellt. Auf dem Deckblatt mussten die Testpersonen zunächst einen individuellen Code (Pseudonym) generieren, sodass jeder einzelnen Person die entsprechenden Messinstrumente zugeordnet werden konnten, ohne die Anonymität zu beeinträchtigen. Anschließend folgten zwei Seiten mit Informationen zur Datenerhebung, sowie zum Umgang mit den erhobenen Daten. Zusätzlich wurden einige persönliche Daten abgefragt (Geschlecht, Alter, geplanter Studiengang, Standort). Zum Schluss folgten die Items zur Messung der Repräsentationskompetenzen und des Fachwissens in einer durchmischten und randomisierten Reihenfolge.

6.2.2 Messung der räumlichen Fähigkeiten

Basierend auf der bisherigen Forschung wurden vier räumliche Faktoren, die einen hohen Zusammenhang mit der Leistung und beim Bearbeiten von Repräsentationen in der Chemie zeigen, für das zweite Forschungsprojekt ausgewählt: *Visualization*, *Speeded Rotation*, *Spatial Relation* und *Flexibility of Closure* (Kapitel 3.3). Zusätzlich wurden die Definitionen der restlichen räumlichen Faktoren im Cattell-Horn-Carroll Modell in Bezug auf die Relevanz für die Naturwissenschaften evaluiert. Da zu diesen nach bestem Gewissen keine Forschung im Kontext der Naturwissenschaften vorliegt, wurden diese nach Augenmaß mit Passung zur Chemie ausgewählt. Insgesamt wurden vier weitere räumliche Faktoren in die Untersuchung aufgenommen: *Visual Memory*, *Closure Speed*, *Perceptual Speed* und *Spatial Scanning* (Tabelle 2). Aus Gründen der Testökonomie, der Zumutbarkeit und der Möglichkeiten zur Datenerhebung im Rahmen der Vorkurse, konnten nicht alle Faktoren aus dem Cattell-Horn-Carroll Modell in die Untersuchung einbezogen werden. Daher wurden die Faktoren *Imagery*, *Length Estimation*, *Perceptual Alternations*, *Perceptual Illusions* und *Serial Perceptual Integration* (Tabelle 2) im Rahmen dieser Studie nicht berücksichtigt.

Da die Struktur des Cattell-Horn-Carroll Modells auf faktoranalytischen Untersuchungen mit standardisierten, validierten psychometrischen Messinstrumenten beruht, konnte auf diese im Rahmen des Projekts zurückgegriffen werden. Dies hat den Vorteil, dass die Ergebnisse des Forschungsanliegens gut mit anderen Studien vergleichbar sind und vor diesem Hintergrund diskutiert werden können. Sieben der acht Messinstrumente wurden vom Educational Testing Service (ETS) in Form des *Kit of Factor-Referenced Cognitive Tests* (Ekstrom et al., 1976a) und des entsprechenden Manuals zur Auswertung der Messinstrumente, dem *Manual for Kit of Factor-Referenced Cognitive Tests* (Ekstrom et al., 1976b), zur Verfügung gestellt. Bei den Dokumenten handelt es sich um gescannte Originaldateien in englischer Sprache. Die Messinstrumente wurden ins Deutsche übersetzt und die Repräsentationen durch Anpassung des Kontrasts, der Schärfe und der Ausrichtung übernommen. Die Anpassung der Repräsentationen war nötig, da Scan-Artefakte wie schwarze Flecken, Unschärfen, Kontrastfehler und falsch ausgerichtete Seiten im Originaldokument vorliegen. Die Repräsentationen an sich wurden nicht verändert.

Die sieben Messinstrumente des *Kit of Factor-Referenced Cognitive Tests* (Ekstrom et al., 1976a) sind vom Aufbau und der Durchführung identisch gestaltet: Zunächst gibt es eine Einleitung in das jeweilige Messinstrument, die die Durchführung mit Aufgabenbeispielen darstellt, die Auswertung und Bewertung transparent macht und Hinweise zum zeitlichen Umfang und der Anzahl der Seiten pro Testteil gibt. Anschließend folgen zwei Testteile, die in ihrer Bearbeitung identisch sind, sich aber in den zu bearbeitenden Items unterscheiden. Da es sich hier ausschließlich um Speed-Tests handelt, gibt es für jedes Messinstrument und jeden Testteil konkrete Zeitvorgaben. Die Instrumente sind teilweise so konzipiert, dass die Testpersonen die Testteile in der vorgegebenen Zeit nicht vollständig bearbeiten können, z. B. der *Hidden Patterns Test*.

Das achte Messinstrument, der *Revised Purdue Spatial Visualization Test: Visualization of Rotation (Revised PSVT:R)* von Yoon (2011b), wird in deutscher Sprache mit entsprechendem Manual zur Auswertung zur Verfügung gestellt. Das Instrument umfasst eine Einleitung, die die Durchführung mit Aufgabenbeispielen beschreibt. Darauf folgt ein einzelner Testteil, der ohne zeitliche Limitierung durchgeführt werden sollte. Aufgrund der Organisation und Bedingungen in den Vorkursen, in denen die Datenerhebung durchgeführt wurde, konnte dies nicht gewährleistet werden. Die Bearbeitungszeit musste auf 35 Minuten begrenzt werden.

Alle Messinstrumente wurden durch ein Deckblatt und zwei Seiten mit allgemeinen Informationen zur Studie und den Messinstrumenten sowie einer Seite zur Erhebung der persönlichen Angaben ergänzt. Auf dem Deckblatt musste ein individueller Code zur Pseudonymisierung und eindeutigen Zuordnung der einzelnen Messinstrumente zu einer Testperson generiert werden.

Die Datenerhebung mit den acht Messinstrumenten ergänzte eine Power-Point-Präsentation. Diese erläuterte das Ziel des Projekts, allgemeine Hinweise zur Datenerhebung und zum Einsatz des jeweiligen Messinstruments. Nach dem Durchlesen der Einleitung zu dem jeweiligen Messinstrument wurden offene Fragen geklärt und anschließend die Testteile durchgeführt. Dafür wurden zeitlich programmierte Start- und Stopp-Symbole genutzt, um die Speed-Tests entsprechend den Vorgaben durchzuführen. Mündliche Angaben der Testleitung unterstützten die Visualisierungen.

Tabelle 12: Übersicht der eingesetzten Messinstrumente zur Erhebung der acht fokussierten räumlichen Faktoren mit Zeitangaben.

Messinstrument	Kürzel	Räumlicher Faktor	Zeit ¹
Building Memory	BM	Visual Memory <i>Die Fähigkeit, eine mentale Repräsentation eines visuellen externen Stimulus zu erstellen, diese abzuspeichern und sie zu einem späteren Zeitpunkt wieder abzurufen.</i>	20 min
Card Rotations Test	CRT	Speeded Rotation <i>Die Fähigkeit, Probleme schnell zu lösen, indem einfache Bilder mental (2-dimensional) rotiert werden.</i>	10 min
Gestalt Completion Test	GCT	Closure Speed <i>Die Fähigkeit, unvollständige oder teilweise verdeckte visuelle Stimuli oder Muster schnell mental zu vervollständigen, ohne im Voraus zu wissen, worum es sich dabei handelt.</i>	8 min
Hidden Patterns Test	HPT	Flexibility of Closure <i>Die Fähigkeit, eine im Voraus bekannte Figur in einem komplexen, ablenkenden Muster zu erkennen und zu identifizieren.</i>	10 min
Identical Pictures Test	IPT	Perceptual Speed <i>Die Fähigkeit, räumlich nebeneinander oder separiert angeordnete visuelle Symbole oder Muster schnell miteinander zu vergleichen.</i>	7 min
Maze Tracing Speed Test	MTSP	Spatial Scanning <i>Die Fähigkeit, ein visuelles Feld oder Muster schnell zu überblicken und einen Weg dadurch zu finden.</i>	10 min
Paper Folding Test	PFT	Visualization <i>Die Fähigkeit, komplexe Muster wahrzunehmen und mental zu simulieren, wie sie aussehen könnten, wenn sie verändert werden.</i>	10 min
Revised Purdue Spatial Visualization Test: Visualization of Rotations	Revised PSVT:R	Spatial Relation <i>Die Fähigkeit, Probleme zeitunabhängig durch die (3-dimensionale) Rotation komplexer Bilder zu lösen.</i>	35 min

¹ Die Zeiten umfassen die gesamte Durchführung, d. h. mit mündlicher Einführung und Lesen der Einleitung.

Im Folgenden werden die eingesetzten Messinstrumente detailliert beschrieben, Alternativen kurz dargestellt und die Wahl der eingesetzten Instrumente begründet. Tabelle 12 gibt einen Überblick über die Messinstrumente und die entsprechenden räumlichen Faktoren. Insgesamt dauert die Bearbeitung aller Messinstrumente etwa 110 Minuten. Die Auswertung der acht Instrumente basiert auf der KTT (Kapitel 5.3.1).

6.2.2.1 Building Memory

Beschreibung und Durchführung – Der *Building Memory* Test misst die Fähigkeit, sich (komplexe) Bilder über einen kurzen Zeitraum einzuprägen (*Visual Memory*). In den zwei Testteilen wird dafür zunächst eine Stadtkarte mit zwölf Gebäuden präsentiert, deren Lokationen sich die Testpersonen einprägen müssen. Die Einprägungsphase dauert vier Minuten. Anschließend sind die Lokationen auf der leeren Stadtkarte anzugeben. Dazu sind auf der Karte dieser die Buchstaben A bis D mehrfach abgebildet, und die zwölf Gebäude werden auf der linken Seite dargestellt. Für jedes Gebäude muss ein Buchstabe eindeutig zugeordnet und entsprechend angekreuzt werden. Die Bearbeitungszeit für diese Aufgabe beträgt ebenfalls vier Minuten (Ekstrom et al., 1976a, 1976b).

Auswertung – Die Punktzahl in diesem Test ergibt sich aus der Anzahl der korrekt angegebenen Lokationen abzüglich der falsch angegebenen. Nicht beantwortete Items werden mit null Punkten bewertet. In jedem Testteil können maximal zwölf Punkte erreicht werden, somit für beide Testteile zusammen 24 Punkte (Ekstrom et al., 1976a, 1976b).

Alternative Messinstrumente – Zur Messung der Fähigkeit *Visual Memory* bieten sich ebenso der *Shape Memory Test* sowie das *Map Memory* als Messinstrumente an. Diese sind in der Durchführung und vom zeitlichen Umfang vergleichbar. Für die drei Instrumente liegen Ergebnisse aus einer Studie mit $542 \leq N \leq 574$ Männern vor. Diese zeigen, dass die Reliabilität für den *Building Memory* Test am höchsten ist, weshalb er für dieses Projekt gewählt wurde (Ekstrom et al., 1976a, 1976b).

6.2.2.2 Card Rotations Test

Beschreibung und Durchführung – Mithilfe des *Card Rotations Test* wird die Fähigkeit gemessen, Probleme schnell durch die zweidimensionale mentale Rotation einfacherer Bilder zu lösen (*Speeded Rotation*). In jedem Testteil werden den Testpersonen insgesamt zehn Figuren auf der linken Seite dargestellt. Rechts von jeder Figur wird die gleiche Figur acht Mal rotiert und ggf. gespiegelt dargestellt. Für jede der acht Figuren muss angegeben werden, ob diese gleich (G) ist, d. h. durch einfache Rotation in Übereinstimmung gebracht werden kann, oder ob diese anders (A) ist, d. h. nicht übereinstimmt. Jeder Testteil umfasst insgesamt 80 Items, wofür drei Minuten zur Verfügung stehen (Ekstrom et al., 1976a, 1976b).

Auswertung – Die Punktzahl ergibt sich aus der Differenz korrekt und falsch beantworteter Items. Nicht beantwortete Items werden mit null Punkten bewertet. Pro Testteil können maximal 80 und im gesamten Test maximal 160 Punkte erreicht werden (Ekstrom et al., 1976a, 1976b).

Alternative Messinstrumente - Für den *Card Rotations Test* stand kein vergleichbares bzw. alternatives Messinstrument zur Verfügung.

6.2.2.3 Gestalt Completion Test

Beschreibung und Durchführung – In diesem Messinstrument werden den Testpersonen unvollständige, lückenhafte Bilder präsentiert, die mental vervollständigt werden müssen. Aufgabe ist es, das dargestellte Objekt mit wenigen Worten zu beschreiben. Dadurch kann die Fähigkeit *Closure Speed* gemessen werden. In jedem Testteil werden zehn unvollständige Bilder dargestellt, wofür jeweils zwei Minuten zur Verfügung stehen (Ekstrom et al., 1976a, 1976b).

Auswertung – Der Testscore ergibt sich aus der Summe der korrekt beantworteten Items. Nicht und falsch beantwortete Items werden mit null Punkten bewertet. Pro Testteil können maximal zehn Punkte und im gesamten Test maximal 20 Punkte erreicht werden (Ekstrom et al., 1976a, 1976b).

Alternative Messinstrumente – Alternativ zum *Gestalt Completion Test* können der *Concealed Words Test* oder die *Snowy Pictures* als Messinstrumente eingesetzt werden. Ersterer könnte jedoch aufgrund der Verwendung von

Buchstaben und Wörtern für Menschen mit Dyslexie aus verschiedenen Gründen eine zusätzliche, ungewollte Herausforderung darstellen (Moojen et al., 2020), weshalb Bilder vorzuziehen sind, um diese Störvariable zu vermeiden. Die *Snowy Pictures* (Ekstrom et al., 1976b, 1976a) konnten aufgrund der mangelhaften Bildqualität nicht verwendet werden.

6.2.2.4 Hidden Patterns Test

Beschreibung und Durchführung – Um die Fähigkeit zu messen, eine bekannte Figur in einem komplexen, ablenkenden Muster zu erkennen (*Flexibility of Closure*), wird der *Hidden Patterns Test* eingesetzt. Die Testpersonen müssen hier eine einfache vorgegebene Figur aus horizontalen, vertikalen und diagonalen Linien in einem Linienmuster identifizieren. Dabei darf die Figur nicht rotiert oder gespiegelt werden. Jeder Testteil besteht aus 200 Mustern, die in drei Minuten auf die bekannte Figur überprüft werden müssen. Wenn die Figur in dem Muster vorhanden ist, muss ein „X“ eingetragen werden, ist sie nicht vorhanden, wird eine „0“ eingetragen. In beiden Testteilen ist die vorgegebene Figur identisch (Ekstrom et al., 1976a, 1976b).

Auswertung – Jedes richtig gesetzte „X“ und jede richtig gesetzte „0“ ergibt einen Punkt, falsch gesetzte „X“ und „0“ ergeben jeweils einen Minuspunkt. Nicht beantwortete Items werden mit null Punkten bewertet. Der Testscore ergibt sich aus der Summe aller Werte. Maximal können pro Testteil 200 Punkte und insgesamt in dem Test 400 Punkte erreicht werden (Ekstrom et al., 1976a, 1976b).

Alternative Messinstrumente – Neben dem *Hidden Patterns Test* bieten sich auch der *Hidden Figures Test* und der *Copying Test* an. Der *Hidden Figures Test* ist mit dem *Hidden Patterns Test* gut vergleichbar, dieser gibt jedoch fünf Figuren vor, von denen eine in komplexen Mustern erkannt werden soll. Insgesamt dauert der Test 24 Minuten, weshalb dieser aus testökonomischen Gründen und in Bezug auf die Zumutbarkeit für die Testpersonen nicht verwendet wird. Der *Copying Test* gibt Figuren aus vier verbundenen Linien vor, die in einem quadratischen Punkteraster (5x5) nachskizziert werden müssen. Da der *Hidden Patterns Test* der Definition von *Flexibility of Closure* formal eher entspricht, wird dieser für das Projekt gewählt (Ekstrom et al., 1976a, 1976b).

6.2.2.5 Identical Pictures Test

Beschreibung und Durchführung – Mit dem *Identical Pictures Test* wird die Fähigkeit erhoben, Figuren schnell visuell zu erfassen und miteinander zu vergleichen (*Perceptual Speed*). In einer linken Spalte wird dafür eine Figur präsentiert, die mit fünf rechts davon stehenden Figuren verglichen werden muss, wobei eine davon identisch ist. Pro Testteil gibt es 48 Items, die in 1.5 Minuten bearbeitet werden müssen (Ekstrom et al., 1976a, 1976b).

Auswertung – Der Testscore ergibt sich aus der Summe der korrekt beantworteten Items abzüglich der falsch beantworteten. Nicht beantwortete Items werden mit null Punkten bewertet. Pro Testteil können somit maximal 48 Punkte erreicht werden, im gesamten Test 96 Punkte (Ekstrom et al., 1976a, 1976b).

Alternative Messinstrumente – Der *Finding A's Test* und der *Number Comparison Test* stellen Alternativen zu dem *Identical Pictures Test* dar. Bei dem *Finding A's Test* müssen Testpersonen pro Testteil in 20 Spalten mit jeweils 41 Wörtern jene fünf Wörter markieren, die den Buchstaben „a“ enthalten (Ekstrom et al., 1976a, 1976b). Dieser Test wird ausgeschlossen, da er auf Englisch verfasst ist und eine Übersetzung auf Deutsch die Testdurchführung und -bedingungen verändert. Zudem könnten hier Testpersonen mit Dyslexie benachteiligt werden, wenn diese beispielsweise eine verminderte Lesegeschwindigkeit haben (Moojen et al., 2020), was eine Störvariable dargestellt hätte. Ähnlich verhält es sich bei dem *Number Comparison Test*, bei dem zwei Zahlen auf Übereinstimmung geprüft werden müssen. Menschen mit Dyskalkulie könnten hier Schwierigkeiten haben und die Aufgabe langsamer bearbeiten (Memon et al., 2021), was sich in diesem Fall nicht auf die Fähigkeit *Perceptual Speed* zurückführen lässt (Ekstrom et al., 1976a, 1976b).

6.2.2.6 Maze Tracing Speed Test

Beschreibung und Durchführung – Durch den *Maze Tracing Speed Test* wird die Fähigkeit gemessen, ein komplexes Feld mit vielen Hindernissen schnell visuell zu erfassen und darin einen Ausweg zu erkennen (*Spatial Scanning*). Dafür nutzt das Instrument pro Testteil 24 aneinandergereihte quadratische Labyrinth, in denen jeweils der einzige durchgängige Weg vom Anfang zum

Ende markiert werden muss. Für jedes Testteil mit 24 Labyrinth stehen drei Minuten zur Verfügung (Ekstrom et al., 1976a, 1976b).

Auswertung – Bei der Auswertung entscheidet man für jedes der 24 Labyrinth, ob der Weg richtig oder falsch ist. „Richtig“ heißt in diesem Fall, dass mindestens ein markierter Weg das Labyrinth durchquert, ohne eine durchgezogene Linie mit der Markierung zu durchbrechen. Führt ein Weg in eine Sackgasse und setzt neu an, gilt er trotzdem als korrekt, ebenso, wenn eine durchgezogene Linie von der Markierung lediglich berührt wird. Richtig durchquerte Labyrinth erhalten einen Punkt. Falsch durchquerte (wenn eine durchgezogene Linie durchbrochen wird), teilweise durchquerte oder nicht bearbeitete Labyrinth erhalten null Punkte. Insgesamt sind pro Testteil maximal 24 Punkte und im gesamten Test maximal 48 Punkte erreichbar (Ekstrom et al., 1976a, 1976b).

Alternative Messinstrumente – Ähnliche Testinstrumente sind *Choosing a Path* und der *Map Planning Test*. In *Choosing a Path* wird ein komplexes Netz aus Kabeln mit zahlreichen Knotenpunkten und einem Zentralpunkt dargestellt. Daran knüpfen fünf Buchsen mit zwei Anschlüssen S (*start*) und F (*finish*) an. Aufgabe ist es, die Buchse zu markieren, bei der ein geschlossener Stromkreis entsteht, der über den Zentralpunkt führt. Für den gesamten Test werden 14 Minuten (ohne Einleitung) benötigt, weshalb dieser Test hier aus testökonomischen Gründen nicht eingesetzt wird. Im *Map Planning Test* werden in jedem Testteil zwei Straßennetze aus 6×7 Quadraten dargestellt. Die Start- und Endpunkte der Straßen sind mit den Buchstaben „A“ bis „Z“ markiert. Innerhalb des Straßennetzes gibt es Blockaden, mit Kreisen markiert, und jeweils zehn Gebäude, die mit Quadraten und den Nummer 1-10 dargestellt sind. Aufgabe ist es, in jedem Straßennetz für zehn Strecken (z. B. von F nach I) den kürzesten Weg zu identifizieren und anzugeben, welches Gebäude auf dieser Strecke passiert wird. Zwar dauert jeder Testteil ebenfalls nur drei Minuten, dieser hat in einer aufgeführten Studie ($605 \leq N \leq 629$) jedoch eine schlechtere Reliabilität als der Maze Tracing Speed Test (*Maze Tracing Speed Test*: $Rel\varphi = .89$, $Rel\sigma = .91$; *Map Planning Test*: $Rel\varphi = .75$, $Rel\sigma = .80$), weshalb auch dieser nicht zum Einsatz kommt (Ekstrom et al., 1976a, 1976b).

6.2.2.7 Paper Folding Test

Beschreibung und Durchführung – Zur Messung der Fähigkeit, komplexe Muster wahrzunehmen und diese mental zu manipulieren (*Visualization*), wird der *Paper Folding Test* eingesetzt. In jedem Testteil werden dafür zehn quadratische Stücke Papier Schritt für Schritt auf unterschiedliche Art und Weise gefaltet. Anschließend wird das gefaltete Stück Papier an genau einer Stelle durch alle Lagen gestanzt. Aufgabe ist es, aus einer Auswahl von fünf Möglichkeiten auszuwählen, wie das Stück Papier aussieht, wenn es wieder vollständig entfaltet wird. Für jeden Testteil haben die Testpersonen drei Minuten Zeit (Ekstrom et al., 1976a, 1976b).

Auswertung – Der Testscore errechnet sich aus der Differenz der korrekt und falsch beantworteten Items. Nicht beantwortete Items werden jeweils mit null Punkten bewertet. In jedem Testteil können somit maximal zehn Punkte erreicht werden und im gesamten Test insgesamt 20 Punkte (Ekstrom et al., 1976a, 1976b).

Alternative Messinstrumente – Der *Form Board Test* bietet eine Alternative. Hier werden eine geometrische Figur und fünf Fragmente vorgegeben. Ein Teil dieser Fragmente ergibt zusammengelegt die geometrische Figur, wobei diese beliebig rotiert, aber nicht gespiegelt werden dürfen. Die Testperson muss die benötigten Fragmente mit einem „+“ und die nicht benötigten Fragmente mit einem „-“ markieren. Pro Testteil stehen acht Minuten zur Verfügung, weshalb der Test aus zeitökonomischen Gründen nicht eingesetzt wird. Aus dem gleichen Grund wird auch der *Surface Development Test* nicht eingesetzt. Hier wird eine Faltvorlage, bei der die Kanten mit Zahlen beschriftet sind, und der daraus gefaltete Körper, bei dem die Kanten mit Buchstaben beschriftet sind, dargestellt. Jeder Zahl muss ein Buchstabe zugeordnet werden. Ein Testteil dauert sechs Minuten (Ekstrom et al., 1976a, 1976b).

6.2.2.8 Revised Purdue Spatial Visualization Test: Visualization of Rotations (PSVT:R)

Beschreibung und Durchführung – Das ursprüngliche Testinstrument, der *Purdue Visualization of Rotations Test* (ROT), ist Bestandteil einer Testbatterie, dem *Purdue Spatial Visualization Test* (Bodner & Guay, 1997; Guay, 1977). Es wird eingesetzt, um die Fähigkeit der dreidimensionalen Rotation zu messen (*Spatial Relation*). Dafür wird eine geometrische, dreidimensionale Figur mit geraden, abgeschrägten und/oder abgerundeten Flächen auf eine bestimmte Art rotiert. Anschließend wird eine neue geometrische Figur dargestellt, für die die Testperson aus fünf vorgegeben Antworten wählen muss, wie die Figur aussieht, wenn sie auf die gleiche Art rotiert wird (Abbildung 9). In dem Test sind jedoch Fehler (zusätzliche oder fehlende Linien und Bestandteile) aufgefallen (Yue, 2008), weshalb dieser von Yoon (2011) korrigiert und neu validiert wurde: Der *Revised Purdue Spatial Visualization Test: Visualization of Rotations* (Revised PSVT:R). Dieser umfasst einen Testteil mit 30 Items. Die Bearbeitung sollte zeitlich nicht begrenzt werden (Yoon, 2011a), was aufgrund der Organisation und Bedingungen im Rahmen der Datenerhebung nicht gewährleistet werden konnte. Die Bearbeitungszeit musste auf 35 Minuten begrenzt werden.

Auswertung – Der Testscore ergibt sich aus der Summe der korrekt beantworteten Items. Falsch beantwortete und nicht beantwortete Items werden mit null Punkten bewertet. Maximal können in diesem Test 30 Punkte erlangt werden.

Alternative Messinstrumente – Der *Mental Rotation Test* (MRT) von Vandenberg und Kuse (1978) ist eine mögliche Alternative für die Messung der Fähigkeit *Spatial Relation*. Allerdings ist der (Revised) PSVT:R im Kontext von MINT-Studienfächern sehr populär und wurde daher für diese Studie als Messinstrument gewählt (Carlisle et al., 2015; Yoon, 2011a; Yue, 2008). Des Weiteren ist der Revised PSVT:R gegenüber dem MRT weniger anfällig für analytische Strategien (Kapitel 3.3.3), weshalb dieser die bessere Wahl ist (Carlisle et al., 2015).

6.3 Psychometrische und statistische Methoden

In einem ersten Schritt wurden die 33 Items des CRI:TIC und die 16 Fachwissensitems noch einmal hinsichtlich der Modellgüte und der Modell- und Itemfit Statistiken geprüft. Ebenso wurde die Struktur der Repräsentationskompetenzen (FF1.1) und die Trennbarkeit dieser vom Fachwissen erneut untersucht (FF1.2), um die Replizierbarkeit der Ergebnisse des ersten Forschungsanliegens zu prüfen. Dazu wurde wie in Kapitel 5.4 beschrieben verfahren. Darüber hinaus wurde eine Analyse der *Differential Item Functioning (DIF)* durchgeführt, um potenzielle Unterschiede in den Antwortmustern zwischen den verschiedenen Erhebungsorten systematisch zu untersuchen (Kapitel 6.4).

DIF-Analyse – Um zu untersuchen, ob Items in einem Test für unterschiedliche Gruppen unterschiedlich gut messen, d. h. um zu prüfen, inwieweit ein Item fair misst und bestimmte Subgruppen nicht benachteiligt oder bevorteilt, werden *Differential Item Functioning (DIF)* Analysen durchgeführt. Eines der bewährtesten und am häufigsten benutzten Verfahren ist die Mantel-Haenszel Statistik (Mantel & Haenszel, 1959). Hier wird für dichotome Items (Fachwissensitems) die *Mantel-Haenszel Methode (MH)* und für polytome Items (Items des CRI:TIC) die *Generalized Mantel-Haenszel Methode (GMH)* verwendet. Beide Vorgehen sind nicht-parametrisch und vergleichen zwei Subgruppen (Referenz- und Fokusgruppe) auf Basis der Anzahl richtiger und falscher Antworten innerhalb der zwei Subgruppen sowie der Stichprobengrößen der zwei Subgruppen. Daher eignen sich diese Verfahren auch im Rahmen von IRT-basierten Analysen. Der Vergleich wird mithilfe eines χ^2 -Test durchgeführt (Fidalgo & Madeira, 2008; French et al., 2019; Huang & Ishii, 2025; Penfield, 2001; Wang & Su, 2004). Um die Effektgröße einzuschätzen, wird der ΔMH -Wert betrachtet, auch als *ETS-Delta Scaling* bekannt. Für $|\Delta MH| < 1.0$ wird eine vernachlässigbare, für $|\Delta MH| < 1.5$ eine moderate und für $|\Delta MH| > 1.5$ eine große Effektgröße angenommen (Ozdemir & AlGhamdi, 2022).

Mit Blick auf das zweite Forschungsanliegen, die Untersuchung des Zusammenhangs zwischen räumlichen Fähigkeiten, Fachwissen und Repräsentationskompetenzen, werden zunächst bivariate Korrelationsanalysen zwischen den erhobenen Konstrukten durchgeführt. Dabei kann zwischen verschiedenen Verfahren unterschieden werden. Klassische Verfahren sind Pearsons Produkt-Moment-Korrelation und Spearmans Rangkorrelation (Döring, 2023a, 2023b; Witte, 2019b).

Produkt-Moment-Korrelationskoeffizient nach Pearson – Pearsons Korrelationskoeffizient r ergibt sich aus dem Quotienten der Kovarianz und dem Produkt der Standardabweichungen der zwei gemessenen Variablen. Diese Berechnung setzt intervallskalierte, normalverteilte und linear zusammenhängende Daten voraus (Witte, 2019b).

Rangkorrelation nach Spearman – Spearmans Rangkorrelationskoeffizient ρ (Rho) nutzt nicht die absoluten Werte der Daten, sondern bildet eine Rangfolge. Dieses Verfahren eignet sich für ordinal skalierte, nicht normalverteilte und nicht notwendigerweise linear zusammenhängende Daten (Witte, 2019b).

Da die Verteilung der Daten eine Rolle bei der Wahl des Verfahrens zur Bestimmung der Korrelation spielt, werden diese im Vorhinein mithilfe des Shapiro-Wilk Tests auf Normalverteilung geprüft. Dieser Test wird als robustester Normalverteilungstest gesehen (Razali & Wah, 2011).

Um einer α -Fehler-Kumulation durch umfangreiche Korrelationsanalysen entgegenzuwirken, wird eine Bonferroni-Korrektur angewendet, sodass $\alpha \leq 0.05$ um die Anzahl gerechneter Korrelation (hier 16) auf $\alpha_{\text{Bonferroni}} \leq 0.003125$ korrigiert wird (Abdi, 2010).

Für Forschungsfrage 2.2 werden Pfadanalysen aus der Familie der Strukturgleichungsmodelle verwendet, um die Prädiktorvariablen für die Repräsentationskompetenzen in dieser Studie zu identifizieren. Die Repräsentationskompetenz stellt die endogene (abhängige) Variable dar, auf die die räumlichen Faktoren (Tabelle 12) und das Fachwissen als exogene (unabhängige) Variablen wirken (Abbildung 10). Diese Wirkrichtung lässt sich aus der Theorie ableiten und begründen. Für die acht räumlichen Faktoren ist sie leicht nachzuvollziehen, da sie sich in zahlreichen Studien bereits als wichtiger Prädiktor für die Leistung in der Chemie und beim Bearbeiten repräsentationsbasierter Chemieaufgaben herausgestellt hat (Buckley et al., 2018; Stieff & Uttal, 2015; Uttal et al., 2013; Uttal & Cohen, 2012; Yang et al., 2020). Zusätzlich wird angenommen, dass die räumlichen Faktoren als Faktoren erster Ordnung der räumlichen Fähigkeiten (Abbildung 8) miteinander zusammenhängen und werden daher im Pfadmodell miteinander korreliert. Mit Blick auf das Fachwissen muss hier betont werden, dass es als Momentaufnahme des aktuellen deklarativen Fachwissens erhoben wird und somit zur Bearbeitung der reprä-

sentationsbasierten Chemieaufgaben bereits zur Verfügung steht. Somit beeinflusst es die Bearbeitung der Items mit mathematisch-symbolischen und visuell-graphischen Repräsentationen (z. B. Talanquer, 2022). Daher wird die Wirkrichtung hier vom Fachwissen auf die Repräsentationskompetenzen modelliert.

Im Gegensatz zu Strukturgleichungsmodellen, in denen latente Konstrukte durch manifeste Variablen operationalisiert werden, arbeitet ein Pfadmodell ausschließlich mit manifesten Variablen und verzichtet auf die explizite Modellierung latenter Konstrukte. Hier fließen für die räumlichen Faktoren die Summenscores und für die Repräsentationskompetenzen sowie das Fachwissen die Personenfähigkeiten aus dem Partial-Credit-Modell ein. Daher werden diese im Pfadmodell unterschiedlich dargestellt. Die Summenscores der räumlichen Faktoren sind Rohwerte, die ordinal und nicht metrisch skaliert sind. Daher sind diese Werte als manifeste, beobachtete Variablen zu definieren, die als Quader abgebildet werden. Die Personenfähigkeiten stellen durch das Partial-Credit-Modell geschätzte, metrisch skalierte latente Merkmale dar, die als Ovale im Pfadmodell abgebildet werden. Daher können Pfadanalysen auch als Kombination simultaner Regressionsanalysen aufgefasst werden (Gäde & Schermelleh-Engel, 2023; Streiner, 2005).

An dieser Stelle ist zu fragen: Warum wird eine Pfadanalyse und kein Strukturgleichungsmodell genutzt? Dies lässt sich mit den Anforderungen der beiden Verfahren begründen. Sie stellen in Bezug auf die Stichprobe hohe Anforderungen, um belastbare Ergebnisse zu erzielen. Pro geschätztem Modellparameter werden mindestens zehn Testpersonen empfohlen (Streiner, 2005), worin der ausschlaggebende Punkt liegt. Im Strukturgleichungsmodell werden mithilfe manifesten Variablen latente Variablen operationalisiert, wodurch u. a. Faktorladungen und Messfehler geschätzt werden müssen, was in der Pfadanalyse nicht der Fall ist, da hier Summenscores bzw. Personenfähigkeiten verwendet werden. Letztere werden nicht durch manifeste Variablen operationalisiert. Dadurch sind Strukturgleichungsmodelle im Vergleich zu Pfadmodellen komplexer und weisen eine höhere Anzahl an zu schätzenden Parametern auf, wodurch diese deutlich höhere Anforderungen an die Stichprobe stellen (Gäde & Schermelleh-Engel, 2023; Streiner, 2005). Diese Anforderungen können hier nicht bedient werden, weshalb die Untersuchungen auf

Pfadanalysen begrenzt werden müssen, wodurch eine Schätzung von Messfehlern beispielsweise nicht möglich ist und das Modell vereinfacht wird. Die Pfadanalyse wird mit *lavaan* (Rosseel et al., 2024) in R durchgeführt.

Um die Güte von Pfadmodellen (und Strukturgleichungsmodellen) zu beurteilen, hat sich eine Vielzahl an Kennwerten etabliert, die dafür herangezogen werden können.

χ^2 -Test – Klassischerweise wird die Kovarianzmatrix der empirisch ermittelten Daten mit der Kovarianzmatrix des geschätzten Modells verglichen, um die Passung der Daten zum Modell zu prüfen. Dieser Test ist sensitiv für große Stichproben, d. h. mit zunehmender Anzahl an Testpersonen wird es wahrscheinlicher, dass der χ^2 -Test einen statistisch signifikanten Unterschied angibt (Zheng & Bentler, 2025). Daher sollten für Stichproben $N > 400$ (Kenny, 2024) andere Parameter zur Bewertung der Modellgüte stärker gewichtet werden, die im Folgenden beschrieben werden.

TLI und CFI – Zur Bewertung der (relativen) Modellgüte kann das untersuchte Modell mit einem Unabhängigkeitsmodell (*Baseline Model*) verglichen werden, wofür der *Comparative-Fit-Index* (CFI) von Bentler (1990) und *Tucker-Lewis-Index* (TLI) von Tucker und Lewis (1973) etablierte Parameter sind. Das Unabhängigkeitsmodell geht davon aus, dass alle betrachteten Variablen im Modell unkorreliert sind, wodurch keine Kovarianzen der Variablen erklärt werden können, was in einer schlechten Passung zu den Daten resultiert. Beide Parameter beschreiben, wie viel besser das untersuchte Modell die empirischen Daten erklären kann und nimmt Werte zwischen 0 und 1 an (Gäde & Schermelleh-Engel, 2023). Werte von CFI und TLI ≥ 0.95 weisen auf eine gute Passung des untersuchten Modells hin (Hu & Bentler, 1999).

RMSEA – Die absolute Modellgüte kann mithilfe des *Root Mean Square Error of Approximation* (RMSEA) bewertet werden. Dieser schätzt die Passung zwischen dem Modell und den empirischen Daten bezogen auf die Stichprobengröße und die Freiheitsgrade. Werte von RMSEA ≤ 0.06 weisen auf eine gute Modellpassung hin (Hu & Bentler, 1999). Browne und Cudeck (1992) beschreiben, dass ein RMSEA ≤ 0.08 auf einen angemessenen Näherungsfehler hinweist und keine Werte von RMSEA > 0.10 verwendet werden sollten.

SRMR – Ebenso wie der RMSEA ist der *Standardized Root Mean Square Residual* (SRMR) ein absolutes Modellgütemaß, das die mittlere standardisierte

Differenz zwischen empirischer und modelltheoretischer Kovarianzmatrix misst (Gäde & Schermelleh-Engel, 2023). Wie bereits in Kapitel 5.4 beschrieben, weisen Werte $SRMR \leq 0.08$ auf einen guten Modellfit hin (Hu & Bentler, 1999), wobei auch Werte ≤ 0.10 noch als akzeptabel angesehen werden (Gäde et al., 2020; Kline, 2016; Lichti, 2019).

Die Parameterschätzung des Pfadmodells beruht auf der Maximum-Likelihood-Schätzung, die formal eine Normalverteilung der Daten voraussetzt. Daher werden die Daten vor der Pfadanalyse mithilfe des Shapiro-Wilk Tests auf Normalverteilung geprüft. Dieser Test wird als robustester Normalverteilungstest gesehen (Razali & Wah, 2011). Bei nicht normalverteilten Daten wird die Yuan-Bentler Korrektur (Yuan & Bentler, 1998) von *lavaan* angewendet, wodurch die Parameter zur Bewertung der Modellgüte entsprechend adjustiert werden (Rosseel et al., 2024).

6.4 Datenerhebung und Stichprobe

Für die Datenerhebung wurden erneut Chemie-Vorkurse in den Fokus genommen. Insgesamt haben $N = 494$ Studierende teilgenommen, die zwischen 17 und 51 Jahren alt waren ($M_{\text{Alter}} = 19.5$ Jahre, $SD_{\text{Alter}} = 2.7$ Jahre) und zuletzt mit Schulnoten zwischen 2 Punkten (5-) und 15 Punkten (1+) in Chemie bewertet wurden ($M_{\text{Note}} = 11.5$ Pkt. (2+), $SD_{\text{Note}} = 2.5$ Pkt.). Die Verteilung der Studierenden auf die drei Standorte sowie die Verteilung der Geschlechter sind in Abbildung 18 dargestellt.

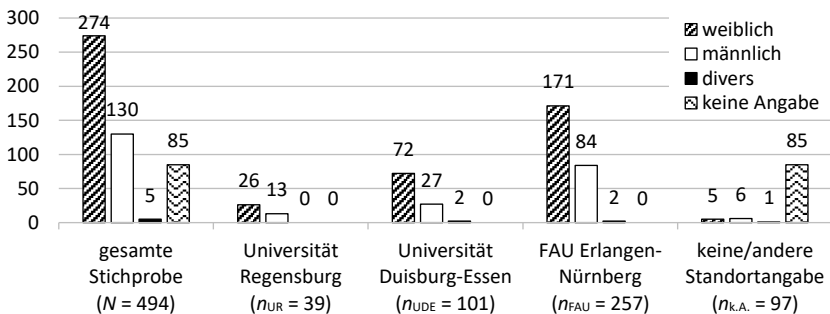


Abbildung 18: Verteilung der Personen getrennt nach Geschlechtern auf die drei Erhebungsorte in der Datenerhebung vom September/Oktober 2024.

Abbildung 19 kann die Verteilung der Studierenden auf die geplanten Studiengänge entnommen werden. Zum Zeitpunkt der Datenerhebung haben noch nicht alle Studierenden eine Zulassung zum intendierten Studiengang erhalten.

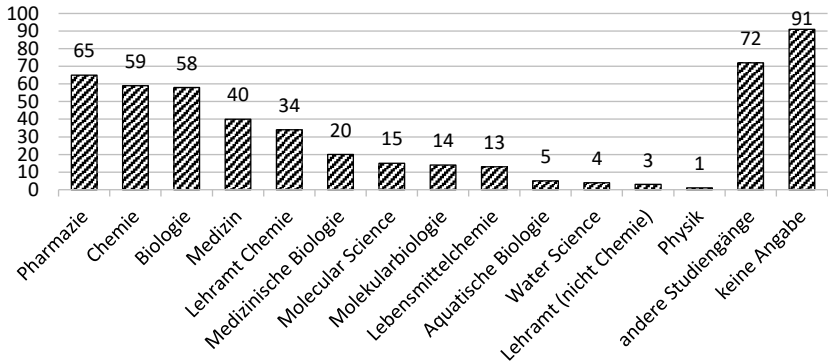


Abbildung 19: Verteilung der Studierenden auf die geplanten Studiengänge in der Datenerhebung vom September/Oktober 2024.

Aufgrund der erhöhten Stichprobenanforderungen der geplanten Analysen wurden Datenerhebungen an mehreren Standorten in vergleichbaren Chemie-Vorkursen durchgeführt. Neben der Friedrich-Alexander-Universität Erlangen-Nürnberg (Bayern) unterstützen auch die Universität Regensburg (Bayern) und die Universität Duisburg-Essen (Nordrhein-Westfalen) die Datenerhebung mit ihren Chemie-Vorkursen. Die drei Standorte unterschieden sich in der Dauer und der Organisation ihrer Vorkurse, weshalb die Datenerhebungen individuell an die Bedingungen vor Ort angepasst wurden. Um eine Beeinflussung der Ergebnisse des CRI:TIC und der Fachwissensitems durch die Inhalte des Vorkurses zu vermeiden, sodass alle Testpersonen unabhängig vom Standort lediglich auf das Schulwissen zurückgreifen konnten, wurden diese Daten jeweils am ersten Tag der Vorkurse erhoben. Bei den räumlichen Faktoren spielten der Erhebungszeitpunkt und die Reihenfolge keine Rolle, da es sich dabei um recht stabile Faktoren handelt, die von den Inhalten der Vorkurse nicht beeinflusst werden. Die individuell angepassten Datenerhebungen werden im Folgenden beschrieben.

Universität Regensburg (UR) – Der Vorkurs hat viertägig vom 02.09. bis zum 05.09.2024 stattgefunden. Am ersten Tag wurde in einem 90-minütigen Zeitfenster das Testheft mit dem CRI:TIC und den Fachwissensitems bearbeitet. Die Zeit war so ausreichend, dass alle Studierenden das Testheft vollständig bearbeiten konnten. Am Folgetag wurden in zwei Etappen die räumlichen Fähigkeiten erhoben. In einem ersten 60-minütigen Durchgang wurden der *Hidden Patterns Test*, der *Gestalt Completion Test*, der *Identical Pictures Test* sowie der *Revised Purdue Spatial Visualization Test: Visualization of Rotations* und in einem zweiten, 50-minütigen Durchgang das *Building Memory*, der *Card Rotations Test*, der *Maze Tracing Speed Test* sowie der *Paper Folding Test* eingesetzt (Tabelle 12). Zwischen den einzelnen Messinstrumenten wurden kurze Pausen eingehalten.

Universität Duisburg-Essen (UDE) – Hier hat der Kurs in der Zeit vom 16.09. bis 27.09.2024 stattgefunden. Die Messung der Repräsentationskompetenzen und des Fachwissens mussten aufgrund der Kursorganisation in zwei Erhebungen am ersten und zweiten Kurstag in jeweils 45 Minuten stattfinden. An den folgenden sieben Tagen wurde jeweils im Anschluss an eine fachinhaltliche Vorlesung ein bis zwei Tests zur Messung der räumlichen Faktoren (Tabelle 12) durchgeführt. Diese wurde dankenswerterweise von den zuvor instruierten Referentinnen in Duisburg-Essen übernommen.

Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg – Der Vorkurs hat wie in Duisburg-Essen über zwei Wochen vom 30.09. bis zum 11.10.2024 stattgefunden. Am ersten Tag wurden in einem 90-minütigen Zeitfenster die Repräsentationskompetenzen und das Fachwissen erfasst. Anschließend wurden in den sieben Folgetagen die räumlichen Faktoren jeweils vor den fachinhaltlichen Vorlesungen erhoben.

Im Vergleich zu der Datenerhebung im Oktober 2023 stand in allen Vorkursen ausreichend Zeit zur Bearbeitung des CRI:TIC und der Fachwissensitems zur Verfügung. Daher wird davon ausgegangen, dass alle unbeantworteten Items aufgrund der Ausprägung des gemessenen Konstrukts nicht beantwortet werden konnten. Diese Fehlwerte werden entsprechend als *Missings Not At Random* (MNAR) und mit 0 Punkten gewertet (Lüdtke et al., 2007; Rubin, 1976).

6.5 Ergebnisse

Zunächst wurden die Untersuchungen aus Forschungsanliegen I wiederholt, um die Replizierbarkeit der Ergebnisse mit einer größeren Stichprobe zu prüfen. Anschließend werden die Ergebnisse für Forschungsanliegen II dargestellt.

6.5.1 Forschungsanliegen I – Wiederholung der Analysen zur Prüfung der Replizierbarkeit

6.5.1.1 Evaluation des CRI:TIC

Basierend auf den Ergebnissen von Forschungsfrage 1.1 wird in den folgenden Analysen ein eindimensionales Partial-Credit-Modell verwendet, um das CRI:TIC zu evaluieren. Insgesamt liegen 251 vollständige Datensätze vor ($n_{UR} = 38$, $n_{UDE} = 89$, $n_{FAU} = 116$, $n_{K.A.} = 8$). Da sich die Lehrpläne der Bundesländer für das Fach Chemie zwar aus den Bildungsstandards ableiten, aber individuell von diesen umgesetzt werden, mussten die Items hinsichtlich ihrer Fairness für unterschiedliche Standorte mit einer *DIF*-Analyse geprüft werden. Der Generalized Mantel-Haenszel Test zeigt fünf Items, für die ein *DIF* mit hohem Effekt ($\Delta MH > 1.5$) in Bezug auf den Standort angenommen werden muss (Tabelle 13). Mit Blick auf die lokale stochastische Unabhängigkeit lassen sich bei vier von 528 (0.76 %) Itempaaren korrelierte Residuen identifizieren ($|aQ3| > 0.20$; Tabelle 13), während das globale Maß bei $MADaQ3 = 0.056$ liegt. Beide Ergebnisse sprechen mit hoher Evidenz für lokale stochastische Unabhängigkeit.

Die absolute Modellgüte liegt mit $SRMR = 0.092$ in einem akzeptablen Bereich und die WLE-Reliabilität lässt sich mit $REL_{RK} = 0.889$ als gut bezeichnen. Mit Blick auf die *infit* und *outfit* Statistik lassen sich jedoch zwei bzw. vier Items mit kritischen Werten erkennen (Tabelle 13). Im Vergleich der durchschnittlichen Personenfähigkeit ($M_{\theta} = 0.02$ logits, $SD_{\theta} = 0.84$ logits) mit der mittleren Itemschwierigkeit ($M_{\xi} = 0.23$ logits, $SD_{\xi} = 1.41$ logits) zeigt sich, dass das CRI:TIC immer noch tendenziell zu schwierig für die Stichprobe ist. Mit Blick auf die Daten vom Oktober 2023 lässt sich jedoch erkennen, dass die Differenz hier deutlich geringer ist.

Auch hier sollten die Werte im Zusammenhang betrachtet werden und darauf basierend die Items noch einmal evaluiert und ggf. einzelne Items aus weiteren Analysen ausgeschlossen werden. In Tabelle 13 lassen sich zwei Items identifizieren (*Int_12* und *Int_13*), die mehrere kritische Parameter aufwiesen. Eine Evaluation der Daten und der Items konnte hier keine logische Erklärung dieser kritischen Werte liefern, weshalb diese aufgrund schlechter *infit* und *outfit* Statistiken, kritischer *aQ3*-Werte und/oder eines signifikanten *DIF* für die weiteren Analysen ausgeschlossen wurden. Weitere Items mit einem einzelnen schlechten Parameter wurden ebenfalls geprüft, aufgrund keiner erkennbaren Auffälligkeiten jedoch für die weiteren Analysen beibehalten. Ein einzelner kritischer Parameter legitimiert den Ausschluss aus dem weiteren Verfahren nicht, weshalb 31 Items für die weiteren Analysen verblieben. Die Ergebnisse legitimieren auch den Zusammenschluss der Daten aus Nordrhein-Westfalen und Bayern.

Tabelle 13: Evaluation der Items des CRI:TIC hinsichtlich der lokalen stochastischen Unabhängigkeit (*aQ3*), *DIF* (*Generalized Mantel-Haenszel X²-Tests*) und des Itemfits (*infit* und *outfit* Statistik).

Item	<i>aQ3</i>	<i>DIF</i>	<i>infit</i>		<i>outfit</i>	
	kritische Werte $ aQ3 > .20$	<i>p</i> -Wert des <i>GMH X²-Tests</i>	<i>MNSQ</i>	<i>ZSTD</i>	<i>MNSQ</i>	<i>ZSTD</i>
Int_01		0.883	0.910	-1.445	0.834	-1.915
Int_03		0.563	0.947	-0.681	0.889	-0.989
Int_04		0.975	0.986	-0.137	0.985	-0.134
Int_05		0.868	0.862	-1.992	0.814	-1.956
Int_06		0.595	0.964	-0.158	1.007	0.113
Int_07		0.392	1.110	1.253	1.267	1.947
Int_09		0.484	0.932	-1.423	0.906	-1.521
Int_10		0.008	0.904	-1.972	0.852	-2.313
Int_11		0.070	1.130	0.812	1.603	2.192
Int_12	-.23	0.000	1.399	2.971	1.664	3.559
Int_13	-.22	0.778	1.653	4.292	3.061	6.852
Int_14	.23	0.947	0.975	-0.183	0.951	-0.236

Trans_01		0.426	1.049	0.596	1.057	0.486
Trans_02		0.091	1.113	2.436	1.145	2.499
Trans_03		0.048	0.964	-0.405	0.904	-0.753
Trans_05		0.230	1.173	1.879	1.193	1.609
Trans_06	-.22	0.016	0.805	-2.570	0.771	-2.854
Trans_07		0.127	0.925	-0.547	0.741	-1.407
Trans_09		0.306	1.081	1.314	1.075	0.878
Trans_10	.24	0.128	0.865	-1.638	0.808	-1.180
Trans_11	.24	0.619	0.934	-0.547	0.876	-1.104
Konst_01	-.23	0.090	0.967	-0.315	1.155	0.941
Konst_02		0.422	0.968	-0.188	0.976	-0.050
Konst_03		0.284	1.051	0.674	1.090	1.031
Konst_05		0.563	0.881	-1.578	0.851	-1.791
Konst_06		0.415	0.984	-0.027	0.866	-0.425
Konst_07		0.031	0.944	-1.230	0.963	-0.634
Konst_08		0.120	0.933	-0.694	0.845	-1.026
Konst_09		0.414	0.941	-0.164	0.462	-1.182
Konst_10	.23	0.175	0.989	-0.065	1.117	0.726
Konst_11		0.333	0.818	-1.879	0.716	-2.008
Konst_12		0.892	1.009	0.114	1.100	0.603
Konst_13		0.960	0.891	-1.434	0.810	-1.323

GMH = Generalized Mantel-Haenszel Test; **fett** gedruckte Werte = kritische Werte; grau hinterlegte Items = ausgeschlossene Items

Im nächsten Schritt wurden die 31 Items wieder für die Schätzung des ein-, zwei- und dreidimensionalen Modells verwendet. Die Ergebnisse der Analysen können zusammengefasst aus Tabelle 14 entnommen werden.

Tabelle 14: Vergleich der lokalen stochastischen Unabhängigkeit (*MADaQ3*, $\alpha Q3$), der WLE-Reliabilität und des Modells (*SRMR*) sowie Itemfits (*outfit* und *infit* *MNSQ*) für die drei Modelle aus FF1.1.

	1D Modell Int + Trans + Konst	2D Modell Int Trans + Konst	3D Modell Int Trans Konst
<i>MADaQ3</i>	0.053	0.053	0.054
$\alpha Q3$	$-0.18 \leq \alpha Q3 \leq 0.24$	$-0.18 \leq \alpha Q3 \leq 0.25$	$-0.21 \leq \alpha Q3 \leq 0.25$
<i>SRMR</i>	0.082	0.082	0.082
WLE-Reliabilität	$Rel = 0.890$	$Rel_{Int} = 0.637$ $Rel_{Trans+Konst} = 0.861$	$Rel_{Int} = 0.637$ $Rel_{Trans} = 0.654$ $Rel_{Konst} = 0.791$
<i>outfit</i> <i>MNSQ</i>	$0.546 \leq outfit \leq 1.814$	$0.558 \leq outfit \leq 1.820$	$0.592 \leq outfit \leq 1.833$
<i>infit</i> <i>MNSQ</i>	$0.806 \leq infit \leq 1.282$	$0.809 \leq infit \leq 1.285$	$0.817 \leq infit \leq 1.241$

Int = Interpretation, Trans = Translation, Konst = Konstruktion

Die lokale stochastische Unabhängigkeit kann für alle drei Modelle angenommen werden, da maximal drei Itempaare (0.65 %) korrelierte Residuen aufweisen und die *MADaQ3*-Werte deutlich für eine globale lokale stochastische Unabhängigkeit sprechen. Die Modellgüte wird durch den *SRMR* für alle Modelle als gut gewertet. Die WLE-Reliabilität ist für das eindimensionale Modell, die Translations- und Konstruktionsitems im zweidimensionalen Modell und die Konstruktionsitems im dreidimensionalen Modell in einem guten Bereich, während die Interpretationsitems im zwei- und dreidimensionalen Modell sowie die Translationsitems im dreidimensionalen Modell in einem noch akzeptablen Bereich liegen. Die *infit* Statistik kann für alle Modelle als gut bewertet werden. Mit Blick auf den *outfit* lässt sich ein Item mit einem *Underfit* (*outfit* < 1.3) erkennen und drei Items mit einem leichten (*outfit* > 1.3) sowie ein Item mit einem starken (*outfit* >> 1.3) *Overfit* erkennen. Da diese jedoch die einzigen kritischen Werte für die jeweiligen Items sind, werden keine weiteren Items ausgeschlossen.

6.5.1.2 Untersuchung der Trennbarkeit der lower-level skills der Repräsentationskompetenzen

Zusammengefasst lässt sich aus den Ergebnissen ableiten, dass diese den Vergleich der Modelle zur Untersuchung der Struktur der Repräsentationskompetenzen legitimieren. Das *AICc* zeigt hier für das ein- und zweidimensionale Modell einen sehr geringen Unterschied ($\Delta AICc < 2$). Das *saBIC* attestiert für alle drei Modelle vergleichbare Werte ($1.02 \leq \Delta saBIC \leq 2.51$). Diese geringen Unterschiede lassen sich so interpretieren, dass es keinen Vorteil für die komplexeren mehrdimensionalen Modelle gibt (Burnham & Anderson, 2004). Der Likelihood Ratio Test attestiert hier jedoch im Vergleich zu dem eindimensionalen Modell eine bessere Passung des zweidimensionalen ($\chi^2(2) = 6.204$, $p = .045$) und des dreidimensionalen Modells ($\chi^2(5) = 15.465$, $p = .008$). Der Vergleich der mehrdimensionalen Modelle zeigt eine bessere Passung des dreidimensionalen Modells ($\chi^2(2) = 9.261$, $p = 0.026$).

Tabelle 15: Vergleich der drei Modelle in der Hauptstudie anhand des *AICc* und des *saBIC*, basierend auf der Log-Likelihood (*LL*), der geschätzten Modellparameter (n_p) und der Stichprobengröße (*N*).

	<i>LL</i>	n_p	<i>N</i>	<i>AICc</i>	<i>saBIC</i>
1D Modell	-5016.014	61	251	10194.05	10175.70
2D Modell	-5012.912	63	251	10194.95	10174.21
3D Modell	-5008.281	68	251	10204.12	10176.72

fett gedruckte Werte = beste Werte

Die Cross Plots mit den Personenfähigkeiten des eindimensionalen Modells auf der Ordinate und den Personenfähigkeiten der einzelnen Dimensionen aus den zwei- und dreidimensionalen Modellen auf der Abszisse unterstützen die Annahme eines eindimensionalen Konstrukts (Abbildung 20). Obwohl hier in drei Diagrammen Datenpunkte außerhalb der Messfehlerlinien ($CI = 95\%$) liegen, sind diese eher als zufällig zu betrachten und reichen nicht zur Unterstützung der Annahme eines mehrdimensionalen Konstrukts. Mit Blick auf die Korrelation zwischen den Dimensionen des dreidimensionalen Modells zeigt sich eine hohe Korrelation. Die Personenfähigkeit der *Konstruktion* korreliert gleich hoch mit denen der *Interpretation* ($r = 0.72$) und *Translation* ($r = 0.72$).

Die Korrelationen zwischen den Personenfähigkeiten von *Interpretation* und *Translation* ist etwas geringer ($r = 0.61$).

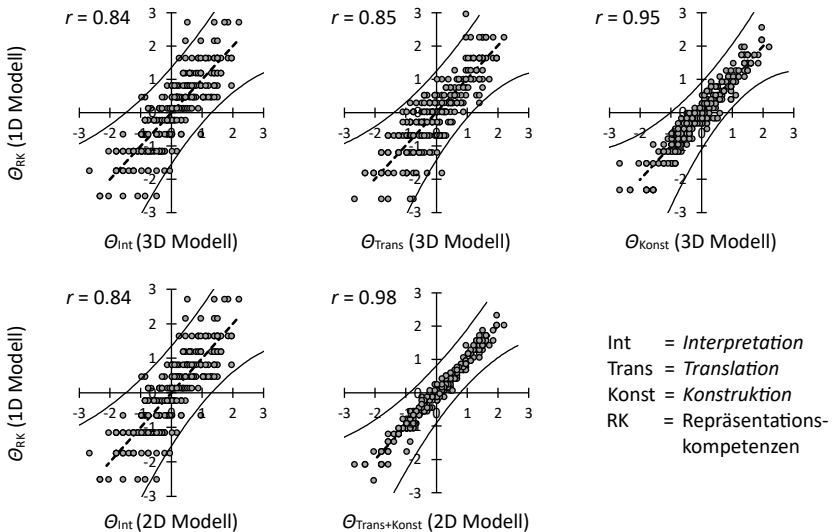


Abbildung 20: Cross Plots mit Messfehlerlinien (CI = 95 %) und Pearson-Korrelation (r) der Personenfähigkeiten (θ) aus dem eindimensionalen Modell (1D) auf der Ordinate und der Personenfähigkeiten aus dem dreidimensionalen Modell (3D, oben) bzw. dem zweidimensionalen Modell (2D, unten) auf der Abszisse (Daten der Hauptstudie, September/Oktober 2024).

6.5.1.3 Evaluation der Fachwissensitems

Ebenso wie in Kapitel 6.5.1.1 beschrieben, wurden die Fachwissensitems erneut evaluiert. Die *DIF*-Analyse mithilfe des Mantel-Haenszel χ^2 -Tests zeigt für vier Items ein *DIF* mit hohem Effekt ($\Delta MH > 1.5$), d. h. diese Items waren für die Standorte Nordrhein-Westfalen und Bayern unterschiedlich schwierig zu beantworten (Tabelle 16). Die *aQ3*-Statistik zeigt, dass eines von 120 Itempaaren (0.83 %) korrelierte Residuen $|aQ3| > 0.20$ aufweist, und auch der globale Wert mit $MADaQ3 = 0.063$ deutet mit hoher Evidenz auf lokale stochastische Unabhängigkeit hin. Die absolute Modellgüte lässt sich mit $SRMR = 0.104$ als noch akzeptabel bezeichnen. Auf der Item-Ebene zeigen alle Items eine gute *infit* Statistik, die *outfit* Statistik belegt jedoch drei Items mit

einem *Overfit*. Bei zwei Items kann dieser aufgrund des *outfit ZSTD* als zufällig betrachtet werden, bei dem dritten Item (Fach_09) wird der *outfit MNSQ* nur sehr leicht überschritten, weshalb auch diese Überschreitung hier als akzeptabel und wenig kritisch gesehen wird (Tabelle 16). Die WLE-Reliabilität ist mit $Rel = 0.680$ ebenfalls noch akzeptabel. Auch die Fachwissensitems sind im Durchschnitt noch zu schwierig ($M_{\xi} = 0.46$ logits, $SD_{\xi} = 1.09$ logits) für die Testpersonen ($M_{\theta} = -0.03$ logits, $SD_{\theta} = 1.16$ logits), wobei auch hier die Differenz deutlich geringer ist als in der Datenerhebung von Oktober 2023.

Betrachtet man die kritischen Werte für die Items in Tabelle 16, lässt sich ein Item mit zwei kritischen Statistiken erkennen (Fach_09). Die *outfit* Statistik wurde jedoch aufgrund der leichten Überschreitung des Cut-Off Kriteriums als wenig kritisch eingestuft und das Item bei den weiteren Analysen einbezogen. Die restlichen Items, die jeweils einen einzelnen kritischen Wert aufzeigen, wurden evaluiert und ebenfalls für die weiteren Analysen berücksichtigt. Dadurch lässt sich auch hier der Zusammenschluss der Datensätze aus Nordrhein-Westfalen und Bayern legitimieren.

Tabelle 16: Evaluation der Fachwissensitems hinsichtlich der lokalen stochastischen Unabhängigkeit ($\alpha Q3$), *DIF* (Mantel-Haenszel χ^2 -Tests) und des Itemfits (*infit* und *outfit* Statistik).

Item	$\alpha Q3$	<i>DIF</i>	<i>infit</i>		<i>outfit</i>	
	kritische Werte $ \alpha Q3 > .20$	p-Wert des MH χ^2 -Tests	<i>MNSQ</i>	<i>ZSTD</i>	<i>MNSQ</i>	<i>ZSTD</i>
Fach_02		0.958	0.925	-1.585	0.894	-1.620
Fach_03		0.987	0.961	-0.653	0.924	-0.890
Fach_04		0.188	1.122	1.998	1.157	1.586
Fach_05		0.000	0.936	-1.334	0.913	-1.307
Fach_06		0.691	1.003	0.067	0.985	-0.157
Fach_07	.21	0.796	0.916	-1.379	0.872	-1.425
Fach_08		0.037	0.918	-1.498	0.897	-1.187
Fach_09		0.005	1.237	4.579	1.322	4.395
Fach_10	.21	0.110	0.948	-0.541	0.902	-0.644
Fach_12		0.925	0.985	-0.286	0.983	-0.229
Fach_13		0.858	0.939	-0.832	0.878	-0.995

Fach_14	0.297	1.139	1.072	1.442	1.827
Fach_15	0.591	0.988	-0.087	0.948	-0.242
Fach_16	0.374	0.898	-2.040	0.846	-2.082
Fach_17	0.000	1.005	0.127	1.016	0.256
Fach_18	0.664	1.124	0.709	1.552	1.631

MH = Mantel-Haenszel Test; **fett** gedruckte Werte = kritische Werte

Im nächsten Schritt wird zur Prüfung der Trennbarkeit von Repräsentationskompetenzen und Fachwissen die Items des CRI:TIC einbezogen und ein ein-dimensionales und zweidimensionales Modell geschätzt (Tabelle 17).

Tabelle 17: Vergleich der lokalen stochastischen Unabhängigkeit (*MADaQ3*, *aQ3*), der WLE-Reliabilität und des Modells (*SRMR*) und Itemfits (*outfit* und *infit* MNSQ) für die zwei Modelle (FF1.2).

	1D Modell RK + FW	2D Modell RK FW
<i>MADaQ3</i>	0.057	0.055
<i>aQ3</i>	$-0.22 \leq aQ3 \leq 0.27$	$-0.21 \leq aQ3 \leq 0.27$
<i>SRMR</i>	0.098	0.096
WLE-Reliabilität	$Rel = 0.90$	$Rel_{RK} = 0.89$ $Rel_{FW} = 0.68$
<i>outfit</i> MNSQ	$0.578 \leq outfit \leq 1.951$	$0.538 \leq outfit \leq 1.798$
<i>infit</i> MNSQ	$0.778 \leq infit \leq 1.365$	$0.801 \leq infit \leq 1.307$

Mit Blick auf die lokale stochastische Unabhängigkeit (*aQ3*, *MADaQ3*), die Modellgüte (*SRMR*) und die WLE-Reliabilität lassen sich alle Werte für beide Modelle als akzeptabel bis gut beschreiben. *Infit* und *outfit* Statistiken zeigen einzelne Items mit *Under*- und *Overfit*, die jedoch durch die entsprechenden *ZSTD* als nicht signifikant relativiert werden bzw. die einzigen kritischen Werte für diese Items darstellen. Somit sind Modell- und Itemfit für beide Modelle akzeptabel.

6.5.1.4 Untersuchung der Trennbarkeit von Fachwissen und Repräsentationskompetenzen

Die vorangegangenen Analysen legitimieren die folgenden Modellvergleiche des ein- und zweidimensionalen Modells hinsichtlich ihrer Passung. Die Informationskriterien *AICc* und *saBIC* favorisieren hier eindeutig das komplexere, zweidimensionale Modell. Dieses Ergebnis wird auch von dem Likelihood Ratio Test ($\chi^2(2) = 61.627, p < 0.001$) gestützt.

Tabelle 18: Vergleich der zwei Modelle in der Hauptstudie anhand des *AICc* und des *saBIC*, basierend auf der Log-Likelihood (*LL*), der geschätzten Modellparameter (n_p) und der Stichprobengröße (*N*).

	<i>LL</i>	n_p	<i>N</i>	<i>AICc</i>	<i>saBIC</i>
1D Modell	-7269.03	77	251	14761.50	14719.43
2D Modell	-7238.22	79	251	14708.36	14662.51

fett gedruckte Werte = beste Werte

Ein Blick auf die Cross Plots unterstützt dieses Bild ebenfalls (Abbildung 21). Mit Blick auf das zweite Diagramm zeigen sich einige Datenpunkte außerhalb der Messfehlerlinien, mehr als in den Cross Plots zur Untersuchung der Trennbarkeit der lower-level skills der Repräsentationskompetenzen (Abbildung 20). Die Korrelation zwischen den Personenfähigkeiten der Repräsentationskompetenzen und des Fachwissens zeigen einen hohen Zusammenhang zwischen den beiden latenten Variablen ($r = 0.60$).

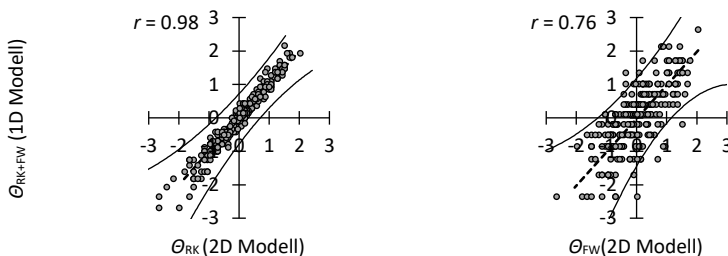


Abbildung 21: Cross Plots mit Messfehlerlinien (CI = 95 %) und Pearson-Korrelation (r) der Personenfähigkeiten (Θ) aus dem eindimensionalen Modell (1D) auf der Ordinate und der Personenfähigkeiten für die Repräsentationskompetenzen (RK) und das Fachwissen (FW) aus dem zweidimensionalen Modell (2D) auf der Abszisse (Daten der Hauptstudie, September/Oktober 2024).

6.5.2 Forschungsanliegen II – Untersuchung des Zusammenhangs zwischen Repräsentations-kompetenzen, Fachwissen und räumlichen Faktoren

Zur Untersuchung des Zusammenhangs zwischen den Repräsentationskompetenzen, des Fachwissens und der räumlichen Faktoren wurden in einem ersten Schritt die Instrumente zur Messung der räumlichen Faktoren evaluiert und anschließend Korrelationsanalysen durchgeführt (FF2.1). In der Folge wurden Pfadanalysen modelliert, um das hypothetisch angenommene Wirkmodell (Abbildung 10) zu prüfen.

6.5.2.1 Evaluation der Messinstrumente zu den räumlichen Faktoren und Korrelationsanalysen

Bei den eingesetzten Messinstrumenten handelt es sich, wie bereits beschrieben, um standardisierte und validierte psychometrische Messinstrumente. Im Rahmen dieses Forschungsanliegens ist es jedoch sinnvoll, die Zuverlässigkeit der Messung über Cronbachs α (interne Konsistenz) zu bewerten, da diese Größe stichprobenabhängig ist.

In Tabelle 19 lassen sich die Werte für Cronbachs α und entsprechende Referenzwerte aus den Manualen und Publikationen zu den Messinstrumenten (Ekstrom et al., 1976b; Maeda et al., 2013) entnehmen. Die ermittelten Werte für Cronbachs α liegen alle in einem guten bis exzellenten Bereich und sind mit $\alpha \geq 0.79$ angemessen für Intelligenztests (Field, 2013). Eine Ausnahme stellt der *Gestalt Completion Test* dar, bei dem Cronbachs α gerade noch akzeptabel ist.

Tabelle 19: Stichprobengröße (N), Anzahl der Items und Cronbachs α mit Referenzwerten für die Messungen der acht räumlichen Faktoren in der Datenerhebung vom September/Oktober 2024.

Messinstrument Räumlicher Faktor	N	Anzahl der Items	Cronbachs α	Referenzwerte (^a Ekstrom et al., 1976b; ^b Maeda et al., 2013)
Building Memory Visual Memory	221	24	0.79	0.80 ^a
Card Rotations Test Speeded Rotation	305	160	0.97	$0.80 \leq \alpha \leq 0.89$ ^a
Gestalt Completion Test Closure Speed	288	20	0.69	0.85 ^a
Hidden Patterns Test Flexibility of Closure	281	400	0.98	$0.81 \leq \alpha \leq 0.91$ ^a
Identical Pictures Test Perceptual Speed	293	96	0.95	$0.81 \leq \alpha \leq 0.87$ ^a
Maze Tracing Speed Test Spatial Scanning	280	48	0.93	$0.89 \leq \alpha \leq 0.94$ ^a
Paper Folding Test Visualization	210	20	0.79	$0.75 \leq \alpha \leq 0.84$ ^a
Revised PSVT:R Spatial Relation	270	30	0.84	0.84 ^b

fett gedruckte Werte = kritische Werte

Mit Blick auf Forschungsfrage 2.1 werden bivariate Korrelationsanalysen mit Bonferroni-Korrektur zur Anpassung des α -Levels (α -Fehlerkumulation) durchgeführt. Hier wird Spearmans Rangkorrelationskoeffizient ρ zur Beantwortung der Forschungsfrage genutzt, da die Daten der räumlichen Faktoren ordinalskaliert- und nur die Personenfähigkeiten der Repräsentationskompetenzen und des Fachwissens intervallskaliert sind. Außerdem sind die empirischen Daten größtenteils nicht normalverteilt (Anhang B.1). Zwar betont Döring (2023d), dass die Verfahren nach Pearson und Spearman häufig zum glei-

chen Ergebnis kommen, hier ist Spearmans Rangkorrelationskoeffizient jedoch das statistisch adäquate Verfahren für die Beantwortung der Forschungsfrage. Die Ergebnisse der Korrelationsanalysen sind in Tabelle 20 dargestellt.

Tabelle 20: Spearmans Rangkorrelationskoeffizient ρ mit Bonferroni-Korrektur zwischen den räumlichen Faktoren und den Repräsentationskompetenzen sowie dem Fachwissen.

	Spearmans Rangkorrelationskoeffizient ρ für ...	
	die Repräsentationskompetenzen	das deklarative Fachwissen
Visual Memory	0.254 #	0.015
Speeded Rotation	0.395 ***	0.226 *
Closure Speed	-0.018	-0.061
Flexibility of Closure	0.328 ***	0.129
Perceptual Speed	0.291 **	0.125
Spatial Scanning	0.204 #	0.110
Visualization	0.446 ***	0.090
Spatial Relation	0.539 ***	0.173

$p \leq 0.100$, * $p \leq 0.050$, ** $p \leq 0.010$, *** $p \leq 0.001$; nicht-signifikante Koeffizienten sind grau dargestellt

Auf den ersten Blick lässt sich erkennen, dass alle räumlichen Faktoren, außer dem Faktor *Closure Speed*, der Fähigkeit, unvollständige visuelle Reize zu erkennen, mindestens marginal signifikant mit den Repräsentationskompetenzen korrelieren. Die Fähigkeiten, sich visuelle Reize über kurze Zeit einzuprägen (*Visual Memory*) und den Weg aus einem visuellen, verzweigten Feld zu finden (*Spatial Scanning*), korrelieren nur schwach und marginal signifikant mit den Repräsentationskompetenzen. Eine ebenfalls schwache, aber signifikante Korrelation mit den Repräsentationskompetenzen zeigen die Fähigkeiten, visuelle Reize schnell miteinander zu vergleichen (*Perceptual Speed*) und bekannte Figuren in komplexen Mustern zu erkennen (*Flexibility of Closure*).

Die Fähigkeiten, Probleme durch die zweidimensionale mentale Rotation einfacher Objekte (*Speeded Rotation*) und die dreidimensionale mentale Rotation komplexer Objekte (*Spatial Relation*) zu lösen und Objekte mental zu manipulieren (*Visualization*), weisen nach Akoglu (2018) eine mittelstarke Korrelation mit den Repräsentationskompetenzen auf.

Das Fachwissen korreliert nur mit dem Faktor *Speeded Rotation*. Ohne Anwendung der Bonferroni-Korrektur zeigen zusätzlich nur die Faktoren *Perceptual Speed* und *Spatial Relation* eine signifikante Korrelation ($p \leq 0.05$). Bei der Betrachtung der Korrelationskoeffizienten lässt sich unschwer erkennen, dass maximal ein schwacher Zusammenhang ($p \leq 0.226$) zwischen den Variablen besteht.

Der *Gestalt Completion Test* weist neben einer fragwürdigen Reliabilität ($\alpha = 0.69$) als einziger Test keine Korrelation mit den Repräsentationskompetenzen auf. Selbst ohne Bonferroni-Korrektur ist hier kein Zusammenhang erkennbar. Daher wird dieser Test noch einmal kritisch evaluiert. Er besteht aus 20 Items, die unvollständige Objekte, beispielsweise das Piktogramm eines Vogels oder eines Fahrrads verwenden. Aufgabe ist es, die Gegenstände zu identifizieren. Es fällt auf, dass einige der dargestellten Objekte kulturell behaftet oder aus Gründen der Aktualität aus heutiger Sicht weniger bekannt und dadurch für bestimmte Populationen schwierig oder gar nicht zu beantworten sind. Beispiele für kulturell behaftete Objekte aus dem *Gestalt Completion Test* sind der Spielzeug Hand- bzw. Bollerwagen (typischerweise in Rot; *red wagon* vom Spielzeughersteller *Radio Flyer*®) und das Western-Windrad, hauptsächlich bekannt im amerikanischen Raum. Das Wählscheibentelefon ist ein Beispiel für eine Darstellung, die aus heutiger Sicht problematisch sein könnte, da dieses kaum noch in Haushalten zu finden ist. In einem Item ist eine Tischlampe mit kegelförmigem Lampenschirm dargestellt, die durch Auslassungen auch als modernes Weinglas identifiziert werden kann. Insgesamt muss somit nicht nur die Reliabilität des Messinstruments in Frage gestellt werden, sondern auch die Validität. Aus diesen Gründen kann der Test im weiteren Vorgehen nicht in die Analysen aufgenommen werden, da die Ergebnisse nicht einwandfrei belastbar wären.

6.5.2.1 Pfadanalysen

Die Korrelationsanalysen geben einen ersten Überblick über Zusammenhänge zwischen den gemessenen Konstrukten. Sodann wird eine Pfadanalyse durchgeführt, um Variablen zu identifizieren, die – auf Grundlage vorangegangener Studien – als Prädiktoren für die Leistung bei der Bearbeitung repräsentationsbasierter Chemieaufgaben angenommen werden. Wie in Kapitel 6.3 beschrieben, werden dazu sämtliche räumliche Faktoren und das deklarative Wissen als exogene (unabhängige) Variable operationalisiert, die auf die Repräsentationskompetenzen (endogene bzw. abhängige Variable) wirken. Die räumlichen Faktoren werden im Modell untereinander korreliert, jedoch nicht mit dem deklarativen Fachwissen (Abbildung 10).

Insgesamt werden für das Modell $n_P = 47$ Modellparameter bei einer Stichprobengröße von $N = 494$ Personen geschätzt. Basierend auf diesem Verhältnis lassen sich die Ergebnisse der Pfadanalyse als belastbar einordnen (Streiner, 2005). Die Daten der Messinstrumente sind zum Großteil nicht normalverteilt (Anhang B.1), weshalb hier die Yuan-Bentler Korrektur (Yuan & Bentler, 1998) auf die Parameter zur Bewertung der Modellgüte angewendet wird. Der χ^2 -Test zeigt, dass sich die empirischen Daten signifikant vom Modell unterscheiden, eine Passung somit nicht gegeben ist. Da die Stichprobe hier sehr groß ist ($N > 400$), sollten die anderen Modellgütekriterien stärker als der χ^2 -Test gewichtet werden, da dieser Test sehr sensitiv für große Stichproben ist (Kenny, 2024; Zheng & Bentler, 2025). Der $CFI = 0.983$ und der $SRMR = 0.066$ weisen eine sehr gute Passung des Modells zu den empirischen Daten auf. $TLI = 0.912$ und $RMSEA = 0.087$ zeigen eine weniger gute, aber noch akzeptable Passung. Insgesamt kann somit eine ausreichende Passung der empirischen Daten zum Modell angenommen werden.

Die Ergebnisse der Pfadanalyse sind in Abbildung 22 dargestellt. Von den räumlichen Faktoren haben sich die Fähigkeiten, Objekte mental zu manipulieren (*Visualization*) und Probleme durch dreidimensionale mentale Rotation zu lösen (*Spatial Relation*) als Prädiktoren für die Repräsentationskompetenzen ergeben. Die Fähigkeit, Probleme durch zweidimensionale Rotation zu lösen (*Speeded Rotation*) ist marginal signifikant aus der Analyse hervorgegangen. Ebenso ist der Einfluss des Fachwissens auf die Repräsentationskompetenzen signifikant geworden. Dieser weist den höchsten standardisierten Pfadkoeffizienten ($\beta = 0.533$) auf. Der räumliche Faktor *Speeded Rotation* hat

den kleinsten Pfadkoeffizienten mit $\beta = 0.096$, gefolgt von den Faktoren *Visualization* ($\beta = 0.146$) und *Spatial Relation* ($\beta = 0.352$). Insgesamt werden durch das Modell 61.1 % der Varianz der Repräsentationskompetenzen aufgeklärt.

Die Kovarianzen der räumlichen Faktoren nehmen Werte von $0.145 \leq r \leq 0.607$ an. Der geringste Wert ist zwischen *Perceptual Speed* und *Spatial Relation* zu finden, der höchste zwischen *Perceptual Speed* und *Flexibility of Closure*.

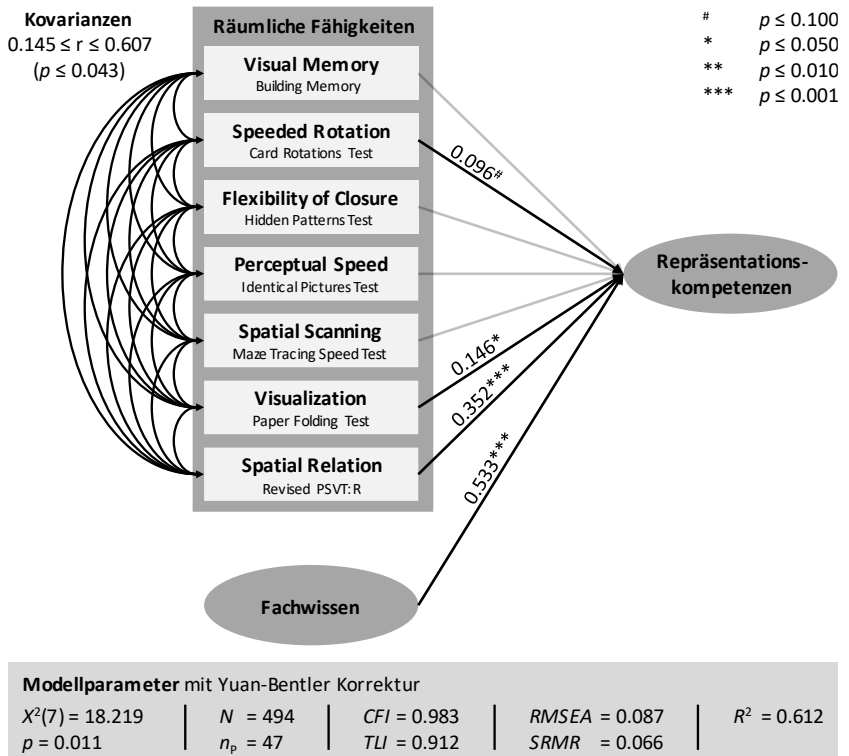


Abbildung 22: Pfadmodell mit standardisierten Pfadkoeffizienten (β) für signifikante Pfade (nicht-signifikante Pfade sind transparent dargestellt) und den Modellparametern zu Beurteilung der Güte.

Mit Blick auf Forschungsanliegen I und Forschungsfrage 1.1 hat sich gezeigt, dass ein eindimensionales Modell für die Repräsentationskompetenzen angenommen werden muss, da die Ergebnisse ein ambivalentes Bild zur Dimensionalität skizzieren. Das Pfadmodell in Abbildung 22 ist mit diesen Ergebnissen konform. Wie in Kapitel 5.7 bereits diskutiert, steht diesen empirischen Befunden eine inhaltlich-theoretische chemiedidaktische Perspektive gegenüber, für die eine getrennte Betrachtung der lower-level skills durchaus sinnvoll ist, besonders mit Blick auf Lehr-Lern-Prozesse. Auch die Zuordnung der Items des CRI:TIC durch verschiedene Rater hat gezeigt, dass sich die Anforderungen in den Items eindeutig einer der lower-level skills *Interpretation*, *Translation* und *Konstruktion* zuordnen lassen ($\kappa_{\text{Fleiss}} = 0.86$). Aufgrund der ambivalenten Ergebnisse bezüglich Forschungsfrage 1.1 und der unterschiedlichen Anforderungen der drei lower-levels skills wird ein zweites Pfadmodell geprüft, in dem die Repräsentationskompetenzen nicht als eindimensionales, sondern als dreidimensionales Konstrukt mit der *Interpretation*, *Translation* und *Konstruktion* als separate, aber interkorrelierte endogene Variablen eingehen. Dadurch soll geprüft werden, ob sich unterschiedliche Prädiktoren für die drei lower-level skills manifestieren und dadurch die unterschiedlichen Anforderungen mit Evidenz untermauern lassen.

Ähnlich wie das erste Pfadmodell weist das zweite Pfadmodell (Abbildung 23) einen signifikanten χ^2 -Test auf – die empirischen Daten passen demnach nicht zu dem Modell. Aber auch hier müssen die anderen Parameter zur Bewertung der Modellgüte stärker gewichtet werden. $CFI = 0.989$ und $SRMR = 0.065$ weisen auf eine sehr gute und $TLI = 0.912$ und $RMSEA = 0.085$ auf eine noch akzeptable Modellpassung hin. Hier ist jedoch klar zu betonen, dass die Ergebnisse des Pfadmodells aufgrund des Verhältnisses der geschätzten Modellparameter ($n_p = 70$) zur Stichprobengröße ($N = 494$) mäßig belastbar sind und lediglich einen explorativen Einblick in die statistischen Zusammenhänge zwischen den drei lower-level skills und den räumlichen Faktoren sowie dem Fachwissen geben, die mit einer gewissen Vorsicht interpretiert werden müssen.

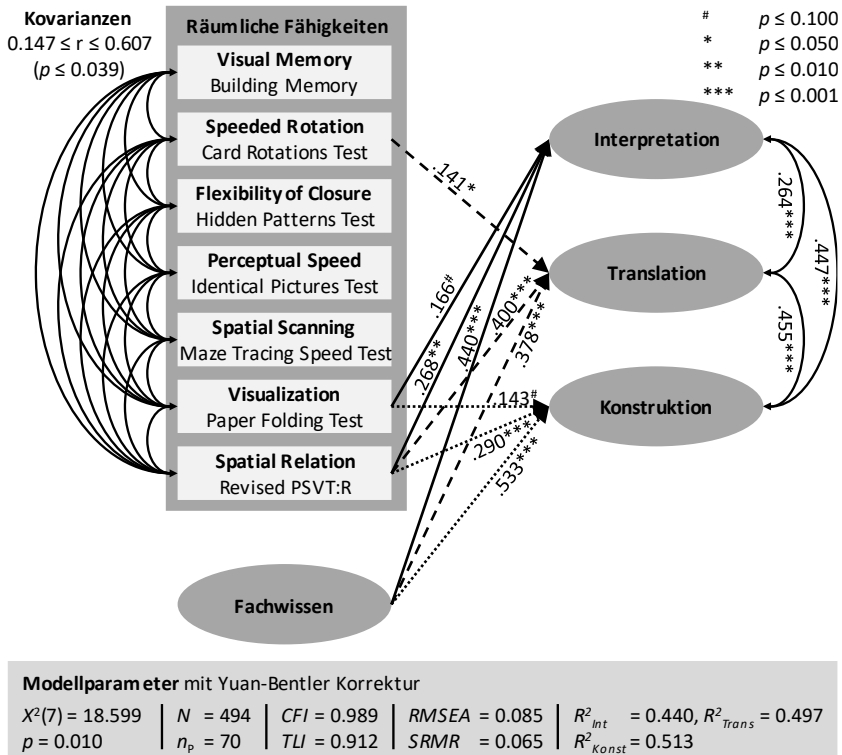


Abbildung 23: Pfadmodell mit standardisierten Pfadkoeffizienten (β) für signifikante Pfade und den Modellparametern zu Beurteilung der Güte für die drei lower-level skills *Interpretation*, *Translation* und *Konstruktion*.

Wie im ersten Pfadmodell haben sich für alle drei lower-level skills das deklarative Fachwissen und der räumliche Faktor *Spatial Relation* als signifikante Prädiktoren ergeben. Hierbei ist interessant, dass bei der *Translation* der standardisierte Pfadkoeffizient für *Spatial Relation* (minimal) höher ist als der des Fachwissens. Bei der *Interpretation* und *Konstruktion* ist es umgekehrt und der Unterschied zwischen den beiden standardisierten Pfadkoeffizienten deutlich größer. Der Pfad vom Faktor *Visualization* zur *Interpretation* ist nur marginal signifikant, zur *Konstruktion* signifikant und fehlt bei der *Translation*. Dieses Muster ist interessant, da der Faktor *Visualization* in Abbildung 22 einer der

zwei signifikanten Pfade der räumlichen Faktoren zur Repräsentationskompetenz ist. Besonders auffällig ist der Pfad vom räumlichen Faktor *Speeded Rotation*, der Fähigkeit, Probleme durch zweidimensionale mentale Rotation zu lösen, zur *Translation*, da dieser hier ein Alleinstellungsmerkmal ist und für die anderen beiden lower-level skills nicht signifikant wird.

Mit Blick auf die Kovarianzen lassen sich bei den räumlichen Faktoren vergleichbare Werte erkennen. Die geringste Kovarianz lässt sich zwischen den räumlichen Faktoren *Perceptual Speed* und *Spatial Relation* ($r_{\min} = 0.152$) erkennen und die höchste zwischen *Visualization* und *Spatial Relation* ($r_{\max} = 0.611$). Bei den lower-level skills ist die Kovarianz zwischen *Interpretation* und *Translation* am geringsten ($r = 0.272$) und zwischen der *Konstruktion* und *Interpretation* bzw. *Translation* vergleichbar hoch ($0.442 \leq r \leq 0.458$). Durch das Modell können für die *Interpretation* 44.0 %, für die *Translation* 49.7 % und für die *Konstruktion* 51.3 % der Varianz aufgeklärt werden.

Ein Vergleich der beiden Modelle anhand von Informationskriterien (z. B. *AIC* oder *BIC*) ist methodisch nicht zulässig, da sich die Modelle hinsichtlich ihrer endogenen Variablen unterscheiden. Ein Vergleich kann daher ausschließlich auf inhaltlicher Ebene erfolgen, etwa durch die Gegenüberstellung und Interpretation der standardisierten Pfadkoeffizienten.

6.6 Zusammenfassung und Diskussion der Ergebnisse

Zunächst zur Betrachtung der Ergebnisse der Analysen zum ersten Forschungsanliegen mit dem neuen Datensatz. Aufgrund eines *Differential Item Functionings* (*DIF*) und schlechter Itemfit Statistiken, sind zwei Items aus dem Bereich *Interpretation* für die weiteren Analysen ausgeschlossen worden. Die restlichen Modell- und Itemfit Statistiken sind jedoch akzeptabel bis gut, was die weiteren Analysen und die Zusammenführung der Daten aus Nordrhein-Westfalen und Bayern legitimiert.

Mit Blick auf die Trennbarkeit der drei lower-level skills *Interpretation*, *Translation* und *Konstruktion* ergeben die Analysen die gleichen Ergebnisse wie im Rahmen des ersten Forschungsanliegens. *AICc* und *saBIC* zeigen für das ein- und zweidimensionale bzw. für alle drei Modelle vergleichbare Werte, wodurch kein Vorteil für eines der komplexeren Modelle attestiert werden kann (Tabelle 15). Der Likelihood Ratio Test hingegen verdeutlicht, dass das

dreidimensionale Modell gegenüber den anderen beiden Modellen die beste Passung hat, ebenso, dass das zweidimensionale Modell besser zu den Daten passt als das eindimensionale Modell. Da die Stichprobengröße und die Anzahl geschätzter Parameter, d. h. die Komplexität des Modells, hier nicht einbezogen wird, sollten *AICc* und *saBIC* stärker gewichtet werden. Zusätzlich favorisieren die Cross Plots das eindimensionale Modell, auch wenn vereinzelt Datenpunkte außerhalb der Messfehlerlinien ($CI = 95\%$) liegen (Abbildung 20). Somit wird auch hier wieder ein ambivalentes Bild skizziert, weshalb das eindimensionale Modell angenommen werden muss. Die Korrelationen zwischen der *Interpretation*, *Translation* und *Konstruktion* des dreidimensionalen Modells zeigen, dass diese in einer starken wechselseitigen Beziehung stehen ($0.61 \leq r \leq 0.72$).

Im Gegensatz dazu sprechen bei der Prüfung der empirischen Trennbarkeit der Repräsentationskompetenzen und des deklarativen Fachwissens alle Ergebnisse dafür, dass es sich um zwei separate Konstrukte handelt. Sowohl die Informationskriterien *AICc* und *saBIC* (Tabelle 18), der Likelihood Ratio Test und die Cross Plots (Abbildung 21) favorisieren das zweidimensionale Modell. Die Konstrukte lassen sich somit auch mit dem neuen Datensatz empirisch trennen, wobei die Korrelation zwischen diesen ebenfalls auf einen hohen Zusammenhang hindeutet ($r = 0.60$).

Die Ergebnisse des ersten Forschungsanliegens lassen sich somit mit dem zweiten größeren Datensatz replizieren, wodurch sich in dieser Arbeit ein einheitliches Bild skizzieren lässt. Die Repräsentationskompetenzen sind empirisch nicht eindeutig zu trennen, weshalb *Interpretation*, *Translation* und *Konstruktion* als eng zusammenhängende, psychometrisch nicht strikt trennbare Konstrukte innerhalb der Repräsentationskompetenzen angenommen werden. Dadurch werden die Ergebnisse vorangegangener Studien in diesem Kontext untermauert, die zu vergleichbaren Ergebnissen kommen (Nitz, Ainsworth, et al., 2014; Scheid et al., 2018; Taskin et al., 2015; Ward et al., 2025). Mit Blick auf das Fachwissen kann dieses empirisch eindeutig von den Repräsentationskompetenzen getrennt werden, was mit den Studien von Nitz, Ainsworth et al. (2014) und Edelsbrunner et al. (2023) konform ist. Dabei sind auch diese beiden Konstrukte hoch korreliert, was darauf schließen lässt, dass sie zwar formal trennbar sind, aber in wechselseitiger Beziehung stehen.

Um den Zusammenhang zwischen Fachwissen und Repräsentationskompetenzen weiter zu untersuchen, erfolgte im Rahmen von Forschungsanliegen II eine Pfadanalyse. Dabei wurde das Fachwissen als Momentaufnahme des aktuellen Wissens betrachtet und somit als Prädiktorvariable für die Repräsentationskompetenzen modelliert. Zentral wurde in diesem Forschungsanliegen jedoch die Rolle der räumlichen Faktoren im Zusammenhang mit dem deklarativen Fachwissen und den Repräsentationskompetenzen untersucht. Dadurch sollten Prädiktoren der räumlichen Faktoren für die Bearbeitung repräsentationsbasierter Chemieaufgaben identifiziert und zusätzlich die Frage untersucht werden, ob und inwieweit sich die unterschiedliche Operationalisierung der lower-level skills *Interpretation*, *Translation* und *Konstruktion* in unterschiedlichen räumlichen Prädiktorvariablen äußert. Zunächst wurden jedoch die Messinstrumente der räumlichen Faktoren evaluiert und die Beziehungen zwischen den räumlichen Faktoren und dem deklarativen Fachwissen sowie den Repräsentationskompetenzen mithilfe von Korrelationsanalysen betrachtet.

FF2.1 Welcher korrelative Zusammenhang lässt sich zwischen Repräsentationskompetenzen, Fachwissen und räumlichen Faktoren in der Chemie identifizieren?

Im ersten Schritt wurden die Messinstrumente zur Erhebung der räumlichen Faktoren hinsichtlich ihrer Reliabilität mithilfe von Cronbachs α untersucht. Mit Ausnahme des *Gestalt Completion Test* zur Messung der Fähigkeit, unvollständige Bilder mental zu ergänzen und zu identifizieren (*Closure Speed*), weisen alle Tests gute Reliabilitäten auf. Die anschließend durchgeführten Korrelationsanalysen zeigten, dass alle räumlichen Faktoren, mit Ausnahme des *Closure Speeds*, mit den Repräsentationskompetenzen korrelieren, zwei davon nur marginal signifikant. Aufgrund dieser Auffälligkeit und mit Blick auf die fragwürdige Reliabilität des Messinstruments, wurde dieses noch einmal genauer evaluiert. Dabei hat sich herausgestellt, dass durch kulturelle behaftete und wenig zeitgemäße Bilder die Validität des Messinstruments in Frage gestellt werden muss. Daher wurde dieses Messinstrument für die weiteren Analysen (FF2.2) ausgeschlossen.

Mit Blick auf die verbleibenden sieben räumlichen Faktoren und die Rangkorrelationskoeffizienten lässt sich erkennen, dass die Fähigkeit zur dreidimensionalen mentalen Rotation (*Spatial Relation*) den höchsten Wert ($\rho = 0.539$) aufweist. Dieser wird gefolgt von der mentalen Manipulation (*Visualization*; $\rho = 0.446$), der zweidimensionalen mentalen Rotation (*Speeded Rotation*; $\rho = 0.395$) und der Identifikation von Figuren in Mustern (*Flexibility of Closure*; $\rho = 0.328$). Die Befunde sind insofern interessant, da diese vier Faktoren bereits als relevant für die Naturwissenschaften und besonders die Chemie identifiziert wurden (Bodner & Guay, 1997; Bodner & McMillen, 1986; Carter et al., 1987; Harle & Towns, 2011; Harris et al., 2013; Hodgkiss et al., 2018; Sorby, 2009; Wai et al., 2009; Wu & Shah, 2004). Die zusätzlich aufgenommenen Fähigkeiten, Bilder schnell vergleichen (*Perceptual Speed*) und sich einprägen (*Visual Memory*) zu können sowie einen Weg aus einem komplexen Feld zu finden (*Spatial Scanning*), zeigen hier einen geringeren Zusammenhang mit den Repräsentationskompetenzen ($0.204 \leq r \leq 0.291$). Dahingehend kann Hypothese 2.1 zum Großteil bestätigt werden – alle räumlichen Faktoren korrelieren, wenn auch teilweise nur marginal signifikant, mit den Repräsentationskompetenzen. Dabei ist der Faktor *Closure Speed* aus beschriebenen Gründen kritisch einzuordnen.

Bezüglich des Zusammenhangs zwischen deklarativem Fachwissen und räumlichen Faktoren fällt auf, dass nur der Faktor *Speeded Rotation* mit dem Fachwissen korreliert, wobei dieser Zusammenhang eher als schwach einzuordnen ist ($\rho = 0.226$). Dieser Befund ist spannend, da bisherige Ergebnisse die räumlichen Fähigkeiten, besonders den Faktor *Spatial Relation* (Bodner & Guay, 1997; Carter et al., 1987; Harle & Towns, 2011; Harris et al., 2013; Sorby, 2009), als wichtigen Prädiktor für die Leistung in der Chemie darstellen. Die vorliegenden Ergebnisse zeigen jedoch, dass sie differenziert werden müssen. Die Korrelationsanalysen lassen den Schluss zu, dass räumliche Fähigkeiten im Zusammenhang mit der Leistung in der Chemie stehen, wenn Repräsentationen dabei einen zentralen Stellenwert einnehmen. Hingegen scheinen räumliche Fähigkeiten eher unabhängig von der Leistung zu sein, wenn Aufgaben vornehmlich deklaratives Fachwissen erfordern. Dies lässt sich am Beispiel der Enantiomerie gut nachvollziehen. Eine Person kann Enantiomere ggf. als Verbindungen, die sich wie Bild und Spiegelbild gleichen, sich aber nicht durch Rotation ineinander überführen lassen, definieren. Dieses Wissen muss aber

nicht bedeuten, dass die Person Enantiomere identifizieren kann, wenn verschiedene Moleküle angegeben werden. Dazu werden räumliche Fähigkeiten benötigt, die es einer Person ermöglichen, diese Definition an den Molekülen durch mentale Rotation (*Speeded Rotation* oder *Spatial Relation*) zu prüfen.

Die Ergebnisse der Korrelationsanalysen unterstützen auf den ersten Blick die bisherige Forschung in Bezug auf den Zusammenhang zwischen räumlichen Fähigkeiten und Repräsentationskompetenzen. Die Korrelationsanalyse hat jedoch den Nachteil, dass die Interkorrelationen zwischen den einzelnen Faktoren nicht berücksichtigt werden und auch das Fachwissen als Prädiktor in diesen Analysen keine Berücksichtigung findet. Daher ist die Pfadanalyse eine sinnvolle und logische Erweiterung zu den Analysen, die in Forschungsfrage 2.2 behandelt werden.

FF2.2 Welche räumlichen Faktoren sind beim Bearbeiten repräsentationsbasierter Aufgaben prädiktiv für die Leistung?

Durch die Modellierung eines Pfadmodells können die Interkorrelation der räumlichen Faktoren und das Fachwissen als Prädiktoren für die Bearbeitung der repräsentationsbasierten Chemieaufgaben berücksichtigt werden. Dies hat den Vorteil, dass räumliche Faktoren zu identifizieren sind, die einen einzigartigen Beitrag zur Varianzaufklärung leisten. Mit Blick auf das modellierte Pfadmodell in Abbildung 22 lässt sich erkennen, dass neben dem Fachwissen die dreidimensionale mentale Rotationsfähigkeit (*Spatial Relation*) und die mentale Manipulationsfähigkeit (*Visualization*) signifikante Pfade haben. Die zweidimensionale mentale Rotationsfähigkeit (*Speeded Rotation*) zeigt einen marginal signifikanten Pfad. Diese Prädiktoren klären einen einzigartigen Anteil der Varianz auf, was durch die restlichen Variablen nicht gelingt.

Bei der Betrachtung der Pfadkoeffizienten lässt sich erkennen, dass das Fachwissen den höchsten Pfadkoeffizienten ($\beta = 0.533$) aufweist, was für einen hohen Einfluss des deklarativen Fachwissens auf die Bearbeitung der repräsentationsbasierten Items spricht (Cohen, 1988; Nieminen, 2022). Das deklarative Fachwissen stellt das zum Zeitpunkt der Messung verfügbare Wissen zur Bearbeitung der Aufgaben dar, weshalb es hier als exogene Variable operationalisiert wird. Daher ist es wenig wunderlich, dass es einen hohen Effekt auf

die Bearbeitung der repräsentationsbasierten Items hat, da sich dieser Zusammenhang im bereits dargestellten und diskutierten Repräsentationsdilemma ausdrückt. Repräsentationskompetenzen und Fachwissen beeinflussen sich wechselseitig. Beim Lernen mit Repräsentationen kann das eine nicht ohne das andere entwickelt werden (Rau, 2017a, 2017c; Rau et al., 2021). Die Ergebnisse unterstreichen somit die zentrale Rolle des deklarativen Fachwissens für die erfolgreiche Bearbeitung repräsentationsbasierter Chemieaufgaben und legen nahe, dass dessen Anwendung eine Voraussetzung für die effektive Nutzung von Repräsentationskompetenzen darstellt. Ein Erklärungsansatz dafür liefern Cook und Kolleg:innen (2008), die zeigen konnten, dass sich Studierende mit hohem Vorwissen beim Bearbeiten von repräsentationsbasierten Aufgaben vornehmlich auf konzeptuell relevante Aspekte der Repräsentationen fokussieren, während sich Studierende mit geringem Vorwissen bei solchen Aufgaben auf oberflächliche Merkmale fokussieren.

Mit Blick auf die drei signifikanten räumlichen Faktoren kann der dreidimensionalen mentalen Rotationsfähigkeit (*Spatial Relation*) ein mittlerer Effekt ($\beta = 0.352$) und der zweidimensionalen mentalen Rotationsfähigkeit (*Speeded Rotation*) sowie der mentalen Manipulationsfähigkeit (*Visualization*) jeweils ein kleiner Effekt ($0.096 \leq \beta \leq 0.146$) zugesprochen werden. Diese drei Faktoren zeigen in den Korrelationsanalysen (FF2.1) die höchsten Rangkorrelationskoeffizienten.

Die dargestellten Ergebnisse lassen sich gut in aktuelle Befunde einordnen. Dem Faktor *Spatial Relation* wird eine hohe Relevanz als Prädiktor für die Leistung in der Chemie in zahlreichen Studien zugesprochen. Der Faktor ist in diesem Kontext der am häufigsten untersuchte (Bodner & Guay, 1997; Carter et al., 1987; Harle & Towns, 2011; Harris et al., 2013; Kenneally & Bentley, 2024; Sorby, 2009; Stieff et al., 2018). Auch die Faktoren *Speeded Rotation* (Harris et al., 2013; Wu & Shah, 2004) und *Visualization* (Harris et al., 2013; Wai et al., 2009) haben sich bereits als wichtige Prädiktoren für die Leistung in den Naturwissenschaften erwiesen. Das Novum dieser Studie ist jedoch, dass zahlreiche räumliche Faktoren im Zusammenhang mit dem deklarativen Fachwissen erhoben wurden und als Einflussgrößen auf die Leistung bei der Bearbeitung repräsentationsbasierter Chemieaufgaben in einem Pfadmodell operationalisiert werden. Die dargelegten Ergebnisse unterstützen die bisherige Forschung mit neuer und belastbarer Evidenz, dass für die Bearbeitung repräsen-

tationsbasierter Aufgaben neben dem Fachwissen, auch die räumlichen Faktoren *Spatial Relation*, *Speeded Rotation* und *Visualization* einen einzigartigen Anteil der Varianz aufklären und eine wichtige – wenn auch dem deklarativen Fachwissen (Vorwissen) nachgeordnete – Rolle bei der Bearbeitung spielen.

Die Definition der Repräsentationskompetenzen deuten bereits an, dass räumliche Fähigkeiten eine relevante Rolle für die Anwendung spielen. Bei der *Interpretation* müssen beispielsweise Muster erkannt werden und Vorgänge auf molekularer Ebene beschrieben werden. Für die *Translation* sind Perspektivwechsel relevant und für die *Konstruktion* die Gestaltung neuer Repräsentationen. Das zeigt deutlich, dass Repräsentationskompetenzen und räumliche Fähigkeiten Hand in Hand gehen (Gilbert, 2008; Harle & Towns, 2011). Gleichzeitig hebt es auch hervor, dass die lower-level skills unterschiedlich operationalisiert sind und somit verschiedene Faktoren für einzelnen Fähigkeiten variierende Bedeutung haben könnten. Forschungsanliegen I zeigt zwar, dass die drei lower-level skills empirisch nicht einwandfrei trennbar sind, dennoch erscheint es sinnvoll, die Pfadanalyse aus Abbildung 22 getrennt für die drei lower-level skills zu wiederholen, um zu prüfen, inwieweit sich unterschiedliche Prädiktorvariablen für die *Interpretation*, *Translation* und *Konstruktion* identifizieren lassen.

Die Ergebnisse in Abbildung 23 zeigen zunächst, dass der Einfluss des räumlichen Faktors *Spatial Relation* und des Fachwissens für alle drei lower-level skills signifikant wird. Gleichzeitig lassen sich dabei einige interessante Unterschiede mit Blick auf die Pfadkoeffizient erkennen. Das Fachwissen zeigt lediglich bei der *Konstruktion* einen hohen Effekt ($\beta = 0.533$), bei der *Interpretation* und *Translation* jedoch nur einen mittleren Effekt ($0.378 \leq \beta \leq 0.440$). Dieses Ergebnis lässt sich durch die Ansprüche der *Konstruktion* plausibel erklären. Bei der *Konstruktion* muss Wissen über Konventionen und chemische Konzepte einfließen, um beispielsweise auf Basis einer Summenformel eine Keilstrichformel zu konstruieren. Wissen zur Bindigkeit der einzelnen Atome, zur Anordnung der Atome und weitere Informationen werden dafür benötigt (Talanquer, 2022).

Bei den beiden anderen lower-level skills ist die Rolle des Fachwissens dagegen weniger gewichtig. Zwar wird auch für die *Interpretation* und die *Translation* Fachwissen benötigt, beispielsweise wenn nach einer Koordinationszahl gefragt wird, die Testperson muss aber den Begriff und das dahinterliegende

Konzept verstehen, um die Aufgabe adäquat lösen zu können. Prinzipiell basieren die Anforderung der *Interpretation* und *Translation* eher auf der Repräsentation selbst, d. h. für die *Interpretation* und *Translation* können in der Regel alle notwendigen Informationen aus der Repräsentation selbst entnommen werden, wie es bei der Identifikation von Mustern und Merkmalen (*Interpretation*) oder einem Perspektivwechsel (*Translation*) gefordert ist.

Bei der *Translation* wird im Rahmen dieser Arbeit deutlich, dass räumliche Fähigkeiten für diesen Prozess eine zentrale Rolle spielen, da hier der Pfadkoeffizient für die dreidimensionale mentale Rotation ($\beta = 0.400$) höher ist als der des Fachwissens ($\beta = 0.378$). Auch dieser Befund ist nicht überraschend, da Perspektivwechsel implizieren, dass ein Objekt rotiert werden muss bzw. auch eine Änderung der eigenen Person zum Objekt stattfinden kann. Letzteres kann durch den Test jedoch nicht abgedeckt werden und erfordert andere Messinstrumente, weshalb hier davon ausgegangen werden kann, dass die dreidimensionale mentale Rotation maßgeblich für den Erfolg bei Translationsleistungen beteiligt ist. Zusätzlich zeigt sich für die *Translation* die zweidimensionale mentale Rotation (*Speeded Rotation*) als signifikanter Prädiktor mit kleinem Effekt. Auch dieser Zusammenhang erscheint logisch, da Perspektivwechsel nicht immer eine dreidimensionale Rotation erfordern müssen.

Der Faktor *Visualization* – die Fähigkeit, Objekte mental zu manipulieren – wird für die *Interpretation* und die *Konstruktion* marginal mit einem kleinen Effekt ($0.143 \leq \beta \leq 0.166$) signifikant. Besonders bei der *Konstruktion* ist dies nicht überraschend, da neue, von der Referenzrepräsentation distinkte Repräsentationen geschaffen werden müssen, beispielsweise Isomere einer chemischen Verbindung, was neben entsprechendem Fachwissen auch die Fähigkeit zur Manipulation der Repräsentation voraussetzt. Auch bei der *Interpretation* lässt sich dies erklären, da eine Repräsentation mental manipuliert werden muss, um implizit dargestellte Informationen aus dieser abzuleiten, wobei das Fachwissen dabei ebenfalls von signifikanter Bedeutung ist.

Mit Blick auf die *Translation*, für die die mentale Manipulationsfähigkeit scheinbar keine signifikante Bedeutung spielt, lassen sich die Ergebnisse ebenso logisch einordnen: Bei der *Translation* wird die Ausgangsrepräsentation nicht verändert, lediglich der Repräsentationstyp, wobei dieser einen vergleichbaren Abstraktionsgrad hat und somit in Bezug auf explizit und implizit dargestellte Informationen vergleichbar ist. Auch für Perspektivwechsel spielt

die mentale Manipulation offenbar eine vernachlässigbare Rolle, was sich aus dem angepassten Definitor der *Translation* ableiten lässt (Tabelle 3). Für die *Translation* sind somit eher mentale Rotationsprozesse von zentraler Bedeutung. Diese Ergebnisse sind spannend und für den aktuellen Diskurs zur Frage nach der Trennbarkeit der Faktoren *Visualization* und *Spatial Relation* interessant (Buckley et al., 2018; Merchant et al., 2013). Die hier vorliegenden Ergebnisse sprechen eher für eine getrennte Betrachtung der beiden Faktoren. Für eine belastbare Aussage sind jedoch weitere faktoranalytische Untersuchungen notwendig.

Zusammenfassend lässt sich sagen, dass die *Interpretation* und *Konstruktion* in dieser Studie die gleichen Prädiktoren aufweisen. Die räumlichen Faktoren *Visualization* und *Spatial Relation* haben auf beide lower-level skills einen vergleichbaren, kleinen Effekt. Das Fachwissen spielt für beide Fähigkeiten eine maßgebliche Rolle, zeigt bei der *Konstruktion* jedoch einen großen und bei der *Interpretation* nur einen mittleren Effekt. Die *Translation* zeigt ein anderes Muster. Hier spielen sowohl zwei- als auch die dreidimensionale Rotationsfähigkeiten eine Rolle, der Faktor *Visualization* jedoch nicht. Besonders auffällig ist, dass dreidimensionale mentale Rotationsfähigkeiten einen (leicht) größeren Effekt auf die *Translation* hat als das Fachwissen. Hypothese 2.2 kann somit nur teilweise bestätigt werden, sowohl für das einfachere als auch für das komplexere Pfadmodell.

Die Ergebnisse des in Abbildung 23 dargestellten Pfadmodells zeigen, dass die *Translation* im Vergleich zu den anderen beiden lower-level skills andere Prädiktoren in den räumlichen Fähigkeiten aufweist. Dadurch wird die Frage eröffnet, ob die *Interpretation* und *Konstruktion* näher verwandte Prozesse sind als bisher angenommen und sich daraus ein bisher nicht geprüftes zweidimensionales Modell ergibt, das die *Translation* der *Interpretation* und *Konstruktion* gegenüberstellt. Auch könnten diese Ergebnisse die Hypothese unterstützen, dass die *Interpretation* die Voraussetzung für die *Konstruktion* ist (Talanquer, 2022; Ward et al., 2025). Basierend auf den hier dargestellten Ergebnissen lassen sich auf diese Fragen jedoch keine Antworten ableiten und erfordern weitere Forschung. Besonders mit Blick auf die hierarchische Strukturierung der Repräsentationskompetenzen sind weiterführende Studien notwendig.

Auch wenn sich hier bereits ein Einblick in die statistischen Zusammenhänge zwischen Fachwissen, Repräsentationskompetenzen und räumlichen Fähigkeiten gewinnen lässt, ist die Anreicherung dieser Ergebnisse mit qualitativen Methoden und Daten sinnvoll. Dadurch kann ein Einblick in zugrundeliegende mentale Prozesse und Wirkmechanismen zwischen den Konstrukten aufgedeckt werden, auf deren Basis die quantitativen Ergebnisse geprüft und tiefergreifender diskutiert werden können. Das im folgenden Kapitel dargestellte Forschungsanliegen III widmet sich diesem Ziel und stellt die Planung, Durchführung und Ergebnisse einer qualitativen Datenerhebung zur differenzierten und tiefergreifenden Untersuchung der lower-level skills mit Fokus auf Hindernissen bei der Bearbeitung repräsentationsbasierter Aufgaben dar.

7

Forschungsanliegen III – Hindernisse beim Bearbeiten repräsentationsbasierter Chemieaufgaben

Forschungsanliegen II liefert Einblicke in die Zusammenhänge zwischen räumlichen Fähigkeiten und Repräsentationskompetenzen sowie dem deklarativen Fachwissen. Forschungsanliegen III zielt darauf ab, ein tiefergehendes Verständnis dieser Ergebnisse zu gewinnen, indem essenzielle kognitive Prozesse und Fähigkeiten für das erfolgreiche Bearbeiten repräsentationsbasierter Chemieaufgaben über die Identifikation von Hindernissen beim Bearbeiten dieser Aufgaben untersucht werden.

7.1 Design

Für das dritte Forschungsanliegen kam eine *qualitative Laborstudie* mit *nicht-experimentellem, explorativem Querschnittsdesign* zum Einsatz. Die Forschungsfrage wurde mithilfe umstrukturierter Datenerhebungsmethoden adressiert. Die Datenerhebung erfolgte unter kontrollierten Bedingungen, so dass Störvariablen ausgeschlossen bzw. kontrolliert werden konnten. Es wurde auf eine konkrete Population an Testpersonen zurückgegriffen, ohne gezielte Variationen in der Ausprägung bestimmter Variablen vorzunehmen. Theoriegeleitet angelegt, erfolgte die Auswertung der Daten nicht hypothesenbasiert, sondern explorativ (Döring, 2023f).

Ziel war es, Einblicke in die kognitiven Prozesse und Fähigkeiten der Testpersonen bei der Bearbeitung repräsentationsbasierter Chemieaufgaben zu gewinnen. Dabei sollte insbesondere untersucht werden, inwieweit sich diese Prozesse und Fähigkeiten bei der *Interpretation*, *Translation* und *Konstruktion* voneinander unterscheiden. Von zentralem Interesse waren dabei kognitive Prozesse und Fähigkeiten, die für die erfolgreiche Bearbeitung dieser Aufgaben notwendig sind. Dadurch sollten die Ergebnisse des zweiten Forschungsanliegens mit zusätzlicher Evidenz angereichert und die Wechselbeziehungen

und Zusammenhänge zwischen Repräsentationskompetenzen und räumlichen Fähigkeiten sowie dem Fachwissen tiefergehend untersucht werden. Das Anliegen wurde in dieser Studie durch die Identifikation von Hindernissen im Bearbeitungsprozess adressiert, d. h. indem nicht direkt die Prozesse betrachtet wurden, die zum erfolgreichen Lösen der repräsentationsbasierten Aufgaben führen, sondern Stellen im Bearbeitungsprozess, die das erfolgreiche Lösen der Aufgaben erschwerten oder verhinderten. Dieses Vorgehen hatte den Vorteil, dass Hindernisse besser und klarer operationalisiert und somit auch besser identifiziert werden konnten. Zudem ließen sich dadurch systematisch kognitive Engpässe der Testpersonen im Bearbeitungsprozess identifizieren. So konnte auf kognitive Prozesse und Fähigkeiten geschlossen werden, deren Ausprägung für den Lösungserfolg eine Schlüsselrolle spielten.

Für diese Untersuchung wurde ein Testheft mit Aufgaben für die *Interpretation*, *Translation* und *Konstruktion* (Tabelle 3) entwickelt. Im Fokus stand die Komplexchemie, die ähnlich wie die Organische Chemie eine sehr visuell-räumlich geprägte Domäne innerhalb der Chemie ist (Dai et al., 2020; Eticha et al., 2023). Dieses Testheft wurde von Testpersonen mit Lautem Denken bearbeitet, wodurch simultan zur Bearbeitung Einblicke in die Hindernisse bei der Bearbeitung sowie die kognitiven Prozesse und Fähigkeiten der Testpersonen gewonnen werden konnten. Die Protokolle des Lauten Denkens wurden nach der inhaltlich-semantischen Transkription basierend auf der qualitativen Inhaltsanalyse (Mayring & Fenzl, 2022) ausgewertet.

7.2 Methodisches Vorgehen und Datenauswertung

7.2.1 Hindernisse im Bearbeitungsprozess

Die Identifikation von Hindernissen im Bearbeitungsprozess hat, wie bereits beschrieben, den Vorteil, dass dadurch systematisch Stellen identifiziert werden können, an denen sich die Ausprägung einer Fähigkeit die Lösung einer Aufgabe erschwert oder verhindert. Dadurch kann direkt darauf geschlossen werden, welche Fähigkeiten für die Lösung einer Aufgabe eine zentrale Rolle spielen. Um Hindernisse zu identifizieren, sind sie zunächst eindeutig zu operationalisieren. Dafür werden im Folgenden zunächst einige wichtige Begriffe in diesem Kontext definiert und voneinander abgegrenzt.

Zunächst muss der Begriff der *Aufgabe* näher beleuchtet werden. Eine solche kann als kognitive Anforderung verstanden werden, bei der eine gegebene Situation (Ist-Zustand) mithilfe bekannter Methoden in eine gewünschte Situation (Soll-Zustand) transformiert werden soll. Tritt dabei eine Schwierigkeit auf, die das Erreichen der gewünschten Situation verhindert, wird von einem *Problem* gesprochen (Dörner, 1987; Tobinski, 2017). *Aufgabe* und *Problem* unterscheiden sich dementsprechend lediglich im Vorhandensein eines Hindernisses, das die Transformation verhindert. Der Begriff *Aufgabe* sollte daher durch den Begriff *Routineaufgaben* abgelöst werden, da hier bekannte Methoden zur Transformation routiniert genutzt werden. *Probleme* sind dementsprechend als *Problemaufgaben* zu bezeichnen, wodurch *Aufgabe* als Oberbegriff für *Routine-* und *Problemaufgaben* zu verstehen ist (Lange, 2013; Pehkonen et al., 2013). Daraus lässt sich ableiten, dass der Aufgabenbegriff interindividuell variabel ist, d. h. eine Aufgabe kann für eine Person eine *Routineaufgabe* darstellen, während die gleiche Aufgabe für eine andere Person eine *Problemaufgabe* ist (Lange, 2013; Pehkonen et al., 2013; Tobinski, 2017).

An dieser Stelle stellt sich die Frage, was genau ein Hindernis ist und ab wann von einem solchen gesprochen werden kann. Zunächst fällt auf, dass im deutschsprachigen Raum je nach Quelle in diesem Zusammenhang unterschiedliche Begriffe genutzt werden, die jedoch als Synonyme zu verstehen sind: Hindernis (Göhner & Krell, 2021), Barriere (Dörner, 1987; Vollrath, 1992) und Schwierigkeit (Vollrath, 1992). In dieser Arbeit wird konsequent der Begriff *Hindernis* verwendet.

Nach Dörner (1987) tritt ein Hindernis auf, wenn die Transformation vom Ist- in den Soll-Zustand aus irgendwelchen Gründen nicht gelingt (Dörner, 1987). Diese Definition wird von Pehkonen et al. (2013) erweitert, indem ein Problem durch die Kombination zusätzlicher Informationen in einem neuen Kontext ergänzt wird. Lange (2013) hat verschiedene Definitionen und Beschreibungen von Hindernissen verglichen und in eine umfassende Definition synthetisiert, die beispielsweise auch bei Göhner und Krell (2021) Grundlage für die Identifikation von Hindernissen im Modellierungsprozess in den Naturwissenschaft ist. Ein Hindernis ist demnach: „Eine Stelle in einem Bearbeitungsprozess, in der rekonstruierbar ist, dass eine Person nichts oder etwas nicht selbstverständlich (im Sinne von nicht sicher, zweifelnd) ausführt und dabei auf nichts in der Aufgabensituation Anwendbares zurückgreifen möchte bzw. zurückgreifen kann [...]“ (Lange, 2013, S. 32).

7.2.2 Entwicklung des Materials

Mit Blick auf das Bearbeiten repräsentationsbasierter Aufgaben in den Naturwissenschaften und speziell in der Chemie zeigt sich, dass Repräsentationskompetenzen, Fachwissen und räumliche Fähigkeiten von zentraler Bedeutung für das erfolgreiche Lösen solcher Aufgaben sind. Hier sollen die *lower-level skills* der Repräsentationskompetenzen näher untersucht werden. Die bisherigen Daten haben gezeigt, dass diese eher als stark interkorrelierte Fähigkeiten eines eindimensionalen Konstrukts zu verstehen sind. Dennoch gibt es Hinweise, dass sich *Interpretation*, *Translation* und *Konstruktion* hinsichtlich der notwendigen kognitiven Prozesse und Fähigkeiten unterscheiden. Daher sollen Hindernisse bei der Bearbeitung von Aufgaben zu den drei *lower-level skills* identifiziert und erörtert werden, um ein tiefergreifendes Verständnis für die kognitiven Prozesse und die nötigen Fähigkeiten für die Anwendung der drei Fähigkeiten zu erlangen. Die Theorie (Kapitel 2 und 3) sowie die Ergebnisse der Forschungsanliegen I und II unterstreichen bereits, dass das deklarative Fachwissen (Vorwissen) maßgeblich zur Umsetzung der *lower-level skills* beiträgt (z. B. Cook et al., 2008; Edelsbrunner et al., 2023; Nitz, Ainsworth, et al., 2014; Rau, 2018; Talanquer, 2022) und auch die räumlichen Faktoren *Speeded Rotation*, *Spatial Relation* und *Visualization* (sowie *Flexibility of Closure*) signifikante Prädiktoren für die Leistung bei der Bearbeitung repräsentationsbasierter Chemieaufgaben sind (z. B. Bodner & Guay, 1997; Bodner & McMillen, 1986; Carter et al., 1987; Harle & Towns, 2011; Harris et al., 2013; Hodgkiss et al., 2018; Sorby, 2009; Wai et al., 2009; Wu & Shah, 2004).

Um das Forschungsanliegen zu untersuchen, mussten die Inhalte zunächst eingegrenzt werden, sodass die Aufgaben und die Ergebnisse vergleichbar waren. Aus diesem Grund eigneten sich die Items des CRI:TIC nicht für dieses Anliegen, da sie durch eine hohe Bandbreite an unterschiedlichen Fachinhalten und Repräsentationen keinen validen Vergleich ermöglichten. Daher musste neues Aufgabenmaterial entwickelt werden, das einen Vergleich der *Interpretation*, *Translation* und *Konstruktion* durch gleiche Fachinhalte und Repräsentationen zuließ. Mit Blick auf die bisherige Forschung zeigt sich, dass sich diese vornehmlich auf die Organische Chemie fokussierte, da sie durch eine enorme Bandbreite an unterschiedlichen Repräsentationen gekennzeichnet ist, die zudem ein hohes Maß an räumlichen Fähigkeiten erfordern (Gurung et al., 2022; Pribyl & Bodner, 1987; Stieff et al., 2018; Ward et al., 2022, 2025). Im Vergleich dazu wurde die Rolle von Repräsentationen beim Lernen

im Feld der Komplexchemie bislang weniger umfassend untersucht, obwohl sie durch vielfältige räumliche Strukturen (z. B. tetraedrisch, quadratisch-planar, oktaedrisch), unterschiedliche Repräsentationen und eine ausgeprägte Isomerie gekennzeichnet ist (Dai et al., 2020; Eticha et al., 2023) und somit vergleichbare Voraussetzungen zur Untersuchung des Forschungsanliegens bietet (Abbildung 24). Daher steht die Komplexchemie im Fokus des dritten Forschungsanliegens.

Exkurs: Grundbegriffe der Komplexchemie

Die Komplexchemie ist ein Teilbereich der Anorganischen Chemie. Komplexverbindungen sind aus einem zentralen Koordinationszentrum und einer umgebenden Ligandenhülle aufgebaut. Das Koordinationszentrum kann hier ein ungeladenes Zentralatom oder ein geladenes Zentralion sein. Die Liganden sind Moleküle oder Ionen, die häufig über mindestens ein freies Elektronenpaar verfügen, über das sie an das Zentralatom/-ion koordinieren. Die Bindung wurde daher früher als koordinative oder dative Bindung bezeichnet und von der klassischen kovalenten Bindung unterschieden, bei der beide Bindungspartner die Bindungselektronen bereitstellen. Heute sind jedoch viele Komplexe bekannt, bei denen die Bindungselektronen vom Zentralatom/-ion oder beiden Partnern stammen, weshalb diese Differenzierung nicht mehr vorgenommen und auch in Komplexen von kovalenten Bindungen gesprochen wird (Binnewies et al., 2016; Riedel & Janiak, 2022).

Liganden können mehrere Koordinationsstellen besitzen, wobei von Zähigkeit gesprochen wird. Ethylendiamin ($\text{H}_2\text{N}-\text{CH}_2-\text{CH}_2-\text{NH}_2$) ist ein Beispiel für einen zweizähligen Liganden, der über die freien Elektronenpaare der beiden Stickstoffatome (N) an ein Zentralatom/-ion koordinieren kann. Die Anzahl der direkt an das Koordinationszentrum gebundenen Atome wird als Koordinationszahl bezeichnet. Aus der Koordinationszahl kann auf die Struktur geschlossen werden, d. h. die räumliche Anordnung der Liganden um das Koordinationszentrum. So sind Komplexe mit der Koordinationszahl zwei linear, solche mit der Koordinationszahl vier tetraedrisch oder quadratisch-planar und solche mit der Koordinationszahl sechs oktaedrisch (Abbildung 24) aufgebaut (Binnewies et al., 2016; Finze & Radius, 2014b, 2014a; Radius & Finze, 2014a; Riedel & Janiak, 2022).

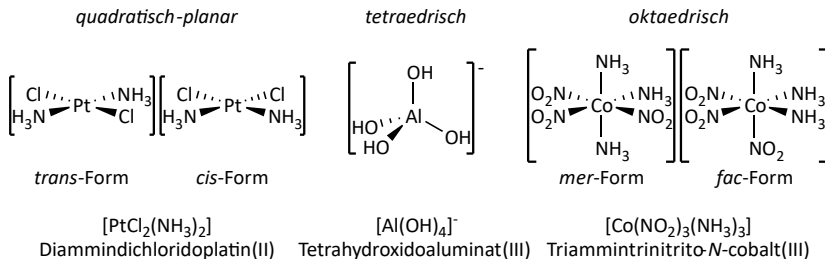


Abbildung 24: Beispiele für Komplexe mit unterschiedlicher Anzahl an Liganden und entsprechend unterschiedlichen Strukturen und Isomeren (erstellt mit ChemDraw Professional, Version 23.1.2.7).

Neben der Beschränkung auf den Fachbereich Komplexchemie war eine Begrenzung der betrachteten Repräsentationen erforderlich. Zwar ist die Anzahl der unterschiedlichen Repräsentationen überschaubarer als in der Organischen Chemie, dennoch erschien es sinnvoll, sie hier auf zwei unterschiedliche Varianten zu begrenzen. Besonders mit Blick auf die Testökonomie und die Zumutbarkeit für die Testpersonen war diese Eingrenzung angebracht. Zudem gewährleistete sie die Vergleichbarkeit zwischen den Aufgaben zur *Interpretation*, *Translation* und *Konstruktion*.

Um authentische Aufgaben und relevante Repräsentationen zu identifizieren, wurden zunächst Modulhandbücher für den Studiengang *B. Sc. Chemie* an verschiedenen Universitäten in Deutschland hinsichtlich Literaturempfehlungen für die Anorganische Chemie bzw. die Komplexchemie (sofern explizit aufgeführt) gesichtet. Die Kapitel zur Komplexchemie wurden anschließend hinsichtlich des Einsatzes visuell-graphischer Repräsentationen untersucht (Tabelle 21). Symbolisch-mathematische Repräsentationen wurden hier nur bei den Aufgaben zur *Konstruktion* als Grundlage für die *Konstruktion* von Repräsentationen berücksichtigt, da diese einen vermeintlich geringeren Bezug zu räumlichen Fähigkeiten aufweisen.

In den Lehrwerken und Kapiteln zur Komplexchemie kommen vornehmlich Keilstrichformeln zur Darstellung von Komplexen zum Einsatz. Ebenso finden Kugel-Stab-Modelle häufig Verwendung oder eine hybride Form aus beiden

Repräsentationen. Daher wurden in diesem Forschungsanliegen Keilstrichformeln und Kugel-Stab-Modelle eingesetzt. Diese Wahl bietet zudem den Vorteil, dass Kugel-Stab-Modelle die Dreidimensionalität des Komplexes direkt darstellen, indem Kugeln derselben Atomart im Vordergrund größer als im Hintergrund dargestellt erscheinen und Objekte im Vordergrund jene im Hintergrund überlagern. Dadurch lässt sich die dreidimensionale Struktur direkt aus ihnen ablesen, auch wenn sie auf einer zweidimensionalen Papierebene gedruckt sind. Anders verhält es sich bei den Keilstrichformeln. Hierbei handelt es sich um zweidimensionale Strukturen, die die Dreidimensionalität der Komplexe durch Striche, gefüllte und gestrichelte Keile andeuten. Die Dreidimensionalität ist daraus nur abzulesen, wenn die Konventionen der Keilstrichformel bekannt sind und interpretiert werden können (Stieff et al., 2016, 2018).

Zur Konstruktion der Aufgaben wurden die Messinstrumente, die als Orientierung für die Entwicklung des CRI:TIC gedient haben und aus denen Items adaptiert wurden, als Grundlage verwendet (Averbeck, 2021; Sim & Daniel, 2014; Taskin et al., 2015). Ferner kamen die Studien von Stieff und Kolleg:innen (2018) und Ward und Kolleg:innen (2022) als Orientierung zur Konstruktion der Aufgaben zum Einsatz. Stieff et al. (2018) untersuchten den Zusammenhang zwischen räumlichen Fähigkeiten (Rotationsfähigkeiten) und Repräsentationskompetenzen mithilfe von Keilstrichformeln und Kugel-Stab-Modellen, während Ward et al. (2022) basierend auf dem Faktormodell zur Interpretation von externen Repräsentationen (Schönborn & Anderson, 2009) die kognitiven Prozesse (*reasoning*) von Studierenden bei der *Interpretation*, *Translation*, *Konstruktion* und *Nutzung* von Keilstrichformeln und Newman-Projektionen in der Organischen Chemie untersuchten. Darauf basierend wurden für diese Studie Aufgaben zur *Interpretation*, *Translation* und *Konstruktion* von Keilstrichformeln und Kugel-Stab-Modellen in der Komplexchemie konstruiert. Diese werden im Folgenden kurz beschrieben. Das vollständige Testheft mit den Aufgaben (Anhang C.1) und das entsprechende Manual zur Auswertung (Anhang C.3) sind dem Anhang zu entnehmen.

Tabelle 21: Übersicht der gesichteten Lehrbücher hinsichtlich des Einsatzes visuell-graphischer Repräsentationen zur Darstellung von Komplexen.

Autor:innen (Jahr)	Buch	Buch-kapitel	Titel	Seiten
Binnewies et al. (2016)	☐	☒	Komplexreaktionen	323-352
Elschenbroich (2008)	☒	☐	Organometallchemie	
Finze & Radius (2014a)	☐	☒	Koordinationschemie: Struktur und Bindung	439-542
Finze & Radius (2014b)	☐	☒	Strukturen von Koordinationsverbindungen	625-684
Gade (1998)	☒	☐	Koordinationschemie	
Hollemann et al. (2016a)	☐	☒	Der Molekülbau. Die chemische Bindung, Teil I	121-205
Hollemann et al. (2016b)	☐	☒	Grundlagen der Komplexchemie	1550-1647
Janiak (2018)	☐	☒	Komplex-/Koordinationschemie	409-624
Mortimer & Müller (2020)	☐	☒	Komplex-Verbindungen	479-496
Radius & Finze (2014a)	☐	☒	Charakterisierung von Koordinationsverbindungen	543-624
Radius & Finze (2014b)	☐	☒	Reaktionen von Koordinationsverbindungen: Kinetik und Mechanismen	685-756
Riedel & Janiak (2022)	☐	☒	Komplexverbindungen	729-762
Schweda (2022a)	☐	☒	Komplexchemie	119-140
	☐	☒	Chemie der Chelatliganden	141-154
Schweda (2022b)	☐	☒	Komplexverbindungstitrationen	93-106
	☐	☒	Komplexverbindungen	352-358
Steudel (1973)	☐	☒	Theorie der chemischen Bindung	61-174
Strähle & Schweda (2006)	☐	☒	Komplexchemie	113-130
Weber (2021)	☒	☐	Koordinationschemie - Grundlagen und aktuelle Trends	

Interpretation – Hier sollten die Testpersonen Informationen aus Keilstrichformeln und Kugel-Stab-Modellen ableiten und diese analysieren. Beispielsweise ging es darum, Zentralatome/-ionen, die Koordinationszahl, die Art und Anzahl der Liganden, ihre Zähigkeit, die Anzahl an Atomen und den Bindungstyp innerhalb des Komplexes und der Molekülliganden präzise anzugeben. Dabei wurde darauf geachtet, dass die dargestellten Komplexe nicht identisch, aber beispielsweise durch eine gleiche Raumstruktur vergleichbar sind. Des Weiteren wurden keine Komplexe mit speziellen, sondern eher „klassischen“ Liganden (H_2O , NH_3 , OH^- , Cl^- usw.) verwendet.

Translation – Zunächst sollten die Testpersonen Kugel-Stab-Modelle mit räumlichen Strukturen in eine entsprechende Keilstrichformel übersetzen und diese skizzieren. Eine umgekehrte Übersetzung war nicht zielführend und nicht authentisch. Da die Repräsentationstypen einen vergleichbaren Abstraktionsgrad aufweisen, kann dieser Prozess hier eindeutig der *Translation* zugeordnet und von der *Konstruktion* abgegrenzt werden. Anschließend sollten die Testpersonen Komplexe vergleichen und angeben, ob die dargestellten Strukturen identisch oder Isomere sind. Zusätzlich sollte die Art der Isomerie bestimmt werden, wobei diese Aufgabe mit einem Tippkärtchen unterstützt wurde, auf dem die unterschiedlichen Arten der Isomerie (Konstitutionsisomerie, Konfigurationsisomerie, Konformationsisomerie usw.) beschrieben und mit eher visuell-graphischen Beispielen aus der Organischen Chemie dargestellt wurden. Die Aufgaben wurden so gestaltet, dass die Testpersonen gleiche Repräsentationstypen (Keilstrichformel mit Keilstrichformel, Kugel-Stab-Modell mit Kugel-Stab-Modell) und unterschiedliche Repräsentationstypen miteinander vergleichen müssen (Keilstrichformel mit Kugel-Stab-Modell und umgekehrt). Insgesamt gab es dazu drei Aufgaben, wobei die ersten beiden ausschließlich Komplexe mit einzähnigen Liganden und in der dritten Aufgabe zusätzlich zweizähnige Liganden beinhalteten. Dadurch konnten auch Enantiomere (Spiegelbilder, die durch Rotation nicht ineinander überführt werden können) bei Komplexen einbezogen werden.

Konstruktion – Hier wurden auch Summenformeln in den Aufgaben verwendet. Diese sollten zunächst interpretiert werden, indem das Zentralatom/-ion sowie die Art und Anzahl der Liganden angegeben werden (Interpretationsaufgabe). Anschließend sollten die Komplexe mit allen theoretisch möglichen Konfigurationsisomeren (Diastereomere, d. h. *cis-trans*- und *mer-fac*-Isomere (Abbildung 24) und Enantiomere) als Keilstrichformeln dargestellt werden.

Dieser Prozess kann eindeutig der *Konstruktion* zugeordnet werden, da der Abstraktionsgrad von Summenformel (mathematisch-symbolisch) zu Keilstrichformel (eher visuell-graphisch) sehr unterschiedlich ist. Insgesamt wurden vier unterschiedliche Komplexe als Summenformel mit entsprechender Raumstruktur angegeben. Die räumliche Struktur musste an dieser Stelle vorgegeben werden, da bei Komplexen mit der Koordinationszahl vier sowohl quadratisch-planare als auch tetraedrische Komplexe in Frage kommen. Zur Klärung der Struktur muss dabei die Ligandenfeldtheorie herangezogen werden, die hier aufgrund der Komplexität und des Ziels der Studie nicht aufgegriffen wird.

Reflexionen – Nach jedem Aufgabenblock wurde eine kurze Reflexion durchgeführt, sodass die Testpersonen nachträglich die Schwierigkeit der Aufgaben diskutieren und Möglichkeiten zur Unterstützung angeben konnten, die ihnen bei der Bearbeitung geholfen hätten.

Das entwickelte Testheft wurde mit drei wissenschaftlichen Hilfskräften pilotiert. Diese sind in Bezug auf Alter, Vorwissen und Semester mit der Stichprobe vergleichbar (Kapitel 7.2.5). Die Pilotierung fand unter den gleichen Bedingungen statt wie die Hauptdatenerhebung. Auf Basis der Ergebnisse und eines Evaluationsgesprächs mit den wissenschaftlichen Hilfskräften wurden Unklarheiten identifiziert und die Aufgaben entsprechend präzisiert und angepasst.

7.2.3 Lautes Denken

Um einen Einblick in kognitive Prozesse und Fähigkeiten zu erhalten, die bei der Bearbeitung repräsentationsbasierter Chemieaufgaben notwendig sind, eignet sich besonders die Methode des Lauten Denkens. Diese Methode stammt aus der Kognitionspsychologie und ermöglicht die Analyse von Denk- und Problemlöseprozessen (Sandmann, 2014). Wie die Bezeichnung Lautes Denken vermuten lässt, soll die Testperson alle Gedanken laut aussprechen, die während einer bestimmten Aktivität aufkommen. Dadurch wird ein Einblick in die kognitiven Vorgänge während dieser Aktivität gewonnen, die durch kaum eine andere Methode möglich ist. Sie kann den unstrukturierten Interviews zugeordnet werden, da die Datenerhebung in einem offenen Format

stattfindet und nicht durch gezielte Fragen oder einen Interviewleitfaden gesteuert wird. Vielmehr wird es der Testperson ermöglicht, das Gespräch zu gestalten und auf für sie zentrale Aspekte einzugehen.

Das Laute Denken tritt im natürlichen Umfeld beispielsweise auf, wenn eine Person sehr konzentriert an etwas arbeitet (Bilandzic, 2017; Döring, 2023c; Sandmann, 2014; Schuttkowski & Staubach, 2022). In der Laborumgebung wird dieser Prozess jedoch forciert und erscheint unnatürlich, wobei sich die Frage stellt, ob und inwieweit das Laute Denken die Ausführung einer Aktivität und die zugrundeliegenden kognitiven Prozesse beeinflusst. Hier liefern Fox et al. (2011) mit den Ergebnissen ihrer Metastudie Evidenz, dass Lautes Denken keine Verzerrung der kognitiven Prozesse oder der Aufgabenbearbeitung bewirkt, was die wissenschaftliche Fundierung dieser Methode weiter stärkt.

Für das Laute Denken können verschiedene Formen differenziert werden. Bei der *Introspektion* (auch *periaktionales Lautes Denken* genannt) werden die Gedanken zeitgleich zur Aktivität verbalisiert, wobei dies maßgeblich Inhalte des Kurzzeitgedächtnisses betrifft. Die *Retrospektion* (oder *postaktionales Lautes Denken*) wird unmittelbar an die Bearbeitung der Aufgaben angeschlossen oder verzögert, teilweise erst einige Tage nach der eigentlichen Aktivität durchgeführt und zielt darauf ab, dass die Testperson ihre Gedankenschritte interpretiert und erklärt (Bilandzic, 2017; Döring, 2023c; Konrad, 2010; Sandmann, 2014). Im Rahmen dieser Studie wird eine Kombination aus *Introspektion* und *Retrospektion* verwendet. Während der Bearbeitung der drei Aufgaben (*Interpretation*, *Translation* und *Konstruktion*) verbalisieren die Testpersonen sämtliche Gedanken zu den Aufgaben und ihrer Bearbeitung. Im Anschluss sollen die Testpersonen ihre eigenen Gedanken und das Vorgehen mit Fokus auf selbst wahrgenommene Hindernisse bei der Bearbeitung reflektieren.

Das Laute Denken wurde in der Studie mit zwei unabhängigen Tools audio- und videographisch aufgezeichnet. Die Audiodatei diente dabei als Sicherungskopie. Die Videos wurden so aufgenommen, dass das Testheft und der Oberkörper, ohne das Gesicht, zu sehen waren. Die Videographie war im Kontext dieser Studie sinnvoll und relevant, da Menschen dazu neigen, ihre kognitiven Prozesse mit Gesten zu unterstützen, insbesondere bei mentalen Manipulations- und Rotationsprozessen. Hier können Studien belegen, dass vor allem Menschen mit schwach ausgeprägten räumlichen Fähigkeiten häufiger

Gesten zur Unterstützung der mentalen Prozesse nutzen als Personen mit gut ausgeprägten räumlichen Fähigkeiten (Göksun et al., 2013; Goldin-Meadow et al., 2012; Hostetter & Alibali, 2008).

Die Videos wurden anschließend mithilfe eines *semantisch-inhaltlichen Systems* transkribiert, das sich durch seine Einfachheit schnell erlernen ließ. Faktoren wie die Prosodie, d. h. Tonhöhe, Lautstärke und Redegeschwindigkeit waren für das Forschungsanliegen nicht relevant, weshalb diese Aspekte hier nicht transkribiert wurden und auf komplexe Transkriptionssysteme verzichtet wurde. Bei der semantisch-inhaltlichen Transkription wird wörtlich transkribiert, wobei beispielsweise Wortverschleifungen geglättet (z. B. „hamma“ zu „haben wir“), Dialekte ins Hochdeutsche übersetzt und die Interpunktion zugunsten der Lesbarkeit angepasst werden (Dresing & Pehl, 2018). Zusätzlich werden Gesten an entsprechenden Stellen im Transkript vermerkt. Zur Transkription wurde MAXQDA Analytics Pro (Version 24) verwendet. Eine ausführliche Anleitung zur semantisch-inhaltlichen Transkription, die im Rahmen dieses Projekts nach Dresing und Pehl (2018) angepasst und angewendet wurde, kann dem Anhang entnommen werden (Anhang C.4).

7.2.4 Qualitative Inhaltsanalyse nach Mayring

Die *Qualitative Inhaltsanalyse* stellt eine Möglichkeit zur Auswertung von Texten dar. Sie ist unter den bibliographischen Auswertungsmethoden die am häufigsten verwendete Methode und ermöglicht die Analyse großer Datenmengen. Die Qualitative Inhaltsanalyse ist kategorien- und streng regelgeleitet, was sie von anderen Methoden wie der eher explorativen *Grounded Theory* (z. B. Strübing, 2022) oder der *Ethnographie* (z. B. Knoblauch & Vollmer, 2022) abgrenzt. Das Material wird mithilfe eines Kategoriensystems (Summe aller Analyseaspekte/Kategorien) analysiert und Textstellen diesen Kategorien zugeordnet. Anschließend können die Kategorien quantifiziert werden. Dies ermöglicht eine Analyse darüber, ob und in welchem Umfang bestimmte Kategorien mehrfach zugeordnet werden können. Daher ist die Bezeichnung „*Qualitative Inhaltsanalyse*“ nicht vollständig adäquat und eine eindeutige Zuordnung zu den qualitativen Methoden nicht korrekt. „*Qualitativ orientierte kategoriengeleitete Textanalyse*“ ist eine bessere Bezeichnung für die Methode (Döring, 2023c; Göhner & Krell, 2020; Mayring & Fenzl, 2022).

Neben einem streng regelgeleiteten Ablaufmodell (Abbildung 25) für die *qualitativ orientierte kategoriengeleitete Textanalyse* zeichnet sich die Systematik durch die Definition von *Analyseeinheiten* aus. Die *Auswertungseinheit* gibt die Materialportion an, auf die ein Kategoriensystem angewendet wird. Mit der *Kontexteinheit* werden die maximalen Testbestandteile bestimmt, die in eine Kategorie fallen dürfen. Die *Kodiereinheit* legt den minimalen Textbestandteil fest, der kodiert werden darf. Durch die Einhaltung dieses Vorgehens kann die inhaltsanalytische Güte mithilfe der Interdecoderübereinstimmung geprüft und sichergestellt werden. Dabei wird die Übereinstimmung der Kategorisierung zwischen unabhängigen Kodierenden bestimmt (Göhner & Krell, 2020; Krebs & Menold, 2022; Mayring & Fenzl, 2022). Hier wird diese Übereinstimmung durch Kategorienüberlappung ermittelt, d. h. ob zwei mit derselben Kategorie kodierten Textabschnitte von unterschiedlichen Kodierenden sich zu einem bestimmten Mindestanteil überlappen. Daraus lässt sich der Übereinstimmungskoeffizient Kappa (κ) nach Brennan und Prediger (1981) ermitteln. Dieser Koeffizient nimmt in der Regel Werte zwischen 0 und 1 an, wobei 1 für eine perfekte Übereinstimmung zwischen den Kodierenden spricht und 0 für eine zufällige Übereinstimmung. Werte von $\kappa \geq 0.60$ werden als substantiell angesehen (Landis & Koch, 1977).

Durch das systematische und regeleleitetete Vorgehen ist die *qualitativ orientierte kategoriengeleitete Textanalyse* gut nachzuvollziehen und eine sehr transparente Methode, die sich im Rahmen dieses Projekts ideal zur Identifikation von Hindernissen bei der Bearbeitung repräsentationsbasierter Chemieaufgaben eignet. Auch durch die Kombination aus theoriegeleitet deduktivem und eher explorativ induktivem Vorgehen bietet sich die Methode hier ideal an.

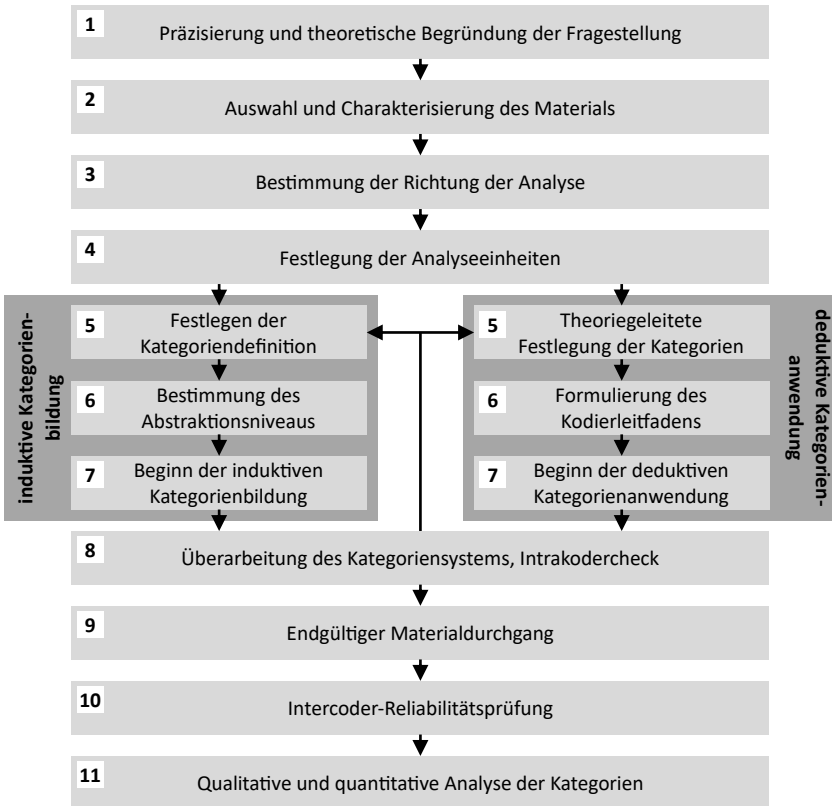


Abbildung 25: Ablauf der qualitativen Inhaltsanalyse mit induktiver Kategorienbildung (links) bzw. deduktiver Kategorienanwendung (rechts), (adaptiert nach Mayring & Fenzl, 2022, S. 699).

7.2.5 Umsetzung der qualitativ orientierten kategoriengeleiteten Textanalyse in dieser Studie (qualitative Inhaltsanalyse)

Zur Beantwortung der Forschungsfragen, die dieses dritte Forschungsanliegen leiten (Kapitel 4), wurde theoriegeleitet und auf Basis vergleichbarer Studien in diesem Kontext Aufgabenmaterial entwickelt, das sich zur Untersuchung des Forschungsanliegens eignete. Dabei traten die drei lower-level skills *Inter-*

pretation, *Translation* und *Konstruktion* in den Vordergrund. Die Aufgaben waren inhaltlich auf die Komplexchemie fokussiert und konzentrierten sich insbesondere auf Keilstrichformeln sowie Kugel-Stab-Modelle (Kapitel 7.2.1). Diese wurden anschließend durch Testpersonen mit der Methode des Lauten Denkens bearbeitet und die Videos inhaltlich-semantic transkribiert (Kapitel 7.2.3). Diese Transkripte stellten die Grundlage für die qualitativ orientierte kategoriengeleitete Textanalyse dar. Für die Bearbeitung der Aufgaben wurden Studierende akquiriert, die im Rahmen ihres Studiums bereits eine Veranstaltung zur Komplexchemie absolviert und dadurch Vorwissen sowie Grundkenntnisse zu den Inhalten der Aufgaben mitbrachten. Dadurch sollte eine bessere Vergleichbarkeit der Ergebnisse der drei Studien erreicht werden.

Basierend auf der Theorie und den bisherigen Ergebnissen ließ sich bereits deduktiv ableiten, dass (deklaratives) Fachwissen (z. B. Cook et al., 2008; Edelsbrunner et al., 2023; Nitz, Ainsworth, et al., 2014; Rau, 2018) und räumliche Fähigkeiten (z. B. Buckley et al., 2018; Harle & Towns, 2011; Sorby, 2009; Uttal & Cohen, 2012) zentrale Rollen bei der Bearbeitung repräsentationsbasierter Chemieaufgaben spielen. Daraus ließen sich zwei Hauptkategorien für die Analyse und die Identifikation von Hindernissen im Bearbeitungsprozess ableiten: Hindernisse durch falsches, lückenhaftes oder nicht vorhandenes Fachwissen und Hindernisse durch schwach ausgeprägte räumliche Fähigkeiten. Besonders für das Fachwissen ließen sich theoriebasiert keine spezifischen bzw. differenzierten Hindernisse für die Oberkategorie ableiten. Für die räumlichen Fähigkeiten bot das Cattell-Horn-Carroll Modell zwar eine geeignete Grundlage, indem durch die räumlichen Faktoren differenzierte und spezifische Kategorien für die Analyse angegeben werden, dennoch ist fraglich, inwieweit sich diese in den Transkripten identifizieren und vor allem trennscharf zuordnen lassen.

Forschungsanliegen II hat bereits gezeigt, dass vor allem die zwei- und dreidimensionale Rotationsfähigkeit (*Speeded Rotation* und *Spatial Relation*) sowie die räumliche Manipulationsfähigkeit (*Visualization*) zentrale Fähigkeiten bei der Bearbeitung repräsentationsbasierter Chemieaufgaben darstellen, was auch mit anderen Studien belegt werden kann. Besonders diese drei Fähigkeiten zeigen jedoch eine besonders hohe Überschneidung, was sich auch im bisherigen Forschungsstand widerspiegelt – es wird immer noch diskutiert, ob *Spatial Relation* ein eigenständiger Faktor ist oder als ein Indikator für den

Faktor *Visualization* angesehen werden kann (Buckley et al., 2018). Aus diesem Grund sollten die Kategorien induktiv beim Durchgehen des Materials abgeleitet werden. Des Weiteren hatte diese explorative Kategorienfindung den Vorteil, dass zusätzlich unspezifizierte Hindernisse identifiziert werden konnten, die sich beispielsweise keiner Hauptkategorie zuordnen ließen.

Vor Beginn der induktiven Kategorienbildung wurde ein ausführliches Kodiermanual (Anhang C.5) erstellt, in welchem die Auswertungs-, Kontext- und Kodiereinheit definiert werden.

Auswertungseinheit – Die Transkripte werden nacheinander der qualitativ orientierten kategoriengeleiteten Textanalyse unterzogen.

Kontexteinheit – Das Testheft ist in drei Aufgaben mit Teilaufgaben strukturiert, die die Kontexteinheit darstellen. Die qualitativ orientierte kategoriengeleitete Textanalyse wird so strukturiert, dass jede Teilaufgabe über ein eigenes Kategoriensystem verfügt, wobei die Kategorien der einzelnen Teilaufgaben identisch sind. Dadurch soll die Analyse und Auswertung erleichtert werden.

Kodiereinheit – Der minimale zu kodierende Textbestandteil sind Phrasen in einem Sinnzusammenhang, beispielsweise ein Hauptsatz in einer Satzreihe oder ein Haupt- bzw. Nebensatz in einem Satzgefüge.

Zusätzlich wurde das Selektionskriterium definiert und erläutert. Es zeigt das Verständnis von Hindernissen in dieser Studie auf und stellt dadurch eine präzise Operationalisierung dar, welche Phrasen und Sinnzusammenhänge als Hindernis zu kategorisieren sind. Dieses leitet sich aus den Definitionen von Kechel (2016), Göhner und Krell (2021) sowie Lange (2013) ab:

Ein externalisiertes Hindernis liegt immer dann vor, wenn eine Person einen Aufgabenteil als schwierig empfindet, an einer bestimmten Stelle nicht weiterweiß oder ratlos ist und dies verbalisiert oder wenn eine außenstehende Person beobachten kann, dass eine Person längere Pausen beim Lauten Denken macht und dann ggf. mit einem anderen Aufgabenteil fortfährt, einen Aufgabenteil unsicher und zweifelnd bearbeitet, lückenhafte bzw. fehlerhafte Informationen und Kenntnisse anwendet, die das erfolgreiche Lösen eines Aufgabenteils verzögern bzw. verhindern, oder Verbalisierungen distinkt zu selbst gestalteten Repräsentationen sind.

Die unterstrichenen Begriffe werden im Kodiermanual näher definiert bzw. präzisiert. Beispielsweise wird hier von Externalisierung gesprochen. Zwar fokussiert sich das Laute Denken auf verbale Äußerungen, hier spielen Gesten jedoch eine besondere Rolle, da durch sie kognitive Prozesse, wie die Rotation oder Manipulation von Objekten unterstützt und ausgedrückt werden können (Göksun et al., 2013; Goldin-Meadow et al., 2012; Hostetter & Alibali, 2008). Somit werden unter Externalisierungen sowohl verbale als auch nonverbale Äußerungen zusammengefasst. Diese Ergänzungen und Erläuterungen sollen eine möglichst präzise und objektive Kategorisierung der Hindernisse ermöglichen.

Auf dieser Basis konnte die induktive Kategorienbildung in MAXQDA Analytics Pro (Version 24) durchgeführt werden. Hierzu wurden zusätzlich zwei wissenschaftliche Hilfskräfte hinzugezogen, ausführlich in die Thematik und das Ziel der Studie eingewiesen und mit dem Vorgehen sowie dem Kodiermanual für die qualitativ orientierten kategoriengeleiteten Textanalysen vertraut gemacht. Besonders in der Phase der induktiven Kategorienbildung wurden drei unabhängige Kodierer als sinnvoll erachtet, da dies eine robustere Kategorienbildung durch tiefergehende Diskussionen der Kategorien ermöglicht, Uneinigkeit durch einen Mehrheitsentscheid gelöst und einer subjektiven Verzerrung entgegengewirkt werden kann. Dies gilt auch für den endgültigen Materialdurchgang, in welchem das Material gleichmäßig unter den drei Kodierenden aufgeteilt wurde (Bengtsson, 2016; Döring, 2023c; O'Connor & Joffe, 2020).

In einem ersten Durchgang gingen alle drei Kodierer individuell und unabhängig dieselben drei zufällig ausgewählten Transkripte durch und identifizierten entsprechend dem Kodiermanual Hindernisse und kategorisierten sie. Dabei konnten Textabschnitte mehreren Kategorien zugeordnet werden. Die kategorisierten Textabschnitte und induktiv abgeleiteten Kategorien wurden anschließend verglichen und diskutiert. Dadurch wurde gemeinsam ein Kategoriensystem erstellt, in dem die Kategorien benannt und definiert wurden und das dem Konsens aller drei Kodierer entsprach. Dieses Kategoriensystem wurde auf ein neues viertes Transkript angewendet, um die Kategorien auf ihre Anwendbarkeit zu prüfen und ggf. neue zu ergänzen. Anschließend wurde die Übereinstimmung der Kodierenden mithilfe von Kappa bestimmt (s. o.). Da die Übereinstimmung mit $\kappa < 0,60$ nicht ausreichend hoch war, wurde das Transkript anschließend erneut mit Blick auf die kodierten Textabschnitte und

zugeordneten Kategorien verglichen und diskutiert. Auf dieser Basis erfolgte eine Überarbeitung des Kategoriensystems und des Kodiermanuals, insbesondere des Selektionskriteriums. Dieses Vorgehen wiederholte sich mit drei weiteren Transkripten nacheinander, sodass insgesamt sieben Transkripte für die Entwicklung des Kategoriensystems genutzt wurden. Nach dem letzten Transkript zeigte sich, dass keine weiteren Kategorien ergänzt wurden und die Definitionen der Kategorien für alle Kodierer prinzipiell anwendbar waren. Dennoch bestand eine nicht ausreichende Übereinstimmung bei der Kodierung ($\kappa < 0,60$).

Diese Ergebnisse waren Anlass für eine Diskussion, inwieweit das Vorgehen angepasst und optimiert werden muss. Ein Blick auf die Transkripte zeigte dabei, dass prinzipiell an ähnlichen Stellen im Text die gleichen Kategorien kodiert wurden, was für die Anwendbarkeit des Kategoriensystems sprach, jedoch begannen und endeten die Kategorien an unterschiedlichen Stellen im Transkript. Das ließ sich beispielsweise durch die Reduktion der Minimalüberlappung auf 50 % für die Bestimmung für Kappa erkennen, da sich dadurch der Wert auf $\kappa > 0,60$ erhöhte. Die Frage „Wann beginnt bzw. endet ein Hindernis?“ war somit nicht ausreichend operationalisiert und bot einen zu großen intersubjektiven Spielraum. Zudem zeigte sich, dass auch das Selektionskriterium nicht eindeutig genug war, um zu bestimmen, wann ein Hindernis vorliegt, da teilweise unterschiedliche Stellen im Transkript als solches von den drei Kodierenden identifiziert wurden.

Darauf basierend planten und setzten die Kodierer ein neues Vorgehen um. Die letzten vier Transkripte kodierte das Team abermals, vergab dabei jedoch nur eine Kategorie, die dem Selektionskriterium entsprach. Somit erfassten sie nur die Stellen in den Transkripten, an denen allgemein ein Hindernis entsprechend des Selektionskriteriums bzw. nach der Auslegung der Definition durch die Kodierer vorlag. Die kodierten Textabschnitte verglichen die Beteiligten erneut und diskutierten jeden kategorisierten Abschnitt hinsichtlich des Selektionskriteriums. Die Einigung wurde für alle vier Transkripte in Form einer Konsenskodierung festgehalten, die anschließend auf die Übereinstimmung mit den individuell kodierten Transkripten der drei Kodierer geprüft wurde. Der Rater mit der höchsten Übereinstimmung ($\kappa = 0,71$) wandte das Selektionskriterium am zuverlässigsten an. Dementsprechend übernahm dieser die Kodierung des übrigen Materials (18 Transkripte), um so zunächst die Textpassagen zu kodieren, an denen allgemein ein Hindernis vorliegt. Eine

Konsenskodierung wurde in diesem Fall nicht weiter durchgeführt, da der damit verbundene Aufwand angesichts des Umfangs der Transkripte als unverhältnismäßig hoch zu bewerten war.

Die Kodierung der allgemeinen Hindernisse war im nächsten Schritt die Grundlage für die drei Kodierenden zur Identifikation und Spezifikation der Hindernisse durch die Anwendung des induktiv abgeleiteten Kategoriensystems. Vier zufällig ausgewählte Transkripte wurden von allen Kodierenden kategorisiert und die Übereinstimmung zwischen den Kodierenden anhand des Kappa-Koeffizienten ermittelt. Die erzielten Werte lagen mit $0.72 \leq \kappa \leq 0.79$ in einem guten Bereich, was auf eine hohe Anwendbarkeit sowie eine zufriedenstellende Objektivität des Kategoriensystems hinweist. Das verbleibende Material (14 Transkripte) wurde daher gleichmäßig unter den drei Kodierenden für die finale Kodierung mit dem Kategoriensystem aufgeteilt. Das angewendete Kategoriensystem wird in Kapitel 7.3 dargestellt und an Textbeispielen erläutert.

Das beschriebene Vorgehen berücksichtigt zentrale und essenzielle Maßnahmen der Qualitätssicherung, um die Bestätigbarkeit (*Objektivität*), die Übertragbarkeit und Vertrauenswürdigkeit (*Validität*) sowie die Zuverlässigkeit (*Reliabilität*) der Ergebnisse sicherzustellen (Döring, 2023e; Kuckartz, 2018; Miles et al., 2014). Zu diesen Maßnahmen zählen die induktive und deduktive Kategorienbildung, die Prüfung der Interrater-Übereinstimmung, die ausführliche Dokumentation des Vorgehens, das systematische und regelgeleitete Vorgehen, die Diskussion des Kategoriensystems und der kodierten Textstellen, die Konsensbildung sowie das Training der Kodierenden (Flick, 2022; Göhner & Krell, 2020; Kuckartz, 2018; Mayring & Fenzl, 2022; Miles et al., 2014).

7.3 Datenerhebung und Stichprobe

Der Ablauf der Datenerhebung gliederte sich in drei Phasen. In der ersten Phase haben die Testpersonen eine Einleitung in das Laute Denken und den generellen Ablauf der Datenerhebung gelesen. Da das Laute Denken eine ungewohnte Aufgabe ist und in dem natürlichen Umfeld eher unbewusst stattfindet, wurde das Vorgehen an drei einfachen fachunabhängigen Aufgaben geübt. Nachdem alle offenen Fragen geklärt waren, wurde abschließend darauf hingewiesen, dass es sich nicht um eine Prüfungssituation handelt. Es

wurde betont, dass das Nichtbearbeiten oder fehlerhafte Bearbeiten einzelner Aufgaben keinerlei negative Konsequenzen habe. Zudem erfolgte die Datenerhebung anonym bzw. pseudonymisiert.

Anschließend startete die zweite Phase, die eigentliche Datenerhebung mit Aufnahme des Lauten Denkens. Dabei rückte die Testleitung vollständig in den Hintergrund, d. h. sie verblieb im Raum, um ggf. aufkommende Fragen zu klären oder Hinweise zu geben (beispielsweise, wenn aufgabenbezogenes Material ignoriert oder übersehen wird), entzog sich jedoch dem Sichtfeld der Testperson. Die Interaktion zwischen Testperson und Testleitung wurde auf ein Minimum reduziert. Im Idealfall sollte die Testperson alle Aufgaben ohne das Einwirken der Testleitung bearbeiten. Dadurch sollte jegliche Ablenkung ausgeschlossen werden, sodass sich die Testperson auf die Aufgaben konzentrieren konnte.

Um sicherzustellen, dass die Testpersonen ohne Unterbrechung sprechen und das Laute Denken nicht unterbrochen wird, wurden die Testpersonen bei jeder Aufgabenstellung daran erinnert, ihre Gedanken zu verbalisieren (Konrad, 2010; Sandmann, 2014; Schuttkowski & Staubach, 2022). Nach Abschluss der letzten Aufgabe wurde die Aufnahme gestoppt. In der abschließenden Erhebungsphase wurden personenbezogene Daten (Alter, Geschlecht, Fachsemester und Studiengang) erfasst. Gleichzeitig wurde das Ziel der Datenerhebung transparent kommuniziert: Die Untersuchung potenzieller Hindernisse im Bearbeitungsprozess von Aufgaben zur *Interpretation*, *Translation* und *Konstruktion* chemischer Repräsentationen, um zentrale kognitive Prozesse und Fähigkeiten zu identifizieren, die mit einer erfolgreichen Aufgabenbewältigung in Zusammenhang stehen.

Die Arbeitsumgebung wurde in allen Testungen gleichgehalten. Auf dem Arbeitsplatz fanden die Testpersonen ihr Testheft, ein Tippkärtchen, das Isomerie erklärt, und eine Erinnerung, dass sie durchgehend laut denken und ununterbrochen reden sollen. Vor dem Arbeitsplatz wurde ein Stativ mit einem Tablett für die Videoaufnahme installiert und ein Tablet auf dem Arbeitsplatz für die Audioaufnahme platziert. Die Testleitung blieb während der gesamten Bearbeitung außerhalb des Sichtfelds der Testperson. Die Bearbeitung der Aufgaben hat zwischen 27 und 85 Minuten gedauert ($M_{\text{Zeit}} = 58.25 \text{ min}$, $SD_{\text{Zeit}} = 15.05 \text{ min}$).

An der Datenerhebung nahmen insgesamt $N = 25$ Lehramtsstudierende ($n_w = 18$ (72 %), $n_m = 7$ (28 %)) mit dem Fach Chemie für das Gymnasium im Frühjahr 2023 teil. Diese waren zwischen 20 und 26 Jahre alt ($M_{\text{Alter}} = 21.8$ Jahre, $SD_{\text{Alter}} = 1.6$ Jahre) und studierten im 4. bis 10. Semester ($M_{\text{Semester}} = 6.2$, $SD_{\text{Semester}} = 1.3$). Alle Testpersonen haben im Laufe ihres Studiums bereits an mindestens einer Veranstaltung teilgenommen, in der die Komplexchemie behandelt wurde, sodass theoretisch die notwendigen Grundbegriffe bekannt waren.

7.4 Ergebnisse

Die induktive Kategorienbildung mit insgesamt fünf Schleifen zur Diskussion und Anpassung des Kategoriensystems durch die drei Kodierenden hat in der Summe zwölf Kategorien hervorgebracht, die sich vier Hauptkategorien zuordnen lassen. Dafür wurden sieben der insgesamt 25 Transkripte verwendet, die von den abschließenden Analysen ausgeschlossen wurden. Die hier präsentierten Ergebnisse beziehen sich auf die 18 verbleibenden Transkripte. Das finale Kategoriensystem ist in Tabelle 22 dargestellt. Im Folgenden werden die vier Hauptkategorien kurz erörtert.

① **Hindernisse mit Bezug zum Fachwissen** – Diese Hauptkategorie wurde theoriebasiert deduktiv abgeleitet. Insgesamt wurden dieser vier Kategorien zugeordnet, die sich durch einen direkten Bezug zum Fachwissen (deklarativ, prozedural oder metakognitiv) auszeichneten. Dieses umfasst chemisches Fachwissen aber auch Wissen über Repräsentationen bzw. die Konventionen von Repräsentationen.

② **Hindernisse mit Bezug zu räumlichen Fähigkeiten** – Auch diese Hauptkategorie wurde deduktiv aus der Theorie abgeleitet. Vor Beginn der induktiven Kategorienbildung hat sich die Frage gestellt, ob und inwieweit das Cattell-Horn-Carroll Modell (Abbildung 8) und die darin abgebildeten räumlichen Faktoren (Tabelle 12) als Kategorien anwendbar sind. Die induktive Kategorienbildung hat schnell gezeigt, dass sich die räumlichen Faktoren nicht als ausreichend differenzierbare Kategorien erwiesen haben. Es konnten zwar mentale Rotations- (zwei- und dreidimensional) und Manipulationsprozesse erkannt werden, diese ließen sich aber nicht eindeutig operationalisieren bzw. zuord-

nen, weshalb diese in einer Kategorie (2.1) zusammengefasst wurden. Die Fähigkeit, Figuren in komplexen Mustern zu erkennen (*Flexibility of Closure*), wurde ebenfalls diskutiert. Eine Abgrenzung zwischen Mustererkennung und dem reinen Beschreiben sichtbarer Merkmale einer Repräsentation konnte im Rahmen dieser Untersuchung nicht ausreichend differenziert und operationalisiert werden, weshalb diese Kategorie nicht in das Kategoriensystem aufgenommen wurde.

③ **Hindernisse mit direktem Bezug zur Repräsentation** – Diese Hauptkategorie wurde nach der induktiven Kategorienbildung aus den Kategorien abgeleitet. Hierunter wurden Hindernisse zusammengefasst, die direkt in Bezug auf Merkmale standen und unmittelbar aus einer Repräsentation abgelesen werden konnten. Diese Hauptkategorie grenzt sich von ① insofern ab, dass hier keine zusätzlichen Informationen bzw. Vorwissen eingebracht werden.

④ **Sonstige Hindernisse** – Wie ③ wurde diese Hauptkategorie nach der induktiven Kategorienbildung abgeleitet. Diese umfasst Hindernisse, die auf sonstige Fähigkeiten und Kompetenzen (mathematische Kompetenzen und Lesekompetenz), das persönliche Wohlbefinden oder erhebungsbedingte Aspekte zurückzuführen waren. Außerdem wurde hier das Nichtverwenden von Aufgabenmaterial oder das Nichtbeantworten von Aufgaben verortet, sofern kein Hindernis externalisiert wurde und diese lediglich ignoriert oder übersprungen wurden.

Tabelle 22: Finales Kategoriensystem zur systematischen Identifikation und Beschreibung von Hindernissen bei der *Interpretation*, *Translation* und *Konstruktion* von Keilstrichformeln und Kugel-Stab-Modellen in der Komplexchemie mit Textbeispielen zur Veranschaulichung.

① Hindernisse mit Bezug zum Fachwissen

1.1) Einsatz inadäquater Fachbegriffe

Eine Person kann nicht auf adäquate Fachbegriffe zurückgreifen oder nutzt falsche, nicht-chemische Fachbegriffe (z. B. Neologismen), wobei aber kein falsches chemisches Fachkonzept zugrunde liegt.

Beispiel: „[...] irgendwie so tetraedrisch-planar ist das hier [...]“ (TWEP29, Abs. 39).

Erläuterung: Gemeint ist hier quadratisch-planar, es wird nur ein falscher Begriff verwendet, der keine Komplexgeometrie beschreibt.

1.2) Beschreibung/Anwendung fehlerhafter Fachkonzepte/-inhalte

Eine Person externalisiert fehlerhafte chemische Fachkonzepte, d. h. nicht korrekte Konzepte bzw. Sachzusammenhänge in einem bestimmten Themenbereich oder sie kombiniert chemisch korrekte Fachbegriffe, die im Zusammenhang inkorrekt sind.

Beispiel: „Also das sind auch Konstitutionsisomere, weil die haben wieder die gleiche Summenformel, aber nicht die gleiche Anzahl ähm Anordnung von den Bindungen. Dann sind das Konstitutionsisomere“ (TWEP29, Abs. 41).

Erläuterung: Die Person kombiniert die Definitionen der Konstitutions- und Konfigurationsisomerie (trotz Verwendung des Tippkärtchens).

1.3) Diskrepanz zwischen Fachkonzepten

Eine Person externalisiert eine Diskrepanz zwischen zwei (bestehenden) explizit genannten Fachkonzepten, die richtig oder falsch sein können, aber im Widerspruch zueinanderstehen.

Beispiel: „soweit ich weiß, müsste das ein Tetraamminaqu-Komplex sein, also eigentlich müsste da auch noch irgendwo Wasser hingebunden haben“ (JKPS07, Abs. 21).

Erläuterung: Der dargestellte Komplex ist (fachlich korrekt) mit vier Ammin-Liganden dargestellt. Die Person geht jedoch davon aus, dass Wasser-Liganden fehlen.

1.4) Beschreibung/Anwendung fehlerhafter Konventionen chemischer Repräsentationen

Eine Person externalisiert nicht-konventionsgemäße chemische Repräsentationen (symbolisch-mathematisch oder visuell-graphisch), Fehler in Bezug auf den Zweck und Nutzen einer Repräsentation, wie die Repräsentation Informationen kommuniziert oder welche Bedeutung spezifischen Repräsentationsmerkmalen zukommt.

Beispiel: „[...] die Größe der Kugeln steht für die Größe des Atomradius, beziehungsweise halt wie viele Elektronen quasi auf den besetzten Schalen sitzen. Quasi für die, ja, Periode mehr oder weniger, aus der sie kommen“ (ALCG65, Abs. 9).

Erläuterung: Die Kugelgröße gibt verhältnismäßig die Atomradien wieder, gibt aber keine Informationen über die Anzahl der Elektronen oder die Anzahl der Schalen.

② Hindernisse mit Bezug zu räumlichen Fähigkeiten

2.1) Hindernis bei mentaler Manipulation chemischer Repräsentationen

Eine Person externalisiert fehlerhafte mentale Manipulations- oder Rotationsprozesse bei der Bearbeitung einer chemischen Repräsentation.

Beispiel: „Wie muss ich es denn gedreht haben? (*Drehende Handbewegungen*) Ich muss es dann einmal um 180 Grad gedreht haben und dann (Hand dreht sich auf Papierebene um 180 Grad nach links) ist dann hier das NO₂. Nein, dann ist das NO₂ ja nach links (dreht Hand wieder). Ich kann es ja auch andersherum gedreht haben“ (TWEP29, Abs. 43).

Erläuterung: Der beschriebene Rotationsprozess ist zwar richtig, um die Aufgabe zu lösen, dabei werden jedoch Positionen von Liganden vertauscht, d. h. die Rotation nicht korrekt beschrieben, wodurch die Aufgabe nicht richtig gelöst wird.

③ Hindernisse mit direktem Bezug zur Repräsentation

3.1) Beschränkung auf Oberflächenmerkmale

Eine Person externalisiert lediglich oberflächliche Merkmale einer Repräsentation, d. h. Informationen, die sich direkt aus einer Repräsentation ablesen lassen, ohne auf zugrunde liegende chemische Konzepte zurückzugreifen, wodurch die Externalisierung fachlich zwar nicht gänzlich falsch ist, aber im Kontext inadäquat/nicht präzise genug ist.

Beispiel: „Also Zentralion ist bei der eins ähm das Aluminium“ (TWEP29, Abs. 30).

Erläuterung: Aluminium lässt sich direkt ablesen, es ist jedoch ein Aluminium(III)-Ion, was erst durch die Interpretation der Repräsentation ersichtlich wird.

3.2) Hindernis beim Vergleichen multipler Repräsentationen

Eine Person externalisiert Hindernisse beim Vergleichen multipler Repräsentationen, die sich auf variable oder ggf. nicht kategorisierbare bzw. identifizierbare Gründe zurückführen lassen.

Beispiel: „Ich würde glaube ich als erstes Mal bei den Molekülkomplexen mir das Hinschreiben, beziehungsweise mir daneben das noch malen in der Keilstrichformel, weil dann ist das einfacher zu vergleichen“ (NWNZ30, Abs. 20).

Erläuterung: Die Person hat Schwierigkeiten beim Vergleich von Keilstrichformel und Kugel-Stab-Modell, ohne dies explizit zu erläutern.

④ Sonstige Hindernisse

4.1) Nicht-Verwenden aufgabenbezogenen Materials

Eine Person externalisiert, aufgabenbezogenes Material nicht zu nutzen, wodurch es im weiteren Verlauf zu selbstinduzierten Verzögerungen oder Fehlern bei der Aufgabenlösung kommt.

Beispiel: „Ich glaube/ oder ich verwende mal das Tipp-Kärtchen“ (TWEP29, Abs. 41).

Erläuterung: Bis zu diesem Punkt hat die Person die Aufgabe ohne das Tippkärtchen bearbeitet und dabei einige Hindernisse gezeigt.

4.2) Nicht-Beantworten einer Teilaufgabe

Die Person beantwortet eine Teilaufgabe aus variablen Gründen gar nicht, d. h. die Teilaufgabe wird zu Beginn zwar vorgelesen, darauf bei der Beantwortung aber nicht eingegangen (die Entsprechende Teilaufgabe ist in der Aufgabenstellung zu kodieren).

4.3) Mangelnde Lesekompetenz

Eine Person liest eine Teilaufgabe falsch vor oder überspringt Aspekte, wodurch die Teilaufgabe ggf. eine neue Bedeutung erhält (die entsprechende Teilaufgabe wird in der Aufgabenstellung kodiert).

Beispiel: „Dann sage ich jetzt, das hier ist ein Konfigurationsisomer“ (TWEP29, Abs. 45).

Erläuterung: Die Aufgabenstellung hat die Konfigurationsisomerie als richtige Lösung bereits ausgeschlossen.

4.4) Mangelnde mathematische Kompetenzen

Eine Person externalisiert Hindernisse bei variablen mathematischen Operationen, z. B. beim Zählen oder Berechnen der Anzahl an Atomen oder Liganden.

Beispiel: „[...]5-mal ähm Stickstoffatome (..) und jeweils 3, 6, 9, 12, 18. 18 Wasserstoffatome, die eben an den Stickstoffatomen liegen“ (OYNK19, Abs. 11).

Erläuterung: Es sind 15 und nicht 18 Wasserstoffatome in der Verbindung.

4.5) Erhebungsbedingte und persönliche Hindernisse

Eine Person externalisiert Hindernisse, die direkt durch die Umfrage entstehen (z. B. durch den Umfang der Erhebung oder die Aufgabenstellung) bzw. auf das persönliche Befinden zurückgeführt werden können (z. B. Müdigkeit, wenig Schlaf oder Kopfschmerzen).

Beispiel: „[...] ich merke gerade, dass es mir sehr schwer fällt bei solchen Aufgaben jetzt gleichzeitig zu reden und zu denken“ (SJMKS2, Abs. 51).

Erläuterung: Das Hindernis entsteht nicht durch mangelnde aufgabenbezogene Kompetenzen oder Fähigkeiten, sondern durch die Methode des lauten Denkens.

Die zwölf Kategorien wurden auf die 18 Transkripte angewendet. Sechs der Kategorien sind bei 70 bis 100 % der Personen zu identifizieren und haben auch die höchste Anzahl an Kodierungen. Die am häufigsten kodierte Kategorie war „2.1) Beschreibung/Anwendung fehlerhafter Fachkonzepte/-inhalte“ mit insgesamt 244 Kodierungen. Dagegen wurde die Kategorie „1.1) Einsatz inadäquater Fachbegriffe“ am wenigsten kodiert (insgesamt zwei Mal).

Mit Blick auf die drei Aufgaben zur *Interpretation* (A1), *Translation* (A2) und *Konstruktion* (A3) lassen sich deutliche Unterschiede bei der Kodierung erkennen. Die Kategorien „1.2) Beschreibung/Anwendung fehlerhafter Fachkonzepte/-inhalte“ und „3.1) Beschränkung auf Oberflächenmerkmale“ wurden mit Abstand am häufigsten bei der *Interpretation* verwendet. Zwar wurde Kategorie 3.1 auch bei der Konstruktionsaufgabe oft kodiert, diese finden sich jedoch ausschließlich in dem Interpretationsteil der Aufgabe wieder. Daher kann gesagt werden, dass dieser Code nahezu ausschließlich bei der *Interpretation* gesetzt wurde.

Die Kategorien „2.1) Hindernis bei mentaler Manipulation chemischer Repräsentationen“ und „3.2) Hindernis beim Vergleichen multipler Repräsentationen“ wurden mit Abstand am häufigsten bei der *Translation*, aber auch bei der *Konstruktion* identifiziert. Eine ausführliche Übersicht über die Häufigkeit der Kodierungen aller Kategorien bei den drei Aufgaben ist in Tabelle 23 dargestellt.

Tabelle 23: Übersicht mit der absoluten Anzahl an Personen für jede Kategorie sowie der Anzahl der Kodierungen jeder Kategorie getrennt nach Aufgaben bei der Introspektion (d. h. ohne die Reflexionsaufgaben mit Retrospektion).

Hindernis	Anzahl der Personen mit dem Hindernis	Anzahl Kodierungen			
		A1	A2	A3	Σ
1) Hindernisse mit Bezug zum Fachwissen					
1.1) Einsatz inadäquater Fachbegriffe	2	1	0	1	2
1.2) Beschreibung/Anwendung fehlerhafter Fachkonzepte/-inhalte	18	134	84	26	244
1.3) Diskrepanz zwischen Fachkonzepten	3	3	0	1	4
1.4) Beschreibung/Anwendung fehlerhafter Konventionen chemischer Repräsentationen	16	5	18	26	68
2) Hindernisse mit Bezug zu räumlichen Fähigkeiten					
2.1) Hindernis bei mentaler Manipulation chemischer Repräsentationen	13	1	23	7	31
3) Hindernisse mit direktem Bezug zur Repräsentation					
3.1) Beschränkung auf Oberflächenmerkmale	18	67	2	65*	134
3.2) Hindernis beim Vergleichen multipler Repräsentationen	17	0	125	48	173
4) Sonstige Hindernisse					
4.1) Nicht-Verwenden aufgabenbezogenen Materials	4	0	4	1	5
4.2) Nicht-Beantworten einer Teilaufgabe	16	55	6	5	66
4.3) Mangelnde Lesekompetenz	7	3	2	5	10
4.4) Mangelnde mathematische Kompetenzen	5	6	0	0	6
4.5) Erhebungsbedingte und persönliche Hindernisse	5	2	7	4	13

Hinweise: A1 = Interpretationsaufgabe, A2 = Translationsaufgaben, A3 = Konstruktionsaufgaben (* ausschließlich Interpretationsteil A3)

Mit Blick auf die einzelnen Testpersonen lässt sich festhalten, dass zwei bis neun der zwölf Kategorien identifiziert werden konnten ($M_{\text{Kat}} = 6.6$, $SD_{\text{Kat}} = 1.5$). Insgesamt wurden zwischen 7 und 52 Hindernisse kodiert ($M_{\text{Code}} = 31.2$, $SD_{\text{Code}} = 12.0$). Was die Bearbeitung der Aufgaben angeht, erreichten die Testpersonen zwischen 39 % und 92 % der Punkte ($M_{\text{Pkt.}} = 60.0$ %, $SD_{\text{Pkt.}} = 13.3$ %). In Tabelle 24 ist eine detaillierte Charakterisierung der 18 Testpersonen zu entnehmen, deren Transkripte für die finale Analyse verwendet werden.

Zur Sicherung der Qualität und der Objektivität bei der Auswertung der Aufgaben wurden 22 % der Testhefte (vier Stück) von einem zweiten Kodierer bewertet. Die Kodierenden zeigen dabei eine Übereinstimmung von $\kappa = 0.85$, was nach Landis und Koch (1977) für eine hohe Objektivität der Bewertung spricht. Mit Blick auf die erreichten Punkte und die Anzahl der kodierten Hindernisse lässt sich deskriptiv ein negativer Zusammenhang erkennen: Je mehr Hindernisse identifiziert werden, desto weniger Punkte werden erreicht und vice versa. Zwar erscheint dieser Zusammenhang logisch, muss jedoch vorsichtig interpretiert werden, da die Transkripte unterschiedlich lang sind und für das Kodieren keine klaren Grenzen (wo beginnt und endet ein Hindernis? s. Kapitel 7.2.4) operationalisiert werden können.

Im Folgenden werden die bisher dargestellten Ergebnisse ausführlicher dargestellt und mit Inhalten gefüllt, um die Art der Hindernisse besser zu verstehen und nachvollziehen zu können. Dazu werden zunächst die Ergebnisse der Introspektion aufgeführt, also die Ergebnisse der Externalisierungen während der Bearbeitung der Aufgaben. Anschließend wird auf die Retrospektion eingegangen, d. h. die Reflexion der bearbeiteten Aufgaben in Bezug auf selbst identifizierte Hindernisse und Unterstützungsbedarfe der Testpersonen. Die Inhalte werden dazu exemplarisch mit Aussagen der Testpersonen untermauert.

Tabelle 24: Charakterisierung der Testpersonen durch persönliche Angaben, erreichte Punkte in den Aufgaben, Anzahl vorhandener Hinderniskategorien und dem Gesamtumfang der Kodierungen bei der Introspektion (d. h. ohne die Reflexionsaufgaben mit Retrospektion).

Person	Gender	Alter	Semester	Letzte Note ²	Dauer in min	Punkte Aufgaben ¹	Anzahl Kategorien	Anzahl Codes
ARBS75	♂	20	6	15	34	93 (0.76)	7	13
BCFH58	♀	20	6	15	54	86 (0.70)	6	19
CTNR37	♀	23	6	11	79	70 (0.57)	8	36
DHPK58	♀	22	6	-	74	48 (0.39)	7	27
ENFP62	♀	22	6	11	27	60 (0.49)	8	38
INNF75	♀	24	4	15	66	61 (0.50)	9	52
KRRH17	♀	21	6	14	72	95 (0.77)	6	37
MRFL66	♂	21	6	11	55	68 (0.55)	8	25
NJBS00	♀	21	6	14	75	68 (0.55)	7	48
NMEM52	♀	21	4	9	52	77 (0.63)	7	40
NSMK54	♂	20	6	5	63	58 (0.47)	7	37
NWNZ30	♀	22	6	11	47	61 (0.50)	6	34
SDSB66	♂	22	6	9	49	69 (0.56)	7	44
SHFS79	♀	20	6	11	45	80 (0.65)	7	32
SJMK52	♀	22	6	13	62	87 (0.71)	6	21
SMRG54	♂	24	8	8	60	77 (0.63)	5	20
STJS58	♂	23	6	11	44	57 (0.46)	6	31
SUF508	♀	22	8	-	85	113 (0.92)	2	7

1 = Erreichte Gesamtpunktzahl (absolut und relativ) in den Aufgaben (basierend auf dem Erwartungshorizont, max. 123 Punkte);

2 = Letzte Chemienote auf einem Schulzeugnis

7.4.1 Hindernisse bei der *Interpretation*

Mit Blick auf Tabelle 23 fällt sogleich auf, dass bei der *Interpretation* (A1) drei Kategorien dominieren. Sie tauchen bei allen oder fast allen Testpersonen auf. Diese drei zentralen Kategorien werden hier näher beleuchtet, um ihre Schlüsselrolle bei der Bearbeitung von A1 zu verdeutlichen. Zwei weitere Kategorien wurden zwar seltener kodiert, bieten aber interessante Einblicke in das Verständnis über Repräsentationen und werden im späteren Verlauf bei den folgenden Aufgaben wieder aufgegriffen.

► **1.2) Beschreiben/Anwenden fehlerhafter Fachkonzepte/-inhalte** – Bei der Beantwortung der Aufgaben fällt besonders auf, dass die Testpersonen Hindernisse in Bezug auf einige Fachbegriffe hatten. Vor allem die Begriffe „Koordinationszahl“ und „Zähnigkeit“ stellten zentrale Hindernisse in Aufgabe 1 dar. So konnten die Testpersonen die Begriffe teilweise nicht definieren und die Aufgabe aus diesem Grund nicht beantworten:

„Zähnigkeit? Das ist jetzt zum Beispiel ein Begriff, den kenne ich nicht“ (SDSB66, Abs. 7).

„Was ist die Koordinationszahl? Well...“ (NMEM52, Abs. 4).

Zudem wurden hier Fachbegriffe und Definitionen verwechselt, wodurch die Aufgaben falsch beantwortet wurden. In einzelnen Fällen wurde für die Zäh-nigkeit die Koordinationszahl angegeben. Dementsprechend konnten einige wenige Testpersonen die Begriffe dem Zentralatom/-ion bzw. den Liganden nicht korrekt zuordnen. Die Koordinationszahl bezieht sich auf das Zentralatom/-ion (Anzahl der daran gebundenen Atome) und die Zähnigkeit auf die Liganden (Anzahl der Koordinationsstellen des Liganden). Häufiger wurde aber die Koordinationszahl mit der Oxidationszahl des Zentralatoms/-ions gleichgesetzt, teilweise auch mit der Gesamtladung des Komplexes.

„Koordinationszahl, ist das diese 3 plus als Koordinationszahl? Ich habe keine Ahnung ey. Aber ich geh jetzt mal davon aus, was wäre es denn davor gewesen dann? Da müsste die Koordinationszahl beim Chrom ja 1 plus sein [es ist Cr^{3+}], oder? Wenn das insgesamt positiv geladen ist“ (NMEM52, Abs. 9).

In den Aufgaben 1.1 bis 1.3 bestimmten insgesamt bis zu zehn Testpersonen die Koordinationszahl der Zentralatome/-ionen korrekt, während bis zu sechs die Zähigkeit der Liganden richtig angaben.

Ähnliche Hindernisse traten auch bei der Benennung des Bindungstyps innerhalb des Komplexes und der Molekülliganden auf. Zunächst bestand bei vielen Testpersonen eine Verwirrung um den Begriff „Bindungstyp“. Darunter werden Metallbindung, Ionenbindung und kovalente Bindung zusammengefasst (Binnewies et al., 2016; Mortimer & Müller, 2020), wobei letztere innerhalb der Molekülliganden und der Bindung zwischen Zentralatom/-ion und Liganden entspricht. Der Bindungstyp wurde jedoch teilweise mit dem Begriff der Mehrfachbindung verwechselt, die zwar zu den kovalenten Bindungen zählt, jedoch nicht die korrekte Antwort darstellt. Häufiger wurde eine ionische Bindung angenommen und die Verbindung als Salz deklariert. Teilweise wurde auch angegeben, dass im Komplex sowohl ionische als auch kovalente Bindungen vorliegen. Auch bei den Molekülen Wasser (A1.1) und Ammoniak (A1.2) konnten die Testpersonen die kovalente Bindung teilweise nicht benennen. Insgesamt beschrieben nur sechs der Personen die Bindung zwischen Zentralatom/-ion und Liganden korrekt als kovalent.

„Also bei Wasser ist es Wasserstoffbrückenbindungen. Ähm (..) bei dem Chlor und dem Chrom müsste es eigentlich eine (.) ionische Bindung sein? Doch, eine ionische Bindung sein. Und das wars auch, mehr gibt es nicht. Doch, die Bindung vom Wasser zum Chrom natürlich noch. Die Bindung vom Wasser zum Chrom müsste (...) eine hm (....)/ Ist das jetzt eine/ das ist ionisch/ ähm das müsste eine kovalente Bindung sein“ (SDSB66, Abs. 11).

► **3.1) Beschränkung auf Oberflächenmerkmale** – Bei der Angabe der Zentralionen und der ionischen Liganden fällt auf, dass diese häufig nicht vollständig korrekt angegeben wurden. Die Oxidationsstufe, d. h. die Ladung der Ionen, wurde bei fast der Hälfte der Testpersonen regelmäßig nicht angegeben. Diese Angabe ist jedoch notwendig und wichtig, da besonders die Metalle der Nebengruppen in verschiedenen Oxidationsstufen vorliegen können (z. B. Fe^{2+} und Fe^{3+} oder Cr^{3+} und Cr^{6+}) und Komplexverbindungen dementsprechend auch unterschiedliche physikalische und chemische Eigenschaften aufweisen. Ebenso wurde von der Hälfte der Testpersonen Chlorid nicht als solches angegeben, sondern als Chlor oder Chloratom. Eine fachsprachliche

Genauigkeit ist jedoch wichtig, da Chlor bzw. Chloratom implizieren, das molekulares Chlor (Cl_2) oder ein radikalisches, hochreaktives Chloratom ($\text{Cl}\cdot$) vorliegen und eben kein anionisches Chlorid (Cl^-). Die Angaben vieler Testpersonen waren somit nicht richtig, lassen sich jedoch so aus den Repräsentationen ableiten. Sie lesen lediglich die oberflächlichen Merkmale daraus ab (z. B. Chrom (Cr) als Zentralatom bei A1.1). Teilweise wussten die Testpersonen, dass ein Zentralion vorliegt, spezifizierten es jedoch nicht (z. B. „das Zentralion ist das Chromion“ (NJB500, Abs. 4)), wobei unklar bleibt, weshalb sie auf eine Präzisierung verzichteten.

Hier fällt im Vergleich zwischen den Aufgaben mit Keilstrichformel und Kugel-Stab-Modell auf, dass bei den Kugel-Stab-Modellen zumindest deskriptiv häufiger die Ladung mit angegeben wurde, sowohl die des Zentralions als auch die der ionischen Liganden. Dies lässt sich jedoch durch die angegebene Legende einfach erklären, da dort explizite Angaben gemacht wurden (z. B., dass die rosafarbene Kugel in A1.2 für ein Cobalt(III)-Ion (Co^{3+}) steht). Dies war bei den Keilstrichformeln nicht der Fall und musste durch die Interpretation der Repräsentation und Vorwissen (z. B., dass Wasser ein neutrales Molekül ist oder dass das Element Chlor in Komplexverbindungen als Chlorid vorliegt) bestimmt werden.

Wie oben bereits beschrieben, können die Kodierungen dieses Hindernisses bei Aufgabe 3 auch der *Interpretation* zugeordnet werden und nicht der *Konstruktion*. Die erste Aufgabe bestand darin, die Zentralatome/-ionen und die Liganden aus der Summenformel anzugeben. Auch hier beschränkte sich mehr als die Hälfte der Testpersonen lediglich auf die Oberflächenmerkmale und gab beispielsweise für den Komplex $[\text{PtCl}_2(\text{NH}_3)_2]$ Platin (Pt) statt Platin(II) (Pt^{2+}) als Zentralatom/-ion und Chlor ($\text{Cl}_2/\text{Cl}\cdot$) statt Chlorid (Cl^-) als Liganden an.

► **4.2) Nicht-Beantworten einer Teilaufgabe** – Dieser Code wurde insgesamt 55-mal gesetzt. Dabei lässt sich in den meisten Fällen nur mutmaßen, weshalb eine Teilaufgabe nicht beantwortet oder auf einen Aspekt nicht eingegangen wurde. Teilweise gaben die Personen zu einem früheren Zeitpunkt an, dass sie beispielsweise nicht wüssten, was Zähigkeit bedeutet und übersprangen diesen Punkt. In manchen Fällen kann jedoch auch auf einen Flüchtigkeitsfehler oder einfach das Vergessen des Aspekts geschlossen werden. Interessant ist

die Teilaufgabe, in der die Testpersonen auf die sichtbaren Merkmale der Repräsentationen eingehen sollten (A1.1.b und A1.2.b). Hier gaben zehn bzw. elf Personen nicht an, wofür Elementsymbole bzw. Kugeln in den Repräsentationen stehen. Dies könnte sich als Selbstverständlichkeit für Personen mit chemischem Hintergrund interpretieren lassen, sodass diese es als überflüssig erachteten, diese Merkmale zu erklären.

► **4.4) Mangelnde mathematische Kompetenzen** – Diese Kategorie wurde ausschließlich bei der ersten Aufgabe kodiert, insgesamt jedoch nur sechs Mal. Sie ist interessant, da die Testpersonen teilweise eine falsche Anzahl an Liganden (z. B. „3 Mal NH_3 “ (NJBS00, Abs. 14) statt vier NH_3) oder an insgesamt vorhandenen Atomen zählten oder berechneten.

► **1.4) Beschreibung/Anwendung fehlerhafter Konventionen chemischer Repräsentationen** – Auch diese Kategorie ist interessant und verdient einen genaueren Blick. Zwar konnten fast alle Testpersonen die Bedeutung von gestrichelten und ausgefüllten Keilen sowie einfachen Strichen in der Keilstrichformel angeben und wussten, dass darüber der räumliche Bau von Verbindungen angedeutet wird, sie konnten jedoch nicht beschreiben, dass es sich dabei um klar geregelte Konvention zur systematischen Darstellung dieser Verbindungen handelt. Auch mit Blick auf die Angabe des Bindungstypen wurden hier einige Defizite deutlich. Sowohl die Keile als auch die Striche symbolisieren kovalente Bindungen und schließen somit ionische Bindungen aus, die von vielen Testpersonen jedoch als Bindungstyp für Komplexe angegeben wurde.

Im Zusammenhang mit dieser Kategorie und der Kategorie „1.3) Diskrepanz zwischen Fachkonzepten“ ist eine Aussage sehr interessant:

„Man erkennt, warum es vielleicht doch keine Komplexbindung ist, weil man erkennt, wie weit die Bindungen reinreichen, also dürfte es ja eigentlich keine dative/ das ist keine dative Bindung das ist eine kovalente Bindung. Weil man erkennt an der Darstellungsweise von den Atomen, die gehen nur bis zur Hälfte, das heißt die teilen sich die Bindung. Das geht jetzt nicht ineinander über, wie's eigentlich bei einem Komplex sein sollte. Ja, aber dann würde es ja auch keine Koordinationszahl geben. Also wir streichen es nochmal durch, das ist auf jeden Fall eine dative Bindung“ (NJBS00, Abs. 13).

Die Person wusste, dass das Chloridion und die Ammoniakmoleküle Elektronenpaardonatoren sind und dass es sich hier um dative Bindungen handeln musste (auch wenn dieser Begriff nicht mehr aktuell ist, wird dieser noch gelehrt und hier als richtig gewertet). Die Repräsentation stand jedoch im Widerspruch zum Vorwissen, da die Bindungen im Kugel-Stab-Modell hier zweifarbig dargestellt wurden und somit implizieren, dass die Bindungselektronen von beiden Bindungspartnern stammen, was nicht der Fall ist. Dies ist eine berechtigte Kritik, weshalb einfarbige Bindungen in dieser Darstellung des Komplexes vorzuziehen wären. Die Testperson zeigte an dieser Stelle durch die kritische Betrachtung der Repräsentation bereits das Erreichen der Fähigkeit *Vergleich und Kritik* aus den higher-level skills (Tabelle 1).

► Die *Retrospektion* im Anschluss an Aufgabe 1 zeigte, dass sich die Testpersonen zum Großteil ihrer Hindernisse bewusst waren. Sie führten diese maßgeblich auf mangelndes Fachwissen zurück. Entsprechend gaben sie an, dass es ihnen helfen würde, wenn sie sich vorher die Grundlagen zur Komplexchemie noch einmal angeschaut und vor allem Fachbegriffe wiederholt hätten:

„Also besser helfen würde natürlich, wenn man mit dem Thema Komplexchemie frischer vertraut wäre, [...] demnach sind Grundbegriffe wahrscheinlich vorhanden, aber ein tieferes Wissen ähm müsste man dann vermutlich ähm aktiv noch einmal nachschauen, damit ist es wieder aktiviert wird“ (CTNR37, Abs. 35).

7.4.2 Hindernisse bei der *Translation*

Der Blick auf Tabelle 23 zeigt, dass bei der *Translation* von Repräsentationen vier Hindernisse zentral bei der Bearbeitung der Aufgaben waren und bei über 70 bis 100 % der Personen auftraten. Diese werden im Anschluss näher beleuchtet.

► **3.2) Hindernisse beim Vergleich multipler Repräsentationen** – In den Aufgaben 2.2 bis 2.4 sollten die Testpersonen Komplexe miteinander vergleichen und angeben, ob es sich um die gleichen Komplexe handelt oder um Isomere, wobei in diesem Fall die Art der Isomerie anzugeben war. Dazu wurde eine Übersicht mit Definitionen zu den verschiedenen Isomerie-Arten mit Beispielen, d. h. Repräsentationen aus der Organischen Chemie, zur Verfügung gestellt. Die Definition der Kategorie (Tabelle 22) wurde bewusst offengehalten.

Dies lässt sich damit begründen, dass in vielen Fällen offensichtlich Hindernisse identifiziert werden konnten, bei denen aber der Grund des Hindernisses nicht klar aus den Externalisierungen der Testpersonen hervorging. Hier ein Beispiel aus Aufgabe 2.3:

„Ähm genau also bei dem unten links handelt es sich auf jeden Fall auch um ein Konstitutionsisomer glaube ich und bei dem unten rechts ähm ist es ein, auch ein Konstitutionsisomer“ (ENPF62, Abs. 53).

Die Testperson gab an, dass die Bezugsrepräsentation und die zwei Vergleichsrepräsentationen *Konstitutionsisomere* seien. Die Repräsentation links unten ist dabei identisch mit der Bezugsrepräsentation, und die Repräsentation rechts unten stellt ein *Diastereomer* dazu (*trans*-Isomer) dar. Die Aufgabe wurde somit falsch beantwortet, wobei jedoch nicht klar wird, warum. Einerseits könnte es sein, dass die Person die Definitionen der Arten von Isomerie nicht verstanden hat und diese nicht anwenden konnte, wofür es an früherer Stelle im Transkript Hinweise gibt. Andererseits konnte nicht ausgeschlossen werden, dass die Testperson Hindernisse hatte, die Repräsentationen mental durch Rotationsprozesse o. Ä. zu vergleichen, wie es beispielsweise bei identischen Komplexen erforderlich wäre. Dies müsste in die Aussage hineininterpretiert werden, was in der Kodierung vermieden werden sollte, um subjektive Einflüsse durch die Kodierenden zu minimieren.

In vielen Fällen ließ sich jedoch eindeutig erkennen, weshalb die Aufgaben nicht korrekt gelöst wurden bzw. weshalb Hindernisse vorlagen. Dementsprechend wurden solche Textabschnitte oft mehrfach kodiert, hauptsächlich mit den Kategorien 1.2) und 3.2).

► **1.2) Beschreiben/Anwenden fehlerhafter Fachkonzepte/-inhalte** – Beim Vergleich von Repräsentationen zeigten die Testpersonen häufig fachliche Defizite, besonders beim Verständnis von Isomerie und der korrekten Anwendung der Definitionen. Obwohl die Aufgabe mit Zusatzmaterial und notwendigen Definitionen unterstützt wurde, zeigten hier nahezu alle Testpersonen Hindernisse. Ein Beispiel war die Referenzrepräsentation und die Vergleichsrepräsentation oben links aus Aufgabe 2.1. Es handelte sich um identische Komplexe, wobei ein Komplex um 180° um die vertikale Achse rotiert dargestellt wurde, was den Eindruck von *Enantiomeren* erweckte. Dieser optische Eindruck reichte aus, dass mehr als die Hälfte der Testpersonen die Komplexe als *Enantiomere* einordnete. Zwar erschienen die Komplexe als Spiegelbilder,

ließen sich jedoch durch die beschriebene Rotation zur Deckung bringen, was bei *Enantiomeren* nicht der Fall ist. Es fiel auf, dass die Personen sich allein auf das Kriterium der Spiegelbilder beschränkten und nicht weiter prüften, inwieweit sie zur Deckung gebracht werden konnten. Dieser Fehler wurde bei nachfolgenden Komplexen ebenfalls des Öfteren identifiziert.

„Das obere Linke und das in der Mitte, das müssten Enantiomere sein. Da man die beiden/ Wenn man einen Spiegel hineinlegen würde (Hand simuliert Spiegel) ähm, wenn man eine Spiegelebene hineinzieht, dann würden die sich spiegeln, das heißt sie sind Enantiomere“ (SDSB62, Abs. 39).

Ähnliche Hindernisse zeigten sich auch bei der *Konstitutionsisomerie* und der *Konformationsisomerie*. Bei der *Konstitutionsisomerie* handelt es sich um Verbindungen mit gleicher Summenformel, aber unterschiedlicher Struktur (*Konstitutionen*). Diese Art der Isomerie kam in den Aufgaben nicht vor. Hier ging es um *Stereoisomere*, bei denen die *Konstitution* identisch ist und sich lediglich die räumliche Anordnung der Liganden unterscheidet. Genauer konnten diese auf *Enantiomere* und *Diastereomere* eingegrenzt werden, da auch *Konformationsisomere* nicht vorkamen, die durch Drehung um eine Einfachbindung ineinander überführt werden können, ohne dabei eine Bindung zu brechen.

Bei der Bearbeitung fiel auf, dass einige der Testpersonen eine unterschiedliche räumliche Anordnung der Liganden, das heißt eine *Stereoisomerie*, mit der *Konstitutionsisomerie* gleichsetzten. Unter *Konformationsisomerie* verstanden einige Testpersonen eine Drehung des gesamten Komplexes, weshalb sie identische Komplexe als *Konformationsisomere* benannten. Auch das Tauschen von Liganden wurde als *Konformationsisomerie* eingeordnet, was jedoch einen Bindungsbruch erfordert und somit der *Konfigurationsisomerie* entspricht (*Diastereomere* oder *Enantiomere*).

„Die ersten beiden Komplexe sind nicht die gleiche Struktur, die sind Isomere. Und quasi dann Konstitutionsisomere. Oder kann auch Konformation sein, weil das entsteht durch Drehung einer NH_3 -Gruppe nach unten und Chlor nach oben“ (STJS58, Abs. 21).

„Demnach würde ich sagen, dass es sich bei dem Molekül ähm links oben und bei dem in der Mitte um Konformationsisomere handelt, ähm da die

Liganden (...) eigentlich gleich angeordnet sind, nur ähm um drei Positionen gewandert sind“ (CNTR37, Abs. 39).

Ein weiterer Aspekt, der zwar nur bei wenigen Testpersonen aufkam, aber dennoch sehr interessant ist, war, dass sie die Komplexe nicht drehten, da sie davon ausgingen, dass ihre Ausrichtung im Raum nicht verändert werden dürfe. Hier zeigte sich ein fehlendes Verständnis dafür, dass es sich bei Molekülen und Verbindungen um dreidimensionale Objekte handelt, die ihre Ausrichtung im Raum ständig ändern können und dass sich der Stoff selbst dadurch nicht verändert. Durch diese Annahme konnten die Studierenden die Aufgabe nicht richtig bearbeiten.

„Ich versteh nicht, ob man das Molekül in der Mitte jetzt irgendwie drehen darf, einfach so“ (NJBS00, Abs. 21).

„[...] ich glaub Komplexe können nicht gedreht werden. Ja deswegen sagen wir das sind Konstitutionsisomere zueinander“ (STJS58, Abs. 19).

Auch bei der Benennung der Raumstruktur der Komplexe bei Aufgabe 2.1 ließen sich Hindernisse erkennen. Bei Komplexen mit vier Liganden können diese quadratisch-planar oder tetraedrisch angeordnet sein, wobei Letzteres häufiger der Fall und den Studierenden daher eher bekannt ist. Eine tetraedrische Anordnung ist darüber hinaus auch aus der Organischen Chemie und vom VSEPR-Modell (*Valence Shell Electron Pair Repulsion*) bekannt, beispielsweise vom Methanmolekül (CH_4). Daher bezeichneten einige Testpersonen den quadratisch-planaren Komplex als tetraedrisch, obwohl sie erkannten, dass dieser planar war. Bei einer Testperson fand auch eine Vermischung der Benennung der Raumstruktur mit der Isomerie statt:

„Ähm das ist vier, das schaut aus wie so ein Methan-Ding, deswegen ist das ein Tetraeder. Oder so quadratisch, sagt man doch auch, irgendwie so was [in Bezug auf den quadratisch-planaren Komplex]“ (INN75, Abs. 37).

„Das würde ich sagen ist eine *cis*-Form von der Tetraederstruktur [in Bezug auf den quadratisch-planaren Komplex]. Hm(bejahend), genau. [...] Und dann hier nach außen, das wäre hier eine *trans*-Form [in Bezug auf den tetraedrischen Komplex]“ (STJS58, Abs. 18).

► 2.1) Hindernisse bei der mentalen Manipulation von Repräsentationen –

Neben den fachlichen Aspekten spielten räumliche Fähigkeiten bei der Aufgabenlösung eine wichtige Rolle. Um zu überprüfen, ob die Komplexe identisch waren und zur Deckung gebracht werden konnten, rotierten viele Studierende die Komplexe mental. Dabei traten viele Hindernisse und Fehler auf, besonders beim Vergleich der *Enantiomere* in Aufgabe 2.4, was u. a. auf die Komplexität durch die zweizähligen Liganden zurückgeführt werden konnte. Hier erkannten nur vier Testpersonen, dass es sich um Enantiomere handelte. Die restlichen Studierenden ordneten eine andere Isomerie zu oder gaben an, dass die Komplexe identisch waren.

„Und wenn ich rechts unten mit rechts oben vergleiche, dann sind das auch die gleichen Strukturen, weil hier eben wieder die/ der eine einzählige Ligand in der axialen Ebene liegt und der andere in der äquatorialen“ (SJKM52, A2.4, Abs. 53).

Auch bei anderen Aufgaben stellten mentale Rotations- und Manipulationsprozesse Hindernisse dar. Von vielen Studierenden wurden Prozesse zum Vergleich der Komplexe beschrieben, die prinzipiell einen korrekten Lösungsansatz darstellten, aber nicht korrekt vollzogen wurden. Hervorzuheben ist aber, dass die einzelnen Testpersonen nicht immer zu einem falschen Ergebnis kamen und fast alle Personen auch korrekte Ergebnisse durch Rotations- und Manipulationsprozesse erzielten.

„Dann würde ich sagen, dass der Komplex oben links ein Enantiomer zu dem Komplex in der Mitte ist, weil wenn ich sie nebeneinander schreiben würde, verhalten die sich wie Bild und Spiegelbild zueinander und sie sind sozusagen nicht ineinander überzuführen [sie sind ineinander überführbar]“ (SJKM52, A2.2, Abs. 51).

Bei Aufgabe 2.1, in der die Testpersonen als Kugel-Stab-Modell dargestellte Komplexe in die Keilstrichformel übersetzen sollten, zeigten sich ebenfalls vereinzelt Hindernisse. Wie oben bereits beschrieben, fehlten den Studierenden häufig die richtigen Fachbegriffe, um die Struktur zu beschreiben. Bei zwei Personen wurde der quadratisch-planare Komplex als Tetraeder dargestellt (Abbildung 26). Die Personen erkannten offenbar nicht, dass die Raumstruktur der beiden Repräsentationen unterschiedlich war. Auch im Zusammenhang mit dem tetraedrischen Komplex auf der gleichen Seite fiel ihnen dieser Unterschied im räumlichen Aufbau nicht auf.

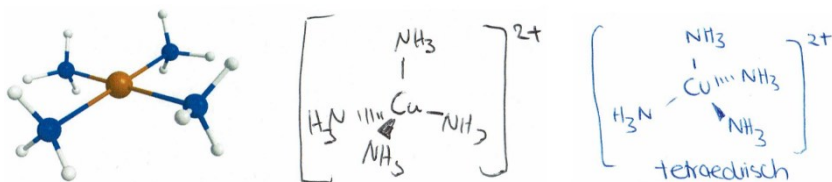


Abbildung 26: Falsche Übersetzung des als Kugel-Stab-Modell dargestellten quadratisch-planaren Komplexes $[\text{Cu}(\text{NH}_3)_4]^{2+}$ (links) in die Keilstrichformel der Testpersonen SHFS79 (mittig) und BCFH58 (rechts).

Mit Blick auf die Übersetzung der Kugel-Stab-Modelle fiel bei dem tetraedrischen Komplex zusätzlich auf, dass die Übersetzung häufig räumlich nicht korrekt wiedergegeben wurde (Abbildung 27). Während die Testpersonen die ersten beiden Komplexe rotierten, um diese gemäß den üblichen Konventionen der Keilstrichformel darzustellen, verzichteten einige der Testpersonen bei dem tetraedrischen Komplex darauf und versuchten, die räumliche Anordnung wie im Original wiederzugeben. Bei der Hälfte der Testpersonen entsprach die skizzierte Struktur keinem Tetraeder.

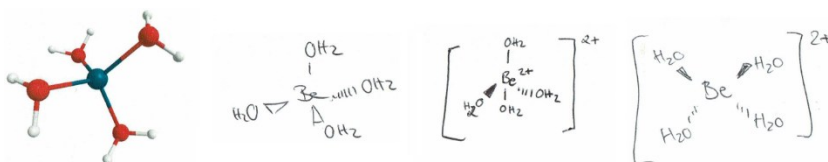


Abbildung 27: Falsche Übersetzung des als Kugel-Stab-Modell dargestellten tetraedrischen Komplexes $[\text{Be}(\text{H}_2\text{O})_4]^{2+}$ (links) in die Keilstrichformel (von links nach rechts: NWNZ30, SJMK52, ENFP62).

► **1.4) Beschreibung/Anwendung fehlerhafter Konventionen chemischer Repräsentationen** – Bei der Übersetzung der Kugel-Stab-Modelle in die Keilstrichformeln in Aufgabe 2.1 zeigte sich, dass einige Testpersonen die üblichen Konventionen nicht berücksichtigten. So koordinierten beispielsweise die Ammoniakmoleküle über das Stickstoffatom und die Wassermoleküle über das Sauerstoffatom an das Zentralion (Abbildung 27, rechte Keilstrichformel). Bei sechs der Personen wurden die eckigen Klammern und die Ladung nicht skizziert bzw. angegeben (Abbildung 27, linke Keilstrichformel), welche jedoch

notwendig sind, um die Koordinationseinheit darzustellen und um Rückschlüsse auf die Oxidationszahl des Zentralions ziehen zu können. An dieser Stelle war jedoch unklar, weshalb die Klammern und Ladungen nicht angegeben wurden. Eine Erklärung hierfür könnte das Fehlen dieser Symbole im Kugel-Stab-Modell sein. Auch fachliche Hindernisse sind denkbar, beispielsweise wenn die Ladungen nicht selbstständig auf Basis des Kugel-Stab-Modells und der entsprechenden Legende bestimmt werden konnten. Eine andere Erklärung könnte sein, dass es für die Testpersonen keine relevanten Informationen waren, auch wenn in Aufgabe 1 bereits Komplexe in Keilstrichformel mit den entsprechenden Informationen angegeben wurden.

► Interessanterweise gaben einige Testpersonen während der *Intro*- und auch während der *Retrospektion* an, dass es für sie schwieriger war, unterschiedliche Repräsentationen miteinander zu vergleichen (Kugel-Stab-Modell und Keilstrichformel) als gleichartige Repräsentationen. Sie kamen mit den Keilstrichformeln ihrer Einschätzung nach besser zurecht und übersetzten teilweise die Kugel-Stab-Modelle in die Keilstrichformel, um diese zu vergleichen – was jedoch trotzdem zu falschen Lösungen führte. Mit Blick auf die Ergebnisse konnte kein Unterschied bei der Aufgabenlösung erkannt werden: Sowohl auf individueller Ebene als auch auf globaler Ebene war die Anzahl korrekter Lösungen beim Vergleich identischer und unterschiedlicher Repräsentationen nahezu gleich.

► Generell berichteten die Testpersonen in der *Retrospektion*, dass für sie die Isomerie, trotz des Zusatzmaterials, Hindernisse bereitete. Des Weiteren erkannten einzelne, dass die Art der Liganden bzw. die Zähigkeit dieser ein schwierigkeiterzeugendes Merkmal war. Die Aufgaben 2.2 und 2.3 wurden demnach als einfacher empfunden als Aufgabe 2.4. Zudem berichteten viele Studierende, dass es ihnen schwerfiel, die Aufgaben zu lösen, da ihnen das räumliche Vorstellungsvermögen fehle und sie die Komplexe mental nicht oder nur mit äußerster Mühe rotieren konnten. Daher wünschten sich viele Studierende Molekülbaukästen bzw. haptische Modelle, um die Aufgabe zu bearbeiten. Hier behalf sich knapp die Hälfte der Studierenden während der Bearbeitung mit ihren Händen, deutete mit den Fingern die Ausrichtung der Liganden im Raum an, und drehte ihre Hände, um die Rotation zu simulieren (Hinweis: aus technischen Gründen musste bei vier Testpersonen die Audio-datei statt des Videos zur Transkription verwendet werden, da bei der Videographie technische Probleme auftraten).

7.4.3 Hindernisse bei der Konstruktion

Der Blick auf Tabelle 23 zeigt vier Hindernisse, die näher betrachtet werden müssen, da diese häufig auftraten und Schlüsselstellen bei der Bearbeitung der Konstruktionsaufgaben darstellten. Die Kategorie „3.1) Beschränkung auf Oberflächenmerkmale“ wird hier nicht weiter vertieft, da diese ausschließlich im Rahmen von Teilaufgabe a) kodiert wurde, bei der die Testpersonen die Summenformeln interpretieren sollten. Auf diese Kategorie wird bei der Interpretationsaufgabe eingegangen (Kapitel 7.3.1).

► **1.4) Beschreibung/Anwendung fehlerhafter Konventionen chemischer Repräsentationen** – In Aufgabe 1 zeigte sich bereits, dass die Testpersonen häufig nicht den Zweck von Keilstrichformeln beschreiben konnten und vereinzelt auch nicht, was ausgefüllte und gestrichelte Keile sowie einfache Striche in der Keilstrichformel bedeuteten. In Aufgabe 2.1 war auffällig, dass die Testpersonen besonders die eckigen Klammern zur Kennzeichnung der Koordinationseinheit und die Gesamtladung des Komplexes nicht darstellten. Dieser Trend setzte sich fort. Ein Drittel der Personen gab die Klammern und die Gesamtladungen bei dem geladenen Komplex $[\text{FeCl}_4]^-$ nicht an, obwohl diese in den Aufgabenstellungen zuvor entsprechend dargestellt wurden und auch im Aufgabenstamm von Aufgabe 3 die eckigen Klammern und die Gesamtladung in der Summenformel angegeben wurden. Dies erhärtet den Eindruck, dass diese Informationen von den Studierenden lediglich als sekundär und nicht relevant wahrgenommen wurden. Weiter fiel auf, dass ein Drittel bis die Hälfte der Testpersonen die Molekülliganden (H_2O und NH_3) bei den Komplexen 3 und 4 nicht über das korrekte Atom an das Zentralatom/-ion binden ließen, obwohl sie es bei Komplex 1 (NH_3) zum Großteil korrekt darstellten. Dies konnte vereinzelt auch bei der Translationsaufgabe 2.1 (Abbildung 27, rechte Keilstrichformel) erkannt werden.

► **1.2) Beschreibung/Anwendung fehlerhafter Fachkonzepte/-inhalte** – Zunächst lässt sich festhalten, dass alle Testpersonen das Zentralatom/-ion und die Liganden korrekt identifizieren konnten (*Interpretation*) und entsprechend in den konstruierten Keilstrichformeln korrekt darstellten. Vereinzelt wurden diese zwar falsch benannt (z. B. Pt als Blei oder Rh als Rhombium), dadurch wurden die Aufgaben jedoch nicht falsch gelöst. Wie in Aufgabe 2, stellte die Isomerie hier auch ein Problem dar. Die Aufgabenstellung gab vor, dass alle *Konfigurationsisomere*, d. h. *Diastereomere* und *Enantiomere*, für die vier

Komplexe skizziert werden sollten. An dieser Stelle wäre zu erwarten gewesen, dass die Testpersonen, die in Aufgabe 2 *Konstitutions-* oder *Konformationsisomere* identifizierten, stutzig werden, dass diese hier nicht skizziert werden sollen. Dies war jedoch nicht der Fall. Dies unterstreicht, dass das Verständnis der Isomerie bei vielen Testpersonen nur sehr oberflächlich vorhanden war. Auch hier führten die falschen Definitionen zu Skizzen „falscher“ Isomere. Beispielsweise wurden *Enantiomere* lediglich als Spiegelbilder charakterisiert und skizziert, obwohl diese durch Rotation ineinander überführbar waren (Abbildung 28, rechts; hier spielten ggf. auch die räumlichen Fähigkeiten eine wichtige Rolle), und *Konformationsisomere* wurden als vollständig rotierte Komplexe dargestellt (Abbildung 28, links).



Abbildung 28: Skizzierte Isomere, die auf falschen bzw. falsch interpretierten Definitionen basieren. Links ein Beispiel für *Konformationsisomerie* (SDSB66) und rechts für die *Enantiomerie* (SHFS79).

Ein weiteres gutes Beispiel ist die Ausführung von CNTR37 zum dritten Komplex. Zunächst skizzierte die Probandin korrekt das *fac*-Isomer von $[\text{RhCl}_3(\text{H}_2\text{O})_3]$. Anschließend skizzierte sie ebenfalls in richtiger Weise das *mer*-Isomer und beschrieb, dass dieses weiter rotiert werden könne, sodass dadurch insgesamt acht *mer*-Isomere entstünden. Sie beschrieb selbst, dass diese durch Rotation entstanden, betonte im Anschluss aber, dass diese alle unterschiedlich und dadurch Isomere seien. Zum einen zeigte sich hier ein inkonsequentes Vorgehen, da nach diesem Verständnis ebenfalls acht *fac*-Isomere existieren müssten; zum anderen wurde erneut impliziert, dass Komplexe bzw. Verbindungen allgemein nur eine spezifische Ausrichtung im Raum haben und entsprechend ihre Ausrichtung nicht ändern können, wie es bereits bei Aufgabe 2 von einzelnen Studierenden angedeutet wurde.

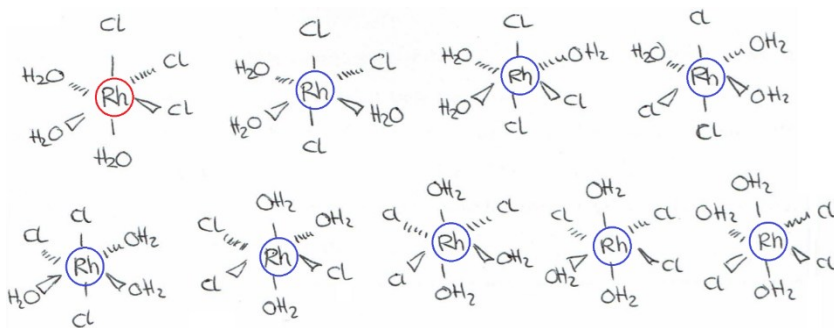


Abbildung 29: *fac*-Isomer (rot umkreist) und acht der Testperson zufolge unterschiedliche *mer*-Isomere (blau umkreist) von $[\text{RhCl}_3(\text{H}_2\text{O})_3]$, die jedoch alle identisch sind und durch Rotation ineinander überführt werden können (CNTR37, Abs. 56).

► **3.2) Hindernisse beim Vergleich multipler Repräsentationen** – Ähnlich wie in Aufgabe 2 hatten die Studierenden große Probleme dabei, identische Komplexe zu erkennen. Dementsprechend skizzierten viele mehr „Isomere“ als theoretisch vorhanden, in diesem Fall identische Komplexe mit anderer Ausrichtung im Raum. Häufig war nicht erkennbar, wo das Hindernis lag, d. h. ob beispielsweise eine falsche Auslegung einer Definition die Ursache war oder mangelnde räumliche Fähigkeiten, die es den Personen nicht ermöglichten, die Repräsentationen mental zu vergleichen. Beispielsweise skizzierten beim quadratisch-planaren $[\text{PtCl}_2(\text{NH}_3)_2]$ 13 Testpersonen korrekt das *cis*- und das *trans*-Isomer. Davon skizzierten vier Personen weitere angebliche Isomere, die sich durch Rotation jedoch ineinander überführen lassen konnten. Beim oktaedrischen $[\text{RhCl}_3(\text{H}_2\text{O})_3]$ skizzierten elf Personen korrekt das *mer*- und das *fac*-Isomer des Komplexes, davon skizzierten sieben jedoch weitere „Isomere“, die lediglich eine andere Ausrichtung im Raum wie die bereits skizzierten Komplexe aufwiesen (vgl. Abbildung 30). Daraus kann abgeleitet werden, dass die Anzahl der Liganden hier ein schwierigkeitserzeugendes Merkmal ist, wie auch die Anzahl der unterschiedlichen Liganden. Bei $[\text{RhCl}_3(\text{H}_2\text{O})_3]$ gibt es zwei unterschiedliche Liganden, während beim vierten Komplex $[\text{CrBrCl}(\text{H}_2\text{O})_3]$ vier unterschiedliche vorliegen. Hier lösten lediglich zwei Personen die Aufgabe mit insgesamt fünf unterschiedlichen Isomeren korrekt. Bei den anderen Testpersonen wurden ebenfalls identische Komplexe skizziert, es wurden aber auch mögliche Isomere nicht erkannt und fehlten dementsprechend. Zusammenfassend lässt sich sagen, dass die Studierenden zum Teil

große Probleme hatten, identische Verbindungen zu identifizieren, aber auch unterschiedliche Verbindungen zu konstruieren.

► 2.1) Hindernisse bei der mentalen Manipulation von Repräsentationen –

Ein Grund für die Hindernisse beim Vergleich der konstruierten Keilstrichformeln waren mangelnde mentale Rotations- und Manipulationsfähigkeiten. In Abbildung 30 sind die Skizzen und die dabei verbalisierten Gedanken von NJBS00 dargestellt, repräsentativ für viele Testpersonen mit diesem Hindernis. Ähnlich wie in Abbildung 29 wurde der erste Komplex hier lediglich immer weiter rotiert, wobei die Person dieses Vorgehen jedoch nicht beschrieb. Beim Skizzieren des letzten „Isomers“ fiel ihr auf, dass dieses mit dem ersten Komplex identisch war, da es durch Kippen, also eine 180°-Rotation, in den ersten Komplex überführt werden konnte. Die Person erkannte jedoch nicht, dass die anderen beiden Komplexe ebenfalls durch Rotation zur Deckung gebracht werden konnten und gab sie entsprechend falsch als Isomere an (Abbildung 30).

„Das heißt Pt in die Mitte, dann müssen alle jetzt so quadratisch da angeordnet werden. Das heißt einmal so das Chlorid nach hinten, einmal das Chlorid nach vorne, einmal das NH_3 nach vorne, einmal das NH_3 nach hinten. Genau und die beiden Elektronenpaare die muss ich ja nicht malen.“

Dann noch einmal so mit Chlorid nach hinten und das andere Chlorid auch nach hinten, NH_3 vor, NH_3 vor. Und jetzt kann eigentlich nur noch eine Möglichkeit geben oder.

Weil ich kann ja auch, ah klar ich kann ja beide NH_3 noch nach hinten machen. NH_3 nach hinten, NH_3 nochmal nach hinten, ähm Chlorid nach vorne, Chlorid nach vorne. [...]

Ich mach NH_3 nach rechts, das Chlor nach links und NH_3 nach vorne, ach nee jetzt kann ich es ja einfach kippen (drehende Handbewegung). [...] Es gibt doch nur die 3, oder? Weil das ist ja das gleiche Molekül eigentlich [in Bezug auf das den ersten Komplex], das streich ich wieder durch.“

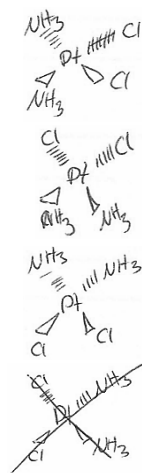


Abbildung 30: Auszug aus dem Transkript von NJBS00 (Abs. 27) mit den entsprechenden Skizzen aus dem Testheft zu den entsprechenden Textabschnitten.

7.5 Zusammenfassung und Diskussion

Forschungsanliegen I und II haben wichtige und interessante Ergebnisse zur Struktur der Repräsentationskompetenzen sowie zum Zusammenhang der lower-level skills *Interpretation*, *Translation* und *Konstruktion* mit dem deklarativen Fachwissen und den unterschiedlichen räumlichen Faktoren gegeben. Über die Definitionen der drei lower-level skills kann bereits abgeleitet werden, dass für diese unterschiedliche kognitive Prozesse und Fähigkeiten eine zentrale Rolle spielen. Dazu liefern die ersten beiden Forschungsanliegen jedoch keine Evidenz und stellen den Fokus in Forschungsanliegen III dar.

Zur Untersuchung der Forschungsfrage wurden Aufgaben zur *Interpretation*, *Translation* und *Konstruktion* von Keilstrichformeln sowie dem Kugel-Stab-Modell in der Komplexchemie konzipiert. Diese wurden von Studierenden mit der Methode des Lauten Denkens bearbeitet. Die Aufnahmen wurden inhaltlich-semantisch transkribiert und einer qualitativen Inhaltsanalyse bzw. eher einer *qualitativ orientierten kategoriengeleiteten Textanalyse* unterzogen, wobei die Kategorienbildung hauptsächlich induktiv am Material stattfand. Ziel dieser Analyse war es, Hindernisse im Bearbeitungsprozess zu identifizieren, d. h. Stellen, an denen Testpersonen lösungsrelevante Fähigkeiten nicht einsetzen konnten bzw. wollten, um darüber einen Einblick in relevante kognitive Prozesse und Fähigkeiten zu erhalten, die für die *Interpretation*, *Translation* und *Konstruktion* unerlässlich sind.

FF3.1 Welche zentralen Hindernisse lassen sich beim *Interpretieren*, *Übersetzen* und *Konstruieren* von chemischen Repräsentationen identifizieren?

FF3.2 Inwieweit unterscheiden sich die Hindernisse bzw. die kognitiven Prozesse und Fähigkeiten bei der *Interpretation*, *Translation* und *Konstruktion* von chemischen Repräsentationen?

Die induktive Kategorienbildung brachte insgesamt zwölf Kategorien hervor, die sich vier Hauptkategorien zuordnen lassen (Tabelle 22). Davon stechen insgesamt fünf Kategorien heraus und lassen sich dadurch als zentrale Hindernisse bei der Bearbeitung ansehen, die bei 70-100 % der Testpersonen auftreten (Tabelle 23). Kategorie 1.2, die *Beschreibung/Anwendung fehlerhafter Fachkonzepte/-inhalte* aufgreift, stellt sich als großes und wichtiges Hindernis

bei allen drei Aufgaben heraus. Dabei sind hauptsächlich Fachbegriffe bzw. die Definitionen von Fachbegriffen Hindernisse für die korrekte Anwendung dieser Aufgaben. Definitionen sind teilweise gar nicht bekannt, werden falschen Fachbegriffen zugeordnet oder vermischt. Des Weiteren werden sie falsch und oberflächlich bzw. nur bei bestimmten Aspekten einer Definition bei der Lösung von Aufgaben angewendet.

Bei der *Interpretation* (A1) spielt neben Kategorie 1.2 die *Beschränkung auf Oberflächenmerkmale* (Kategorie 3.2) eine maßgebliche Rolle. Dabei beschränken sich die Testpersonen vorwiegend auf unmittelbar ablesbare Merkmale der Repräsentationen, wodurch wesentliche Informationen ausgelassen werden, sofern diese nicht explizit dargestellt sind. Selbst explizit angegebene Informationen werden häufig nicht berücksichtigt. Dies zeigt sich vor allem bei der Angabe von Ionen, wobei Ladungen bzw. Oxidationszahlen bewusst weggelassen werden, die jedoch zur Differenzierung von Komplexen wichtig sind, da viele der Zentralatome/-ionen in unterschiedlichen Oxidationsstufen vorliegen können (z. B. Cr^{3+} und Cr^{6+} oder Co^{2+} und Co^{3+}) und somit unterschiedliche physikalische und chemische Eigenschaften damit einhergehen. Auffällig ist, dass bei Kugel-Stab-Modellen deskriptiv häufiger die Ladung der Ionen angegeben wird als bei der Interpretation der Keilstrichformeln.

Bei der *Translation* und der *Konstruktion* ergeben sich neben der Hinderniskategorie 1.2 drei weitere relevante Hindernisse. Beim Skizzieren von Keilstrichformeln zeigt sich bei beiden Aufgaben, dass diese von vielen Testpersonen nicht konventionsgemäß skizziert werden (Kategorie 1.4) – vornehmlich fehlen die Gesamtladung und die eckigen Klammern zur Kennzeichnung der Koordinationseinheit.

Auch bei der *Interpretation* ist bei einigen Personen ein Hindernis identifizierbar: Sie können eine systematische Darstellung von Verbindungen nicht als Ziel der Keilstrichformel benennen und teilweise nicht angeben, dass es sich bei Keilen und Strichen in der Keilstrichformel ausschließlich um kovalente Bindungen handelt. Teilweise geben die Testpersonen an, dass es ionische Bindungen sind (vor allem zwischen Chlorid und Zentralion). Bei nahezu allen Personen tritt ein *Hindernis beim Vergleichen multipler Repräsentationen* auf (Kategorie 3.2). Dabei wird häufig nicht klar, welche Ursache das Hindernis hat. Fakt ist jedoch, dass die Testpersonen Hindernisse haben, identische Komplexe und unterschiedliche Komplexe zu identifizieren, unabhängig von der

Angabe der Isomerie bei unterschiedlichen Komplexen. Bei der Bestimmung der Isomerie ist Kategorie 1.2 das größte Hindernis. Daneben tritt häufig ein Hindernis bei der mentalen Manipulation chemischer Repräsentationen (Kategorie 2.1) auf. Sowohl die mentale Rotation als auch die mentale Manipulation wird von vielen Testpersonen nicht korrekt vollzogen, wodurch sie zu falschen Schlüssen kommen und die Aufgabe nicht korrekt lösen können.

Die restlichen Kategorien treten eher vereinzelt auf und stellen somit vornehmlich individuelle Hindernisse dar, die zwar auch interessante Einblicke in die Bearbeitungsprozesse geben, zum Beantworten der Fragestellung jedoch nicht beitragen. Des Weiteren sind viele Hindernisse, besonders aus Hauptkategorie ④ (*Sonstige Hindernisse*), keine aufgabenbezogenen Hindernisse und tragen somit nicht zur Beantwortung der Forschungsfragen bei. Der Vollständigkeit halber werden die Hindernisse jedoch kodiert und für die Transparenz im Kategoriensystem (Tabelle 22) sowie der Übersicht der Kodierungen (Tabelle 23) mit angegeben.

Die in dieser Studie identifizierten Hindernisse decken sich mit den Ergebnissen anderer Untersuchungen. Die *Beschränkung auf Oberflächenmerkmale* (Kategorie 3.2) ist ein gut belegtes Hindernis beim Bearbeiten von externen Repräsentationen – vor allem bei Noviz:innen (Ainsworth, 2006). Die Testpersonen gehen kaum auf tiefgehende Strukturen, d. h. implizit dargestellte Informationen ein und beziehen sich vornehmlich auf explizit dargestellte Merkmale. Dies lässt sich hier beispielsweise daran erkennen, dass Zentralionen und ionische Liganden bei den Kugel-Stab-Modellen deskriptiv häufiger konkret angegeben werden als bei den Keilstrichformeln, da die Informationen dort explizit in der Legende dargestellt sind (Ainsworth, 2006; Graulich, 2015; Rodemer et al., 2021; Talanquer, 2022).

Die *Beschreibung/Anwendung fehlerhafter Fachkonzepte/-inhalte* macht deutlich, dass fehlendes Fachwissen ein wesentliches Hindernis beim Bearbeiten chemischer Repräsentationen ist, das überwunden werden muss, was sich u. a. im Repräsentationsdilemma widerspiegelt (Kapitel 2.2.3; z. B. Rau, 2018). Nur wer zentrale Begriffe definieren und anwenden kann, ist in der Lage, mit Repräsentationen zu arbeiten, was beispielsweise Talanquer (2022) betont. Darin liegt ein zweiter wichtiger Aspekt. Deklaratives und prozedurales Fachwissen müssen miteinander verknüpft werden, um Probleme erfolgreich zu

lösen. Bei kognitiv weniger anspruchsvollen Aufgaben, beispielsweise der Definition (deklaratives Fachwissen) und Bestimmung (prozedurales Fachwissen) der Koordinationszahl eines Komplexes, gelingt diese Integration häufig. Die Definition und Bestimmung der Isomerie ist dagegen kognitiv anspruchsvoller. Hier müssen verschiedene Definitionen voneinander abgegrenzt werden. Darüber hinaus sind die einzelnen Isomeriebegriffe durch mehrere spezifische Kriterien charakterisiert, die bei der Überprüfung systematisch berücksichtigt werden müssen. Dies bewerkstelligen viele Testpersonen nicht. Andererseits gelingt die Anwendung und somit die Integration von deklarativem und prozeduralem Fachwissen nicht adäquat.

Die Integration von deklarativem und prozeduralem Fachwissen ist somit von zentraler Bedeutung, was u. a. auch Lutter et al. (2019) hervorheben. Im Rahmen anderer Studien konnte gezeigt werden, dass in der universitären Lehre maßgeblich deklaratives Fachwissen vermittelt wird, während das prozedurale Fachwissen vernachlässigt wird (z. B. Friege & Lind, 2006; Solaz-Portolés & López, 2008). Eine Integration kann somit kaum stattfinden, womit sich die hier gefundenen Ergebnisse erklären und begründen lassen. Diese Hypothese ist jedoch vorläufig und bedarf weiterer Untersuchungen. Eindeutig ist, dass die Studierenden Defizite in Bezug auf das prozedurale Fachwissen haben – es fehlt ihnen an Strategien, um die Aufgaben erfolgreich zu bearbeiten, was beispielsweise auch Ward et al. (2022) in einer vergleichbaren Studie zu Denkprozessen bei der *Interpretation, Translation, Konstruktion* und *Nutzung* von Keilstrichformeln und Newman-Projektionen in der Organischen Chemie zeigen konnten.

Des Weiteren fällt auf, dass die Testpersonen schwache und nicht konsequente Argumentationsstrategien nutzen, beispielsweise bei der Bestimmung von Enantiomeren. Sie treffen vorschnell Entscheidungen, wenn etwa Spiegelbilder erkannt werden und die Prüfung auf Überführbarkeit der Komplexe übersprungen wird. Diese heuristische Strategie konnten auch Maeyer und Talanquer (2013) in ihrer Studie identifizieren und bezeichnen dieses Vorgehen als *One-Reason Decision Making*. In Abbildung 29 ist ein gutes Beispiel für nicht konsequente und nicht schlüssige Argumentationen dargestellt, was sich u. a. auf die fehlende Integration von deklarativem und prozeduralem Wissen zurückführen lässt.

Neben dem Fachwissen haben sich mentale Manipulations- und Rotationsprozesse als Hindernisse beim Bearbeiten herausgestellt, was beispielsweise auch Harle und Towns (2011) in ihrem Review zu räumlichen Fähigkeiten in der Chemie unterstreichen. Hier sind diese Probleme vor allem bei dem Vergleich von Repräsentationen relevant. Die Befunde decken sich mit den Ergebnissen anderer Studien. Ähnlich wie bei Stieff und Kolleg:innen (2012) können verschiedene Strategien zur Bewältigung der Aufgaben beobachtet werden.

Ein Großteil der Testpersonen nutzt die *spatial-imagistic strategy*, d. h. mentale Manipulations- und Rotationsprozesse, um zu überprüfen, ob es sich bei den Strukturen um identische Komplexe oder Isomere handelt. Vereinzelt wird auch die *spatial-diagrammatic strategy* genutzt, d. h. Kugel-Stab-Modelle werden zunächst in die Keilstrichformel übersetzt und skizziert, ehe es zum Vergleich der Strukturen kommt. Seltener wird die *spatial-analytic strategy* verwendet, bei der strukturelle Merkmale ohne Rotations- oder Manipulationsprozesse verglichen werden. Daneben werden von vielen Testpersonen die Hände genutzt, um die mentalen Prozesse zu unterstützen, was als Indikator für schwach ausgeprägte räumliche Fähigkeiten bzw. Manipulations- und Rotationsprozesse zu werten ist (Göksun et al., 2013; Goldin-Meadow et al., 2012; Hostetter & Alibali, 2008). Diese Befunde sind für Noviz:innen ebenfalls bekannt. Sie nutzen hauptsächlich die *spatial-imagistic strategy*, wohingegen Expert:innen eher analytische Strategien anwenden (Stieff, 2007; Stieff et al., 2012).

Mit Blick auf die *Konstruktion* fällt in dieser Studie auf, dass die Kategorie zwar verwendet wird, aber weitaus weniger häufig als bei der *Translation*. Dennoch spielen hier mentale Manipulations- und Rotationsprozesse eine bedeutende Rolle, da für die *Konstruktion* neuer und zu den bereits skizzierten Repräsentationen distinkter Komplexe (Isomere) skizziert werden müssen. Diese sind mental zu vergleichen, zu manipulieren und zu rotieren, um die Aufgabe erfolgreich zu lösen. Die wenigen Kodierungen können hier u. a. auf die Methode des Lauten Denkens zurückgeführt werden. Denn auch wenn die Testpersonen ihre Gedanken verbalisieren sollen, kann es sein, dass diese nicht vollständig wiedergegeben werden.

Auch die nicht konventionsgemäße Beschreibung und Skizzierung von Repräsentationen liefert hier interessante und wichtige Einblicke. Das systematische Auslassen wichtiger Bestandteile von Repräsentationen (hier die eckigen

Klammern und Ladungen der Komplexe) zeigt, dass ein Drittel der Testpersonen diesen Merkmalen keine Relevanz zuordnet bzw. zuordnen kann. Daraus lässt sich die Hypothese ableiten, dass Studierende im Laufe ihres Studiums zwar mit den Repräsentationen arbeiten und grundlegende Repräsentationsmerkmale beschreiben, die Zweckmäßigkeit und die Relevanz einzelner Merkmale jedoch nicht erläutern können. Diese Fähigkeiten lassen sich u. a. den higher-level skills zuordnen (Tabelle 1). Auch in anderen Studien wird berichtet, dass Testpersonen zwar Repräsentationen kennen, ihre Bedeutung, die Zweckmäßigkeit und den Zusammenhang mit anderen Repräsentationen aber nicht erörtern können. Dies lässt sich oft auf die Lehre zurückführen, in der Repräsentationskompetenzen nicht gezielt vermittelt werden bzw. ein Fokus auf den lower-level skills liegt, wenn diese adressiert werden (z. B. Popova & Jones, 2021).

Mit Blick auf die Retrospektion und die Reflexion der Aufgabenbearbeitung lässt sich erkennen, dass sich die Testpersonen der Hindernisse weitgehend bewusst sind, sie jedoch vornehmlich ihr deklaratives Wissen auffrischen würden, um die Aufgaben zu bearbeiten. Ein Nachholbedarf beim prozeduralen Fachwissen wird kaum angegeben. Interessant ist dabei, dass sich die Testpersonen häufig haptische Modelle bzw. Molekülbaukästen zur Bearbeitung der Translations- und Konstruktionsaufgaben wünschen. Stieff und Kolleg:innen (2012) können hierfür Evidenz liefern, dass der Einsatz solcher Modelle zielführend ist und zu einer besseren Leistung bei der Bearbeitung räumlicher Probleme führen kann.

Wie bereits beschrieben, unterscheiden sich die Hindernisse bei den drei lower-level skills. Während Fachwissen für alle drei Fähigkeiten von zentraler Bedeutung ist, lassen sich fast ausschließlich bei der *Interpretation* Hindernisse durch eine Beschränkung auf Oberflächenmerkmale identifizieren. Dieses Hindernis beeinflusst jedoch nicht die *Translation* bzw. *Konstruktion* von Repräsentationen, was beispielsweise auch Ward et al. (2022) in ihrer Studie festhalten. Dagegen sind Hindernisse beim Vergleich und der mentalen Manipulation und Rotation sowie das Wissen und Anwenden von repräsentationsbezogenen Konventionen fast ausschließlich bei der *Translation* und *Konstruktion* zu identifizieren.

Diese Befunde decken sich auch mit anderen Studien (z. B. Stieff et al., 2018; Stull et al., 2012) und lassen sich bereits aus den Definitionen und Anforderungen an die *Translation* ableiten (Tabelle 1). Daraus lässt sich schließen, dass die notwendigen Fähigkeiten und kognitiven Prozesse für die *Interpretation*, *Translation* und *Konstruktion* zwar einige Parallelen aufweisen, sich aber in vielen notwendigen Fähigkeiten und Prozessen auch unterscheiden, was auch Ward et al. (2022) aus ihren Ergebnissen folgern. Ebenso spielt die Art der Repräsentation eine Rolle dabei, d. h. auch unterschiedliche Repräsentationen erfordern unterschiedliche Denkprozesse und Fähigkeiten zur Bearbeitung. Die zentralen Befunde und die Einordnung einschlägiger empirischer Befunde in diesem Kontext sind zusammengefasst Abbildung 31 zu entnehmen.

Letztlich muss ein Blick auf den Kontext der Komplexchemie geworfen werden. Eine Vielzahl der hier herangezogenen Studien fokussiert auf die Organische Chemie (z. B. Graulich, 2015; Stieff, 2007; Stieff et al., 2018; Stull et al., 2016; Ward et al., 2022, 2025), da sie sich einer Vielzahl unterschiedlicher Repräsentationen mit teilweise hohen räumlichen Ansprüchen bedient. Die Komplexchemie steht dagegen in diesem Kontext kaum im Fokus der Forschung, obwohl sie ebenso unterschiedliche Repräsentationen mit hohen Anforderungen an das räumliche Denken nutzt (Dai et al., 2020; Eticha et al., 2023). Die Ergebnisse der vorliegenden Studie sind insofern wichtig und interessant, als sich die Ergebnisse aus den Studien mit Bezug zur Organischen Chemie nahezu identisch auf die Komplexchemie übertragen lassen (Abbildung 31). Somit kann hier gefolgert werden, dass sich die kognitiven Prozesse bei der *Interpretation*, *Translation* und *Konstruktion* von chemischen Repräsentationen sowie die mentalen Rotations- und Manipulationsprozessen in der Organischen Chemie und der Komplexchemie nicht unterscheiden und für beide Fachbereiche von hoher Bedeutung sind.

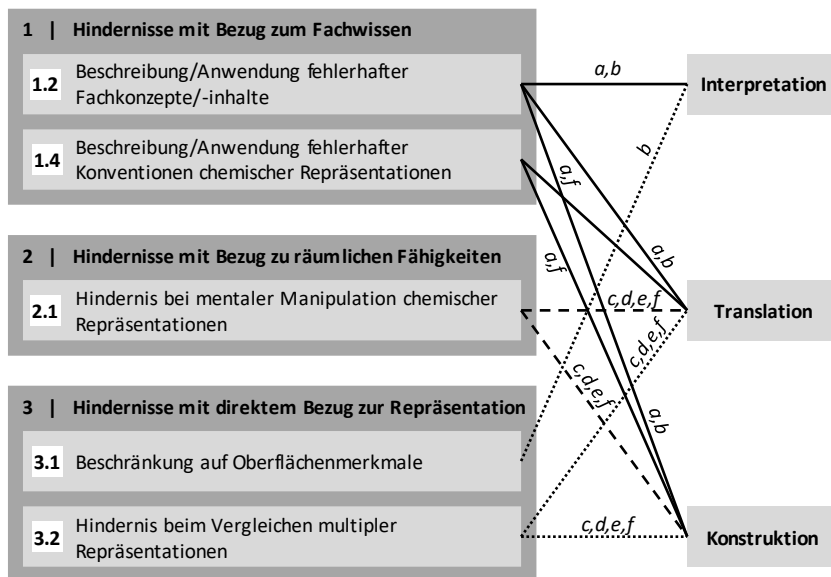
^a Talanquer (2022)^c Stieff (2007)^e Stull et al. (2018)^f Stieff et al. (2018)^b Ward et al. (2022)^d Stieff et al. (2012)

Abbildung 31: Zentrale Hindernisse bei der *Interpretation*, *Translation* und *Konstruktion* von repräsentationsbasierten Komplexchemieaufgaben in Forschungsanliegen III und Einordnung einschlägiger empirischer Befunde (vornehmlich aus der Organischen Chemie) in diesem Kontext. Studien mit allgemeineren Aussagen zu den aufgeführten Hindernissen in Bezug auf Repräsentationen und nicht spezifisch für die *Interpretation*, *Translation* und *Konstruktion* werden hier nicht aufgenommen.

8

Integration der Forschungsanliegen

Unter den Naturwissenschaften ist besonders die Chemie vom Gebrauch (multipler) externer Repräsentationen geprägt. Dadurch wird die submikroskopische Ebene der Atome, Moleküle und Ionen visualisiert, die sonst nicht erfassbar ist. Somit dienen Repräsentationen der Vermittlung von und der Kommunikation über chemische Konzepte. Das impliziert jedoch ein Repräsentationsdilemma: Unbekannte Fachkonzepte müssen mit unverständlichen Repräsentationen erlernt werden – das Lernen *mit* Repräsentationen steht also dem Lernen *über* Repräsentationen gegenüber.

Um fachliche Inhalte mit Repräsentationen zu lernen, müssen Fähigkeiten erworben werden, die es ermöglichen, erfolgreich mit Repräsentationen umzugehen. Kozma und Russell (1997, 2005) haben im Rahmen ihrer Noviz:innen-Expert:innen-Vergleiche entsprechende Fähigkeiten abgeleitet, die sich unter dem Begriff Repräsentationskompetenzen zusammenfassen lassen. Diese werden weiter differenziert (Gurung et al., 2022; Nitz, 2012): Auf der einen Seite stehen die lower-level skills, die die *Interpretation*, *Translation* und *Konstruktion* von Repräsentationen umfassen. Diese bilden die Grundlage für das Arbeiten mit chemischen Repräsentationen und werden auch von anderen Autor:innen als zentral in diesem Kontext angesehen (Gilbert, 2005, 2008). Auf der anderen Seite stehen die higher-level skills, die den kritischen Vergleich, die epistemologische Betrachtung von Repräsentationen und deren Nutzen in Diskursen umfassen und daher auch als metarepresentational skills bezeichnet werden (Tabelle 1).

Um die higher-level skills zu entwickeln, müssen zunächst die lower-level skills erworben werden (Gurung et al., 2022), weshalb sie in dieser Arbeit im Fokus stehen. Im Kontext des Lernens *mit* und *über* Repräsentationen gibt es aktuell jedoch noch ungeklärte Fragen und teilweise gegensätzliche Befunde zur Struktur und Differenzierbarkeit der lower-level skills *Interpretation*, *Translation* und *Konstruktion* (FF1.1) sowie zur empirischen Abgrenzung von Repräsentationskompetenzen und Fachwissen (FF1.2). Diese Fragen stellen den Gegenstand des ersten Forschungsanliegens dar.

Zur Untersuchung des ersten Forschungsanliegens wurde das *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC) entwickelt. Das Instrument misst die Fähigkeiten der *Interpretation*, *Translation* und *Konstruktion* differenziert und weitgehend unabhängig vom Fachwissen. Diese Aspekte werden von bestehenden Messinstrumenten zum Zeitpunkt der Entwicklung des CRI:TIC nicht ausreichend abgedeckt, weshalb diese Instrumente in ihrer ursprünglichen Form für das Forschungsanliegen ungeeignet sind. Einzelne Items aus diesen Instrumenten wurden jedoch im Rahmen der Entwicklung des CRI:TIC adaptiert. Zur Messung wurden geschlossene und halboffene Items mit symbolisch-mathematisch und visuell-graphischen Repräsentationen eingesetzt, daneben Multiple Choice Items mit rein textlichen Repräsentationen zur Messung des deklarativen Fachwissens. Das CRI:TIC und die Fachwissensitems wurden in einem Chemievorkurs für unterschiedliche Studiengänge mit $N = 185$ Studierenden verwendet. Die gewonnenen Daten waren mithilfe eines (multidimensionalen) Partial-Credit-Rasch Modells auszuwerten, das die unterschiedlich skalierten Items berücksichtigt und Itemschwierigkeiten sowie Personenfähigkeiten auf derselben linearen Skala abbildet.

Die Evaluation des CRI:TIC zeigt nach Ausschluss von fünf Items mit schlechter Passung zum Modell, inhaltlichen Diskrepanzen oder einer zu hohen Itemschwierigkeit gute Item- und Modellfit Statistiken, die das CRI:TIC als geeignetes Instrument zur Messung der Repräsentationskompetenzen legitimieren (final: 33 Items). Gleiches gilt für die Fachwissensitems nach Ausschluss von zwei Items mit zu hoher Schwierigkeit (final: 16 Items). Der statistische und graphische Vergleich dreier theoretisch abgeleiteter Modelle zur Trennbarkeit der *Interpretation*, *Translation* und *Konstruktion* zeigt ein ambivalentes und uneindeutiges Bild. Aus diesem Grund wird in dieser Studie das einfachste Modell angenommen, das die drei lower-level skills als Fähigkeiten eines eindimensionalen Konstrukts betrachtet, die stark interkorreliert sind. Dieses Modell wurde anschließend dem deklarativen Fachwissen gegenübergestellt und ein zusammenfassendes eindimensionales Modell mit einem zweidimensionalen Modell, das Repräsentationskompetenzen und Fachwissen als getrennte Konstrukte behandelt, verglichen. Dieser Vergleich zeigt ein eindeutiges Bild, das die beiden Konstrukte als empirisch trennbar erscheinen lässt. Diese Ergebnisse lassen sich mit einer zweiten, größeren Stichprobe mit

$N = 251$ Studierenden aus drei Chemievorkursen replizieren (die Daten stammen aus dem zweiten Forschungsanliegen).

Neben Repräsentationskompetenzen und Fachwissen haben sich räumliche Fähigkeiten als wichtige Prädiktoren in den Naturwissenschaften und vor allem der Chemie herausgestellt. Diese lassen sich im Cattell-Horn-Carroll Modell als ein Faktor von Intelligenz (*g*-Faktor) beschreiben (Abbildung 8), auf den eine Vielzahl an räumlichen Faktoren lädt (Tabelle 2). Besonders die Fähigkeit zur zweidimensionalen (*Speeded Rotation*) und dreidimensionalen (*Spatial Relation*) mentalen Rotation, zu mentaler Manipulation (*Visualization*) und zur Erkennung von Figuren in komplexen Mustern (*Flexibility of Closure*) haben sich in der Chemie und im Kontext des Arbeitens mit chemischen Repräsentationen als wichtige räumliche Faktoren herausgestellt (Bodner & Guay, 1997; Bodner & McMillen, 1986; Carter et al., 1987; Harle & Towns, 2011; Harris et al., 2013; Hodgkiss et al., 2018; Sorby, 2009; Wai et al., 2009; Wu & Shah, 2004). Aktuell fehlt es jedoch an Studien, die eine hohe Bandbreite an räumlichen Faktoren berücksichtigen, um den korrelativen Zusammenhang zwischen Repräsentationskompetenzen bzw. deklarativem Fachwissen und den räumlichen Faktoren (FF2.1) zu untersuchen. Des Weiteren muss die prädiktive Wirkung einzelner räumlicher Faktoren im Kontext mit anderen räumlichen Faktoren untersucht werden, da die räumlichen Fähigkeiten ein Zusammenspiel verschiedener räumlicher Faktoren darstellen (FF2.2). Diese Untersuchungen stellen den Fokus in Forschungsanliegen II dar.

Zur Untersuchung der Fragestellungen wurden das CRI:TIC und die Fachwissensitems aus Forschungsanliegen I sowie acht standardisierte, validierte psychometrische Instrumente zur Messung acht unterschiedlicher räumlicher Faktoren eingesetzt. Letztere wurden im Gegensatz zu den anderen Instrumenten mittels klassischer Testtheorie gemäß den Manualen ausgewertet. An der Datenerhebung nahmen insgesamt $N = 494$ Studierende teil, wobei die Datenerhebung über mehrere Tage stattfand und somit nicht alle Testhefte von allen Studierenden bearbeitet wurden. Mit Ausnahme des *Gestalt Completion Tests* zur Messung der Fähigkeit, unvollständige Bilder zu erkennen (*Closure Speed*), zeigen die restlichen sieben Tests zur Messung der räumlichen Faktoren eine angemessene interne Konsistenz ($\alpha \geq 0.79$). Da auch die Validität des *Gestalt Completion Tests* infrage gestellt werden muss (Verzerrung der Messung durch die Verwendung kulturell behafteter und nicht zeit-

gemäßer Piktogramme), wurde der Test aus den weiteren Analysen ausgeschlossen. Die Korrelationsanalysen (Spearman Rangkorrelationskoeffizient ρ) mit Bonferroni-Korrektur zeigen eine schwache bis starke Korrelation ($0.204 \leq \rho \leq 0.529$) aller räumlicher Faktoren mit den Repräsentationskompetenzen. Dagegen zeigt sich beim Fachwissen lediglich mit der mentalen zweidimensionalen Rotationsfähigkeit (*Speeded Rotation*) eine schwache Korrelation ($\rho = 0.226$).

In einem Pfadmodell wurden anschließend die sieben räumlichen Faktoren und das Fachwissen als Prädiktoren für Leistung beim Bearbeiten repräsentationsbasierter Aufgaben modelliert. Dadurch können die Interkorrelationen zwischen den räumlichen Faktoren berücksichtigt und Variablen, die einen einzigartigen Beitrag zur Varianzaufklärung leisten, identifiziert werden. Das Modell hat eine gute Passung und zeigt, dass das Fachwissen den höchsten standardisierten Pfadkoeffizienten aufweist ($\beta = 0.533$). Daneben zeigen die zwei- und dreidimensionale Rotationsfähigkeit (*Speeded Rotation* und *Spatial Relation*) sowie die mentale Manipulationsfähigkeit (*Visualization*) signifikante bzw. marginal signifikante Pfade ($0.096 \leq \beta \leq 0.352$; Abbildung 22).

In einem zweiten Modell werden die drei lower-level skills *Interpretation*, *Translation* und *Konstruktion* getrennt voneinander betrachtet. Hier fällt auf, dass bei der *Interpretation* und *Konstruktion* das Fachwissen den höchsten standardisierten Pfadkoeffizienten aufweist und zusätzlich die mentale dreidimensionale Rotationsfähigkeit sowie die mentale Manipulationsfähigkeit signifikant werden. Bei der *Translation* weist die mentale dreidimensionale Rotationsfähigkeit den höchsten Pfadkoeffizienten auf, gefolgt vom Fachwissen und der mentalen zweidimensionalen Rotationsfähigkeit (Abbildung 23). Zwar sind diese Ergebnisse aufgrund der hohen Anzahl geschätzter Modellparameter bei einer Stichprobe von $N = 494$ Studierenden mit Vorsicht zu interpretieren, zeigen jedoch, dass für die *Interpretation*, *Translation* und *Konstruktion* möglicherweise unterschiedliche kognitive Prozesse und Fähigkeiten benötigt werden, was sich auch aus den Definitionen der lower-level skills ableiten lässt.

Aus der psychometrischen Perspektive liegen Evidenzen zur Annahme der lower-level skills als eindimensionales Konstrukt vor – *Interpretation*, *Translation* und *Konstruktion* stellen stark interkorrelierte, nicht trennbare Fähigkei-

ten dar. Dagegen lässt sich das Fachwissen von den Repräsentationskompetenzen in dieser Studie und mit der hier vorliegenden Operationalisierung empirisch abgrenzen. Aus der fachdidaktisch-inhaltlichen Perspektive ist eine differenzierte Betrachtung dennoch sinnvoll, besonders im Kontext von Lehr-Lern-Prozessen. Der Einbezug der räumlichen Fähigkeiten in Forschungsanliegen II unterstützt diese Perspektive, da es Hinweise darauf gibt, dass für die drei lower-level skills verschiedene kognitive Prozesse und Fähigkeiten eine Rolle spielen, wobei diese unterschiedlich stark ins Gewicht fallen. Daher wird in Forschungsanliegen III eine qualitative Studie zur vertieften Untersuchung der drei lower-level skills durchgeführt, um die kognitiven Prozesse und Fähigkeiten bei der *Interpretation*, *Translation* und *Konstruktion* genauer zu untersuchen. Diese sollen über die Identifikation von Hindernissen im Bearbeitungsprozess von Repräsentationen abgeleitet werden, da Hindernisse eindeutiger zu operationalisieren sind und kognitive Engpässe im Bearbeitungsprozess darstellen, an denen Fähigkeiten zur Lösung der Aufgabe fehlen. In diesem Zusammenhang stellen sich die Fragen, welche zentralen Hindernisse bei der Bearbeitung repräsentationsbasierter Chemieaufgaben mit Fokus auf die lower-level skills auftreten (FF3.1) und inwieweit sich die Hindernisse bzw. die damit verbundenen kognitiven Prozesse und Fähigkeiten bei der *Interpretation*, *Translation* und *Konstruktion* von chemischen Repräsentationen unterscheiden (FF3.2).

Bei der Bearbeitung von Komplexchemieaufgaben zur *Interpretation*, *Translation* und *Konstruktion* von Keilstrichformeln und Kugel-Stab-Modellen mit der Methode des Lauten Denkens haben sich insgesamt zwölf Hindernisse induktiv identifizieren lassen. Davon spielen fünf Hindernisse eine zentrale Rolle und treten bei 70 % bis 100 % der Testpersonen auf. Dabei fällt auf, dass sich die Hindernisse bei den drei Fähigkeiten teilweise unterscheiden, woraus sich folgern lässt, dass die kognitiven Prozesse und Fähigkeiten bei der *Interpretation*, *Translation* und *Konstruktion* von Repräsentationen unterschiedlich sind.

Aber auch bei der Arbeit mit unterschiedlichen Repräsentationen unterscheiden sich die Denkprozesse der Testpersonen. Zwei Hindernisse, die bei allen drei lower-level skills auftreten, sind mangelndes, falsches oder lückenhaftes Fachwissen (Kategorie 1.2) bzw. Wissen über Konventionen chemischer Repräsentationen (Kategorie 1.4), wobei letzteres maßgeblich bei der *Translation* und *Konstruktion* auftritt.

Nahezu ausschließlich bei der *Interpretation* spielt die Beschränkung auf Oberflächenmerkmale (Kategorie 3.1) eine zentrale Rolle, d. h. die Testpersonen gehen nur auf explizite, nicht aber auf implizit dargestellte Merkmale und Informationen in den Repräsentationen ein, wodurch wichtige und lösungsrelevante Inhalte nicht berücksichtigt werden. Bei der *Translation* und *Konstruktion* stellt vor allem der Vergleich multipler externer Repräsentationen (Kategorie 3.2) ein zentrales Hindernis dar, wobei teilweise nicht erkennbar ist, welche Art von Hindernis bei den Personen genau vorliegt. Ein Großteil dieser Hindernisse kann jedoch mit der Ausprägung des Fachwissens erklärt werden, während sich ein weiterer Teil durch gering ausgeprägte mentale Manipulations- und Rotationsfähigkeiten (Kategorie 2.1) erklären lässt.

8.1 Integrative Gesamtdiskussion und Limitationen

Die Ergebnisse der drei Forschungsanliegen werden in den jeweiligen Kapiteln ausführlich dargestellt, analysiert und unter Einbezug einschlägiger, kontextnaher Studien vergleichend diskutiert sowie theoretisch eingeordnet. Im Folgenden erfolgt eine zusammenführende Diskussion der drei Forschungsanliegen, mit dem Ziel, ein integratives Fazit für die vorliegende Arbeit abzuleiten.

Forschungsanliegen I hat gezeigt, dass die lower-level skills *Interpretation*, *Translation* und *Konstruktion* im Rahmen dieser Studie und bei der zugrunde gelegten Operationalisierung eher als stark interkorrelierte Fähigkeiten eines eindimensionalen Konstrukts anzunehmen sind. Die empirischen Befunde zeigen ein ambivalentes Bild, das weder die Annahme noch die Ablehnung eines der geprüften Modelle zulässt, weshalb hier das einfachste Modell, das eindimensionale Modell, angenommen wird. Die Ergebnisse treten nicht nur im Kontext des ersten Forschungsanliegens ($N = 185$) auf, sondern lassen sich darüber hinaus im Datensatz des zweiten Forschungsanliegens ($N = 251$) in nahezu identischer Weise replizieren, was auf eine hohe Übertragbarkeit der Befunde hinweist.

Im Zusammenhang mit anderen Studien in diesem Kontext und aus weiteren naturwissenschaftlichen Domänen sind die Ergebnisse vergleichbar (z. B. Scheid et al., 2018; Taskin et al., 2015). Besonders die Studie von Ward et al. (2025) liefert mit einem vergleichbaren, wenn auch methodisch anderen Vorgehen, nahezu identische Ergebnisse: Die lower-level skills stellen mit hoher

Evidenz stark interkorrelierte Fähigkeiten eines eindimensionalen Konstrukts dar. In diesem Zusammenhang müssen aber auch die Unterschiede der Studien betont werden, die sich vor allem (a) bei der Operationalisierung der lower-level skills, (b) der Konstruktion der Items und (c) der Abgrenzung vom Fachwissen zeigen.

- (a) Die von Kozma und Russell (1997, 2005) abgeleiteten Fähigkeiten sind relativ unpräzise und dadurch nicht trennscharf definiert (Tabelle 1). Daher werden sie im Rahmen dieser Studien mit der Kategorisierung von Repräsentationen (Abbildung 2) synthetisiert, um sie eindeutiger zu operationalisieren (Tabelle 3). Dadurch können Items mit hoher Präzision den Fähigkeiten zugeordnet werden ($K_{\text{Fleiss}} = 0.86$). Die offenen Definitionen von Kozma und Russell (1997, 2005) lassen einen großen Spielraum für die Auslegung der Fähigkeiten, was sich im Vergleich mit den anderen Studien zeigt. Ward et al. (2022, 2025) verstehen beispielsweise unter der *Translation* reine Selektionsaufgaben, während in diesem Projekt auch das Skizzieren von neuen Repräsentationen dazugehört, sofern der Abstraktionsgrad vergleichbar ist und die Referenzrepräsentation an sich nicht verändert wird (Tabelle 3). Global betrachtet lässt sich jedoch festhalten, dass hinsichtlich der Kategorisierung und Definition der lower-level skills eine weitgehende Übereinstimmung besteht, wodurch ein Vergleich der Studienergebnisse grundsätzlich legitimiert wird. Gleichzeitig zeigen sich jedoch Unterschiede in der konkreten Operationalisierung. Vor diesem Hintergrund erscheint eine konsensorientierte Klärung der Kategorien und Definitionen auf Grundlage der Arbeiten von Kozma und Russell (1997, 2005) erstrebenswert.
- (b) Das *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC) nutzt zur Messung der Fähigkeiten neben Items im Multiple Choice: Single Select Format auch halboffene Items, wobei beispielsweise Kurzantworten gegeben oder Skizzen von Strukturformeln angefertigt werden müssen. Diese Kombination erscheint adäquat, da der Umgang mit Repräsentationen auch das Anfertigen von Skizzen umfasst, wie es in den Repräsentationskompetenzen unschwer herauszulesen ist (Tabelle 1). Ähnlich haben beispielsweise auch Taskin et al. (2015) das *Chemical Representation Inventory* und Sim und Daniel (2014) den *Test of Representational Competence* konzipiert, aus denen hier u. a. Items adaptiert wurden. Kritisch zu hinterfragen sind die unterschiedlichen kognitiven

Anforderungen der unterschiedlichen Aufgabenformate: Multiple Choice: Single Select Aufgaben beruhen eher auf der Reproduktion von Wissen, während halboffene Aufgaben kognitiv anspruchsvollere Anforderungen an die Testpersonen stellen. Daher ist in diesem Kontext das Partial-Credit-Rasch Modell (Masters, 1982) eine geeignete Grundlage für die Datenauswertung, da dieses Vorgehen Itemschwierigkeiten und Personenfähigkeiten auch bei unterschiedlich skalierten Items auf der gleichen linearen Skala schätzt und für polytome Items Schwellenwerte für das Erreichen jeder Stufe schätzt (Boone et al., 2014; Döring, 2023c; Kelava & Moosbrugger, 2020). Ward et al. (2025) hingegen nutzen ausschließlich Items im Multiple Choice: Single Select Format. Dieses Vorgehen hat den Nachteil, dass die Anforderungen der Repräsentationskompetenzen so nicht vollständig abgedeckt werden. Ein Vorteil ist jedoch, dass die kognitiven Anforderungen zwischen den Items konstant gehalten werden und die Ergebnisse in Bezug auf die lower-level skills dadurch vergleichbarer sind. Ein weiterer Unterschied zeigt sich, wie bereits diskutiert, in Bezug auf die Inhalte und Repräsentationen. Während in dieser Studie und beispielsweise auch bei Taskin et al. (2015) und Sim und Daniel (2014) Inhalte aus verschiedenen Teilbereichen der Chemie sowie unterschiedliche Repräsentationen abgedeckt werden, beschränkt sich der ORCA von Ward et al. (2025) ausschließlich auf die Organische Chemie und auf ausgewählte, für die Organische Chemie typische Repräsentationen. Dadurch erreicht der ORCA eine hohe Vergleichbarkeit der Items innerhalb und zwischen den einzelnen Fähigkeiten der Repräsentationskompetenzen. Für diese Studie wird jedoch der Vorteil in der Verwendung einer hohen Bandbreite an Repräsentationen gesehen, da so allgemeingültigere Aussagen auf Basis der Daten getroffen werden können.

- (c) Die Abgrenzung vom Fachwissen stellt eine inhaltlich-methodische Schwäche aller bisher genannten Studien dar (Nitz, Ainsworth, et al., 2014; Scheid et al., 2018; Taskin et al., 2015; Ward et al., 2025). Die Messinstrumente grenzen die beiden Konstrukte nicht ab, wodurch eine Vielzahl der Items in den Instrumenten potenziell eher Fachwissen misst als Repräsentationskompetenzen. Dieser Aspekt wird im CRI:TIC durchweg berücksichtigt und durch eine klare Operationalisierung von Repräsentationskompetenzen und Fachwissen umgesetzt, was als Stärke des entwickelten Messinstruments hervorzuheben ist.

Zusammengefasst lässt sich zwar eine sehr hohe Überschneidung in den zum Vergleich herangezogenen Studien erkennen, genauso lassen sich aber auch Unterschiede mit Blick auf die Operationalisierung der Repräsentationskompetenzen, der Item-Konstruktion und den Inhalten erkennen. Dennoch kommen alle Studien zum gleichen Ergebnis: Die lower-level skills der Repräsentationskompetenzen lassen sich empirisch nicht trennen und stellen interkorrelierte Fähigkeiten dar.

Die Ergebnisse aus Forschungsanliegen II liefern in diesem Zusammenhang weitere Erkenntnisse, indem unterschiedliche räumliche Faktoren im Kontext des Bearbeitens repräsentationsbasierter Chemieaufgaben berücksichtigt werden. Global betrachtet lässt sich eine Korrelation zwischen allen acht bzw. sieben (Ausschluss eines Faktors aufgrund mangelnder Reliabilität und Validität) erhobenen räumlichen Faktoren und den Repräsentationskompetenzen erkennen. In der Pfadanalyse wird jedoch nur der Einfluss der räumlichen Faktoren *Speeded Rotation*, *Spatial Relation* und *Visualization* (zwei- und dreidimensionale mentale Rotations- und Manipulationsprozesse) signifikant, d. h. nur diese drei Faktoren leisten einen einzigartigen Beitrag zur Varianzaufklärung (Abbildung 22). Diese Befunde decken sich mit den Ergebnissen einer Vielzahl an Studien (z. B. Buckley et al., 2018; Harle & Towns, 2011; Stieff et al., 2012, 2016; Stull et al., 2012), wobei hier betont werden sollte, dass diese Studie als erste eine solche Bandbreite an räumlichen Fähigkeiten im Kontext des Arbeitens mit chemischen Repräsentationen erhebt und auswertet. Auf Basis der geschätzten Modellparameter ($n_p = 47$) und der zugrunde gelegten Stichprobengröße ($N = 494$) können die Ergebnisse als belastbar angesehen werden.

Ein differenzierter Blick auf die Fähigkeiten *Interpretation*, *Translation* und *Konstruktion* (Abbildung 23) verdeutlicht, dass sich die räumlichen Faktoren bei den drei Fähigkeiten leicht unterscheiden oder unterschiedlich stark ins Gewicht fallen (Abbildung 23). Bei der *Interpretation* und der *Konstruktion* zeigt sich die mentale dreidimensionale Rotations- und Manipulationsfähigkeit als wichtiger räumlicher Faktor, bei der *Translation* werden dagegen die zwei- und dreidimensionale mentale Rotationsfähigkeit signifikant, wobei die Regressionskoeffizienten hier deutlich höher sind, vor allem bei der dreidimensionalen mentalen Rotationsfähigkeit. Diese Ergebnisse müssen jedoch vorsichtiger interpretiert werden, da die Anzahl der geschätzten Parameter hier deutlich höher ist ($n_p = 70$).

Neben den empirischen Befunden, der *psychometrischen* Perspektive, dass die lower-level skills empirisch nicht trennbar sind, lässt sich aus der Operationalisierung der Fähigkeiten ableiten (Tabelle 3), dass diese unterschiedliche kognitive Anforderungen und Fähigkeiten erfordern. Daher steht der *psychometrischen* Perspektive eine *inhaltlich-fachdidaktische* Perspektive gegenüber, die eine getrennte Betrachtung der drei Fähigkeiten dennoch als sinnvoll ansieht, besonders mit Blick auf Lehr-Lern-Prozesse zur Förderung der Repräsentationskompetenzen. Diese Perspektive wird durch die Ergebnisse der Pfadanalysen aus Forschungsanliegen II unterstützt. Die im Rahmen von Forschungsanliegen III gewonnenen Daten untermauern diese Annahme weiter. Die identifizierten Hindernisse bei der Bearbeitung der repräsentationsbasierten Chemieaufgaben zeigen, dass insbesondere bei der *Translation* mentale Manipulations- und Rotationsprozesse von hoher Relevanz bei der Aufgabenlösung sind, was auch durch einschlägige Literatur untermauert wird (z. B. Buckley et al., 2018; Harle & Towns, 2011; Stieff et al., 2012, 2016; Stull et al., 2012). In der Pfadanalyse zeigt sich eine ähnliche Tendenz, da hier der Regressionskoeffizient für die dreidimensionale mentale Rotation bei der *Translation* am höchsten ist. Hingegen weist Forschungsanliegen III eine eher untergeordnete, wenn nicht vernachlässigbare Rolle der räumlichen Fähigkeiten bei der *Interpretation* von chemischen Repräsentationen aus, was auf den ersten Blick im Widerspruch zu den Ergebnissen der Pfadanalyse steht. Da die lower-level skills jedoch stark interkorreliert sind, ist ein Zusammenhang zwischen den Konstrukten nicht überraschend.

Forschungsanliegen III liefert wertvolle Einblicke in die relevanten kognitiven Prozesse und Fähigkeiten, die für die *Interpretation*, *Translation* und *Konstruktion* chemischer Repräsentationen notwendig sind. Diese unterstreichen zudem die *inhaltlich-fachdidaktische* Perspektive, da sich die kognitiven Prozesse und Fähigkeiten für die drei lower-level skills unterscheiden und dadurch voneinander abgrenzen lassen. Ein Aspekt, der jedoch alle drei lower-level skills miteinander verbindet, ist die zentrale Bedeutung des Fachwissens. Im Rahmen von Forschungsanliegen I kann gezeigt werden, dass die Repräsentationskompetenzen stark mit dem deklarativen Fachwissen korrelieren. Dennoch kann hier Evidenz geliefert werden, dass die Repräsentationskompetenzen empirisch eindeutig vom deklarativen Fachwissen getrennt werden können, was auch im zweiten Forschungsanliegen repliziert sowie in anderen Studien gezeigt werden kann (z. B. Edelsbrunner et al., 2023; Nitz, Ainsworth,

et al., 2014). Diese Befunde sind besonders mit Blick auf das Repräsentationsdilemma von hoher Relevanz (Rau, 2017a, 2017b, 2017c, 2018; Rau et al., 2021), da dadurch die Argumente zur stärkeren Berücksichtigung und Vermittlung der Repräsentationskompetenzen im Lehr-Lern-Geschehen untermauert werden. Darüber hinaus zeigt die erste Pfadanalyse in Forschungsanliegen II, dass der standardisierte Pfadkoeffizient des deklarativen Fachwissens von allen signifikanten Einflüssen am höchsten ausgeprägt ist (Abbildung 22). In dem Pfadmodell mit der getrennten Betrachtung der lower-level skills zeigt sich dieses Bild ebenfalls für die *Interpretation* und *Konstruktion* – die standardisierten Pfadkoeffizienten sind für das deklarative Fachwissen deutlich höher als für die räumlichen Faktoren. Bei der *Translation* ist dies jedoch nicht der Fall. Hier ist der Pfadkoeffizient für die dreidimensionale mentale Rotationsfähigkeit (*Spatial Relation*) am höchsten, auch wenn die Differenz zum deklarativen Fachwissen nur gering ist (Abbildung 23). Im dritten Forschungsanliegen zeichnet sich ein vergleichbares Bild ab. Fehlendes, mangelhaftes oder falsches Fachwissen stellt das am stärksten vertretene Hindernis für alle drei lower-level skills dar. Dabei wird das deklarative Fachwissen abgedeckt, das auch im Rahmen der ersten zwei Forschungsanliegen eine Rolle spielt. Ferner zeigt sich aber auch, dass das prozedurale Fachwissen, besonders bei komplexeren und kognitiv anspruchsvolleren Aufgaben von sehr hoher Relevanz ist. Eine Integration des deklarativen und prozeduralen Wissens ist dabei ebenfalls maßgeblich für die Lösung verantwortlich.

Ein weiterer wichtiger Aspekt ist die Übertragbarkeit der Hindernisse beim Bearbeiten repräsentationsbasierter Aufgaben in der Organischen Chemie (z. B. Graulich, 2015; Stieff, 2007; Stieff et al., 2018; Stull et al., 2016; Ward et al., 2022, 2025) auf den Fachbereich der Komplexchemie. Die Befunde deuten darauf hin, dass räumliche Fähigkeiten in verschiedenen Bereichen der Chemie von zentraler Bedeutung sind, ebenso wie fachliches Wissen, das jedoch kontextspezifisch ist. Hindernisse im Umgang mit Repräsentationen lassen sich somit in unterschiedlichen Fachgebieten der Chemie replizieren, was erneut die Relevanz und Dringlichkeit betont, die damit verbundenen Kompetenzen zum erfolgreichen Umgang mit Repräsentationen in Lehr-Lern-Prozessen neben dem Fachwissen gezielt zu adressieren und zu fördern.

8.2 Implikationen für Praxis und Forschung

Die Ergebnisse dieses Projekts ermöglichen zum einen die Ableitung wichtiger Implikationen für Lehr-Lern-Prozesse, zum anderen eröffnen sie Perspektiven für weiterführende Forschungsfragen, die sich aus den gewonnenen Erkenntnissen ableiten lassen.

Implikationen für Lehr-Lern-Prozesse - Repräsentationskompetenzen spielen im Lehr-Lern-Prozess sowie in Lehr- und Sachbüchern eine dem Fachwissen untergeordnete Rolle und werden nicht ausreichend gefördert (Aydin et al., 2014; Gurung et al., 2022; Nyachwaya & Gillaspie, 2016; Popova & Jones, 2021; Rau et al., 2021; Sim & Daniel, 2014; Wu & Rau, 2018). Mit Blick auf das Repräsentationsdilemma (z. B. Rau, 2018) lässt sich jedoch ableiten, dass beide eine zentrale Rolle im Lehr-Lern-Prozess einnehmen müssen. Diese Studie untermauert diese Aussage, indem Evidenz geliefert wird, dass Repräsentationskompetenzen und deklaratives Fachwissen zwar korrelieren, aber empirisch trennbare Konstrukte sind. Repräsentationskompetenzen müssen daher stärker in den Lehr-Lern-Prozess einbezogen werden, um die Lernenden zu befähigen, einzelne und multiple Repräsentationen reflektiert zu nutzen, über chemische Phänomene zu kommunizieren und die zugrundeliegenden Prozesse zu visualisieren. Erst dadurch können Lernende Repräsentationen zur Entwicklung fachlicher Konzepte effektiv nutzen, wie es auch andere Autor:innen hervorheben (z. B. Kozma & Russell, 2005; Rau, 2018; Sim & Daniel, 2014; Ward et al., 2022). Zwar werden im Rahmen dieser Studie Studierende als Testpersonen herangezogen, dies lässt sich jedoch auch auf Kinder und Jugendliche in der Schule beziehen, da sie dort bereits mit den gleichen Repräsentationen konfrontiert werden und daher mit deren Einführung die Repräsentationskompetenzen entwickeln müssen. Dennoch spielt das Fachwissen eine zentrale Rolle. Besonders die Integration von deklarativem und prozeduralem Fachwissen muss gefördert werden. Hier liegen Hinweise vor, dass der Fokus bei den Studierenden auf deklarativem Wissen liegt und sie dieses nicht adäquat in Problemsituationen anwenden können, was auch in anderen Studien belegt wird und daher gefördert werden muss (z. B. Friege & Lind, 2006; Solaz-Portolés & López, 2008). Diese Fördermaßnahmen sollten auch Strategien umfassen, wie Lernende implizit dargestellte Informationen aus Repräsentationen herausarbeiten können, sodass sie sich nicht nur auf Oberflächenmerkmale und explizit dargestellte Repräsentationsmerkmale beziehen (z. B. Ainsworth, 2006; Graulich, 2015).

In Bezug auf die Struktur und den Zusammenhang der lower-level skills der Repräsentationskompetenzen lässt sich hier zwar aus der *psychometrischen* Perspektive festhalten, dass diese stark miteinander korreliert sind, aus der *inhaltlich-fachdidaktischen* Perspektive ist eine getrennte Thematisierung und Betrachtung der Fähigkeiten im Lehr-Lern-Prozess dennoch sinnvoll. Die vorliegenden Ergebnisse zeigen, dass für die lower-level skills unterschiedliche kognitive Prozesse und Fähigkeiten benötigt werden, was die getrennte Thematisierung weiter unterstützt. Diese Ergebnisse zeigen beispielsweise Ward et al. (2022, 2025) und sprechen ähnliche Empfehlungen aus. In diesem Zusammenhang sollte auch die Bedeutsamkeit des *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC) noch einmal betont werden, das sich als geeignetes Instrument für die Erfassung der lower-level skills der Repräsentationskompetenzen erwiesen hat und besonders durch eine Kontrolle und Berücksichtigung des Fachwissens von vergleichbaren Messinstrumenten abhebt. Dadurch wird eine valide Messung der Repräsentationskompetenzen ermöglicht.

Räumliche Fähigkeiten, konkreter zwei- und dreidimensionale mentale Rotations- und Manipulationsfähigkeiten, spielen besonders bei der *Translation* und *Konstruktion* chemischer Repräsentationen eine Schlüsselrolle, was hier in den Forschungsanliegen II und III gezeigt werden konnte. Barke und Engida (2001) empfehlen beispielsweise, so früh wie möglich räumlich-visuelle Repräsentationen in den Unterricht zu integrieren, sodass räumliche Fähigkeiten optimal entwickelt werden können. Darüber hinaus sollten haptische (z. B. Molekülbaukästen) oder virtuelle (z. B. Augmented Reality) Modelle bei der Bearbeitung räumlicher Aufgaben angeboten und genutzt werden, da diese die kognitiven Prozesse bei der Problemlösung unterstützen (z. B. Bogomolova et al., 2020; Habig, 2020; Kenneally & Bentley, 2024; Stieff et al., 2016; Stull et al., 2012).

Auch die Vermittlung spezifischer Strategien zur Bearbeitung und Handhabung räumlicher Aufgaben kann in diesem Zusammenhang eine effektive Möglichkeit darstellen (Maresch, 2014; Wang & Carr, 2014). Zahlreiche Studien konnten darüber hinaus belegen, dass Trainings die räumlichen Fähigkeiten von Lernenden fördern, was daher ein vielversprechender Ansatz ist, um Lernende bei der Bearbeitung repräsentationsbasierter Chemieaufgaben zu unterstützen und ihre Leistung zu verbessern (Carlisle et al., 2015; Hawes et

al., 2022; Merchant et al., 2013; Sorby, 2009; Stieff & Uttal, 2015; Uttal et al., 2013; Yang et al., 2020).

Implikationen und Ausblick auf weitere Forschung – Im Fokus dieser Arbeit stehen die drei lower-level skills *Interpretation*, *Translation* und *Konstruktion*. Zwar können dieses Projekt und vergleichbare Studien Ergebnisse zum Zusammenhang der Fähigkeit liefern und dass diese ein eher eindimensionales Konstrukt und stark interkorrelierte Fähigkeiten darstellen, gleichzeitig gibt es jedoch auch Hinweise, dass diese in einer eher hierarchischen Beziehung stehen und beispielsweise Repräsentationen zunächst interpretiert, ehe sie übersetzt oder neue Repräsentationen konstruiert werden müssen (z. B. Ward et al., 2025). Das methodische Vorgehen in dieser Studie kann dazu keinen Beitrag liefern und muss daher in weiteren Untersuchungen mit adäquaten Methoden adressiert werden. Denkbar wäre eine Interventionsstudie im Prä-Post-Design mit Messung der Fähigkeiten mithilfe des CRI:TIC und Interventionen zur Förderung der einzelnen lower-level skills. Dadurch könnte u. a. untersucht werden, inwieweit eine Förderung der *Interpretation* die Fähigkeiten *Translation* und *Konstruktion* beeinflusst.

Die higher-level skills *Vergleich und Kritik*, *Epistemologie* und *Argumentation* werden im Rahmen dieser Studie nicht untersucht, sind jedoch notwendig, um erfolgreich mit Repräsentationen arbeiten zu können, indem Metawissen über die Repräsentationen genutzt wird, beispielsweise, um begründet spezifische Repräsentationen in Problemsituationen zur Problemlösung heranzuziehen. Studien belegen, dass diese Fähigkeiten im Lehr-Lern-Prozess kaum Beachtung finden und wenn eher die lower-level skills gefördert werden, die gleichzeitig die Grundlage für die Entwicklung der higher-level skills darstellen (z. B. Aydin et al., 2014; Gurung et al., 2022; Nyachwaya & Gillaspie, 2016; Popova & Jones, 2021). Auch in der bisherigen Forschung finden die higher-level skills nicht viel Beachtung. Daher sollte sich zukünftige Forschung vermehrt auf die Erfassung dieser fokussieren und auch die Struktur und den Zusammenhang zwischen diesen untersuchen. Dafür fehlt es aktuell jedoch an geeigneten Messinstrumenten. Genauso muss die Wechselbeziehung zwischen den lower- und higher-level skills weiter untersucht werden.

Die Ergebnisse zur Relevanz der zwei- und dreidimensionalen mentalen Rotations- und Manipulationsfähigkeit im Kontext des Arbeitens, genauer des Übersetzens und Konstruierens chemischer Repräsentationen geben hier

neue Anknüpfungspunkte für das Trainieren räumlicher Fähigkeiten. Auf dieser Basis können gezielt Trainings für die drei Faktoren *Speeded Rotation*, *Spatial Relation* und *Visualization* entwickelt werden. Hierfür bietet sich ein außerfachliches Training an, um die Förderung fachlicher Inhalte auszuschließen, sodass der Einfluss auf die Förderung der räumlichen Faktoren zurückgeführt werden kann. Der Einfluss des Trainings auf die *Translation* und *Konstruktion* muss anschließend evaluiert und kritisch geprüft werden.

9

Verzeichnisse

9.1 Literaturverzeichnis

- Abdi, H. (2010). Holm's Sequential Bonferroni Procedure. In N. Salkind (Hrsg.), *Encyclopedia of Research Design*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288.n178>
- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3), 183–198. <https://doi.org/10.1016/j.learninstruc.2006.03.001>
- Ainsworth, S. (2008). The Educational Value of Multiple-representations when Learning Complex Scientific Concepts. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Hrsg.), *Visualization: Theory and Practice in Science Education* (S. 191–208). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5267-5_9
- Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction*, 14(3), 241–255. <https://doi.org/10.1016/j.learninstruc.2004.06.002>
- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18(3), 91–93. <https://doi.org/10.1016/j.tjem.2018.08.001>
- Anderson, L. W., & Krathwohl, D. R. (Hrsg.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed). Longman.
- Averbeck, D. (2021). *Zum Studienerfolg in der Studieneingangsphase des Chemiestudiums: Der Einfluss kognitiver und affektiv-motivationaler Variablen*. Logos Verlag.
- Aydin, S., Sinha, S., Izci, K., & Volkmann, M. (2014). Turkish, Indian, and American Chemistry Textbooks Use of Inscriptions to Represent 'Type-sof Chemical Reactions'. *EURASIA Journal of Mathematics, Science and Technology Education*, 10(5). <https://doi.org/10.12973/eurasia.2014.1060a>

- Baddeley, A. D. (2001). *Human memory: Theory and practice* (Rev. ed). Psychology Press.
- Barke, H.-D., & Engida, T. (2001). Structural chemistry and spatial ability in different cultures. *Chem. Educ. Res. Pract.*, 2(3), 227–239. <https://doi.org/10.1039/B1RP90025K>
- Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, 2, 8–14. <https://doi.org/10.1016/j.npls.2016.01.001>
- Bent, H. A. (1984). Uses (and abuses) of models in teaching chemistry. *Journal of Chemical Education*, 61(9), 774. <https://doi.org/10.1021/ed061p774>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, 101, 150–167. <https://doi.org/10.1016/j.compedu.2016.06.005>
- Bilandzic, H. (2017). Lautes Denken. In L. Mikos & C. Wegener (Hrsg.), *Qualitative Medienforschung: Ein Handbuch* (2., völlig überarbeitete und erweiterte Auflage, S. 406–413). UVK Verlagsgesellschaft mbH. <https://doi.org/10.36198/9783838586472>
- Binnewies, M., Finze, M., Jäckel, M., Schmidt, P., Willner, H., & Rayner-Canham, G. (2016). *Allgemeine und Anorganische Chemie*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-45067-3>
- Bodner, G. M., & Guay, R. B. (1997). The Purdue Visualization of Rotations Test. *The Chemical Educator*, 2(4), 1–17. <https://doi.org/10.1007/s00897970138a>
- Bodner, G. M., & McMillen, T. L. B. (1986). Cognitive restructuring as an early stage in problem solving. *Journal of Research in Science Teaching*, 23(8), 727–737. <https://doi.org/10.1002/tea.3660230807>
- Bogomolova, K., Van Der Ham, I. J. M., Dankbaar, M. E. W., Van Den Broek, W. W., Hovius, S. E. R., Van Der Hage, J. A., & Hierck, B. P. (2020). The Effect of Stereoscopic Augmented Reality Visualization on Learning Anatomy and the Modifying Effect of Visual-Spatial Abilities: A Double-Center Randomized Controlled Trial. *Anatomical Sciences Education*, 13(5), 558–567. <https://doi.org/10.1002/ase.1941>

- Boluarte-Carbajal, A., Salazar-Conde, M., Alata Vasquez, S., & Zegarra-López, A. (2023). Psychometric review of the perceived stress scale under CFA and Rasch models in Lima, Peru. *Frontiers in Psychology*, 14, 1160466. <https://doi.org/10.3389/fpsyg.2023.1160466>
- Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch Analyses in the Human Sciences*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-43420-5>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Netherlands. <https://doi.org/10.1007/978-94-007-6857-4>
- Brandt, H., & Moosbrugger, H. (2020). Planungsaspekte und Konstruktionsphasen von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 39–66). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4_3
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educational and Psychological Measurement*, 41(3), 687–699. <https://doi.org/10.1177/001316448104100307>
- Brewer, M. J., Butler, A., & Cooksley, S. L. (2016). The relative performance of AIC, AIC_c and BIC in the presence of unobserved heterogeneity. *Methods in Ecology and Evolution*, 7(6), 679–692. <https://doi.org/10.1111/2041-210X.12541>
- Brody, N. (2000). History of Theories and Measurements of intelligence. In R. J. Sternberg (Hrsg.), *Handbook of Intelligence* (1. Aufl., S. 16–33). Cambridge University Press. <https://doi.org/10.1017/CBO9780511807947.003>
- Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2), 230–258. <https://doi.org/10.1177/0049124192021002005>
- Bruner, J. S. (1966). *Toward a theory of instruction*. The Belknap press of Harvard university press.
- Bucat, B., & Mocerino, M. (2009). Learning at the Sub-micro Level: Structural Representations. In J. K. Gilbert & D. Treagust (Hrsg.), *Multiple Representations in Chemical Education* (Bd. 4, S. 11–29). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8872-8_2

- Buckley, J., Seery, N., & Canty, D. (2018). A Heuristic Framework of Spatial Ability: A Review and Synthesis of Spatial Factor Literature to Support its Translation into STEM Education. *Educational Psychology Review*, 30(3), 947–972. <https://doi.org/10.1007/s10648-018-9432-z>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Cardellini, L. (2012). Chemistry: Why the Subject is Difficult? *Educación Química*, 23, 305–310. [https://doi.org/10.1016/S0187-893X\(17\)30158-1](https://doi.org/10.1016/S0187-893X(17)30158-1)
- Carlisle, D., Tyson, J., & Nieswandt, M. (2015). Fostering spatial skill acquisition by general chemistry students. *Chemistry Education Research and Practice*, 16(3), 478–517. <https://doi.org/10.1039/C4RP00228H>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Carter, C. S., Larussa, M. A., & Bodner, G. M. (1987). A study of two measures of spatial ability as predictors of success in different levels of general chemistry. *Journal of Research in Science Teaching*, 24(7), 645–657. <https://doi.org/10.1002/tea.3660240705>
- Castro-Alonso, J. C., & Sweller, J. (2021). The Modality Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (3. Aufl., S. 261–267). Cambridge University Press. <https://doi.org/10.1017/9781108894333.026>
- Castro-Alonso, J. C., Wong, M., Adesope, O. O., Ayres, P., & Paas, F. (2019). Gender Imbalance in Instructional Dynamic Versus Static Visualizations: A Meta-analysis. *Educational Psychology Review*, 31(2), 361–387. <https://doi.org/10.1007/s10648-019-09469-1>
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40(3), 153–193. <https://doi.org/10.1037/h0059973>
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1–22. <https://doi.org/10.1037/h0046743>
- Cheng, P. C.-H. (1999). Unlocking conceptual learning in mathematics and science with effective representational systems. *Computers & Education*, 33(2–3), 109–130. [https://doi.org/10.1016/S0360-1315\(99\)00028-7](https://doi.org/10.1016/S0360-1315(99)00028-7)

- Chiu, M.-H., & Wu, H.-K. (2009). The Roles of Multimedia in the Teaching and Learning of the Triplet Relationship in Chemistry. In J. K. Gilbert & D. Treagust (Hrsg.), *Multiple Representations in Chemical Education* (Bd. 4, S. 251–283). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8872-8_12
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q_3 : Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement*, 41(3), 178–194. <https://doi.org/10.1177/0146621616677520>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.
- Contreras, M. J., Colom, R., Hernández, J. M., & Santacreu, J. (2003). Is Static Spatial Performance Distinguishable From Dynamic Spatial Performance? A Latent-Variable Analysis. *The Journal of General Psychology*, 130(3), 277–288. <https://doi.org/10.1080/00221300309601159>
- Cook, M., Wiebe, E. N., & Carter, G. (2008). The influence of prior knowledge on viewing and interpreting graphics with macroscopic and molecular representations. *Science Education*, 92(5), 848–867. <https://doi.org/10.1002/sce.20262>
- Dai, R., Laureanti, J. A., Kopelevich, M., & Diaconescu, P. L. (2020). Developing a Virtual Reality Approach toward a Better Understanding of Coordination Chemistry and Molecular Orbitals. *Journal of Chemical Education*, 97(10), 3647–3651. <https://doi.org/10.1021/acs.jchemed.0c00469>
- Davidson, J. E., & Downing, C. L. (2000). Contemporary Models of Intelligence. In R. J. Sternberg (Hrsg.), *Handbook of Intelligence* (1. Aufl., S. 34–50). Cambridge University Press. <https://doi.org/10.1017/CBO9780511807947.004>
- Davidson, J. E., & Kemp, I. A. (2011). Contemporary Models of Intelligence. In R. J. Sternberg & S. B. Kaufman (Hrsg.), *The Cambridge Handbook of Intelligence* (1. Aufl., S. 58–82). Cambridge University Press. <https://doi.org/10.1017/CBO9780511977244.005>
- Demars, C. E. (2013). A comparison of confirmatory factor analysis and multidimensional Rasch models to investigate the dimensionality of test-taking motivation. *Journal of Applied Measurement*, 14(2), 179–196.

- Dickmann, T. (with Universität Duisburg-Essen). (2019). *Visuelles Modellverständnis und Studienerfolg in der Chemie: Zwei Seiten einer Medaille*. Logos Verlag.
- D’Oliveira, T. C. (2004). Dynamic Spatial Ability: An Exploratory Analysis and a Confirmatory Study. *The International Journal of Aviation Psychology*, 14(1), 19–38. https://doi.org/10.1207/s15327108ijap1401_2
- Döring, N. (2023a). Bestimmung von Teststärke, Effektgröße und optimalem Stichprobenumfang. In N. Döring, *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 789–846). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64762-2_14
- Döring, N. (2023b). Datenanalyse. In N. Döring, *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 587–766). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64762-2_12
- Döring, N. (2023c). Datenerhebung. In N. Döring, *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 321–570). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64762-2_10
- Döring, N. (2023d). Operationalisierung. In N. Döring, *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 223–291). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64762-2_8
- Döring, N. (2023e). Qualitätskriterien in der empirischen Sozialforschung. In N. Döring, *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 79–118). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64762-2_3
- Döring, N. (2023f). Untersuchungsdesign. In N. Döring, *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 183–221). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64762-2_7
- Dörner, D. (1987). *Problemlösen als Informationsverarbeitung* (3. Aufl.). Kohlhammer.
- Dreher, A., & Kuntze, S. (2015). Teachers Facing the Dilemma of Multiple Representations Being Aid and Obstacle for Learning: Evaluations of Tasks and Theme-Specific Noticing. *Journal Für Mathematik-Didaktik*, 36(1), 23–44. <https://doi.org/10.1007/s13138-014-0068-3>

- Dresing, T., & Pehl, T. (2018). *Praxisbuch Interview, Transkription & Analyse: Anleitungen und Regelsysteme für qualitativ Forschende* (8. Auflage). Eigenverlag.
- Edelsbrunner, P. A., Malone, S., Hofer, S. I., Küchemann, S., Kuhn, J., Schmid, R., Altmeyer, K., Brünken, R., & Lichtenberger, A. (2023). The relation of representational competence and conceptual knowledge in female and male undergraduates. *International Journal of STEM Education*, 10(1), 44. <https://doi.org/10.1186/s40594-023-00435-6>
- Eilam, B., & Gilbert, J. K. (2014). The Significance of Visual Representations in the Teaching of Science. In B. Eilam & J. K. Gilbert (Hrsg.), *Science Teachers' Use of Visual Representations* (Bd. 8, S. 3–28). Springer International Publishing. https://doi.org/10.1007/978-3-319-06526-7_1
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976a). *Kit of Factor-Referenced Cognitive Tests*. Educational Testing Service.
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976b). *Manual for Kit of Factor-Referenced Cognitive Tests*. Educational Testing Service.
- Elschenbroich, C. (2008). *Organometallchemie* (6., überarbeitete Auflage). Vieweg+Teubner. <https://doi.org/10.1007/978-3-8351-9223-2>
- Estes, Z., & Felker, S. (2012). Confidence Mediates the Sex Difference in Mental Rotation Performance. *Archives of Sexual Behavior*, 41(3), 557–570. <https://doi.org/10.1007/s10508-011-9875-5>
- Eticha, A. T., Amde, M., Bogale, Y., & Bekana, D. (2023). The effect of context-based close packing supported with the 3D-virtual model of crystals structure on students' achievement and attitude. *Chemistry Teacher International*, 5(3), 311–323. <https://doi.org/10.1515/cti-2022-0047>
- Feuerstahler, L., & Wilson, M. (2019). Scale Alignment in Between-Item Multidimensional Rasch Models. *Journal of Educational Measurement*, 56(2), 280–301. <https://doi.org/10.1111/jedm.12209>
- Fidalgo, A. M., & Madeira, J. M. (2008). Generalized Mantel-Haenszel Methods for Differential Item Functioning Detection. *Educational and Psychological Measurement*, 68(6), 940–958. <https://doi.org/10.1177/0013164408315265>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock „n“ roll* (4th edition). Sage.
- Finze, M., & Radius, U. (2014a). Koordinationschemie: Struktur und Bindung. In J. Huheey, E. Keiter, & R. Keiter, *Anorganische Chemie* (S. 439–542). DE GRUYTER. <https://doi.org/10.1515/9783110307955.439>

- Finze, M., & Radius, U. (2014b). Strukturen von Koordinationsverbindungen. In J. Huheey, E. Keiter, & R. Keiter, *Anorganische Chemie* (S. 625–684). DE GRUYTER. <https://doi.org/10.1515/9783110307955.625>
- Fiorella, L., & Mayer, R. E. (2021). Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In R. E. Mayer & L. Fiorella (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (3. Aufl., S. 185–198). Cambridge University Press. <https://doi.org/10.1017/9781108894333.019>
- Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Hrsg.), *Encyclopedia of Special Education* (1. Aufl.). Wiley. <https://doi.org/10.1002/9781118660584.es0431>
- Flick, U. (2022). Gütekriterien qualitativer Sozialforschung. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 533–547). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-37985-8_34
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, 137(2), 316–344. <https://doi.org/10.1037/a0021663>
- French, B. F., Finch, W. H., & Immekus, J. C. (2019). Multilevel Generalized Mantel-Haenszel for Differential Item Functioning Detection. *Frontiers in Education*, 4, 47. <https://doi.org/10.3389/feduc.2019.00047>
- Friege, G., & Lind, G. (2006). Types and Qualities of Knowledge and their Relations to Problem Solving in Physics. *International Journal of Science and Mathematics Education*, 4(3), 437–465. <https://doi.org/10.1007/s10763-005-9013-8>
- Fulmer, G. W., Hwang, J., Ding, C., Hand, B., Suh, J. K., & Hansen, W. (2021). Development of a questionnaire on teachers' knowledge of language as an epistemic tool. *Journal of Research in Science Teaching*, 58(4), 459–490. <https://doi.org/10.1002/tea.21666>
- Gäde, J. C., & Schermelleh-Engel, K. (2023). Strukturgleichungsmodelle. In N. Döring, *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 923–950). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64762-2_17

- Gäde, J. C., Schermelleh-Engel, K., & Brandt, H. (2020). Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 615–659). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4_24
- Gade, L. H. (1998). *Koordinationschemie* (1. Aufl.). Wiley. <https://doi.org/10.1002/9783527663927>
- Galton, F. (1879). Generic images. *Nineteenth Century*, 6(1), 157–169.
- Galton, F. (1880). Mental imagery. *Fortnightly Review*, 28(2), 312–324.
- Gebhard, U., Höttecke, D., & Rehm, M. (2017). *Pädagogik der Naturwissenschaften*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-19546-9>
- Gilbert, J. K. (2005). Visualization: A Metacognitive Skill in Science and Science Education. In J. K. Gilbert (Hrsg.), *Visualization in Science Education* (S. 9–27). Springer Netherlands. https://doi.org/10.1007/1-4020-3613-2_2
- Gilbert, J. K. (2008). Visualization: An Emergent Field of Practice and Enquiry in Science Education. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Hrsg.), *Visualization: Theory and Practice in Science Education* (S. 3–24). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5267-5_1
- Gilbert, J. K., & Treagust, D. F. (2009). Introduction: Macro, Submicro and Symbolic Representations and the Relationship Between Them: Key Models in Chemical Education. In J. K. Gilbert & D. Treagust (Hrsg.), *Multiple Representations in Chemical Education* (Bd. 4, S. 1–8). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8872-8_1
- Göhner, M., & Krell, M. (2020). Qualitative Inhaltsanalyse in naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von Gütekriterien: Ein Review. *Zeitschrift für Didaktik der Naturwissenschaften*, 26(1), 207–225. <https://doi.org/10.1007/s40573-020-00111-0>
- Göhner, M., & Krell, M. (2021). Was ist schwierig am Modellieren? Identifikation und Beschreibung von Hindernissen in Modellierungsprozessen von Lehramtsstudierenden naturwissenschaftlicher Fächer. *Zeitschrift für Didaktik der Naturwissenschaften*, 27(1), 155–180. <https://doi.org/10.1007/s40573-021-00131-4>
- Göksun, T., Goldin-Meadow, S., Newcombe, N., & Shipley, T. (2013). Individual differences in mental rotation: What does gesture tell us? *Cognitive Processing*, 14(2), 153–162. <https://doi.org/10.1007/s10339-013-0549-1>

- Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*, 15(6), 876–884. <https://doi.org/10.1111/j.1467-7687.2012.01185.x>
- Graulich, N. (2015). Intuitive Judgments Govern Students' Answering Patterns in Multiple-Choice Exercises in Organic Chemistry. *Journal of Chemical Education*, 92(2), 205–211. <https://doi.org/10.1021/ed500641n>
- Guay, R. (1977). *Purdue Spatial Visualization Tests*. Purdue Research Foundation.
- Gurung, E., Jacob, R., Bunch, Z., Thompson, B., & Popova, M. (2022). Evaluating the Effectiveness of Organic Chemistry Textbooks for Promoting Representational Competence. *Journal of Chemical Education*, 99(5), 2044–2054. <https://doi.org/10.1021/acs.jchemed.1c01054>
- Habig, S. (2020). Who can benefit from augmented reality in chemistry? Sex differences in solving stereochemistry problems using augmented reality. *British Journal of Educational Technology*, 51(3), 629–644. <https://doi.org/10.1111/bjet.12891>
- Hamel, J.-F., Sébille, V., Challet-Bouju, G., & Hardouin, J.-B. (2016). Partial Credit Model: Estimations and Tests of Fit with Pmodel. *The Stata Journal: Promoting Communications on Statistics and Stata*, 16(2), 464–481. <https://doi.org/10.1177/1536867X1601600212>
- Harle, M., & Towns, M. (2011). A Review of Spatial Ability Literature, Its Connection to Chemistry, and Implications for Instruction. *Journal of Chemical Education*, 88(3), 351–360. <https://doi.org/10.1021/ed900003n>
- Harris, J., Hirsh-Pasek, K., & Newcombe, N. S. (2013). Understanding spatial transformations: Similarities and differences between mental rotation and mental folding. *Cognitive Processing*, 14(2), 105–115. <https://doi.org/10.1007/s10339-013-0544-6>
- Hawes, Z. C. K., Gilligan-Lee, K. A., & Mix, K. S. (2022). Effects of spatial training on mathematics performance: A meta-analysis. *Developmental Psychology*, 58(1), 112–137. <https://doi.org/10.1037/dev0001281>
- Hegarty, M. (2018). Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin & Review*, 25(3), 1212–1219. <https://doi.org/10.3758/s13423-017-1347-z>

- Hodgkiss, A., Gilligan, K. A., Tolmie, A. K., Thomas, M. S. C., & Farran, E. K. (2018). Spatial cognition and science achievement: The contribution of intrinsic and extrinsic spatial skills from 7 to 11 years. *British Journal of Educational Psychology*, 88(4), 675–697. <https://doi.org/10.1111/bjep.12211>
- Höfler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17(6), 722–738. <https://doi.org/10.1016/j.learninstruc.2007.09.013>
- Holleman, A. F., Wiberg, E., & Wiberg, N. (2016a). *Anorganische Chemie: Grundlagen und Hauptgruppenelemente*. De Gruyter. <https://doi.org/10.1515/9783110495850>
- Holleman, A. F., Wiberg, N., & Wiberg, E. (2016b). *Anorganische Chemie: Nebengruppenelemente, Lanthanoide, Actinoide, Transactinoide*. De Gruyter. <https://doi.org/10.1515/9783110495904>
- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities* [Unveröffentlichte Dissertation]. University of Illinois.
- Horn, J. L., & Blankson, A. N. (2012). Foundations for Better Understanding of Cognitive Abilities. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed, S. 73–98). Guilford Press.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57(5), 253–270. <https://doi.org/10.1037/h0023816>
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15(3), 495–514. <https://doi.org/10.3758/PBR.15.3.495>
- Hotchkiss, J. T., Cook-Cottone, C. P., Wong, M. Y. C., Guyker, W., & Garcia, A. C. M. (2023). Intercultural Validation of the Mindful Self-Care Scale— Rasch and Factor Analysis of 16 Studies Representing Five Continents. *Mindfulness*, 14(8), 2055–2072. <https://doi.org/10.1007/s12671-023-02196-3>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

- Huang, S., & Ishii, H. (2025). Enhancing Precision in Predicting Magnitude of Differential Item Functioning: An M-DIF Pretrained Model Approach. *Educational and Psychological Measurement*, 85(2), 384–400. <https://doi.org/10.1177/00131644241279882>
- Janiak, C. (2018). Komplex-/Koordinationschemie. In H.-J. Meyer (Hrsg.), *Moderne Anorganische Chemie* (S. 409–624). De Gruyter. <https://doi.org/10.1515/9783110441635-003>
- Jansen, M., Schneider, R., Schipolowski, S., & Henschel, S. (2019). Motivationale Schülermerkmale im Fach Mathematik und in den naturwissenschaftlichen Fächern. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich, & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 337–354). Waxmann.
- Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, 7(2), 75–83. <https://doi.org/10.1111/j.1365-2729.1991.tb00230.x>
- Johnstone, A. H. (2000). Teaching of Chemistry—Logical or Psychological? *Chem. Educ. Res. Pract.*, 1(1), 9–15. <https://doi.org/10.1039/A9RP90001B>
- Justi, R., Gilbert, J. K., & Ferreira, P. F. M. (2009). The Application of a ‘Model of Modelling’ to Illustrate the Importance of Metavisualisation in Respect of the Three Types of Representation. In J. K. Gilbert & D. Treagust (Hrsg.), *Multiple Representations in Chemical Education* (Bd. 4, S. 285–307). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8872-8_13
- Kechel, J.-H. (2016). *Schülerschwierigkeiten beim eigenständigen Experimentieren: Eine qualitative Studie am Beispiel einer Experimentieraufgabe zum Hooke’schen Gesetz*. Logos Verlag.
- Kelava, A., & Moosbrugger, H. (2020). Einführung in die Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 369–409). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4_16
- Kenneally, C. D., & Bentley, B. (2024). A Cognitive Load Approach to Molecular Geometries: Augmented Reality Technology and Visuospatial Abilities in Chemistry. *Education Sciences*, 14(9), 1036. <https://doi.org/10.3390/educsci14091036>

- Kenny, D. A. (2024). *Measuring Model Fit*. Measuring Model Fit.
<https://davidakenny.net/cm/fit.htm>
- Kircher, E. (2015a). Modellbegriff und Modellbildung in der Physikdidaktik. In E. Kircher, R. Girwidz, & P. Häußler (Hrsg.), *Physikdidaktik* (S. 783–807). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41745-0_27
- Kircher, E. (2015b). Ziele und Kompetenzen im Physikunterricht. In E. Kircher, R. Girwidz, & P. Häußler (Hrsg.), *Physikdidaktik* (S. 75–105). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41745-0_3
- Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R., J. (2018). From Cognitive Load Theory to Collaborative Cognitive Load Theory. *International Journal of Computer-Supported Collaborative Learning*, 13(2), 213–233. <https://doi.org/10.1007/s11412-018-9277-y>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (T. D. Little, Hrsg.; Fourth edition). The Guilford Press.
- Knoblauch, H., & Vollmer, T. (2022). Ethnographie. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 659–676). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-37985-8_41
- Kohl, P. B., & Finkelstein, N. (2017). Understanding and Promoting Effective Use of Representations in Physics Learning. In D. F. Treagust, R. Duit, & H. E. Fischer (Hrsg.), *Multiple Representations in Physics Education* (Bd. 10, S. 231–254). Springer International Publishing.
https://doi.org/10.1007/978-3-319-58914-5_11
- Konrad, K. (2010). Lautes Denken. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 476–490). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92052-8_34
- Kozma, R., & 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%253C949::AID-TEA7%253E3.0.CO;2-U](https://doi.org/10.1002/(SICI)1098-2736(199711)34:9%253C949::AID-TEA7%253E3.0.CO;2-U)
- Kozma, R., & Russell, J. (2005). Students Becoming Chemists: Developing Representationl Competence. In J. K. Gilbert (Hrsg.), *Visualization in Science Education* (S. 121–145). Springer Netherlands.
https://doi.org/10.1007/1-4020-3613-2_8

- Krebs, D., & Menold, N. (2022). Gütekriterien quantitativer Sozialforschung. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 549–565). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-37985-8_35
- Krey, O., & Schwanewedel, J. (2018). Lernen mit externen Repräsentationen. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 159–175). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56320-5_10
- Krüger, D., Kauertz, A., & Upmeyer Zu Belzen, A. (2018). Modelle und das Modellieren in den Naturwissenschaften. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 141–157). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56320-5_9
- Krumm, S., Schmidt-Atzert, L., & Amelang, M. (2021). Grundlagen diagnostischer Verfahren. In L. Schmidt-Atzert, S. Krumm, & M. Amelang (Hrsg.), *Psychologische Diagnostik* (S. 39–207). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61643-7_2
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (4., überarbeitete Aufl.). Beltz.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Lange, D. (2013). *Inhaltsanalytische Untersuchung zur Kooperation beim Bearbeiten mathematischer Problemaufgaben* (Online-Ausg.). Waxmann.
- Larson, G. E. (1996). Mental rotation of static and dynamic figures. *Perception & Psychophysics*, 58(1), 153–159. <https://doi.org/10.3758/BF03205484>
- Latscha, H. P., Kazmaier, U., & Klein, H. (2016). *Organische Chemie: Chemie-Basiswissen II*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-46180-8>
- Lichtenberger, A., Kokkonen, T., & Schalk, L. (2024). Learning with multiple external representations in physics: Concreteness fading versus simultaneous presentation. *Journal of Research in Science Teaching*, 61(9), 2258–2290. <https://doi.org/10.1002/tea.21947>
- Lichti, M. (2019). *Funktionales Denken fördern: Experimentieren mit gegenständlichen Materialien oder Computer-Simulationen*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-23621-2>

- Linn, M. C., & Petersen, A. C. (1985). Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis. *Child Development*, 56(6), 1479. <https://doi.org/10.2307/1130467>
- Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex Differences in Mental Rotation and Line Angle Judgments Are Positively Associated with Gender Equality and Economic Development Across 53 Nations. *Archives of Sexual Behavior*, 39(4), 990–997. <https://doi.org/10.1007/s10508-008-9460-8>
- Liu, X., & McKeough, A. (2005). Developmental growth in students' concept of energy: Analysis of selected items from the TIMSS database. *Journal of Research in Science Teaching*, 42(5), 493–517. <https://doi.org/10.1002/tea.20060>
- Lohman, D. F. (1979). *Spatial Ability: A Review and Reanalysis of the Correlational Literature*. School of Education, Stanford University.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58(2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>
- Lutter, J. C., Hale, L. V. A., & Shultz, G. V. (2019). Unpacking graduate students' knowledge for teaching solution chemistry concepts. *Chemistry Education Research and Practice*, 20(1), 258–269. <https://doi.org/10.1039/C8RP00205C>
- Maaß, M., Lanfermann, P., & Habig, S. (2022). Trendbericht: Chemiedidaktik 2022. *Nachrichten Aus Der Chemie*, 70(12), 17–25. <https://doi.org/10.1002/nadc.20224125445>
- Maeda, Y., Yoon, S. Y., Kim-Kang, K., & Imbrie, P. K. (2013). Psychometric Properties of the Revised PSVT: R for Measuring First Year Engineering Students' Spatial Ability. *International Journal of Engineering Education*, 29(3), 763–776.
- Maeyer, J., & Talanquer, V. (2013). Making predictions about chemical reactivity: Assumptions and heuristics. *Journal of Research in Science Teaching*, 50(6), 748–767. <https://doi.org/10.1002/tea.21092>
- Magritte, R. (1929). *The Treachery of Images (This is Not a Pipe)* [Öl auf Leinwand]. <https://collections.lacma.org/node/239578>

- Maier, U., Kleinknecht, M., Metz, K., & Bohl, T. (2010). Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben. *Beiträge zur Lehrerbildung*, 28(1), 84–96.
<https://doi.org/10.25656/01:13734>
- Mansour, N. (2024). Like or Not Like Studying Science: Exploring Students' Personal and Cultural Characteristics. *Canadian Journal of Science, Mathematics and Technology Education*, 24(3–4), 353–379.
<https://doi.org/10.1007/s42330-025-00352-6>
- Mantel, N., & Haenszel, W. (1959). Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease. *JNCI: Journal of the National Cancer Institute*. <https://doi.org/10.1093/jnci/22.4.719>
- Maresch, G. (2014). Strategies for Assessing Spatial Ability Tasks. *Journal for Geometry and Graphics*, 18(1), 125–132.
- Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Mayer, R. E. (2021). Cognitive Theory of Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (3. Aufl., S. 57–72). Cambridge University Press.
<https://doi.org/10.1017/9781108894333.008>
- Mayring, P., & Fenzl, T. (2022). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 691–706). Springer Fachmedien Wiesbaden.
https://doi.org/10.1007/978-3-658-37985-8_43
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>
- Meehl, P. E. (2013). The Power of Quantitative Thinking. In N. G. Waller, L. J. Yonce, W. M. Grove, D. Faust, & M. F. Lenzenweger (Hrsg.), *A Paul Meehl Reader: Essays on the Practice of Scientific Psychology* (S. 433–445). Taylor and Francis.
- Menon, V., Padmanabhan, A., & Schwartz, F. (2021). Cognitive neuroscience of dyscalculia and math learning disabilities. In K. Cohen Kadosh (Hrsg.), *The Oxford Handbook of Developmental Cognitive Neuroscience*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198827474.013.23>

- Merchant, Z., Goetz, E. T., Keeney-Kennicutt, W., Cifuentes, L., Kwok, O., & Davis, T. J. (2013). Exploring 3-D virtual reality technology for spatial ability and chemistry achievement. *Journal of Computer Assisted Learning*, 29(6), 579–590. <https://doi.org/10.1111/jcal.12018>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (Third edition). SAGE Publications, Inc.
- Moojen, S. M. P., Gonçalves, H. A., Bassôa, A., Navas, A. L., De Jou, G., & Miguel, E. S. (2020). Adults with dyslexia: How can they achieve academic success despite impairments in basic reading and writing abilities? The role of text structure sensitivity as a compensatory skill. *Annals of Dyslexia*, 70(1), 115–140. <https://doi.org/10.1007/s11881-020-00195-w>
- Moosbrugger, H., & Brandt, H. (2020). Antwortformate und Itemtypen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 91–117). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4_5
- Moosbrugger, H., Gäde, J. C., Schermelleh-Engel, K., & Rauch, W. (2020). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 275–304). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4_13
- Moosbrugger, H., & Kelava, A. (2020). Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien“). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 13–38). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4_2
- Moosbrugger, H., Schermelleh-Engel, K., Gäde, J. C., & Kelava, A. (2020). Testtheorien im Überblick. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 251–273). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4_12
- Moreau, D. (2013). Differentiating two- from three-dimensional mental rotation training effects. *Quarterly Journal of Experimental Psychology*, 66(7), 1399–1413. <https://doi.org/10.1080/17470218.2012.744761>
- Mortimer, C. E., & Müller, U. (2020). *Chemie: Das Basiswissen der Chemie* (13. vollständig überarbeitete Auflage). Thieme.

- Müller, A., Hettmannsperger, R., Scheid, J., & Schnotz, W. (2017). Representational Competence, Understanding of Experiments, Phenomena and Basic Concepts in Geometrical Optics: A Representational Approach. In D. F. Treagust, R. Duit, & H. E. Fischer (Hrsg.), *Multiple Representations in Physics Education* (Bd. 10, S. 209–229). Springer International Publishing. https://doi.org/10.1007/978-3-319-58914-5_10
- Nehring, A., & Schwichow, M. (2020). Was ist Wissen, was ist Können? Deutungen des Kompetenzbegriffs und deren psychometrische Konsequenzen im Kontext von Fachwissen und Variablenkontrollstrategie. *Zeitschrift für Didaktik der Naturwissenschaften*, 26(1), 73–87. <https://doi.org/10.1007/s40573-020-00113-y>
- Newcombe, N. S. (2010). Picture This: Increasing Math and Science Learning by Improving Spatial Thinking. *American Educator*, 34(2), 29–35.
- Newcombe, N. S., Booth, J. L., & Gunderson, E. A. (2019). Spatial Skills, Reasoning, and Mathematics. In J. Dunlosky & K. A. Rawson (Hrsg.), *The Cambridge Handbook of Cognition and Education* (1. Aufl., S. 100–123). Cambridge University Press. <https://doi.org/10.1017/9781108235631.006>
- Newcombe, N. S., & Shipley, T. F. (2015). Thinking About Spatial Thinking: New Typology, New Assessments. In J. S. Gero (Hrsg.), *Studying Visual and Spatial Reasoning for Design Creativity* (S. 179–192). Springer Netherlands. https://doi.org/10.1007/978-94-017-9297-4_10
- Newton, P. E., & Shaw, S. D. (2013). Standards for talking and thinking about validity. *Psychological Methods*, 18(3), 301–319. <https://doi.org/10.1037/a0032969>
- Nickel, S., Brockmüller, S., Boone, W. J., & Habig, S. (2025). Measuring representational competence – analyses of dimensionality and the relationship to general content knowledge. *International Journal of Science Education*, 1–25. <https://doi.org/10.1080/09500693.2025.2515621>
- Nicklin, C., & Vitta, J. P. (2022). Assessing Rasch measurement estimation methods across R packages with yes/no vocabulary test data. *Language Testing*, 39(4), 513–540. <https://doi.org/10.1177/02655322211066822>
- Nieminen, P. (2022). Application of Standardized Regression Coefficient in Meta-Analysis. *BioMedInformatics*, 2(3), 434–458. <https://doi.org/10.3390/biomedinformatics2030028>

- Nitz, S. (2012). *Fachsprache im Biologieunterricht: Eine Untersuchung zu Bedingungsfaktoren und Auswirkungen* [Dissertation]. Christian-Albrechts-Universität zu Kiel.
- Nitz, S., Ainsworth, S. E., Nerdel, C., & Prechtel, H. (2014). Do student perceptions of teaching predict the development of representational competence and biological knowledge? *Learning and Instruction*, 31, 13–22. <https://doi.org/10.1016/j.learninstruc.2013.12.003>
- Nitz, S., Prechtel, H., & Nerdel, C. (2014). Survey of classroom use of representations: Development, field test and multilevel analysis. *Learning Environments Research*, 17(3), 401–422. <https://doi.org/10.1007/s10984-014-9166-x>
- Nyachwaya, J. M., & Gillaspie, M. (2016). Features of representations in general chemistry textbooks: A peek through the lens of the cognitive load theory. *Chemistry Education Research and Practice*, 17(1), 58–71. <https://doi.org/10.1039/C5RP00140D>
- O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, 19, 1609406919899220. <https://doi.org/10.1177/1609406919899220>
- Ozdemir, B., & AlGhamdi, H. M. (2022). Investigating the Distractors to Explain DIF Effects Across Gender in Large-Scale Tests With Non-Linear Logistic Regression Models. *Frontiers in Education*, 6, 748884. <https://doi.org/10.3389/feduc.2021.748884>
- Paas, F., & Sweller, J. (2021). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (3. Aufl., S. 73–81). Cambridge University Press. <https://doi.org/10.1017/9781108894333.009>
- Pehkonen, E., Näveri, L., & Laine, A. (2013). On teaching problem solving in school mathematics. *CEPS Journal*, 3(4), 9–23. <https://doi.org/10.25656/01:8498>
- Pellegrino, J. W., Hunt, E. B., Abate, R., & Farr, S. (1987). A computer-based test battery for the assessment of static and dynamic spatial reasoning abilities. *Behavior Research Methods, Instruments, & Computers*, 19(2), 231–236. <https://doi.org/10.3758/BF03203790>

- Penfield, R. D. (2001). Assessing Differential Item Functioning Among Multiple Groups: A Comparison of Three Mantel-Haenszel Procedures. *Applied Measurement in Education*, 14(3), 235–259. https://doi.org/10.1207/S15324818AME1403_3
- Peoples, S. M., & O'Dwyer, L. M. (2014). The Nature of Science Instrument-Elementary (NOSI-E): The end of the road? *Journal of Applied Measurement*, 15(4), 338–358.
- Popova, M., & Jones, T. (2021). Chemistry instructors' intentions toward developing, teaching, and assessing student representational competence skills. *Chemistry Education Research and Practice*, 22(3), 733–748. <https://doi.org/10.1039/D0RP00329H>
- Pospeschill, M. (2022). *Testtheorie, Testkonstruktion, Testevaluation* (2., aktualisierte Auflage). Ernst Reinhardt Verlag. <https://doi.org/10.36198/9783838558332>
- Pribyl, J. R., & Bodner, G. M. (1987). Spatial ability and its role in organic chemistry: A study of four organic courses. *Journal of Research in Science Teaching*, 24(3), 229–240. <https://doi.org/10.1002/tea.3660240304>
- Radius, U., & Finze, M. (2014a). Charakterisierung von Koordinationsverbindungen. In J. Huheey, E. Keiter, & R. Keiter, *Anorganische Chemie* (S. 543–624). DE GRUYTER. <https://doi.org/10.1515/9783110307955.543>
- Radius, U., & Finze, M. (2014b). Reaktionen von Koordinationsverbindungen: Kinetik und Mechanismen. In J. Huheey, E. Keiter, & R. Keiter (Hrsg.), *Anorganische Chemie* (S. 685–756). DE GRUYTER. <https://doi.org/10.1515/9783110307955.685>
- Rapp, D. N. (2005). Mental Models: Theoretical Issues for Visualizations in Science Education. In J. K. Gilbert (Hrsg.), *Visualization in Science Education* (S. 43–60). Springer Netherlands. https://doi.org/10.1007/1-4020-3613-2_4
- Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.
- Rau, M. A. (2017a). A Framework for Educational Technologies that Support Representational Competencies. *IEEE Transactions on Learning Technologies*, 10(3), 290–305. <https://doi.org/10.1109/TLT.2016.2623303>
- Rau, M. A. (2017b). Conditions for the Effectiveness of Multiple Visual Representations in Enhancing STEM Learning. *Educational Psychology Review*, 29(4), 717–761. <https://doi.org/10.1007/s10648-016-9365-3>

- Rau, M. A. (2017c). Do Knowledge-Component Models Need to Incorporate Representational Competencies? *International Journal of Artificial Intelligence in Education*, 27(2), 298–319. <https://doi.org/10.1007/s40593-016-0134-8>
- Rau, M. A. (2018). Making connections among multiple visual representations: How do sense-making skills and perceptual fluency relate to learning of chemistry knowledge? *Instructional Science*, 46(2), 209–243. <https://doi.org/10.1007/s11251-017-9431-3>
- Rau, M. A., Alevén, V., & Rummel, N. (2015). Successful learning with multiple graphical representations and self-explanation prompts. *Journal of Educational Psychology*, 107(1), 30–46. <https://doi.org/10.1037/a0037211>
- Rau, M. A., Zahn, M., Misback, E., Herder, T., & Burstyn, J. (2021). Adaptive support for representational competencies during technology-based problem solving in chemistry. *Journal of the Learning Sciences*, 30(2), 163–203. <https://doi.org/10.1080/10508406.2021.1888733>
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1), 21–33.
- Reilly, D., Neumann, D. L., & Andrews, G. (2017). Gender Differences in Spatial Ability: Implications for STEM Education and Approaches to Reducing the Gender Gap for Parents and Educators. In M. S. Khine (Hrsg.), *Visual-spatial Ability in STEM Education* (S. 195–224). Springer International Publishing. https://doi.org/10.1007/978-3-319-44385-0_10
- Rexig, E., Kuhn, J., Becker, S., & Malone, S. (2024). The More the Better? A Systematic Review and Meta-Analysis of the Benefits of More than Two External Representations in STEM Education. *Educational Psychology Review*, 36(4), 124. <https://doi.org/10.1007/s10648-024-09958-y>
- Riedel, E., & Janiak, C. (2022). *Anorganische chemie* (10. Auflage). De Gruyter.
- Robitzsch, A. (2021). A Comprehensive Simulation Study of Estimation Methods for the Rasch Model. *Stats*, 4(4), 814–836. <https://doi.org/10.3390/stats4040048>
- Robitzsch, A., Kiefer, T., & Wu, M. (2024, Februar 19). *Package ‘TAM’*. <https://cran.r-project.org/web/packages/TAM/TAM.pdf>

- Rodemer, M., Eckhard, J., Graulich, N., & Bernholt, S. (2021). Connecting explanations to representations: Benefits of highlighting techniques in tutorial videos on students' learning in organic chemistry. *International Journal of Science Education*, 43(17), 2707–2728. <https://doi.org/10.1080/09500693.2021.1985743>
- Roff, M. (1952). A Factorial Study of Tests in the Perceptual Area. *Psychometric Monographs No.8*, 8(1), 1–41.
- Romaní-Romaní, F. (2024). Adaptación cultural al español del Perú y validación psicométrica de un instrumento para medir el alfabetismo científico en estudiantes de medicina [Cultural adaptation to Peruvian Spanish and psychometric validation of an instrument to measure scientific literacy in medical students]. *Anales de la Facultad de Medicina*, 85(1), 34–42. <https://doi.org/10.15381/anales.v85i1.26534>
- Rosseel, Y., De Wilde, L., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Rockwood, N., Scharf, F., Du, H., Jamil, H., & Classe, F. (2024). *Package 'lavaan'*. <https://cran.r-project.org/web/packages/lavaan/lavaan.pdf>
- Rost, J. (2004). *Lehrbuch Testtheorie, Testkonstruktion* (2., vollständig überarbeitete und erweiterte Aufl). H. Huber.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Russell, J. W., Kozma, R. B., Jones, T., Wykoff, J., Marx, N., & Davis, J. (1997). Use of Simultaneous-Synchronized Macroscopic, Microscopic, and Symbolic Representations To Enhance the Teaching and Learning of Chemical Concepts. *Journal of Chemical Education*, 74(3), 330. <https://doi.org/10.1021/ed074p330>
- Salame, I. I., & Kabir, S. A. (2022). Examining Students' Spatial Ability and Its Impact on the Learning of Stereochemistry. *Interdisciplinary Journal of Environmental and Science Education*, 18(4), e2288. <https://doi.org/10.21601/ijese/12099>
- Sandmann, A. (2014). Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 179–188). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37827-0_15

- Schanze, S., & Girwidz, R. (2018). Lernen mit digitalen Medien. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Theorien in der naturwissenschafts-didaktischen Forschung* (S. 177–192). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56320-5_11
- Scheid, J., Müller, A., Hettmannsperger, R., & Schnotz, W. (2018). Representational Competence in Science Education: From Theory to Assessment. In K. L. Daniel (Hrsg.), *Towards a Framework for Representational Competence in Science Education* (Bd. 11, S. 263–277). Springer International Publishing. https://doi.org/10.1007/978-3-319-89945-9_13
- Schermelleh-Engel, K., & Gade, J. C. (2020). Modellbasierte Methoden der Reliabilitätsschätzung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 335–368). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4_15
- Schmuck, C. (2013). *Basisbuch Organische Chemie*. Pearson Deutschland.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll Model of Intelligence. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues* (3. ed, S. 99–144). Guilford Press.
- Schnotz, B., Baadte, C., & Müller, A. (2010). Creative thinking and problem solving with depictive and descriptive representations. In L. Verschaffel, E. De Corte, T. De Jong, & J. Elen (Hrsg.), *Use of Representations in Reasoning and Problem Solving* (0 Aufl., S. 11–35). Routledge. <https://doi.org/10.4324/9780203847824>
- Schnotz, W. (2021). Integrated Model of Text and Picture Comprehension. In R. E. Mayer & L. Fiorella (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (3. Aufl., S. 82–99). Cambridge University Press. <https://doi.org/10.1017/9781108894333.010>
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)
- Schönborn, K. J., & Anderson, T. R. (2009). A Model of Factors Determining Students' Ability to Interpret External Representations in Biochemistry. *International Journal of Science Education*, 31(2), 193–232. <https://doi.org/10.1080/09500690701670535>
- Schuttkowski, C., & Staubach, K. (2022). Lautes Denken. In M. Beisswenger, L. Lemnitzer, & C. Müller-Spitzer (Hrsg.), *Forschen in der Linguistik: Eine Methodeneinführung für das Germanistik-Studium* (S. 338–349). UTB.

- Schweda, E. (2022a). *Jander/Blasius—Anorganische Chemie I: Theoretische Grundlagen und Qualitative Analyse* (19., völlig neu bearbeitete Auflage). Hirzel Verlag.
- Schweda, E. (2022b). *Jander/Blasius—Anorganische Chemie II: Quantitative Analyse und Präparate* (18., völlig neu bearbeitete Auflage). Hirzel Verlag.
- Sclove, S. L. (1987). Application of Model-Selection Criteria to Some Problems in Multivariate Analysis. *Psychometrika*, 52(3), 333–343.
<https://doi.org/10.1007/BF02294360>
- Sim, J. H., & Daniel, E. G. S. (2014). Representational competence in chemistry: A comparison between students with different levels of understanding of basic chemical concepts and chemical representations. *Cogent Education*, 1(1), 991180.
<https://doi.org/10.1080/2331186X.2014.991180>
- SINUS. (2024). *Was motiviert zum MINT-Lernen? Eine qualitativ-quantitative SINUS-Studie unter 10- bis 16-Jährigen* (Deutsche Telekom Stiftung, Hrsg.).
- Solaz-Portolés, J. J., & López, V. S. (2008). Types of knowledge and their relations to problem solving in science: Directions for practice. *Sísifo. Educational Sciences Journal*, 6, 105–112.
- Sorby, S. (2009). Educational Research in Developing 3-D Spatial Skills for Engineering Students. *International Journal of Science Education*, 31(3), 459–480. <https://doi.org/10.1080/09500690802595839>
- Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, 26, 20–29.
<https://doi.org/10.1016/j.lindif.2013.03.010>
- Spearman, C. (1904). „General Intelligence,“ Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201.
<https://doi.org/10.2307/1412107>
- Spearman, C. (1927). *The Abilitie of Man—Their Nature and Measurement*. Macmillan.
- Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (Hrsg.). (2020). *Bildungsstandards im Fach Chemie für die Allgemeine Hochschulreife: Beschluss der Kultusministerkonferenz vom 18.06.2020* (1. Auflage). Carl Link Verlag.

- Steinbach, M., Eitemüller, C., Rodemer, M., & Walpuski, M. (2024). The influence of representations on task difficulty in organic chemistry: An exploration using a novel paired-items test instrument. *International Journal of Science Education*, 1–23.
<https://doi.org/10.1080/09500693.2024.2378218>
- Steudel, R. (1973). *Chemie der Nichtmetalle*. De Gruyter.
<https://doi.org/10.1515/9783112419380>
- Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, 17(2), 219–234. <https://doi.org/10.1016/j.learninstruc.2007.01.012>
- Stieff, M., & DeSutter, D. (2021). Sketching, not representational competence, predicts improved science learning. *Journal of Research in Science Teaching*, 58(1), 128–156. <https://doi.org/10.1002/tea.21650>
- Stieff, M., Origenes, A., DeSutter, D., Lira, M., Banevicius, L., Tabang, D., & Cabel, G. (2018). Operational constraints on the mental rotation of STEM representations. *Journal of Educational Psychology*, 110(8), 1160–1174. <https://doi.org/10.1037/edu0000258>
- Stieff, M., Ryu, M., Dixon, B., & Hegarty, M. (2012). The Role of Spatial Ability and Strategy Preference for Spatial Problem Solving in Organic Chemistry. *Journal of Chemical Education*, 89(7), 854–859.
<https://doi.org/10.1021/ed200071d>
- Stieff, M., Scopelitis, S., Lira, M. E., & Desutter, D. (2016). Improving Representational Competence with Concrete Models. *Science Education*, 100(2), 344–363. <https://doi.org/10.1002/sce.21203>
- Stieff, M., & Uttal, D. (2015). How Much Can Spatial Training Improve STEM Achievement? *Educational Psychology Review*, 27(4), 607–615.
<https://doi.org/10.1007/s10648-015-9304-8>
- Strähle, J., & Schweda, E. (2006). *Jander/Blasius—Lehrbuch der analytischen und präparativen anorganischen Chemie* (16., überarb. Aufl). Hirzel.
- Streiner, D. L. (2005). Finding Our Way: An Introduction to Path Analysis. *The Canadian Journal of Psychiatry*, 50(2), 115–122.
<https://doi.org/10.1177/070674370505000207>
- Strübing, J. (2022). Grounded Theory und Theoretical Sampling. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 587–606). Springer Fachmedien Wiesbaden.
https://doi.org/10.1007/978-3-658-37985-8_37

- Stull, A. T., Gainer, M., Padalkar, S., & Hegarty, M. (2016). Promoting Representational Competence with Molecular Models in Organic Chemistry. *Journal of Chemical Education*, 93(6), 994–1001. <https://doi.org/10.1021/acs.jchemed.6b00194>
- Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational Translation With Concrete Models in Organic Chemistry. *Cognition and Instruction*, 30(4), 404–434. <https://doi.org/10.1080/07370008.2012.719956>
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections: Further analysis of the data by akaike' s. *Communications in Statistics - Theory and Methods*, 7(1), 13–26. <https://doi.org/10.1080/03610927808827599>
- Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation* (Bd. 55, S. 37–76). Elsevier. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Taber, K. S. (2009). Learning at the Symbolic Level. In J. K. Gilbert & D. Treagust (Hrsg.), *Multiple Representations in Chemical Education* (Bd. 4, S. 75–105). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8872-8_5
- Talanquer, V. (2011). Macro, Submicro, and Symbolic: The many faces of the chemistry “triplet”. *International Journal of Science Education*, 33(2), 179–195. <https://doi.org/10.1080/09500690903386435>
- Talanquer, V. (2022). The Complexity of Reasoning about and with Chemical Representations. *JACS Au*, 2(12), 2658–2669. <https://doi.org/10.1021/jacsau.2c00498>
- Taskin, V., Bernholt, S., & Parchmann, I. (2015). An inventory for measuring student teachers' knowledge of chemical representations: Design, validation, and psychometric analysis. *Chemistry Education Research and Practice*, 16(3), 460–477. <https://doi.org/10.1039/C4RP00214H>

- 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 : ZfDN*, 18, 7–28.
<https://doi.org/10.25656/01:31747>
- Thorndike, E. L. (1927). *The Measurement of Intelligence*. Teachers College, Columbia University.
- Thurstone, L. L. (1938). *Primary Mental Abilities*. The University of Chicago Press.
- Tobinski, D. (2017). *Kognitive Psychologie: Problemlösen, Komplexität und Gedächtnis* (1. Aufl. 2017). Springer Berlin Heidelberg.
<https://doi.org/10.1007/978-3-662-53948-4>
- Treagust, D., Chittleborough, G., & Mamiala, T. (2003). The role of submicroscopic and symbolic representations in chemical explanations. *International Journal of Science Education*, 25(11), 1353–1368.
<https://doi.org/10.1080/0950069032000070306>
- Trendtel, M., Pham, G., & Yanagida, T. (2016). Skalierung und Linking. In S. Breit & C. Schreiner (Hrsg.), *Large-Scale Assessment mit R - Methodische Grundlagen der österreichischen Bildungsstandardüberprüfung [Large-Scale Assessment with R - Methodological principles of the Austrian educational standard review]* (S. 185–224). Facultas.
- Tsui, C.-Y. (2003). *Teaching and Learning Genetics with Multiple Representations* [Unveröffentlichte Dissertation, Curtin University of Technology].
https://espace.curtin.edu.au/bitstream/handle/20.500.11937/1614/14027_Tsui_combined%20aha.pdf?sequence=2&isAllowed=y
- Tsui, C.-Y., & Treagust, D. F. (2013). Introduction to Multiple Representations: Their Importance in Biology and Biological Education. In D. F. Treagust & C.-Y. Tsui (Hrsg.), *Multiple Representations in Biological Education* (Bd. 7, S. 3–18). Springer Netherlands. https://doi.org/10.1007/978-94-007-4192-8_1
- Tucker, L. R., & Lewis, C. (1973). A Reliability Coefficient for Maximum Likelihood Factor Analysis. *Psychometrika*, 38(1), 1–10.
<https://doi.org/10.1007/bf02291170>

- Uttal, D. H., & Cohen, C. A. (2012). Spatial Thinking and STEM Education. In B. H. Ross (Hrsg.), *Psychology of Learning and Motivation* (Bd. 57, S. 147–181). Elsevier. <https://doi.org/10.1016/B978-0-12-394293-7.00004-2>
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139(2), 352–402. <https://doi.org/10.1037/a0028446>
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills*, 47(2), 599–604. <https://doi.org/10.2466/pms.1978.47.2.599>
- Vollhardt, K. P. C., & Schore, N. E. (2011). *Organische Chemie* (5. Aufl). Wiley-VCH.
- Vollrath, H.-J. (1992). Zur Rolle des Begriffs im Problemlöseprozeß des Beweissens. *Mathematische Semesterberichte*, 39(2), 127–136. <https://doi.org/10.1007/BF03186465>
- Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117(2), 250–270. <https://doi.org/10.1037/0033-2909.117.2.250>
- Vukić, Đ., Martinčić-Ipšić, S., & Meštrović, A. (2020). Structural Analysis of Factual, Conceptual, Procedural, and Metacognitive Knowledge in a Multidimensional Knowledge Network. *Complexity*, 2020, 1–17. <https://doi.org/10.1155/2020/9407162>
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817–835. <https://doi.org/10.1037/a0016127>
- Wang, L., & Carr, M. (2014). Working Memory and Strategy Use Contribute to Gender Differences in Spatial Ability. *Educational Psychologist*, 49(4), 261–282. <https://doi.org/10.1080/00461520.2014.960568>
- Wang, W.-C., & Su, Y.-H. (2004). Factors Influencing the Mantel and Generalized Mantel-Haenszel Methods for the Assessment of Differential Item Functioning in Polytomous Items. *Applied Psychological Measurement*, 28(6), 450–480. <https://doi.org/10.1177/0146621604269792>

- Ward, L., Rotich, F., Hoang, J., & Popova, M. (2022). Representational Competence Under the Magnifying Glass—The Interplay Between Student Reasoning Skills, Conceptual Understanding, and the Nature of Representations. In N. Graulich & G. Shultz (Hrsg.), *Student Reasoning in Organic Chemistry* (1. Aufl., S. 36–56). The Royal Society of Chemistry. <https://doi.org/10.1039/9781839167782-00036>
- Ward, L., Rotich, F., Raker, J. R., Komperda, R., Nedungadi, S., & Popova, M. (2025). Design, development, and evaluation of the organic chemistry representational competence assessment (ORCA). *Chemistry Education Research and Practice*, 26(1), 244–258. <https://doi.org/10.1039/D3RP00188A>
- Warm, T. A. (1989). Weighted Likelihood Estimation of Ability in Item Response Theory. *Psychometrika*, 54(3), 427–450. <https://doi.org/10.1007/BF02294627>
- Weber, B. (2021). *Koordinationschemie: Grundlagen und aktuelle Trends*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-63819-4>
- Williams, R. H., Zimmerman, D. W., Zumbo, B. D., & Ross, D. (2003). Charles Spearman: British Behavioral Scientist. *The Human Nature Review*, 3, 114–118.
- Willis, J. O., Dumont, R., & Kaufman, A. S. (2011). Factor-Analytic Models of Intelligence. In R. J. Sternberg & S. B. Kaufman (Hrsg.), *The Cambridge Handbook of Intelligence* (1. Aufl., S. 39–57). Cambridge University Press. <https://doi.org/10.1017/CBO9780511977244.004>
- Witte, K. (2019a). Grundlagen der Testtheorie und Testkonstruktion. In K. Witte, *Angewandte Statistik in der Bewegungswissenschaft (Band 3)* (S. 139–158). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-58360-9_10
- Witte, K. (2019b). Statistische Verfahren zur Überprüfung von Zusammenhangshypothesen bei zwei Stichproben. In K. Witte, *Angewandte Statistik in der Bewegungswissenschaft (Band 3)* (S. 79–95). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-58360-9_7
- Wu, H., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, 88(3), 465–492. <https://doi.org/10.1002/sce.10126>

- Wu, H.-K., & Puntambekar, S. (2012). Pedagogical Affordances of Multiple External Representations in Scientific Processes. *Journal of Science Education and Technology*, 21(6), 754–767. <https://doi.org/10.1007/s10956-011-9363-7>
- Wu, S. P. W., & Rau, M. A. (2018). Effectiveness and efficiency of adding drawing prompts to an interactive educational technology when learning with visual representations. *Learning and Instruction*, 55, 93–104. <https://doi.org/10.1016/j.learninstruc.2017.09.010>
- Yang, C.-C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis*, 50(4), 1090–1104. <https://doi.org/10.1016/j.csda.2004.11.004>
- Yang, W., Liu, H., Chen, N., Xu, P., & Lin, X. (2020). Is Early Spatial Skills Training Effective? A Meta-Analysis. *Frontiers in Psychology*, 11, 1938. <https://doi.org/10.3389/fpsyg.2020.01938>
- Yen, W. M. (1984). Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model. *Applied Psychological Measurement*, 8(2), 125–145. <https://doi.org/10.1177/014662168400800201>
- Yoon, S. Y. (2011a). *Psychometric properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (the Revised PSVT:R)*. Purdue University.
- Yoon, S. Y. (2011b). *Revised Purdue Spatial Visualization Test: Visualization of Rotations (Revised PSVT:R)* [Psychometric Instrument].
- Yuan, K.-H., & Bentler, P. M. (1998). Structural Equation Modeling with Robust Covariances. *Sociological Methodology*, 28(1), 363–396. <https://doi.org/10.1111/0081-1750.00052>
- Yue, J. (2008). Spatial Visualization by Realistic 3D Views. *Engineering Design Graphics Journal*, 72(1), 28–38.
- Zhang, J. (2001). External Representations in Complex Information Processing Tasks. In A. Kent & C. M. Hall (Hrsg.), *Encyclopedia of library and information science* (S. 164–180). M. Dekker.
- Zheng, B. Q., & Bentler, P. M. (2025). Enhancing Model Fit Evaluation in SEM: Practical Tips for Optimizing Chi-Square Tests. *Structural Equation Modeling: A Multidisciplinary Journal*, 32(1), 136–141. <https://doi.org/10.1080/10705511.2024.2354802>

9.2 Abbildungsverzeichnis

Abbildung 1:	La trahison des images (Der Verrat der Bilder) mit dem Schriftzug „Ceci n'est pas une pipe“ („Dies ist keine Pfeife“) von René Magritte (1929).	12
Abbildung 2:	Synthese der Kategorisierung von Repräsentationen und Charakterisierung durch den Abstraktionsgrad (¹ Krey & Schwanewedel, 2018; ² Schnotz & Bannert, 2003; ³ Talanquer, 2022; ⁴ Tsui, 2003; ⁵ Wu & Puntambekar, 2012), adaptiert nach Nickel et al. (2025).	15
Abbildung 3:	Unterschiedliche Repräsentation eines (S)-Milchsäuremoleküls ((S)-2-Hydroxypropansäure), charakterisiert und geordnet nach dem Abstraktionsgrad (adaptiert nach Nickel et al., 2025).	19
Abbildung 4:	S_N1 -Mechanismus der Reaktion von (S)-3-Chlor-3-methylhexan mit Hydroxidionen (OH^-) über ein tertiäres Carbeniumion als Zwischenstufe zu (R)- und (S)-3-Methylhexan-3-ol (adaptiert nach Latscha et al., 2016).	20
Abbildung 5:	Modell der <i>Cognitive Theory of Multimedia Learning</i> nach Mayer (2021).	24
Abbildung 6:	Titrationsskurve der Titration von 50 mL Essigsäurelösung ($c = 0.1 \text{ mol} \cdot \text{L}^{-1}$) mit Natronlauge ($c = 0.1 \text{ mol} \cdot \text{L}^{-1}$) mit markierten charakteristischen Punkten und Bereichen (adaptiert nach Mortimer & Müller, 2020, S. 310 f.).	29
Abbildung 7:	Skelettformel, Keilstrichformel und Kugel-Stab-Modell eines Ethanolmoleküls (C_2H_5OH) mit Informationen zur Interpretation spezifischer Repräsentationsmerkmale (verändert nach Schmuck, 2013).	37
Abbildung 8:	Hierarchische Struktur des Cattell-Horn-Carroll Modells mit Hervorhebung der räumlichen Fähigkeiten (<i>Visual Processing</i> , <i>Gv</i>) nach Schneider und McGrew (2012) sowie Buckley et al. (2018).	42

Abbildung 9:	Beispiel-Item aus dem Revised Purdue Spatial Visualization Test: Visualization of Rotation (Yoon, 2011b) zur Messung der dreidimensionalen Rotationsfähigkeit.....	46
Abbildung 10:	Theoretisch abgeleitetes Modell, dass im Rahmen von Forschungsanliegen II untersucht und geprüft wird.	58
Abbildung 11:	Item Int_01 als Beispielitem für die Interpretation (selbst gestaltet).....	78
Abbildung 12:	Item Trans_07 als Beispielitem für die Translation (selbst gestaltet).....	79
Abbildung 13:	Item Konst_05 als Beispielitem für die Konstruktion (selbst gestaltet).....	80
Abbildung 14:	Item Fach_07 als Beispielitem für die ein Fachwissens-item (Averbeck, 2021).	82
Abbildung 15:	Verteilung der Studierenden auf die geplanten Studiengänge der Datenerhebung für das Forschungsanliegen I im Oktober 2023. „Sonstiges“ umfasst unter anderem die Studienfächer Biologie und Physik, weitere lehramts-bezogene Studiengänge sowie zusätzliche, nicht näher spezifizierte Studiengänge.	96
Abbildung 16:	Cross Plots mit Messfehlerlinien (CI = 95 %) und Pearson-Korrelation (r) der Personenfähigkeiten (θ) aus dem eindimensionalen Modell (1D) auf der Ordinate und der Personenfähigkeiten aus dem dreidimensionalen Modell (3D, oben) bzw. dem zweidimensionalen Modell (2D, unten) auf der Abszisse.	104
Abbildung 17:	Cross Plots mit Messfehlerlinien (CI = 95 %) und Pearson-Korrelation (r) der Personenfähigkeiten (θ) aus dem eindimensionalen Modell (1D) auf der Ordinate und der Personenfähigkeiten für die Repräsentationskompetenzen (RK) und das Fachwissen (FW) aus dem zweidimensionalen Modell (2D) auf der Abszisse.....	110

Abbildung 18:	Verteilung der Personen getrennt nach Geschlechtern auf die drei Erhebungsorte in der Datenerhebung vom September/ Oktober 2024.	138
Abbildung 19:	Verteilung der Studierenden auf die geplanten Studiengänge in der Datenerhebung vom September/Oktober 2024.....	139
Abbildung 20:	Cross Plots mit Messfehlerlinien (CI = 95 %) und Pearson-Korrelation (r) der Personenfähigkeiten (θ) aus dem eindimensionalen Modell (1D) auf der Ordinate und der Personenfähigkeiten aus dem dreidimensionalen Modell (3D, oben) bzw. dem zweidimensionalen Modell (2D, unten) auf der Abszisse (Daten der Hauptstudie, September/Oktober 2024).	146
Abbildung 21:	Cross Plots mit Messfehlerlinien (CI = 95 %) und Pearson-Korrelation (r) der Personenfähigkeiten (θ) aus dem eindimensionalen Modell (1D) auf der Ordinate und der Personenfähigkeiten für die Repräsentationskompetenzen (RK) und das Fachwissen (FW) aus dem zweidimensionalen Modell (2D) auf der Abszisse (Daten der Hauptstudie, September/Oktober 2024).	149
Abbildung 22:	Pfadmodell mit standardisierten Pfadkoeffizienten (β) für signifikante Pfade (nicht-signifikante Pfade sind transparent dargestellt) und den Modellparametern zu Beurteilung der Güte.	155
Abbildung 23:	Pfadmodell mit standardisierten Pfadkoeffizienten (β) für signifikante Pfade und den Modellparametern zu Beurteilung der Güte für die drei lower-level skills Interpretation, Translation und Konstruktion.	157
Abbildung 24:	Beispiele für Komplexe mit unterschiedlicher Anzahl an Liganden und entsprechend unterschiedlichen Strukturen und Isomeren (erstellt mit ChemDraw Professional, Version 23.1.2.7).	174

Abbildung 25: Ablauf der qualitativen Inhaltsanalyse mit induktiver Kategorienbildung (links) bzw. deduktiver Kategorienanwendung (rechts), (adaptiert nach Mayring & Fenzl, 2022, S. 699).	182
Abbildung 26: Falsche Übersetzung des als Kugel-Stab-Modell dargestellten quadratisch-planaren Komplexes $[\text{Cu}(\text{NH}_3)_4]^{2+}$ (links) in die Keilstrichformel der Testpersonen SHFS79 (mittig) und BCFH58 (rechts).....	207
Abbildung 27: Falsche Übersetzung des als Kugel-Stab-Modell dargestellten tetraedrischen Komplexes $[\text{Be}(\text{H}_2\text{O})_4]^{2+}$ (links) in die Keilstrichformel (von links nach rechts: NWNZ30, SJMK52, ENFP62).	207
Abbildung 28: Skizzierte Isomere, die auf falschen bzw. falsch interpretierten Definitionen basieren. Links ein Beispiel für Konformationsisomerie (SDSB66) und rechts für die Enantiomerie (SHFS79).	210
Abbildung 29: <i>fac</i> -Isomer (rot umkreist) und acht der Testperson zufolge unterschiedliche <i>mer</i> -Isomere (blau umkreist) von $[\text{RhCl}_3(\text{H}_2\text{O})_3]$, die jedoch alle identisch sind und durch Rotation ineinander überführt werden können (CNTR37, Abs. 56).	211
Abbildung 30: Auszug aus dem Transkript von NJBS00 (Abs. 27) mit den entsprechenden Skizzen aus dem Testheft zu den entsprechenden Textabschnitten.	212
Abbildung 31: Zentrale Hindernisse bei der Interpretation, Translation und Konstruktion von repräsentationsbasierten Komplexchemieaufgaben in Forschungsanliegen III und Einordnung einschlägiger empirischer Befunde (vornehmlich aus der Organischen Chemie) in diesem Kontext. Studien mit allgemeineren Aussagen zu den aufgeführten Hindernissen in Bezug auf Repräsentationen und nicht spezifisch für die Interpretation, Translation und Konstruktion werden hier nicht aufgenommen.	220

Abbildung 32: Tippkärtchen „Isomerie – Was war das noch einmal?“ als unterstützendes Material für die Bearbeitung der Aufgaben 2 (Translation) und 3 (Konstruktion) in Forschungsanliegen III (Binnewies et al., 2016; Mortimer & Müller, 2020; Riedel & Janiak, 2022; Schmuck, 2013; Weber, 2021).	324
---	-----

9.3 Tabellenverzeichnis

Tabelle 1:	Übersicht der Repräsentationskompetenzen nach ¹ Kozma und Russell (1997; 2005), kategorisiert nach ² Nitz (2012) und ³ Gurung et al. (2022).	33
Tabelle 2:	Definition der elf räumlichen Faktoren, die auf die räumlichen Fähigkeiten im Cattell-Horn-Carroll Modells laden (Flanagan & Dixon, 2014; McGrew, 2009; Schneider & McGrew, 2012).	43
Tabelle 3:	Operationalisierung der lower-level skills (¹ Gurung et al., 2022; ² Kozma & Russell, 1997, 2005; ³ Nitz, 2012) mithilfe der Charakterisierung von Repräsentationen (Abbildung 2).	73
Tabelle 4:	Herkunft der 38 Items des CRI:TIC und Zuordnung zu den drei lower-level skills.	76
Tabelle 5:	Vergleich der durchschnittlichen Personenfähigkeiten (θ) und Itemschwierigkeiten (ξ) für die Interpretation, Translation und Konstruktion des dreidimensionalen Grundmodells.	97
Tabelle 6:	Evaluation der Items des CRI:TIC hinsichtlich der lokalen stochastischen Unabhängigkeit ($\alpha Q3$), Itemschwierigkeit (ξ) und Itemfit (<i>infit</i> und <i>outfit</i>).	98
Tabelle 7:	Vergleich der lokalen stochastischen Unabhängigkeit ($MAD\alpha Q3$, $\alpha Q3$), der WLE-Reliabilität und des Modells ($SRMR$) sowie Itemfits (<i>outfit</i> und <i>infit MNSQ</i>) für die drei Modelle (FF1.1).	101
Tabelle 8:	Vergleich der drei Modelle anhand des $AICc$ und des $saBIC$, basierend auf der Log-Likelihood (LL), der geschätzten Modellparameter (n_p) und der Stichprobengröße (N).	103
Tabelle 9:	Evaluation der Fachwissensitems hinsichtlich der lokalen stochastischen Unabhängigkeit ($\alpha Q3$), Itemschwierigkeit (ξ) und Itemfit (<i>infit</i> und <i>outfit</i>).	106

Tabelle 10: Vergleich der lokalen stochastischen Unabhängigkeit (<i>MADaQ3</i> , <i>aQ3</i>), der WLE-Reliabilität und des Modells (<i>SRMR</i>) und Itemfits (<i>outfit</i> und <i>infit MNSQ</i>) für die zwei Modelle (FF1.2).	107
Tabelle 11: Vergleich der zwei Modelle anhand des <i>AICc</i> und des <i>saBIC</i> , basierend auf der Log-Likelihood (<i>LL</i>), der geschätzten Modellparameter (n_p) und der Stichprobengröße (<i>N</i>).	109
Tabelle 12: Übersicht der eingesetzten Messinstrumente zur Erhebung der acht fokussierten räumlichen Faktoren mit Zeitangaben. .	126
Tabelle 13: Evaluation der Items des CRI:TIC hinsichtlich der lokalen stochastischen Unabhängigkeit (<i>aQ3</i>), <i>DIF</i> (Generalized Mantel-Haenszel χ^2 -Tests) und des Itemfits (<i>infit</i> und <i>outfit Statistik</i>).	142
Tabelle 14: Vergleich der lokalen stochastischen Unabhängigkeit (<i>MADaQ3</i> , <i>aQ3</i>), der WLE-Reliabilität und des Modells (<i>SRMR</i>) sowie Itemfits (<i>outfit</i> und <i>infit MNSQ</i>) für die drei Modelle aus FF1.1.	144
Tabelle 15: Vergleich der drei Modelle in der Hauptstudie anhand des <i>AICc</i> und des <i>saBIC</i> , basierend auf der Log-Likelihood (<i>LL</i>), der geschätzten Modellparameter (n_p) und der Stichprobengröße (<i>N</i>).	145
Tabelle 16: Evaluation der Fachwissensitems hinsichtlich der lokalen stochastischen Unabhängigkeit (<i>aQ3</i>), <i>DIF</i> (Mantel-Haenszel χ^2 -Tests) und des Itemfits (<i>infit</i> und <i>outfit Statistik</i>).	147
Tabelle 17: Vergleich der lokalen stochastischen Unabhängigkeit (<i>MADaQ3</i> , <i>aQ3</i>), der WLE-Reliabilität und des Modells (<i>SRMR</i>) und Itemfits (<i>outfit</i> und <i>infit MNSQ</i>) für die zwei Modelle (FF1.2).	148
Tabelle 18: Vergleich der zwei Modelle in der Hauptstudie anhand des <i>AICc</i> und des <i>saBIC</i> , basierend auf der Log-Likelihood (<i>LL</i>), der geschätzten Modellparameter (n_p) und der Stichprobengröße (<i>N</i>).	149

Tabelle 19: Stichprobengröße (N), Anzahl der Items und Cronbachs α mit Referenzwerten für die Messungen der acht räumlichen Faktoren in der Datenerhebung vom September/Okttober 2024.	151
Tabelle 20: Spearmans Rangkorrelationskoeffizient ρ mit Bonferroni-Korrektur zwischen den räumlichen Faktoren und den Repräsentationskompetenzen sowie dem Fachwissen.....	152
Tabelle 21: Übersicht der gesichteten Lehrbücher hinsichtlich des Einsatzes visuell-graphischer Repräsentationen zur Darstellung von Komplexen.	176
Tabelle 22: Finales Kategoriensystem zur systematischen Identifikation und Beschreibung von Hindernissen bei der Interpretation, Translation und Konstruktion von Keilstrichformeln und Kugel-Stab-Modellen in der Komplexchemie mit Textbeispielen zur Veranschaulichung.	190
Tabelle 23: Übersicht mit der absoluten Anzahl an Personen für jede Kategorie sowie der Anzahl der Kodierungen jeder Kategorie getrennt nach Aufgaben bei der Introspektion (d. h. ohne die Reflexionsaufgaben mit Retrospektion).....	195
Tabelle 24: Charakterisierung der Testpersonen durch persönliche Angaben, erreichte Punkte in den Aufgaben, Anzahl vorhandener Hinderniskategorien und dem Gesamtumfang der Kodierungen bei der Introspektion (d. h. ohne die Reflexionsaufgaben mit Retrospektion).....	197
Tabelle 25: Übersicht der im CRI:TIC vorkommenden Repräsentationen mit Informationen zu deren Konventionen, die dem Messinstrument als unterstützendes Material beilagen.	303
Tabelle 26: Manual zur Auswertung und Bepunktung des <i>Chemical Representation Inventory: Translation, Interpretation, Construction</i> (CRI:TIC, s. Anhang A.1).	306
Tabelle 27: Deskriptive Statistiken und Ergebnisse des Shapiro-Wilk Tests auf Normalverteilung für die acht Instrumente zur Messung der räumlichen Faktoren in Forschungsanliegen II.	310

Tabelle 28:	Deskriptive Statistiken und Ergebnisse des Shapiro-Wilk Tests auf Normalverteilung für die Repräsentationskompetenzen und das deklarative Fachwissen in Forschungsanliegen II.	311
--------------------	---	-----

9.4 Abkürzungsverzeichnis

1PL	Ein-Parameter-Logistisches-Modell
2PL	Zwei-Parameter-Logistisches-Modell
3PL	Drei-Parameter-Logistisches-Modell
AIC	Akaike Information Criterion
AICc	Bias-corrected Akaike Information Criterion
BIC	Bayesian Information Criterion
CFI	Comparative-Fit-Index
CMLE	Conditional Maximum Likelihood Estimator
	<i>Chemical Representation Inventory: Translation, Interpretation, Construction</i>
CRI:TIC	
EAP	Expected a-posteriori (Reliabilität)
FAU	Friedrich-Alexander-Universität Erlangen-Nürnberg
GMH	Generalized Mantel-Haenszel Methode
IC	Informationskriterium
IC-Funktion	Itemcharakteristische Funktion
Int	<i>Interpretation</i>
IRT	Item-Response Theorie
	y_i Manifeste Variable des Items i
	γ_i Rateparameter des Items i
	θ Personenfähigkeitsparameter
	ξ_i Schwierigkeitsparameter des Items i
JMLE	Joint Maximum Likelihood Estimator
Konst	<i>Konstruktion</i>
KTT	Klassische Testtheorie
	Testwert einer Person (ν) für einen Test oder ein Item (i) in der
	$y_{\nu i}$ KTT
	Trennschärfeparameter für einen Test oder ein Item (i) in der
	α_i KTT
	Messfehler einer Person (ν) für einen Test oder ein Item (i) in
	$\varepsilon_{\nu i}$ der KTT
	η Latente Variable
	λ_i Modellparameter für einen Test oder ein Item (i) in der KTT
	Wahrer Wer einer Person (ν) für einen Test oder ein Item (i) in
	$\tau_{\nu i}$ der KTT
LL	Log-Likelihood

<i>M</i>	Mittelwert
<i>MADaQ3</i>	Mean Absolute Deviation of the adjusted <i>Q3</i> statistic
<i>MH</i>	Mantel-Haenszel Methode
ML-Verfahren	Maximum-Likelihood-Verfahren
MMLE	Marginal Maximum Likelihood Estimator
<i>MNSQ</i>	Mean Squares
<i>N</i>	Stichprobengröße
<i>n</i>	Teilstichprobengröße
<i>n_P</i>	Geschätzte Modellparameter
PCM	Partial-Credit-Modell
<i>r</i>	Korrelationskoeffizient
<i>RMSEA</i>	Root Mean Square Error of Approximation
<i>saBIC</i>	sample size adjusted Bayes Information Criterion
<i>SD</i>	Standardabweichung
<i>SRMR</i>	Standardized Root Mean <i>Square Residual</i>
<i>TLI</i>	Tucker-Lewis-Index
Trans	<i>Translation</i>
UDE	Universität Duisburg-Essen
UR	Universität Regensburg
WLE	Weighted Likelihood Estimator
<i>ZSTD</i>	z-standardisierter Wert
α	Cronbachs Alpha
ρ	Spearman's Rangkorrelationskoeffizient (Rho)

10

Anhang

A	Anhang zu Forschungsanliegen I	280
A.1	Das Chemical Representation Inventory: Translation, Interpretation, Construction (CRI:TIC)	280
A.2	Konventionen der im CRI:TIC vorkommenden Repräsentationen	303
A.3	Manual zur Auswertung des CRI:TIC	306
B	Anhang zu Forschungsanliegen II	310
B.1	Deskriptive Statistiken und Normalverteilung zu den Messinstrumenten	310
C	Anhang zu Forschungsanliegen III	312
C.1	Testheft – Interpretation, Translation und Konstruktion von Keilstrichformeln und Kugel-Stab-Modellen in der Komplexchemie	312
C.2	Tippkärtchen: Isomerie – Was war das noch einmal?	324
C.3	Manual zur Auswertung des Testhefts	325
C.4	Semantisch-inhaltliches Transkriptionssystem	335
C.5	Kodiermanual – Identifikation zentraler Hindernisse bei der Bearbeitung repräsentationsbasierter Chemieaufgaben	337

A Anhang zu Forschungsanliegen I

A.1 Das *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC)

Liebe Studierende,

vielen Dank, dass Sie mich bei meinem Dissertationsprojekt unterstützen! Ich möchte in dem Projekt herausfinden, welche Faktoren räumlicher Fähigkeiten für das Bearbeiten repräsentationsbasierter Chemieaufgaben wichtig sind.

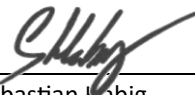
Aber was sind räumliche Fähigkeiten? Die gängigste Definition beschreibt sie als die Fähigkeit, abstrakte visuelle Bilder zu erzeugen, mental zu speichern und zu manipulieren (Lohmann, 1979). Diese Fähigkeit lässt sich in verschiedene Faktoren gliedern, die mithilfe von psychometrischen Tests erhoben werden. Durch die Bearbeitung eines Testinstruments mit repräsentationsbasierten Chemieaufgaben kann ich mir im Anschluss einen Überblick über den Zusammenhang zwischen den Faktoren räumlicher Fähigkeiten und der erbrachten Leistung im Test mit repräsentationsbasierten Chemieaufgaben verschaffen.

Warum interessiere ich mich dafür? Chemie ist ein Fach, das von Repräsentationen lebt: Wir benötigen sie, um neue fachliche Konzepte zu lernen und für Problemlöseprozesse. Gleichzeitig benötigen wir dafür aber auch räumliche Fähigkeiten, z. B. wenn wir Reaktionsmechanismen verstehen möchten oder die absolute Konfiguration eines Moleküls bestimmen. Welche Faktoren räumlicher Fähigkeiten dabei wichtig sind, ist aber bisher nur oberflächlich bekannt. Durch die Ergebnisse dieses Projekts ergibt sich somit möglicherweise ein Ansatzpunkt, um gezielt spezifische Faktoren räumlicher Fähigkeiten zu trainieren und die Leistung von Chemiestudierenden bzw. von Studierenden chemienaher Fächer zu verbessern. Dafür bin ich auf Ihre Unterstützung angewiesen!

Herzlichen Dank für Ihre Teilnahme!



Sebastian Nickel



Prof. Dr. Sebastian Habig

Wichtige Hinweise zu dieser Umfrage



Die Teilnahme an dieser Umfrage ist **freiwillig und vollständig anonym**. Alle erhobenen Daten werden sensibel behandelt, nicht an Dritte weitergegeben und lediglich für das beschriebene Projekt verwendet. Bei Nichtteilnahme entstehen Ihnen **keine Nachteile**.

Hinweise zum Testheft

In diesem Testheft bearbeiten Sie Chemieaufgaben aus verschiedenen Bereichen der Chemie. Die meisten Aufgaben sind Multiple Choice Aufgaben zum Ankreuzen. Teilweise können mehrere Antwortmöglichkeiten richtig sein. Des Weiteren finden Sie einige halboffene Aufgaben, bei denen Sie etwas markieren oder kurze Antworten notieren müssen.

Sollten Sie eine Antwort nicht wissen, raten Sie bitte nicht und lassen das Item unbeantwortet. Sollten Sie eine Antwort korrigieren wollen, streichen Sie die erste Antwort eindeutig durch und setzen Sie ein neues Kreuz bzw. notieren die neue Antwort wie in diesem Beispiel (keine Chemieaufgabe):

Geben Sie den Farbstoff an, der Laubblätter grün färbt.

- a) β -Carotin
- b) Chlorophyll
- ☒ c) Cyanidin
- d) Delphinidin

Geben Sie den Farbstoff an, der Laubblätter grün färbt.

- a) β -Carotin
- ☒ b) Chlorophyll
- ☒ c) Cyanidin
- d) Delphinidin

In den Aufgaben werden Sie mit verschiedenen Arten von chemischen Formeln und Modellen arbeiten müssen, wovon einige ggf. neu für Sie sind. Auf der beigegefügt Seite „Übersicht zu chemischen Formeln und Modellen“ finden Sie die wichtigsten Informationen zu diesen, sodass Sie die Aufgaben lösen können. Diese Übersicht können Sie während der gesamten Zeit nutzen. Bitte geben Sie bei dem letzten Item in dem Testheft an, ob Sie diese Informationen genutzt haben.

Angaben zur Person

Geschlecht

☐

weiblich

☐

männlich

☐

divers

Alter

_____ Jahre

Wo studieren Sie?

☐Duisburg-Essen
(NRW)☐Erlangen-
Nürnberg
(BY)☐Regensburg
(BY)

woanders:

Was planen sie zu studieren?

☐ B. Sc. Chemie☐ B. Sc. Physik☐ B. Sc. Biologie☐ B. Sc. Molecular Science☐ B. Sc. Water Science☐ B. Sc. Aquatische Biologie☐ B. Sc. Molekularbiologie☐ B. Sc. Medizinische Biologie☐ Lebensmittelchemie☐ Pharmazie☐ Medizin☐ Anderes:

Lehramt ☐ Chemie für☐ Gymnasium (BY)☐ Gymnasium, Gesamtschule (NRW)☐ Realschule (BY)☐ Haupt-, Real-, Gesamtschule (NRW)☐ Mittelschule (BY)☐ Grundschule (BY, NRW)☐ Berufskolleg/-schule (BY, NRW)☐ Andere Schulform oder Fächer

In welchem Semester sind Sie?

☐

1

☐

2

☐

3

☐

4

☐

5

☐

6

☐

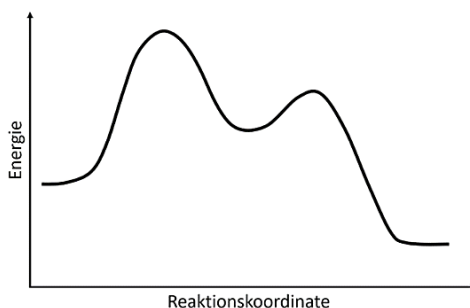
höher

Letzte Chemie Zeugnisnote:

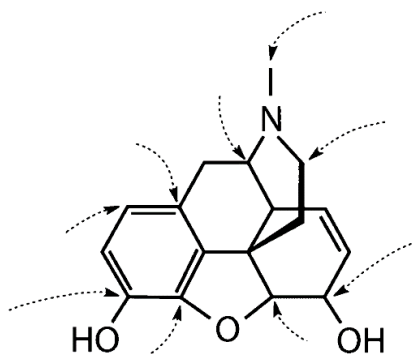
☐ 1+☐ 2+☐ 3+☐ 4+☐ 5+☐ 6☐ 1☐ 2☐ 3☐ 4☐ 5☐ 1-☐ 2-☐ 3-☐ 4-☐ 5-

Interpretation 01**Selbst erstellt**

Während vieler chemischer Reaktionen werden Zwischenprodukte gebildet, die in Energiediagrammen als lokale Energieminima erkenntlich sind. **Markieren** Sie in dem abgebildeten Energiediagramm das Zwischenprodukt/die Zwischenprodukte.

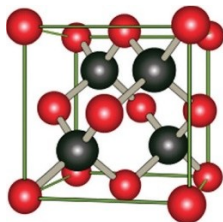
**Interpretation 03****Taskin et al. (2015)**

Geben Sie die Anzahl der Wasserstoffatome an, die an die mit Pfeilen markierte Kohlenstoff-Atome gebunden sind.



Interpretation 04**Adaptiert von Taskin et al. (2015)**

Elementarzellen enthalten alle notwendigen Informationen, die zur Konstruktion eines unendlichen Salzgitters notwendig sind. Die folgende Abbildung stellt eine solche Elementarzelle eines Salzes dar. **Geben** Sie die Koordinationszahl (Anzahl der direkten Nachbarn) der in dem Salzgitter enthaltenen Ionen **an**.

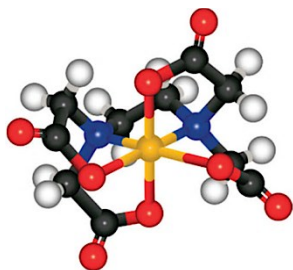
**Koordinationszahl der Ionen**

rot:

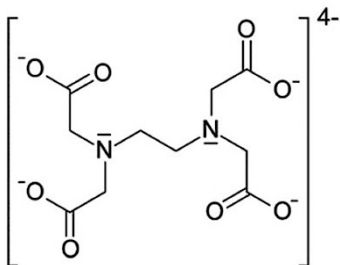
schwarz:

Interpretation 05**Adaptiert von Taskin et al. (2015)**

Bindet ein $[\text{EDTA}]^{4-}$ -Ion koordinativ an ein Kupfer(II)-Ion, so entsteht ein Kupfer(II)-EDTA Komplex, der als Kugel-Stab-Modell in der linken Abbildung dargestellt ist (das Kupfer(II)-Ion ist gelb dargestellt). Auf der rechten Seite ist das $[\text{EDTA}]^{4-}$ -Ion als Skelettformel abgebildet. **Markieren** Sie in der rechten Abbildung die Stellen im $[\text{EDTA}]^{4-}$ -Ion, die im Kupfer(II)-EDTA Komplex an das Kupfer(II)-Ion koordinieren.

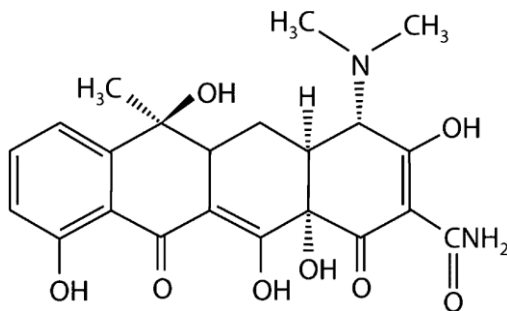


Kupfer(II)-EDTA Komplex

 $[\text{EDTA}]^{4-}$ -Ion

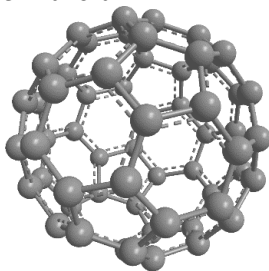
Interpretation 06**Adaptiert von Taskin et al. (2015)**

Markieren Sie in dem abgebildeten Molekül die Kohlenstoffatome, die Stereozentren sind (Kohlenstoffatome mit vier verschiedenen Substituenten).

**Interpretation 07****Selbst erstellt**

In der Abbildung ist die Struktur einer Kohlenstoffmodifikation (C₆₀-Fulleren) abgebildet. **Geben** Sie **an**, wie viele Bindungspartner jedes Kohlenstoffatom besitzt.

*Es gibt genau **eine** richtige Antwort.*

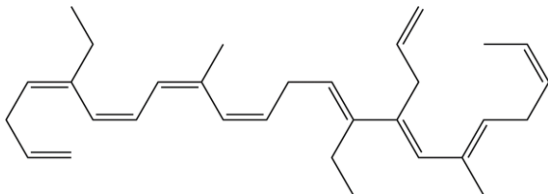


- a) 3
- b) 4
- c) 5
- d) 6

Interpretation 09

Selbst erstellt

Markieren Sie das größte konjugierte π -Elektronensystem (abwechselnd C-C-Einfach- und C=C-Doppelbindungen, erstes und letztes C-Atom muss eine C=C-Doppelbindung aufweisen) in dem dargestellten Molekül.

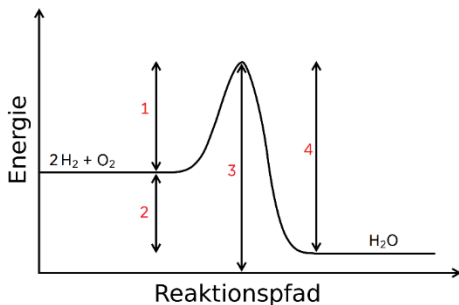


Interpretation 10

Selbst erstellt

Wasserstoff und Sauerstoff reagieren stark exotherm miteinander, allerdings benötigt es dafür einen „zündenden Funken“, die Aktivierungsenergie (E_a). Ohne diese findet keine Reaktion statt. **Kreuzen** Sie **an**, welcher der Pfeile die Aktivierungsenergie in dem Energiediagramm darstellt.

*Es gibt genau **eine** richtige Antwort.*



- a) Pfeil 1
- b) Pfeil 2
- c) Pfeil 3
- d) Pfeil 4

Interpretation 11

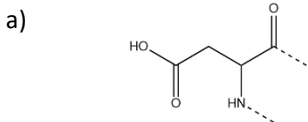
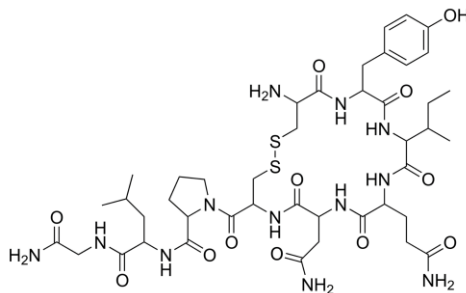
Selbst erstellt

Oxytocin ist ein Hormon aus der Gruppe der Peptidhormone. Es besteht aus insgesamt 8 verschiedenen der 20 proteinogenen Aminosäuren (eine Aminosäure ist doppelt vorhanden, daher in der Summe 9 Aminosäuren). Die Bindung zwischen zwei Aminosäuren bezeichnet man als Peptidbindung, die aus der Bindung der Aminogruppe ($-NH_2$) einer Aminosäure und der Carboxylgruppe ($-COOH$) einer anderen Aminosäure unter Abspaltung von Wasser (H_2O) hervorgeht.

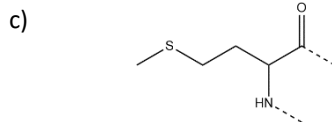
Betrachten Sie die Struktur des Oxytocins und **geben** Sie **an**, welche der vier angegebenen Aminosäuren im Oxytocin eingebaut ist.

Hinweis: Die vier Aminosäuren sind nicht mit der Carboxyl- und Aminogruppe dargestellt, sondern so, wie sie in einer Peptidbindung vorliegen. Beachten Sie, dass die Moleküle a) – d) auch gedreht werden können.

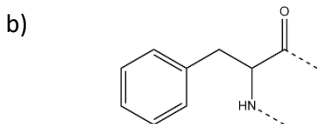
Es gibt genau **eine** richtige Antwort.



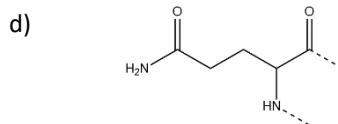
(Asparaginsäure)



(Methionin)



(Phenylalanin)



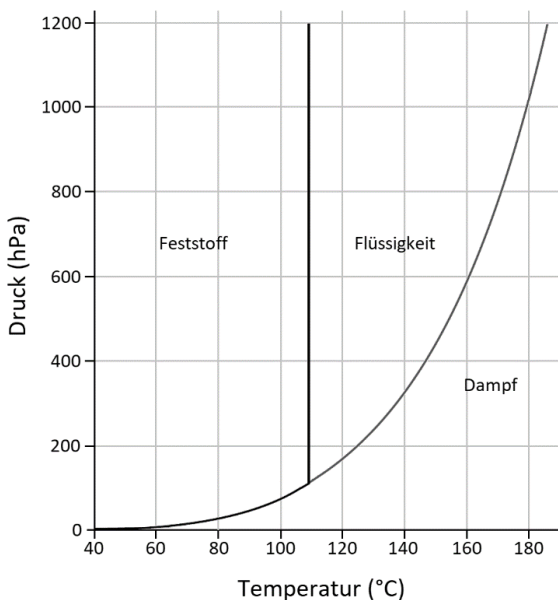
(Glutamin)

Interpretation 12**Selbst erstellt**

In der Abbildung ist das Phasendiagramm von Iod dargestellt. Aus diesem kann abgelesen werden, welchen Aggregatzustand (fest, flüssig, gasförmig) Iod bei welcher Temperatur und welchem Druck hat. **Geben** Sie ...

- a) die Siedetemperatur von Iod bei Normaldruck (1013 hPa) **an**:
_____ °C
- b) den Aggregatzustand bei 500 hPa und 170 °C **an**:

- c) den Druck und die Temperatur **an**, an dem Iod sowohl fest, flüssig, als auch gasförmig vorliegt:
_____ hPa _____ °C

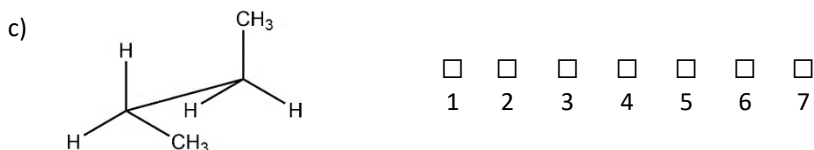
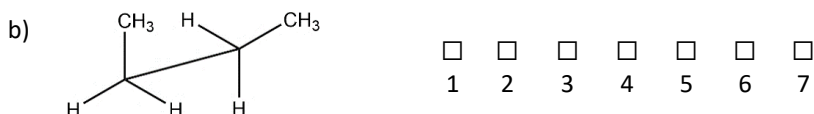
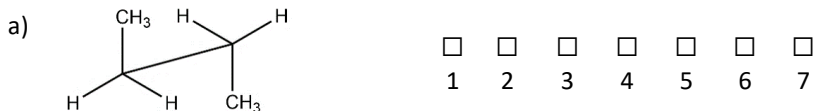
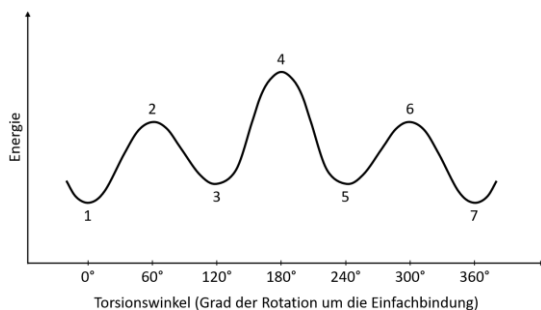


Interpretation 13**Selbst erstellt**

Moleküle können unterschiedliche Konformationen einnehmen, indem sie um eine Einfachbindung rotiert werden. Die unterschiedlichen Konformere, die durch diese Rotation entstehen, haben unterschiedliche Energiezustände. Energetisch ist es besonders günstig, wenn alle Substituenten den größten möglichen Abstand zueinander haben und die sterische Hinderung möglichst gering ist.

Ordnen Sie den angegebenen Molekülen ein Energieniveau (1-7) in dem abgebildeten Diagramm zu.

Es gibt ggf. **mehrere** richtige Antworten.

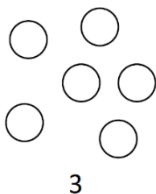
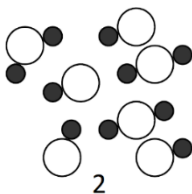
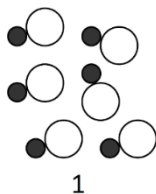


Interpretation 14**Adaptiert von Sim & Daniel (2014)**

Geben Sie an, in welcher Reihenfolge die Abbildung die folgende Reihe angibt:

reines Element > reine Verbindung > Stoffgemisch aus Elementen > Stoffgemisch aus Verbindungen

Es gibt genau **eine** richtige Antwort.



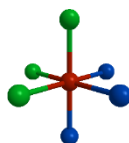
- a) 4, 3, 2, 1
- b) 3, 4, 1, 2
- c) 3, 1, 4, 2
- d) 1, 2, 3, 4

Translation 01

Taskin et al. (2015)

Im Folgenden sind oktaedrische Komplexe vom Typ MA_3X_3 dargestellt, wobei M ein Metall und A und X einzähnige Liganden sind. **Kreuzen** Sie **an**, welche der dargestellten Komplexe mit Komplex A identisch ist.

Es gibt **mehrere** richtige Antworten.



Komplex A

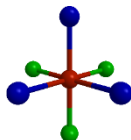
a)



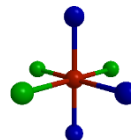
b)



c)



d)

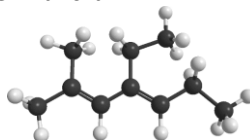


Translation 02

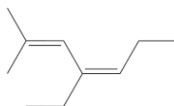
Selbst erstellt

Kreuzen Sie die Skelettformel **an**, die dem abgebildeten Kugel-Stab-Modell entspricht.

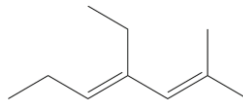
Es gibt genau **eine** richtige Antwort.



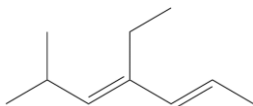
a)



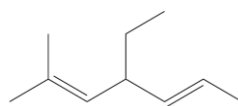
c)



b)



d)

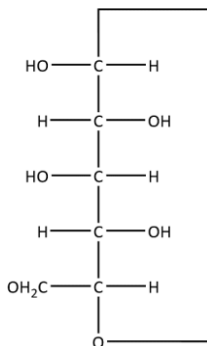


Translation 03

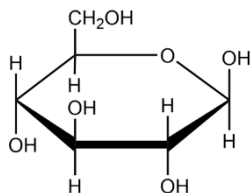
Selbst erstellt

Wählen Sie die Haworth-Projektion **aus**, die das in der Fischer-Projektion dargestellte Kohlenhydrat-Molekül abbildet.

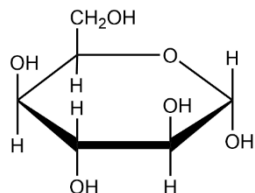
Es gibt genau **eine** richtige Antwort.



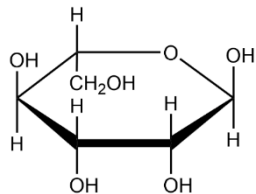
a)



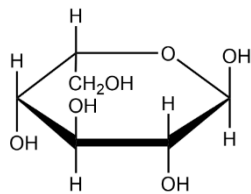
c)



b)



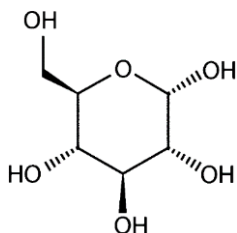
d)



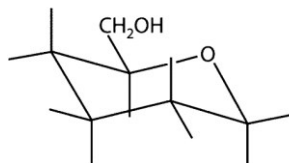
Translation 05

Adaptiert von Taskin et al. (2015)

Die stabilste Form der Glucose ist ein Sechsring in Sesselkonformation. **Ergänzen** Sie in der Sesselkonformation auf der rechten Seite die OH-Gruppen mithilfe der auf der linken Seite abgebildeten Skelettformel der Glucose an der korrekten Position.



Skelettformel der Glucose

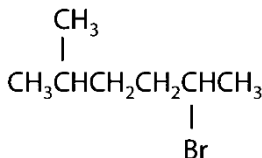


Sesselkonformation der Glucose

Translation 06

Taskin et al. (2015)

Skizzieren Sie die Skelettformeln der zwei dargestellten stark kondensierten Halbstrukturformeln.

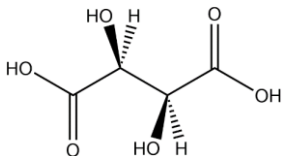


Translation 07

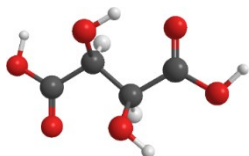
Selbst erstellt

Wählen Sie das Kugel-Stab-Modell, das der abgebildeten Keilstrichformel des Weinsäure-Moleküls entspricht.

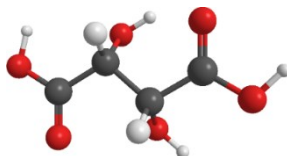
Es gibt genau **eine** richtige Antwort.



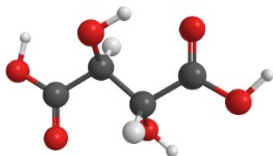
a)



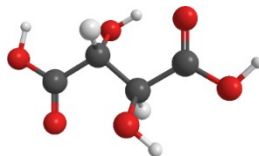
c)



b)



d)

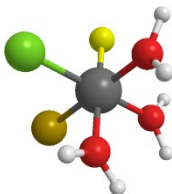


Translation 09

Selbst erstellt

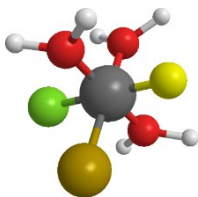
Markieren Sie den Komplex, der zu Komplex A identisch ist, d. h. sich durch Rotation in Komplex A überführen lässt.

*Es gibt genau **eine** richtige Antwort.*

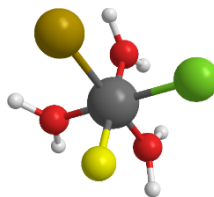


Komplex A

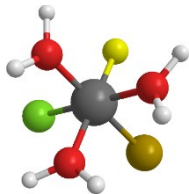
a)



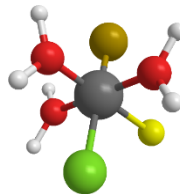
c)



b)



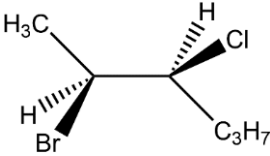
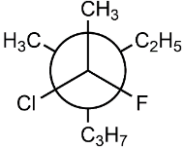
d)



Translation 10

Selbst erstellt

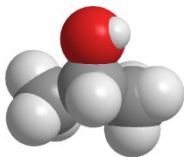
Skizzieren Sie Molekül A in der Newman-Projektion und Molekül B als Keilstrichformel.

	Molekül A	Molekül B
Keilstrichformel		
Newman Projektion		

Translation 11

Selbst erstellt

Skizzieren Sie die im Kalottenmodell dargestellte Verbindung als Keilstrichformel, in der Sie alle Atome darstellen.



Kalottenmodell

Keilstrichformel

(● Kohlenstoff, ● Sauerstoff, ● Wasserstoff)

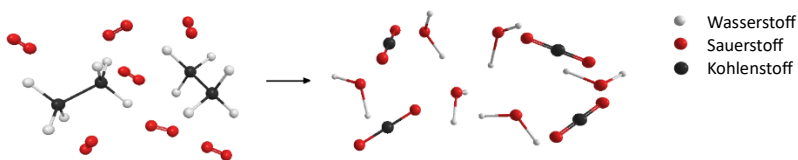
Konstruktion 01**Selbst erstellt**

Stellen Sie eine ausgeglichene Reaktionsgleichung zur Verbrennung von Butan (C_4H_{10}) mit Luftsauerstoff (O_2) zu Kohlenstoffdioxid (CO_2) und Wasser (H_2O) **auf**, sodass auf beiden Seiten des Reaktionspfeils die gleiche Anzahl an Atomen jeder Atomsorte vorliegt. **Nutzen** Sie nur ganzzahlige Koeffizienten/Vorfaktoren.

Konstruktion 02**Selbst erstellt**

Formulieren Sie die Reaktionsgleichung zu der in der Abbildung dargestellten Reaktion.

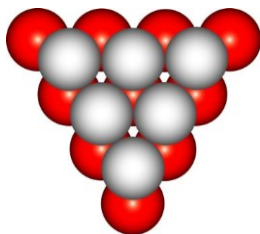
*Es gibt genau **eine** richtige Antwort.*



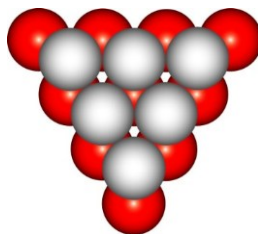
- a) $2 C_2H_6 + 7 O_2 \rightarrow 6 H_2O + 4 CO$
- b) $2 C_2H_6 + 7 O_2 \rightarrow 6 H_2O + 4 CO_2$
- c) $2 C_2H_6 + 7 O_2 \rightarrow 4 H_3O + 4 CO_2$
- d) $2 C_2H_6 + 7 O_2 \rightarrow 4 H_3O + 4 CO_3$

Konstruktion 03**Taskin et al. (2015)**

In den folgenden Abbildungen sehen Sie ausschnittsweise jeweils zwei Schichten einer dichtesten Kugelpackung. **Zeichnen** Sie in beide Abbildungen die Umriss dreier Kugeln, die Sie in der dritten Schicht einer dichtesten Kugelpackung mit der Schichtenfolge ABABA... bzw. ABCABC... vorfinden.



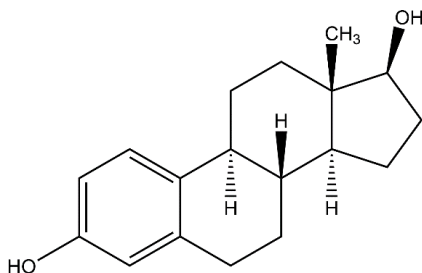
Schichtenfolge
ABABA...



Schichtenfolge
ABCABC...

Konstruktion 05**Selbst erstellt**

Geben Sie die Summenformel der abgebildeten Verbindung (Estradiol) in der Form $C_xH_yO_z$ an.

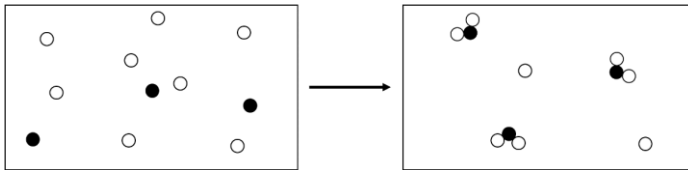


Summenformel: _____

Konstruktion 06**Adaptiert von Taskin et al. (2015)**

Kreuzen Sie **an**, welche der unten aufgeführten Reaktionsgleichungen ein Chemiker wählen würde, um die Reaktion zwischen den Elementen X (●-Kreise) und Y (○-Kreise) zum Ausdruck zu bringen.

*Es gibt genau **eine** richtige Antwort.*



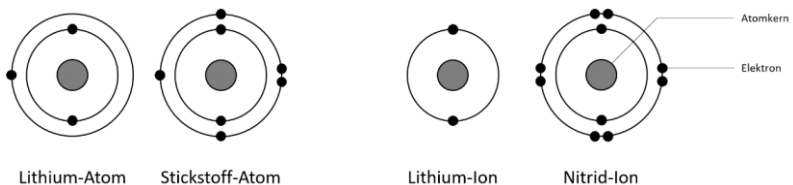
- a) $3 X + 8 Y \rightarrow X_3 Y_8$
- b) $3 X + 6 Y \rightarrow X_3 Y_6$
- c) $X + 2 Y \rightarrow XY_2$
- d) $3 X + 8 Y \rightarrow 3 XY_2 + 2 Y$

Konstruktion 07**Adaptiert von Sim und Daniel (2014)**

Lithium reagiert mit Stickstoff zu dem Salz Lithiumnitrid, das aus Lithium- und Nitrid-Ionen besteht. Die Reaktion ist eine Redoxreaktion, es werden also Elektronen ausgetauscht. In der Abbildung sind die Atome von Lithium und Stickstoff, sowie die entstehenden Ionen im Bohr-Modell dargestellt.

Leiten Sie daraus die Summenformel des Lithiumnitrids **ab**.

*Es gibt genau **eine** richtige Antwort.*



- a) LiN
- b) LiN_3
- c) Li_2N_3
- d) Li_3N

Konstruktion 08**Selbst erstellt**

Skizzieren Sie eine acyclische (Struktur frei von Ringen) und eine cyclische (Struktur mit Ring(en)) Verbindung Ihrer Wahl als Skelettformel mit der Summenformel C_6H_{10} .



acyclisch

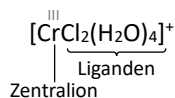


cyclisch

Konstruktion 09**Selbst erstellt**

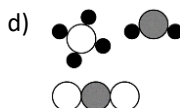
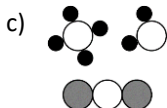
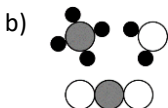
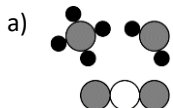
Skizzieren Sie alle theoretisch möglichen Konfigurationsisomere (Verbindungen mit gleicher Struktur, aber unterschiedlicher räumlicher Anordnung der Liganden) des angegebenen oktaedrischen Komplexes als Keilstrichformeln.

Tetraaquadichlorido-chrom(III)-Ion

**Konstruktion 10****Adaptiert von Sim und Daniel (2014)**

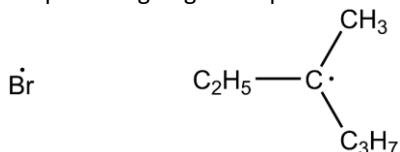
Wählen Sie die Abbildung aus, die ein Methan-Molekül (CH_4), ein Wasser-Molekül (H_2O) und ein Kohlenstoffdioxid-Molekül (CO_2) darstellt.

Es gibt genau **eine** richtige Antwort.



Konstruktion 11**Selbst erstellt**

Bei der radikalischen Bromierung von 3-Methylhexan entstehen durch Kettenstart und Kettenfortpflanzung folgende Spezies an Radikalen:

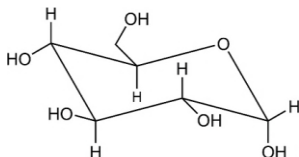


Diese Radikale besitzen ungepaarte Elektronen ($\bullet$) und sind sehr reaktiv. Im Kettenabbruch können diese Radikale miteinander reagieren, indem die ungepaarten Elektronen von zwei Radikalen eine Elektronenpaarbindung ($\bullet\bullet$ oder $-$) ausbilden.

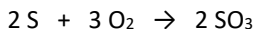
Skizzieren Sie alle Moleküle als Keilstrichformel, die durch die Reaktion zweier Radikale entstehen können. **Beachten** Sie, dass gleiche und ungleiche Radikale miteinander reagieren können und **achten** Sie auch auf die Stereochemie, vor allem den räumlichen Aufbau der Moleküle.

Konstruktion 12**Selbst erstellt**

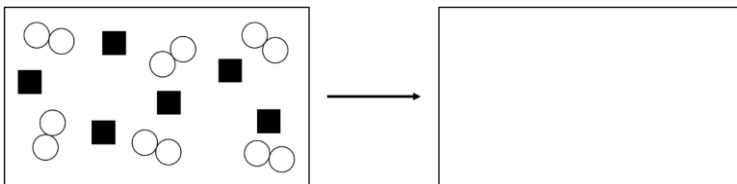
Skizzieren Sie das Enantiomer (Spiegelbild) des angegebenen Kohlenhydrats. **Wählen** Sie eine Spiegelebene und **zeichnen** Sie diese ein!

**Konstruktion 13****Adaptiert von Sim und Daniel (2014)**

In einem geschlossenen System (es können keine Atome entweichen oder hinzukommen) liegen Schwefel (S als ■) und Sauerstoff (O_2 als ○○) vor und reagieren nach folgender Reaktionsgleichung miteinander:



Skizzieren Sie auf Teilchenebene mit den vorgegebenen Symbolen, wie der Inhalt des geschlossenen Systems nach vollständig abgelaufener Reaktion aussieht.



A.2 Konventionen der im CRI:TIC vorkommenden Repräsentationen

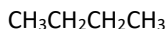
Tabelle 25: Übersicht der im CRI:TIC vorkommenden Repräsentationen mit Informationen zu deren Konventionen, die dem Messinstrument als unterstützendes Material beiliegen.

Summenformel



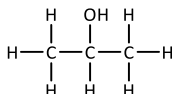
Die Zahl im Index bezieht sich immer auf das vorangehende Elementsymbol. Keine Zahl im Index steht immer für eine „1“. H_2O besteht also aus zwei H-Atomen und einem O-Atom.

Halbstrukturformel



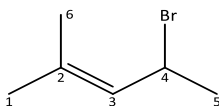
Diese Formel wird vor allem für Kohlenwasserstoffe genutzt und gibt Informationen darüber, welche Atome aneinandergebunden sind. Bei dem Beispiel handelt es sich um eine Kohlenstoffkette mit vier Kohlenstoffatomen, die linear aufgebaut ist. Alle Kohlenstoffatome sind mit einer maximalen Anzahl an Wasserstoffatomen gesättigt. Somit handelt es sich um *n*-Butan.

Valenzstrichformel



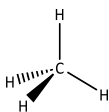
Valenzstrichformeln veranschaulichen chemische Bindungen innerhalb eines Moleküls auf zweidimensionalem Weg. Es kann also abgelesen werden, welche Atome miteinander binden und auch Mehrfachbindungen können erkannt werden. Der räumliche Bau kann aber nicht abgelesen werden (die 3D Moleküle werden für die Valenzstrichformel sozusagen plattgedrückt).

Skelettformel



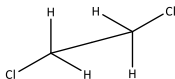
Skelettformeln sind stark reduzierte, vereinfachte und leicht zu zeichnende Formeln für Moleküle. Jede Ecke in der Formel (1-6) steht für ein Kohlenstoffatom, jeder einzelne Strich für eine Einfachbindung, d. h. zwischen 2 und 3 liegt eine Doppelbindung vor. Alle Atome, außer Kohlenstoff- und Wasserstoffatome, werden explizit dargestellt, z. B. das Bromatom bei 4. Ansonsten ist jedes Kohlenstoffatom mit Wasserstoffatomen abgesättigt: Bei 1, 6 und 5 sind drei, bei 3 und 4 ein und bei 2 kein Wasserstoffatom gebunden.

Keilstrichformel



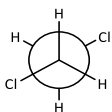
Die Keilstrichformel gibt nicht nur Informationen über die chemischen Bindungen innerhalb des Moleküls, sie gibt auch Informationen über den räumlichen Aufbau: Einfache Striche stellen Bindungen in der Papierebene dar, gestrichelte Keile zeigen von der Papierebene nach hinten und ausgefüllte Keile aus der Papierebene nach vorne.

Sägebock-Projektion



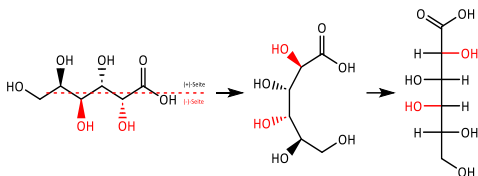
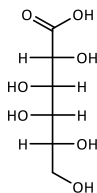
Die Sägebock-Projektion gibt Einblick in den räumlichen Bau zweier benachbarter Kohlenstoffatome. Sie veranschaulicht, wie die Gruppen räumlich zueinanderstehen.

Newman-Projektion

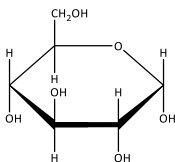


Geht man von der Sägebock-Projektion (s. o.) aus, so betrachtet man das Molekül aus einer anderen Perspektive. Man schaut auf das Molekül entlang der C-C-Einfachbindung. Diese Darstellung wird verwendet, um die räumliche Stellung der gebundenen Substituenten und den Winkel zwischen ihnen zu veranschaulichen.

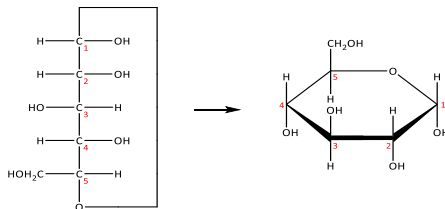
Fischer-Projektion



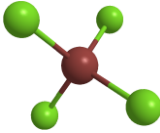
Haworth-Projektion



In der Haworth-Projektion werden ringförmige fünf- und sechsgliedrige Moleküle dargestellt. Beim Übersetzen der Fischer-Projektion zeigen alle Substituenten, die auf der linken Seite stehen nach oben, und alle die auf der rechten Seite stehen nach unten.

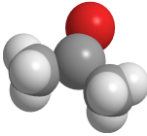


Kugel-Stab-Modell



Das Kugel-Stab-Modell ist ein dreidimensionales Modell, aus dem die räumliche Anordnung von Molekülen ersichtlich wird, ähnlich wie bei Molekülmodellen aus dem Molekülbaukasten. Farben stellen unterschiedliche Atomsorten dar (meist aus einer Legende ablesbar) und die Kugelgrößen der unterschiedlichen Atomsorten veranschaulichen den Atomradius.

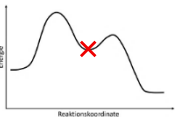
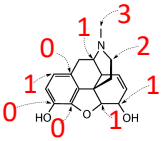
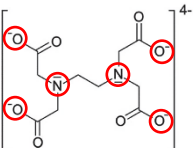
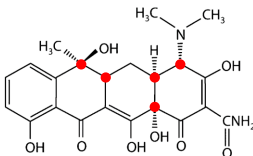
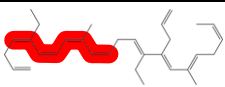
Kalottenmodell

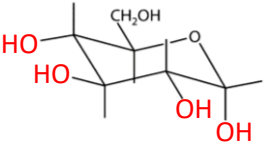
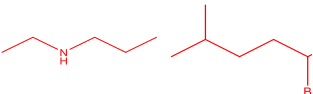
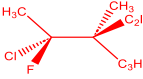


Ähnlich wie das Kugel-Stab-Modell stellt das Kalottenmodell den dreidimensionalen Aufbau eines Moleküls dar, wobei die Relationen der Atomgrößen, Bindungswinkel und Bindungslängen den tatsächlichen Verhältnissen entsprechen.

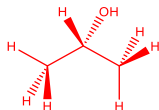
A.3 Manual zur Auswertung des CRI:TIC

Tabelle 26: Manual zur Auswertung und Bepunktung des *Chemical Representation Inventory: Translation, Interpretation, Construction* (CRI:TIC, s. Anhang A.1).

Item	Lösung(en)	max. Pkt.	Hinweise
Interpretation			
Int 01		1	Bei mehr Kreuzen ist die Aufgabe falsch gelöst.
Int 03 ^a		2	9 richtige Antworten = 2 Pkt. 8 richtige Antworten = 1 Pkt. <8 richtige Antworten = 0 Pkt.
Int 04 ^a	Koordinationszahl rot: 4 Koordinationszahl schwarz: 4	2	2 richtige Antworten = 2 Pkt. 1 richtige Antwort = 1 Pkt.
Int 05 ^a		2	Alles richtig markiert = 2 Pkt. Nur O- oder N-Atome richtig = 1 Pkt.
Int 06 ^a		2	Differenz aus korrekten und falschen Antworten bilden. Dann: 5 richtig = 2 Pkt. 4 richtig = 1 Pkt. <4 richtig = 0 Pkt.
Int 07	Antwort a	1	
Int 09		1	Muss exakt der Markierung entsprechen.
Int 10	Antwort a	1	

Int 11	Antwort d	1	
Int 12	a) $175\text{ °C} \leq \mathbf{X} \leq 180\text{ °C}$ b) Dampf/gasförmig c) $95\text{ hPa} \leq \mathbf{X} \leq 115\text{ hPa}$	4	Je korrekter Antwort ein Punkt.
	$105\text{ °C} \leq \mathbf{X} \leq 115\text{ °C}$		
Int 13	a) 1 und 7 b) 3 und 4 c) 2 und 6	6	Jedes richtige Kreuz und richtig nicht gesetzte Kreuz 1 Pkt (für a) - c) je max. 7 Pkt.). Anschließend umcodieren: 7 richtig = 2 Pkt. 6 richtig = 1 Pkt. <6 richtig = 0 Pkt. Anschließend Summe bilden.
Int 14 ^b	Antwort c	1	
Translation			
Trans 01 ^a	Antworten b und c	1	Alles muss richtig sein.
Trans 02	Antwort c	1	
Trans 03	Antwort a	1	
Trans 05 ^a		1	Alles muss richtig sein.
Trans 06 ^a		2	Je richtiger Skelettformel 1 Pkt.
Trans 07	Antwort a	1	
Trans 09	Antwort b	1	
Trans 10	 Molekül A  Molekül B	2	Je richtiger Translation 1 Pkt. Andere Lösungen möglich (z. B. durch andere Ordnung oder Rotation).

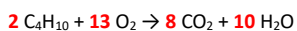
Trans 11



1

Konstruktion

Konst 01



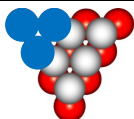
4

Jeder richtige Koeffizient gibt 1 Pkt.
(Vielfaches auch korrekt)

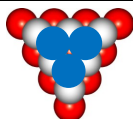
Konst 02

Antwort **b**

1

Konst 03^a

ABABAB...

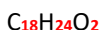


ABCABC...

2

Je richtiger Skizze 1 Pkt.

Konst 05



3

Je korrektem Index 1 Pkt.

Konst 06^aAntwort **c**

1

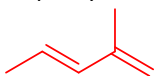
Konst 07^bAntwort **d**

1

Konst 08

Bsp. acyclisch

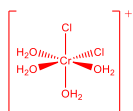
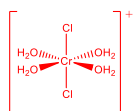
Bsp. cyclisch



2

Je richtiger Skelettformel 1 Pkt.
Abweichende Lösungen möglich und
zu prüfen.

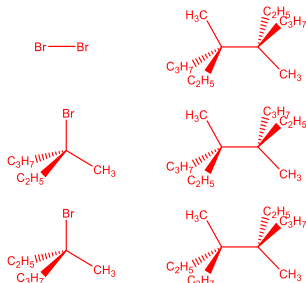
Konst 09



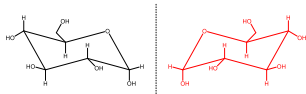
4

1-mal cis/trans richtig = 2 Pkt.
>1-mal cis/trans richtig = 1 Pkt.
Anschließend Summe bilden.Konst 10^bAntwort **b**

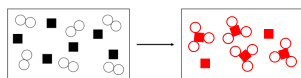
1

Konst 11

- 6 Je richtiger Keilstrichformel 1 Pkt.
Abweichende Lösungen möglich und zu prüfen.

Konst 12

- 1 Abweichende Lösungen möglich und zu prüfen.

Konst 13^b

- 2 Je SO₃ (auch linear, max. 4) = 1 Pkt.
Je S (max. 2) = 1 Pkt.
Falsche/mehr Moleküle = -1 Pkt.
Differenz aus korrekten und falschen Antworten bilden. Dann:
6 richtig = 2 Pkt.
5 richtig = 1 Pkt.
<5 richtig = 0 Pkt.

Hinweis: Die Lösungen der adaptierten Items wurden teilweise aus den entsprechenden Quellen (^aSim & Daniel, 2014; ^bTaskin et al., 2015) übernommen bzw. angepasst.

B Anhang zu Forschungsanliegen II

B.1 Deskriptive Statistiken und Normalverteilung zu den Messinstrumenten

Tabelle 27: Deskriptive Statistiken und Ergebnisse des Shapiro-Wilk Tests auf Normalverteilung für die acht Instrumente zur Messung der räumlichen Faktoren in Forschungsanliegen II.

Räumlicher Faktor Messinstrument	N	Rohdaten				relative Pkt.		Verteilung		
		min.	max.	M	SD	M	SD	Schiefe	Kurtosis	Shapiro-Wilk
Visual Memory Building Memory	221	-3	24	17.8	6.2	0.74	0.26	-1.32	1.32	< 0.001
Speeded Rotation Card Rotations Test	320	-55	158	113.8	30.0	0.72	0.18	-1.30	3.71	< 0.001
Closure Speed Gestalt Completion Test	288	3	18	14.1	2.8	0.78	0.15	-0.80	0.84	< 0.001
Flexibility of Closure Hidden Patterns Test	279	25	302	198.2	46.7	0.50	0.12	-0.80	1.18	< 0.001
Perceptual Speed Identical Pictures Test	294	-24	96	64.4	25.1	0.67	0.26	-0.80	-0.07	< 0.001
Spatial Scanning Maze Tracing Speed Test	280	13	48	29.0	6.7	0.61	0.14	0.24	-0.05	0.154
Visualization Paper Folding Test	210	-16	20	8.7	6.5	0.43	0.33	-0.95	0.92	< 0.001
Spatial Relation Revised PSVT:R	270	0	30	20.8	5.6	0.69	0.19	-0.90	1.10	< 0.001

Tabelle 28: Deskriptive Statistiken und Ergebnisse des Shapiro-Wilk Tests auf Normalverteilung für die Repräsentationskompetenzen und das deklarative Fachwissen in Forschungsanliegen II.

Konstrukt Messinstrument	N	Personenfähigkeiten (in logits)				Verteilung		
		min.	max.	M	SD	Schiefe	Kurtosis	Shapiro-Wilk
RK 31 Items des CRI:TIC	251	-2.68	2.18	0.00	0.96	-0.16	-0.26	0.129
Interpretation 10 <i>Int</i> -Items des CRI:TIC	251	-3.81	2.72	0.00	1.15	-0.10	-0.12	< 0.001
Translation 9 <i>Trans</i> -Items des CRI:TIC	251	-3.92	2.97	0.02	1.09	-0.17	0.41	0.002
Konstruktion 12 <i>Konst</i> -Items des CRI:TIC	251	-3.83	2.57	0.00	1.06	-0.30	0.17	0.017
Fachwissen 16 Fachwissensitems	251	-3.58	2.65	-0.03	1.16	-0.74	1.06	< 0.001

RK = Repräsentationskompetenzen

C Anhang zu Forschungsanliegen III

C.1 Testheft – *Interpretation, Translation* und *Konstruktion* von Keilstrichformeln und Kugel-Stab-Modellen in der Komplexchemie

Lautes Denken - Wie funktioniert das?

DENKEN SIE BITTE LAUT!

Erzählen Sie bitte alles, was Sie denken, wenn Sie die Aufgaben zum ersten Mal sehen. Erzählen Sie so lange, bis alle aufgeworfenen Fragestellungen für Sie vollständig geklärt sind.

LESEN Sie zu diesem Zweck LAUT.

Sie sollten im Idealfall ohne Unterbrechung, also möglichst PAUSENLOS Ihre Gedanken zu den gerade bearbeiteten Textstellen SPRECHEN.

Sie sollten Ihre Äußerungen vor dem Sprechen jedoch nicht besonders ordnen oder Ihre Gedanken besonders verständlich wiedergeben und auch nicht dem Versuchsleiter das Problem oder Ihre Gedanken erklären.

Stellen Sie sich vor, Sie wären GANZ ALLEIN IM RAUM und sprechen zu sich selbst. Wichtig ist, dass Sie möglichst IMMER REDEN.

Die Anleitung wurde von Sandmann (2014, S. 184) adaptiert.

Ablauf der Datenerhebung

Auf der nächsten Seite finden Sie drei Beispielaufgaben, um sich an das laute Denken zu gewöhnen und kurz zu üben.

Im Hauptteil gibt es insgesamt drei Aufgabenblöcke, die sich z. T. in Teilaufgaben untergliedern. Bearbeiten Sie diese Aufgaben nacheinander und haken Sie die einzelnen (Teil-)Aufgaben ab, wenn Sie diese bearbeitet haben, um eine vollständige Bearbeitung zu gewährleisten. Denken Sie daran die Aufgaben zunächst laut vorzulesen und alle Ihre Gedanken zu der Aufgabe laut auszusprechen.

Nach jeder der drei Aufgaben werden Sie gebeten noch einige Fragen zur Bearbeitung der Aufgaben zu beantworten. Auch dabei ist lautes Denken gewünscht.

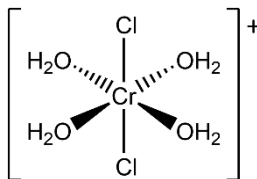
Beispielaufgaben zum lauten Denken

- a) Multiplizieren Sie bitte 12 mal 13 im Kopf und erzählen Sie alles, was Ihnen dabei durch den Kopf geht.
- b) Versuchen Sie zu schätzen, wie viel Fenster und Türen Sie in Ihrer Wohnung haben und sprechen Sie alles aus, was Ihnen dabei durch den Kopf geht.
- c) Erzähle bitte, was Sie bei der Bearbeitung der Aufgaben gedacht haben.

Die Beispielaufgaben wurde von Sandmann (2014, S. 185) adaptiert.

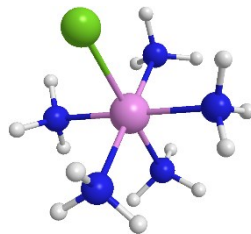
Aufgabe 1**Aufgabe 1.1**

- a) Beschreiben Sie, was in der Abbildung dargestellt ist und erzählen Sie alles, was Ihnen durch den Kopf geht.
- b) Interpretieren Sie alle sichtbaren Merkmale der Strukturformel, indem Sie angeben, was Striche, Keile, Symbole usw. bedeuten und sprechen Sie dabei alle Ihre Gedanken aus.
- c) Geben Sie
- ▶ das Zentralion und seine Koordinationszahl (KZ),
 - ▶ die Anzahl der unterschiedlichen Liganden, sowie
 - ▶ die Art der Liganden (z. B. Bromid) und ihre Zähnigkeit an.
- Erzählen Sie alles, was Ihnen dabei durch den Kopf geht.
- d) Geben Sie die Anzahl der Atome jeder Atomsorte an. Verbalisieren Sie alle Ihre Gedanken.
- e) Benennen Sie den Bindungstypen innerhalb des Komplexes und der mehratomigen Liganden.
- f) Denken Sie laut über den Zweck der Abbildung nach: Warum und wofür wird diese Art der Darstellung verwendet?



Aufgabe 1.2

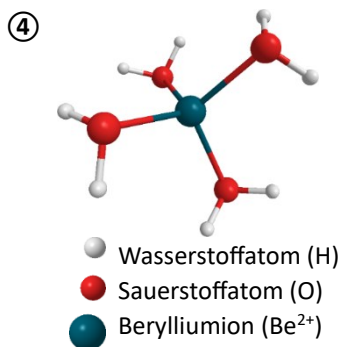
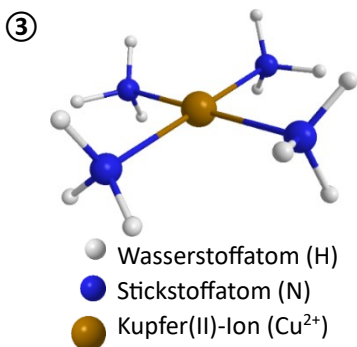
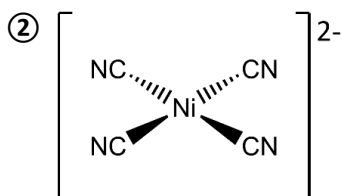
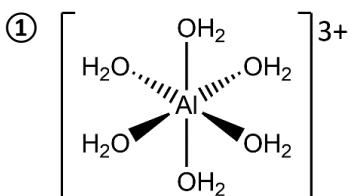
- a) Beschreiben Sie, was in der Abbildung dargestellt ist und erzählen Sie alles, was Ihnen durch den Kopf geht.
- b) Interpretieren Sie alle sichtbaren Merkmale der Struktur, indem Sie angeben, was Striche, Farben, Kugeln, Kugelgrößen usw. bedeuten und sprechen Sie dabei alle Ihre Gedanken aus.
- c) Geben Sie
- ▶ das Zentralion und seine Koordinationszahl (KZ),
 - ▶ die Anzahl der unterschiedlichen Liganden, sowie
 - ▶ die Art der Liganden (z. B. Bromid) und ihre Zähigkeit an.
- Erzählen Sie alles, was Ihnen dabei durch den Kopf geht.
- d) Geben Sie die Anzahl der Atome jeder Atomsorte an. Verbalisieren Sie alle Ihre Gedanken.
- e) Benennen Sie den Bindungstypen innerhalb des Komplexes und der mehratomigen Liganden.
- f) Denken Sie laut über den Zweck der Abbildung nach: Warum und wofür wird diese Art der Darstellung verwendet?



- Wasserstoffatom (H)
- Stickstoffatom (N)
- Chloridion (Cl^-)
- Cobalt(III)-Ion (Co^{3+})

Aufgabe 1.3

- a) Geben Sie
- das Zentralion und seine Koordinationszahl (KZ),
 - die Anzahl der unterschiedlichen Liganden, sowie
 - die Art der Liganden (z. B. Bromid) und ihre Zähnnigkeit an.
- Erzählen Sie alles, was Ihnen dabei durch den Kopf geht.
- b) Geben Sie die Anzahl der Atome jeder Atomsorte an. Verbalisieren Sie alle Ihre Gedanken.



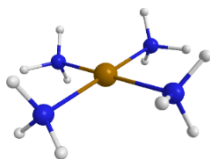
Reflexion der Aufgaben 1.1 bis 1.3

- Möchten Sie nachträglich noch etwas zu einer der Aufgaben ergänzen, dass Ihnen vielleicht noch eingefallen ist? Erzählen Sie alle Ihre Gedanken und Ideen.
- Denken Sie laut darüber nach, was Ihnen bei der Bearbeitung der Aufgaben Probleme bereitet hat, was Ihnen besonders schwergefallen ist und was Ihnen leichtgefallen ist.
- Was würde Ihnen dabei helfen die Aufgaben in Zukunft besser zu lösen? Erzählen Sie alles, was Ihnen dazu durch den Kopf geht.

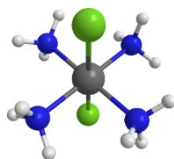
Aufgabe 2

Aufgabe 2.1

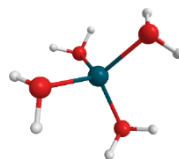
Skizzieren Sie die dargestellten Komplexe als Keilstrichformel und geben Sie die räumliche Anordnung der Liganden um das Zentralion an (linear, trigonal-bipyramidal, o. ä.). Erzählen Sie dabei alle Ihre Gedanken.



- Wasserstoffatom (H)
- Stickstoffatom (N)
- Kupfer(II)-Ion (Cu^{2+})



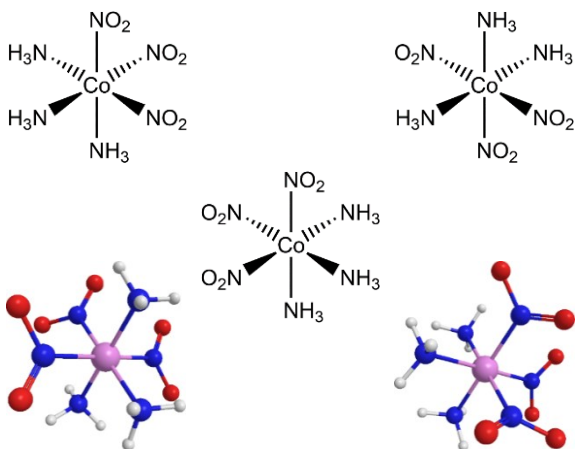
- Wasserstoffatom (H)
- Stickstoffatom (N)
- Chloridion (Cl^-)
- Chrom(III)-Ion (Cr^{3+})



- Wasserstoffatom (H)
- Sauerstoffatom (O)
- Berylliumion (Be^{2+})

Aufgabe 2.2

Vergleichen Sie den in der Mitte abgebildeten Komplex mit den vier umgebenden Komplexen und geben Sie an, ob es die gleichen Strukturen oder Isomere sind. Sollten die Strukturen Isomere sein, geben Sie die Art der Isomerie an (Konstitutionsisomer, Konformationsisomerie, Enantiomerie oder Diastereomerie). Nutzen Sie das Tippkärtchen „Isomerie – Was war das noch einmal?“ und erzählen Sie alles, was Ihnen durch den Kopf geht.

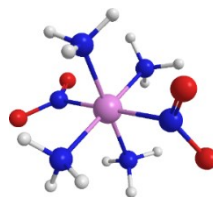
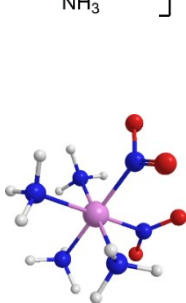
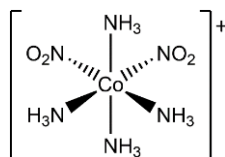
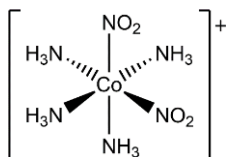


 Wasserstoff-Atom (H)
 Stickstoff-Atom (N)

 Sauerstoff-Atom (O)
 Cobalt(III)-Ion (Co^{3+})

Aufgabe 2.3

Vergleichen Sie den in der Mitte abgebildeten Komplex mit den vier umgebenden Komplexen und geben Sie an, ob es die gleichen Strukturen oder Isomere sind. Sollten die Strukturen Isomere sein, geben Sie die Art der Isomerie an (Konstitutionsisomer, Konformationsisomerie, Enantiomerie oder Diastereomerie). Nutzen Sie das Tippkärtchen „Isomerie – Was war das noch einmal?“ und erzählen Sie alles, was Ihnen durch den Kopf geht.

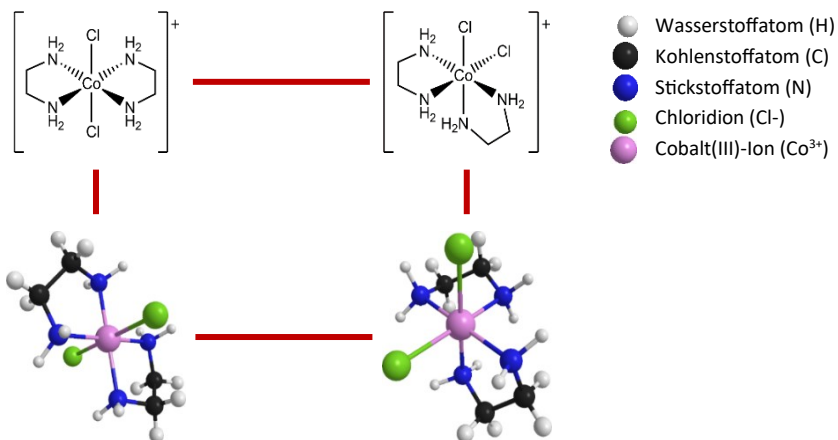


 Wasserstoff-Atom (H)
 Stickstoff-Atom (N)

 Sauerstoff-Atom (O)
 Cobalt(III)-Ion (Co^{3+})

Aufgabe 2.4

Vergleichen Sie die durch rote Striche verbundenen Komplexe miteinander und geben Sie an, ob es die gleichen Strukturen oder Isomere sind. Sollten die Strukturen Isomere sein, geben Sie die Art der Isomerie an (Konstitutionsisomer, Konformationsisomerie, Enantiomerie oder Diastereomie). Nutzen Sie das Tippkärtchen „Isomerie – Was war das noch einmal?“ und erzählen Sie alles, was Ihnen durch den Kopf geht.



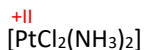
Reflexion der Aufgaben 2.1 bis 2.4

- Möchten Sie nachträglich noch etwas zu einer der Aufgaben ergänzen, dass Ihnen vielleicht noch eingefallen ist? Erzählen Sie alle Ihre Gedanken und Ideen.
- Denken Sie laut darüber nach, was Ihnen bei der Bearbeitung der Aufgabe Probleme bereitet hat, was Ihnen besonders schwergefallen ist und was Ihnen leichtgefallen ist.
- Was würde Ihnen dabei helfen die Aufgabe in Zukunft besser zu lösen? Erzählen Sie alles, was Ihnen dazu durch den Kopf geht.

Aufgabe 3

- a) Geben Sie zunächst das Zentralatom/-ion, die Liganden und Anzahl der (verschiedenen) Liganden des Komplexes ① an und erzählen Sie alles, was Ihnen dabei durch den Kopf geht.
- b) Skizzieren Sie Komplex ① als Keilstrichformel und alle theoretisch möglichen Konfigurationsisomere. Nutzen Sie das Tippkärtchen „Isomerie – Was war das noch einmal?“ und erzählen Sie alles, was Ihnen durch den Kopf geht.
- c) Gehen Sie für die Komplexe ② bis ④ genauso vor.

①



quadratisch planarer Komplex

②



tetraedrischer Komplex

③



④

**Reflexion der Aufgabe 3**

- a) Möchten Sie nachträglich noch etwas zu einer der Aufgaben ergänzen, dass Ihnen vielleicht noch eingefallen ist? Erzählen Sie alle Ihre Gedanken und Ideen.
- b) Denken Sie laut darüber nach, was Ihnen bei der Bearbeitung der Aufgabe Probleme bereitet hat, was Ihnen besonders schwergefallen ist und was Ihnen leichtgefallen ist.
- c) Was würde Ihnen dabei helfen die Aufgabe in Zukunft besser zu lösen? Erzählen Sie alles, was Ihnen dazu durch den Kopf geht.

Angaben zur Person

Geschlecht

☐

weiblich

☐

männlich

☐

divers

Alter

_____ Jahre

Was studieren Sie?

☐ B. Sc. Chemie

Lehramt Chemie für

☐ Gymnasien☐ B. Sc. Molecular Science☐ Realschulen☐ Lebensmittelchemie☐ Mittelschulen☐ Pharmazie☐ Grundschulen☐ Sonstiges: _____

In welchem Semester sind Sie?

☐☐☐☐☐☐☐☐☐☐☐☐

1

2

3

4

5

6

7

8

9

10

höher

Letzte Chemie Zeugnisnote:

☐ 1+☐ 2+☐ 3+☐ 4+☐ 5+☐ 6☐ 1☐ 2☐ 3☐ 4☐ 5☐ 1-☐ 2-☐ 3-☐ 4-☐ 5-

Anmerkung: Das Testheft wurde in der Studie in einem anderen Format eingesetzt, sodass die dargestellten Repräsentationen größer und besser zu erkennen sind. Die Aufgaben wurden hier in das Design der Arbeit eingebettet und Abbildungen in der Größe angepasst. Die Aufgaben an sich wurden nicht verändert.

C.2 Tippkärtchen: Isomerie – Was war das noch einmal?

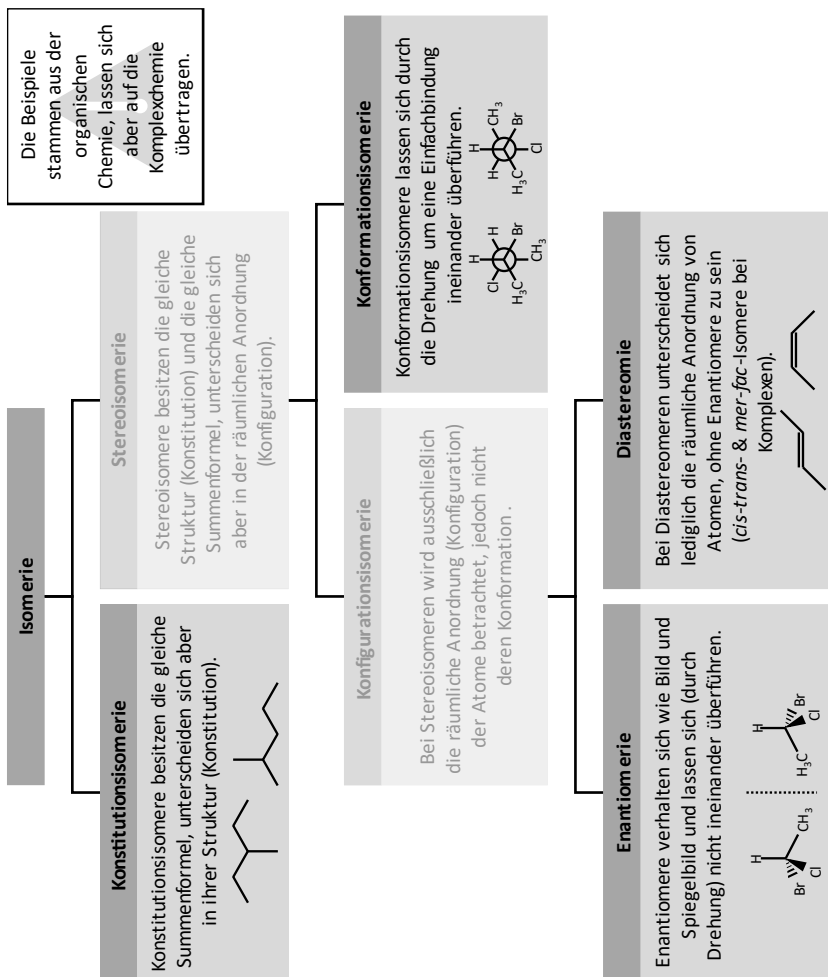


Abbildung 32: Tippkärtchen „Isomerie – Was war das noch einmal?“ als unterstützendes Material für die Bearbeitung der Aufgaben 2 (*Translation*) und 3 (*Konstruktion*) in Forschungsanliegen III (Binnewies et al., 2016; Mortimer & Müller, 2020; Riedel & Janiak, 2022; Schmuck, 2013; Weber, 2021).

C.3 Manual zur Auswertung des Testthefts

Aufgabe 1

Aufgabe 1.1

a) Von der Bewertung ausgeschlossen.

b) Interpretieren Sie alle sichtbaren Merkmale der Strukturformel, indem Sie angeben, was Striche, Keile, Symbole usw. bedeuten und sprechen Sie dabei alle Ihre Gedanken aus.

- Striche: Bindungen in der (Papier)Ebene 1
- Ausgefüllte (fette) Keile: Bindungen aus der (Papier)Ebene nach vorne 1
- Gestrichelte (dünne) Keile: Bindungen aus der (Papier)Ebene nach hinten 1
- Buchstaben: Elementsymbole aus dem PSE stellen Atome und Ionen dar (ggf. Angabe der Atom- oder Ionensorte) 1
- Klammer und Plus: Angabe der Gesamtladung des Komplexes, einfach positive Ladung 1

c) Geben Sie das Zentralion und seine Koordinationszahl (KZ), die Anzahl der unterschiedlichen Liganden, sowie die Art der Liganden (z. B. Bromid) und ihre Zähigkeit an.

- Zentralion: Chrom(III), Cr^{3+} , Chrom drei plus; 1
- Koordinationszahl: 6 1
- 2 unterschiedliche Liganden oder 2 x Chlorid/ Cl^- /Chlorido & 4 x Wasser/ H_2O /Aqua 1
- 2 x Chlorid/ Cl^- /Chlorido & 4 x Wasser/ H_2O /Aqua 1+1
- Beide Liganden sind einzähnig/monodentat 1

d) Geben Sie die Anzahl der Atome jeder Atomsorte an.

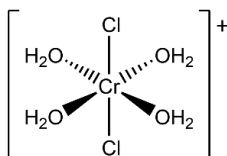
- 2 x Chlor/Chlorid/ Cl^- /Chlorido 1
- 1 x Chrom/Chrom(III), Cr^{3+} , Chrom drei plus 1
- 4 x Sauerstoff/O 1
- 8 x Wasserstoff/H 1

e) Benennen Sie den Bindungstypen innerhalb des Komplexes und der mehratomigen Liganden.

- Komplex: kovalente Bindung (auch koordinative Bindung / dative Bindung / Komplexbindung / Donator-Akzeptor-Bindung) 1
- Wasser: kovalente Bindung/Atombindung/Elektronenpaarbindung 1

f) Warum und wofür wird diese Art der Darstellung verwendet?

- Darstellung 3-dimensionaler Strukturen auf einfache, systematische und 2-dimensionale Art und Weise 1
- Visuelle Klarheit der Struktur durch konkret definierte Bedeutung von Strichen und Keilen 1
- Veranschaulichung der räumlichen Anordnung von Substituenten 1
- Essenziell für die Stereochemie und die Vorhersage der stereoselektivität (OC) 1

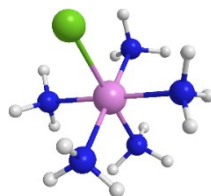
**Aufgabe 1.2**

a) Von der Bewertung ausgeschlossen.

b) Interpretieren Sie alle sichtbaren Merkmale der Struktur, indem Sie angeben, was Striche, Farben, Kugeln, Kugelgrößen usw. bedeuten und sprechen Sie dabei alle Ihre Gedanken aus.

- Striche: Bindungen 1
- Farben: Unterschiedliche Atom-/Ionensorten (ggf. konkrete Angabe aus der Legende) 1
- Kugeln: Analogie zu Atomen als kleine Kugeln 1
- Kugelgrößen: Stellen die Atomradien der einzelnen Atomsorten im Verhältnis dar; unterschiedliche Kugelgrößen gleicher Atomsorten deuten die räumliche Orientierung an (größere Kugeln sind weiter vorne im Raum als kleiner Kugeln) 1

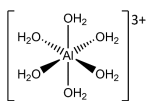
c) <i>Geben Sie das Zentralion und seine Koordinationszahl (KZ), die Anzahl der unterschiedlichen Liganden, sowie die Art der Liganden (z. B. Bromid) und ihre Zähigkeit an.</i>	
• Zentralion: Cobalt(III), Co^{3+} , Cobalt drei plus	1
Koordinationszahl: 6	1
• 2 unterschiedliche Liganden oder 1 x Chlorid/ Cl^- /Chlorido & 5 x Ammoniak/ NH_3 / Ammin	1
• 1 x Chlorid/ Cl^- /Chlorido & 5 x Ammoniak/ NH_3 / Ammin	1+1
• Beide Liganden sind einzähnig/monodentat	1
d) <i>Geben Sie die Anzahl der Atome jeder Atomsorte an.</i>	
• 1 x Chlor/Chlorid/ Cl^- /Chlorido	1
• 1 x Cobalt/Cobalt(III)/ Co^{3+} /Cobalt drei plus	1
• 5 x Stickstoff/N	1
• 15 x Wasserstoff/H	1
e) <i>Benennen Sie den Bindungstypen innerhalb des Komplexes und der mehratomigen Liganden.</i>	
• Komplex: koordinative Bindung / dative Bindung / Komplexbindung / Donator-Akzeptor-Bindung	1
• Ammoniak: kovalente Bindung / Atombindung / Elektronenpaarbindung	1
• Mögliche Kritik: Die koordinative Bindung sind falsch dargestellt, da die Bindungselektronen allein vom Liganden kommen (Donator-Akzeptor-Bindung)	
f) <i>Warum und wofür wird diese Art der Darstellung verwendet?</i>	
• Darstellung 3-dimensionaler Strukturen zur Veranschaulichung der Geometrie, Bindungswinkel und relative Anordnung der Atome in einer Verbindung	1
• Visuelle Klarheit über räumliche Beziehungen zwischen Atomen und Bindungen	1
• Zeigt exakte Bindungswinkel und -längen	



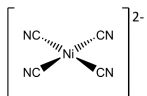
- Wasserstoffatom (H)
- Stickstoffatom (N)
- Chloridion (Cl^-)
- Cobalt(III)-Ion (Co^{3+})

Aufgabe 1.3

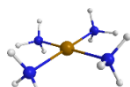
a) Geben Sie das Zentralion und seine Koordinationszahl (KZ), die Anzahl der unterschiedlichen Liganden, sowie die Art der Liganden (z. B. Bromid) und ihre Zähnnigkeit an.



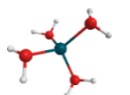
- Zentralion: Aluminium(III)/Al³⁺/Aluminium drei plus 1
- Koordinationszahl: 6 1
- Eine Ligandenart: Wasser/H₂O/Aqua 1
- Einzähniger/monodentater Ligand 1



- Zentralion: Nickel(II)/Ni²⁺/Nickel zwei plus 1
- Koordinationszahl: 4 1
- Eine Ligandenart: Cyanid/CN⁻/Cyanido/Cyano 1
- Einzähniger/monodentater Ligand 1

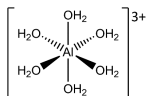


- Zentralion: Kupfer(II)/Cu²⁺/Kupfer zwei plus 1
- Koordinationszahl: 4 1
- Eine Ligandenart: Ammoniak/NH₃/Ammin 1
- Einzähniger/monodentater Ligand 1

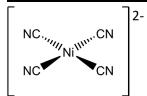


- Zentralion: Beryllium(II)/Be²⁺/Beryllium zwei plus 1
- Koordinationszahl: 4 1
- Eine Ligandenart: Wasser/H₂O/Aqua 1
- Einzähniger/monodentater Ligand 1

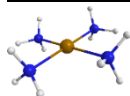
b) Geben Sie die Anzahl der Atome jeder Atomsorte an.



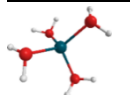
- 1 x Aluminium/Aluminium(III)/Al³⁺/Aluminium 3 plus 1
- 6 x Sauerstoff/O 1
- 12 x Wasserstoff/H 1



- 1 x Nickel/Nickel(II)/Ni²⁺/Nickel zwei plus 1
- 4 x Kohlenstoff/C 1
- 4 x Stickstoff/N 1



- 1 x Kupfer/Kupfer(II)/Cu²⁺/Kupfer zwei plus 1
- 4 x Stickstoff/N 1
- 12 x Wasserstoff/H 1



- 1 x Beryllium/Beryllium(II)/Be²⁺/Beryllium zwei plus 1
- 4 x Sauerstoff/O 1
- 8 x Wasserstoff/H 1

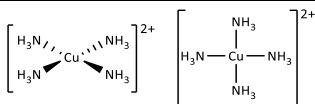
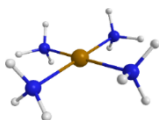
Reflexion der Aufgaben 1.1 bis 1.3

- a) Von der Bewertung ausgeschlossen.
 b) Von der Bewertung ausgeschlossen.
 c) Von der Bewertung ausgeschlossen.

Aufgabe 2

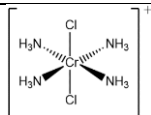
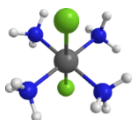
Aufgabe 2.1

Skizzieren Sie die dargestellten Komplexe als Keilstrichformel und geben Sie die räumliche Anordnung der Liganden um das Zentralion an (linear, trigonal-bipyramidal, o. ä.). Erzählen Sie dabei alle Ihre Gedanken.



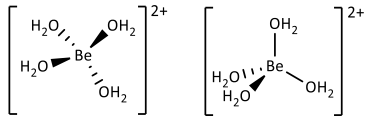
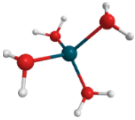
Weiter Lösungen möglich und jeweils zu bewerten (z. B. durch Rotation).

- „Cu“ steht in der Mitte 1
- NH₃ koordiniert immer über das N 1
- Alle NH₃ liegen in einer quadratischen Ebene 1
- Klammern um den Komplex 1
- Ladung ist angegeben 1



Weiter Lösungen möglich und jeweils zu bewerten (z. B. durch Rotation).

- „Cr“ steht in der Mitte 1
- NH₃ koordiniert immer über das N 1
- Es handelt sich um einen trans-Komplex 1
- Klammern um den Komplex 1
- Ladung ist angegeben 1



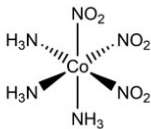
Weiter Lösungen möglich und jeweils zu bewerten (z. B. durch Rotation). 1

- „Be“ steht in der Mitte 1
- H₂O koordiniert immer über das O 1
- Die H₂O sind tetraedrisch angeordnet 1
- Klammern um den Komplex 1
- Ladung ist angegeben 1

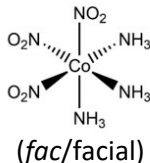
Aufgabe 2.2

Vergleichen Sie den in der Mitte abgebildeten Komplex mit den vier umgebenden Komplexen und geben Sie an, ob es die gleichen Strukturen oder Isomere sind. Sollten die Strukturen Isomere sein, geben Sie die Art der Isomerie an (Konstitutionsisomer, Konformationsisomerie, Enantiomerie oder Diastereomerie). Nutzen Sie das Tippkärtchen „Isomerie – Was war das noch einmal?“ und erzählen Sie alles, was Ihnen durch den Kopf geht.

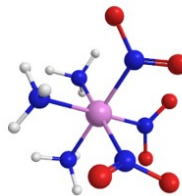
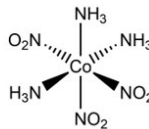
Identisch
(*fac*/*facial*)



Diastereomer
(*mer*/*meridional*)



Diastereomer
(*mer*/*meridional*)



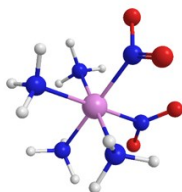
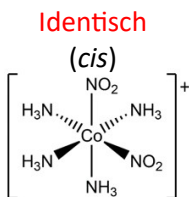
Identisch
(*fac*/*facial*)

1+1

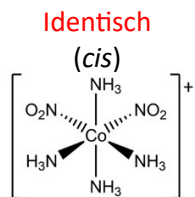
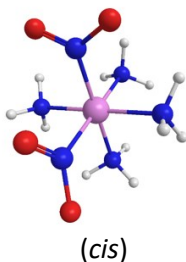
1+1

Aufgabe 2.3

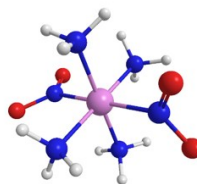
Vergleichen Sie den in der Mitte abgebildeten Komplex mit den vier umgebenden Komplexen und geben Sie an, ob es die gleichen Strukturen oder Isomere sind. Sollten die Strukturen Isomere sein, geben Sie die Art der Isomerie an (Konstitutionsisomer, Konformationsisomerie, Enantiomerie oder Diastereomerie). Nutzen Sie das Tippkärtchen „Isomerie – Was war das noch einmal?“ und erzählen Sie alles, was Ihnen durch den Kopf geht.



Identisch
(cis)



1+1

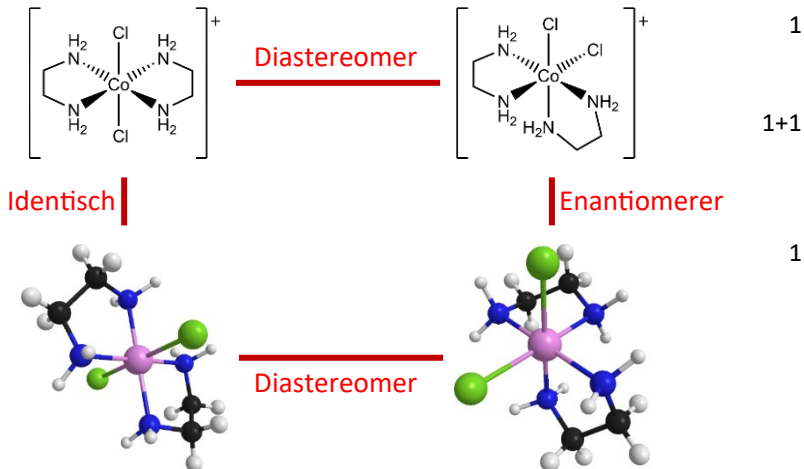


1+1

Diastereomer
(trans)

Aufgabe 2.4

Vergleichen Sie die durch rote Striche verbundenen Komplexe miteinander und geben Sie an, ob es die gleichen Strukturen oder Isomere sind. Sollten die Strukturen Isomere sein, geben Sie die Art der Isomerie an (Konstitutionsisomer, Konformationsisomerie, Enantiomerie oder Diastereomerie). Nutzen Sie das Tippkärtchen „Isomerie – Was war das noch einmal?“ und erzählen Sie alles, was Ihnen durch den Kopf geht.

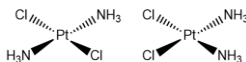
**Reflexion der Aufgaben 2.1 bis 2.4**

- a) Von der Bewertung ausgeschlossen.
- b) Von der Bewertung ausgeschlossen.
- c) Von der Bewertung ausgeschlossen.

Aufgabe 3

- ① a) • Zentralion: Platin(II)/Pt²⁺/Platin zwei plus 1
 • Liganden: 4 Stück, 2 x Chlorid/Cl⁻/Chlorido, 2 x Ammoniak/NH₃/Ammin 1

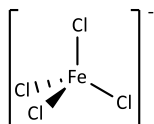
b)



- „Pt“ steht immer in der Mitte
 - Skizziert das *trans*-isomer 1
 - Skizziert das *cis*-Isomer 1
 - NH₃ koordiniert über das N 1
 - Skizziert genau ein *cis*- und ein *trans* Isomer 1
- Weiter Lösungen möglich und jeweils zu bewerten (z. B. durch Rotation). 1

- ② a) • Zentralion: Eisen(III)/Fe³⁺/Eisen drei plus 1
 • Liganden: 4 Stück, 4 x Chlorid/Cl⁻/Chlorido 1

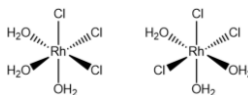
b)



- „Fe“ steht immer in der Mitte 1
 - „Cl“ sind tetraedisch angeordnet 1
 - Skizziert die Klammern 1
 - Gibt die Ladung an 1
 - Skizziert genau einen Komplex 1
- Weiter Lösungen möglich und jeweils zu bewerten (z. B. durch Rotation).

- ③ a) • Zentralion: Rhodium(III)/Rh³⁺/Rhodium drei plus 1
 • Liganden: 6 Stück, 3 x Chlorid/Cl⁻/Chlorido, 3 x Wasser/H₂O/Aqua 1

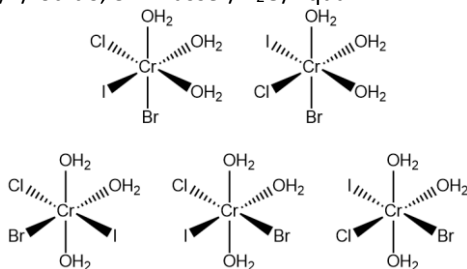
b)



- „Rh“ steht immer in der Mitte
 - Skizziert das *fac*-isomer 1
 - Skizziert das *mer*-isomer 1
 - NH₃ koordiniert immer über das N 1
 - Skizziert genau einen *mer*- und ein *fac*-Isomer 1
- Weiter Lösungen möglich und jeweils zu bewerten (z. B. durch Rotation). 1

- ④ a) • Zentralion: Chrom(III)/Cr³⁺/Chrom drei plus 1
- Liganden: 6 Stück, 1 x Chlorid/Cl⁻/Chlorido, 1 x Bromid/Br⁻/Bromido, 1
- 1 x Iodid/I⁻/Iodido, 3 x Wasser/H₂O/Aqua

b)



- „Cr“ steht immer in der Mitte 1
- Skizziert 2 *fac*-Isomere 1
- Skizziert 3 *mer*-Isomere 1
- H₂O koordiniert immer über das O 1
- Skizziert genau 3 *mer*- und 2 *fac*-Isomere 1

Weiter Lösungen möglich und jeweils zu bewerten (z. B. durch Rotation).

Reflexion der Aufgabe 3

a) Von der Bewertung ausgeschlossen.

b) Von der Bewertung ausgeschlossen.

c) Von der Bewertung ausgeschlossen.

C.4 Semantisch-inhaltliches Transkriptionssystem

Die Anleitung zur semantisch-inhaltlichen Transkription wurde von Dresing und Pehl (2018) adaptiert und zum Teil im Wortlaut übernommen.

Inhaltlich-semantische Transkription

1. Es wird wörtlich transkribiert, also nicht lautsprachlich oder zusammenfassend.
2. Wortverschleifungen werden an das Schriftdeutsch angenähert. „So ‘n Buch“ wird zu „so ein Buch“. Die Satzform wird beibehalten, auch wenn sie syntaktische Fehler beinhaltet, z.B.: „Bin ich nach Kaufhaus gegangen.“
3. Dialekte werden möglichst wortgenau ins Hochdeutsche übersetzt.
4. Umgangssprachliche Partikel wie „gell, gelle, ne“ werden transkribiert.
5. Stottern wird geglättet bzw. ausgelassen, abgebrochene Wörter werden ignoriert. Wortdoppelungen werden nur erfasst, wenn sie als Stilmittel zur Betonung genutzt werden: „Das ist mir sehr, sehr wichtig.“
6. Wort- und Satzabbrüche werden mit dem Abbruchzeichen „/“ gekennzeichnet.
7. Interpunktion wird zugunsten der Lesbarkeit geglättet, das heißt, bei kurzem Senken der Stimme oder nicht eindeutiger Betonung wird eher ein Punkt als ein Komma gesetzt.
8. Rezeptionssignale wie „aha, ja, genau“ werden dann transkribiert, wenn sie als direkte Antwort auf eine Frage genannt werden, sonst nicht. Nach der Partikel „hm“ wird eine Beschreibung der Betonung in Klammern festgehalten. Zu nutzen sind: bejahend, verneinend, nachdenkend, fragend, wohlfühlend, z.B. „hm (bejahend)“.
9. Pausen werden je nach Länge durch Auslassungspunkte in Klammern markiert. Hierbei steht „(.)“ für circa eine Sekunde, „(..)“ für circa zwei Sekunden, „(...)“ für circa drei Sekunden und „(Zahl)“ für mehr als drei Sekunden.
10. Besonders betonte Wörter oder Äußerungen werden durch VERSALIEN gekennzeichnet.

11. Emotionale nonverbale Äußerungen der befragten Person und des Interviewers, welche die Aussage unterstützen oder verdeutlichen (wie lachen oder seufzen), werden beim Einsatz in Klammern notiert.
12. Unverständliche Wörter werden mit „(unv.)“ gekennzeichnet. Längere unverständliche Passagen werden möglichst mit der Ursache versehen: „(unv., Mikrofon rauscht)“. Vermutet man einen Wortlaut, wird die Passage mit einem Fragezeichen in Klammern gesetzt, z.B. „(Axt?)“.
13. Bei einer Interaktion wird die interviewende Person durch ein „I:“, die befragte Person durch ein „B:“ gekennzeichnet.

Hinweise zur einheitlichen Schreibweise

1. Die Partikeln „hm“ werden unabhängig von der Betonung immer „hm“ geschrieben (nicht: „hnhhm“, „mhm“, „hnh“).
2. Zögerungslaute werden immer „ähm“ geschrieben (nicht: „äm“, „ehm“, „öhm“).
3. (Maß-)Einheiten werden ausgeschrieben, z. B. Euro, Prozent, Meter.
4. Gesprochene Zeichen werden ausgeschrieben, z. B. „ät“, „Paragraf“.
5. Abkürzungen werden nur getippt, wenn sie explizit so gesprochen wurden („etc.“ wird nur getippt bei gesprochenem „e te ce“).
6. Wird in der Aufnahme wörtliche Rede zitiert, wird das Zitat in Anführungszeichen gesetzt: „Und ich sagte dann ‚Na, dann schauen wir mal‘“.
7. Wortverkürzungen wie „runtergehen“ statt „heruntergehen“ oder „mal“ statt „einmal“ werden genauso geschrieben, wie sie gesprochen werden.
8. Englische Begriffe werden nach deutschen Rechtschreibregeln in Groß- und Kleinschreibung behandelt.
9. Personalpronomen der zweiten Person (du und ihr) werden kleingeschrieben, die Personalpronomen der Höflichkeitsform (Sie und Ihnen) werden großgeschrieben.
10. Auch Redewendungen/Idiome werden wörtlich wiedergegeben, z. B. „übers Ohr hauen“ (statt: über das Ohr hauen).

C.5 Kodiermanual – Identifikation zentraler Hindernisse bei der Bearbeitung repräsentationsbasierter Chemieaufgaben

Das Kodiermanual wurde von Kechel (2016, S. 255 ff.) adaptiert, d. h. auf das Forschungsanliegen angepasst und Teilweise wörtlich übernommen, besonders mit Blick auf das Abstraktionsniveau, Kategorienformulierung und besondere Kodierregeln wurden Inhalte wörtlich übernommen.

Um einen Einblick in die kognitiven Prozesse und notwendige Fähigkeiten bei der *Interpretation, Translation und Konstruktion* (Kozma & Russell, 1997, 2005) von Studierenden beim Bearbeiten repräsentationsbasierter Chemieaufgaben zu erhalten, werden Hindernisse im Bearbeitungsprozess identifiziert. Dazu bearbeiten Studierende repräsentationsbasierte Chemieaufgaben, die die drei lower-level Skills der Repräsentationskompetenzen (Kozma & Russell, 1997, 2005) in den Fokus rücken und sich auf Keilstrichformeln (2D) und Kugel-Stab-Modelle (3D) von Komplexen begrenzen. Der Bearbeitungsprozess wird videographiert und die Transkripte hinsichtlich externalisierter Hindernisse entsprechend einer induktiven qualitativ orientierten kategoriengeleiteten Textanalyse analysiert, wobei vornehmlich induktiv Kategorien abgeleitet werden (Mayring & Fenzl, 2022).

1 Formulierung von Kategorien verschiedener Hindernisse

1.1 Kategorienbildung

Im Folgenden werden die zentralen Analyseschritte gemäß des Ablaufmodells induktiver und deduktiver Kategorienbildung (Abbildung 25) beschrieben (Mayring & Fenzl, 2022). Dafür werden folgende Analyseeinheiten definiert:

Auswertungseinheit – Die Transkripte werden einzeln nacheinander der qualitativ orientierten kategoriengeleiteten Textanalyse unterzogen.

Kontexteinheit – Das Testheft ist in drei Aufgaben mit Teilaufgaben strukturiert. Die einzelnen Teilaufgaben stellen die Kontexteinheit dar. Die qualitativ orientierte kategoriengeleitete Textanalyse wird so strukturiert, dass jede Teilaufgabe über ein eigenes Kategoriensystem verfügt, wobei die Kategorien der

einzelnen Teilaufgaben identisch sind. Dadurch soll die Analyse und Auswertung erleichtert werden.

Kodiereinheit – Der minimale zu kodierende Textbestandteil sind Phrasen in einem Sinnzusammenhang, beispielsweise ein Hauptsatz in einer Satzreihe oder ein Haupt- bzw. Nebensatz in einem Satzgefüge.

Festlegung des Selektionskriterium zur Identifikation von Hindernissen

Ein externalisiertes¹ Hindernis² liegt immer dann vor, wenn eine Person einen Aufgabenteil³ als schwierig empfindet, an einer bestimmten Stelle nicht weiterweiß oder ratlos ist und dies verbalisiert⁴ oder wenn eine außenstehende Person beobachten kann, dass eine Person längere Pausen⁵ beim lauten Denken macht und dann ggf. mit einem anderen Aufgabenteil fortfährt, einen Aufgabenteil unsicher und zweifelnd⁶ bearbeitet, lückenhafte bzw. fehlerhafte Informationen und Kenntnisse⁷ anwendet, die das erfolgreiche Lösen eines Aufgabenteils verzögern bzw. verhindern, oder Verbalisierungen distinkt zu selbst gestalteten Repräsentationen⁸ sind.

Nähere Erläuterungen zum Selektionskriterium

1) „externalisiertes“

In diesem Kontext meint externalisieren (1) das direkte Verbalisieren von wahrgenommenen Schwierigkeiten und Problemen (= Hindernis) beim Bearbeiten der Aufgaben (z. B. „Oder ist Koordinationszahl doch das wie viele Liganden gebunden sind? Ich habe absolut keine Ahnung“ (KRHR17, Abs. 14)), (2) das indirekte Äußern von Hindernissen durch entsprechendes Verhalten (s. u.), (3) das Anfertigen von schriftlichen Notizen bzw. Repräsentationen und (4) das Nutzen von Gesten.

2) „Hindernis“

Ein Hindernis (auch Barriere, Problem, Hürde, Schwierigkeit) ist „eine Stelle in einem Bearbeitungsprozess, in der rekonstruierbar ist, dass eine Person nichts oder etwas nicht selbstverständlich (im Sinne von nicht sicher, zweifelnd) ausführt und dabei auf nichts in der Aufgabensituation Anwendbares zurückgreifen möchte bzw. zurückgreifen kann [...]“ (Lange, 2013, S. 32). In dieser Studie stellt die Aufgabensituation das Lösen der Teilaufgaben dar. Hindernisse liegen

immer dann vor, wenn das Ziel, also das Lösen der Teilaufgabe nicht oder nur mit großer Mühe gelingt.

3) „Aufgabenteil“

Ein Aufgabenteil ist einen Aspekt einer Teilaufgabe (z. B. A 1.1 a)) oder eine gesamte Teilaufgabe.

4) „verbalisiert“

Eine Verbalisierung stellt eine direkte Aussage einer Person dar und wird direkt aus dem Transkript abgelesen. Es werden keine Interpretationen dieser Aussage vorgenommen.

5) „längere Pausen“

Pausen ab einer Länge von vier Sekunden werden als längere Pausen gewertet (z. B. „Warte mal (8). Ähm, [...]“ (KRHR17, Abs. 23)).

6) „unsicher und zweifelnd“

Unsicherheiten und Zweifel werden hier kodiert, wenn mind. drei kurze Pausen (ab einer Sekunde) in einer Phrase aufeinander folgen, mind. drei Füllwörter (wie ähm) aufeinanderfolgend genutzt werden oder viele Satzbrüche aufeinanderfolgen (z. B. „Hm (..) und (...) okay (7) okay, ähm also was ich/ [...]“ (KRHR17, Abs. 22) oder „Ähm (.), na na na (.) / durch Drehungen um eine Einfachbindung/ ne [...]“ (JKPS07, Abs. 27)).

7) „Informationen und Kenntnisse“

Informationen und Kenntnisse können zum einen direkt aus einer Repräsentation oder der Aufgabe abgeleitet werden und zum anderen deklaratives Wissen, d. h. beispielsweise Faktenwissen über chemische Konzepte, Definitionen usw. darstellen.

8) „selbst gestalteten Repräsentationen“

Von der Person skizzierte Strukturformeln, aufgestellte Summenformeln, Stichpunkte usw.

1.2 Abstraktionsniveau & Kategorienformulierung

Die herauszuarbeitenden Kategorien sollen so konkret formuliert werden, dass sich aus den Labels ableiten lässt, welcher Art die kognitiven Prozesse bzw. Fähigkeiten sind und was deren Inhalt ist. Die Formulierung der Kategorien sollte möglichst kurz und prägnant sein. Anzustreben sind (substantivische) Formulierungen, aus denen das vorliegende Hindernis deutlich wird.

Beispiel: „Wie muss ich es denn gedreht haben? (Drehende Handbewegungen) Ich muss es dann einmal um 180 Grad gedreht haben und dann (Hand dreht sich auf Papierebene um 180 Grad nach links) ist dann hier das NO2. Nein, dann ist das NO2 ja nach links (dreht Hand wieder). Ich kann es ja auch andersherum gedreht haben“ (TWEP29, Abs. 43).

⇒ Hindernis bei mentaler Manipulation chemischer Repräsentationen

1.3 Besondere Kodierregeln

- Die Kodierung externalisierter Hindernisse erfolgt in den Transkripten der Videos/Audios mithilfe des Programms MAXQDA®.
- Fällt die Entscheidung schwer, ob und wie ein Abschnitt kodiert werden soll, kann dieser zunächst unter dem Label „Label diskutieren“ mit einem entsprechenden Memo vermerkt werden, um diesen im Anschluss gemeinsam einzuordnen bzw. zu verwerfen.
- Es werden nur Abschnitte und Phrasen kodiert, in denen Hindernisse identifiziert werden. Alle anderen Abschnitte und Phrasen werden nicht kodiert.
- Zur Schaffung einer besseren Transparenz ist bei jeder ersten Kodierung eines Hindernisses zu vermerken, warum es sich bei der kodierten Phrase im Sinne der o. g. Selektionskriteriums tatsächlich um ein entsprechendes Hindernis handelt.
- Tritt ein Hindernis mehrfach in einem Absatz (Kontexteinheit) auf, so wird jede der entsprechenden Phrasen mit derselben Kategorie kodiert.
- Sind Hindernisse direkt im Testheft zu beobachten (beispielsweise eine fehlerhafte Skizze o. ä.), so sind diese im Transkript zu dem Zeitpunkt zu kodieren, an dem die Person die entsprechende Antwort in das Begleitheft

einträgt. Dabei ist zu prüfen, ob es während der Bearbeitung eine Aussage oder Diskussion durch die Person gibt.

- Liegt ein Hindernis nach o. g. Selektionskriterium vor, das allerdings im späteren Verlauf der Bearbeitung korrigiert wird, so ist dies dennoch als Hindernis zu werten, wenn sich hierdurch Verzögerungen im Lösungsprozess ergeben. Wird ein Hindernis unmittelbar im Anschluss daran (max. zwei Sätze danach) selbst von der Person korrigiert, so ist dies nicht als Hindernis zu kodieren.
- Tritt ein Hindernis auf, das durch die empirische Untersuchung bedingt ist, so wird diese Stelle ebenfalls kodiert und unter dem Label „4.5) Erhebungsbedingte und persönliche Hindernisse“ eingeordnet.
- Um zu vermeiden, dass identische Hindernisse in verschiedenen Transkripten mit unterschiedlichen Kategorien bezeichnet werden, wird das Kategoriensystem von Kodierung zu Kodierung fortgesetzt. Es wird also nicht bei jedem neuen Transkript ein komplett neues Kategoriensystem erstellt, sondern vielmehr das bereits erstellte Kategoriensystem erweitert.
- Um das Vorhandensein einer Kategorie einer Teilaufgabe zu prüfen, wird für jede Teilaufgabe ein separates Kategoriensystem in MAXQDA® angelegt, wobei die Kategoriensysteme der einzelnen Teilaufgaben identisch sind.

Bisher erschienene Bände der Reihe
Studien zum Physik- und Chemielernen

ISSN 1614-8967

Vollständige Übersicht auf unserer Website



<https://www.logos-verlag.de/spcl>

Aktuelle Bände

- 311 Marvin Rost (2021): Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht der Sekundarstufe I. Entwicklung und quantitative Dimensionalitätsanalyse eines Testinstruments aus epistemologischer Perspektive
ISBN 978-3-8325-5256-5 44.00 EUR (open access)
- 312 Christina Kobl (2021): Förderung und Erfassung der Reflexionskompetenz im Fach Chemie
ISBN 978-3-8325-5259-6 41.00 EUR (open access)
- 313 Ann-Kathrin Beretz (2021): Diagnostische Prozesse von Studierenden des Lehramts. eine Videostudie in den Fächern Physik und Mathematik
ISBN 978-3-8325-5288-6 45.00 EUR (open access)
- 314 Judith Breuer (2021): Implementierung fachdidaktischer Innovationen durch das Angebot materialgestützter Unterrichtskonzeptionen. Fallanalysen zum Nutzungsverhalten von Lehrkräften am Beispiel des Münchener Lehrgangs zur Quantenmechanik
ISBN 978-3-8325-5293-0 50.50 EUR (open access)
- 315 Michaela Oettle (2021): Modellierung des Fachwissens von Lehrkräften in der Teilchenphysik. Eine Delphi-Studie
ISBN 978-3-8325-5305-0 57.50 EUR (open access)
- 316 Volker Brüggemann (2021): Entwicklung und Pilotierung eines adaptiven Multistage-Tests zur Kompetenzerfassung im Bereich naturwissenschaftlichen Denkens
ISBN 978-3-8325-5331-9 40.00 EUR (open access)
- 317 Stefan Müller (2021): Die Vorläufigkeit und soziokulturelle Eingebundenheit naturwissenschaftlicher Erkenntnisse. Kritische Reflexion, empirische Befunde und fachdidaktische Konsequenzen für die Chemielehrer*innenbildung
ISBN 978-3-8325-5343-2 63.00 EUR
- 318 Laurence Müller (2021): Alltagsentscheidungen für den Chemieunterricht erkennen und Entscheidungsprozesse explorativ begleiten
ISBN 978-3-8325-5379-1 59.00 EUR

- 319 Lars Ehlert (2021): Entwicklung und Evaluation einer Lehrkräftefortbildung zur Planung von selbstgesteuerten Experimenten
ISBN 978-3-8325-5393-7 41.50 EUR (open access)
- 320 Florian Seiler (2021): Entwicklung und Evaluation eines Seminarkonzepts zur Förderung der experimentellen Planungskompetenz von Lehramtsstudierenden im Fach Chemie
ISBN 978-3-8325-5397-5 47.50 EUR (open access)
- 321 Nadine Boele (2021): Entwicklung eines Messinstruments zur Erfassung der professionellen Unterrichtswahrnehmung von (angehenden) Chemielehrkräften hinsichtlich der Lernunterstützung
ISBN 978-3-8325-5402-6 46.50 EUR
- 322 Franziska Zimmermann (2022): Entwicklung und Evaluation digitalisierungsbezogener Kompetenzen von angehenden Chemielehrkräften
ISBN 978-3-8325-5410-1 49.50 EUR
- 323 Lars-Frederik Weiß (2021): Der Flipped Classroom in der Physik-Lehre. Empirische Untersuchungen in Schule und Hochschule
ISBN 978-3-8325-5418-7 51.00 EUR
- 324 Tilmann Steinmetz (2021): Kumulatives Lehren und Lernen im Lehramtsstudium Physik. Theorie und Evaluation eines Lehrkonzepts
ISBN 978-3-8325-5421-7 51.50 EUR
- 325 Kübra Nur Celik (2022): Entwicklung von chemischem Fachwissen in der Sekundarstufe I. Validierung einer Learning Progression für die Basiskonzepte „Struktur der Materie“, „Chemische Reaktion“ und „Energie“ im Kompetenzbereich „Umgang mit Fachwissen“
ISBN 978-3-8325-5431-6 55.00 EUR
- 326 Matthias Ungermann (2022): Förderung des Verständnisses von Nature of Science und der experimentellen Kompetenz im Schüler*innen-Labor Physik in Abgrenzung zum Regelunterricht
ISBN 978-3-8325-5442-2 55.50 EUR
- 327 Christoph Hoyer (2022): Multimedial unterstütztes Experimentieren im webbasierten Labor zur Messung, Visualisierung und Analyse des Feldes eines Permanentmagneten
ISBN 978-3-8325-5453-8 45.00 EUR
- 328 Tobias Schüttler (2022): Schülerlabore als interesselördernde authentische Lernorte für den naturwissenschaftlichen Unterricht nutzen
ISBN 978-3-8325-5454-5 50.50 EUR
- 329 Christopher Kurth (2022): Die Kompetenz von Studierenden, Schülerschwierigkeiten beim eigenständigen Experimentieren zu diagnostizieren
ISBN 978-3-8325-5457-6 58.50 EUR
- 330 Dagmar Michna (2022): Inklusiver Anfangsunterricht Chemie. Entwicklung und Evaluation einer Unterrichtseinheit zur Einführung der chemischen Reaktion
ISBN 978-3-8325-5463-7 49.50 EUR

- 331 Marco Seiter (2022): Die Bedeutung der Elementarisierung für den Erfolg von Mechanikunterricht in der Sekundarstufe I
ISBN 978-3-8325-5471-2 66.00 EUR
- 332 Jörn Hägele (2022): Kompetenzaufbau zum experimentbezogenen Denken und Arbeiten. Videobasierte Analysen zu Aktivitäten und Vorstellungen von Schülerinnen und Schülern der gymnasialen Oberstufe bei der Bearbeitung von fachmethodischer Instruktion
ISBN 978-3-8325-5476-7 56.50 EUR (open access)
- 333 Erik Heine (2022): Wissenschaftliche Kontroversen im Physikunterricht. Explorationsstudie zum Umgang von Physiklehrkräften und Physiklehramtsstudierenden mit einer wissenschaftlichen Kontroverse am Beispiel der Masse in der Speziellen Relativitätstheorie
ISBN 978-3-8325-5478-1 48.50 EUR (open access)
- 334 Simon Goertz (2022): Module und Lernzirkel der Plattform FLExKom zur Förderung experimenteller Kompetenzen in der Schulpraxis. Verlauf und Ergebnisse einer Design-Based Research Studie
ISBN 978-3-8325-5494-1 66.50 EUR
- 335 Christina Toschka (2022): Lernen mit Modellexperimenten. Empirische Untersuchung der Wahrnehmung und des Denkens in Analogien beim Umgang mit Modellexperimenten
ISBN 978-3-8325-5495-8 50.00 EUR (open access)
- 336 Alina Behrendt (2022): Chemiebezogene Kompetenzen in der Übergangsphase zwischen dem Sachunterricht der Primarstufe und dem Chemieunterricht der Sekundarstufe I
ISBN 978-3-8325-5498-9 40.50 EUR (open access)
- 337 Manuel Daiber (2022): Entwicklung eines Lehrkonzepts für eine elementare Quantenmechanik. Formuliert mit In-Out Symbolen
ISBN 978-3-8325-5507-8 48.50 EUR
- 338 Felix Pawlak (2022): Das Gemeinsame Experimentieren (an-)leiten. Eine qualitative Studie zum chemiespezifischen Classroom-Management
ISBN 978-3-8325-5508-5 46.50 EUR
- 339 Liza Dopatka (2022): Konzeption und Evaluation eines kontextstrukturierten Unterrichtskonzeptes für den Anfangs-Elektrizitätslehreunterricht
ISBN 978-3-8325-5514-6 69.50 EUR
- 340 Arne Bewersdorff (2022): Untersuchung der Effektivität zweier Fortbildungsformate zum Experimentieren mit dem Fokus auf das Unterrichtshandeln
ISBN 978-3-8325-5522-1 39.00 EUR (open access)

- 341 Thomas Christoph Münster (2022): Wie diagnostizieren Studierende des Lehramtes physikbezogene Lernprozesse von Schüler*innen?. Eine Videostudie zur Mechanik
ISBN 978-3-8325-5534-4 44.50 EUR (open access)
- 342 Ines Komor (2022): Förderung des symbolisch-mathematischen Modellverständnisses in der Physikalischen Chemie
ISBN 978-3-8325-5546-7 46.50 EUR
- 343 Verena Petermann (2022): Überzeugungen von Lehrkräften zum Lehren und Lernen von Fachinhalten und Fachmethoden und deren Beziehung zu unterrichtsnahem Handeln
ISBN 978-3-8325-5545-0 47.00 EUR (open access)
- 344 Jana Heinze (2022): Einfluss der sprachlichen Konzeption auf die Einschätzung der Qualität instruktionaler Unterrichtserklärungen im Fach Physik
ISBN 978-3-8325-5553-5 42.00 EUR (open access)
- 345 Jannis Weber (2022): Mathematische Modellbildung und Videoanalyse zum Lernen der Newtonschen Dynamik im Vergleich
ISBN 978-3-8325-5566-5 68.00 EUR (open access)
- 346 Fabian Sterzing (2022): Zur Lernwirksamkeit von Erklärvideos in der Physik. Eine Untersuchung in Abhängigkeit von ihrer fachdidaktischen Qualität und ihrem Einbettungsformat
ISBN 978-3-8325-5576-4 52.00 EUR (open access)
- 347 Lars Greitemann (2022): Wirkung des Tablet-Einsatzes im Chemieunterricht der Sekundarstufe I unter besonderer Berücksichtigung von Wissensvermittlung und Wissenssicherung
ISBN 978-3-8325-5580-1 50.00 EUR
- 348 Fabian Poensgen (2022): Diagnose experimenteller Kompetenzen in der laborpraktischen Chemielehrer*innenbildung
ISBN 978-3-8325-5587-0 48.00 EUR
- 349 William Lindlahr (2023): Virtual-Reality-Experimente. Entwicklung und Evaluation eines Konzepts für den forschend-entwickelnden Physikunterricht mit digitalen Medien
ISBN 978-3-8325-5595-5 49.00 EUR
- 350 Bert Schlüter (2023): Teilnahmemotivation und situationales Interesse von Kindern und Eltern im experimentellen Lernsetting KEMIE
ISBN 978-3-8325-5598-6 43.00 EUR
- 351 Katharina Nave (2023): Charakterisierung situativer mentaler Modellkomponenten in der Chemie und die Bildung von Hypothesen. Eine qualitative Studie zur Operationalisierung mentaler Modellkomponenten für den Fachbereich Chemie
ISBN 978-3-8325-5599-3 43.00 EUR
- 352 Anna B. Bauer (2023): Experimentelle Kompetenz Physikstudierender. Entwicklung und erste Erprobung eines performanzorientierten Kompetenzstrukturmodells unter Nutzung qualitativer Methoden
ISBN 978-3-8325-5625-9 47.00 EUR (open access)

- 353 Jan Schröder (2023): Entwicklung eines Performanztests zur Messung der Fähigkeit zur Unterrichtsplanung bei Lehramtsstudierenden im Fach Physik
ISBN 978-3-8325-5655-6 46.50 EUR
- 354 Susanne Gerlach (2023): Aspekte einer Fachdidaktik Körperpflege. Ein Beitrag zur Standardentwicklung
ISBN 978-3-8325-5659-4 45.00 EUR
- 355 Livia Murer (2023): Diagnose experimenteller Kompetenzen beim praktisch-naturwissenschaftlichen Arbeiten. Vergleich verschiedener Methoden und kognitive Validierung eines Testverfahrens
ISBN 978-3-8325-5657-0 41.50 EUR (open access)
- 356 Andrea Maria Schmid (2023): Authentische Kontexte für MINT-Lernumgebungen. Eine zweiteilige Interventionsstudie in den Fachdidaktiken Physik und Technik
ISBN 978-3-8325-5605-1 57.00 EUR (open access)
- 357 Julia Ortmann (2023): Bedeutung und Förderung von Kompetenzen zum naturwissenschaftlichen Denken und Arbeiten in universitären Praktika
ISBN 978-3-8325-5670-9 37.00 EUR (open access)
- 358 Axel-Thilo Prokop (2023): Entwicklung eines Lehr-Lern-Labors zum Thema Radioaktivität. Eine didaktische Rekonstruktion
ISBN 978-3-8325-5671-6 49.50 EUR
- 359 Timo Hackemann (2023): Textverständlichkeit sprachlich variiert physikbezogener Sachtexte
ISBN 978-3-8325-5675-4 41.50 EUR (open access)
- 360 Dennis Dietz (2023): Vernetztes Lernen im fächerdifferenzierten und integrierten naturwissenschaftlichen Unterricht aufgezeigt am Basiskonzept Energie. Eine Studie zur Analyse der Wirksamkeit der Konzeption und Implementation eines schulinternen Curriculums für das Unterrichtsfach „Integrierte Naturwissenschaften 7/8“
ISBN 978-3-8325-5676-1 49.50 EUR
- 361 Ann-Katrin Krebs (2023): Vielfalt im Physikunterricht. Zur Wirkung von Lehrkräftefortbildungen unter Diversitätsaspekten
ISBN 978-3-8325-5672-3 65.50 EUR (open access)
- 362 Simon Kaulhausen (2023): Strukturelle Ursachen für Klausurmisserfolg in Allgemeiner Chemie an der Universität
ISBN 978-3-8325-5699-0 37.50 EUR (open access)
- 363 Julia Eckoldt (2023): Den (Sach-)Unterricht öffnen. Selbstkompetenzen und motivationale Orientierungen von Lehrkräften bei der Implementation einer Innovation untersucht am Beispiel des Freien Explorierens und Experimentierens
ISBN 978-3-8325-5663-1 48.50 EUR (open access)
- 364 Albert Teichrow (2023): Physikalische Modellbildung mit dynamischen Modellen
ISBN 978-3-8325-5710-2 58.50 EUR (open access)

- 365 Sascha Neff (2023): Transfer digitaler Innovationen in die Schulpraxis. Eine explorative Untersuchung zur Förderung der Implementation
ISBN 978-3-8325-5687-7 59.00 EUR (open access)
- 366 Rahel Schmid (2023): Verständnis von Nature of Science-Aspekten und Umgang mit Fehlern von Schüler*innen der Sekundarstufe I. Am Beispiel von digital-basierten Lernprozessen im informellen Lernsetting Smartfeld
ISBN 978-3-8325-5722-5 53.50 EUR (open access)
- 367 Dennis Kirstein (2023): Individuelle Bedingungs- und Risikofaktoren für erfolgreiche Lernprozesse mit kooperativen Experimentieraufgaben im Chemieunterricht. Eine Untersuchung zum Zusammenhang von Lernvoraussetzungen, Lerntätigkeiten, Schwierigkeiten und Lernerfolg beim Experimentieren in Kleingruppen der Sekundarstufe I
ISBN 978-3-8325-5729-4 50.50 EUR (open access)
- 368 Frauke Düwel (2024): Argumentationslinien in Lehr-Lernkontexten. Potenziale englischer Fachtexte zur Chromatografie und deren hochschuldidaktische Einbindung
ISBN 978-3-8325-5731-7 63.00 EUR (open access)
- 369 Fabien Güth (2023): Interessenbasierte Differenzierung mithilfe systematisch variierten Kontextaufgaben im Fach Chemie
ISBN 978-3-8325-5737-9 48.00 EUR (open access)
- 370 Oliver Grewe (2023): Förderung der professionellen Unterrichtswahrnehmung und Selbstwirksamkeitsüberzeugungen hinsichtlich sprachsensibler Maßnahmen im naturwissenschaftlichen Sachunterricht. Konzeption und Evaluation einer video- und praxisbasierten Lehrveranstaltung im Masterstudium
ISBN 978-3-8325-5738-6 44.50 EUR (open access)
- 371 Anna Nowak (2023): Untersuchung der Qualität von Selbstreflexionstexten zum Physikunterricht. Entwicklung des Reflexionsmodells REIZ
ISBN 978-3-8325-5739-3 59.00 EUR (open access)
- 372 Dominique Angela Holland (2023): Bildung für nachhaltige Entwicklung (BNE) kooperativ gestalten. Vergleich monodisziplinärer und interdisziplinärer Kooperation von Lehramtsstudierenden bei der Planung, Durchführung und Reflexion von Online-BNE-Unterricht
ISBN 978-3-8325-5760-7 47.00 EUR (open access)
- 373 Renan Marcello Vairo Nunes (2024): MINT-Personal an Schulen. Eine Untersuchung der Arbeitssituation und professionellen Kompetenzen von MINT-Lehrkräften verschiedener Ausbildungswege
ISBN 978-3-8325-5778-2 51.00 EUR (open access)
- 374 Mats Kieserling (2024): Digitalisierung im Chemieunterricht. Entwicklung und Evaluation einer experimentellen digitalen Lernumgebung mit universeller Zugänglichkeit
ISBN 978-3-8325-5786-7 45.50 EUR
- 375 Cem Aydin Salim (2024): Die Untersuchung adaptiver Lernsettings im Themenbereich „Schwimmen und Sinken“ im naturwissenschaftlichen Unterricht
ISBN 978-3-8325-5787-4 49.00 EUR (open access)

- 376 Novid Ghassemi (2024): Evaluation eines Lehramtsmasterstudiengangs mit dem Profil Quereinstieg im Fach Physik. Erkenntnisse zu Eingangsbedingungen, professionellen Kompetenzen und Aspekten individueller Angebotsnutzung
ISBN 978-3-8325-5789-8 41.50 EUR (open access)
- 377 Martina Flurina Cavelti (2024): Entwicklung und Validierung eines Messinstruments zur Erfassung der Schülerkompetenzen im Bereich des wissenschaftlichen Skizzierens im Fach Chemie in der Sekundarstufe I
ISBN 978-3-8325-5829-1 45.00 EUR (open access)
- 378 Tom Bleckmann (2024): Formatives Assessment auf Basis von maschinellem Lernen. Eine Studie über automatisiertes Feedback zu Concept Maps aus dem Bereich Mechanik
ISBN 978-3-8325-5842-0 46.50 EUR (open access)
- 379 Jana Marlies Rehberg (2024): Das physikspezifische Mindset zum Studienbeginn. Fragebogenentwicklung und Aufbau einer Online-Intervention
ISBN 978-3-8325-5850-5 59.50 EUR (open access)
- 380 Florian Trauten (2024): Entwicklung und Evaluation von automatisierten Feedbackschleifen in Online-Aufgaben im Fach Chemie
ISBN 978-3-8325-5859-8 46.00 EUR (open access)
- 381 Johanna Dejanovikj (2024): Reflexion als Lerngegenstand im Unterricht zur Förderung von Bewertungskompetenz
ISBN 978-3-8325-5860-4 41.00 EUR (open access)
- 382 Katharina Flieler (2024): Verständlichkeit physikalischer Sachtexte. Untersuchungen zum Wirkungsgefüge zwischen sprachlicher Textgestaltung und der Behaltensleistung sowie der Textwahrnehmung im Schulfach Physik
ISBN 978-3-8325-5858-1 44.00 EUR (open access)
- 383 Stephanie Neppel (2024): Perspektivenübernahme im Physikunterricht. Explorative Interviewstudie zu einer Seminarkonzeption mit dem Schwerpunkt Perspektivenübernahme bei der Planung von Physikunterricht
ISBN 978-3-8325-5865-9 48.00 EUR (open access)
- 384 Katja Plicht (2024): Ein Physikübungskonzept zur Förderung der Problemlösekompetenz. Entwicklung und empirische Evaluation eines Strategietrainings auf der Basis von Expertisemerkmalen
ISBN 978-3-8325-5875-8 45.00 EUR (open access)
- 385 Svenja Boegel (2024): Feedback beim Experimentieren: Zur Rolle von Cognitive Load und Motivation
ISBN 978-3-8325-5911-3 45.50 EUR (open access)
- 386 Laura Pannullo (2025): Wahlmöglichkeiten beim Experimentieren: Entwicklung und Erprobung eines Konzeptes für Experimente in inklusiven Lerngruppen im Physikunterricht
ISBN 978-3-8325-5916-8 46.50 EUR (open access)

- 387 Matthias Schweinberger (2025): „Stumme Videos“. Aufmerksamkeitslenkende Moderation von Demonstrationsexperimenten
ISBN 978-3-8325-5930-4 50.50 EUR
- 388 Rike Große-Heilmann (2025): Entwicklung fachdidaktischen Wissens zum Einsatz digitaler Medien im Fach Physik
ISBN 978-3-8325-5943-4 51.00 EUR (open access)
- 389 Lion Cornelius Glatz (2025): Eine vergleichende Betrachtung der Überzeugungskraft von Experimenten zum Teilchenmodell
ISBN 978-3-8325-5954-0 42.00 EUR (open access)
- 390 Rasmus Viefers (2025): Förderung der Variablenkontrollstrategie im Sachunterricht
ISBN 978-3-8325-5968-7 44.50 EUR (open access)
- 391 Luc Broder Albrecht (2025): Kritisches Denken in der Chemielehrer*innenbildung. Eine konzeptionell-empirische Studie zur Integration und Operationalisierung
ISBN 978-3-8325-5970-0 57.00 EUR (open access)
- 392 Anna Weißbach (2025): Reflexion von Physikunterricht. Entwicklung und Validierung eines Testinstruments mit Assessment-Feedback zur Reflexionsfähigkeit Physik-Lehramtsstudierender
ISBN 978-3-8325-5977-9 51.50 EUR (open access)
- 393 Robert Gieske (2025): Förderung fachwissenschaftlicher und kommunikativer Kompetenzen von Schüler*innen durch Disaggregate Instruction im Fach Chemie. Konzeption und Durchführung einer vergleichenden Interventionsstudie zur Analyse der Wirksamkeit sprachsensibler Unterrichtsansätze in der Sekundarstufe I
ISBN 978-3-8325-5984-7 51.00 EUR (open access)
- 394 Thomas Sean Weatherby (2025): Talking Circuits. The Development and Assessment of a Digitally-Scaffolded, Collaborative Method for Teaching and Learning Electrical Circuits in Early Secondary Schools
ISBN 978-3-8325-6008-9 57.00 EUR (open access)
- 395 Julia Welberg (2025): Wen interessiert's? Quantitative Untersuchungen der Zusammenhänge der Neigungen zu empathisierender und systematisierender Denkweise und Interesse an Physik
ISBN 978-3-8325-6021-8 42.00 EUR (open access)
- 396 Sebastian Nickel (2026): Repräsentationskompetenzen im Fach Chemie. Eine empirische Untersuchung zur Dimensionalität des Konstrukts und zu dessen Zusammenhang mit Fachwissen und räumlichen Fähigkeiten
ISBN 978-3-8325-6074-4 55.00 EUR (open access)
- 397 Manuel Becker (2026): Der Energie-Feld-Ansatz: Entwicklung und Evaluation eines Konzeptes für den Energieunterricht der Sekundarstufe II
ISBN 978-3-8325-6079-9 53.50 EUR (open access)

- 398 Sevan Khagy (2026): Lernen durch Erklären – Entwicklung und Evaluation eines Seminarkonzepts zur Förderung der Sach- und Erklärkompetenz von Chemie-Lehramtsstudierenden
ISBN 978-3-8325-6081-2 52.50 EUR (open access)

Vollständige Übersicht unter: <https://www.logos-verlag.de/spcl>

Alle erschienenen Bücher können unter der angegebenen ISBN direkt online (<http://www.logos-verlag.de>) oder telefonisch (030 - 42 85 10 90) beim Logos Verlag Berlin bestellt werden.

Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf und Mathias Ropohl

Die Reihe umfasst inzwischen eine große Zahl von wissenschaftlichen Arbeiten aus vielen Arbeitsgruppen der Physik- und Chemiedidaktik und zeichnet damit ein gültiges Bild der empirischen physik- und chemiedidaktischen Forschung im deutschsprachigen Raum.

Die Herausgeber laden daher Interessenten zu neuen Beiträgen ein und bitten sie, sich im Bedarfsfall an den Logos-Verlag oder an ein Mitglied des Herausgeberteams zu wenden.

Kontaktadressen:

Univ.-Prof. Dr. Martin Hopf
Universität Wien,
Österreichisches Kompetenzzentrum
für Didaktik der Physik,
Porzellangasse 4, Stiege 2,
1090 Wien, Österreich,
Tel. +43-1-4277-60330,
e-mail: martin.hopf@univie.ac.at

Prof. Dr. Mathias Ropohl
Didaktik der Chemie,
Fakultät für Chemie,
Universität Duisburg-Essen,
Schützenbahn 70, 45127 Essen,
Tel. 0201-183 2704,
e-mail: mathias.ropohl@uni-due.de

Repräsentationen sind im Fach Chemie unverzichtbar, stellen Lernende jedoch zugleich vor Herausforderungen. Ein erfolgreicher Umgang mit ihnen erfordert Repräsentationskompetenzen, fachliches Wissen sowie räumliche Fähigkeiten. Die Struktur und die wechselseitigen Beziehungen dieser drei Faktoren sind bislang jedoch nicht hinreichend untersucht.

Im Rahmen des vorliegenden Projekts wurde zunächst das Chemical Representation Inventory: Translation, Interpretation, Construction (CRI:TIC) entwickelt und anhand einer Stichprobe von $N = 185$ Studierenden aus einem Chemievorkurs evaluiert. In einer Folgestudie wurden das überarbeitete CRI:TIC sowie Instrumente zur Erfassung des deklarativen Fachwissens und verschiedener räumlicher Faktoren erneut in Chemievorkursen eingesetzt ($N = 494$). Zur vertieften Analyse der Zusammenhänge zwischen den drei Konstrukten wurden die quantitativen Daten durch qualitative Erhebungen mit $N = 25$ Teilnehmenden ergänzt.

Die Ergebnisse zeigen unter anderem, dass mentale Rotations- und Manipulationsfähigkeiten neben dem deklarativen Fachwissen eine zentrale Rolle bei der erfolgreichen Bearbeitung repräsentationsbasierter Chemieaufgaben einnehmen. Für eine Optimierung von Lehr-Lern-Prozessen mit chemischen Repräsentationen sollten daher Repräsentationskompetenzen sowie relevante räumliche Faktoren gezielt und systematisch – ergänzend zum (deklarativen) Fachwissen – gefördert und entwickelt werden.

Logos Verlag Berlin

ISBN 978-3-8325-6074-4