Studien zum Physik- und Chemielernen

M. Hopf und M. Ropohl [Hrsg.]

369

Fabien Güth

Interessenbasierte Differenzierung mithilfe systematisch variierter Kontextaufgaben im Fach Chemie

λογος

Studien zum Physik- und Chemielernen

Band 369

Fabien Güth

Interessenbasierte Differenzierung mithilfe systematisch variierter Kontextaufgaben im Fach Chemie

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 Attribution 4.0 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 2023

ISBN 978-3-8325-5737-9
ISSN 1614-8967
DOI 10.30819/5737

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

Interessenbasierte Differenzierung mithilfe systematisch variierter Kontextaufgaben im Fach Chemie

Dissertation

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

- Dr. rer. nat -

vorgelegt von

Fabien Louis Güth

geboren in Gelsenkirchen

Fakultät für Chemie

der Universität Duisburg-Essen

Mai 2023

Die vorliegende Arbeit wurde im Zeitraum von September 2019 bis April 2023 in der Arbeitsgruppe von Prof. Dr. Elke Sumfleth unter Betreuung von Dr. Helena van Vorst in der Didaktik der Chemie der Universität Duisburg-Essen durchgeführt.

1. Gutachterin: Prof. Dr. Elke Sumfleth

2. Gutachter: Prof. Dr. Sebastian Habig

3. Gutachterin: Prof. Dr. Mirjam Steffensky

Vorsitzende: Jun.-Prof. Dr. Anzhela Galstyan

Tag der Disputation: 20.09.2023

Für meinen Großvater Dr. Wulf Jordan

Danksagungen

An dieser Stelle möchte ich mich bei den vielen Menschen bedanken, die mich im Verlauf meiner Promotion maßgeblich unterstützt und zum Gelingen dieser Arbeit beigetragen haben.

Zunächst danke ich Frau Prof. Dr. Elke Sumfleth: Liebe Elke, ich danke dir vielmals für dein Vertrauen und die Möglichkeit, in deiner Arbeitsgruppe zu promovieren. Du hast mir die besten Arbeitsbedingungen geboten, die man sich nur wünschen kann. Danke für deine Unterstützung und die guten Ratschläge in den letzten Jahren. Für mich war es immer eine große Freude, in deiner Arbeitsgruppe zu arbeiten.

Außerdem möchte ich mich bei Herrn Prof. Dr. Sebastian Habig für anregende Hinweise, interessante Diskussionen und die Erstellung des Zweitgutachtens bedanken.

Frau Prof. Dr. Mirjam Steffensky möchte ich ganz herzlich für die Bereitschaft danken, ein drittes Gutachten für meine Arbeit zu erstellen.

Ich danke Frau Jun.-Prof. Dr. Anzhela Galstyan für die Übernahme des Prüfungsvorsitzes.

Mein besonderer Dank gilt Frau Dr. Helena van Vorst. Ich danke dir, liebe Helena, nicht nur für die Mitbetreuung dieser Arbeit, sondern für die gesamte Begleitung meiner bisherigen akademischen Laufbahn. Du hast mir unglaublich viele Freiheiten bei der Gestaltung dieser Arbeit gelassen, so dass ich viele meiner Ideen in dieser Arbeit umsetzen konnte. Zugleich hast du mich stets ermutigt, auch Neues auszuprobieren. Du warst immer für mich da und ich konnte mich immer auf deine Unterstützung verlassen. Vielen Dank, dass ich in den letzten Jahren so viel von dir lernen durfte!

Ich möchte mich bei allen Schulen, Lehrkräften sowie Schülerinnen und Schülern bedanken, die auch während einer Pandemie dazu bereit waren, an meinen Studien teilzunehmen. Dies hat die Durchführung meiner Arbeit überhaupt erst möglich gemacht.

Mein Dank gilt Marie Schwantes, Sonja Dieterich und Anna Liskes, die mich durch ihre Masterarbeiten bzw. durch ihre Tätigkeit als wissenschaftliche Hilfskräfte bei der Durchführung dieser Arbeit unterstützt haben.

Daneben möchte ich mich bei Christoph Pelka für die technische Unterstützung bei der Datenerhebung bedanken. Ohne Janina Schmidt und Katrin Falchi hätte ich vermutlich nicht eine Tagung besucht, da ich am Dienstreiseantrag gescheitert wäre. Hermann Vielhauer und Tessa de Kock danke ich für die Unterstützung während der Lehre.

Allen Kollegen und Kolleginnen an der Schützenbahn danke ich für eine unvergessliche Promotionszeit. Ich hatte unglaublich viel Spaß mit euch bei Konferenzen, Weihnachtsfeiern, Betriebsausflügen, Abenden im Sausalitos und in den Mittagspausen. Ich danke euch allen! Insbesondere möchte ich mich aber bei meinen Bürokollegen Martin Steinbach, Dennis Kirstein und Simon Kaulhausen bedanken. Danke für den fachlichen Austausch, aber vor allem

danke ich euch für all den Unsinn, die vielen gemeinsamen Stunden des Lachens und eure Freundschaft!

Ein ganz besonderer Dank gilt meiner Familie, meiner Mutter Indra Melian und meinem Stiefvater Ralf Melian, meinem Bruder Leon Melian, meinen Großeltern Doris Klein und Wulf Jordan sowie meinen Schwiegereltern Axel Splithöfer und Margret Steden. Danke für euren Rückhalt und eure Unterstützung, nicht nur während des Studiums.

Schließlich möchte ich bei dem wichtigsten Menschen in meinem Leben und meiner größten Unterstützerin bedanken. Kira, danke für dein unglaubliches Verständnis und dafür, dass du immer wieder Entbehrungen hinnimmst, damit ich meine Ziele erreiche. Vielen, vielen Dank für deine Unterstützung seit so vielen Jahren.

Inhaltsverzeichnis

1 Einleitung ... 1

2 Individuelle Förderung im Unterricht ... 5

2.1 Differenzierung im Unterricht ... 6

2.2 Forschungslage zur Differenzierung im Unterricht ... 10

2.3 Kritische Würdigung der Forschungslage ... 13

3 Interessenbasierte Differenzierung: Aufgabenwahl als Differenzierungsansatz ... 15

3.1 Theoretische Grundlagen ... 16

3.1.1 Die Selbstbestimmungstheorie der Motivation ... 16

3.1.2 Das Erwartungs-Wert-Modell leistungsbezogener Wahlentscheidungen ... 20

3.1.3 Die Person-Gegenstands-Konzeption des Interesses ... 22

3.2 Empirische Befunde zur Bereitstellung von Wahloptionen ... 24

3.3 Implikationen für das Fach Chemie ... 27

4 Kontextorientiertes Lernen im naturwissenschaftlichen Unterricht ... 29

4.1 Zum Kontextbegriff ... 30

4.2 Kontextmerkmale ... 33

4.3 Unterrichtliche Gestaltung von kontextorientiertem Lernen ... 36

4.4 Empirische Befunde zum kontextorientierten Lernen ... 38

4.4.1 Effektivität von kontextorientiertem Lernen ... 38

4.4.2 Einfluss der Kontextmerkmale ... 40

4.5 Fazit zum Einsatz von Kontexten zur interessenbasierten Differenzierung ... 44

5 Ziel, Forschungsfragen und Hypothesen ... 47

6 Studiendesign ... 51

6.1 Teilstudie I ... 51

6.2 Teilstudie II ... 53

7 Psychometrische und Statistische Methoden ... 57

7.1 Testtheorie ... 57

7.2 Clusteranalyse ... 63

7.3 Klassische inferenzstatistische Verfahren ... 65

7.4 Pfadanalyse ... 68

7.5 Mehrebenenmodelle ... 69

8 Teilstudie I: Untersuchung der Kontextwahl ... 73

8.1 Entwicklung des Lernmaterials ... 73

8.2 Beschreibung der Messinstrumente ... 80
8.2.1 Erfassung der kognitiven und affektiv-motivationalen Personenmerkmale ... 80
8.2.2 Erfassung der prozessbezogenen Variablen ... 87
8.2.3 Überprüfung der Merkmalszugehörigkeit der Kontextaufgaben ... 89

8.3 Datenerhebung ... 90

8.4 Stichprobe ... 92

8.5 Ergebnisse und Diskussion ... 93
8.5.1 Analyse der Messinstrumente ... 93
8.5.2 Überprüfung der Merkmalszugehörigkeit der Kontextaufgaben ... 104
8.5.3 Auswahl relevanter Personenmerkmale ... 107
8.5.4 Identifizierung der Lernendengruppen ... 112
8.5.5 Kontextwahl der Lernendengruppen ... 117
8.5.6 Evaluation der Kontextwahl ... 120

8.6 Zusammenfassung und Implikationen für die weitere Untersuchung ... 131

9 Entwicklung des Vorhersagemodells ... 133

9.1 Maschinelles Lernen ... 133

9.2 Algorithmen zum maschinellen Lernen ... 139

9.3 Modellierung ... 145
9.3.1 Datenvorverarbeitung ... 145
9.3.2 Optimierung der Hyperparameter ... 148
9.3.3 Evaluation im Test-Datensatz ... 150

9.4 Überprüfung der Vorhersagegüte ... 152
9.4.1 Beschreibung der Messinstrumente ... 153
9.4.2 Datenerhebung und Stichprobe ... 153
9.4.3 Ergebnisse und Diskussion ... 154

9.5 Implikationen und Limitationen ... 158

10 Teilstudie II: Effektivität interessenbasierter Differenzierung ... 161

10.1 Weiterentwicklung des Lernmaterials ... 161

10.2 Beschreibung der Messinstrumente ... 165
10.2.1 Erfassung der Kontrollvariablen ... 165
10.2.2 Erfassung der abhängigen Variablen ... 166

10.3 Datenerhebung ... 170

10.4 Stichprobe ... 173
10.5 Ergebnisse und Diskussion ... 173
10.5.1 Analyse der Messinstrumente ... 174
10.5.2 Vergleichbarkeit der Untersuchungsgruppen ... 185
10.5.3 Überprüfung der Vorhersagegüte ... 187
10.5.4 Einfluss auf die Zufriedenheit ... 189
10.5.5 Einfluss auf das situationale Interesse ... 200
10.5.6 Einfluss auf die kognitive Belastung ... 207
10.5.7 Einfluss auf den Fachwissenszuwachs ... 212
10.5.8 Konsolidierung der Ergebnisse ... 217
10.6 Zusammenfassende Diskussion der Ergebnisse ... 221

11 Zusammenfassung und Limitationen ... 229

12 Implikationen für die Schulpraxis und Forschung ... 239

13 Verzeichnisse ... 243
13.1 Literaturverzeichnis ... 243
13.2 Abbildungsverzeichnis ... 265
13.3 Tabellenverzeichnis ... 269
13.4 Abkürzungsverzeichnis ... 274

14 Anhang ... 277

1 Einleitung

Das schulische Lernen ist maßgeblich durch die individuellen Voraussetzungen der Lernenden geprägt (Hardy et al., 2019; Subban, 2006; Tomlinson et al., 2003). Wie beträchtlich die Unterschiede in den Lernvoraussetzungen der Schülerinnen und Schüler tatsächlich sind, zeigen beispielsweise die Ergebnisse der PISA-Studie (*Programme for International Student Assessmen*t) in regelmäßigen Abständen. Unterschiede lassen sich hierbei nicht nur in Bezug auf kognitive Lernvoraussetzungen, wie die naturwissenschaftliche Kompetenz, sondern auch im Hinblick auf affektiv-motivationale Lernvoraussetzungen wie das Interesse an naturwissenschaftlichen Themen oder die Motivation im Unterricht herausstellen, wie Ergebnisse aus dem Jahr 2015 zeigen (Schiepe-Tiska, Rönnebeck et al., 2016; Schiepe-Tiska, Simm & Schmidtner, 2016). Nicht zuletzt aufgrund der Schulschließungen im Zuge der Covid-19-Pandemie scheint sich diese Lage zunehmend zu verschärfen. Erste empirische Ergebnisse deuten an, dass die metaphorische Bildungsschere zwischen den Schülerinnen und Schülern durch die pandemiebedingten Schulschließungen noch größer wurde (Hammerstein et al., 2021). Zukünftig scheint es damit immer wichtiger zu werden, daraus resultierende individuelle Förderbedarfe im Rahmen von unterrichtlichen Bildungsprozessen zu identifizieren und aufzugreifen.

Die Abstimmung des Unterrichts auf die unterschiedlichen Lernbedarfe der Schülerinnen und Schüler ist eine Schlüsselaufgabe der Lehrkräfte, die zudem ein Merkmal für gute Unterrichtsqualität ist (Helmke, 2017; H. Meyer, 2011). Bemühungen, den Unterricht zunehmend adaptiv zu gestalten, werden in der Unterrichtspraxis und in der wissenschaftlichen Diskussion häufig unter dem Begriff der individuellen Förderung adressiert, der sich zumeist in mehr oder weniger unterschiedliche Konzepte ausdifferenziert (z. B. Dumont, 2019).

Ein verbreitetes Konzept zur individuellen Förderung im Unterricht ist die Differenzierung der Lerngruppe anhand individueller Lernvoraussetzungen (Bohl et al., 2012). Die meisten Differenzierungsansätze fokussieren hierbei vorwiegend leistungsbezogene Lernvoraussetzungen zur adaptiven Gestaltung unterrichtlicher Lehr-Lernprozesse. Evaluationsergebnisse entsprechender Ansätze verweisen aufgrund diverser Forschungszugänge und konträrer Ergebnisse auf keine eindeutige Befundlage (Smale-Jacobse et al., 2019; Steenbergen-Hu et al., 2016). Für das Fach Chemie wurden bislang auch nur vereinzelt Studien durchgeführt, um potenzielle Effekte näher aufzuklären, wobei auch hier ein Schwerpunkt auf der Evaluation leistungsadaptiver Unterrichtsgestaltung liegt (Anus, 2015; Hauerstein, 2019; Kallweit & Melle, 2017). Gleichwohl in der Literatur auch die affektiv-motivationale Heterogenität im Unterricht diskutiert wird (Lazarides & Schiepe-Tiska, 2022), nutzen die meisten Forschungsansätze – insbesondere im Fach Chemie – interindividuell unterschiedliche affektiv-motivationale Lernvoraussetzungen nicht zur adaptiven Unterrichtsgestaltung. Aus diesem Grund befasst sich die vorliegende Arbeit mit der Entwicklung und Evaluation eines interessenbasierten Differenzierungsansatzes für das Fach Chemie.

Die Berücksichtigung affektiv-motivationaler Lernvoraussetzungen kann durch die Öffnung des Unterrichts im Hinblick auf die Auswahl der bereitgestellten Lernmaterialien ermöglicht werden, die hinsichtlich des Themas oder des Anwendungsbezugs variieren (Gijsbers et al., 2020; Morgan & Wagner, 2013; Tomlinson et al., 2003). Aus lehr-lernpsychologischer Perspektive führt die Bereitstellung von Wahloptionen zur Befriedigung des psychologischen Grundbedürfnisses nach Autonomie, welches sich gemäß der Selbstbestimmungstheorie positiv auf die intrinsische Motivation und die Leistung der Lernenden auswirkt (Ryan & Deci, 2020). Zugleich betonen Forschungsarbeiten aber auch die Bedeutsamkeit der Passung zwischen individuellen Merkmalen der Lernenden und den spezifischen Merkmalen der Aufgaben (Wilde et al., 2018). Für das Fach Chemie ist bislang unklar, wie Lernmaterialien und die konkrete Lehr-Lernsituation im Unterricht gestaltet werden können, um affektiv-motivationale Lernvoraussetzungen im Sinne eines interessenbasierten Differenzierungsansatzes aufzugreifen. Forschungsarbeiten aus dem kontextorientierten Lernen im naturwissenschaftlichen Unterricht verweisen hierbei auf das Potential des Einsatzes systematisch variierter Kontexte (z. B. Broman et al., 2018; Habig et al., 2018; Sevian, Hugi-Cleary et al., 2018; van Vorst & Aydogmus, 2021).

Das kontextorientierte Lernen ist im naturwissenschaftlichen Lernen sehr verbreitet und im Chemieunterricht verpflichtend umzusetzen (z. B. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2019). Zentral ist die Idee, durch den Einsatz von lebensweltlichen Kontexten die Relevanz der Unterrichtsinhalte zu steigern und damit das Interesse an naturwissenschaftlichen Fächern zu erhöhen (Gilbert, 2006; Pilot & Bulte, 2006). Empirische Befunde deuten darauf hin, dass unterschiedliche Kontexte für unterschiedliche Gruppen von Lernenden geeignet sind, ohne jedoch konkret aufzuschlüsseln, welche Kontexte für welche Lernenden genutzt werden sollten (Habig et al., 2018; Podschuweit & Bernholt, 2018; van Vorst & Aydogmus, 2021). Insgesamt besteht damit aktuell ein Desiderat hinsichtlich der Adaptivität der Kontexte für einzelne Lernendengruppen, wobei dieses Desiderat für die adaptive Unterrichtsgestaltung in einem interessenbasierten Differenzierungsansatz zwingend aufzulösen ist.

Im Zentrum der ersten Projektphase dieser Arbeit stand deshalb die Untersuchung der Kontextwahl und der Adaptivität der von den Lernenden gewählten Kontexte. Hierfür wurde eine explanative Studie mit Schülerinnen und Schülern aus dem dritten Lernjahr im Fach Chemie an Gymnasien und Gesamtschulen durchgeführt. Relevante Personenmerkmale wurden hier mithilfe von psychometrischen Leistungs- und Persönlichkeitstests erhoben. Chemielernende haben eine selbstgewählte kontextorientierte Aufgabe zum Inhaltsfeld *saure und alkalischen Lösungen* bearbeitet, die in systematisch variierte Kontexte eingebettet war. Die Adaptivität der selbstgewählten Kontexte wurde durch die Erhebung der aufgabenbezogenen Zufriedenheit, des situationalen Interesses und der kognitiven Belastung in den Blick genommen.

Die aus dieser ersten Teilstudie abgeleiteten Implikationen zur Adaptivität der Kontexte wurden genutzt, um die konkrete Ausgestaltung der unterrichtlichen Lehr-Lernsituation im Rahmen der interessenbasierten Differenzierung näher zu untersuchen. Ein besonderer Fokus lag darauf, die Bedeutung der Kontextwahl und der Passung zwischen den Kontexten und den Lernvoraussetzungen der Lernenden für den Lernprozess zu analysieren. Dazu wurde eine experimentelle Studie im Prä-Post-Design in den regulären Chemieunterricht von Chemielernenden des dritten Lernjahres an Gymnasien und Gesamtschulen integriert. Kontrastiert wurden drei Untersuchungsgruppen, die gezielt im Hinblick auf die Kontextwahl und die Adaptivität zwischen Kontext und den Lernvoraussetzungen der Lernenden variiert wurden. Dafür wurde vorab ein Vorhersagemodell zur Zuweisung von passgenauen Kontextaufgaben mithilfe von maschinellem Lernen entwickelt, dessen Prädiktionsgenauigkeit im Projektverlauf mehrfach überprüft wurde. Die Lernenden arbeiteten im Verlauf einer kontextorientierten Lerneinheit an drei sequenzierten Teilaufgaben zum Inhaltsfeld *saure und alkalische Lösungen*, die in systematisch variierte Kontexte integriert wurden. Dabei wurden die aufgabenbezogene Zufriedenheit, das situationale Interesse und die kognitive Belastung lernprozessbegleitend nach jeder Teilaufgabe sowie der erreichte Fachwissenszuwachs am Ende der vollständigen Lerneinheit erfasst.

2 Individuelle Förderung im Unterricht

In der Abschlussempfehlung des Forums Bildung zum Umgang mit den eklatanten Ergebnissen der PISA-Studie wird die individuelle Förderung erstmals an exponierter Stelle als eine von zwölf Empfehlungen zur Reform des Bildungssystems genannt (Forum Bildung, 2002). Als Reaktion auf diese Empfehlung wurde individuelle Förderung der Schülerinnen und Schüler durch einen Großteil der Bundesländer in ihren Schulgesetzen verankert, ohne dass diesen Gesetzestexten ein einheitliches Konzept der individuellen Förderung zugrunde lag (Fischer, 2014). Der fehlende Konsens in Bezug auf ein gültiges Konzept der individuellen Förderung zeigt sich auch im wissenschaftlichen Diskurs (Klieme & Warwas, 2011; Wischer, 2014). Hier wird der Begriff bisweilen sogar als „…(politische[r]) *Containerbegriff* [bezeichnet]…, der unterschiedlichste und nahezu beliebige Konzepte, Handlungsprogramme und normative Orientierungen umfasst“ (Maykus et al. (2011, S. 140) bezeichnet.

Inhaltlich erscheint es sinnvoll, sich einer Definition über die Bedeutung der beiden Begrifflichkeiten zu nähern. Der Begriff *individuell* bezieht sich auf das Individuum, welches im Lateinischen für das Unteilbare steht. Das Individuum steht in der Bildungssprache häufig für den Menschen als Einzelwesen (Ritter, 1976). Der Begriff der *Förderung* bezeichnet im Allgemeinen die Tätigkeit, jemanden zu unterstützen, um Leistung, Anstrengung und Entwicklung anzuregen (Werner, 2011). Aus der Perspektive der Erziehungswissenschaft kann Förderung als zielorientierte und intendierte pädagogische Intervention beschrieben werden, die von einem Zusammenhang zwischen dem pädagogischen Handeln und dem Kompetenzerwerb ausgeht (Maykus et al., 2011). Dieser Prozess ist vielschichtig, umfasst unterschiedliche Bereiche, Lernendengruppen sowie diverse Ziele. So können insbesondere fachspezifische Kompetenzen oder motivationale Aspekte adressiert werden. Weiterhin können spezifische Lernendengruppen, wie Hochbegabte oder Leistungsschwache, gefördert werden. Ebenso ist eine geschlechterspezifische Förderung denkbar (Wischer, 2014). Die Förderung ergibt sich dabei aus den unterschiedlichen Lernvoraussetzungen und der Tatsache, dass nicht alle Schülerinnen und Schüler die Ziele des Unterrichts erreichen. Förderung stellt damit eine Optimierung der schulischen Lernprozesse dar (Arnold & Richert, 2008).

Folglich scheint bereits der Begriff der Förderung die individuellen Voraussetzungen der Lernenden zu fokussieren. Klieme und Warwas (2011, S. 807) merken diesbezüglich an, dass ein erziehungswissenschaftlicher Theoriediskurs über den Begriff der individuellen Förderung obsolet ist, da „...sie keinen spezifischen Begriffsinhalt hat“. Trotz dessen liefert Dumont (2019) in einem Übersichtsartikel zu individueller Förderung eine Definitionsgrundlage, in der verschiedene Definitionen kontrastiert werden. Hierbei arbeitet die Autorin heraus, dass alle Definitionen darin übereinstimmen, die Bedürfnisse aller Schülerinnen und Schüler in den Blick zu nehmen. Weiterführend kann die Berücksichtigung individueller Dispositionen sowie die optimale Unterstützung der Lernenden im Lernprozess als zentrales Merkmal indi-

vidueller Förderung herausgearbeitet werden. Unterschiede liegen hier besonders in der Umsetzung. So konstatiert auch C. Fischer (2014), dass es sich bei individueller Förderung um einen Sammelbegriff handelt, der verschiedene Konzepte und Ansätze zusammenfasst.

Unter Betrachtung der dargelegten Diskussion zum Begriff der individuellen Förderung sollen im Rahmen dieser Arbeit unter individueller Förderung alle Konzepte und Ansätze verstanden werden, die allen Schülerinnen und Schülern unter Berücksichtigung der individuellen Lernvoraussetzungen eine optimale Potenzialentfaltung im Sinne der Persönlichkeits- und Kompetenzentwicklung ermöglichen.

2.1 Differenzierung im Unterricht

Zum Umgang mit Heterogenität und zur individuellen Förderung aller Schülerinnen und Schüler haben sich mittlerweile eine Vielzahl von Ansätzen entwickelt. Diese Ansätze sind aus unterschiedlichen wissenschaftlichen Traditionen hervorgegangen, wodurch sie zum Teil voneinander abweichende Schwerpunkte aufweisen, aber nicht immer klar abgrenzbar sind, sodass verschiedene Ansätze mit gleichen Begrifflichkeiten beschrieben werden (Bohl et al., 2012; Dumont, 2019; C. Fischer, 2014; Vock & Gronostaj, 2017).

Das Konzept des adaptiven Unterrichts (oder der Adaptivität) stammt aus der Lehr-Lernpsychologie und kann als Ursprung verschiedener Konzepte der individuellen Förderung angesehen werden (Bohl et al., 2012; Wischer & Trautmann, 2012). Theoretisch geprägt wurde das Konzept des adaptiven Unterrichts besonders durch den *Aptitude-Treatment-Interaction (ATI)*-Ansatz von Cronbach und Snow (1981). Die zentrale Grundannahme ist hierbei, dass die Wirkung von bestimmten Unterrichtsmaßnahmen (engl. *treatment*) maßgeblich durch die Merkmale der Lernenden (engl. *aptitude*) beeinflusst wird. So wird beispielsweise angenommen, dass Lernende mit ungünstigen Lernvoraussetzungen (z. B. niedrige Intelligenz, geringes Vorwissen) von strukturierten und gesteuerten Lernprozessen profitieren. Andererseits profitieren Lernende mit günstigen Lernvoraussetzungen eher von selbstreguliertem Lernen in einer offenen Lernumgebung (R. E. Snow, 1989).

In Anbetracht dieser Ergebnisse liegt die Vermutung nahe, dass ein sogenannter *one-size-fits-all*-Ansatz, d. h. allen Lernenden das gleiche Lernangebot zu unterbreiten, nur wenig zielführend ist, um alle Lernende in einem Lernprozess angemessen zu adressieren. Adaptiver Unterricht bietet Lernangebote, die an die individuellen Lernvoraussetzungen der Schülerinnen und Schüler anknüpfen (Corno, 2008). In diesem Zusammenhang bezeichnen Hardy et al. (2011) alle Lerngelegenheiten als adaptiv, die zu einer Annäherung von leistungsschwachen und leistungsstarken Lernenden beitragen und zu überdurchschnittlichen Effekten hinsichtlich kognitiver sowie affektiv-motivationaler Lernergebnisse führen. Wirksame Lernprozesse

resultieren dabei aus methodisch-didaktischen Angeboten, die den individuellen Lernvoraussetzungen angepasst sind und die individuell unterschiedlich von den Schülerinnen und Schülern genutzt werden.

Diese Anpassung kann auf zwei unterschiedlichen Ebenen stattfinden (Corno & Snow, 1986). Mikroadaptionen finden während der Lerngelegenheit statt. Sie betreffen vorrangig Anpassungen in der Interaktion zwischen Lehrenden und Lernenden, die meistens eher kurzfristig sind. Weit verbreitet ist etwa eine kontinuierliche Rückmeldung zum erreichten Lernfortschritt im Lernprozess in Form von Feedback (Black & Wiliam, 2009). Daneben können auch Strukturierungshilfen und weitere Unterstützungshandlungen der Lehrenden mikroadaptiv wirken (Corno, 2008; Parsons et al., 2018). Makroadaptionen betreffen dagegen eher übergeordnete Anpassungen des Unterrichts, die zudem langfristig angelegt sind. Darunter werden überwiegend Anpassungen für kleinere Gruppen von Lernenden mit ähnlichen Lernvoraussetzungen gefasst. Eine zentrales makroadaptives Konzept ist die aus der Schulpädagogik bzw. Didaktik stammende Differenzierung (Wischer & Trautmann, 2012).

Der terminologische Ursprung des Begriffs Differenzierung liegt im lateinischen Wort *differe*, was so viel wie „verschieden sein, sich unterscheiden“ bedeutet. In diesem Zusammenhang adressiert der Begriff – wie auch andere Konzepte zur individuellen Förderung – die bereits vielfach aufgeführte Heterogenität der Lernenden. Wegen der unterrichtspraktischen Ausrichtung des Konzeptes liegt der Schwerpunkt nicht auf der möglicherweise illusorischen Annahme, jedem und jeder Lernenden ein individualisiertes Unterrichtsangebot zu machen, sondern zielt auf unterschiedliche Gruppen von Lernenden. Kerngedanke ist die Bereitstellung eines Lernangebots, welches an kleinere Lernendengruppen angepasst ist, die vorher auf Grundlage bestimmter Merkmale (z. B. der Leistungsfähigkeit) zusammengesetzt wurden (Bohl et al., 2012). Mit Bezug auf Arnold und Richert (2008) sei angemerkt, dass ein derartiges Angebot zwar adaptiv für die jeweilige Lernendengruppe sein kann, aber nicht zwangsläufig für jeden Schüler oder jede Schülerin dieser Gruppe.

Die Lernenden können auf verschiedenen Ebenen differenziert werden, was durch die äußere und innere Differenzierung beschrieben wird. Beide Begriffe gehen im Wesentlichen auf Klafki und Stöcker (1976) zurück, die das Konzept besonders in den 1970er-Jahren publik machten. Die Autoren definieren wie folgt:

> „Innere Differenzierung meint dabei all jene Differenzierungsformen, die *innerhalb* einer gemeinsam unterrichteten Klasse oder Lerngruppe vorgenommen werden, im Unterschied zu allen Formen sog. *äußerer* Differenzierung, in der Schülerpopulationen nach irgendwelchen Gliederungs- oder Auswahlkriterien – z. B. den Gesichtspunkten unterschiedlichen Leistungsniveaus oder unterschiedlicher Interessen – in Gruppen aufgeteilt werden, die räumlich getrennt und von verschiedenen Personen bzw. zu verschiedenen Zeiten unterrichtet werden.“ (Klafki & Stöcker, 1976, S. 497)

Maßnahmen der äußeren Differenzierung betreffen damit vor allem die Verteilung der Lernenden auf unterschiedliche Schulformen, aber auch die vorzeitige Einschulung oder Zurückstellungen zählen zu äußerer Differenzierung. Innerhalb der Schule findet äußere Differenzierung durch Wahl(pflicht)kurse (z. B. Naturwissenschaften oder weitere Fremdsprachen), der Fachleistungsdifferenzierung in G- und E-Kurse an Gesamtschulen oder der Wahl von Grund- und Leistungskursen in der gymnasialen Oberstufe statt (Vock & Gronostaj, 2017). Die innere Differenzierung (auch Binnendifferenzierung) zielt dagegen auf die Berücksichtigung der Bedarfe von Lernenden innerhalb der Lerngruppe, indem die Lernenden anhand vorab definierter Kriterien in kleinere Gruppen aufgeteilt werden, denen dann passende Lernangebote unterbreitet werden (Trautmann & Wischer, 2009)

Im internationalen Forschungsraum findet man das Konzept der inneren Differenzierung unter Stichwörtern wie *within-class grouping* oder *ability grouping*, wobei letzteres als leistungsbezogene Differenzierung zu verstehen ist. Die Gruppierung kann hier so organisiert werden, dass die Gruppen hinsichtlich ausgewählter Merkmale gleich (*homogenous*) oder verschieden sind (*heterogeneous*). Es bestehen auch gewisse Analogien zum Konzept der *differentiated instruction*, welches maßgeblich von Tomlinson (z. B. 1999) geprägt wurde. Das Konzept ist allerdings nicht nur auf die Anpassung der Lernangebote für Lernendengruppen beschränkt, sondern nimmt explizit auch die Bedürfnisse einzelner in den Blick (Tomlinson et al., 2003). Damit kann es eher als ein Oberbegriff beschrieben werden, der verschiedene Ansätze individueller Förderung umfasst.

In der Literatur sind zahlreiche Möglichkeiten zur Differenzierung beschrieben. Im Wesentlichen lassen sich diese Möglichkeiten auf die Gruppierung von Lernenden im Hinblick auf bestimmte Lernvoraussetzungen und die darauf abgestimmte Anpassung des Unterrichtsangebots reduzieren.

Tomlinson (1999) nennt die Leistungsfähigkeit, das Interesse sowie den Lernstil als mögliche Lernvoraussetzungen, die zur Gruppierung genutzt werden können. Der Lernstil ist hierbei die bevorzugte Lernweise des Lernenden (z. B. visuell, auditiv oder motorisch-kinästhetisch). Meyer (2011) benennt die (vermutete) Leistungsfähigkeit, die diagnostizierte Kompetenzstufe, den individuellen Förderbedarf, das Geschlecht sowie das Interesse als weitere personenbezogene Kriterien. Hierbei nimmt die Diagnose der Lernbedürfnisse der Schülerinnen und Schüler einen zentralen Stellenwert ein. Die Diagnose der affektiv-motivationalen Lernvoraussetzungen fällt den Lehrkräften deutlich schwerer als die der leistungsbezogenen Lernvoraussetzungen (Karing, 2009). Aus diesem Grund können beispielsweise auch standardisierte Testverfahren zum Einsatz kommen, die objektivere Urteile zulassen (Hardy et al., 2019; Wagner, 2016). Die Diagnose von leistungsbezogenen Lernvoraussetzungen wurde zur Differenzierung schon häufiger mithilfe von standardisierten Instrumente durchgeführt (z. B. Anus, 2015). Dagegen stehen für die Diagnose von affektiv-motivationalen Lernvoraussetzungen im Unterricht bislang keine standardisierten Instrumente zur Verfügung. Demzufolge

wird bei der Differenzierung der Lerngruppe anhand affektiv-motivationaler Lernvoraussetzungen in der Regel keine standardisierte Diagnose der Lernvoraussetzungen durchgeführt. Stattdessen werden den Lernenden häufig mehrere Aufgaben angeboten, aus denen sie auswählen können (siehe Kapitel 3).

Die Anpassung des Unterrichtsangebots wird auch als didaktische Differenzierung bezeichnet und umfasst die Auswahl passender Inhalte, Methoden und Ziele (Meyer, 2011). Diese Unterteilung findet sich unter den Begriffen *content*, *process* und *product* auch bei Tomlinson (1999) wieder. Weitere Möglichkeiten zur Differenzierung des Unterrichtsangebots finden sich in der Literatur (z. B. bei Brüning & Saum, 2010; Klafki & Stöcker, 1976), lassen sich aber meist einem der drei Oberbegriffe zuordnen.

Zunächst können unterschiedliche Inhalte für die Lernenden eingesetzt werden. Hierbei können Inhalte genutzt werden, die hinsichtlich ihrer Komplexität (bzw. dem Anforderungsniveau) an die Leistungsfähigkeit der Lernenden adaptiert sind (Brüning & Saum, 2010; Klafki & Stöcker, 1976). Dies wird zumeist unter leistungsbezogener Differenzierung bzw. ability grouping gefasst. Weiterhin können Inhalte in Bezug auf das Thema variiert werden, um unterschiedliche Interessen der Lernenden anzusprechen (Meyer, 2011). Im Sprachunterricht können die Lernenden beispielsweise aus unterschiedlichen Lesetexten wählen, die zu ihren Interessen passen (z. B. Morgan & Wagner, 2013). Aber auch im mathematisch-naturwissenschaftlichen Unterricht können fachliche Konzepte mit unterschiedlichen Anwendungsbezügen aus dem alltäglichen Leben der Schülerinnen und Schüler verknüpft werden, um unterschiedliche Interessen zu adressieren (Gijsbers et al., 2020). Ein derartiger Ansatz ist als interessenbasierte Differenzierung zu verstehen.

Inhalte können methodisch-didaktisch unterschiedlich vermittelt werden. Unter Berücksichtigung unterschiedlicher Lernprofile können hier variierende Formen kooperativer Arbeit, unterschiedliche Lerntempi oder auch weitere methodische Zugänge angeboten werden (Tomlinson, 1999). In diesem Zusammenhang ist auch die Öffnung des Unterrichts als zentraler Baustein zu nennen. Hierbei werden zwei gegensätzliche Pole für die Differenzierung unterschieden. Nach Heymann (2010) wird das Lernen in geschlossenen Differenzierungsformen so organisiert, dass den Lernenden ein passendes Lernangebot zugewiesen wird. In offenen Differenzierungsformen können Lernende dagegen selbstständig ihren Lernweg und ein passendes Lernangebot auswählen.

Letztlich können die angepassten Inhalte und deren methodisch-didaktische Umsetzung in divergente Lernziele für die Schülerinnen und Schüler resultieren. Besonders die Verbindlichkeit der Curricula begrenzt allerdings die Anpassung der Lernziele (Riedel, 2008). In diesem Zusammenhang bietet die Ausdifferenzierung der Bildungsstandards in Mindest-, Regel- und Optimalstandards einen ersten Anhaltspunkt, um angepasste Lernziele festzusetzen (Vock & Gronostaj, 2017). So können die Lernenden beispielsweise unterschiedliche Lernprodukte am Ende des Lernprozesses hervorbringen (Tomlinson, 1999).

2.2 Forschungslage zur Differenzierung im Unterricht

Die Ergebnisse der PISA-Studie aus dem Jahr 2015 verweisen darauf, dass es einen Zusammenhang zwischen der Testleistung der Schülerinnen und Schüler und der Häufigkeit des Einsatzes adaptiver Maßnahmen im Unterricht gibt (OECD, 2016). Aufgrund der Anlage der Befragung können diese Effekte allerdings nicht eindeutig auf ein bestimmtes Konzept individueller Förderung zurückgeführt werden. Konkrete Aussagen lassen sich aus der viel beachteten (Meta-)Metaanalyse von Hattie (2012) zum Einfluss unterschiedlicher Merkmale auf den Lernerfolg ableiten. Hier zeigen sich klare Tendenzen zugunsten von Unterrichtsanpassungen, welche die Lehrenden-Lernenden-Interaktion im Sinne von Mikroadaptionen betreffen. So weisen insbesondere die formative Evaluation ($d = 0.90$) und Feedback ($d = 0.73$) eine der höchsten Effektstärken abseits der Merkmale der Lernenden (z. B. Selbsteinschätzung des eigenen Leistungsniveaus) auf. Makroadaptive Maßnahmen wie die interne Differenzierung ($d = 0.16$) zeigen eher kleine Effekte.

Die Effektivität von Differenzierungsmaßnahmen wurde vorrangig im Zuge des within-class grouping und ability grouping untersucht (Dumont, 2019). Folglich betrachten die meisten Studien die Auswirkungen eines leistungsgestützten Differenzierungsansatzes. Die Forschungslage ist in diesem Zusammenhang als äußerst diffus zu bezeichnen, da viele Studien konträre Ergebnisse zeigen.

Zahlreiche Untersuchungen deuten auf positive Effekte zugunsten einer homogenen Gruppierung im Vergleich zur heterogenen Gruppierung hin, wobei die Anpassung des Unterrichtsangebots an die Bedürfnisse der unterschiedlichen Gruppen als entscheidender Faktor bei der homogenen Gruppierung identifiziert werden kann (Lou et al., 1996; Slavin, 1986; Kulik & Kulik, 1982; Smale-Jacobse et al., 2019). Eine aktuellere Metaanalyse von Steenbergen-Hu et al. (2016) fasst – ähnlich wie Hattie – die Ergebnisse von fünf Metaanalysen zusammen und kommt zum Schluss, dass die innere Differenzierung der Lerngruppe in homogene Teilgruppen anhand der Leistungsfähigkeit einen kleinen, aber statistisch signifikanten Effekt auf die Lernleistung hat. Diese Effekte zeigen sich konsistent über den Primar- und Sekundarstufenbereich in älteren und aktuelleren Metaanalysen (Kulik & Kulik, 1982, 1992; Slavin, 1987; Lou et al., 1996; Smale-Jacobse et al., 2019). Auffällig ist hierbei allerdings, dass die Effektstärken der einzelnen Untersuchungen, die in die Metaanalyse einflossen, zum Teil stark divergente Effektstärken zeigen. So weisen die vier von Smale-Jacobse et al. (2019) ausgewählten Studien zur homogenen Differenzierung kleine bis große Effektstärken ($0.28 \leq d \leq 4.50$) auf. Die Studien mit den höchsten Effektstärken nutzten neben der Leistungsfähigkeit auch den bevorzugten Lernstil (Bal, 2016) und die dominante Intelligenzdomäne (Altıntaş & Özdemir, 2015a, 2015b; siehe auch Gardner, 2016 für die Theorie der multiplen Intelligenzen) zur Differenzierung, weshalb eine eindeutige Kausalinterpretation der Studien problematisch ist.

Innerhalb der Untersuchungen zur leistungsbasierten Differenzierung deuten sich Interaktionseffekte des Differenzierungsansatzes mit der Gruppenzusammensetzung und der Leistungsfähigkeit auf die Lernleistung der Lernenden an. Lou et al. (1996) geben hier an, dass Lernende mit einer geringen Leistungsfähigkeit eher von heterogenen Gruppierungen profitieren, während Lernende mit einer mittleren Leistungsfähigkeit von homogenen Gruppierungen profitieren. Für das Lernen von leistungsstarken Schülerinnen und Schülern ist die Gruppenzusammensetzung irrelevant. Parallel belegen die von Lou et al. (1996) berücksichtigten Studien eine höhere Effektivität von leistungsbasierten Differenzierungsmaßnahmen für leistungsschwache Lernende ($d = 0.37$) im Vergleich zu Lernenden mit mittlerer ($d = 0.19$) und hoher Leistungsfähigkeit ($d = 0.28$). Dagegen indizieren die von Kulik und Kulik (1992) untersuchten Studien, dass leistungsstarke ($d = 0.30$) hinsichtlich der Lernleistung mehr als durchschnittliche ($d = 0.18$) und leistungsschwache ($d = 0.16$) Schülerinnen und Schüler von leistungsbasierter Differenzierung in homogenen Lernendengruppen profitieren. Steenbergen-Hu et al. (2016) konkludieren sogar, dass die Wirksamkeit leistungsbasierter Differenzierungsmaßnahmen auf die Lernleistung unabhängig vom Leistungsniveau der Schülerinnen und Schüler ist.

Im Gegensatz zu potenziellen Leistungseffekten werden die Auswirkungen von leistungsbasierter Differenzierung auf affektive Merkmale seltener berücksichtigt (Smale-Jacobse et al., 2019). Hier gibt es allerdings Hinweise auf positive Effekte der leistungsbasierten Differenzierung in Bezug auf die Einstellung gegenüber dem Unterrichtsfach. Das Selbstkonzept der Lernenden wird dagegen nicht beeinflusst (Kulik & Kulik, 1982; Lou et al., 1996).

Zu interessengestützten Differenzierungsmaßnahmen finden sich bislang keine Metaanalysen, welche die Ergebnisse mehrerer Studien aggregieren. Demzufolge erwähnt auch Hattie (2012) keine Anpassungen im Unterricht gemäß unterschiedlicher Interessen der Lernenden. Im Bereich der Sprachdidaktik wurden einzelne Studien durchgeführt, die unterschiedliche Interessen der Lernenden durch die Bereitstellung von unterschiedlichen Lesetexten berücksichtigen. Diesen Studien kann meistens allerdings kein klarer Ansatz der individuellen Förderung im Sinne der Individualisierung oder Differenzierung zugeordnet werden. Im Gegensatz zu den primär quantitativ angelegten Studien, die in den Metaanalysen zur leistungsbasierten Differenzierung berücksichtigt werden, haben viele Studien hier einen qualitativen Forschungszugang. Morgan und Wagner (2013) haben eine dreiwöchige Unterrichtseinheit an einer High-School in den USA durchgeführt, in der die Lernenden eigene Bücher zum Lesen auswählen konnten. Durch Gespräche mit den Lernenden kamen Morgan und Wagner zu dem Schluss, dass die Lernenden durch die selbstständige Wahl engagierter beim Lesen waren. Die Untersuchungen von Ivey und Johnston (2013) sowie Allred und Cena (2020) berichten von ähnlichen Effekten durch selbst gewählte Lesematerialien im Englischunterricht aufgrund von Interviews. Guthrie et al. (2007) haben in einer Mixed-Methods-Studie bei Grundschulkindern untersucht, welche Effekte eine zwölfwöchige Leseintervention hat, bei

der die Lernenden unter anderem auch Bücher zum Lesen auswählen konnten. Die quantitative Datenanalyse zeigt hierbei, dass ungefähr 22 % der Varianz in der Zunahme des Leseverständnisses durch die Wahl der Lesetexte aufgeklärt werden können. Gijsbers et al. (2020) implementierten einen interessenbasierten Differenzierungsansatz in den Mathematikunterricht von Lernenden der Sekundarstufe II. Die Lernenden hatten während der Intervention die Möglichkeit, Mathematikaufgaben auszuwählen und zu bearbeiten, die in drei unterschiedliche Anwendungsbezüge von Mathematik (Wissenschaft, Biomedizin und Sozioökonomie) eingebettet waren. Die Autorinnen und Autoren schlussfolgern, dass die Wahl der unterschiedlichen Anwendungsbezüge beim Mathematiklernen die wahrgenommene Relevanz von Mathematik bei Lernenden positiv beeinflusst, insbesondere bei Lernenden, die Mathematik bisher als eher unwichtig wahrgenommen haben. Bei der Interpretation dieses Ergebnisses sollte jedoch die fehlende Kontrollgruppe berücksichtigt werden.

Die methodisch-didaktische Anpassung des Unterrichts unter Berücksichtigung individueller Lernendenmerkmale wurde besonders im Kontext der ATI-Forschung untersucht (Cronbach & Snow, 1981). Hattie (2012) schreibt der Anpassung der Unterrichtsmethoden im Hinblick auf die Merkmale der Lernenden nur einen geringen Effekt zu (d = .19). Diese Effektstärke kann allerdings nicht ohne weitere Einschränkungen interpretiert werden, da die berücksichtigten Studien zum Teil methodische Mängel aufweisen. In diesem Zusammenhang schreiben Corno und Snow (1977, S. 492): „…no Aptitude x Treatment interactions are so well confirmed that they can be used directly as guides to instruction". Daran anschließend zeigt die Berücksichtigung der bevorzugten Lernstile durch die methodisch-didaktische Anpassung des Unterrichts laut Hattie (2012) eine höhere Effektstärke (d = .41), die allerdings auf Metaanalysen basiert, die mit ähnlichen Kritikpunkten konfrontiert sind. Die von Hattie (2012) berücksichtigten Metaanalysen enthalten keine Interventionsstudien, die sich an Lernstilen ausrichten. Hattie (2012) verweist darauf, dass die Metaanalyse von Dunn et al. (1995) Studien umfasst, die Fehler in der statistischen Auswertung aufweisen. Folglich ist auch die auf Grundlage dieser Studien von Hattie (2012) berechnete Effektstärke nicht zweifelsfrei interpretierbar.

Bislang finden sich nur wenige Studien, welche die Effektivität von Differenzierungsmaßnahmen explizit im Fach Chemie untersuchen. In den wenigen Studien, die bisher durchgeführt wurden, besteht ein Fokus auf leistungsgestützte Maßnahmen der individuellen Förderung, die sich teilweise nicht eindeutig der Individualisierung oder Differenzierung zuordnen lassen.

Anus (2015) evaluierte für den Chemieunterricht eine individuelle Fördermaßnahme durch den Einsatz adaptiver Lernmaterialien. Hierbei diente ein Multiple-Choice-Test als Diagnoseinstrument zur Erfassung des Fachwissens. Darauf aufbauend erhielt eine Untersuchungsgruppe eine individuelle Aufgabenabfolge entsprechend des diagnostizierten Förderbedarfs. Die anderen beiden Untersuchungsgruppen erhielten eine zufällige Aufgabenabfolge oder

eine sachlogische Abfolge mit steigendem Schwierigkeitsgrad. Es zeigten sich keine Haupteffekte der variierten Aufgabenabfolge, allerdings konnten signifikante Interaktionseffekte zwischen den Gruppen und den kognitiven Leistungsniveaus der Lernenden identifiziert werden. Kognitiv leistungsstarke und leistungsschwache Lernende profitieren von der diagnosegestützten Aufgabensortierung und kognitiv durchschnittliche Lernende erreichten höhere Lernzuwächse durch die sachlogisch strukturierte oder zufällige Aufgabenabfolge. Wie die Autorin selbst anmerkt, sind diese Interaktionseffekte nur schwierig interpretierbar.

In einer weiteren Studie wurde die Effektivität des Einsatzes eines Selbsteinschätzungsbogens mit anschließender Bearbeitung differenzierter Aufgaben untersucht (Kallweit & Melle, 2017). Dabei wurden in einer experimentellen Feldstudie zwei Gruppen miteinander verglichen: Die Interventionsgruppe erhielt nach Verwendung des Selbsteinschätzungsbogens passende Übungsaufgaben, wohingegen die Kontrollgruppe selbstständig Aufgaben zur Bearbeitung auswählte. Die Interventionsgruppe profitierte hinsichtlich des Fachwissenszuwachses und der Beurteilung des Unterrichts signifikant von der Arbeit mit dem Selbsteinschätzungsbogen.

Hauerstein (2019) untersuchte in ihrem Dissertationsprojekt, welche Auswirkungen Strukturierung und Differenzierung durch eine Lernleiter im Chemieunterricht der Sekundarstufe I haben. Die Lernleiter wurde innerhalb der Untersuchung als Strukturierungshilfe genutzt, die zudem die Implementation von differenziertem Material erlaubt, welches hinsichtlich des Anforderungsniveaus systematisch variiert wurde. Die Ergebnisse verweisen auf keine signifikanten Effekte der Differenzierung – weder auf den Lernzuwachs noch andere affektive Variablen wie das situationale Interesse. Dieses Ergebnis wird von der Autorin auf den geringen zeitlichen Umfang der differenzierten Übungsphase zurückgeführt, die während der gesamten Intervention von etwa 12 Einzelstunden nur dreimal stattfand.

2.3 Kritische Würdigung der Forschungslage

Die individuelle Förderung der Lernenden kann im Unterricht durch verschiedene Konzepte realisiert werden, für die bis dato auch eine Vielzahl an empirischen Ergebnissen zur Wirksamkeit vorliegen. Die Forschung zur individuellen Förderung kann dabei als ähnlich diffus bezeichnet werden wie schon der theoretische Diskurs. Aufgrund der vielen unterschiedlichen Konzepte zur Umsetzung von individueller Förderung im Unterricht – die teilweise synonym verwendet werden – wird die systematische Einordnung der Befunde erheblich erschwert (Bohl et al., 2012). Damit verbunden sind ungleiche Operationalisierungen in den Studien und divergente Forschungszugänge (Smale-Jacobse et al., 2019; Trautmann & Wischer, 2007). So kann insgesamt die Forschungslage zur Effektivität von individueller Förderung – trotz einer relativ hohen Anzahl von Studien – als unzureichend bezeichnet werden. Als Kon-

sens kann herausgestellt werden, dass die Berücksichtigung individueller Lernvoraussetzungen durch die Anpassung des Unterrichts grundsätzlich lernförderlich ist. Hierbei sind Mikroadaptionen wie Feedback besonders effektiv (Hattie, 2009). Die Datenlage in Bezug auf makroadaptive Maßnahmen ist als vielschichtiger zu bezeichnen, wobei auch aufgrund qualitativ hochwertiger Metaanalysen, die zudem randomisiert kontrollierte Studien einschließen, festgehalten werden kann, dass durch Individualisierung und Differenzierung nur kleine Effekte zu erwarten sind (z. B. Hattie, 2012; Smale-Jacobse et al., 2019; Steenbergen-Hu et al., 2016). Aufgrund der Vielzahl von Studien bietet sich quantitative Aggregation der Ergebnisse in Metaanalysen zwar an, allerdings erschwert dieser Zugang zugleich die finale Interpretation der gefundenen Effektstärken. Teilweise erfordert die Vielschichtigkeit des Differenzierungsbegriffs die dezidierte Analyse der Untersuchungen und insbesondere deren Methodik, um die Effekte zu interpretieren. Dies ist nicht in jeder Metaanalyse der Fall. Vor diesem Hintergrund sind auch teilweise konträre Effekte einzuordnen.

Mit Bezug auf die dargelegte Evidenz zur Wirksamkeit von individueller Förderung im Chemieunterricht sei auf die geringe Anzahl an Studien verwiesen, welche die abschließende Beurteilung erheblich erschwert. Diese Problematik verschärft sich infolge der unterschiedlichen Konzepte der individuellen Förderung, die in den unterschiedlichen Untersuchungen zugrunde gelegt wurden und den vermutlich daraus resultierenden konträren Ergebnissen.

Obwohl die Berücksichtigung unterschiedlicher Lernvoraussetzungen (z. B. Lernstile oder Interesse bei Tomlinson, 1999) in der Literatur ausführlich beschrieben ist, berücksichtigen Forschungsarbeiten zur individuellen Förderung im Unterricht – auch für den Chemieunterricht – fast ausschließlich die Anpassung des Unterrichts hinsichtlich unterschiedlicher Leistungsniveaus. Zwar werden in manchen Untersuchungen neben der Leistungsfähigkeit der Lernenden auch andere Merkmale zur individuellen Förderung genutzt, allerdings lassen sich potenzielle Effekte dieser Maßnahmen aufgrund mangelnder interner Validität der Untersuchung kaum abschätzen. Für interessenbasierte Differenzierungsmaßnahmen stehen lediglich Einzelstudien zur Verfügung, die häufig einen qualitativen Forschungszugang haben. Insbesondere für das Fach Chemie fehlt es gänzlich an Untersuchungen, wodurch die Effektivität interessenbasierter Differenzierung kaum beurteilt werden kann. Aus diesem Grund widmet sich die vorliegende Arbeit dieser Forschungslücke.

3 Interessenbasierte Differenzierung: Aufgabenwahl als Differenzierungsansatz

Im Laufe der menschlichen Ontogenese entwickelt jeder Mensch eine eigene Persönlichkeitsstruktur, deren zentraler Kern die menschliche Identität repräsentiert. Bei der Entwicklung dieser Persönlichkeitsstruktur bilden sich individuelle Interessen aus, d. h. es bilden sich individuelle Bewertungspräferenzen hinsichtlich unterschiedlicher Objekte, Sachverhalte oder Ereignisse heraus. So ist anzunehmen, dass am Ende dieses Entwicklungsprozesses interindividuell unterschiedlich ausgeprägte affektiv-motivationale Lernvoraussetzungen bestehen, die auch im Hinblick auf schulische Bildungsprozesse relevant sind (Krapp, 1992b, 2002).

Die Heterogenität affektiv-motivationaler Merkmale im Unterrichtskontext wird von Lazarides und Schiepe-Tiska (2022) beschrieben. In ihrer Publikation differenzieren die Autorinnen zwischen affektiv-motivationaler Heterogenität auf Ebene der Person und auf der Ebene der Interaktion zwischen Person und Situation. Auf Personenebene äußern sich interindividuell unterschiedliche affektiv-motivationale Lernvoraussetzungen insbesondere in Unterschieden in der Ausprägung der fachspezifischen Motivation. Naturwissenschaftliche Unterrichtsfächer – mit Ausnahme des Faches Biologie – motivieren im Durchschnitt weniger als andere Fächer (Merzyn, 2013). Auf der Ebene der Interaktion zwischen Person und Situation steht im Zentrum, inwieweit unterschiedliche motivationale Erlebensqualitäten in Bezug auf spezifische Lernsituationen oder Themen bestehen (Lazarides & Schiepe-Tiska, 2022). In verschiedenen Untersuchungen konnte gezeigt werden, dass sich das Interesse an typischen naturwissenschaftlichen Tätigkeiten zwischen verschiedenen Lernendengruppen unterscheidet (Blankenburg et al., 2016; Dierks et al., 2016). Bedeutende Unterschiede zwischen Lernendengruppen zeigen sich auch hinsichtlich der motivationalen Erlebensqualität naturwissenschaftlicher Themen und Kontexte (z. B. Elster, 2007; Schiepe-Tiska, Simm & Schmidtner, 2016). Vor dem Hintergrund dieser Ergebnisse scheint es unabdingbar, diese Unterschiede zu berücksichtigen, um jeden Lernenden optimal im Lernprozess zu fördern.

In der Literatur werden interindividuell unterschiedliche affektiv-motivationale Lernvoraussetzungen häufig durch die Öffnung des Unterrichts im Hinblick auf die Auswahl der bereitgestellten Lernmaterialien berücksichtigt. So können die Lernenden im Sprachunterricht Lesetexte mit unterschiedlichen Themen (z. B. Morgan & Wagner, 2013) oder im Mathematikunterricht Aufgaben mit unterschiedlichen Anwendungsbezügen zur Bearbeitung auswählen, die zu ihren Interessen passen (z. B. Gijsbers et al., 2020). Deswegen werden interessenbasierte Differenzierungsmaßnahmen im Rahmen dieser Arbeit als interessenbasierte Aufgabenwahlen betrachtet, bei denen die Lernenden aus einer begrenzten Anzahl an Aufgaben wählen können, die hinsichtlich des Anwendungsbezugs variiert wurden, aber deren Anforderungsniveau nicht verändert wurde (siehe Kapitel 4 zum kontextorientierten Lernen).

Gleichwohl es bislang an Studien mangelt, welche die Effekte interessenbasierter Differenzierungsmaßnahmen in den Blick nehmen, wurden die Auswirkungen der Bereitstellung von

Wahloptionen aus der Perspektive der Lehr-Lernpsychologie schon mehrfach empirisch untersucht. Diese Forschungsarbeiten können als erste Annäherung verstanden werden, um die Auswirkungen interessenbasierter Wahlen im Chemieunterricht abzuschätzen.

Bevor ausgewählte Befunde zur Bereitstellung von Wahloptionen vorgestellt werden, sollen interessengeleitete Wahlentscheidungen zunächst aus theoretischer Perspektive betrachtet werden, um potenzielle Effekte zu verorten.

3.1 Theoretische Grundlagen

3.1.1 Die Selbstbestimmungstheorie der Motivation

Die Selbstbestimmungstheorie (engl. *Self-determination theory,* SDT) von Deci und Ryan (1985) lässt sich als eine der einflussreichsten Theorien zur Erklärung der menschlichen Motivation bezeichnen. Dem Erleben vollkommener Selbstbestimmung werden zahlreiche positive Effekte zugeschrieben, die durch eine Vielzahl von Studienbelegen aus unterschiedlichen Disziplinen empirisch gestützt werden. Dazu gehören eine gesteigerte intrinsische Motivation und Leistungsfähigkeit sowie positive Auswirkungen auf die psychische Gesundheit (Ryan et al., 2021).

Die Konzeptualisierung der Selbstbestimmung in der SDT ist maßgeblich durch die Arbeiten von Heider (1958) und de Charms (1968) geprägt. Heider vertritt die Annahme, dass menschliches Verhalten durch den *perceived locus of causality* (PLOC) erklärbar ist. Entscheidend ist hier die Überzeugung, für den Ausgang eines Ereignisses verantwortlich zu sein (*personally caused*) oder am Ergebnis unbeteiligt zu sein (*impersonally caused*). Diese Kausalitätszuschreibung ist von dem Nachweis einer Handlungsintention abhängig, die sich als Ausdruck von Anstrengung und dem Einsatz von Fähigkeiten interpretieren lässt. Das Konzept von Heider wurde von de Charms weiter ausdifferenziert, welcher postulierte, dass nicht jede intentionale Handlung zwangsläufig freiwillig ausgeführt wird. Nach de Charms haben Handlungen einen *external perceived locus of causality (EPLOC)*, wenn externer Druck oder äußere Anreize handlungswirksam sind und bestimmte Verhaltensweisen als aufgezwungen erlebt werden. Konträr dazu besitzen Handlungen, die willentlich ausgeführt werden und bei denen das sogenannte *Selbst* der Ursprung der Handlung ist, einen *internal perceived locus of causality* (IPLOC) (Ryan & Deci, 2017). Derartige Handlungen können als vollkommen selbstbestimmt angesehen werden, da sie vollständig durch das Selbst organisiert sind (Ryan & Deci, 2006; Ryan et al., 2021).

Gleichwohl das Selbst damit eine zentrale Stellung in der SDT einnimmt, wird nicht unmittelbar deutlich, was unter dem Selbst zu verstehen ist. Ryan und Deci (z. B. 2017) merken diesbezüglich an, dass das Selbst nicht mit der eigentlichen Person gleichgesetzt werden kann. So kann beispielsweise auch innerpsychischer Druck innerhalb der Person als kontrollierend und damit nicht als Teil des Selbst erlebt werden. Das Selbst ist „…not primarily an object of

perception or evaluation but, rather, is phenomenally accessed as the sense of activity in contacting, relating, assimilating, constructing, and caring in the world.“ (Ryan & Deci, 2017, S. 53). Die Autoren schlagen diesbezüglich vor, nicht das Selbst als solches zu betrachten, sondern eher unterschiedliche Verhaltensweisen als Ausdruck des Selbst zu kontrastieren, um ein besseres Verständnis für das Selbst zu entwickeln.

In diesem Zusammenhang können unterschiedliche Arten von Motivation als Verhaltensweisen beschrieben werden, die sich als Ausdruck des Selbst manifestieren. Sie unterscheiden sich insbesondere im Hinblick auf ihre Verhaltensregulierung, dem perceived locus of causality sowie dem Grad der Selbstbestimmung (Ryan & Deci, 2000; Ryan et al., 2021). Die von Deci und Ryan entwickelte Taxonomie der Motivation ist in Abbildung 1 dargestellt.

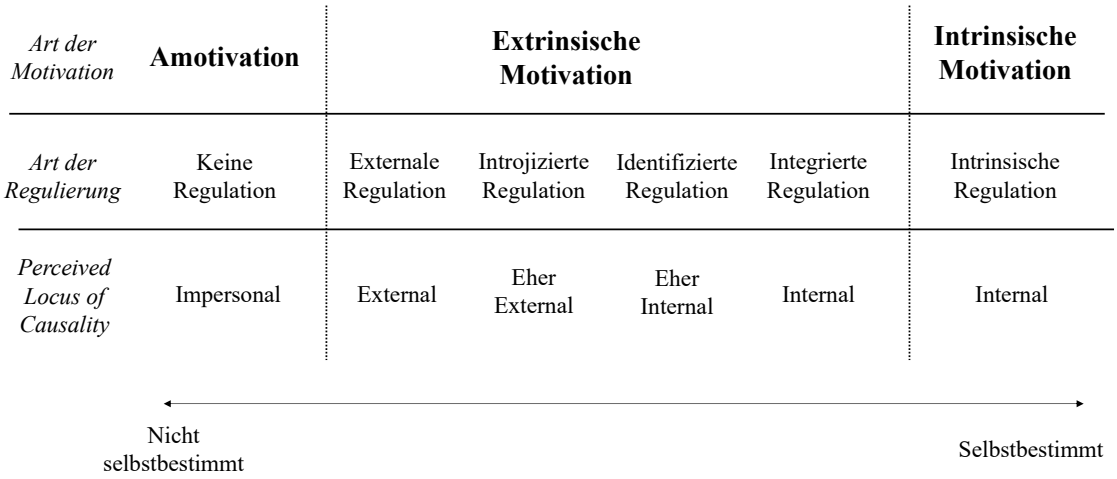


Abbildung 1: Das Selbstbestimmungs-Kontinuum mit der Art der Motivation, der Regulierung und dem perceived locus of causality (in Anlehnung an Ryan & Deci, 2000)

Am linken Ende des Kontinuums befindet sich der Zustand der *Amotivation*. Dieser Zustand bezieht sich auf die vollkommene Abwesenheit von Intentionalität. Darunter fallen häufig Verhaltensweisen, die kein erkennbares Ziel verfolgen und sich als Ausdruck von fehlendem Interesse, Kompetenz und Anstrengung beschreiben lassen (Deci & Ryan, 1993; Ryan & Deci, 2020). Die Attribuierung des Verhaltens ist *impersonally caused* (Heider, 1958). Amotiviertes Verhalten ist damit durch einen vollständigen Mangel an Selbstbestimmung gekennzeichnet (Deci & Ryan, 2000).

Alle anderen Punkte auf dem Selbstbestimmungskontinuum lassen sich als motiviert bezeichnen (Ryan & Deci, 2002). Motiviertes Verhalten ist grundsätzlich durch ein hohes Maß an Intentionalität gekennzeichnet, d. h. es ist zielgerichtet (Deci & Ryan, 1993; Heider, 1958; Rheinberg & Vollmeyer, 2012). Wie auch andere Autoren (u. a. Rheinberg & Vollmeyer, 2012), differenzieren Deci und Ryan zwischen intrinsischer und extrinsischer Motivation. In

extrinsisch motivierten Verhaltensweisen werden äußere Handlungsanreize wirksam. Derartige Verhaltensweisen werden aus anderen Gründen als der inneren Bedürfnisbefriedigung ausgeführt, zielen aber auf das Erreichen einer bestimmten Handlungskonsequenz ab (Ryan & Deci, 2020). Nach Deci und Ryan können Inhalt und Eigenschaften extrinsischer Verhaltensweisen stark variieren, lassen sich aber im Wesentlichen auf vier unterschiedliche Subkategorien reduzieren. Verhaltensweisen, die sich dem Typ der *externalen Regulation* zuordnen lassen, können als der Inbegriff extrinsischer Motivation beschrieben werden und sind damit vollkommen durch externe Zwänge getrieben. Im Mittelpunkt steht hierbei der Erhalt einer Belohnung oder das Entgehen einer Strafe (Deci & Ryan, 1993). Diese Art der Verhaltensregulierung kann durch einen external perceived locus of causality beschrieben werden (Ryan & Deci, 2017). *Introjiziert regulierte* Verhaltensweisen erfolgen nicht aufgrund eines äußeren, sondern eines inneren Drucks. Solche Verhaltensweisen sind häufig mit einer Erhöhung des Selbstwertes oder mit Androhung von Schamgefühlen verbunden. Der perceived locus of causality von introjiziert reguliertem Verhalten ist nach wie vor als external einzustufen, da der Ursprung der Handlung nicht aus dem Selbst hervorgeht, sondern fremdbestimmt ist (Deci & Ryan, 2000; Ryan & Deci, 2017). Kommt es zur Assimilation von bestimmten Verhaltensweisen mit dem Selbst, werden jene als *identifiziert reguliert* bezeichnet. „Identifications are defined by a conscious endorsement of values and regulations" (Ryan & Deci, 2017, S. 187). Die bewusste Annahme von Regeln und Werten drückt sich in einer höheren Selbstbestimmung aus und einem stärkeren internal perceived locus of causality. Gleichwohl derartige Verhaltensweisen relativ selbstbestimmt sind, lassen sie sich als extrinsisch bezeichnen, da der Handlungsanreiz nach wie vor in der Überzeugung begründet ist, dass sie für ein bestimmtes Ziel lohnend sind (Ryan & Deci, 2017; Ryan et al., 2021). Die autonomste Form der extrinsischen Motivation bezeichnen Deci und Ryan als *integrierte Regulation*. In das Selbst integrierte Verhaltensweisen oder Werte stehen in vollständiger Kongruenz mit dem Selbst und werden folglich akzeptiert (Deci & Ryan, 2000; Ryan & Deci, 2017). Trotz der Integration bestimmter Werte oder Verhaltensweisen in das Selbst liegt der Handlungsanreiz solcher Tätigkeiten nach wie vor in der Überzeugung begründet, dass sie für ein bestimmtes Ziel lohnend sind (Ryan et al., 2021). Im Gegensatz dazu sind intrinsisch motivierte Handlungen vollkommen frei von externen Anreizen. Sie werden um ihrer selbst willen vollzogen und fokussieren nicht die Erlangung einer bestimmten Konsequenz. Intrinsische Verhaltensweisen sind vorwiegend durch Interesse, Neugier und Spontaneität gekennzeichnet. Sie zeichnen sich durch den höchsten Grad an Selbstbestimmung aus und besitzen einen vollständigen internal perceived locus of causality (Deci & Ryan, 1993, 2000; Ryan & Deci, 2020).

Ein zentrales Postulat der SDT ist, dass der Mensch ein natürliches Bestreben besitzt, extrinsisch motivierte in selbstbestimmte Verhaltensweisen zu überführen (Deci & Ryan, 1993; Ryan & Deci, 2000). Dazu müssen entsprechende Verhaltensweisen zunächst internalisiert werden, d. h. „… people's „taking in" a value or regulation…" (Ryan & Deci, 2000, S. 71).

Internalisierte Werte und Verhaltensweisen sind im Selbst integriert. Der Prozess der Internalisierung kann unterschiedlich erfolgreich ablaufen, wodurch unterschiedliche Arten von Motivation entstehen (Deci & Ryan, 2000). Gleichwohl dieser Prozess einem natürlichen Bestreben folgt, bedarf er weiterer Unterstützung, um erfolgreich abzulaufen. Von entscheidender Bedeutung ist hierbei das Ausmaß, in dem die drei psychologischen Grundbedürfnisse nach Kompetenz, sozialer Eingebundenheit und Autonomie befriedigt werden (Deci & Ryan, 1993, 2000).

Das Erleben von *Kompetenz* bezieht sich auf das Gefühl, in sozialen Interaktionen die eigenen Fähigkeiten und Fertigkeiten einbringen und ausdrücken zu können. Im Vordergrund stehen hierbei nicht die tatsächlich erworbenen Fähigkeiten oder Fertigkeiten, sondern viel mehr das Gefühl selbstwirksam zu sein und dem eigenen Handeln vertrauen zu können (Ryan & Deci, 2002). Das Bedürfnis nach *sozialer Eingebundenheit* beschreibt „…the tendency to connect with and be integral to and accept by others" (Ryan & Deci, 2002, S. 7). Demnach streben Menschen besonders danach, ein Gefühl der Zugehörigkeit zu erfahren sowie sich in eine Gemeinschaft einzubringen, in der sie umsorgt werden und sich um andere kümmern können (Ryan & Deci, 2017). Das für die SDT zentrale Grundbedürfnis ist das nach *Autonomie*. Deci und Ryan (2017) bezeichnen es als „…vehicle through which the organization of personality proceeds and through which other psychological needs are actualized." (S. 97). Das Grundbedürfnis nach Autonomie adressiert den Wunsch, das eigene Verhalten als vollständig selbstbestimmt zu erleben. Im Wesentlichen wird damit die angeborene Tendenz des Menschen angesprochen, dem Selbst Ausdruck zu verleihen. Das Gegenteil von Autonomie – Heteronomie – bezieht sich auf Verhaltensregulierungen, die außerhalb des Selbst liegen. Dies können zum einen innere Impulse oder externe Anreize wie Belohnungen oder Bestrafungen sein (Ryan & Deci, 2006).

Innerhalb der SDT wird die Bereitstellung von Wahloptionen häufig als Maßnahme zur Autonomieunterstützung interpretiert. Die SDT postuliert damit, dass Individuen einen höheren internal perceived locus of causality im eigenen Verhalten erleben und selbstbestimmter sind, wenn sie zwischen mehreren Optionen wählen können. Dies kann wiederum zu einer erhöhten intrinsischen Motivation führen (Ryan & Deci, 2000, 2017). Dies ist aber keine zwangsläufige Konsequenz. Ryan und Deci (2002) konstatieren, dass insbesondere die Passung der Wahloptionen mit eigenen Werten und Vorstellungen entscheidend für das Erleben von Autonomie ist. Der Wert einer spezifischen Wahloption wird durch Erwartungs-Wert-Modelle näher beschrieben.

3.1.2 Das Erwartungs-Wert-Modell leistungsbezogener Wahlentscheidungen

In der Motivationspsychologie sind sogenannte Erwartungs-Wert-Theorien weit verbreitet und allgemein akzeptiert, um Wahlentscheidungen zu erklären. Das Paradigma derartiger Ansätze besteht darin, dass der subjektive Wert einer Handlungsalternative und die Erwartung des Erfolgs determinieren, welche Alternative (z. B. welche Aufgabe) ausgewählt wird (Beckmann & Heckhausen, 2018; Rheinberg & Vollmeyer, 2012; Urhahne, 2008).

Für die pädagogische Psychologie haben Eccles et al. (1983) grundlegende Annahmen der Erwartungs-Wert-Theorie von Atkinson (1957) übernommen und durch die Integration sozialkultureller und psychologischer Konstrukte besonders die Erwartungs- und Wert-Komponente weiter ausdifferenziert. Sowohl die Erwartungs- als auch die Wertkomponente werden in dem Modell durch ein komplexes Wirkungsnetzwerk aus soziokulturellen Einflüssen, früheren Erfahrungen sowie weiteren affektiven und kognitiven Merkmalen der Person beeinflusst. Das so entstandene Erwartungs-Wert-Modell leistungsbezogener Wahlentscheidungen (engl. *expectancy value model of achievement-related choices*) wurde in der Vergangenheit durch eine Vielzahl an Untersuchungen empirisch validiert und eignet sich damit zur Erklärung von Leistung und Wahlentscheidungen im schulischen Kontext. Eccles und Wigfield (2020) bezeichnen ihr Modell in neueren Veröffentlichungen als *situated expectancy-value-theory* (SEVT), um zu betonen, dass sowohl die Erwartungs- als auch die Wertkomponente situationsspezifisch sind. Ein reduziertes Modell ist in Abbildung 2 dargestellt.

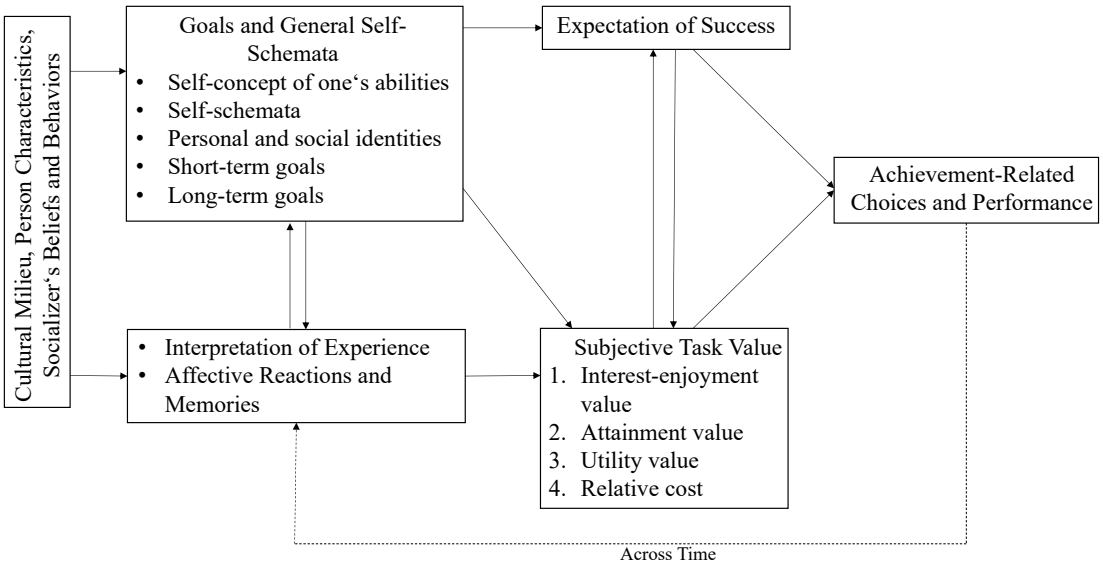


Abbildung 2: Das reduzierte Modell der SEVT nach Wigfield und Eccles (2020)

Die Erfolgserwartung wird von Eccles (2002, S. 119) definiert als „individuals' beliefs about how well they will do on upcoming tasks, either in the immediate or longer term future". Damit weist dieses Konstrukt eine große konzeptionelle Nähe zur *efficacy expectation* von Bandura (1977) und seiner Arbeit zur Selbstwirksamkeitserwartung auf (Wigfield & Eccles, 2000). In Analogie zur Erfolgserwartung wird darunter die eigene Überzeugung gefasst, anstehende Aufgaben bewältigen zu können. Die Erfolgserwartung wird maßgeblich durch das akademische Selbstkonzept, die Selbstschemata und die Ziele der jeweiligen Person beeinflusst (Wigfield & Eccles, 2020). In Abgrenzung zur Erfolgserwartung ist das akademische Selbstkonzept eher stabil und auf die Gegenwart gerichtet, während die Erfolgserwartung stärker von Zeit und Aufgabe abhängig ist und zukünftige Ereignisse fokussiert. In vergangenen Forschungsarbeiten konnten die Erfolgserwartung und das Selbstkonzept nicht als empirisch trennbare Konstrukte ausgewiesen werden und werden deshalb häufig als ein Konstrukt operationalisiert (Eccles & Wigfield, 2020; Wigfield & Eccles, 2020).

Der Wert einer Aufgabe wird im Wesentlichen durch den intrinsic value, den attainment value, den utility value und die relative costs bestimmt (Eccles et al., 1983).

Die *relative costs* beziehen sich auf alle negativen Aspekte, die mit der Wahl einer Aufgabe einhergehen (Eccles, 2002). Eccles und Wigfield differenzieren zwischen drei Arten von Kosten: effort costs, opportunity costs und emotional costs. Die *effort costs* beziehen sich auf den Aufwand, der für die Bewältigung einer Aufgabe betrieben werden muss. *Opportunity costs* sind die Opportunitätskosten, d. h. der entgangene Ertrag der nicht gewählten Aufgabe. Die emotionalen und sozialen Ängste, die mit dem Scheitern bei der Aufgabenbearbeitung verbunden sind, werden von den Autoren unter dem Begriff der *emotional costs* gefasst (Eccles & Wigfield, 2020).

Der *utility value* einer Aufgabe beschreibt, wie nützlich eine Aufgabe in Bezug auf ein persönliches Ziel ist (Wigfield & Eccles, 2000). Zum Beispiel kann die Wahl einer medizinbezogenen Aufgabe mit einem hohen utility value einhergehen, da ein Medizinstudium angestrebt wird. Folglich steht die Komponente des utility value in Verbindung mit extrinsischer Motivation innerhalb der SDT. Gleichwohl konstatieren Eccles und Wigfield (2020), dass Aufgaben mit einem hohen utility value auch einen großen Bezug zu persönlichen Zielen haben können, die wiederum Ausdruck des Selbst sind und damit eher intrinsischen Charakter aufweisen.

Die subjektive Bedeutsamkeit, eine Aufgabe erfolgreich zu bearbeiten, wird unter dem Wertaspekt des *attainment value* beschrieben (Eccles et al., 1983). Eccles und Wigfield (2020, S. 5) postulieren: „…attainment value derives from the assumed fit of perceived task characteristics with the individual's core self-schema, social and personal identities, and ought selves". Dieser Aspekt wird auch in der SDT im Hinblick auf das Autonomieerleben aufgegriffen.

Aufgaben mit einem ausgeprägten *interest-enjoyment value* bereiten Freude bei der Bearbeitung (Wigfield & Eccles, 2000). Dabei versteht Eccles (2005) Freude im Sinne der Flow-Theorie von Csikszentmihalyi (1988) als vollständiges Aufgehen in der Tätigkeit. Im Verständnis der SDT handelt es sich dabei um internalisierte Verhaltensweisen, die intrinsisch motiviert sind. Weiterhin bestehen konzeptionelle Überschneidungen zum Interessenkonzept (Eccles & Wigfield, 2020).

3.1.3 Die Person-Gegenstands-Konzeption des Interesses

Die Person-Gegenstands-Konzeption (engl. *person-object theory of interest, POI*) beschreibt Interesse als eine besondere Beziehung zwischen einer Person und einem Gegenstand. Diese Gegenstands- und Inhaltszentrierung grenzt das Interessenkonstrukt von anderen Motivationskonstrukten ab (Krapp, 1992b, 1999a, 2002; Krapp & Prenzel, 2011).

Ein Interessengegenstand kann als kognitive Repräsentation der erfahrbaren Umwelt beschrieben werden, die mehr oder weniger voneinander abgrenzbar und unterschiedlich akzentuiert sind. Gegenstände können in diesem Zusammenhang konkrete Dinge, bestimmte Themen oder auch abstrakte Ideen sein. Ebenso können Unterrichtsfächer als Interessengegenstand dienen (Krapp, 1992b, 2002; Krapp & Prenzel, 2011). In Bezug auf einzelne Unterrichtsfächer (z. B. Physik oder Chemie) kann hinsichtlich bestimmter Themen, Kontexte oder Aktivitäten ein unterschiedliches Maß an Interesse bestehen (Häußler, 1987; Häußler & Hoffmann, 2000).

Die Beziehung zwischen Person und Gegenstand ist sowohl durch affektive als auch kognitive Merkmale geprägt (Hidi et al., 2004). Krapp (1999b, S. 398) bezeichnet vorwiegend zwei Bewertungstendenzen als „…zentrale Aspekte des Interessenkonstrukts“: die emotionale und wertbezogene Valenz (siehe auch Schiefele, 1991). Die *emotionale Valenz* bezieht sich hierbei auf positive Gefühle, die durch die Auseinandersetzung mit dem Interessengegenstand einhergehen. Diese können insbesondere Freude, Spannung oder hohes Engagement sein (Krapp, 1992b, 2002). Unter besonders günstigen Umständen kann es zu einem vollkommenen Aufgehen in der Interessenhandlung im Sinne des Flow-Erlebens kommen (Csikszentmihalyi, 1988). Die besondere Bedeutsamkeit des Interessengegenstands wird durch die *wertbezogene Valenz* beschrieben. Diese resultiert hauptsächlich aus einem hohen Grad an Identifikation sowie zugeschriebener Relevanz des Gegenstandes (Krapp, 1999b, 2002). Daneben ist das Interesse auch durch eine *kognitiv-epistemische Komponente* geprägt. Im Vordergrund der Handlungsausführung steht stets auch der Erwerb von neuem Wissen über den Interessengegenstand oder die weitere Ausdifferenzierung vorhandener Wissensstrukturen (Krapp, 2002). Durch die emotionale und wertbezogene Bedeutung der interessengeleiteten Handlung

ist jene als selbstintentional zu bezeichnen. Folglich ist die Beschäftigung mit dem Interessengegenstand frei von externen Zwängen und intrinsisch motiviert (Krapp, 1992b; Krapp & Prenzel, 2011).

Das Interessenkonstrukt kann auf unterschiedlichen Ebenen betrachtet werden (Krapp, 1992a). Ein erster theoretischer Zugang ergibt sich aus der Interpretation als motivationale Disposition, die zeitlich überdauernd und als Persönlichkeitsmerkmal aufzufassen ist. Dies wird als *individuelles Interesse* bezeichnet. Daneben wird in der Literatur noch ein *themenbezogenes Interesse* als weiteres habituelles Merkmal beschrieben, welches sich stärker an spezifischen Themenbereichen ausrichtet (Schiefele & Krapp, 1996). Werden individuelle Interessen in einer konkreten Situation handlungswirksam, kann man den daraus resultierenden psychischen Zustand als *aktualisiertes (individuelles) Interesse* beschreiben. Das *situationale Interesse* resultiert dagegen aus externen Faktoren, die in einer spezifischen Situation aufgrund der *Interessantheit* des Gegenstandes wirken. In diesem Zusammenhang ist das situationale Interesse eher instabil und an die konkrete Situation gebunden. Die angeführten Ebenen des Interessenkonstrukts sind keinesfalls als konträre Konzepte des Interesses zu verstehen, sondern repräsentieren Subkonzepte, die in einer relationalen Beziehung zueinanderstehen (Abb. 3).

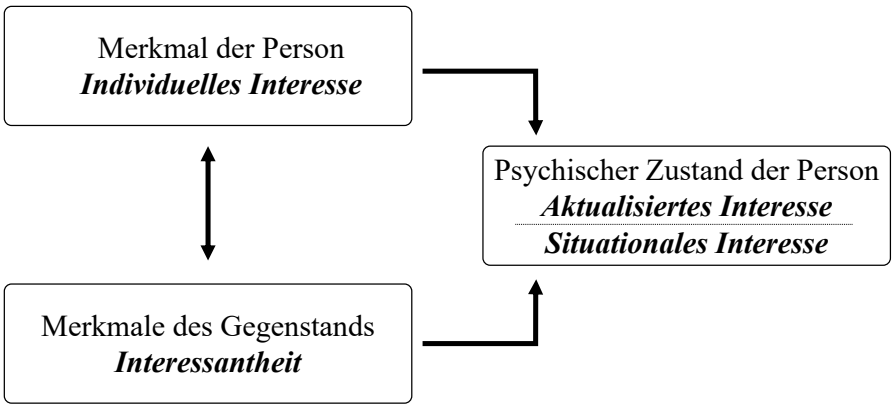


Abbildung 3: Die relationale Struktur des Interessenkonstrukts nach Krapp (1992a)

Überdies kommt dem situationalen Interesse eine zentrale Bedeutung im Prozess der Interessengenese zuteil. Gemäß dem Vier-Phasen-Modell von Hidi und Renninger (2006) wird erst durch situationales Interesse die Ausbildung von individuellen Interessen ermöglicht. Situationales Interesse muss dabei zunächst durch sogenannte *catch*-Faktoren angeregt werden. Unter catch-Faktoren versteht Mitchell (1993) sensorische und kognitive Stimuli, die eine kurzfristige Änderung in der affektiven und kognitiven Verarbeitung hervorrufen. Sensorische Stimuli können dabei beispielsweise auffällige optische oder akustische Anreize in der

Umwelt sein. Dagegen betreffen kognitive Stimuli eher Diskrepanzen zwischen vorhandenen kognitiven Strukturen und der Umwelt. Diese Diskrepanzen können wiederrum einen kognitiven Konflikt auslösen. In diesem Zusammenhang können beispielsweise auch überraschende Informationen oder persönliche Bezüge in einem Text als catch-Faktoren dienen (Anderson et al., 1987). Neben catch-Faktoren sind nach Mitchell (1993) besonders *hold*-Faktoren zur Aufrechterhaltung des situationalen Interesses notwendig. Als hold-Faktoren können Lernumgebungen oder Aufgaben dienen, welche besonders die wertbezogene Valenz des Interesses adressieren und damit persönlich Bedeutsam sind (Hidi & Renninger, 2006).

3.2 Empirische Befunde zur Bereitstellung von Wahloptionen

Schraw et al. (2001) führen die Bereitstellung von Wahloptionen als Unterrichtsstrategie und damit als möglichen catch-Faktor zur Förderung des situationalen Interesses an. Im Einklang mit der SDT ist dies als Maßnahme zur Unterstützung der Autonomie zu interpretieren, die zu höherer Selbstbestimmung im Unterricht führen sollte. Palmer (2009) kommt durch Interviews mit Lernenden nach einer experimentellen Unterrichtsstunde zu dem Ergebnis, dass neben Neugier vor allem auch die Möglichkeit zur Wahl situationales Interesse auslöst. Auch unter Lehrkräften ist die Vorstellung weitverbreitet, dass die Lernenden motivierter im Unterricht sind, wenn sie zwischen verschiedenen Optionen wählen können, wie die Untersuchung von Flowerday und Schraw (2000) zeigt.

Die (pädagogisch-)psychologische Forschung hat sich hierbei verstärkt mit der Frage beschäftigt, ob die Bereitstellung von Wahloptionen tatsächlich mit positiven Effekten verbunden ist. Die Forschungslage lässt sich hierbei als „…equivocal and confusing“ bezeichnen (Katz & Assor, 2007, S. 429). Um bisherige Forschungsergebnisse zu kondensieren, haben Patall et al. (2008) eine Metaanalyse mit 41 experimentellen Studien durchgeführt. Die Autorinnen und Autoren konkludieren: „…when individuals are allowed to affirm their sense of autonomy through choice they experience enhanced motivation, persistence, performance, and production.” (2008, S. 298). Es konnte eine kleine Effektstärke von $d = 0.30$ für den Einfluss von Wahlmöglichkeiten auf die intrinsische Motivation identifiziert werden. Ferner konnten positive Effekte auf die Leistung ($d = 0.32$) und die Anstrengungsbereitschaft ($d = 0.22$) der Probandinnen und Probanden ausgemacht werden.

Bei genauerer Betrachtung der in der Metaanalyse berücksichtigten Studien ist auffällig, dass sich auch gegensätzliche Ergebnisse zeigen. 13 der 91 Effektstärken indizieren einen negativen Einfluss auf die intrinsische Motivation. In einer experimentellen Laboruntersuchung mit Studierenden konnte keine Wirkung durch die Auswahl eines Lesetextes im Hinblick auf das Interesse oder die Leistung in einem Multiple-Choice-Test identifiziert werden (Schraw et al., 1998). Ebenso scheint die Wahl zwischen einem Essay oder einem Kreuzworträtsel keinen Einfluss auf die Leistung oder affektive Einstellungen zu haben (Flowerday & Schraw, 2003).

D'Ailly (2004) konnte in einer Studie mit Fünft- und Sechstklässlern zeigen, dass die Auswahl von Namen, Tieren oder Zahlen beim Vokabellernen einer Fremdsprache nicht mit lernförderlichen Effekten verbunden ist.

In Anbetracht dieser Ergebnisse scheint es so, als wäre „… the relationship between choice and motivation be more complex than the commonsense belief that choice is unequivocally beneficial" (Patall et al., 2008, S. 272). Aber unter welchen Bedingungen führt die Bereitstellung von unterschiedlichen Wahloptionen zu lernförderlichen Effekten?

Zu Beginn der 2000er-Jahre hat besonders Schwartz (2000) die These formuliert, dass die Bereitstellung einer erhöhten Anzahl von Wahloptionen mit negativen Auswirkungen verbunden ist. Bei seiner Argumentation bezog er sich u. a. auf die Studie von Iyengar und Lepper (2000), die durch Labor- und Feldstudien zeigen konnten, dass eine erhöhte Anzahl an Wahloptionen – 24 oder 30 im Vergleich zu sechs Optionen – die Motivation und Zufriedenheit nach der Aufgabenbearbeitung vermindert. Die Autoren bezeichnen die Überforderung durch die hohe Anzahl an Wahloptionen als *choice overload*. Dies ist primär mit den Kosten entgangener Wahloptionen verbunden. Durch einen zusammenfassenden Blick auf die Literatur zeigen sich positive Auswirkungen zugunsten von drei bis fünf Wahloptionen (Patall et al., 2008).

Zusätzlich erschöpfen Wahlentscheidungen als volitionale und selbstregulatorische Handlungen eine begrenzte Ressource des Selbst. Aus der Erschöpfung dieser Ressource resultiert schließlich ein mentaler Zustand, der als *ego depletion* bezeichnet wird (Baumeister et al., 1998). Dies führt – wie Baumeister et al. in einem Laborexperiment zeigen – zu einer verminderten Beharrlichkeit, ein unlösbares Puzzle zu lösen. Die Autoren schließen daraus, dass Wahlentscheidungen per se in dem Zustand der ego depletion münden können.

Moller et al. (2006) empfehlen einen differenzierten Blick auf den Zusammenhang zwischen Wahlentscheidungen und dem Zustand der ego depletion. Wie eingangs erwähnt, akzentuiert die SDT die Passung der Wahloptionen zu den eigenen Interessen, Werten und Zielen oder aus der Perspektive der SDT formuliert: dem eigenen Selbst. Wenn in einer Wahlentscheidung das Selbst handlungswirksam wird, ist die Handlung durch einen IPLOC gekennzeichnet und es handelt sich um eine autonome Wahlentscheidung, die positive Auswirkungen auf die Motivation hat. Wahlentscheidungen, die durch externen Druck herbeigeführt werden oder auf Optionen beruhen, die nicht zum Selbst passen, sind nicht autonom, sondern werden extern reguliert und führen zur ego depletion und anderen negativen Effekten. Diese Hypothese konnte durch drei Laborexperimente bestätigt werden, indem das Studiendesign von Baumeister et al. um eine sogenannte *autonomous-choice condition* ergänzt wurde (Moller et al., 2006, S. 1028). Auch Reeve et al. (2003) zeigen durch drei unterschiedliche Studien, dass lediglich Wahlentscheidungen mit einem IPLOC motivationsförderliche Effekte zeigen.

Deswegen empfehlen Moller et al. (2006) ausschließlich autonome Wahlentscheidungen als *choice* zu bezeichnen. Unter dem Begriff der *decision* sollen alle Wahlentscheidungen mit einem EPLOC verstanden werden. Eine ähnliche Begriffsdifferenzierung findet man bei Ullmann-Margalit und Morgenbesser (1977), die weitaus früher zwischen *choosing* und *picking* unterscheiden:

> „We speak of *choosing* among alternatives when the act of taking (doing) one of them is determined by the differences in one's preferences over them. When preferences are completely symmetrical, where one is strictly indifferent with regard to the alternatives, we shall refer to the act of taking (doing) one of them as an act of *picking*." (Ullmann-Margalit & Morgenbesser, 1977, S. 757)

Die Relevanz dieser Unterscheidung wird durch zahlreiche empirische Studien gestützt. Katz und Assor (2007) kommen in ihrem Literaturreview zu dem Ergebnis, dass lediglich die Auswahl zwischen Optionen, die einen Bezug zu eigenen Interessen und Werten haben, motivationsförderlich ist. Hier konnte Patall (2013) die Bedeutsamkeit des Interesses für die Effektivität von Wahlmöglichkeiten stärker herausarbeiten. Die Autorin kommt zu dem Ergebnis, dass Probanden lieber zwischen Wahloptionen wählen, an denen sie auch interessiert waren. Demzufolge geben die Studierenden in ihrer Studie ein höheres situationales Interesse nach der Aufgabenbearbeitung an, wenn sie ein hohes individuelles Interesse hatten und eine Aufgabe wählen konnten. Studierende mit einem geringeren individuellen Interesse profitierten nicht von der selbstständigen Wahl, sondern von zugewiesenen Aufgaben.

Insbesondere das situationale Interesse an der spezifischen Aufgabe scheint für motivations- und lernförderliche Effekte entscheidend zu sein und nicht die Auswahl zwischen den Optionen. Flowerday et al. (2004) haben in ihrer Untersuchung Versuchspersonen ein versiegeltes Paket mit Lesetexten zugewiesen oder wählen lassen. Den Versuchspersonen war das Thema der Lesetexte folglich nicht bekannt. Die Regressionsanalyse verweist darauf, dass das Leseengagement hierbei lediglich vom situationalen Interesse am Text und nicht von der Möglichkeit zur Wahl beeinflusst wird. In einem weiteren psychologischen Experiment haben Studierende unter Laborbedingungen entweder zwischen drei unterschiedlichen Texten zur Bearbeitung gewählt oder ihnen wurde ein Text zugewiesen, der zu ihren Interessen gepasst oder nicht gepasst hat (Flowerday & Shell, 2015). Die Datenanalyse durch ein Pfadmodell zeigt, dass die Wahl eines Textes keinen Einfluss auf das situationale Interesse hat. Hier ist lediglich entscheidend, ob die Probandinnen und Probanden mit einem Text gearbeitet haben, den sie vorher auch als interessant eingestuft haben. Dass dieses Ergebnis auch im unterrichtlichen Feld repliziert werden konnte, zeigt eine quasi-experimentelle Studie von Wilde et al. (2018) für den Biologieunterricht. Die Lernenden wurden vor der Intervention hinsichtlich ihrer Präferenz für ein Tier (Maus oder Küchenschabe) befragt. Daraufhin wurde den Lernenden entweder das präferierte oder nicht präferierte Tier für die anschließende vierstündige Unterrichtsreihe zugewiesen. Eine dritte Gruppe konnte erneut zwischen beiden Tieren für

die Unterrichtsreihe wählen, nachdem die Tiere näher vorgestellt wurden. Die Autoren berichten, dass die Lernenden die höchste intrinsische Motivation und das höchste Flow-Erleben zeigen, wenn sie nicht wählen mussten, aber mit dem präferierten Tier lernen konnten. Auf der Basis von Kontrastanalysen ergeben sich hierbei allerdings nur kleine Effektstärken ($.02 \leq \eta^2 \leq .04$).

Die Untersuchung von S. Schneider et al. (2018) mit Lernenden aus den Jahrgangsstufen 8, 9 und 10 verweist außerdem darauf, dass sich Wahleffekte nicht bloß auf affektive Variablen beschränken, sondern auch auf den Lernzuwachs sowie die kognitive Belastung wirken. In einem experimentellen 2x2-Design wurde untersucht, welchen Einfluss lernrelevante und lernirrelevante Wahlen auf affektive und kognitive Merkmale haben. Bei der lernrelevanten Wahl konnten die Schülerinnen und Schüler zwischen zwei Texten wählen, die hinsichtlich des Themas variiert wurden. Die lernirrelevante Wahl umfasste die Entscheidung zwischen zwei Musikstücken, die während der Aufgabenbearbeitung abgespielt wurden. Lernrelevante Wahlen erhöhen die Transferleistung und vermindern die intrinsische kognitive Belastung der Lernenden während der Aufgabenbearbeitung.

Weiterführend scheint die Effektivität der Bereitstellung von Wahloptionen durch die empfundene Kompetenz moderiert zu werden. Patall et al. (2014) zeigen hier, dass Personen lieber in Situationen wählen, in denen sie sich kompetent fühlen. In einer experimentellen Studie mit Studierenden haben die Autoren außerdem untersucht, ob die Wahleffekte durch die von den Studierenden empfundene Kompetenz moderiert werden. Die Ergebnisse belegen, dass Wahlen besonders motivierend wirken, wenn sich die wählende Person kompetent fühlt. Dieser Befund ist zudem noch abhängig von der wahrgenommenen Aufgabenschwierigkeit. Die motivationsförderlichen Effekte werden vermindert, wenn die Aufgabe eine hohe wahrgenommene Aufgabenschwierigkeit hat.

3.3 Implikationen für das Fach Chemie

Die Ergebnisse der psychologisch-pädagogischen Forschung deuten lern- und motivationsförderliche Effekte durch die Wahl zwischen Optionen an, die einen Bezug zu eigenen Interessen und Werten haben. Etwas deutlicher formuliert scheint insbesondere die Passung der Wahloptionen und nicht der bloße Auswahlprozess entscheidend zu sein. Auf Grundlage der durchgeführten Untersuchungen sind hier kleine Effekte zu erwarten.

Mit Bezug auf die chemiedidaktische Forschung ist allerdings anzumerken, dass viele Studien unter Laborbedingungen und nicht im unterrichtlichen Feld durchgeführt wurden. Eine Ausnahme bildet die Studie von Wilde et al. (2018) mit Blick auf den Biologieunterricht. Ferner beschränkt sich die Auswahl der Versuchspersonen häufig auf Studierende, womit unklar

bleibt, inwieweit die Ergebnisse beispielsweise für Lernende der Sekundarstufe I gelten. Zudem sind viele Studien auf die Wahl von Lesetexten limitiert. Zusammenfassend ist damit unklar, inwieweit die Ergebnisse auf den Chemieunterricht übertragbar sind.

Trotzdem kann als zentrales Ergebnis herausgearbeitet werden, dass ein differenzierter Blick auf die Ausgestaltung der Wahloptionen und die Merkmale der wählenden Person notwendig ist. Der Passung zwischen Wahloptionen und den Interessen, Zielen und Werten einer Person kommt hier eine entscheidende Rolle zuteil, die empirisch hinreichend belegt ist (u. a. Flowerday & Shell, 2015; Katz & Assor, 2007; Wilde et al., 2018). Hier ist für das Fach Chemie bislang unklar, wie Aufgaben variiert werden können, um unterschiedliche Interessen zu adressieren. Insbesondere in der Sprachdidaktik ist die Bereitstellung von Lesetexten mit unterschiedlichen Themen verbreitet. Im naturwissenschaftlichen Unterricht ist es dagegen gängig, naturwissenschaftliche Inhalte mit außerfachlichen Anwendungsbezügen zu verknüpfen. Das Lernen in sogenannten *Kontexten* wird in vielen Curricula sogar explizit verordnet (z. B. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2019). Die Forschung zum kontextorientierten Lernen deutet an, dass sich unterschiedliche Kontexte zur interessenbasierten Differenzierung eignen könnten.

4 Kontextorientiertes Lernen im naturwissenschaftlichen Unterricht

Als Ausgangspunkt für die Entwicklung des kontextorientierten Lernens werden gemeinhin die Probleme des traditionellen naturwissenschaftlichen Unterrichts der Jahrtausendwende genannt, die von Gilbert (2006) zusammengefasst wurden. Demnach sind die naturwissenschaftlichen Curricula mit Inhalten überladen, wodurch häufig nur isolierte Fakten von den Lernenden akkumuliert werden, die kaum auf neue Probleme angewandt werden können. Das Lernen von chemischen Inhalten ist für die meisten Lernenden außerdem nur wenig relevant, da nur ein geringer Fokus auf die Frage gelegt wurde, warum bestimmte Inhalte erlernt werden sollten. Der Schwerpunkt des traditionellen naturwissenschaftlichen Unterrichts lag damit eher auf der Ausbildung fachwissenschaftlicher Grundlagen, um weiter fortgeschrittene Inhalte der jeweiligen Wissenschaft studieren zu können.

Diese Schwerpunktsetzung des Unterrichts ist allerdings nicht nur problematisch vor dem Hintergrund der begrenzten Anzahl der Lernenden, die ein naturwissenschaftliches Studium aufnehmen, sondern vor allem, da die naturwissenschaftliche Grundbildung aller Lernenden zentrales Anliegen des naturwissenschaftlichen Unterrichts ist. Die naturwissenschaftliche Grundbildung (*Scientific Literacy*) umfasst nach Bybee und McCrae (2011) sowie Schiepe-Tiska, Rönnebeck et al. (2016) folgende Fähigkeiten:

- Naturwissenschaftliches Wissen anwenden, um Fragestellungen zu erkennen, sich neues Wissen anzueignen, naturwissenschaftliche Phänomene zu beschreiben und zu erklären sowie aus Belegen Schlussfolgerungen zu ziehen.
- Die charakteristischen Eigenschaften der Naturwissenschaften als eine Form menschlichen Wissens und Forschens zu verstehen.
- Zu erkennen und sich darüber bewusst zu sein, wie Naturwissenschaften und Technik unsere materielle, intellektuelle und kulturelle Umwelt formen.
- Die Bereitschaft, sich mit naturwissenschaftlichen Ideen und Themen zu beschäftigen und sich reflektierend mit ihnen auseinanderzusetzen.

Ein zentraler Aspekt der naturwissenschaftlichen Grundbildung besteht folglich darin, diese Fähigkeiten einzusetzen, um in Alltagssituationen mit naturwissenschaftlich-technischen Fragestellungen und Problemen umzugehen (Schiepe-Tiska, Rönnebeck et al., 2016). Zur Förderung der naturwissenschaftlichen Grundbildung und zur Überwindung der aufgeführten Probleme des traditionellen naturwissenschaftlichen Unterrichts werden seit einiger Zeit kontextorientierte Lehrverfahren entwickelt und in den Unterricht implementiert (Bennett & Lubben, 2006; Taconis et al., 2016). Diese verfolgen im Wesentlichen ähnliche Zielsetzungen. Durch die Implementation von kontextorientierten Ansätzen in den Unterricht soll die Verbindung zwischen fachlichen Inhalten und der Alltagswelt der Lernenden akzentuiert werden. Ein übergeordnetes Ziel liegt darin, die fachlichen Inhalte für die Lernenden relevanter zu gestalten und eine Antwort auf die Frage zu geben, warum es notwendig ist (*need to know*-Prinzip), die fachlichen Inhalte zu erlernen (Gilbert et al., 2011). Infolgedessen sollen die

Lernenden motivierter im Unterricht sein und ein höheres Interesse an naturwissenschaftlichen Inhalten ausbilden (Bennett, 2016). Dabei steht insbesondere der Aufbau einer vernetzten Wissensstruktur im Vordergrund (Bennett et al., 2005; Nentwig et al., 2007; Waddington, 2005). Damit verbunden ist die Absicht, die Lernleistung der Schülerinnen und Schüler zu verbessern.

Erste Kontextprojekte finden sich bereits in den 1960er-Jahren. Dort begann Freudenthal (1968) in den Niederlanden Alltagssituationen für die mathematische Bildung zu nutzen. In den nachfolgenden Jahren wurden in zahlreichen Ländern Projekte gestartet, welche die Idee der Anwendung von mathematisch-naturwissenschaftlichem Wissen in Alltagssituationen verfolgten (Taconis et al., 2016). Im nordamerikanischen Raum wurden entsprechende Projekte unter dem *Science-Technology-Society* (*STS*)-Konzept durchgeführt. Zentral für das STS-Konzept ist die Betonung eines gesellschaftlichen oder sozialen Problems und deren Verbindung zu Naturwissenschaften und Technologie (Aikenhead, 1994; Fensham, 1988). Einen ähnlichen Ansatz verfolgt das *Salters*-Projekt aus dem Vereinigten Königreich, welches seit den 1980er-Jahren kontextbasierte Kurse für verschiedene Altersgruppen entwickelt. Analog zum STS-Konzept akzentuiert auch das Salters-Projekt die Verknüpfung von fachwissenschaftlichen Konzepten mit der Lebenswelt der Schülerinnen und Schüler (Bennett & Lubben, 2006). Die unterschiedlichen Begrifflichkeiten sind im Wesentlichen auf den unterschiedlichen Ursprung der Projekte zurückzuführen. In Nordamerika wird häufiger von STS gesprochen, während in Europa der Begriff *context-based* (kontextorientiert oder kontextbasiert) verbreiteter ist (Bennett et al., 2007). Für den Physikunterricht in den Niederlanden wurde das *PLON*-Projekt gestartet, welches kontextbasierte Lehrmaterialien entwickelte (Kortland, 2005). In Deutschland wurde ein Projekt unter dem Namen *Chemie im Kontext (ChiK)* initiiert (Parchmann et al., 2006). Analoge Projekte wurden jeweils mit *Physik im Kontext* und mit *Biologie im Kontext* umgesetzt. Weitere Projekte finden sich außerdem in Südafrika, Israel, Irland, Türkei, Australien und zahlreichen weiteren Ländern (Bennett & Holman, 2002; Taconis et al., 2016).

4.1 Zum Kontextbegriff

Gleichwohl bislang zahlreiche Projekte zum kontextorientierten Lernen entwickelt wurden und das kontextorientierte Lernen auch Einzug in die Curricula fast aller Bundesländer erhalten hat (z. B. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2019), ist der wissenschaftliche Diskurs immer noch von einer gewissen Uneinigkeit hinsichtlich des Kontextbegriffes geprägt. Vor rund 30 Jahren äußerten Duranti und Goodwin bereits aus der Perspektive der Lingustik ihre Zweifel an einer allgemeingültigen Definition des Kontextbegriffs: „…it does not seem possible at the present time to give a single, precise, technical definition of context, and eventually we might have to accept that such a definition may not

be possible." (1992, S. 2). Bis heute findet sich auch in den Naturwissenschaftsdidaktiken keine einheitliche Definition (Bennett & Holman, 2002).

Die etymologische Herkunft des Wortes lässt sich im Lateinischen auf das Substantiv *contextus* zurückführen, was so viel wie „Zusammenhang" oder „enge Verknüpfung" bedeutet. McDermott nutzt die Analogie mit einem Behälter, um den Begriff näher zu definieren: „In all common sense of the term, context refers to an empty slot, a container into which other things are placed." (1993, S. 282). Für die Naturwissenschaftsdidaktik hat Finkelstein (2005) ein Modell mit mehreren Ebenen postuliert, um den Kontextbegriff näher zu beschreiben.

Die äußerste Ebene des Kontextbegriffs bezeichnet Finkelstein (2005) mit Bezug auf Fine (1987) als *idioculture* einer Lerngruppe. Diese umfasst spezifische Verhaltensweisen und Rituale einer Lerngruppe. Die idioculture umrahmt damit die *situation*, welche die nächste Ebene des Modells darstellt. Auf dieser Ebene befinden sich die Lernenden, die Lernaufgaben sowie deren Ziel, aber auch die dazugehörigen Hilfsmittel zur Lösung der Aufgabe. Weiterführend ist die Situation als Kontext durch die gewählte Sozialform geprägt, in der die Lernenden arbeiten. Die innerste Ebene stellt die eigentliche Aufgabenstellung dar. Diese lässt sich durch das Wechselspiel zwischen Lernenden, der Aufgabe sowie dem fachlichen Konzept charakterisieren. Die Aufgabe im engeren Sinne ist dabei der Kontext des Problems, wobei der Kontext hierbei als „…storyline of a problem" fungiert (Mestre, 2002, S. 2). In Anlehnung an Chi et al. (1981) und Mestre (2002) verortet Finkelstein diese *storyline* auf der Oberflächenebene des Problems. Auf der Tiefenebene befinden sich dagegen die fachlichen Konzepte.

Gilbert (2006) greift in seiner viel beachteten Publikation zum Kontextbegriff die Aufgaben- und Situationsebene von Finkelstein (2005) auf und definiert einen Kontext mit Bezug auf Duranti und Goodwin (1992) als „…a focal event embedded in its cultural setting." (S. 960). Das *focal event* kann ein Foto, ein animiertes Modell oder auch ein Diagramm sein; es ist vorwiegend ein Ereignis, welches Aufmerksamkeit erregt (vgl. catch-Faktoren bei Mitchell, 1993). Das focal event ist eingebettet in ein *cultural setting*, den sozialen, räumlichen und zeitlichen Rahmen (Gilbert, 2006; Gilbert et al., 2011).

Gilbert (2006) leitet vier Attribute ab, die ein Kontext zur Implementation in eine Lernumgebung aufweisen sollte:

(1) Der Kontext muss ein focal event aufweisen, welches in einen für die Lernenden relevanten Bezugsrahmen (cultural setting) integriert ist. Dieser Bezugsrahmen sollte aus der Lebenswelt, der sozialen oder industriellen Umwelt stammen und bedeutsam für die Lernenden sein. Dadurch soll der Kontext für das Lernen wertgeschätzt werden.
(2) Die Lernaufgabe sollte eine spezielle Lernumgebung aufweisen, die Probleme thematisiert, welche mithilfe von chemischen Konzepten bearbeitet werden können.

(3) Der Kontext sollte vielfältige Anlässe zur Verwendung der chemischen Fachsprache bieten, wodurch die kohärente Verwendung erlernt wird.
(4) Es müssen Gelegenheiten geschaffen werden, damit die Lernenden das Wissen aus dem focal event herauslösen und mit dem eigenen Vorwissen vernetzen (dekontextualisieren). Das neu erworbene Wissen wird auf ein neues focal event angewandt (rekontextualisiert). Der Kontext soll damit den Transfer von Fachwissen ermöglichen.

Darauf aufbauend beschreibt der Autor vier unterschiedliche Modelle für das kontextorientierte Lernen, welche diese Kriterien mehr oder weniger hinreichend umsetzen. Das erste Modell beschreibt den Kontext als direkte Anwendung der Konzepte. Dieses Modell impliziert eine einseitige Beziehung zwischen Kontext und dem fachlichen Konzept, wobei ausschließlich die Anwendung der Konzepte in dem Kontext beispielsweise am Ende der Unterrichtsstunde fokussiert wird. Dieses Modell wird häufig im Unterricht umgesetzt – selbst wenn der sonstige Unterricht nicht kontextorientiert ist. Dagegen wird der Kontext im zweiten Modell als wechselseitige Verbindung zwischen Anwendung und Konzept dargestellt. Er fungiert als „…vehicle through which key concepts can be taught (Gilbert et al., 2011, S. 823). Als Ergänzung zum ersten Modell wird hier der Kontext nach der Konzeptbildung nicht nur als Anwendungsbeispiel der Konzepte präsentiert, sondern schon als Ausgangspunkt für die Vermittlung neuer Konzepte genutzt. Es wird angenommen, dass sich Kontext und Konzept wechselseitig beeinflussen. Das dritte Modell beschreibt den Kontext als das Ergebnis persönlicher geistiger Aktivität, wie sie sich beispielsweise aus der alleinigen Arbeit eines Lernenden mit Lernaufgaben ergibt. Durch die mentale Aktivität des Lernenden wird ein Kontext aus einer Situation (bzw. dem setting) konstruiert, wobei bestehende mentale Modelle dem Kontext Bedeutung verleihen. Dieser Prozess verläuft vollständig internal (Gilbert et al., 2011). Dieses Modell basiert auf der Annahme, dass sich die Entwicklung von Wissen vollständig innerhalb der Person vollzieht (Pope & Keen, 1981). Das letzte Modell definiert den Kontext auf der Grundlage des sozialen Konstruktivismus. Der Kontext wird dabei als soziales Umfeld verstanden, in dem die Lehrenden und Lernenden gemeinsam und über einen längeren Zeitraum an einer Reihe von focal events problemzentriert arbeiten.

Die ersten beiden Modelle des Kontextbegriffes beschreiben damit, in welcher Beziehung die Anwendungssituation zum fachlichen Konzept steht. Dahingegen stellen die Modelle 3 und 4 den Kontext als das Ergebnis eines internalen oder sozialen Konstruktionsprozesses dar. Gilbert et al. (2011) merken diesbezüglich an, dass lediglich das vierte Modell allen Anforderungen an das kontextorientierte Lernen genügt, wobei auch das dritte Modell tragfähig scheint. Die meisten Kontextprojekte nutzen den Kontextbegriff entsprechend dem dritten oder vierten Modell (Pilot & Bulte, 2006). Dementgegen stellt Bennet (2016) fest, dass die meisten Studien ihrer Metaanalyse den Kontextbegriff gemäß dem zweiten Modell verwenden.

Innerhalb der vorliegenden Arbeit wird der Kontextbegriff auf der innersten Ebene – der Aufgabenebene – gemäß der Taxonomie von Finkelstein (2005) betrachtet. Auf der Ebene der Lernaufgabe fungieren Anwendungssituationen für fachliche Konzepte als Strukturierungshilfe für das Lernen. Diese Anwendungssituationen entspringen der Lebenswelt, der sozialen oder industriellen Umwelt und weisen ein *focal event* auf, welches problemzentriert durch fachliche Konzepte bearbeitet werden kann. Die Anwendungssituation wird hierbei nicht bloß als „…starting point for the development of scientific ideas" (Bennet, 2016, S. 2) oder als "…Ausgangspunkt für die Erarbeitung des fachlichen Inhalts…" (van Vorst et al., 2015, S. 30) verstanden. Viel mehr dient sie als „…red thread along which the investigation of the issues in questions develops." (Nentwig, 2007, S. 1441). Folglich kommt der Anwendungssituation nicht nur am Anfang oder Ende des Lernprozesses eine herausragende Bedeutung zuteil, sondern rahmt den gesamten Lernprozess. Der Kontext entwickelt sich dabei als Resultat der mentalen Aktivität mit der Anwendungssituation in der Lernaufgabe, wobei vorhandene mentale Strukturen dem Kontext Bedeutung verleihen.

4.2 Kontextmerkmale

Kontexte können durch unterschiedliche Merkmale näher beschrieben werden. Hierbei können zunächst Situationen – aus denen die Kontexte konstruiert werden – anhand des Themas systematisiert werden. Auf dieser Grundlage lassen sich Kontexte beispielsweise den Themen Umweltschutz, Gesundheit oder auch Technik zuordnen (Elster, 2007; Schiepe-Tiska, Rönnebeck et al., 2016). Diesen Themen lassen sich in einem größeren Bezugsrahmen außerfachliche oder innerfachliche Situationen zuordnen. Innerfachliche Situationen entstammen in der Regel nicht der Lebenswelt der Schülerinnen und Schüler, sondern sind stattdessen typisch für das entsprechende Fach. Im Fach Chemie könnte das beispielsweise eine Situation aus dem Labor sein (Fechner et al., 2015). Im Gegensatz dazu stammen außerfachliche Situationen nicht aus dem entsprechenden Fach. Sie können in der Lebenswelt der Lernenden angesiedelt sein oder aus der sozialen, technischen oder industriellen Umwelt entspringen (Gilbert, 2006).

Zur Beschreibung von außerfachlichen Kontexten haben van Vorst et al. (2015) ein Modell entwickelt, welches Kontextmerkmale basierend auf einem Literaturreview zusammenfasst und systematisiert. In diesem Zusammenhang differenzieren die Autorinnen und Autoren zwischen Merkmalen auf der Seite der Lernenden, Merkmalen auf Seiten des Kontextes und Merkmalen, die der Interaktion zwischen Lernenden und Kontext zuordenbar sind. Das so konzeptualisierte Modell ist in Abbildung 4 dargestellt.

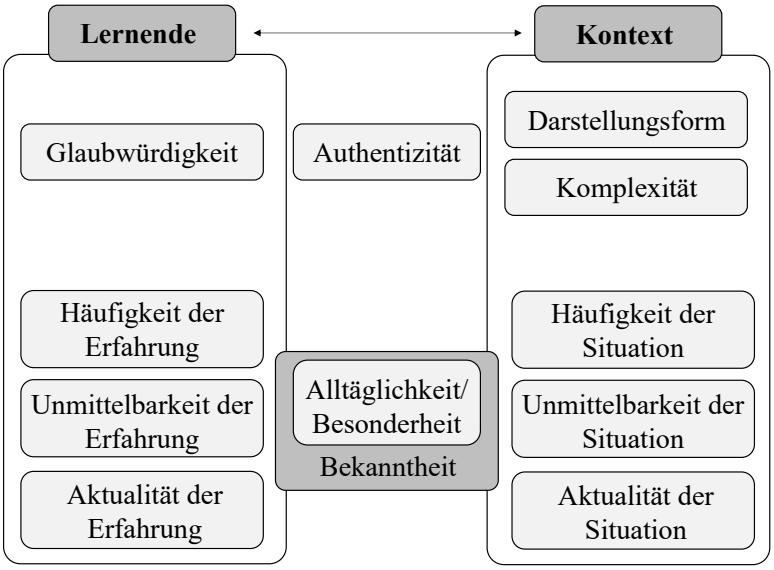


Abbildung 4: Modell zur Operationalisierung der Kontextmerkmale (van Vorst et al., 2015)

Das erste Kontextmerkmal, welches der Interaktion zwischen Lernenden und Kontext zugeordnet werden kann, ist das der *Authentizität*. Im Allgemeinen bezeichnet es die „Echtheit" oder „Glaubwürdigkeit" einer Person, eines Gegenstandes oder eines Ereignisses. Ein authentischer Kontext bezieht sich demnach auf realistische und echte Erfahrungen, welche die Lernenden tatsächlich durchleben könnten (Weiss & Müller, 2015). Aus diesem Grund wird die Authentizität aus der Perspektive der Lernenden besonders über die *Glaubwürdigkeit* des Kontextes bestimmt (Dorschu, 2013). Wenn die außerfachliche Situation oder der Gegenstand für die Lernenden unglaubwürdig ist, werden die Lernenden keinen für sie authentischen Kontext konstruieren (van Vorst et al., 2015). Weiterführend beeinflusst die *Darstellungsform* des Kontextes die Authentizität. Es ist davon auszugehen, dass besonders Darstellungsformen, wie sie von publizistischen Medien genutzt werden, die Glaubwürdigkeit eines Kontextes erhöhen und ihn damit authentischer erscheinen lassen (J. Kuhn, 2010). Authentische Darstellungsformen führen in aller Regel dazu, dass sich die *Komplexität* des Kontextes erhöht. Sie ist maßgeblich durch die Pluralität möglicher Lösungsansätze, die Informationsdichte sowie die Anzahl und Beziehung der Teilprobleme geprägt (Parchmann et al., 2008; van Vorst et al., 2015).

Die *Bekanntheit* eines Kontextes ergibt sich aus den Vorerfahrungen der Lernenden und der Vertrautheit mit der Situation. Die Bekanntheit ist eng verknüpft mit der Authentizität, da sich beide Facetten wechselseitig bedingen (van Vorst et al., 2015). Wie die Authentizität ist auch die Bekanntheit sowohl vom Kontext als auch von den Lernenden abhängig. Überdies

wird die Bekanntheit im Wesentlichen durch die Kontextmerkmale *Alltäglich* und die *Besonders* determiniert. Alltägliche Kontexte sind den Lernenden aus ihrem Alltag bekannt. Es handelt sich dabei um Situationen, Gegenstände oder Handlungen aus der Lebenswelt der Lernenden (Broman & Simon, 2015; Campbell & Lubben, 2000; George & Lubben, 2002). Demgegenüber stehen außergewöhnliche Situationen, Gegenstände oder Handlungen, die nicht oder nur äußerst selten in der Lebenswelt der Lernenden vorkommen. Derartige Kontexte lassen sich als *besonders* bezeichnen (Kasanda et al., 2005). Sie weisen dementsprechend ein geringes Maß an Bekanntheit auf. Die beiden Kontextmerkmale *Alltäglich* und *Besonders* sind als gegensätzliche Kategorien zu verstehen.

Auf Grundlage dieser Definitionsmerkmale ist anzunehmen, dass die Zugehörigkeit zu den Merkmalspaaren *Alltäglich* und *Besonders* sehr individuell und vor allem durch die Vorerfahrungen der Lernenden geprägt ist. Von übergeordneter Bedeutung ist in diesem Zusammenhang die *Häufigkeit*, in der bestimmte Situationen und Erfahrungen durchlebt wurden. Das häufige Erleben sollte hierbei zur mentalen Konstruktion eines alltäglichen Kontextes führen, wohingegen seltenes Erleben zu einem besonderen Kontext führt. Darüber hinaus werden die beiden Merkmale durch die *Unmittelbarkeit* der Erfahrung und Situation beeinflusst. Es ist anzunehmen, dass Primärerfahrungen zu einer höheren Bekanntheit führen als Sekundärerfahrungen (van Vorst et al., 2015). Situationen und Erfahrungen sind daneben durch einen gewissen Grad an *Aktualität* geprägt, die sich aus der Bedeutsamkeit für das gegenwärtige Geschehen ableitet. Die Aktualität ist demnach durch die öffentliche Berichterstattung in den Medien, anstehende Feiertage oder sonstige Ereignisse geprägt.

Der *Relevanz* eines Kontextes wird in der Literatur eine signifikante Bedeutung beigemessen. Dabei werden zwei unterschiedliche Perspektiven auf den Relevanzbegriff unterschieden (van Vorst et al., 2015). Das Kontextmerkmal ist einerseits eng verflochten mit der Kritik am traditionellen naturwissenschaftlichen Unterricht, die Relevanz der fachlichen Inhalte zu wenig zu akzentuieren (Gilbert, 2006). Durch den Einsatz von Kontexten soll die Bedeutsamkeit sowie die Notwendigkeit des Erlernens der fachlichen Inhalte herausgestellt werden (Bennett, 2016; Nentwig et al., 2007; Pilot & Bulte, 2006; Waddington, 2005). Andererseits ist neben der Relevanz des Fachinhalts, die durch den Kontexteinsatz akzentuiert werden soll, auch die Relevanz des Kontextes selbst von Bedeutung. De Jong (2006) unterscheidet in diesem Zusammenhang, ob ein Kontext aus dem persönlichen, dem sozial-gesellschaftlichen, dem beruflichen oder wissenschaftlich-technologischen Bereich stammt und der Kontext so auf einer oder mehreren Ebenen relevant wird. Daran anknüpfend verorten Stuckey et al. (2013) den Relevanzbegriff auf drei Ebenen: der individuellen Ebene, der gesellschaftlichen Ebene und der beruflichen Ebene. Innerhalb der PISA-Studie wird unterschieden, ob ein Kontext persönliche, regionale bzw. nationale oder globale Relevanz besitzt (Schiepe-Tiska, Rönnebeck et al., 2016). Kontexte aus dem persönlichen Bereich bzw. von persönlicher Relevanz lassen sich häufig dem Kontextmerkmal *Alltäglich* zuordnen. Van Vorst et al. (2015) stufen die Re-

levanz des Kontextes nicht als ein operationalisierbares Kontextmerkmal ein. Vielmehr handelt es sich bei der Relevanz um eine abhängige Variable, die durch den Kontext und seinen Merkmalen beeinflusst wird. Mit Bezug zu Keller (1987) wird die Relevanz neben dem Interesse, der Zufriedenheit sowie der Erwartung als eine Komponente zur Förderung der Motivation betrachtet. Daher wurde das Kontextmerkmal nicht in das Modell aufgenommen.

Als grundlegende Voraussetzung für die Integration von Kontexten in den naturwissenschaftlichen Unterricht wird außerdem die Adaptivität des jeweiligen Kontextes genannt (van Vorst, 2013). Bei der Auswahl von Kontexten für den Unterricht sind demzufolge die Charakteristika des Fachinhalts sowie die individuellen Lernvoraussetzungen der Lernenden zu berücksichtigen (George & Lubben, 2002). Bislang wurde bei der Konstruktion von kontextualisierten Aufgaben insbesondere die Passung des Kontextes zu den korrespondierenden Fachinhalten fokussiert. Weniger beachtet wurde dagegen die Passung der Kontexte zu den individuellen Lernvoraussetzungen der Schülerinnen und Schüler. Insbesondere unter Betrachtung der Selbstbestimmungstheorie sowie der zugehörigen Evidenz (siehe Kapitel 3.2) erscheint dieses Vorgehen unzureichend, da der Passung zu den Lernvoraussetzungen eine übergeordnete Bedeutung für den Lernprozess zuteilwird. Aus diesem Grund widmet sich die vorliegende Arbeit explizit der Untersuchung des Kontextmerkmals der Adaptivität.

4.3 Unterrichtliche Gestaltung von kontextorientiertem Lernen

Kontextorientierte Unterrichtsansätze unterscheiden sich hinsichtlich der Bedeutung des Kontextbegriffs und auch im Hinblick auf die eingesetzten Kontexte; sie lassen sich aber auf ein gemeinsames konzeptionelles Fundament reduzieren. Angelehnt an die deutsche Konzeption *Chemie im Kontext* kann zwischen verschiedenen Unterrichtphasen in einem kontextorientierten Lernprozess unterschieden werden (Nentwig et al., 2007; Parchmann et al., 2008).

Ausgangspunkt für die Entwicklung einer kontextorientierten Lerneinheit stellt die Auswahl eines geeigneten Kontextes dar. Dieser sollte vorzugsweise ausgehend von den im Lehrplan formulierten Kompetenzerwartungen beziehungsweise den zu erarbeitenden Basiskonzepten festgelegt werden, um den kumulativen Aufbau von fachlichem Wissen und der Basiskonzepte zu gewährleisten (Parchmann et al., 2008). Basiskonzepte organisieren als allgemeine Prinzipien den kumulativen Wissensaufbau und strukturieren das fachliche Wissen (Parchmann et al., 2006). Die Auswahl des Kontextes sollte sich hierbei nicht nur am Lehrplan orientieren, sondern auch die Motivierung der Lernenden in den Blick nehmen. Trotzdem ist sicherzustellen, dass der fachliche Bezug nicht in den Hintergrund tritt (Parchmann et al., 2008).

In der Phase der Begegnung werden die Lernenden mit dem Kontext vertraut gemacht. Dabei werden die Lernenden mit der Problemstellung oder den inhaltlichen Zusammenhängen des

Kontextes konfrontiert. In der Begegnungsphase können etwa Filmsequenzen, Zeitungsartikel oder auch Bilder eingesetzt werden, um den Kontext einzuführen (Parchmann et al., 2008). Empirische Ergebnisse deuten darauf hin, dass für die Lernenden an dieser Stelle mitunter nicht klar abgrenzbar ist, was Kontext und was fachlicher Inhalt ist (Pilot & Bulte, 2006). So kann Wasser beispielsweise sowohl als Fachinhalt als auch als Kontext angesehen werden (Kölbach & Sumfleth, 2013). Fachinhalte können u. a. Konzepte, Prinzipien oder auch Heuristiken sein, die zur Lösung fachlicher Probleme herangezogen werden. Sie umfassen damit bereichsspezifisches Wissen (Barnett & Ceci, 2002).

Im nächsten Schritt sollen die Lernenden in einer Phase der Neugier relevante Fragestellungen aus dem Kontext entwickeln. Damit stellt sie einerseits ein wichtiges Bindeglied zwischen den Vorerfahrungen der Lernenden und der nachfolgenden Bearbeitung der Aufgabe dar, andererseits fokussiert sie problemlösendes Arbeiten (Parchmann et al., 2008; Nentwig et al., 2007). Die Lehrkraft sollte im Unterrichtsverlauf mögliche Fragestellungen der Schülerinnen und Schüler systematisieren und mit vorab geplanten Unterrichtsinhalten verknüpfen (Parchmann et al., 2008).

Die Phase der Erarbeitung fokussiert die selbstständige Bearbeitung des aufgeworfenen Problems durch die Lernenden (Parchmann et al., 2006). Das zentrale Paradigma des kontextorientierten Lernens ist dabei, dass die fachlichen Konzepte situiert innerhalb des Kontextes in der Erarbeitungsphase erworben werden. Als eine der zentralen Theorien zur Entwicklung von ChiK werden deshalb die Forschungsarbeiten zum situierten Lernen genannt (Nentwig et al., 2007). Der Konsens dieser Theorien ist, dass der Wissenserwerb stets an die Lernsituation gebunden ist, in der er stattgefunden hat. Die Lernenden müssen deshalb beim Transfer des Wissens unterstützt werden. Im Lernprozess wechseln die Lernenden zwischen der Kontext- und Konzeptebene, was King und Ritchie (2013) als *fluid transition* bezeichnen. Die wiederholte Dekontextualisierung und Rekontextualisierung des fachlichen Wissens unterstützt hierbei die Konzeptbildung der Lernenden. Fachliche Konzepte sollen in dieser Phase deshalb in unterschiedlichen Kontexten mit steigendem fachlichem Niveau aufgegriffen werden, um den kumulativen Aufbau von Wissen zu fördern. Dies wird nach Bennett et al. (2005) als *drip-feed* bezeichnet und orientiert sich an dem Konzept des Spiralcurriculums von Bruner (2009). Zur Erarbeitung werden vornehmlich vielfältige Lehrmethoden eingesetzt, welche Lernenden zur aktiven Auseinandersetzung mit dem Kontext anregen sollen und sich als *student-centred* bezeichnen lassen. Dies wird u. a. durch Diskussionen in Kleingruppen oder mit Problemlöseaufgaben erzielt (Bennett et al., 2005; Parchmann et al., 2006). Insgesamt sind Lernumgebungen gefordert, die selbstständiges Arbeiten ermöglichen.

Die letzte Phase widmet sich der Abstraktion und Vernetzung des kontextualisiert erworbenen Wissens. Durch die stärkere Ausrichtung an lebensweltlichen Anwendungsbeispielen muss der Transfer des Wissens unterstützt werden (Parchmann et al., 2008). Für den Transfer

in eine neue Lernsituation ist es notwendig, erworbene fachliche Konzepte zu dekontextualisieren, um sie in einer neuen Lernsituation erneut anzuwenden. Dazu können nach Parchmann et al. (2008) sogenannten Brückenaufgaben genutzt werden, welche die Dekontextualisierung fördern, um ein vorwiegend anwendbares Wissensfundament zu entwickeln.

4.4 Empirische Befunde zum kontextorientierten Lernen

Bisher wurde durch eine Vielzahl von Studien untersucht, inwieweit die Ziele kontextorientierter Unterrichtsansätze tatsächlich erreicht werden. Eine Mehrheit der durchgeführten Studien weist allerdings methodische Mängel auf, sodass sowohl die interne als auch die externe Validität der Studien eingeschränkt sind. So kann in Literaturreviews und Metaanalysen auch häufig nur eine geringe Anzahl an Studien berücksichtigt werden, da in vielen Studien beispielsweise keine Kontrollgruppe in das Untersuchungsdesign implementiert ist oder Messinstrumente genutzt werden, die hinsichtlich der Testgütekriterien nicht analysiert wurden (Bennett et al., 2007; Taasoobshirazi & Carr, 2008). Kontextorientierter Unterricht wird dabei fast ausschließlich mit traditionellem Unterricht (conventional courses) kontrastiert. Ültay und Çalık (2012, S. 687) beschreiben diesen Stereotypen wie folgt: „In traditional chemistry instruction, teachers write on the blackboard, students listen, memorize the facts or rules and answer when they are asked (concisely, chalk and talk)". Aufgrund unzureichender Beschreibungen der Vergleichsgruppe in den Untersuchungen bleibt häufig jedoch ungeklärt, wie der Unterricht in der Vergleichsgruppe tatsächlich gestaltet wurde. Demzufolge wird nicht ersichtlich, womit der kontextorientierte Unterricht verglichen wird. Vor dem Hintergrund dieser Einschränkungen wird die Forschungslage nachfolgend dargelegt.

4.4.1 Effektivität von kontextorientiertem Lernen

Ein zentrales Anliegen vieler Evaluationsstudien ist es, die Effekte kontextorientierter Unterrichtsansätze auf affektive Merkmale aufzuklären, da angenommen werden kann, dass auch die Beeinflussung von kognitiven Merkmalen (z. B. dem Lernzuwachs) über affektive Merkmale vermittelt wird. Fechner (2009) konnte in diesem Zusammenhang für Schülerinnen ein Modell bestätigen, in dem der Effekt zwischen der Kontextorientierung und dem Lernzuwachs über das situationale Interesse mediiert wird.

Die meisten Übersichtsarbeiten zeigen positive Effekte auf affektive Merkmale durch kontextorientiertes Lernen im Vergleich zu traditionellem Unterricht (Bennett et al., 2007; Taasoobshirazi & Carr, 2008; Ültay & Çalık, 2012). In dem Review von Bennett et al. (2007) für die Naturwissenschaftsdidaktik verweisen sieben von neun Studien darauf, dass die Einstellung gegenüber Naturwissenschaften allgemein und bezogen auf die Schule durch kontextorientiertes Lernen verbessert wird. Lediglich die Studien von Ben-Zvi (1999) sowie Wierstra und Wubbels (1994) deuten auf negative Effekte hin. Auf Basis der Daten von drei

der neun Studien, in denen eine angemessene statistische Analyse durchgeführt wurde, kann ein mittlerer Effekt von $d = 0.67$ berichtet werden. Weiterhin deuten zahlreiche empirische Ergebnisse auf eine interessen- und motivationsfördernde Wirkung kontextorientierter Lernumgebungen beim Chemielernen hin (King, 2012). Die Resultate der wissenschaftlichen Begleitung des Salters-Projektes wurden durch Bennett und Lubben (2006) zusammengefasst und zeigen, dass Salters-Kurse im tertiären Bildungswesen häufiger aufgrund von Interesse gewählt werden als traditionelle Chemiekurse. Salters-Kurse nutzen im Gegensatz zu traditionellen Chemiekursen eine sogenannte *chemical storyline,* um den Lernprozess zu gestalten. Die Ergebnisse einer unveröffentlichten Masterarbeit verdeutlichen hierbei sowohl durch qualitative als auch quantitative Daten positive Auswirkungen auf das Interesse der Lernenden durch den Salters-Kurs (Barber, 2001, zitiert nach Bennett & Lubben, 2006). Innerhalb des ChiK-Projektes hat sich gezeigt, dass die Motivation der ChiK-Gruppe weniger stark abnahm als die Motivation der Lernenden, die an einem traditionellen Unterricht teilgenommen hatten. Weiterhin war das Interesse der Lernenden aus dem ChiK-Projekt nach zwei Jahren signifikant höher als das der Vergleichsgruppe (Parchmann et al., 2006). Die Befragung von 495 Lernenden in Schweden ergab, dass fast 60 % der Lernenden überzeugt waren, Lehrkräfte könnten durch eine stärkere Ausrichtung am alltäglichen Leben das Fach Chemie interessanter und relevanter gestalten (Broman & Simon, 2015). Weitere interessenförderliche Ergebnisse werden auch von Ültay und Çalık (2012) für den Chemieunterricht bestätigt. In Bezug auf das kontextorientierte Lernen im Physikunterricht zeigen Taasoobshirazi und Carr (2008) ähnliche Wirkzusammenhänge. Ihre Ergebnisse geben insbesondere Hinweise darauf, dass die Relevanz der Inhalte durch kontextorientierte Lernumgebungen erhöht werden kann.

Kontextorientierte Unterrichtsansätze müssen sich häufig der Kritik stellen, den Fokus auf fachliche Inhalte aufgrund des starken lebensweltlichen Bezugs des Unterrichts aufzugeben (Sevian & Talanquer, 2014). Unter Einbezug der empirischen Ergebnisse von kontextorientiertem Lernen auf kognitive Merkmale kann diese Kritik allerdings als unberechtigt zurückgewiesen werden. Im Vergleich zu Lernenden, die an einem traditionell ausgerichteten naturwissenschaftlichen Unterricht teilgenommen haben, kann eine vergleichbare Lernleistung im kontextorientierten Unterricht ausgemacht werden. Wenige Studien zeigen sogar eine verbesserte Lernleistung. Basierend auf den Ergebnissen von *t*-Tests können für die Studien von Winthur und Volk (1994) und Yager und Weld (1999) Effektstärken von $d = 0.63$ und $d = 1.52$ berechnet werden (Bennett et al., 2007). Dieses eher inkonsistente Bild zieht sich auch durch andere Übersichtsarbeiten. So berichtet King (2012) von drei Studien, die auf vergleichbare Effekte durch kontextorientiertes Lernen verweisen und von vier Studien, die einen Vorteil anzeigen. Bemerkenswert ist zudem, dass selbst Studien, die das gleiche Kontextprojekt begleiteten, auf divergente Ergebnisse hindeuten. In Bezug auf das ChiK-Projekt können Lange und Parchmann (2003) keinen signifikanten Unterschied in der Lernleistung identifizieren. In einer weiteren Untersuchung stellen Parchmann und Schmidt (2003) jedoch fest, dass ChiK-Lernende nach dem kontextorientierten Unterricht Beobachtungen sehr viel

häufiger durch das Teilchenmodell erklärten als die Vergleichsgruppe. Die ChiK-Gruppe hat in einem Fachwissenstest außerdem etwas besser abgeschnitten (Parchmann et al., 2006). Es ist zu vermuten, dass diese Ergebnisse teilweise auf unterschiedliche methodische Zugänge zurückzuführen sind; sie illustrieren dennoch die Kontroverse hinsichtlich der Lernleistung.

Fehlende Effekte auf die Lernleistung sind möglicherweise vor dem Hintergrund des unzureichenden Transfers des erworbenen Wissens aus der kontextorientierten Lernumgebung zu erklären. Es lassen sich verschiedene Evidenzen anführen, die auf Schwierigkeiten der Lernenden hindeuten, situiert erworbenes Wissen zu übertragen (Gilbert et al., 2011). Kehne (2019) hat in diesem Zusammenhang in einer quasi-experimentellen Studie im Chemieunterricht untersucht, von welcher Art der Dekontextualisierung die Lernenden im Lernprozess profitieren. Die Autorin hat drei Treatmentgruppen miteinander verglichen: In der ersten Gruppe wurden die Lernenden bei der Arbeit in kontextorientierten Lerneinheiten dazu aufgefordert, die verschiedenen Kontexte miteinander zu vergleichen. Die zweite Gruppe arbeitete abwechselnd in dekontextualisierten und kontextualisierten Lerneinheiten und wurde instruiert, die entsprechenden Konzepte aus den Kontexten herauszulösen. Die letzte Gruppe sollte das zugrundeliegende Schema (z. B. das Donator-Akzeptor-Prinzip) aus den kontextualisierten Lerneinheiten herauslösen. Durch die Auswertung der quantitativen und qualitativen Daten zeigt sich, dass Lernende in ihrer Transferleistung besonders von einer impliziten Dekontextualisierung durch den Vergleich multipler Quellkontexte profitieren. Dies gilt besonders für kognitiv schwächere Schülerinnen und Schüler.

Zusammenfassend fallen die empirischen Belege hinsichtlich der Wirkung des kontextorientierten Lernens auf affektive Merkmale relativ eindeutig aus: Kontextorientiertes Lernen hat einen positiven Einfluss auf das Interesse und die Motivation im naturwissenschaftlichen Unterricht. Weiterführend scheint auch die Einstellung gegenüber Naturwissenschaften insgesamt von der Kontextorientierung zu profitieren. Gleichermaßen gewinnen ebenso die fachlichen Inhalte an Relevanz für die Lernenden. Es können allerdings nur wenige Nachweise für eine gesteigerte Lernleistung durch kontextorientiertes Lernen angeführt werden. Der Konsens in der Wissenschaft ist vielmehr, dass Lernende im Unterricht durch die Kontextorientierung hinsichtlich der Lernleistung nicht benachteiligt werden. In diesem Zusammenhang stellt sich insbesondere die Frage, ob die in den Studien verwendeten Testinstrumente zur Erfassung des Fachwissenszuwachses hinsichtlich der im Vergleich zu traditionellem Unterricht veränderten Ziele im kontextorientierten Lernen valide sind.

4.4.2 Einfluss der Kontextmerkmale

In Anbetracht der dargelegten Forschungslage scheint es unumstritten, dass die Implementation von kontextorientierten Lernumgebungen in den Unterricht ertragreich sein kann – besonders im Hinblick auf affektive Variablen. Entscheidend ist hier die Ausgestaltung der

Lernmaterialien in der Lernumgebung (Pilot & Bulte, 2006; Prins et al., 2018). Der Einsatz von unterschiedlichen Kontexten scheint diesbezüglich zu unterschiedlichen Lernergebnissen zu führen. Sevian, Hugi-Cleary et al. (2018) haben zwei Hochschulkurse miteinander verglichen, die durch zwei unterschiedliche Kontexte in die kinetische Gastheorie einführten. Im ersten Kurs wurden die Lernenden mit Ballons konfrontiert, die bei gleicher Temperatur und gleichem Druck eine unterschiedliche Masse zeigten. Zur Identifizierung der Gase in den Ballons führte der Kurs eine kinästhetische Aktivität durch, bei der sie selbst als menschliches Modell eines Gases dienten. Die Lernenden aus dem zweiten Kurs wurden dagegen gebeten, einen Ansatz zur Verringerung der CO_2-Konzentration in der Atmosphäre zu entwerfen. Dazu konnten die Lernenden mit Molekulardynamiksimulationen arbeiten. Die Studierenden aus dem ersten Kurs zeigten nach der Intervention ein höheres Verständnis für die Partikelbewegungen, während die Studierenden aus dem zweiten Kurs ein differenzierteres mechanistisches Denken sowie eine höhere Transferleistung durch die kohärente Nutzung der chemischen Fachsprache aufwiesen. Ähnliche Ergebnisse konnten Podschuweit und Bernholt (2018) für Lernende aus der Sekundarstufe I identifizieren. Sie untersuchten, ob Schülerinnen und Schüler von einem experimentellen Lernsetting profitieren, welches heterogen oder homogen kontextualisiert ist. In dem homogen kontextualisierten Lernsetting lernten die Schülerinnen und Schüler mit Kontexten, die sich alle der Kategorie Kraftwerke zuordnen ließen. Die heterogen kontextualisierte Lernumgebung nutzte dagegen Kontexte, die keiner gemeinsamen Kategorie zurechenbar waren. Im Gegensatz zur homogen kontextualisierten Lernumgebung, die nur Inhalte der Physik beinhaltete, wurden hier auch Inhalte der Biologie und Chemie aufgegriffen. Die Autoren nehmen an, dass variierte Kontexte in einer Lernsituation geeigneter sind, um den Bedarfen unterschiedlicher Lernenden gerecht zu werden, statt nur Kontexte einer einzigen Kategorie zu nutzen. Das Lernen in der heterogenen Gruppe von Kontexten führt dabei zu einer gesteigerten Transferleistung. In einer weiteren qualitativen Studie wurden halb-strukturierte Interviews mit Schülerinnen und Schülern der Jahrgangsstufen 10 bis 12 durchgeführt, um mithilfe der Methode des lauten Denkens nähere Einblicke in den Problemlöseprozess von kontextbasierten Aufgaben zu erhalten, die hinsichtlich des Themas und des Kontextes variiert wurden. Die qualitative Auswertung ergab, dass sich die Antworten der Lernenden beim Problemlösen in Abhängigkeit von Thema und Kontext unterscheiden (Broman & Parchmann, 2014). Die aufgeführten Untersuchungen machen damit deutlich, dass ein differenzierter Blick auf die genutzten Kontexte notwendig ist, wobei allerdings nur unzureichend bestimmt wird, welche Kontexte wirksamer sind als andere (Podschuweit & Bernholt, 2018; Sevian, Dori & Parchmann, 2018; Sevian, Hugi-Cleary et al., 2018; Taasoobshirazi & Carr, 2008).

Erste Hinweise liefern in diesem Zusammenhang einige Large-Scale-Studien, die die Interessantheit von Kontexten für den Unterricht in den Naturwissenschaften untersuchten. Als eine der ersten Studien in diesem Feld lässt sich die Interessenstudie des *Leibniz-Instituts für die Pädagogik der Naturwissenschaften und Mathematik* (IPN) nennen, die in den 1980er Jahren

durchgeführt wurde. Für das Fach Chemie wurden diesbezüglich Daten im Jahr 1990 und 2008 erhoben, die Gräber (2011) in seiner Publikation gegenüberstellt. Für die Interessenbefragung wurden Kontexte thematisch zusammengefasst. Es zeigen sich bedeutende Geschlechterunterschiede: Mädchen etwa interessieren sich dafür, wie Chemie im Haushalt genutzt werden kann, während Jungen ein hohes Interesse an Kontexten aus dem Bereich Technik angeben. Beide Geschlechter interessieren sich gleichermaßen für Anwendungen von Chemie, die gefährlich oder in der Zukunft nützlich sein können. Etwas konkreter wird es in der international durchgeführten Studie *The Relevance of Science Education* (ROSE). Hier wurden Lernende aus knapp 40 Nationen hinsichtlich ihrer affektiv-motivationalen Einstellung gegenüber Naturwissenschaften und naturwissenschaftlichem Unterricht befragt (Sjøberg & Schreiner, 2010). Die in Deutschland und Österreich erhobenen Daten von 1247 Lernenden der Sekundarstufe I wurden hinsichtlich des Interesses an naturwissenschaftlichen Inhalten und Kontexten von Elster (2007) analysiert. Auch hier wurden Kontexte wieder thematisch gruppiert. Bedeutende Geschlechterunterschiede zeigen sich am Themenbereich Fitness. Hier geben Mädchen im Vergleich zu Jungen ein sehr viel höheres Interesse an. Allerdings interessieren sich sowohl Mädchen als auch Jungen für Kontexte aus den Themenbereichen Jugend und Gesundheit. Es ist auffällig, dass es sich dabei vornehmlich um Kontexte mit einem Bezug zur Medizin handelt (exemplarisches Item: Was wir über Krebs wissen und wie man ihn behandeln kann). Ein ähnlicher Trend zeichnet sich in den Daten der PISA-Studie aus dem Jahr 2015 ab. Gemäß den Ergebnissen interessieren sich Lernende insbesondere dafür, wie Naturwissenschaften helfen können, Krankheiten zu verhindern. Außerdem werden Kontexte zum Universum und seine Geschichte als interessant bewertet. Dies gilt für Jungen und Mädchen in Deutschland sowie im gesamten OECD-Durchschnitt (Schiepe-Tiska, Simm & Schmidtner, 2016).

Im Zentrum der angeführten Large-Scale-Untersuchungen stehen damit vor allem Themenbereiche, die spezifische Kontexte beinhalten. Auf Grundlage dieser Untersuchungen können allerdings nur wenige Erkenntnisse darüber gewonnen werden, welche übergeordneten Kontextmerkmale geeigneter für das Lernen im Fach Chemie sind. Fechner (2009) hat dieses Desiderat aufgegriffen, indem sie die Effekte außerfachlicher und innerfachlicher Kontexte in einer experimentellen Studie kontrastierte. Die Lernenden arbeiteten dabei kooperativ an fünf experimentellen Aufgaben zum Fachinhalt Säuren und Basen, die in problemorientierte Kontexte eingebettet waren. In Bezug auf ihr situationales Interesse profitierten die Lernenden über alle fünf Aufgaben hinweg von den außerfachlichen Kontexten verglichen mit den innerfachlichen Kontexten. Schülerinnen und Schüler, die mit den außerfachlichen Kontexten lernten, zeigten außerdem einen höheren Lernzuwachs am Ende der Intervention. In einer weiteren experimentellen Studie wurde anschließend untersucht, ob die positiven Effekte des außerfachlichen Kontextes auch unabhängig vom fachlichen Inhalt in individuellen Lernphasen gelten (Kölbach & Sumfleth, 2013). Die Autorinnen konnten einen empirischen Beleg für die interessenförderlichen Effekte des außerfachlichen Kontextes erbringen: Das Lernen mit

einem außerfachlichen Kontext führt unabhängig vom fachlichen Inhalt zu einem höheren situationalen Interesse am Ende der individuellen Lerneinheit mit Lösungsbeispielen. Beim Vergleich der Effektstärken der variierten Fachinhalte ist allerdings auffällig, dass der Effekt des außerfachlichen Kontextes umso größer ausfällt, je größer die Distanz zwischen Fachinhalt und Kontext ist. Im Gegensatz zu Fechner (2009) konnten Kölbach und Sumfleth (2013) keinen Effekt auf den Lernzuwachs identifizieren. Broman und Kollegen (2018, 2020) kamen durch zwei qualitativ angelegte Studien zu vergleichbaren Ergebnissen, indem sie den Lernenden Kontexte zur Wahl anboten, die aus dem persönlichen, gesellschaftlichen oder beruflichen Bereich stammten. 30 von 40 Lernenden wählten einen Kontext aus dem persönlichen Bereich, d. h. einen Kontext mit einem starken lebensweltlichen Bezug. Lediglich fünf Lernende entschieden sich für einen Kontext aus dem beruflichen Bereich, der eher innerfachlich orientiert war (Broman et al., 2018). Dieses Ergebnis konnte auch in einer größeren Stichprobe repliziert werden (Broman et al., 2020).

Van Vorst et al. (2018) untersuchten erstmals differenzierter die Wirkung von außerfachlichen Kontexten ausgehend von den Merkmalen *Alltäglich* und *Besonders* sowie *Aktualität*. Die Ergebnisse zeigen, dass die Schülerinnen und Schüler gegenüber besonderen Kontexten eine höhere emotionale Valenz – als Teilfacette des situationalen Interesses – empfinden als gegenüber alltäglichen Kontexten. Das Kontextmerkmal Aktualität zeigt keinen Effekt auf die emotionale Valenz der Lernenden. Ferner beeinflusst keines der Kontextmerkmale die empfundene wertbezogene Valenz des situationalen Interesses. Diese Ergebnisse müssen allerdings unter der Prämisse betrachtet werden, dass die Lernenden ihr Interesse an den Kontexten lediglich aufgrund von Kontextbeschreibungen angaben. Habig et al. (2018) untersuchten aus diesem Grund in einem Experimentaldesign, welchen Einfluss u. a. das Kontextmerkmal in drei sequenzierten experimentellen Lernaufgaben hat. Hier zeigen sich gegensätzliche Effekte des Kontextmerkmals: Die emotionale Valenz der Lernenden, die in einem alltäglichen Kontext lernten, nimmt im Vergleich zum besonderen Kontext weniger stark ab. Eine tiefergehende Datenanalyse verweist auf einen Moderationseffekt des individuellen Interesses und der Chemienote: Schülerinnen und Schüler mit einem hohen Interesse und einer guten Chemienote profitieren hinsichtlich der emotionalen Valenz von besonderen Kontexten. Dementgegen empfinden Lernende mit einem geringeren Interesse und einer schlechteren Chemienote eine höhere emotionale Valenz beim Lernen mit alltäglichen Kontexten. Folglich scheinen Lernende in Abhängigkeit von ihren individuellen Merkmalen unterschiedliche Kontextaufgaben für das Lernen im Fach Chemie zu präferieren. Diese Hypothese wurde näher von van Vorst und Aydogmus (2021) untersucht, indem sie Schülerinnen und Schülern aus der neunten Jahrgangsstufe Aufgaben aus den Themenbereichen Wellness und Urlaub, Gesundheit sowie Umwelt und Technik vorlegten, aus denen eine Aufgabe zur Bearbeitung ausgewählt werden konnte. Mit zunehmendem Interesse und einer besseren Chemienote wählten die Lernenden häufiger einen besonderen Kontext zum Lernen aus. Eine hierarchische Clusteranalyse verweist daneben auf drei unterschiedliche Gruppen von Lernenden, die

Aufgaben mit unterschiedlichen Kontextmerkmalen aufgrund verschiedener Motive wählten: Die Lernenden der ersten Gruppe entschieden sich häufiger für besondere Kontexte (72.5 %) aufgrund von überraschenden Informationen sowie Neugier und Interesse. Die zweite Gruppe zeigte eine leichte Tendenz zur Wahl alltäglicher Kontexte (53.1 %) und gab als Wahlmotiv den persönlichen Bezug an. Die dritte Lernendengruppe bewertete alle Wahlmotive deutlich niedriger als die übrigen Cluster und zeigte tendenziell keine Präferenz für Kontexte mit einem bestimmten Merkmal. Im Anschluss an die Bearbeitung der gewählten kontextualisierten Aufgabe zeigte sich zudem ein höherer Anteil von Schülerinnen und Schülern des dritten Clusters unzufrieden mit der gewählten Aufgabe und würde retrospektiv keinen der zur Wahl gestellten Kontexte bearbeiten. Im Vergleich dazu war ein höherer Anteil der Lernenden aus dem ersten und zweiten Cluster zufrieden mit der gewählten Kontextaufgabe. Schülerinnen und Schüler des ersten und zweiten Clusters begründeten ihre Antwort durch affektive Aufgabenmerkmale (z. B. das Interesse an der Aufgabe), während Lernende des dritten Clusters vornehmlich kognitive Begründungen (z. B. die hohe Aufgabenschwierigkeit) angaben. Es ist folglich anzunehmen, dass sowohl das situationale Interesse an der Aufgabe und die kognitive Belastung im Lernprozess die Zufriedenheit nach der Aufgabenbearbeitung beeinflussen.

4.5 Fazit zum Einsatz von Kontexten zur interessenbasierten Differenzierung

Ein besonderer Fokus der chemiedidaktischen Forschung besteht bislang auf der Untersuchung der Wirksamkeit von kontextorientiertem Lernen. Obwohl das Modell zur Operationalisierung der Kontextmerkmale von van Vorst et al. (2015) einen ersten theoretischen Zugang liefert, um die Effekte einzelner Kontextmerkmale auf das Chemielernen zu untersuchen, stehen nur vereinzelt empirische Ergebnisse zur Verfügung. Insbesondere das Kontextmerkmal der Adaptivität wird von van Vorst et al. (2015) zwar angeführt, allerdings münden bisherige Überlegungen zur Adaptivität in der Passung zwischen Kontext und Fachinhalt. Kaum beachtet ist die Adaptivität einzelner Kontexte für unterschiedliche Gruppen von Lernenden. Einige Large-Scale-Studien konnten in diesem Zusammenhang bedeutende Geschlechtereffekte aufklären, die sich allerdings vornehmlich auf die Themen der Kontexte beschränken und übergeordnete Kontextmerkmale kaum in den Blick nehmen. Dieses Bild zieht sich auch durch die weiterführende Forschung: In einzelnen Studien werden unterschiedliche Kontexte zwar eingesetzt, allerdings werden diese häufig nur unsystematisch oder lediglich in Bezug auf spezifische Themen (z. B. Broman et al., 2020; Podschuweit & Bernholt, 2018; Sevian, Hugi-Cleary et al., 2018). Auch hier werden übergeordnete Kontextmerkmale kaum fokussiert. In diesem Zusammenhang mangelt es außerdem an einer hinreichenden Charakterisierung der Lernenden, um festzuhalten, welche Lernenden(-gruppen) von welchen Kontexten profitieren. Habig et al. (2018) konnten hier erstmalig herausarbeiten, dass unterschiedliche

Lernende im Hinblick auf die emotionale Valenz des situationalen Interesses von unterschiedlichen Kontexten profitieren. Diese Effekte scheinen sich auch in der Kontextwahl der Lernenden widerzuspiegeln, wie die Studie von van Vorst und Aydogmus (2021) zeigt. Insgesamt deutet sich damit an, dass sich der Einsatz von variierten Kontexten zur interessenbasierten Differenzierung eignen könnte. Ein besonderes Forschungsdesiderat besteht allerdings nach wie vor in der Untersuchung der Adaptivität spezifischer Kontexte sowohl für die Lernenden individuell als auch für einzelne Lernendengruppen im Kollektiv. Die Schließung dieser Forschungslücke ist jedoch notwendig, um einen interessenbasierten Differenzierungsansatz zu entwickeln und zu evaluieren, wie die Selbstbestimmungstheorie und die darauf aufbauende Forschung zeigen.

5 Ziel, Forschungsfragen und Hypothesen

Die individuelle Förderung der Lernenden ist eine politische Forderung, die sich in der schulischen und wissenschaftlichen Praxis in unterschiedliche Ansätze ausdifferenziert, welche sich zu einem großen Teil an dem Leistungsniveau der Lernenden orientieren. Es werden Lernmaterialien mit unterschiedlichen Anforderungsniveaus bereitgestellt, mit dem Ziel, den Lernprozess adaptiv zu gestalten. Obwohl in der Literatur auch die Anpassung des Lernmaterials in Hinblick auf das Thema der Lerninhalte beschrieben wird, um unterschiedliche Interessen der Lernenden zu adressieren (z. B. Tomlinson et al., 2003), wurde ein interessenbasierter Differenzierungsansatz in der empirischen Forschung bislang selten aufgegriffen. Dies gilt sowohl für die Unterrichtsforschung allgemein als auch für die chemiedidaktische Forschung im Speziellen.

Die Konzeptualisierung von interessenbasierten Differenzierungsansätzen durch die Wahl von Aufgaben mit variierten Themen, deutet infolge der Erkenntnisse der pädagogisch-psychologischen Forschung auf positive Lerneffekte hin (u. a. Patall et al., 2008). Die dargelegten Erkenntnisse in Bezug auf die Effektivität des kontextorientierten Lernens und zum Einfluss der Kontextmerkmale belegen, dass Lernende von unterschiedlichen Kontexten profitieren und unterschiedliche Kontexte zum Lernen auswählen. Dabei scheint das vorgeschlagene Modell zur Operationalisierung der Kontextmerkmale von van Vorst et al. (2015) einen geeigneten theoretischen Rahmen zu eröffnen, um Kontexte systematisch zu variieren. Insbesondere die Differenzierung zwischen außerfachlichen und innerfachlichen sowie alltäglichen und besonderen Kontexten bietet vielversprechende Anknüpfungspunkte zur interessenbasierten Differenzierung im Fach Chemie. Bis zu diesem Zeitpunkt wurde ein derartiger Ansatz noch nicht durch die empirische Forschung verfolgt. Daher ist es das Ziel der vorliegenden Arbeit, einen interessenbasierten Differenzierungsansatz durch systematisch variierte Kontextaufgaben im Fach Chemie zu entwickeln und Effekte auf lernrelevante Variablen aufzuklären.

Die Selbstbestimmungstheorie, Erwartungs-Wert-Modelle sowie die empirischen Evidenzen der Forschung zum Wahlverhalten verweisen auf die Bedeutsamkeit der Passung zwischen den Wahloptionen und Interessen, Werten und Zielen – oder etwas weiter gefasst – den Merkmalen einer Person. In Bezug auf das kontextorientierte Lernen ist allerdings nicht abschließend geklärt, welche Kontexte mit welchen Kontextmerkmalen zu unterschiedlichen Lernenden passen. Die bisherigen Studien sind besonders limitiert in Bezug auf die Charakterisierung der Lernenden, die einen bestimmten Kontext zum Lernen auswählen. Hier wurden die Merkmale der Lernenden entweder nicht oder nur in limitierter Anzahl berücksichtigt. Diese Situation erschwert mit Blick auf die Adaptivität der Kontexte die Auswahl von sowohl individuell wie für bestimmte Lernendengruppen geeigneten Kontexten.

Zudem gibt es Hinweise darauf, dass einige Lernende nach der Wahl und der Bearbeitung der Kontextaufgabe unzufrieden waren, was als Indikator für eine unzureichende Passung interpretiert werden kann. Hier scheinen sowohl das situationale Interesse als auch die kognitive Belastung mit der Zufriedenheit nach der Aufgabenbearbeitung in Verbindung zu stehen. Demzufolge erscheint es notwendig, sowohl die nach der Bearbeitung einer selbstgewählten Kontextaufgabe angegebene Zufriedenheit als auch das situationale Interesse und die empfundene kognitive Belastung zu berücksichtigen, um die Passung zwischen Person und Kontext zu überprüfen. Daher wird das Forschungsvorhaben in der ersten Teilstudie durch folgende Forschungsfragen geleitet:

FF1: Welche Lernendengruppen lassen sich hinsichtlich ihrer Personenmerkmale bei der Kontextwahl unterscheiden und welchen Kontext wählen diese zum Lernen?

FF2: Wie bewerten die Lernenden ihre Wahlentscheidung nach der Bearbeitung der Aufgabe im Hinblick auf Zufriedenheit, situationales Interesse und kognitive Belastung?

Auf Grundlage der bestehenden Forschungslage (z. B. Habig et al., 2018; van Vorst & Aydogmus, 2021) kann angenommen werden, dass sich Lernende bei der Kontextwahl im Hinblick auf ihre Leistung in Chemie (H1.1), ihr Interesse an Chemie (H1.2) und ihre Wahlmotive (H1.3) unterscheiden. Bezüglich weiterer Personenmerkmale lassen sich auf Basis der zur Verfügung stehenden Literatur keine verlässlichen Hypothesen formulieren. Hinsichtlich der Kontextwahl besteht die Vermutung, dass sich die Mehrheit der Lernenden für einen außerfachlichen Kontext entscheidet (H1.4), der im Vergleich zu innerfachlichen Kontexten einen höheren Lebensweltbezug aufweist (z. B. Broman et al., 2018; Fechner, 2009; Kölbach & Sumfleth, 2013). Ergebnisse von Habig et al. (2018) und van Vorst (2021) legen nahe, dass Lernende mit hoher Leistung in Chemie und hohem Interesse an Chemie vermutlich besondere Kontexte wählen (H1.5), während Lernende mit geringer Leistung in Chemie und geringem Interesse an Chemie alltägliche Kontexte wählen (H1.6). Mit hoher Wahrscheinlichkeit wird es eine dritte Lernendengruppe geben, die innerfachliche Kontexte wählt. Allerdings lassen sich aus der Literatur keine konkreten Hypothesen über die Ausprägung spezifischer Personenmerkmale in dieser Gruppe von Lernenden ableiten. Aller Voraussicht nach werden sich in den Lernendengruppen Unterschiede im Hinblick auf die Zufriedenheit (H2.1), das situationale Interesse (H2.2) und die kognitive Belastung (H2.3) nachweisen lassen, da sich einige Lernende retrospektiv für das falsche Kontextmerkmal entschieden haben. Hinweise dafür finden sich insbesondere bei van Vorst und Aydogmus (2021). Hierbei ist allerdings fraglich, welche Lernenden das sein könnten.

Die Beantwortung der ersten und zweiten Forschungsfrage führt somit zu weiteren Implikationen für die Adaptivität einzelner Kontexte für unterschiedliche Lernendengruppen. Im weiteren Verlauf wird besonders die Bedeutung der Passung zwischen Kontextmerkmal und den Bedarfen der Lernenden in den Blick genommen und untersucht, inwiefern die selbstständige Wahl einer kontextualisierten Aufgabe gegenüber einer systematisch zugeordneten Aufgabe Einfluss auf die Effektivität einer Lernsituation im Chemieunterricht hat. Im Zentrum der zweiten Teilstudie steht damit die Beantwortung der folgenden Forschungsfrage:

FF3: Welchen Einfluss hat die Passung zwischen Kontext und Lernenden sowie die selbstständige Kontextwahl auf den Fachwissenszuwachs, das situationale Interesse, die Zufriedenheit und die kognitive Belastung?

Auf Grundlage der Selbstbestimmungstheorie und der daran anknüpfenden Forschung (Flowerday & Shell, 2015; Katz & Assor, 2007; Wilde et al., 2018) wird angenommen, dass das Lernen mit einem passenden Kontext den Fachwissenszuwachs (H3.1a), die Zufriedenheit (H3.1b) und das situationale Interesse (H3.1c) steigert sowie die kognitive Belastung (H3.1d) vermindert. Die selbstständige Kontextwahl hat keinen Einfluss auf den Fachwissenszuwachs (H3.2a), die Zufriedenheit (H3.2b), das situationale Interesse (H3.2c) sowie die kognitive Belastung (H3.2d), wenn mit einem passenden Kontext gelernt wird.

Patall (2013) konnten in diesem Zusammenhang zeigen, dass hoch interessierte Personen von selbstgewählten Aufgaben und gering interessierte Personen von zugewiesenen Aufgaben hinsichtlich des situationalen Interesses profitieren. Entsprechend wird postuliert, dass der Einfluss der selbstständigen Kontextwahl und Passung zwischen Kontextaufgabe und Person auf die aufgabenbezogene Zufriedenheit (H3.3b) und das situationale Interesse (H3.3c) durch das individuelle Interesse der Lernenden moderiert wird. Es wird angenommen, dass demzufolge auch der Fachwissenszuwachs vom individuellen Interesse moderiert wird (H3.3a), da sich das situationale Interesse an kontextorientierten Aufgaben auf den Fachwissenszuwachs auswirkt (z. B. Habig, 2017).

Moderierende Effekte der Bereitstellung von Wahlmöglichkeiten und der selbsteingeschätzten Kompetenz auf affektive Merkmale wurden von Patall et al. (2014) nachgewiesen. Dabei scheinen Personen mit hohem Selbstkonzept von selbstgewählten Aufgaben und Personen mit niedrigem Selbstkonzept von zugewiesenen Aufgaben zu profitieren. Es wird daher angenommen, dass sowohl der Einfluss des selbstgewählten Kontexts als auch der Aufgabenpassung auf die aufgabenbezogene Zufriedenheit (H3.4b) und das situationale Interesse (H3.4c) durch das chemiebezogene Selbstkonzept moderiert wird. Aufgrund des engen Zusammenhangs zwischen dem situationalen Interesse und dem Fachwissenszuwachs wird auch

für den Fachwissenszuwachs ein Moderationseffekt durch das Selbstkonzept postuliert (H3.4a).

In Bezug auf die kognitive Belastung können keine theoriegeleiteten Hypothesen über mögliche Moderationseffekte mit dem individuellen Interesse oder Selbstkonzept der Lernenden formuliert werden.

6 Studiendesign

Um den Einfluss der Passung sowie der selbstständigen Kontextwahl im Rahmen der dritten Forschungsfrage zu analysieren, war es zunächst notwendig, die Kontextwahl und die Passung der Kontexte zu den Lernenden zu untersuchen. Deshalb musste die Gesamt-untersuchung in zwei Teilstudien zergliedert werden.

6.1 Teilstudie I

Im Zentrum der ersten Teilstudie standen die Untersuchung der Kontextwahl und die Beantwortung der ersten beiden Forschungsfragen. Zur Beantwortung dieser beiden Forschungsfragen wurde eine quantitative Querschnittstudie durchgeführt. Insbesondere quantitative Erhebungsmethoden eignen sich zur präzisen Erfassung von Leistungs- und Persönlichkeitsmerkmalen (Döring & Bortz, 2016; Hussy, 2013), die mit der Kontextwahl assoziiert sind. Bei der durchgeführten Untersuchung handelt es sich um eine explanative Studie mit einem nicht-experimentellen Design, die zur Gewährleistung der externen Validität im unterrichtlichen Feld durchgeführt wurde. Ein nicht-experimentelles Design erscheint geeignet, um die Kontextwahl verschiedener Schülerinnen und Schüler zu untersuchen (Döring & Bortz, 2016).

In Anlehnung an die Forschung zum kontextorientierten Lernen und der *situated expectancy-value-theory* (z. B. Eccles & Wigfield, 2020) wurden für die Kontextwahl potenziell relevante kognitive und affektiv-motivationale Personenmerkmale ausgewählt, die mithilfe von Leistungs- und Persönlichkeitstests erfasst wurden. Basierend auf bisherigen Arbeiten scheinen sowohl die Leistung in Chemie als auch das individuelle Interesse an Chemie in Bezug auf die Kontextwahl relevant zu sein (Habig et al., 2018; van Vorst & Aydogmus, 2021). Als Indikatoren für weitere Interessen der Lernenden wurden die Berufswahlperspektive sowie das Interesse an spezifischen Freizeitaktivitäten erhoben, da sie möglicherweise einen Erklärungsbeitrag für die Wahl bestimmter Kontexte liefern. Hierbei könnte die Wahl eines Kontextes passend für bestimmte Berufswünsche oder Freizeitaktivitäten sein (Eccles & Wigfield, 2020). Bisher blieb die Untersuchung solcher Zusammenhänge in der Kontextforschung unberücksichtigt. Weiterführend lassen sich auf Grundlage der SEVT das Selbstkonzept sowie die Selbstwirksamkeitserwartung als wesentliche Einflussfaktoren für die Aufgabenwahl herausarbeiten (Eccles & Wigfield, 2020). Die Untersuchung von van Vorst und Aydogmus (2021) zeigt, dass sich Lernende aufgrund ihrer Wahlmotive klassifizieren lassen und die Wahlmotive mit dem gewählten Kontextmerkmal in Verbindung stehen, was die Berücksichtigung der Wahlmotive erforderlich macht. Daneben wurde das Geschlecht als weiteres relevantes Merkmal erhoben. Außerdem wurde das Leseverständnis als Kontrollvariable erfasst, da anzunehmen ist, dass das Leseverständnis den Lernprozess und damit die Einschätzung der prozessbezogenen Variablen im letzten Schritt der Studie maßgeblich beeinflusst.

Für die Untersuchung der Kontextwahl wurden geeignete Aufgaben mit variierten Kontextmerkmalen konzipiert, aus denen die Lernenden wählen mussten. In der bisherigen Forschung deutet sich an, dass unterschiedliche Lernende alltägliche oder besondere Kontexte präferieren (Habig et al., 2018). Aus diesem Grund wurden die Kontextaufgaben hinsichtlich der beiden Kontextmerkmale variiert. Für die Kontextmerkmale *Alltäglich* und *Besonderheit* wurden explizit zwei Kontexte entwickelt, um zu gewährleisten, dass mindestens einer der beiden Kontexte dem angenommenen Kontextmerkmal entspricht. Darüber hinaus scheinen einige Lernende keine Präferenz für eines der beiden Kontextmerkmal zu zeigen (van Vorst & Aydogmus, 2021). Es ist unklar, welche Kontexte für diese Lernenden geeignet sein könnten, allerdings ist durch die Arbeiten von Broman und Kollegen (2018, 2020) bekannt, dass einige Lernende Kontexte mit einem starken innerfachlichen Bezug auswählen. Deshalb wurde zusätzlich ein innerfachlicher Kontext zur Wahl gestellt. Diesbezüglich war es notwendig, die Merkmalszugehörigkeit der kontextbasierten Aufgaben zu überprüfen. Insbesondere die wahrgenommene Zugehörigkeit zu den Kontextmerkmalen *Alltäglich* und *Besonders* ist subjektiv und abhängig von vorherigen Erfahrungen (van Vorst et al., 2015). Demnach kommt der Überprüfung der Merkmalszugehörigkeit eine tragende Bedeutung im Hinblick auf die Validität der Ergebnisse zu teil.

Insgesamt können die Lernenden damit zwischen fünf unterschiedlichen Kontextaufgaben wählen. Die Anzahl der Wahloptionen wird durch die bisherige Forschung zum Wahlverhalten als optimal angesehen, um choice overload zu vermeiden (Patall et al., 2008). Alle Kontextaufgaben wurden so gestaltet, dass sie vollkommen identisch hinsichtlich des Fachinhalts, des Themas sowie äußerer Merkmale waren. Durch die Kontrolle dieser Aufgabenmerkmale sollte die interne Validität der Aufgabenwahl hinsichtlich des zugrundeliegenden Kontextmerkmals sichergestellt werden. Insbesondere das Thema der Kontexte (z. B. Umweltschutz, Gesundheit oder Technik, siehe Elster, 2007 oder Schiepe-Tiska, Rönnebeck et al., 2016) sollte in den Aufgaben vergleichbar gestaltet werden, um auszuschließen, dass ein spezifischer Kontext aufgrund des adressierten Themas gewählt wird. Die Festlegung des Fachinhalts und des Themas für die Kontexte sowie die konkrete Ausgestaltung des kontextbasierten Lernmaterials für die erste Teilstudie wird in Kapitel 8.1 detailliert dargestellt.

Im Anschluss an die Kontextwahl und die Bearbeitung der Kontextaufgabe wurden die Zufriedenheit, das situationale Interesse und die kognitive Belastung als Prozessvariablen und Indikatoren für die Passung der kontextorientierten Aufgabe erhoben. Hierbei wird angenommen, dass die Lernenden nur auf Grundlage eines tatsächlichen stattfindenden Lernprozesses beurteilen können, ob der Kontext passend war (Habig et al., 2018; van Vorst & Aydogmus, 2021). Kooperative Arbeitsformen wurden hierbei ausgeschlossen, damit die Bewertung der Prozessvariablen nicht durch Gruppenprozesse während der Arbeitsphase gestört wird. Die Lernenden haben die kontextorientierte Aufgaben in Einzelarbeit bearbeitet. In Kapitel 8.3 wird der Ablauf der Datenerhebung für die erste Teilstudie dargestellt (siehe auch Abbildung 6).

Aus forschungsökonomischen Gründen wurde im Rahmen der ersten Teilstudie eine Gelegenheitsstichprobe aus der Metropolregion Rhein-Ruhr in Nordrhein-Westfalen rekrutiert (Bortz & Döring, 2016). Hierbei wurden sowohl Schülerinnen und Schüler vom Gymnasium als auch von der Gesamtschule untersucht. Da bisherige Untersuchungen zum kontextorientierten Lernen im Fach Chemie fast ausschließlich Lernende von Gymnasien umfassten (z. B. Habig et al., 2018; Kehne, 2019; Kölbach & Sumfleth, 2013; van Vorst & Aydogmus, 2021; van Vorst et al., 2018), sollten im Rahmen dieser Untersuchung erstmalig auch Lernende von Gesamtschulen rekrutiert werden. Es wird angenommen, dass die Stichprobe so repräsentativer für die Gesamtpopulation aller Chemielernenden in Deutschland ist. Aufgrund des festgelegten Inhaltsfeldes *saure und alkalische Lösungen*, war es notwendig, Lernende im dritten Lernjahr zu untersuchen, da das Inhaltsfeld im dritten Lernjahr verortet ist. An Gymnasien in Nordrhein-Westfalen findet das dritte Lernjahr i. d. R. in der Jahrgangsstufe 9, an Gesamtschulen in der Jahrgangsstufe 10 statt. Zur Beantwortung der ersten beiden Forschungsfragen wird ein Stichprobenumfang von $N = 500$ Lernenden als zielführend angesehen. Diese scheint sowohl für die psychometrische Analyse der Messinstrumente (siehe Kapitel 7.1), die geplante Clusteranalyse (siehe Kapitel 7.2) als auch die daran anknüpfende inferenzstatistische Analyse (siehe Kapitel 7.3) ausreichend. Beispielsweise ergibt eine Power-Analyse eine notwendige Stichprobengröße von $N = 246$ ($\alpha = .05$, $1\text{-}\beta = .80$) beim Mittelwertvergleich dreier Gruppen (Lernende mit alltäglicher, besonderer und innerfachlicher Kontextwahl) mit univariaten Varianzanalysen, wenn eine kleine bis mittlere Effektstärke ($f = .20$) angesetzt wird.

6.2 Teilstudie II

Die aus der ersten Teilstudie abgeleiteten Implikationen für die Adaptivität der Kontexte wurden zur Beantwortung der dritten Forschungsfrage in der zweiten Teilstudie genutzt. Hierzu wurde eine randomisiert kontrollierte Studie im Prä-Post-Design durchgeführt. Experimentelle Studien eignen sich insbesondere zur Prüfung von Kausalhypothesen, wie sie im Rahmen der dritten Forschungsfrage formuliert wurden (Döring & Bortz, 2016; Hussy et al., 2013). Angelehnt an die Untersuchung von Wilde et al. (2018) wurden drei Treatmentgruppen in die Studie integriert, die hinsichtlich der Passung der Kontextaufgabe und der selbstständigen Wahl manipuliert wurden:

(1) Die Lernenden wählen eine Kontextaufgabe selbstständig (*choose & match*)
(2) Die Lernenden bekommen eine passende Kontextaufgabe zugewiesen (*no choose & match*)
(3) Die Lernenden bekommen eine nicht passende Kontextaufgabe zugewiesen (*no choose & no match*)

Den Lernenden aus der zweiten und dritten Treatmentgruppe wurde zu diesem Zweck eine Kontextaufgabe mithilfe eines Vorhersagemodells zugewiesen, welches mithilfe von maschinellem Lernen entwickelt wurde (siehe Kapitel 9). Für die Zuordnung der Kontextaufgaben wurden im Sinne eines datengestützten Differenzierungsansatzes vorher erfasste individuelle Merkmale der Lernenden herangezogen. Dieses Vorgehen hat mehrere Vorteile: Zum einen mussten die Kontexte den Lernenden der zweiten und dritten Treatmentgruppe im Gegensatz zur Studie von Wilde et al. (2018) nicht vorab zur Auswahl vorgelegt werden, um den individuell passenden Kontext zu identifizieren. Damit wurde den Schülerinnen und Schülern eine bereits gewährte Wahlmöglichkeit nicht wieder entzogen, was sich nachweislich negativ auf das Autonomieerleben und die intrinsische Motivation auswirkt (Patall et al., 2008). Zum anderen kannten die Lernenden der no choose-Treatmentgruppen so die anderen Kontexte nicht. Auch hier gibt es Hinweise darauf, dass bekannte, aber nicht zugeordnete Alternativen retrospektiv attraktiver erscheinen, was wiederum die Motivation senkt (Patall et al., 2008). Darüber hinaus ist fraglich, inwieweit es sich tatsächlich um eine no choose & match-Treatmentgruppe handelt, wenn sich die Probandinnen und Probanden im Vorfeld doch für eine Aufgabe entschieden haben und lediglich eine gewisse zeitliche Distanz zwischen Wahl und Zuordnung bestand. Daher wird die Zuweisung für die Treatmentgruppen ohne Wahlmöglichkeit aufgrund der individuellen Merkmale und des konzipierten Modells als zielführend angesehen.

Gemäß der Unterscheidung in der ersten Teilstudie wurden die Kontextaufgaben im Hinblick auf das Kontextmerkmal variiert. Infolgedessen standen eine alltägliche, eine besondere und eine innerfachliche Kontextaufgabe für die Zuweisung bzw. die Wahl zur Verfügung. Aufbauend auf dem Lernmaterial aus der ersten Teilstudie wurden weitere Teilaufgaben für die Verwendung in der zweiten Teilstudie konzipiert, um Neuigkeitseffekte aufzudecken (Theyßen, 2014). Das für die zweite Teilstudie weiterentwickelte Lernmaterial wird in Kapitel 10.1 vorgestellt. Zum Ausschluss von Störfaktoren haben die Lernenden auch hier in Einzelarbeit an den kontextorientierten Aufgaben gearbeitet.

Zur Prüfung der im letzten Kapitel festgelegten Hypothesen wurden das Fachwissen im Inhaltsfeld *saure und alkalische Lösungen*, die aufgabenbezogene Zufriedenheit, das situationale Interesse und die kognitive Belastung erhoben. Um den Fachwissenszuwachs der drei Treatments zu vergleichen, wurde das Fachwissen vor und nach der Bearbeitung der kontextorientierten Lerneinheit gemessen. Die Zufriedenheit, das situationale Interesse und die kognitive Belastung wurden als prozessbezogene Variablen nach Bearbeitung jeder Teilaufgabe gemessen, um Veränderungen während des Lernprozesses aufzuklären.

Neben der für eine experimentelle Studie charakteristischen Randomisierung der Versuchspersonen, wurden weitere mögliche personenbezogene Kontrollvariablen zur Gewährleistung der internen Validität des Experiments kontrolliert. Hierbei wurde das Leseverständnis der

Schülerinnen und Schüler erhoben, da anzunehmen ist, dass insbesondere der Fachwissenszuwachs durch das Leseverständnis als Teilfacette der Lesekompetenz beeinflusst wird (z. B. Artelt et al., 2007). Daneben wurden das Vorwissen in Chemie (z. B. Gruber & Stamouli, 2009), das individuelle Interesse an Chemie (z. B. Schiefele et al., 1992) und das chemiebezogene Selbstkonzept (z. B. Möller & Trautwein, 2009) erfasst, da anzunehmen ist, dass diese einen bedeutenden Einfluss auf die abhängigen Variablen haben. Im Unterschied zum Fachwissen ist das Vorwissen in Chemie nicht auf das Inhaltsfeld *saure und alkalischen Lösungen* beschränkt, sondern beschreibt das Fachwissen in den drei Basiskonzepten, welches in den ersten beiden Lernjahren erlernt werden sollte (siehe Kapitel 8.2.1).

Zur Vermeidung von Versuchsleitungseffekten wurde die Studie fast ausschließlich durch die gleiche testleitende Person durchgeführt. Lediglich zwei Termine wurden aufgrund von Terminüberschneidungen von einer anderen geschulten Testleitung übernommen. Darüber hinaus wurden die Lernenden über den eigentlichen Zweck des Experiments nicht in Kenntnis gesetzt, um potenzielle Erwartungseffekte zu eliminieren (Döring & Bortz, 2016). Es wurde insbesondere vermieden, die Lernenden darüber zu informieren, dass einige Lernende eine Kontextaufgabe wählen dürfen, während anderen diese Wahl verwehrt blieb. Die experimentelle Studie wurde in den regulären Chemieunterricht integriert, um eine hohe externe Validität zu gewährleisten. Der Ablauf der Datenerhebung wird detailliert im Kapitel 10.3 vorgestellt (siehe auch Abbildung 30).

Für die zweite Teilstudie wurde erneut eine Gelegenheitsstichprobe von Lernenden aus dem dritten Lernjahr an Gymnasien und Gesamtschulen in der Metropolregion Rhein-Ruhr in Nordrhein-Westfalen festgelegt. Die Stichprobenumfangsplanung erfolgte gemäß der Empfehlung von Maas und Hox (2005) bei Mehrebenenmodellen mindestens 50 Einheiten auf der zweiten Ebene (im vorliegenden Design die Personen, siehe Kapitel 7.5 und 10.5.4), aber so viele Einheiten wie möglich zu erfassen (Snijders, 2005).

7 Psychometrische und Statistische Methoden

In den beiden quantitativen Studien wurden eine Reihe empirischer Testwerte erhoben, die zur Beantwortung der Forschungsfragen skaliert werden müssen. Aus der Perspektive der Psychometrie ist hierfür die Festlegung einer Testtheorie zwingende Voraussetzung (Kapitel 7.1). Die skalierten Testwerte werden im Rahmen einer quantitativen Datenanalyse genutzt, um mittels statistischer Verfahren (Kapitel 7.2-7.5) die Forschungsfragen zu beantworten (Döring & Bortz, 2016).

7.1 Testtheorie

In der Psychometrie sind die klassische Testtheorie (KTT) und die probabilistische Testtheorie (häufig auch Item Response Theory, IRT) die zwei gebräuchlichsten Messtheorien zur Skalierung von empirischen Testwerten (Moosbrugger, Schermelleh-Engel et al., 2020). Im Folgenden sollen die Grundzüge beider Testtheorien kurz beschrieben werden, um Anwendungspunkte für die Verwendung beider Testtheorien in dieser Arbeit zu begründen.

Die klassische Testtheorie

Die KTT ist der historisch vorherrschende Ansatz zur Konstruktion und Skalierung von psychologischen Testinstrumenten. Bei der KTT handelt es sich um eine Messfehlertheorie (Eid et al., 2017; Moosbrugger, Gäde et al., 2020; Raykov & Marcoulides, 2011). Der Messwert *y* einer Person *v* auf einer Variable *i* setzt sich aus dem Messfehler ε_{vi} und dem wahren Wert τ_{vi} zusammen:

$$y_{vi} = \tau_{vi} + \varepsilon_{vi} \qquad (1)$$

Hierbei wird ein linearer Zusammenhang zwischen der latenten Variable η und dem wahren Wert τ angenommen (Moosbrugger, Gäde et al., 2020).

Die Grundgleichung der KTT (Gleichung 1) ist empirisch nicht überprüfbar. Allerdings können Annahmen über die Messeigenschaften der Items – insbesondere über die Schwierigkeit und Trennschärfe – in einem Messmodell formuliert und empirisch geprüft werden. Hierbei können u. a. drei verschiedene Formen der Messäquivalenz unterschieden werden: τ-Kongenerität, essentielle τ-Äquivalenz, essentielle τ-Parallelität (u. a. Eid & Schmidt, 2014; Moosbrugger, Gäde et al., 2020; Raykov & Marcoulides, 2011). Im τ-kongenerischen Messmodell werden keine Restriktionen hinsichtlich des Diskriminationsparameters, der Schwierigkeit und der Messfehlervarianzen spezifiziert. Das τ-kongenerischen Messmodell prüft somit lediglich die Eindimensionalität einer Skala unter der Annahme unkorrelierter Messfehler und ist in diesem Zusammenhang das vermutlich für die meisten Tests passendste Messmodell (Gäde et al., 2020). Dementsprechend wird es auch im Rahmen dieser Arbeit als geeignet

betrachtet. Führt man dagegen Gleichheitsrestriktionen bezüglich des Diskriminationsparameters ein, wird das Modell essentieller τ-Äquivalenz geprüft. Zusätzliche Modellrestriktionen bezüglich der Messfehlervarianzen sind im essentiell τ-Parallelen Messmodell definiert. Empirische Daten entsprechen den strengen Annahmen essentieller τ-Äquivalenz und essentieller τ-Parallelität häufig nicht (Moosbrugger, Gäde et al., 2020), weshalb beide Annahmen im Rahmen dieser Arbeit auch nicht explizit geprüft werden.

Im Rahmen einer konfirmatorischen Faktorenanalyse (CFA) können psychometrische Messmodelle spezifiziert und deren Gültigkeit überprüft werden. Darin besteht ein wesentlicher Unterschied zur weiter verbreiteten explorativen Faktorenanalyse (EFA), die keine Strukturprüfung erlaubt. Die EFA eignet sich zur Untersuchung der faktoriellen Struktur, sofern keine spezifische Hypothese über die Anzahl der Faktoren oder die Zugehörigkeit der Indikatoren zu den Faktoren besteht (Brandt, 2020; Eid et al., 2017). Die Gültigkeit der in der CFA spezifizierten Messmodelle kann anhand von deskriptiven und inferenzstatistischen Gütekriterien beurteilt werden. Gäde et al. (2020) haben gängige Cut-off-Werte für die am häufigsten genutzten Gütekriterien zusammengefasst (Tab. 1).

Tabelle 1: Gütekriterien und Cut-off-Werte zur Beurteilung des Modellfit (Gäde et al., 2020)

Gütekriterium	Guter Modellfit	Akzeptabler Modellfit
χ^2-Wert	χ^2 / df ≤ 2	χ^2 / df ≤ 3
CFI	CFI ≥ .95	CFI ≥ .90
TLI	TLI ≥ .95	CFI ≥ .90
RMSEA	RMSEA ≤ .05	RMSEA ≤ .08
SRMR	SRMR ≤ .05	SRMR ≤ .10

Fehlspezifikationen des Messmodells können anhand der Modifikationsindizes identifiziert werden. Diese geben an „…um wie viel sich der χ^2-Wert ungefähr verringern würde, wenn ein vorher fixierter oder restringierter Parameter freigesetzt wird, während das restliche Modell gleichzeitig unverändert bleibt." (Gäde et al., 2020, S. 651). Durch die Modifikationsindizes können so Fehlerkovarianzen identifiziert werden, die als Hinweis auf die Verletzung der Annahme der Eindimensionalität zu interpretieren sind. Gäde et al. (2020) empfehlen Fehlerkovarianzen kritisch zu überprüfen. Sofern es sich dabei um konstruktrelevante Fehlerkovarianzen handelt, kann ein zusätzlicher Faktor im Modell spezifiziert werden. Konstrukttirrelevante Fehlerkovarianzen können dagegen auf einen Methodeneffekt hindeuten, der u. a. durch ähnliche Itemformulierungen entstehen kann. Solche Methodeneffekte können über Fehlerkovarianzen modelliert werden (Schermelleh-Engel & Gäde, 2020). Hermida (2015)

empfiehlt, Fehlerkovarianzen nicht zu modellieren. Stattdessen können problematische Items auch aus der weiteren Analyse ausgeschlossen werden (Weiber & Mühlhaus, 2014).

Werden Messinstrumente über verschiedene Messzeitpunkte hinweg eingesetzt, um beispielsweise die Veränderung eines Merkmals über die Zeit zu erfassen, so ist auf Invarianz der Messung zu prüfen (Gäde et al., 2020; Mackinnon et al., 2022). Hierbei können verschiedene Stufen der Messinvarianz unterschieden werden. Konfigurale Messinvarianz setzt die Gleichheit der Modellstruktur über die Messzeitpunkte voraus. Im Messmodell metrischer Messinvarianz werden die Faktorladungen der Items über die Messzeitpunkte restringiert, wodurch geprüft wird, ob die Faktoren über die Zeit dieselbe Bedeutung haben. Skalare Invarianz impliziert die Restriktion der Intercepts der Items über die Zeit, sodass die Mittelwerte der Items nicht über die Zeit variieren. Werden die Fehlervarianzen für jedes Item restringiert, so impliziert man strikte Invarianz über die Zeit. Unter Gültigkeit strikter Invarianz kann angenommen werden, dass Unterschiede in dem Faktor allein auf Unterschiede in den manifesten Indikatoren zurückzuführen sind und nicht auf den Einfluss externer Faktoren (Gäde et al., 2020).

Die Messinvarianztestung erfolgt mittels CFA und der schrittweisen Einführung der beschriebenen Restriktionen hinsichtlich der Modellparameter (Gäde et al., 2020). Der Vergleich der spezifizierten Modelle kann anhand einiger der in Tabelle 1 angeführten Gütekriterien durchgeführt werden. Bei ausreichender Stichprobengröße ($N > 300$) zeigen Änderungen von $\Delta CFI \geq -.010$, $\Delta SRMR \geq .030$ und $\Delta RMSEA \geq .015$ eine signifikant schlechtere Modellpassung an, wodurch die Annahme über die spezifizierte Messinvarianzstufe zu verwerfen ist. Für die Testung auf skalare oder strikte Invarianz sind strengere Kriterien von $\Delta CFI \geq -.010$, $\Delta SRMR \geq .010$ und $\Delta RMSEA \geq .015$ anzusetzen (F. F. Chen, 2007).

Zeigt der Vergleich zweier spezifizierter Modelle eine signifikante Verschlechterung der Modellpassung an, so kann zunächst auf partielle Invarianz geprüft werden. Dafür werden die Parameterrestriktionen für einzelne Items aufgegeben (Byrne et al., 1989). Steenkamp und Baumgartner (1998) verweisen darauf, dass zwei Indikatoren mit metrischer und skalarer Invarianz ausreichend sind.

Um manifeste Mittelwertunterschiede zwischen Gruppen auf die Ausprägung des latenten Merkmals zurückzuführen, ist skalare Messinvarianz notwendig (Gäde et al., 2020). Allerdings handelt es sich dabei um eine äußerst strenge Annahme, die nur selten erfüllt werden kann (Dimitrov, 2010). Steenkamp und Baumgartner (1998) sehen konfigurale Messinvarianz als ausreichend an.

Unter Gültigkeit der Messmodelle können verschiedene modellbasierte Reliabilitätsmaße bestimmt werden, die klassischen Methoden der Reliabilitätsbestimmung überlegen sind. Im Gegensatz zur klassischen Reliabilitätsbestimmung basieren modellbasierte Verfahren auf den geschätzten Modellparametern und nicht auf der empirischen Varianz und Kovarianz der

Itemvariablen (Schermelleh-Engel & Gäde, 2020). Insbesondere Cronbachs α ist ein weit verbreitetes Reliabilitätsmaß, welches allerdings essentielle τ-Äquivalenz der Items voraussetzt, um präzise und unverzerrte Schätzungen der Reliabilität vorzunehmen (Gäde et al., 2020; Raykov & Marcoulides, 2011; Schermelleh-Engel & Gäde, 2020). Allerdings wird das „…Zutreffen der implizit vorausgesetzten Modellrestriktionen wie Eindimensionalität, essentielle τ-Äquivalenz oder essentielle τ-Parallelität der Messungen … in der Praxis meist nicht explizit überprüft, obwohl diese Prüfung zwingend erforderlich wäre." (Schermelleh-Engel & Gäde, 2020, S. 337). Sofern die Messung der latenten Personenmerkmale anhand von τ-kongenerischen Items vorgenehmen wurde, liefert McDonalds ω eine angemessenere Schätzung der Reliabilität. In mehrdimensionalen Modellen kann McDonalds ω für jede Subskala berechnet werden (Schermelleh-Engel & Gäde, 2020; Watkins, 2017). Normalerweise unterscheiden sich Cronbachs α und McDonalds ω nur gering voneinander (Hayes & Coutts, 2020). Unterschiede bestehen vor allem dann, wenn die durchschnittlichen Faktorladungen nicht hoch sind und wenn unterschiedliche Faktorladungen vorliegen (Raykov & Marcoulides, 2015).

Gleichwohl unterschiedliche Richtlinien für die Interpretation der Reliabilität bestehen, werden normalerweise Reliabilitäten von mindestens .70 als angemessen betrachtet (Gäde et al., 2020). Trotzdem sollte von der unreflektierten Verwendung solcher Heuristiken abgeraten werden. Taber (2018) liefert in diesem Zusammenhang einen gelungenen Überblick über die Verwendung von Cronbachs α in der naturwissenschaftsdidaktischen Forschung und empfiehlt stattdessen einen differenzierten Blick auf das Testinstrument und das zu messende Konstrukt einzunehmen. Demnach können beispielsweise bei Test-instrumenten, die einen breiten Wissensbereich abdecken, keine hohen Reliabilitäten erwartet werden. In diesem Fall können auch Reliabilitäten unter .70 ggf. als ausreichend betrachtet werden. Nichtsdestotrotz sind eine hohe Reliabilität und die Eindimensionalität der Skala wesentliche Voraussetzung für die in der KTT üblichen Aufsummierung der manifesten Itemvariablen zur Schätzung der individuellen Merkmalsausprägung (Moosbrugger, Gäde et al., 2020), wie sie auch im Rahmen dieser Arbeit genutzt wird.

Die Item-Response-Theory

Im Gegensatz zur KTT adressiert die IRT nicht den Messfehler, sondern beschreibt in erster Linie das Zustandekommen der Antworten auf die Items (Bond et al., 2021; Embretson, 2009). Ähnlich wie in der KTT werden latente Personenmerkmale als Ursache für die Antworten postuliert. Die IRT modelliert hierbei die bedingte Wahrscheinlichkeit, mit welcher eine Person mit gegebener latenter Personenfähigkeit η eine richtige Antwort *Y* auf das Item *i* gibt (unter der Annahme dichotomer Items):

$$P(Y_i = 1 \mid \eta) \qquad (2)$$

Der Zusammenhang zwischen dem latenten Personenmerkmal und der bedingten Wahrscheinlichkeit ist dabei logistisch (Kelava & Moosbrugger, 2020).

Die IRT hat zahlreiche Messmodelle hervorgebracht, die sich in Bezug auf die Kategorienanzahl der Antwortvariablen und hinsichtlich der zu schätzenden Itemparameter (z. B. Diskriminations- und Schwierigkeitsparameter) unterscheiden (eine Einführung in unterschiedliche IRT-Modelle findet sich z. B. in Bond et al., 2021). Das wohl bekannteste Modell ist das Rasch-Modell, welches sich für dichotome Antwortformate eignet und häufig in der Leistungsdiagnostik eingesetzt wird. Im Rasch-Modell wird lediglich die Itemschwierigkeit als einziger Itemparameter geschätzt, weshalb das Rasch-Modell auch als 1PL-Modell bezeichnet wird. Die Lösungswahrscheinlichkeit eines Items hängt so nur von der Ausprägung des latenten Merkmals und der Itemschwierigkeit ab:

$$P(Y_i = 1 \mid \eta) = \frac{e^{(\eta - \beta_i)}}{1 + e^{(\eta - \beta_i)}} \tag{3}$$

Durch Gleichung 3 ergibt sich die für das Rasch-Modell spezifische Itemcharakteristik, bei der die logistische Funktion die gleiche Form aufweist und lediglich horizontal entlang der x-Achse verschoben ist (Bond et al., 2021).

Inwieweit sich diese spezifische Itemcharakteristik tatsächlich in den empirischen Daten zeigt, kann anhand von Fit-Statistiken beurteilt werden. In diesem Zusammenhang kann zwischen dem ungewichteten Outfit (MNSQ) und dem gewichteten Infit (wMNSQ) unterschieden werden. Sowohl Outfit als auch Infit werden häufig auch in standardisierter Form berücksichtigt, indem sie in die korrespondierenden *t*-Werte transformiert werden. Meistens wird ausschließlich der Infit zur Beurteilung der Modellpassung herangezogen, da der Outfit sensitiv gegenüber Abweichungen von unerwarteten Antwortmustern ist (Neumann, 2014). Im Rahmen dieser Arbeit werden die empfohlenen Cut-Off-Kriterien von $0.80 \leq \text{wMNSQ} \leq 1.20$ genutzt (Adam & Wu, 2002). Für die *t*-Werte werden die üblichen 95 %-Konfidenzintervalle der Standardnormalverteilung von $-1.96 \leq t \leq 1.96$ herangezogen (Bond et al., 2021).

Durch die IRT können modellbasierte Schätzungen der latenten Personenwerte vorgenommen werden. Im Rahmen dieser Arbeit werden die Personenparameter anhand der *Weighted-Maximum-Likelihood-Methode* geschätzt, da die *Weighted-Maximum-Likelihood-Estimates* (WLEs*)* als beste Punktschätzer für die Personenparameter gelten (Rost, 2004). Auf Grundlage dieser WLEs kann ein WLE-Reliabilitätskoeffizient bestimmt werden, der ähnlich wie andere Reliabilitätskoeffizienten zu interpretieren ist (siehe Diskussion zur Reliabilität im Rahmen der KTT).

IRT-skalierte Messinstrumente, die zu verschiedenen Messzeitpunkten eingesetzt werden, um Veränderungsmessungen durchzuführen, sind – ähnlich wie in der KTT – auf Messinvarianz zu prüfen. Voraussetzung dafür ist die Verknüpfung der Messungen zu verschiedenen Zeitpunkten. Dafür stehen im Rahmen der IRT verschiedene Methoden zur Verfügung, von denen

einige von L. Fischer et al. (2021) kontrastiert werden. Eine dieser Methoden ist die *fixed parameter calibration* (FPC), bei der zwei Messungen durch eine Reihe gemeinsamer Items (Ankeritems) und deren fixierten Parametern miteinander verbunden werden. Im Rahmen dieser Arbeit wird dazu ein Vorgehen genutzt, welches Averbeck (2021), basierend auf den Verfahren von Pentecost und Barbera (2013), Wright (2003), Stahl und Muckle (2007) und O'Neill et al. (2013) zusammengestellt hat. Hierfür wird zunächst ein Rasch-Modell für jeden Messzeitpunkt mit frei geschätzten Itemparametern bestimmt und anschließend hinsichtlich der Rasch-Konformität überprüft. Anschließend werden die Itemparameter auf einen der Messzeitpunkte restringiert und ein neues Rasch-Modell wird für den Messzeitpunkt berechnet, der nicht zur Fixierung der Itemparameter genutzt wurde. Items mit schlechter Modellpassung zeigen in diesem Zusammenhang Messvarianz an. In diesem Fall kann das Item entfernt oder frei geschätzt werden. Passt das Item auch mit frei geschätzten Itemparametern nicht zum Modell, sollte es entfernt werden. Auf Grundlage dieses reduzierten Datensatzes wird dann ein neues Rasch-Modell mit vollständig bzw. teilweise fixierten Itemparametern berechnet. Auf Grundlage der fixierten Itemparameter sind die WLEs der Personenfähigkeit direkt miteinander vergleichbar.

Vergleich der KTT und IRT

Es ist eine weit verbreitete Annahme, dass es sich bei der IRT um die konzeptionell überlegene Testtheorie handelt. Diese Annahme basiert vornehmlich darauf, dass es sich bei der KTT um eine axiomatisch festgelegte Testtheorie handelt, deren Zutreffen empirisch nicht überprüft werden kann (Raykov & Marcoulides, 2011). Allerdings können sowohl in der IRT als auch in der KTT modellimplizierte Annahmen empirisch überprüft werden. Tatsächlich haben sich beide Testtheorien in der Vergangenheit immer weiter angenähert, so dass inzwischen nur noch geringe Unterschiede bestehen. Folglich ist die „…IRT nicht als Alternative zur KTT zu sehen, sondern eher als Ergänzung." (Moosbrugger, Schermelleh-Engel et al., 2020, S. 268). In der Forschungspraxis ist es demzufolge auch nicht unüblich beide Testtheorien zur Skalierung unterschiedlicher Testinstrumente zu verwenden (z. B. Dorfner et al., 2018; Höft et al., 2019).

In der IRT stellt die lokale stochastische Unabhängigkeit ein zentrales Kriterium zur Modellkonformität dar, welches in der KTT lediglich auf Basis der weniger strengen Annahme unkorrelierter Messfehler überprüft wird. Die lokale stochastische Unabhängigkeit ist hierbei eine wesentliche Voraussetzung für spezifisch objektive Vergleiche, welche die Schätzung von Personenparametern bei unterschiedlicher Vorgabe von Items erlaubt. Zwar erlaubt auch die KTT spezifische objektive Vergleiche, sofern Konformität mit dem Messmodell essentieller τ-Parallelität besteht (Moosbrugger, Schermelleh-Engel et al., 2020). Allerdings werden aufgrund der strengeren Annahme in der IRT, spezifisch objektive Vergleiche häufig anhand

von Personenparametern durchgeführt, die durch IRT-Modelle geschätzt wurden. Demzufolge wird auch im Rahmen dieser Arbeit ein IRT-Modell zur Schätzung der Personenparameter genutzt, wenn es aus testökologischen Gründen nicht möglich war, dass jede Versuchsperson jedes Item bearbeitet. Dies betrifft insbesondere den Leistungstest zur Erfassung des Vorwissens in Chemie (siehe Kapitel 8.2.1). Darüber hinaus setzen die in der KTT üblichen Summenscores voraus, dass ein linearer Zusammenhang zwischen der latenten Fähigkeitsausprägung und der relativen Anzahl gelöster Items vorliegt. Diese Annahme gilt bei dichotomen Items nur bei konstanten Itemschwierigkeiten (Rost, 2004). Insbesondere Leistungstests werden aber explizit mit unterschiedlich schwierigen Items konzipiert, um das gesamte Fähigkeitsspektrum einer Population zu erfassen (z. B. Walpuski et al., 2010). Aus diesem Grund wird auch der Leistungstest zur Erfassung des Fachwissens im Inhaltsfeld saure und alkalische Lösungen mithilfe eines IRT-Modells untersucht.

Die meisten in dieser Arbeit erfassten affektiv-motivationalen Merkmale sind *nicht* eindimensional (siehe auch Reise & Henson, 2003) und lassen sich demzufolge auch nicht mit eindimensionalen IRT-Modellen (z. B. dem Partial-Credit- oder Rating-Scale-Modell für geordnete Antwortkategorien) analysieren. Stattdessen müssten mehrdimensionale IRT-Modelle zur Untersuchung der affektiv-motivationalen Testinstrumente berechnet werden (Kelava et al., 2020). Diese benötigen allerdings Stichproben von mindestens 500 Personen, um zuverlässige Schätzungen der Itemparameter zu erhalten (Jiang et al., 2016). Da die in dieser Arbeit untersuchten Stichproben kleiner sind, wird auf die Berechnung mehrdimensionaler IRT-Modelle verzichtet. Im Gegensatz dazu werden für die Analyse von Messinstrumenten mithilfe der CFA Stichprobenumfänge benötigt, die fünfmal so groß wie die Anzahl der zu schätzenden Parameter sind (Kline, 2005; Weiber & Mühlhaus, 2014). Die dafür notwendigen Stichprobenumfänge liegen i. d .R. unter 500 Personen[1]. Darüber hinaus empfehlen Wirth und Edwards (2007) die CFA (und damit die KTT), wenn es darum geht, den strukturellen Aufbau eines Messinstruments zu untersuchen – was häufig bei Messinstrumenten erforderlich ist, die mehrdimensionale Merkmale erfassen. Deshalb werden die Messinstrumente zur Erfassung affektiv-motivationaler Merkmale mithilfe der beschriebenen Methoden der KTT untersucht. Die durch die KTT und IRT skalierten Personenparameter bilden damit die Grundlage für die weitere statistische Analyse.

7.2 Clusteranalyse

Insbesondere clusteranalytische Verfahren bieten sich zur Klassifikation von Personen anhand individueller Merkmale an. Das primäre Ziel ist dabei das Auffinden einer empirischen Klassifikation, die hinsichtlich der ausgewählten Merkmale möglichst homogen ist (Bacher

[1] Zum Beispiel werden in einem vier-faktoriellen τ-kongenerischen-Messmodell mit 27 Items in einer CFA 87 Parameter geschätzt. Das ergibt eine notwendige Stichprobengröße von 435 Personen.

et al., 2010; Backhaus et al., 2020; Everitt et al., 2011; Schendera, 2010). Damit können clusteranalytische Verfahren eingesetzt werden, um Lernende anhand von individuellen Merkmalen zu klassifizieren (siehe Kapitel 8.5.4).

Nach Bacher et al. (2010) eignen sich Mittelwertverfahren, das Ward-Verfahren, das K-Means-Verfahren und probabilistische Verfahren zur Klassifikation von Objekten. Insbesondere probabilistische Verfahren werden von den Autoren empfohlen, sofern eine ausreichend große Stichprobe zur Verfügung steht. Als Richtwert werden hier Stichprobengrößen von $N = 500$ aufgrund von Simulationsstudien berichtet. Bei kleineren Stichproben empfehlen Bacher et al. (2010) das K-Means-Verfahren. Dieses berechnet im Gegensatz zu hierarchischen Clusteranalysen eine optimale Lösung unter gegebener Clusterzahl. Hierarchische Clusteranalysen wie das Ward-Verfahren sind für mehr als 250 Fälle ungeeignet (Schendera, 2010).

Zunächst müssen für die Clusteranalyse Variablen ausgewählt werden, die relevant für den untersuchten Sachverhalt sind (Backhaus et al., 2020; Schendera, 2010). Dafür können Verfahren zum Vergleich von Mittelwerten genutzt werden. Die so ausgewählten Clustervariablen sollten außerdem keine zu hohe Korrelation aufweisen, um bestimmte Aspekte bei der Klassifikation nicht über zu gewichten (Backhaus et al., 2020).

Nach der Auswahl von relevanten Variablen muss eine geeignete Anzahl an Clustern ausgewählt werden. Dieser Schritt wird von Bacher et al. (2010, S. 470) als „…»Achillesverse« der Clusteranalyse“ bezeichnet. Zur Bestimmung der Clusterzahl beim K-Means-Verfahren stehen unterschiedliche Maßzahlen zur Verfügung. Bacher et al. (2010) empfehlen, insbesondere die prozentuale Veränderung der Streuungsquadratsumme in den K Clustern gegenüber einer vorausgehenden Clusterlösung zu betrachten. Diese ist wie folgt definiert:

$$PRE_K = 1 - \frac{SQ_{in}(K)}{SQ_{in}(K - 1)} \tag{4}$$

Eine weit verbreitete Methode ist außerdem das sogenannte Elbow-Kriterium, bei der die Gesamtstreuungsquadratsumme gegen die Anzahl der Cluster aufgetragen wird. Es wird die Clusterlösung gewählt, nach der die nachfolgende Clusterlösung keinen erheblichen Abfall der Gesamtstreuungsquadratsumme erbringt. Dieser zeigt sich bei der grafischen Auftragung als „Ellbogen“ (Backhaus et al., 2020).

Bacher et al. (2010) empfehlen bei der Festlegung der Clusteranzahl nicht zu restriktiv zu sein und zu überprüfen, ob die Kriterien an eine gute Klassifikation erfüllt sind. Dieser Schritt umfasst die sogenannte formale Gültigkeitsprüfung, bei der die gefundene Clusterlösung hinsichtlich der Homogenität in den Clustern, der Heterogenität zwischen den Clustern, der Erklärungsgüte sowie der Stabilität der Cluster untersucht wird. Zur Beurteilung der Homogenität in einem Cluster kann die Varianz in einem Cluster in Beziehung zur gesamten Streuung gesetzt werden:

$$Homo_k = 1 - \frac{s_k^2}{s_{ges}^2} \quad (5)$$

Je geringer die Streuung in einem Cluster, desto ähnlicher sind sich die Objekte und der Homogenitätsindex strebt gegen 1 (Bacher et al., 2010).

Analog zum Homogenitätsindex lässt sich der Heterogenitätsindex als Maß für die Unähnlichkeit der Cluster berechnen. Hierbei wird die Gesamtstreuung auf die Distanz zwischen den Clusterzentren normiert:

$$Hetero_k = 1 - \frac{s_{ges}^2}{d_{kk}^2} \quad (6)$$

Bei großen Distanzen zwischen den Clusterzentren wird die Zahl im Nenner sehr groß, wodurch der Heterogenitätsindex gegen 0 tendiert (Bacher et al., 2010).

Durch die gefundene Clusterstruktur soll ein großer Teil der Variation in den Daten erklärt werden. Als Maß für die erklärte Varianz kann η^2_K berechnet werden (Bacher et al., 2010):

$$\eta_K^2 = 1 - \frac{SQ_{in}(K)}{SQ_{ges}} \quad (7)$$

Die identifizierte Clusterstruktur kann als formal zulässig betrachtet werden, wenn die Cluster stabil sind. Zur Beurteilung der Stabilität kann das Verfahren von Hennig (2007) genutzt werden. Hierbei wird eine Clusterlösung basierend auf Bootstrap-Stichproben der ursprünglichen Stichprobe berechnet. Der Jaccard-Koeffizient γ wird als Kennzahl zur Beurteilung der Übereinstimmung zwischen der Bootstrap-Clusterlösung und der ursprünglich berechneten Clusterlösung herangezogen. Clusterlösungen gelten nach dem Autor als stabil, wenn $\bar{\gamma} > .75$ gilt.

Eine formal gültige Clusterlösung muss darüber hinaus interpretierbar sein. Grundlage für die inhaltliche Beschreibung und Interpretation der Cluster können Vergleiche der Mittelwerte der Variablen sein, die in der Clusteranalyse genutzt wurden (Bacher et al., 2010; Schendera, 2010). Zuletzt kann eine Validitätsprüfung der Cluster erfolgen. Hierbei können Hypothesen über die gefundenen Cluster formuliert und mithilfe von Variablen überprüft werden, die nicht bei der Clusteranalyse genutzt wurden. Hierfür können klassische Verfahren der Inferenzstatistik zum Einsatz kommen (Bacher et al., 2010).

7.3 Klassische inferenzstatistische Verfahren

Im Gegensatz zur deskriptiven Statistik ermöglicht die Inferenzstatistik die Überprüfung von statistischen Hypothesen. Hierbei wird anhand von Stichprobendaten auf die Population geschlossen (Bortz & Schuster, 2010; Döring & Bortz, 2016). Zur Prüfung von Hypothesen werden häufig statistische Signifikanztests genutzt. Die passende Auswahl eines Signifikanz-

tests hängt im Wesentlichen von der Art der Hypothese, der Anzahl der abhängigen Variablen, dem Skalenniveau der Variablen sowie den Verteilungseigenschaften der Variablen ab (Döring & Bortz, 2016).

Zur Prüfung von Zusammenhangshypothesen bei nominalen Variablen kann der χ^2-Test genutzt werden, sofern die erwarteten Häufigkeiten der Kontingenztabelle über 5 liegen. Andernfalls entspricht die Stichprobenverteilung der Teststatistik nicht mehr einer approximativen χ^2-Verteilung, wodurch die Teststatistik und damit der Signifikanztest ungenau werden. In diesem Fall kann die exakte Teststatistik mit dem exakten Test nach Fisher berechnet werden (Field et al., 2012). Als Maß für die Stärke des Zusammenhangs kann Cramér‘s V berechnet werden, welches gemäß den Richtlinien von Cohen (1988) interpretiert werden kann (Tab. 2).

Tabelle 2: Interpretation von Cramér‘s V gemäß Cohen (1988)

Effektstärke	Wertebereich
Kleiner Effekt	V = .10
Mittlerer Effekt	V = .30
Großer Effekt	V = .50

Eine Unterschiedshypothese hinsichtlich einer metrischen abhängigen Variable kann bei mehr als zwei unabhängigen Stichproben mithilfe der univariaten Varianzanalyse (engl. *Analysis of Variance*, ANOVA) überprüft werden. Die berechnete *F*-Statistik liefert zuverlässige Ergebnisse, wenn die Varianz der abhängigen Variable in jeder Gruppe vergleichbar ist (Homoskedastizität) und die abhängige Variable in jeder Gruppe normalverteilt ist (Field et al., 2012; Rasch et al., 2010b). Zur Prüfung auf Homoskedastizität kann der Levene-Test berechnet werden (Field et al., 2012). Bei größeren Stichproben ($n > 30$) approximiert die Verteilung der Stichprobe gemäß dem zentralen Grenzwertsatz an die Normalverteilung, weshalb die Normalverteilungsannahme hier meistens nicht explizit überprüft wird (Field et al., 2012; Rasch et al., 2010a). So wird auch im Rahmen dieser Arbeit lediglich auf Normalverteilung geprüft[2], wenn eine der zu vergleichenden Gruppen eine Stichprobengröße aufweist, die kleiner als 30 ist. Die Normalverteilungsannahme kann hier u. a. über die standardisierten Werte der Schiefe und Kurtosis geprüft werden:

$$z_{Schiefe} = \frac{Schiefe - 0}{2 \cdot SE_{Schiefe}} \qquad z_{Kurtosis} = \frac{Kurtosis - 0}{2 \cdot SE_{Kurtosis}} \tag{8}$$

[2] Die Prüfung der Verteilungsannahmen findet sich im Anhang.

Bei standardisierten Werten der Schiefe und Kurtosis, die über 1 oder unter -1 liegen, sollte die Normalverteilungsannahme bei kleineren Stichproben verworfen werden (Field et al., 2012).

Der berechnete *F*-Test verhält sich relativ robust bei Heteroskedastizität und fehlender Normalverteilung, wenn die Gruppengröße vergleichbar ist (Field et al., 2012). Demzufolge wird im Rahmen dieser Arbeit auf Homoskedastizität geprüft, wenn die Gruppen unterschiedlich groß sind. Sofern die Varianzen unterschiedlich sind, aber die Normalverteilungsannahme nicht verletzt ist, können die residualen Freiheitsgrade sowie die *F*-Statistik gemäß dem Vorgehen von Welch adjustiert werden. Bei fehlender Normalverteilung kann der Kruskal-Wallis-Test als nicht-parametrische Alternative berechnet werden (Field et al., 2012).

Da es sich bei der ANOVA um einen sogenannten Omnibus-Test handelt, wird lediglich auf einen Gesamteffekt getestet, der keine Informationen darüber liefert, welche Gruppen signifikante Unterschiede aufweisen. Zum paarweisen Vergleich der Gruppen können Post-Hoc-Tests durchgeführt werden. Hierbei wird auf den Post-Hoc-Test nach Bonferroni zurückgegriffen, der sich besonders zur Kontrolle des Typ I-Fehlers eignet (Field et al., 2012). Als nicht-parametrischer Post-Hoc-Test kann ein paarweiser Wilcoxon-Test mit Bonferroni-Korrektur durchgeführt werden (Field et al., 2012).

Als Maß für die Effektstärke der ANOVA bzw. des Kruskal-Wallis-Tests wird η^2 berechnet, welches gemäß Cohen (1988) interpretiert wird (Tab. 3).

Tabelle 3: Interpretation von η^2 gemäß Cohen (1988)

Effektstärke	Wertebereich
Kleiner Effekt	$\eta^2 = .01$
Mittlerer Effekt	$\eta^2 = .06$
Großer Effekt	$\eta^2 = .14$

Zur Analyse mehrerer abhängiger Variablen können multivariate Analyseverfahren genutzt werden. Dadurch können Zusammenhänge zwischen mehreren abhängigen Variablen modelliert und die Typ-I-Fehler-Kumulierung durch multiples Testen vermieden werden (Eid et al., 2017).

Die Prüfung von Unterschiedshypothesen mehrerer metrischer abhängiger Variablen kann mithilfe einer multivariaten Varianzanalyse (engl. *Multivariate Analysis of Variance*, MANOVA) durchgeführt werden. Die Durchführung einer MANOVA empfiehlt sich immer dann, wenn eine begründete Annahme über einen theoretischen und empirischen Zusammenhang der abhängigen Variablen besteht. Andernfalls können multiple ANOVAs durchgeführt

werden (Eid et al., 2017). Zudem sollten nicht mehr als zehn abhängige Variablen zugleich modelliert werden (Field et al., 2012).

Für die Durchführung der MANOVA muss auf multivariate Normalverteilung und die Gleichheit der Kovarianzmatrix geprüft werden, sofern die Gruppengrößen unterschiedlich sind. Andernfalls verhält sich die MANOVA weitgehend robust gegenüber Abweichungen der Verteilungsannahmen, wenn die MANOVA mit der Pillai-Statistik berechnet wurde (Eid et al., 2017; Field et al., 2012). Multivariate Normalverteilung kann mithilfe des multivariaten Shapiro-Tests überprüft werden. Die Gleichheit der Kovarianzmatrix kann mithilfe des Box-M-Tests untersucht werden, allerdings reagiert der Box-M-Test sensitiv gegenüber der Verletzung der multivariaten Normalverteilungsannahme und zeigt so auch bei Gleichheit der Kovarianzmatrix ein signifikantes Ergebnis an (Field et al., 2012). Deshalb wird bei fehlender multivariater Normalverteilung nicht auf Gleichheit der Kovarianzmatrix geprüft. Stattdessen kann eine nicht-parametrische MANOVA berechnet werden. Diese ermittelt die multivariaten Teststatistiken auf der Basis von Rängen und Permutationstests. Dadurch entfallen Verteilungsannahmen (Burchett et al., 2017). Klassische inferenzstatistische Verfahren wurden im Rahmen dieser Arbeit vornehmlich zur Auswahl relevanter Personenmerkmale für die Clusteranalyse (siehe Kapitel 8.5.3) und zur Evaluation der Kontextwahl (siehe Kapitel 8.5.6) genutzt.

7.4 Pfadanalyse

Neben der MANOVA können multivariate Strukturbeziehungen zwischen verschiedenen Variablen insbesondere durch Strukturgleichungsmodellierung (engl., *Structural equation modeling*, SEM) geprüft werden (Reinecke, 2014; Weiber & Mühlhaus, 2014). Ein vollständiges Strukturgleichungsmodell besteht aus einem Messmodell und einem Strukturmodell. In Messmodellen werden latente Variablen regressionsanalytisch durch beobachtbare Indikatorvariablen (Items) spezifiziert. Es wird angenommen, dass ein latenter Faktor ursächlich für die beobachteten Indikatorwerte ist (Gäde et al., 2020; Weiber & Mühlhaus, 2014). Diese Hypothese wird im Rahmen der CFA über die Kovarianz der Items geprüft (siehe Kapitel 7.1). Im Strukturmodell werden dann die Beziehungen der spezifizierten latenten Variablen analysiert (Eid et al., 2017).

Ein vollständiges Strukturgleichungsmodell erfordert die Schätzung einer großen Anzahl von Parametern, da sowohl die Parameter des Messmodells als auch die des Strukturmodells geschätzt werden müssen. Wie bereits angeführt, werden in der Literatur üblicherweise Stichprobenumfänge angegeben, die fünfmal so groß wie die Anzahl der zu schätzenden Parameter sind (Kline, 2005; Weiber & Mühlhaus, 2014). Alternativ können Pfadmodelle verwendet werden, um Wechselwirkungen zwischen manifesten Variablen zu formulieren und empirisch zu prüfen. Da Pfadmodelle nur ein Strukturmodell enthalten, reduziert sich die Anzahl der zu

schätzenden Parameter. Im Gegenzug erlaubt die Pfadmodellierung allerdings nicht die Kontrolle unsystematischer Messfehler (Eid et al., 2017; Reinecke, 2014).

Der erste Schritt in der Pfadanalyse ist die Spezifikation des Modells. Endogene Variablen sind hierbei immer Kriteriumsvariablen, die durch einen oder mehrere Prädiktoren erklärt werden. Prädiktorvariablen werden im Rahmen der Strukturgleichungsmodellierung als exogene Variablen bezeichnet; sie werden durch das Modell nicht erklärt. Intervenierende Variablen sind dagegen sowohl Prädiktor als auch Kriterium (Kline, 2005; Weiber & Mühlhaus, 2014). Ähnlich wie bei der CFA werden die Parameter des Modells dann iterativ – häufig über die Maximum-Likelihood-Methode – bestimmt. Ziel ist die Differenz zwischen der empirischen und modellimplizierten Kovarianzmatrix zu minimieren (Reinecke, 2014). Die Analyse der Modellpassung kann dann über die in Tabelle 1 aufgeführten Gütekriterien und Interpretationsempfehlungen vorgenommen werden (siehe auch Weiber & Mühlhaus, 2014 für eine Auflistung von Gütekriterien im Rahmen der SEM).

Zur Beurteilung der Modellpassung müssen Restriktionen in das Modell eingeführt werden. Modelle, in denen keine Parameter restringiert sind und in denen alle möglichen Pfade zwischen den enthaltenen Variablen spezifiziert sind, werden als saturiert bezeichnet. Saturierte Modelle beschreiben die Daten zwar perfekt, aber haben einen geringen Informationsgehalt (Reinecke, 2014; Weiber & Mühlhaus, 2014). Um Hypothesen bezüglich bestimmter Kausaleffekte zu testen, können Pfade aufgrund theoretischer Annahmen oder empirischer Ergebnisse auf null fixiert werden. Dieser Ansatz wird von Kline (2005, S. 145) als *model trimming* bezeichnet. Wenn beispielswiese a priori angenommen wird, dass eine spezifische Variable keinen Einfluss auf eine weitere Variable hat, kann dieser Pfad auf null restringiert werden. Alternativ können auch nicht signifikante Pfade schrittweise auf null fixiert werden (Kline, 2005; Weiber & Mühlhaus, 2014). Um zu überprüfen, ob die Restriktion zulässig ist, kann das restringierte Modell, mit dem nicht restringierten Modell durch einen Likelihood-Ratio-Test verglichen werden. Ein signifikantes Testergebnis impliziert hierbei, dass das restriktivere Modell zu bevorzugen ist (Byrne, 2013; Kline, 2005).

7.5 Mehrebenenmodelle

In experimentellen Studien – wie Teilstudie II – wird eine unabhängige Variable gezielt manipuliert, um Effekte auf eine oder mehrere abhängige Variablen zu messen, die häufig zu mehreren Messzeitpunkten erfasst werden. Studien mit einem solchen Design kombinieren einen *between-subject*- mit einem *within-subject*-Faktor und weisen somit ein sogenanntes *mixed-design* auf (Field et al., 2012). Traditionell werden die Daten solcher Studien in der frequentistischen Statistik mit Verfahren ausgewertet, die auf dem allgemeinen linearen Modell beruhen (z. B. ANOVA). Eine der Voraussetzungen für die Anwendung von Verfahren auf der Grundlage des allgemeinen linearen Modells ist die Unabhängigkeit der Residuen

(Backhaus et al., 2020; Field et al., 2012). Insbesondere bei Studien, in denen ein Merkmal bei einer Person zu verschiedenen Zeitpunkten gemessen wird, kann von einer Verletzung der Unabhängigkeitsannahme und einer Korrelation der Messwerte einer Person ausgegangen werden (u. a. Eid et al., 2017; Magezi, 2015; Meteyard & Davies, 2020; Singmann & Kellen, 2019). Mithilfe der Intraklassen-Korrelation (*Intraclass-correlation*, ICC) ρ kann quantitativ überprüft werden, inwieweit die Unabhängigkeitsannahme tatsächlich verletzt ist (Eid et al., 2017). Die Verletzung der Unabhängigkeitsannahme führt in der frequentistischen Statistik zu erhöhten Typ-I-Fehlern und zu kleinen Standardfehlern (Backhaus et al., 2020).

In sogenannten Mehrebenenmodellen[3] (engl. *Multilevel models*) können Abhängigkeiten zwischen Datenpunkten explizit modelliert werden. Solche Abhängigkeiten bestehen immer dann, wenn beispielsweise Messungen (Level 1) in Personen (Level 2) geschachtelt sind, aber auch wenn Lernende (Level 1) in Schulklassen (Level 2) geschachtelt sind (Snijders & Bosker, 2012). Mehrebenenmodelle basieren nicht auf der Annahme unabhängiger Residuen und liefern effiziente und unverzerrte Effektschätzungen, eine höhere statistische Power und keine erhöhten Typ-I-Fehler (Singmann & Kellen, 2019).

In Mehrebenenmodellen können sowohl zufällige als auch feste Effekte modelliert werden. Bei festen Effekten handelt es sich um Effekte, die im Rahmen des Experiments durch gezielt variierte Faktoren hervorgerufen wurden. Dabei kann es sich sowohl um Haupt- als auch Interaktionseffekte handeln (Field et al., 2012; Singmann & Kellen, 2019). Diese festen Effekte können sowohl auf der ersten Ebene (z. B. Effekt der Zeit auf wiederholte Messungen) als auch auf der zweiten Ebene (z. B. Effekt der experimentellen Variation auf Personen) verortet sein. Im Rahmen von experimentellen Studien wird häufig geprüft, ob die experimentelle Variation auf der zweiten Ebene den zeitlichen Verlauf der abhängigen Variable auf der Ebene der Messungen (Level 1) beeinflusst. Da sich die Variablen auf unterschiedlichen Ebenen befinden, spricht man auch von *Cross-Level Interaction* (Eid et al., 2017).

Effekte sind dagegen zufällig, wenn sie durch Faktoren ausgelöst werden, die eine zufällige Auswahl der Grundgesamtheit darstellen, über die im Rahmen des Experiments generalisiert wird (z. B. Personen oder Klassen). In der Literatur sind verschiedene Mehrebenenmodelle beschrieben, um die zufällige Effektstruktur zu spezifizieren (z. B. in Snijders & Bosker, 2012). Zur Modellierung der Abhängigkeit zwischen Messwerten einer Person können sogenannte Random-Intercepts in das Modell eingeführt werden (Random-Intercept-Modell). Dadurch werden personenspezifische Abweichungen vom Gesamtmittelwert der abhängigen Variablen modelliert, die berücksichtigen, dass einige Personen eine generell höhere Merkmalsausprägung in der abhängigen Variable zeigen als andere (Eid et al., 2017; Singmann &

[3] In der Literatur werden Mehrebenenmodelle auch als linear gemischte Modelle (linear mixed models) oder hierarchisch lineare Modelle (hierarchical linear models) bezeichnet.

Kellen, 2019). Wenn darüber hinaus angenommen wird, dass sich die Ausprägung der abhängigen Variable für jede Person zwischen den Messzeitpunkten ändert, können solche Annahmen durch Random-Slopes in das Modell aufgenommen werden. Dieses Modell wird auch als Random-Coefficients-Modell bezeichnet (Eid et al., 2017).

Eine besondere Herausforderung stellt die Spezifikation der zufälligen Effektstruktur des Modells dar, da die fehlende Berücksichtigung zufälliger Effekte die Wahrscheinlichkeit von Typ-I-Fehlern erhöht. Barr et al. (2013) empfehlen diesbezüglich, die aufgrund des Designs maximal mögliche zufällige Effektstruktur zu modellieren. Demzufolge sollen zunächst Random-Intercepts für jede Untersuchungseinheit (z. B. Personen) zum Modell hinzugefügt werden, die für fehlende Unabhängigkeit in den Daten sorgen. Anschließend werden Random-Slopes für jeden Innersubjektfaktor hinzugefügt (z. B. die Zeit). Zuletzt werden Random-Slopes für die Interaktionen zwischen den Innersubjektfaktoren berücksichtigt (Brauer & Curtin, 2018). Bei Konvergenzproblemen des Modells wird empfohlen, die zufällige Effektstruktur schrittweise zu reduzieren (Matuschek et al., 2017).

Statistische Hypothesentests können im Rahmen der Mehrebenenmodelle durch verschiedene Methoden durchgeführt werden (siehe Singmann & Kellen, 2019 sowie Kuznetsova et al., 2017). Weit verbreitet ist der Vergleich von geschachtelten Modellen mittels Likelihood-Ratio-Test (Meteyard & Davies, 2020). Der Likelihood-Ratio-Test liefert allerdings antikonservative *p*-Werte, wenn die Daten nicht balanciert sind oder wenn kleine Stichproben vorliegen (Kuznetsova et al., 2017; Luke, 2017). Stattdessen wird die Kenward-Roger-Approximation zur Berechnung von *p*-Werten für die festen Effekte empfohlen, bei der die Freiheitsgrade für einen *F*-Test geschätzt werden. Für die Kenward-Roger-Approximation müssen die Modellparameter mit der restricted Maximum-Likelihood Methode bestimmt werden (Luke, 2017; Meteyard & Davies, 2020; Singmann & Kellen, 2019). Im Rahmen der Mehrebenenmodellierung steht kein einheitliches Effektstärkemaß für einzelne Modellterme (z. B. Haupt- und Interaktionseffekte) zur Verfügung, was u. a. auf die Varianzzerlegung im Mehrebenenmodell zurückzuführen ist (siehe Rights & Sterba, 2019 für weitere Herausforderungen bei der Berechnung von Effektstärkemaßen bei Mehrebenenmodellen). Anstatt ein globales Effektstärkemaß für einzelne Modellterme zu berechnen, können zunächst Kontrastanalysen durchgeführt werden, um zu untersuchen, zwischen welchen Faktorstufen der festen Effekte signifikante Unterschiede bestehen (Field et al., 2012; Schad et al., 2020; Singmann & Kellen, 2019). Als Maß für die Effektstärke kann anschließend Pearson‘s *r* anhand der *t*-Werte und Freiheitsgrade in der Kontrastanalyse berechnet werden (Field et al., 2012). Pearson‘s *r* kann gemäß den Richtlinien von Cohen (1988) interpretiert werden (Tab. 4).

Tabelle 4: Interpretation von r gemäß Cohen (1988)

Effektstärke	Wertebereich
Kleiner Effekt	$r = .10$
Mittlerer Effekt	$r = .30$
Großer Effekt	$r = .50$

Diese Analysen sollten auf der Basis von *estimated marginal means* durchgeführt werden, welche die Abhängigkeiten in den Daten berücksichtigen (Singman & Kellen, 2019).

8 Teilstudie I: Untersuchung der Kontextwahl

8.1 Entwicklung des Lernmaterials

Für die erste Teilstudie wurden insgesamt fünf unterschiedliche Lernaufgaben entwickelt, die hinsichtlich des Kontextmerkmals variiert wurden. Das Lernmaterial wurde als digitales Lernmaterial konzipiert, da die Datenerhebung aufgrund der pandemiebedingten Schulschließungen teilweise im Onlineunterricht stattfinden musste (siehe Kapitel 8.3). Zur Entwicklung des Lernmaterials wurde die Software *iBooks Author*[4] der Firma Apple genutzt, mit der elektronische Bücher gestaltet werden können. Dadurch konnte das Lernmaterial auf einem Tabletcomputer bearbeitet werden.

Ausgangspunkt für die Entwicklung der kontextorientierten Lerneinheit ist nach Parchmann et al. (2008) die Festlegung der zu erwerbenden Kompetenzen. Zunächst musste ein geeigneter Fachinhalt ausgewählt werden, welcher in unterschiedlichen Kontexten erarbeitet werden kann. Es wurde explizit ein interessanter Fachinhalt ausgewählt, um zu gewährleisten, dass eine ausreichende Anzahl von Lernenden den innerfachlichen Kontext wählt. Die zentrale Annahme dabei ist, dass die Wahl des innerfachlichen Kontextes maßgeblich mit der Interessantheit des fachlichen Inhalts verknüpft ist. Da Habig (2017) durch eine Befragung von 907 Lernenden an Gymnasien und Gesamtschulen zeigen konnte, dass der Inhaltsbereich Säuren und Basen zu den interessantesten gehört, wurde das Inhaltsfeld *saure und alkalische Lösungen* für die Aufgabenkonstruktion ausgewählt (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2019).

Zur Auswahl geeigneter Kompetenzen wurde eine Analyse der bis dahin geltenden Lehrpläne in Nordrhein-Westfalen durchgeführt. Diese umfassten im Wesentlichen die Curricula Chemie für das G8- (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2008) und G9-Gymnasium (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2019) sowie den Lehrplan Naturwissenschaften für die Gesamtschule (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2013). Bei der Analyse wurden lediglich Kompetenzerwartungen aus dem Kompetenzbereich *Umgang mit Fachwissen* berücksichtigt. Die Kompetenzerwartungen aus den Lehrplänen wurden systematisiert, um einen möglichen Weg der Kompetenzentwicklung zu formulieren. Ein zentrales Ziel der Kompetenzentwicklung im Inhaltsfeld *saure und alkalische Lösungen* ist es, von einer phänomenologischen Betrachtung saurer und alkalischer Lösungen zu deren submikroskopischer Beschreibung zu gelangen. Am Ende der Konzeptentwicklung steht das Brönsted-Konzept, welches auf der Grundlage von Donator-Akzeptor-Wechselwirkungen prägend für den gesamten weiteren Verlauf des Chemieunterrichts ist (Pfeifer & Sommer, 2019). Für die Aufgabe wurde deshalb folgende zentrale Kompetenzerwartung formuliert:

[4] iBooks Author wird seit dem 01.07.2020 nicht mehr von Apple unterstützt.

„Die Lernenden können Säuren als Stoffe beschreiben, deren wässrige Lösungen Wasserstoff-Ionen enthalten und die Eigenschaften von sauren Lösungen mit dem Vorhandensein von charakteristisch hydratisierten Ionen erklären“

Diese Kompetenzerwartung erscheint aus mehreren Gründen geeignet: Zum einen müssten alle Lernenden über die notwendigen Lernvoraussetzungen verfügen, da die makroskopischen Eigenschaften von sauren Lösungen im dritten Lernjahr bereits mehrfach untersucht worden sein müssten. Zum anderen knüpft die Kompetenzerwartung damit an das oben formulierte Ziel an, makroskopische Phänomene auf submikroskopischer Ebene zu erklären. Durch die Förderung der Kompetenzerwartung wird ein Beitrag zur Verknüpfung der beiden Repräsentationsebenen und der Entwicklung des Stoff-Teilchen-Konzeptes geleistet.

Die bisherige Forschung zum kontextorientierten Lernen fokussiert in besonderem Maß die Effekte unterschiedlicher Themenfelder (siehe Kap. 4.4.2) und verweist auf bedeutende Unterschiede hinsichtlich der Interessantheit. In Bezug auf diese Untersuchung sollte das ausgewählte Themenfeld zwei Voraussetzungen erfüllen: Zum einen sollte das Themenfeld genügend Anknüpfungspunkte bieten, um merkmalsgetreue Kontexte zu entwickeln. Zum anderen sollten Geschlechtereffekte möglichst vermieden werden, um sowohl die Interessen von Jungen als auch Mädchen zu adressieren. Auf Grundlage der bestehenden Forschungslage wurde das Themenfeld *Körper und Krankheiten* ausgewählt, da es beiden Kriterien genügt. Sowohl Elster (2007) als auch Schiepe-Tiska, Simm und Schmidtner (2016) weisen darauf hin, dass Unterschiede zwischen den Geschlechtern in Bezug auf dieses Themenfeld (bzw. in Bezug auf medizinische Kontexte) geringer ausfallen als in anderen Themenfeldern. Weiterführend erlaubt das Themenfeld die Konstruktion von alltäglichen und besonderen Kontexten.

Auf dieser Grundlage wurden fünf unterschiedliche Lernaufgaben zum Inhaltsfeld *saure und alkalische Lösungen* und zum Themenfeld *Körper und Krankheiten* entwickelt, die jeweils in alltägliche und besondere Kontexte integriert wurden. Außerdem wurde ein innerfachlicher Kontext entwickelt. Die Titel der Kontexte und deren Merkmalszugehörigkeit sind in Tabelle 5 aufgeführt.

Tabelle 5: Titel der entwickelten Kontexte und deren Merkmalszugehörigkeit

Kontextmerkmal	Kontext
Alltäglich	Warum Zähneputzen so wichtig ist
	Unsere Verdauung
Besonders	Die Folgen der chronischen Typ-A-Gastritis
	Schädigung der Knochen durch Übersäuerung
Innerfachlich	Säuren im Labor

Die Entwicklung des Lernmaterials wurde mithilfe der Gestaltungsprinzipien der *Cognitive Theory of Multimedia Learning* (CTML) vorgenommen (Mayer, 2020). Diese Theorie bietet sich vor allem als Grundlage zur Gestaltung von Lernmaterialien an, um informationsverarbeitende Prozesse beim Lernen zu unterstützen. Dadurch soll die Kapazität des Arbeitsgedächtnisses weniger stark belastet werden, wodurch ein höherer Lernzuwachs erzielt werden soll. Zentrale theoretische Grundlage ist die *Cognitive Load Theory* (CLT) von Sweller et al. (1998). Bei der Entwicklung wurden folgende Gestaltungsmerkmale der CTML berücksichtigt (Mayer, 2020):

- *Multimedia Principle*: Das Lernen mit integriertem Text-Bild-Material fördert den Lernzuwachs mehr als nur Text-Material.
- *Spatial Contiguity Principle*: Die Lernenden profitieren von Lernmaterial, in dem korrespondierende Abbildungen und Textabschnitte räumlich nah beieinander präsentiert werden.
- *Signaling Principle*: Das Lernen mit multiplen Repräsentationen kann durch Hervorhebungen, Pfeile oder andere Signale unterstützt werden, um die Aufmerksamkeit der Lernenden auf relevante Informationen zu lenken. Die zu häufige Verwendung von signaling kann die kognitive Belastung allerdings erhöhen.
- *Coherence Principle*: Es sollen nur Abbildungen in das Lernmaterial aufgenommen werden, die für das Lernen relevante Informationen enthalten. Abbildungen, die für das Lernen irrelevante Informationen enthalten, können den Lernzuwachs reduzieren.
- *Segmenting Principle*: Informationsverarbeitende Prozesse bei der Bearbeitung des Lernmaterials können durch die Einteilung in kleinere Sinnabschnitte unterstützt werden, welche die Lernenden in eigener Geschwindigkeit bearbeiten können.
- *Personalisation Principle*: Die Lernenden profitieren beim Lernen von einer direkten Ansprache in der ersten und zweiten Person. Mayer (2020) weist allerdings darauf hin, dass eine zu häufige Nutzung auch nachteilig wirken kann.

Das Lernmaterial wurde gemäß dem *Segmenting Principle* durch drei Abschnitte strukturiert, die sich an der unterrichtlichen Gestaltung von kontextorientiertem Lernen (Parcmann et al., 2008) ausrichten: Begegnungsphase, Aneignungsphase und Übungsphase. Auf die Integration einer Phase der Neugier wurde verzichtet, da potenzielle Fragen der Lernenden durch die Lehrkraft systematisiert werden müssen, um den Lernprozess zu strukturieren. Um Effekte durch die Versuchsleitung oder Lehrkraft zu minimieren, wurde davon abgesehen. Jeder Abschnitt wurde letztendlich durch ein Piktogramm gekennzeichnet, wodurch das Lernmaterial stärker strukturiert wurde. Dafür wurde das jeweilige Piktogramm in die Kopfzeile eingefügt (siehe Lernmaterial im Anhang).

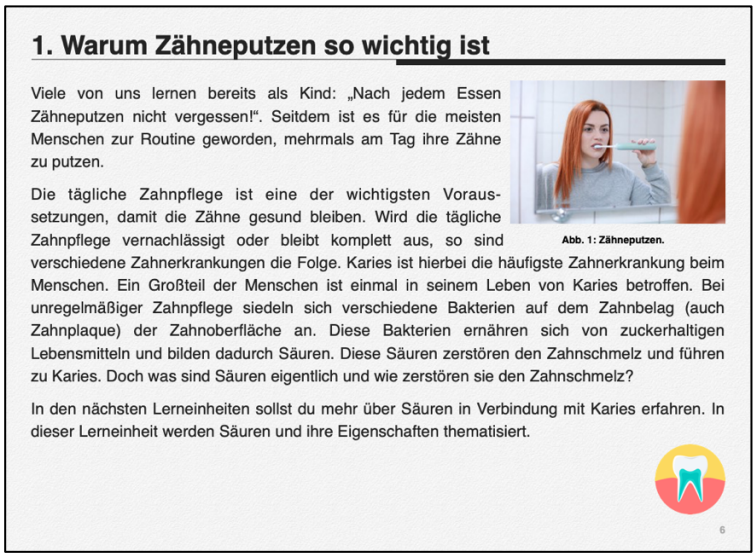
Vorab wurden die Lernenden über den Ablauf der Lerneinheit instruiert. Hier wurde explizit darauf hingewiesen, dass sich die zur Wahl gestellten Aufgaben lediglich im Hinblick auf den Kontext unterscheiden und mit jeder Aufgabe die gleichen chemischen Inhalte erlernt werden. Anschließend wurden die Lernenden durch fünf Kontextbeschreibungen in den Kontext eingeführt (Phase der Begegnung). Die Kontextbeschreibung wurden mithilfe der Standardisierungskriterien von van Vorst (2013) entwickelt, um den Einfluss externer Aufgabenmerkmale auf die Kontextwahl zu kontrollieren (Tab. 6).

Tabelle 6: Standardisierungskriterien bei der Entwicklung der Kontextbeschreibung

Kriterium	Ausgestaltung
Inhaltsfeld	Saure und alkalische Lösungen
Themenfeld	Körper und Krankheiten
Textlänge	Textblock 1: 2 Sätze
	Textblock 2: 7-8 Sätze
	Textblock 3: 2 Sätze
Optisches Layout	Schriftart: Helvetica, Schriftgröße: 22 pt, Schriftfarbe: schwarz, Zeilenabstand: 1,2-fach
	Drei Textblöcke
	Abbildung mit Gegenstand bzw. Situation des Kontextes in der oberen rechten Ecke
	Icon in der unteren rechten Ecke
Textschwierigkeit	Satzbau: Hauptsatz mit maximal einem Nebensatz

Die Kontextbeschreibung besteht aus drei Textblöcken und einer Abbildung in der oberen rechten Ecke, in der die Situation oder der Gegenstand des Kontextes dargestellt ist (Abb. 5).

Im ersten Textblock erfolgt eine kurze Einleitung in den Kontext. Daran anschließend wird im zweiten Textblock der Kontext inhaltlich dargestellt. Es werden nur Hintergrundinformationen zum Kontext und keine chemischen Fachinhalte erläutert. Der zweite Abschnitt endet mit der aus dem Kontext resultierenden Problemstellung. Der letzte Textblock ist in allen Beschreibungen identisch und weist noch einmal auf den entsprechenden Kontext sowie die nachfolgende Verwendung im Unterricht hin. Die Kontextbeschreibungen haben nahezu die gleiche Textlänge und sind hinsichtlich der Textschwierigkeit auf Hauptsätze mit maximal einem Nebensatz beschränkt. Ferner ist das optische Layout der Kontextbeschreibungen vollkommen identisch. Nachdem sich die Lernenden für einen Kontext entschieden haben, sind sie durch einen entsprechenden Link direkt zur dazugehörigen Kontextaufgabe gelangt, indem sie das Icon in der unteren rechten Ecke anklickten.



1. Warum Zähneputzen so wichtig ist

Viele von uns lernen bereits als Kind: „Nach jedem Essen Zähneputzen nicht vergessen!“. Seitdem ist es für die meisten Menschen zur Routine geworden, mehrmals am Tag ihre Zähne zu putzen.

Abb. 1: Zähneputzen.

Die tägliche Zahnpflege ist eine der wichtigsten Voraussetzungen, damit die Zähne gesund bleiben. Wird die tägliche Zahnpflege vernachlässigt oder bleibt komplett aus, so sind verschiedene Zahnerkrankungen die Folge. Karies ist hierbei die häufigste Zahnerkrankung beim Menschen. Ein Großteil der Menschen ist einmal in seinem Leben von Karies betroffen. Bei unregelmäßiger Zahnpflege siedeln sich verschiedene Bakterien auf dem Zahnbelag (auch Zahnplaque) der Zahnoberfläche an. Diese Bakterien ernähren sich von zuckerhaltigen Lebensmitteln und bilden dadurch Säuren. Diese Säuren zerstören den Zahnschmelz und führen zu Karies. Doch was sind Säuren eigentlich und wie zerstören sie den Zahnschmelz?

In den nächsten Lerneinheiten sollst du mehr über Säuren in Verbindung mit Karies erfahren. In dieser Lerneinheit werden Säuren und ihre Eigenschaften thematisiert.

6

Abbildung 5: Exemplarische Kontextbeschreibung (Teilstudie I)

Nachdem die Lernenden einen Kontext zum Lernen ausgewählt hatten (Begegnungsphase), wurden die relevanten chemischen Fachinhalte in der Aneignungsphase vermittelt. Diese orientiert sich zwar konzeptionell an der Phase der Erarbeitung aus ChiK, allerdings wurde auf den Einsatz kooperativer Arbeitsformen verzichtet, um potenzielle Störeinflüsse auf die abschließende Bewertung individueller, prozessbezogener Variablen zu minimieren. Im Zentrum steht trotzdem die schülerzentrierte Bearbeitung einer aufgeworfenen Problemstellung.

Die Aneignungsphase aller Kontextaufgaben wurde inhaltlich identisch gestaltet: Zunächst wird eine fachliche Problemstellung aufgeworfen, die aus dem jeweiligen Kontext resultiert. Die Herleitung dieser Problemstellung unterscheidet sich von Kontext zu Kontext, allerdings ist das fachliche Problem identisch. Nach der Einführung in die Problemstellung wird der fachliche Inhalt von den Lernenden erarbeitet. Die fachlichen Erklärungen unterscheiden sich

nicht in Abhängigkeit vom Kontext und sind identisch im Hinblick auf Textlänge, Wortlaut sowie Layout gestaltet. Zur Unterstützung des Lernprozesses werden in der Aneignungsphase die beschriebenen Gestaltungsprinzipien der CTML genutzt. Dem *Multimedia Principle* entsprechend wird der Text durch verschiedene Repräsentationen unterstützt. Diese umfassen vorrangig Modelle, Wortgleichungen und Reaktionsgleichungen, die räumlich nah an den korrespondierenden Text eingefügt sind (*Spatial Contiguity Principle*). In diesen Repräsentationen werden wichtige Merkmale gemäß dem *Signaling Principle* hervorgehoben. Die Repräsentationen in den Kontextaufgaben unterscheiden sich teilweise aufgrund des unterschiedlichen Kontexts voneinander, allerdings sind sie – wenn möglich – vergleichbar gestaltet. Darüber hinaus werden die Lernenden gemäß dem *Personalization Principle* während der Lerneinheit im Material direkt angesprochen. Zum Ende der Aneignungsphase wird das eingangs aufgeworfene Problem mithilfe des vermittelten Fachinhalts auf Kontextebene beantwortet.

Nach der Aneignungsphase sollten die Lernenden das gelernte Wissen in einer Übungsphase festigen. Hierfür wurden drei Übungsaufgaben mit gestuftem Anforderungsniveau entwickelt. Zur systematischen Variation des Anforderungsniveaus der Aufgaben wurde das ESNaS-Kompetenzstrukturmodell genutzt (Walpuski et al., 2010). Das Kompetenzstrukturmodell wurde ursprünglich für die Entwicklung von Testaufgaben entwickelt, hat sich in der Vergangenheit aber auch für die Entwicklung von Lernaufgaben bewährt (z. B. Hauerstein, 2019).

Im ESNaS-Kompetenzstrukturmodell können Aufgaben über drei Dimensionen beschrieben werden: Kompetenzbereiche, kognitive Prozesse, Komplexität. Die Kompetenzbereiche unterscheiden lediglich zwischen Aufgabenkategorien. Die Aufgabenschwierigkeit wird im Wesentlichen durch die erforderlichen kognitiven Prozesse und die Komplexität determiniert.

Die kognitiven Prozesse leiten sich aus Informationsverarbeitungsstrategien ab, die auf die Bearbeitung von Aufgaben übertragen wurden. Walpuski et al. (2010) beschreiben vier kumulative Stufen der kognitiven Prozesse: Reproduzieren, Selegieren, Organisieren und Integrieren. Ropohl (2010) konnte hierbei zeigen, dass Aufgaben, die systematisch im Hinblick auf die kognitiven Prozesse variiert wurden, in der Lösungswahrscheinlichkeit keine signifikanten Unterschiede zwischen dem *Reproduzieren* und *Selegieren* sowie *Organisieren* und *Integrieren* zeigen. Demnach werden die Prozesse *Reproduzieren* und *Selegieren* zum übergeordneten Prozess *Wiedergeben* zusammengefasst. Analog bilden die kognitiven Prozesse *Organisieren* und *Integrieren* den übergeordneten Prozess *Anwenden*.

Die Aufgabenschwierigkeit beim übergeordneten kognitiven Prozess *Wiedergeben* wird maßgeblich davon beeinflusst, ob die Informationen zum Lösen der Aufgabe im Aufgabenstamm enthalten sind oder nicht (Ropohl, 2010). Die entsprechenden Lernaufgaben wurden so konzipiert, dass die Informationen zum Lösen der Aufgabe zwar nicht im Aufgabenstamm, aber in der Aneignungsphase enthalten sind.

Die Komplexität beschreibt „…welchen Umfang die zu bearbeitenden Inhaltsstrukturen haben…“ (Walpuski et al., 2010, S. 176). Hier differenzieren die Autoren zwischen einem Fakt, zwei Fakten, einem Zusammenhang, zwei Zusammenhängen und dem übergeordneten Konzept. Für die Aufgabenkonstruktion in dieser Arbeit wird auf der Grundlage von Ropohl (2010) und Ferber (2014) lediglich zwischen Fakt, Zusammenhang und übergeordnetem Konzept unterschieden. Fakten umfassen dabei (Fach-)Begriffe, Eigenschaften und Symbole. Zusammenhänge adressieren funktionale Beziehungen zwischen zwei Fakten. Konzepte sind übergeordnete Regeln, Gesetze oder Prinzipien (Ferber, 2014; Ropohl, 2010; Walpuski et al., 2010).

Daraus ergibt sich ein Konstruktionsdesign für die Teilaufgaben, welches in Tabelle 7 dargestellt ist. Alle Teilaufgaben fokussieren die Förderung der festgelegten Kompetenzerwartung aus dem Inhaltsfeld saure und alkalische Lösungen. Jede Teilaufgabe besteht aus einem Aufgabenstamm und der eigentlichen Aufgabenstellung. Im Aufgabenstamm wird ein Bezug zum Kontext der jeweiligen Lernaufgabe hergestellt. Dementsprechend unterscheidet sich der Aufgabenstamm in den fünf Lernaufgaben. Hierbei wurden ähnliche Textumfänge und Satzstrukturen verwendet, um die Teilaufgaben vergleichbar zu gestalten. Die eigentliche Aufgabenstellung jeder Teilaufgabe ist in allen Lernaufgaben vollkommen identisch.

Tabelle 7: Konstruktionsdesign auf Basis der kognitiven Prozesse und Komplexität im ESNaS-Kompetenzstrukturmodell

	Wiedergeben	Anwenden
Fakt	Übungsaufgabe 1	
Zusammenhang	Übungsaufgabe 2	
Konzept		Übungsaufgabe 3

Anmerkung: Zu der ausgegrauten Zelle ist die Konstruktion von Aufgaben nicht sinnvoll, da Fakten nicht angewandt werden können.

In der ersten Übungsaufgabe müssen Fakten wiedergegeben werden, die in der Aneignungsphase dargestellt wurden. Mit Bezug auf das Arrhenius-Konzept sollen die Lernenden hier lediglich beschreiben, in welche Teilchen Säuremoleküle im Wasser aufgespalten werden. Die Wiedergabe eines Zusammenhangs ist für die Lösung der zweiten Teilaufgabe notwendig. Hier gilt es, den Zusammenhang zwischen der Dissoziation der Säuremoleküle und den Eigenschaften von Säuren herzustellen. In der letzten Teilaufgabe muss das Gesamtkonzept auf einen Spezialfall angewandt werden. Die Lernenden müssen in dieser Aufgabe die Dissoziation eines bisher nicht thematisierten Säuremoleküls durch eine Wortgleichung oder Reaktionsgleichung beschreiben und die Folgen für den jeweiligen Kontext erklären. Um die

Lernenden dabei zu unterstützen, werden die Strukturformel des jeweiligen Säuremoleküls abgebildet und der Säurerest rot markiert. Die genutzten Säuremoleküle unterscheiden sich von Kontext zu Kontext, da aus inhaltlichen Gründen nicht für jeden Kontext das gleiche Molekül genutzt wird.

Die Schülerinnen und Schüler können eine Musterlösung der Übungsaufgabe auf der jeweils nachfolgenden Seite einsehen. Es ist anzunehmen, dass der selbstständige Vergleich der eigenen Antwort mit der Musterlösung den Kompetenzerwerb positiv beeinflusst (Hattie, 2012; Raguse & Weber-Peukert, 2019).

8.2 Beschreibung der Messinstrumente

Die erste Teilstudie dieses Forschungsvorhabens erforderte die Operationalisierung zahlreicher Personenmerkmale, die mit dem gewählten Kontextmerkmal in Verbindung stehen könnten. Zusätzlich wurden Kontrollvariablen erfasst, die die Bewertung der Prozessvariablen beeinflussen können. Prozessbezogene Variablen wurden durch geeignete Messinstrumente operationalisiert, die als Indikatoren für die Passung des gewählten Kontextmerkmals dienen. Es musste außerdem ein geeignetes Instrument zur Überprüfung der Merkmalszugehörigkeit der Kontextaufgaben adaptiert werden. Nachfolgend werden alle Messinstrumente ausführlich beschrieben.

8.2.1 Erfassung der kognitiven und affektiv-motivationalen Personenmerkmale

Demografische Daten: Als relevante demografische Daten der Versuchspersonen wurden der Name der Schule, das Geschlecht, das Alter sowie die letzte Zeugnisnote im Fach Chemie erhoben. Insbesondere die Zeugnisnote im Fach Chemie hat sich in anderen Untersuchungen als relevantes Personenmerkmal für die Wahl eines Kontextes erwiesen (Habig et al., 2018). Angesichts der bestehenden Forschungslage scheint auch das Geschlecht von Bedeutung für die Kontextwahl zu sein. Ferner wurden die demografischen Daten zur Beschreibung der Stichprobe genutzt. Die demografischen Daten wurden durch offene Items (Name der Schule, Alter) oder Multiple-Choice-Items (Geschlecht, Zeugnisnote) erhoben.

Leseverständnis: Als einzige Kontrollvariable wurde das Leseverständnis als Teilfacette der Lesekompetenz mithilfe des Lesegeschwindigkeits- und -verständnistests (LGVT) von W. Schneider et al. (2017) gemessen. Der LGVT eignet sich zur Messung der Lesegeschwindigkeit, der Lesegenauigkeit und des Leseverständnisses in den Klassenstufen 5 bis 13. Für die vorliegende Arbeit wurde lediglich der Testwert zur Erfassung des Leseverständnisses berücksichtigt. Die Autoren stellen drei Paralleltexte zur Verfügung, die zur Messung genutzt werden können. Im Rahmen dieser Arbeit wurde der Text Laufbursche verwendet. Während

der Testung haben die Versuchspersonen sechs Minuten Zeit, den gesamten Text zu lesen. Dabei gilt es an 47 Textstellen eins von drei vorgegebenen Wörtern auszuwählen, welches am besten in den Sinnzusammenhang des Textes passt. Die Versuchspersonen werden vor der eigentlichen Testsituation anhand von zwei Beispielen mit dem Testformat vertraut gemacht.

Vorwissen in Chemie: Die bisherige Forschung weist das Vorwissen als relevantes Personenmerkmal für die Passung von Kontextmerkmal und Lernenden aus (Habig et al., 2018; van Vorst & Aydogmus, 2021). Bislang wurde das Vorwissen allerdings unzureichend über die letzte Zeugnisnote im Fach Chemie operationalisiert. Dieses Forschungsdesidarat wird aufgegriffen, indem das Vorwissen im Fach Chemie durch einen psychometrischen Leistungstest erfasst wird. Hierfür wird auf einen Fachwissenstest von Celik (2022) zurückgegriffen.

Der Fachwissenstest von Celik (2022) erfasst 57 Kernideen aus dem Kompetenzbereich *Umgang mit Fachwissen* entlang der drei Basiskonzepte *Struktur der Materie*, *Chemische Reaktion* und *Energie*, die in den ersten beiden Lernjahren erlernt werden sollten (Tab. 8). Eine Kernidee ist dabei „…eine grundlegende Vorstellung oder Annahme der Chemie…, die zum Verständnis eines chemiespezifischen Themen- oder Inhaltsfeldes beiträgt." (Celik, 2022, S. 42).

Tabelle 8: Anzahl der Kernideen pro Basiskonzept und Lernjahr mit ausgewähltem Beispiel

Basiskonzept	Lernjahr	Anzahl	Beispiel für eine Kernidee
Struktur der Materie	1	13	Stoffe bestehen aus Teilchen.
	2	19	Atome können Ionen bilden.
Chemische Reaktion	1	6	Bei chemischen Reaktionen werden die Atome neu gruppiert.
	2	7	Atome reagieren so, dass sie Edelgaskonfiguration erreichen.
Energie	1	7	Es gibt verschiedene Formen von Energie.
	2	5	Durch Energieumwandlung wird elektrische Energie bereitgestellt.

Im ursprünglichen Messinstrument werden die 57 Kernideen insgesamt durch 348 Test-items überprüft. Bei den Testitems handelt es sich um *Multiple-Choice-Single-Select-Items,* die neben dem richtigen Attraktor, drei Distraktoren beinhalten. Für diese Studie wird die Verwendung von einem Item pro Kernidee als hinreichend angesehen, um das Vorwissen aus den ersten beiden Lernjahren valide zu erfassen. Dementsprechend war es notwendig, ein Item pro Kernidee auszuwählen.

Die Auswahl von geeigneten Testitems erfolgte auf Grundlage der berechneten Schwierigkeitsparameter und der Kennwerte des Rasch-Modells aus dem Datensatz von Celik (2022). Hier wurde besonders darauf geachtet, Testitems auszuwählen, die ein breites Kompetenzspektrum abbilden. Außerdem wurde auf die Auswahl von Items verzichtet, die sich in der Analyse von Celik (2022) als nicht Rasch-konform erwiesen haben. Die Liste der so selektierten Items ist im Anhang einzusehen.

Trotz der so deutlich reduzierten Itemzahl war es nicht möglich, dass alle Lernenden jedes Testitem bearbeiten konnten. Ähnlich wie in vielen Large-Scale-Studien (z. B. PISA) wurde den Lernenden jeweils nur ein Teil der Testitems vorgelegt, die über ein *Balanced-Incomplete-Block-Design* (BIBD) verknüpft waren (Frey et al., 2009). Das Vorwissen der Lernenden konnte so mithilfe eines Rasch-Modells geschätzt werden. Für das BIBD wurden die 57 Testitems auf zehn Blöcke verteilt. Jeder Block bestand aus sechs Items zum Basiskonzept *Struktur der Materie*, zwei Items zum Basiskonzept *Chemische Reaktion* und zwei Items zum Basiskonzept *Energie*. Jeweils die Hälfte der Items pro Basiskonzept prüfen Kernideen aus dem ersten und zweiten Lernjahr. Der Anteil der Items pro Block orientierte sich dabei am Anteil der Items pro Kernidee und Lernjahr im gesamten Itempool. Aufgrund der ungleichen Itemzahl pro Kernidee war es konzeptionell nicht möglich, die Items gleichmäßig auf die Blöcke zu verteilen. So ist jedes Item ein bis dreimal in den zehn Blöcken zu finden (siehe Anhang). Bei diesem Design wird jeder Block dann in zwei Testheften berücksichtigt (Tab. 9). Den Lernenden wurde zufällig eines der zehn Testhefte zugeteilt.

Tabelle 9: Verteilung der Itemblöcke auf die Testhefte

Block	Testheft				
	T1	T2	T3	…	T10
B1	X				X
B2	X	X			
B3		X	X		
B4			X	…	
…				…	
B10					X

Individuelles Interesse und Motivation im Fach Chemie: Neben dem Vorwissen scheint auch das individuelle Interesse an Chemie für die Kontextwahl von Bedeutung zu sein (Habig et al., 2018; van Vorst & Aydogmus, 2021). Daneben beschreibt die SEVT den interest-enjoyment value als einen Faktor, der die Aufgabenwahl determiniert und Überschneidungen zum Interessenkonzept aufweist (Eccles & Wigfield, 2020). Das individuelle Interesse an Chemie wurde deshalb entsprechend der Personen-Gegenstands-Konzeption als habituelles Personenmerkmal durch drei psychometrische Skalen operationalisiert (Tabelle 10). Dabei handelt es sich um Skalen, die sich in vielen Forschungsarbeiten bewährt haben (Fechner, 2009; Habig et al., 2018; Kölbach & Sumfleth, 2013; van Vorst & Aydogmus, 2021; van Vorst et al., 2018). Gemäß der Operationalisierung in der IPN-Interessenstudie wird hierbei zwischen dem Fachinteresse und dem Sachinteresse unterschieden. Da diese beiden Skalen besonders die emotionale Valenz des individuellen Interesses adressieren, wurde das Instrument durch eine Skala zur wertbezogenen Valenz ergänzt. Das Instrument wurde maßgeblich von Fechner (2009) konzipiert, die dabei auf verschiedene Quellen zurückgegriffen hat. Die Items zur Erfassung des Fachinteresses stammen ursprünglich von Sumfleth und Wild (2005), die wiederrum Items von Wild und Krapp (1995) für das Fach Chemie adaptierten. Die Skala von Wild und Krapp (1995) ist eine übersetzte Version des Instruments von Ryan und Connell (1989). Die Items zum Sachinteresse wurden dem Potsdamer-Motivations-Inventar von Rheinberg und Wendland (2002) entnommen. Alle Items zur wertbezogenen Valenz wurden von Fechner (2009) konzipiert.

In der SEVT wird der attainment value (die Wichtigkeit) als eine weitere Determinante der Aufgabenwahl beschrieben (Eccles & Wigfield, 2020). Deshalb wurde eine Skala zur Erfassung des attainment value von Trautwein et al. (2012) für die Verwendung in dieser Studie ins Deutsche übersetzt und für das Fach Chemie adaptiert. Da die Autoren und Autorinnen in ihrer Untersuchung einen sehr hohen Zusammenhang zum interest-enjoyment value nachweisen konnten (r = .97), wird die Skala in das Instrument zur Erfassung des individuellen Interesse an Chemie aufgenommen.

Weiterhin ist anzunehmen, dass das themenspezifische Interesse von Bedeutung für den gewählten Kontext ist. Aus diesem Grund wurde jeweils eine Skala zur Erfassung des individuellen Interesses am Inhaltsfeld saure und alkalische Lösungen, am Themenfeld Körper und Krankheiten sowie an alltäglichen und besonderen Kontexten ergänzt. Auf die Erhebung des Interesses am innerfachlichen Kontext wurde verzichtet, da anzunehmen ist, dass das Interesse am innerfachlichen Kontext sehr stark mit dem Interesse am Inhaltsfeld zusammenhängt. Zur Erfassung des Interesses am Inhaltsfeld saure und alkalische Lösungen wurden zwei Items von Habig (2017) adaptiert, ergänzend wurde ein weiteres Item neu entwickelt. Dadurch sollte sichergestellt werden, dass die psychometrische Skala aus drei Items besteht (Döring & Bortz, 2016). Die Items zum Interesse am Themenbereich Körper und Krankheiten gehen auf Items von Elster (2007) und Schiepe-Tiska, Simm und Schmidtner (2016) zurück.

Für die Erfassung des Interesses an besonderen und alltäglichen Kontexten konnten Items von van Vorst (2013), Fechner (2009) und Haugwitz (2009) übernommen werden.

Tabelle 10: Überblick über die eingesetzten Skalen zur Erfassung des individuellen und themenspezifisches Interesses an Chemie

Skala	Itemanzahl	Beispiel
Individuelles Interesse		
Fachinteresse (*fin*)	5	Im Chemieunterricht fühle ich mich wohl.
Sachinteresse (*sin*)	6	Was wir im Chemieunterricht machen, interessiert mich.
Wertbezogene Valenz (*wert*)	3	Das, was man im Chemieunterricht lernt, kann man gut gebrauchen.
Wichtigkeit (*imp*)	3	Ich bin sehr daran interessiert, viel in Chemie zu lernen.
Themenspezifisches Interesse		
Individuelles Interesse am Inhaltsfeld saure und alkalische Lösungen (*sb*)	3	Ich interessiere mich für die Eigenschaften von sauren und basischen Lösungen.
Individuelles Interesse an Kontexten zum Thema Körper und Krankheiten (*kk*)	3	Ich finde es spannend, den menschlichen Körper aus chemischer Sicht zu betrachten.
Individuelles Interesse an besonderen Kontexten (*bes*)	3	Ich finde es interessant, wenn wir im Chemieunterricht etwas über außergewöhnliche Dinge lernen.
Individuelles Interesse an alltäglichen Kontexten (*alk*)	3	Ich finde es gut, wenn ich etwas darüber lerne, was Chemie mit meinem eigenen Leben zu tun hat.

Die Motivation im Chemieunterricht wurde insgesamt durch neun Items erfasst (Tab. 11). Entsprechend der Taxonomie von Deci und Ryan wurde zwischen der intrinsischen und extrinsischen Motivation unterschieden. Während die intrinsische Motivation konzeptionelle Überschneidung zum interest-enjoyment value der SEVT zeigt, kann angenommen werden,

dass die extrinsische Motivation den utitility value der SEVT abbildet (Eccles & Wigfield, 2020). Zur Messung der intrinsischen Motivation wurden vier Items von Fechner (2009) genutzt, die auf Rheinberg und Wendland (2002) zurückgehen. Die Items zur extrinsischen Motivation wurden dem *Science Motivation Questionnaire* von Glynn und Koballa (2006) entnommen. Für die Verwendung in dieser Studie wurden die Items ins Deutsche übersetzt.

Tabelle 11: Überblick über die eingesetzten Skalen zur Erfassung der Motivation im Chemieunterricht

Skala	Itemanzahl	Beispiel
Intrinsische Motivation (*inm*)	4	Wenn ich mich mit Chemie beschäftige, vergesse ich alles um mich herum.
Extrinsische Motivation (*exm*)	5	Es ist für mich persönlich wichtig, eine gute Note in Chemie zu bekommen.

Alle Items wurden von den Probanden und Probandinnen auf einer vierstufigen Likert-Skala beantwortet. Die Ausprägungen reichen dabei von *stimmt gar nicht* (1) bis *stimmt völlig* (4).

Berufswahlperspektive: Da die zukünftige Berufswahl möglicherweise mit dem gewählten Kontextmerkmal zusammenhängt, wurde die Berufswahlperspektive mithilfe von zwei Items erhoben, die in der PISA-Studie 2000 genutzt wurden (Kunter et al., 2002). Anhand des ersten Items wurde erfragt, ob zum jetzigen Zeitpunkt ein spezifischer Berufswunsch vorliegt. Dieses Item konnte von den Lernenden mit Ja oder Nein beantwortet werden. Wenn die Lernenden das erste Item mit Ja beantworteten, konnten sie in einem weiteren offenen Item den spezifischen Berufswunsch angeben.

Die Systematisierung der angegebenen Berufe erfolgte analog zum Vorgehen von Hülsmann (2015). Für die Auswertung wurde zwischen naturwissenschaftlichen, medizinischen und sonstigen Berufen unterschieden, da das gewählte Kontextmerkmal möglicherweise nützlich für Berufsaspirationen im naturwissenschaftlichen oder medizinischen Bereich sein könnte. Dabei wurden die Berufe anhand der Systematiknummer (KldB 2010) der Bundesagentur für Arbeit geordnet (Bundesagentur für Arbeit, 2021). Hierfür wurden die Berufshauptgruppen genutzt. Durch dieses Vorgehen ergeben sich die in Tabelle 12 dargestellten Kategorien.

Tabelle 12: Kodierung der angegeben Berufsaspirationen mit Kategorie, KldB und Beispiel

	Kategorie	KldB	Beispiel
1	Kein Berufswunsch	-	-
2	Naturwissenschaftlicher Berufswunsch	[41]	Chemikant, Chemielaborant
3	Medizinischer Berufswunsch	[81]	Krankenschwester, Ärztin
4	Sonstiger Berufswunsch	Rest	Flugbegleiter, Polizist

Freizeitinteresse und -aktivitäten: Die Freizeitinteressen und -aktivitäten wurden durch Items erfasst, die auf der Basis eines Instruments aus der Shell-Jugendstudie 2019 konzipiert wurden (Albert et al., 2019). Das ursprüngliche Instrument erfragt die Freizeitinteressen und -aktivitäten auf einer Nominalskala. Für die Verwendung in dieser Studie wurde ein Instrument konzipiert, welches die spezifischen Freizeitinteressen und -aktivitäten anhand von 15 Aussagen (z. B. „In meiner Freizeit lese ich gerne Bücher“) erfasst, die auf einer vierstufigen Likert-Skala von *stimmt gar nicht* (1) bis *stimmt völlig* (4) zu bewerten sind.

Chemiebezogenes Selbstkonzept: Basierend auf der SEVT können die Erfolgserwartung sowie das darauf wirkende Selbstkonzept als wesentliche Determinanten für die Aufgabenwahl herausgestellt werden. Im Gegensatz zum Selbstkonzept ist die Erfolgserwartung kein zeitlich überdauerndes Personenmerkmal, sondern expliziter an spezifische Aufgaben und Situationen gebunden (Eccles & Wigfield, 2020). Demzufolge wird im Rahmen dieser Studie nur das Selbstkonzept als Personenmerkmal berücksichtigt – besonders, da das Selbstkonzept und die Erfolgserwartung in vorangegangenen Studien nicht als empirisch trennbare Konstrukte operationalisiert werden konnten. Neben dem Selbstkonzept wurde außerdem die Selbstwirksamkeitserwartung erfasst. Sowohl die Items zum Selbstkonzept als auch zur Selbstwirksamkeitserwartung stammen von Habig (2017), der die Items für das Fach Chemie von Hoffmann et al. (1998) adaptierte (Tab. 13). Auch hier wurden alle Items auf einer vierstufigen Likert-Skala von *stimmt gar nicht* (1) bis *stimmt völlig* (4) bewertet.

Tabelle 13: Überblick über die eingesetzten Skalen zur Erfassung des chemiebezogenen Selbstkonzepts

Skala	Itemanzahl	Beispiel
Selbstkonzept (*sk*)	8	Chemie fällt mir leicht.
Selbstwirksamkeitserwartung (*swe*)	3	Wenn ich mich anstrenge, komme ich im Chemieunterricht problemlos mit.

Motive zur Wahl einer Kontextaufgabe: In der Untersuchung von van Vorst und Aydogmus (2021) erwiesen sich die Wahlmotive als bedeutendes Merkmal zur Unterscheidung von Lernendengruppen bei der Kontextwahl. Folglich wurde die von den Autorinnen entwickelte Skalen auch im Rahmen dieser Studie zur Erfassung der Wahlmotive eingesetzt. Durch Interviews mit Lernenden der neunten Jahrgangsstufe sowie einer explorativen Faktorenanalyse wurden vier Wahlmotive bei der Kontextwahl extrahiert: Interesse und Neugier, Überraschende Informationen, Persönlicher Bezug sowie Bild und Überschrift. Jedes Wahlmotiv wurde durch eine Skala berücksichtigt (Tab. 14). Die Wahlmotive wurden auf einer vierstufigen Likert-Skala von *stimmt gar nicht* (1) bis *stimmt völlig* (4) bewertet.

Tabelle 14: Überblick über die eingesetzten Skalen zur Erfassung der Wahlmotive

Skala	Itemanzahl	Beispiel
Interesse & Neugier (*inneu*)	6	…weil mich das Thema der Aufgabe interessiert.
Überraschungsmoment (*uber*)	5	…weil mich einige Informationen im Text sehr überrascht haben.
Persönlicher Bezug (*per*)	4	…weil die Aufgabe über ein Thema handelt, das mir auch im Alltag begegnet.
Bild & Überschrift (*biub*)	4	…weil ich die Überschrift sehr interessant finde.

8.2.2 Erfassung der prozessbezogenen Variablen

Zufriedenheit: Zum Zeitpunkt der Konstruktion der Erhebungsinstrumente stand noch kein erprobtes Instrument zur Erfassung der Zufriedenheit mit kontextbasierten Aufgaben zur Verfügung. Demzufolge wurde im Rahmen dieser Studie ein neues Testinstrument konzipiert. Dafür musste das interessierende Merkmal zunächst spezifiziert werden (Brandt & Moosbrugger, 2020; Eid & Schmidt, 2014).

Bisher wurde in der Forschung zum kontextorientierten Lernen noch keine hinreichende Definition zur Zufriedenheit mit kontextbasierten Aufgaben erarbeitet. Deshalb wird auf eine Definition zur Beschreibung der Zufriedenheit mit Konsumentenprodukten zurückgegriffen. Hierin wird Zufriedenheit wie folgt definiert: „…A summary affective response of varying intensity…with a time-specific point of determination and limited duration…directed toward focal aspects of product aquisition and/or consumption." (Giese & Cote, 2000, S. 15). Auf dieser Grundlage können zentrale Merkmale für eine kontextaufgabenbezogene Zufriedenheit

abgeleitet werden: Es handelt sich um eine übergeordnete affektive Reaktion, die zeitlich begrenzt aus zentralen Aspekten der Aufgabenbearbeitung resultiert. Der zentrale Aspekt bei der Aufgabenbearbeitung ist dabei das variierte Kontextmerkmal. Insgesamt besteht damit eine gewisse Überschneidung zum situationalen Interesse, welches aus der Interessantheit des Gegenstandes entsteht, allerdings stärker als psychischer Zustand konzeptualisiert wird, während die Zufriedenheit etwas unspezifischer ist.

Daran anknüpfend wurde die Zufriedenheit durch vier umschreibende Aussagen in Form von Testitems operationalisiert. Hierbei wurde u. a. die erneute Wahl der Kontextaufgabe als ein zentraler Indikator für die Zufriedenheit mit der Kontextaufgabe zur Itemkonstruktion genutzt (van Vorst & Aydogmus, 2021). Der Kontextbegriff wurde hierbei mit dem Begriff *Thema* umschrieben, da anzunehmen ist, dass die Lernenden mit dem Kontextbegriff nicht vertraut sind. Der Titel des gewählten Kontextes wurde zur Gewährleistung der Inhaltsvalidität der Items immer hinter der Bezeichnung *Thema* eingeblendet (Beispielitem: Ich würde das Thema (Warum Zähneputzen so wichtig ist) erneut wählen). Um Verzerrungseffekte durch Zustimmungstendenzen zu vermeiden, wurde ein Item negativ formuliert (Moosbrugger & Brandt, 2020). Die vier Aussagen mussten auf einer vierstufigen Likert-Skala von *stimmt gar nicht* (1) bis *stimmt völlig* (4) bewertet werden.

Situationales Interesse: Das situationale Interesse wurde gemäß der Person-Gegenstands-Konzeption durch drei Skalen operationalisiert (Tab. 15), welche die jeweils postulierten Subdimensionen erfassen (siehe Kapitel 3.1.3).

Tabelle 15: Überblick über die eingesetzten Skalen zur Erfassung des situationalen Interesse

Skala	Itemanzahl	Beispiel
Emotionale Valenz (*emo*)	3	Es hat mir Spaß gemacht dieses Thema zu bearbeiten
Wertbezogene Valenz (*wert*)	3	Das Thema der Aufgabe war für mich von persönlicher Bedeutung
Epistemische Komponente (*epi*)	5	Ich würde gerne mehr über die Themen lernen, die in der heutigen Aufgabe behandelt wurden.

Zur Messung der emotionalen und wertbezogenen Valenz wurden jeweils drei Items genutzt, die von Habig (2017) adaptiert wurden. Die Formulierung eines Items zur wertbezogenen Valenz wurde für die Verwendung in dieser Studie angepasst. Ferner konnte die epistemische

Komponente durch eine weitere Skala erfasst werden, die aus fünf Items besteht. Auch bei den Items zur Erfassung des situationalen Interesses wurde der Titel des gewählten Kontextes hinter der Bezeichnung *Thema* eingeblendet. Alle Items stammen ursprünglich von Engeln (2004) und sind auf einer vierstufigen Likert-Skala von *stimmt gar nicht* (1) bis *stimmt völlig* (4) zu bewerten.

Kognitive Belastung: Die kognitive Belastung wurde mithilfe von zwei Single-Item-Skalen erfasst. Mit dem ersten Item wurde in Anlehnung an Kalyuga et al. (1999) die wahrgenommene Aufgabenschwierigkeit (Item: Wie leicht oder schwer waren die Aufgaben insgesamt zu verstehen?) erhoben, das zweite Item erhob die investierte Denkanstrengung nach Paas (1992) (Item: Beim Bearbeiten und Verstehen der Aufgaben war meine geistige Denk-Anstrengung insgesamt…). In vielen Untersuchungen wurde statt der ursprünglich neunstufigen Skala eine reduzierte siebenstufige Skala genutzt. Auch in dieser Studie wurde eine übersetzte siebenstufige Version der Items eingesetzt (z. B. Schmeck et al., 2015). Die Skala reichte für die wahrgenommene Aufgabenschwierigkeit von (1) *sehr leicht* bis (7) *sehr schwer* und für die investierte Denkanstrengung von (1) *sehr gering* bis (7) *sehr hoch*.

8.2.3 Überprüfung der Merkmalszugehörigkeit der Kontextaufgaben

Die Merkmalszugehörigkeit der entwickelten Kontextaufgaben wurde mithilfe eines semantischen Differenzials überprüft. Hierbei wurde den Lernenden eine Reihe von bipolaren Adjektivpaaren vorgelegt, mithilfe derer die gewählten Kontextaufgaben auf einer siebenstufigen Ratingskala beurteilt wurden. Zentral für die Beurteilung war die eingeschätzte Affinität des Kontextes zu den Adjektiven. Da die Kontextaufgaben im Hinblick auf die gegensätzlichen Merkmalspaare *Außerfachlich*/*Innerfachlich* sowie *Alltäglich*/*Besonders* variiert wurden, mussten Adjektive identifiziert werden, die jedes Kontextmerkmal möglichst exakt beschreiben (Tab. 16).

Zur Unterscheidung des Merkmalspaars *Alltäglich*/*Besonders* wurden die Schlüsselwortpaare von van Vorst (2013) übernommen, die durch eine Expertenbefragung validiert wurden. Für die Einschätzung des Merkmalspaars *Außerfachlich* und *Innerfachlich* mussten dagegen eigene Schlüsselwortpaare konzipiert werden. Eine Übersicht findet sich ebenfalls in Tabelle 16.

Tabelle 16: Überblick über die ausgewählten Schlüsselwortpaare für die Merkmale Alltäglich/Besonders und Außerfachlich/Innerfachlich

Alltäglich – Besonders		Außerfachlich – Innerfachlich	
Alltäglich	Besonders	Außerfachlich	Innerfachlich
üblich	ungewöhnlich	lebensweltlich	chemisch
vertraut	fremd	außerfachlich	innerfachlich
normal	außergewöhnlich	allgemein	fachlich
gängig	unüblich	nicht aus dem Chemieunterricht	aus dem Chemieunterricht
häufig	selten		

8.3 Datenerhebung

Die Datenerhebung wurde zwischen Oktober 2020 und April 2021 im regulären Chemieunterricht durchgeführt. Zu dieser Zeit war die Studiendurchführung im Präsenzunterricht aufgrund der pandemiebedingten Schulschließungen nur eingeschränkt möglich. Deshalb war es notwendig, einen Teil der Daten im Onlineunterricht zu erheben.

Wie bereits beschrieben, wurde das Lernmaterial als digitales Lernmaterial entwickelt, welches den Schülerinnen und Schülern als PDF-Datei zur Verfügung gestellt wurde. In dieser PDF-Datei waren außerdem Verknüpfungen zu den Testinstrumenten eingefügt. Zu diesem Zweck wurden die Testinstrumente in die Online-Umfrage-Applikation *LimeSurvey* integriert (LimeSurvey Project Team, 2012). Der Ablauf der Datenerhebung ist in Abbildung 6 dargestellt.

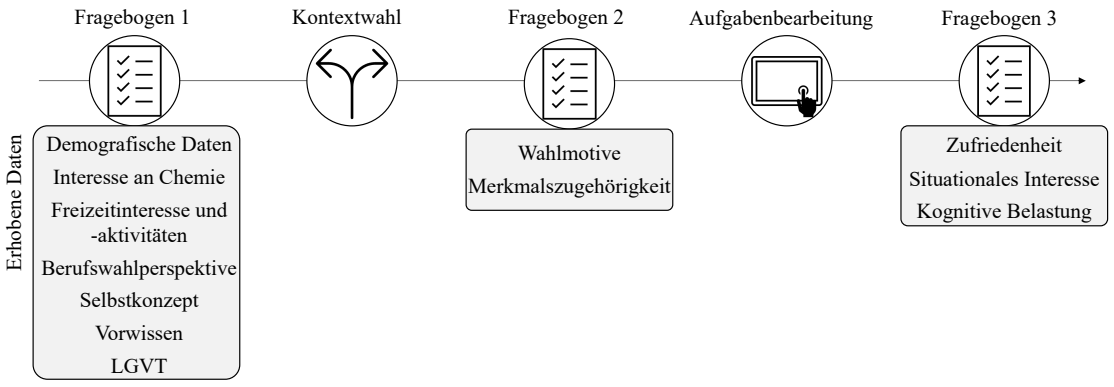


Abbildung 6: Ablauf der Datenerhebung (Teilstudie I)

Nach einer kurzen Instruktion durch die testleitende Person, konnten die Lernenden die PDF-Datei öffnen und einen individuellen Identifikationscode formulieren, der eine eindeutige Zuordnung der Testinstrumente und Lernmaterialien ermöglichte. Danach wurden die Lernenden über einen entsprechenden Hyperlink zum ersten Fragebogen geleitet, mit dem die für die Kontextwahl bedeutsamen kognitiven und affektiv-motivationalen Personenmerkmale erhoben wurden. Das Vorwissen in Chemie wurde im Fragebogen nach den affektiv-motivationalen Merkmalen erhoben, um die Bewertung der affektiven Merkmale im Vorfeld nicht zu beeinflussen. Die Instruktion und der erste Fragebogen nahmen eine Unterrichtsstunde in Anspruch (ca. 45 Minuten).

In der zweiten Schulstunde konnten die Lernenden eine Kontextaufgabe auswählen (ca. 5 Minuten) und den zweiten Fragebogen bearbeiten (ca. 5 Minuten). Mithilfe dieses Fragebogens wurden die Wahlmotive erfasst und die Merkmalszugehörigkeit der gewählten Kontextaufgabe wurde von den Lernenden eingeschätzt. Danach hatten die Schülerinnen und Schüler ungefähr 25 Minuten zur Verfügung, um die gewählte Kontextaufgabe zu bearbeiten. Für den letzten Fragebogen zur Beurteilung der Zufriedenheit standen etwa 10 Minuten zur Verfügung. Die Befragung wurde stets durch dieselbe testleitende Person begleitet, um den Einfluss durch die Testleitung zu kontrollieren. Insgesamt nahm die Datenerhebung damit ungefähr 90 Minuten in Anspruch, wobei zwei 45-minütige Schulstunden für die Befragung genutzt wurden. Diese haben in Abhängigkeit vom Stundenplan entweder unmittelbar hintereinander oder mit wenigen Tagen Abstand stattgefunden. Bei einem 60-minütigen Stundenmodell hat die Lehrkraft die übrigen 15 Minuten der Schulstunde gestaltet, um die Bearbeitungszeit für alle Lernenden zu kontrollieren. Trotz der dargelegten Standardisierungsversuche hat sich die Datenerhebung in Abhängigkeit von der Befragungsform in einigen wenigen Punkten unterschieden, die im Folgenden dargelegt werden.

Im Onlineunterricht wurde die Datenerhebung innerhalb einer Videokonferenz durchgeführt, die zu diesem Zeitpunkt dem gängigen Unterrichtsformat entsprach. Die Lernenden haben die Testinstrumente in LimeSurvey an ihren eigenen Endgeräten bearbeitet. Ebenso wurde den Schülerinnen und Schülern das Lernmaterial als PDF-Datei zur Verfügung gestellt, welches ebenfalls auf den Schüler-Endgeräten bearbeitet wurde. Hier konnten die Lernenden die Antworten in das PDF-Dokument schreiben und beispielsweise wichtige Stellen markieren.

Die Datenerhebung im Präsenzunterricht wurde im gewohnten Unterrichtsraum durchgeführt. Um die Datenerhebung vergleichbar zu gestalten, haben die Schülerinnen und Schüler das Lernmaterial und die Testinstrumente mit zur Verfügung gestellten Tabletcomputern und einem Tablet-Stift bearbeitet. Hierfür wurden dieselbe LimeSurvey-Umfrage und PDF-Datei genutzt.

8.4 Stichprobe

Im Rahmen der ersten Teilstudie wurde eine Gelegenheitsstichprobe von sieben Schulen aus der Metropolregion Rhein-Ruhr in Nordrhein-Westfalen untersucht. Dabei wurden vier Gymnasien und drei Gesamtschulen in die Untersuchung einbezogen. Insgesamt nahmen 26 verschiedene Klassen bzw. Kurse des dritten Lernjahres an der Studie teil. Schlussendlich konnte eine Stichprobe von $N = 495$ Lernenden untersucht werden.

Der Anteil der fehlenden Werte beträgt für den ersten Fragebogen 5.66 %. Für den zweiten Fragebogen liegt der Anteil fehlender Werte zwischen 17.37 % und 18.79 %. Im dritten Fragebogen sind zwischen 20.81 % und 21.82 % der Daten fehlend. Zum Umgang mit fehlenden Werten (*Missings*) stehen in der empirisch-psychologischen Forschung verschiedene Methoden zur Verfügung, deren Angemessenheit von den Annahmen über den ursächlichen Ausfallprozess abhängig sind (Lüdtke et al., 2007). Missings sind completely at random (MCAR), wenn das Auftreten der fehlenden Werte weder von der Ausprägung der Variablen selbst noch von der Ausprägung dritter Variablen im Datensatz abhängt. Besteht ein Zusammenhang zwischen dem Fehlen eines Wertes und der Ausprägung einer anderen Variable, aber nicht mit der Ausprägung der Variable selbst, so wird dies als missing at random (MAR) bezeichnet. Fehlende Werte sind dagegen missing not at random (MNAR), wenn das Fehlen des Wertes von der Ausprägung der Variable selbst abhängt (Rubin, 1976).

Der größte Teil der fehlenden Werte in dieser Studie resultiert aus dem Fehlen der Lernenden an einer der beiden Schulstunden. Zum Zeitpunkt der Datenerhebung kam es aufgrund der Covid-19-Pandemie zu erhöhten krankheitsbedingten Ausfällen. Folglich wird angenommen, dass es sich bei den fehlenden Werten um MCAR handelt. Lüdtke et al. (2007) empfehlen den fallweisen Ausschluss gegenüber anderen traditionellen Verfahren zum Umgang mit fehlenden Werten in der pädagogisch-psychologischen Forschung, da der fallweise Ausschluss unter der Annahme von MCAR nicht zu verzerrten Parameterschätzungen führt. Allerdings geht der fallweise Ausschluss mit einem Verlust an statistischer Teststärke einher. Fehlende Werte werden in der vorliegenden Arbeit fallweise ausgeschlossen, wenn davon ausgegangen werden kann, dass die fehlenden Werte unter der Annahme MCAR entstanden sind.

Nach dem listenweisen Fallausschluss liegen $N = 347$ vollständige Datensätze vor. Davon wurden ungefähr 25.07 % im Onlineunterricht befragt. Das Alter der Lernenden liegt zwischen 13 und 17 Jahren ($M = 14.76$, $SD = 0.89$). Außerdem sind 49.86 % der Lernenden weiblich und 46.11 % männlich, 4.03 % der Lernenden gaben kein Geschlecht an. 62.54 % der Lernenden haben ein Gymnasium besucht. Die übrigen Lernenden haben eine Gesamtschule besucht.

8.5 Ergebnisse und Diskussion

Die Datenanalyse wurde mithilfe der freien Programmiersprache R (Version 4.2.3) in der Entwicklungsumgebung RStudio durchgeführt (R Core Team, 2022). Für die Analyse wurden zahlreiche Pakete genutzt, welche die Funktionen von R erheblich erweitern. Die genutzten Pakete sind in Tabelle 17 aufgeführt.

Tabelle 17: Übersicht über die verwendeten R-Pakete zur Auswertung der ersten Teilstudie

Paket	Funktion	Version	Autor-/innen
tidyverse	Datenaufbereitung	1.3.2	Wickham et al. (2019)
TAM	IRT-Analysen	4.0-16	Robitzsch et al. (2022)
lavaan	CFA & Pfadmodellierung	0.6-12	Rosseel (2012)
semTools	Reliabilitätsbestimmung	0.5-6	Jorgenson et al. (2022)
fpc	Clusterstabilität	2.2-10	Hennig (2023)
npmv	MANOVA (non-parametrisch)	2.4.0	Burchett et al. (2017)

8.5.1 Analyse der Messinstrumente

Zur Erfassung der kognitiven und affektiv-motivationalen Personenvariablen und der Prozessvariablen wurden eine Reihe von psychometrischen Messinstrumenten eingesetzt, die insbesondere im Hinblick auf ihre faktorielle Validität und Reliabilität geprüft werden mussten. Dies erscheint notwendig, da bestehende Messinstrumente für diese Studie angepasst oder neu entwickelt wurden. Der Leistungstest zur Erfassung des Vorwissens in Chemie wurde mithilfe von einem eindimensionalen Rasch-Modell untersucht. Die übrigen Messinstrumente wurden mithilfe einer CFA hinsichtlich ihrer Qualität geprüft. Zur Schätzung der Modellparameter wurde die robuste Maximum-Likelihood-Methode genutzt, die auch bei fehlender Normalverteilung effiziente Schätzungen der Modellparameter liefert (Gäde et al., 2020). Fehlende Werte wurden bei Beurteilung der Messqualität nicht ausgeschlossen. Stattdessen wurden die Modelle mithilfe des Full-Information-Maximum-Likelihood-Ansatzes geschätzt, sodass alle Informationen in der Berechnung berücksichtigt wurden (Weiber & Mühlhaus, 2014). Der LGVT kann aufgrund der Testkonzeption nicht durch klassische Methoden der Reliabilitätsschätzung untersucht werden. Allerdings konnten im Rahmen der Testkonstruktion zufriedenstellende Werte für die Retest- und Paralleltest-Reliabilität von W. Schneider et al. (2017) ausgemacht werden. Da es sich bei den beiden Items zur Erfassung der kognitiven Belastung um Single-Item-Skalen handelt, können auch diese nicht hinsicht-

lich der Reliabilität untersucht werden. Beide Skalen sind allerdings erprobt und sehr verbreitet zur Messung der kognitiven Belastung (z. B. Schmeck et al., 2015). Ferner wurde das Instrument zur Erfassung der Berufswahlperspektive nicht weiter untersucht, da es sich um keine psychometrische Skala im engeren Sinne handelt (Döring & Bortz, 2016).

Vorwissen in Chemie: Der Leistungstest zur Erfassung des Vorwissens in Chemie wurde mithilfe eines eindimensionalen Rasch-Modells untersucht. Die Kennwerte des berechneten Rasch-Modells sind in Tabelle 18 dargestellt.

Tabelle 18: Kennwerte des Rasch-Modells für den Vorwissenstest (Teilstudie I)

Modellmerkmal	Kennwert
$N_{Personen}$	465
N_{Items}	57
Parameterzahl	58
Deviance	11699.81
Varianz	0.44
WLE-Reliabilität	.62
Itemparameter	-1.71 – 1.46
Personenparameter	-2.66 – 2.71
wMNSQ	0.86 – 1.14
t-Werte (wMNSQ)	-2.26 – 1.63

Die Daten zeigen eine ausreichende Passung zum eindimensionalen Rasch-Modell. Alle wMNSQ-Werte liegen in dem geforderten Bereich ($0.80 \leq wMNSQ \leq 1.20$). Entsprechend der Vorgabe von Bond et al. (2021) sind die korrespondierenden *t*-Werte nur zu betrachten, wenn die wMNSQ-Werte außerhalb des vorgesehenen Intervalls liegen. Demzufolge wurden alle Testitems im Datensatz belassen. Die WLE-Reliabilität liegt mit .62 unter der allgemein verbreiteten unteren Grenze von .70. Insgesamt ist die Varianz der Personenverteilung auch als zu gering einzustufen. Als erstrebenswert werden hier Werte größer als 1 angesehen (Neumann, 2014). Ursächlich für die geringe Varianz könnte die geringe Passung der Aufgaben auf die Personen sein. Zur Untersuchung der Passung von Aufgabenschwierigkeit und Personenfähigkeiten kann die Wright-Map betrachtet werden (Abb. 7).

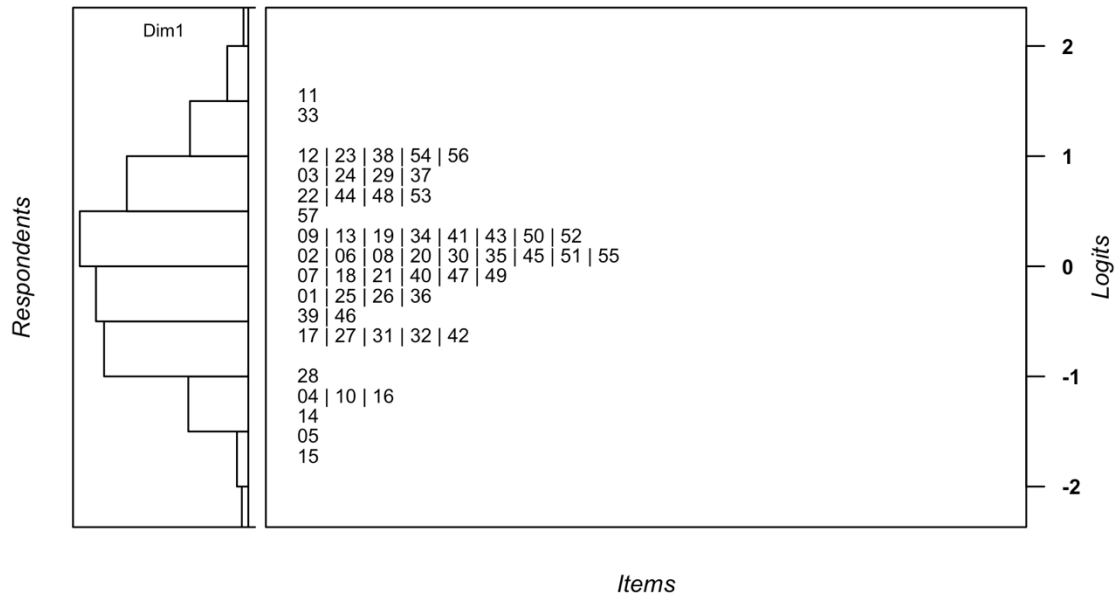


Abbildung 7: Wright-Map zum Vorwissenstest (Teilstudie I)

Die Wright-Map zeigt, dass die Testitems das Fähigkeitsspektrum der Lernenden gut abbilden. Demzufolge unterscheiden sich auch die mittleren Personen[5]- (M = 0.00, SD = 0.82) und Itemparameter (M = 0.02, SD = 0.72) nur geringfügig. Zusammenfassend ist damit unklar, wie die geringe Varianz und die geringe Reliabilität zu erklären sind.

In Bezug auf die Höhe der WLE-Reliabilität darf nicht unberücksichtigt bleiben, dass es sich bei dem Vorwissen aus den ersten beiden Lernjahren um ein sehr weites Konstrukt handelt, wodurch der Grenzwert von .70 für die Reliabilität nicht unreflektiert angewandt werden kann. Gerade unter Anbetracht der deutlich verringerten Itemzahl im Vergleich zum ursprünglichen Instrument (57 statt 348 Testitems) von Celik (2022), kann die Reliabilität von .62 als ausreichend angesehen werden (siehe Kapitel 7.1).

Interesse an Chemie und Motivation im Fach Chemie: Zur Erfassung des Interesses an Chemie und der Motivation im Chemieunterricht wurde im Rahmen dieser Arbeit ein neues Instrument zusammengestellt. Auf Basis bisher durchgeführter Arbeiten lassen sich keine eindeutigen Hypothesen über die faktorielle Struktur und die Zugehörigkeit der Items ableiten. Hier zeigten sich in explorativen Faktorenanalysen teilweise unterschiedliche Ergebnisse – insbesondere im Hinblick auf Querladungen und nicht zuordnungsbare Items (z. B. van Vorst, 2013; Habig, 2017; Hauerstein, 2019). Da zusätzlich weitere Skalen und adaptierte Items in

[5] Das Rasch-Modell wurde fallzentriert berechnet, weshalb die mittlere Personenfähigkeit auf der Logit-Skala gleich 0 ist.

das Instrument aufgenommen wurden, bestehen keine eindeutigen Hypothesen über die Anzahl der latenten Faktoren, die faktorielle Struktur und die Ladungsmatrix der Items. Insbesondere aufgrund der hohen Anzahl an Subskalen (Tabelle 10), die zudem auch theoretisch eng miteinander korrespondieren, können keine eindeutigen Hypothesen über die Faktorstruktur abgeleitet werden. Bei neu konzipierten Instrumenten bietet sich deshalb zunächst die explorative statt der konfirmatorischen Faktorenanalyse an (Eid et al., 2017). Aus diesem Grund wird zunächst eine explorative Faktorenanalyse durchgeführt, um die faktorielle Struktur des Instruments zu untersuchen. Die identifizierte Faktorstruktur wird anschließend konfirmatorisch geprüft (Gäde et al., 2020)

Auf Grundlage des Bartlett-Tests auf Sphärizität kann die Nullhypothese unkorrelierter Variablen und gleicher Varianzen verworfen werden ($\chi^2(703) = 8812.041$, $p < .001$). Jedes Item weist einen MSA > .7 auf, wobei der durchschnittliche $\overline{\text{MSA}} = .95$ ist. Demnach bestehen hohe Interkorrelationen zwischen den Items, wodurch die explorative Faktorenanalyse legitimiert ist (Eid & Schmidt, 2014; Field et al., 2012).

Sowohl der MAP-Test als auch die Parallelanalyse indizieren eine drei-faktorielle Struktur. Deshalb wurde eine Hauptachsenfaktorenanalyse mit Promax-Rotation mit drei Faktoren durchgeführt, da korrelierte Faktoren angenommen werden (Field et al., 2012). Durch die drei Faktoren können 42 % der Varianz aufgeklärt werden (Tabelle 19).

Das individuelle Interesse an Chemie lässt sich als erster latenter Faktor identifizieren. Wie in anderen Forschungsprojekten, besteht der Faktor vorrangig aus Items zum Fachinteresse, zum Sachinteresse, zur wertebezogenen Valenz und zur intrinsischen Motivation (Fechner, 2009; van Vorst, 2013; Habig et al., 2018). Ferner laden Items zur Wichtigkeit von Chemie auf diesen Faktor, was aufgrund des hohen Zusammenhangs zwischen dem interest-enjoyment value und dem attainment value in der Literatur zu erwarten war (Trautwein et al., 2012). Außerdem lassen sich die Items zum Interesse am Inhaltsfeld saure und alkalische Lösungen diesem Faktor zuordnen.

Die Subdimensionen zum themenspezifischen Interesse lassen sich empirisch nicht trennen. Der zweite Faktor wird aus Items zum Interesse an Kontexten aus dem Themenfeld Körper und Krankheiten sowie aus Items zum Interesse an alltäglichen und besonderen Kontexten gebildet. Dieser Faktor lässt sich als themenspezifisches Interesse bezeichnen.

Tabelle 19: Ergebnis der Hauptachsenfaktorenanalyse für das individuelle Interesse an Chemie

		Faktoren		
Skala	Item	1	2	3
Individuelles Interesse an Chemie	*fin2*	.84		
	fin3	.72		
	fin4	.61		
	fin5	.80		
	sin1	.69		
	sin3	-.68		
	sin4	.86		
	wert2	.69		
	wert3	.40		
	imp1	.68		
	imp2	-.69		
	inm1	.65		
	inm2	.77		
	inm3	.68		
	inm4	.66		
	sb1	.71		
	sb2	.54		
	sb3	.72		
Themenspezifisches Interesse	*kk1*		.70	
	kk2		.74	
	kk3		.77	
	alk3		.64	
	bek3		.48	
Extrinsische Motivation in Chemie	*exm1*			.50
	exm2			.76
	exm4			.42
	imp3			.69

Anmerkung: Items mit Querladungen oder Faktorladungen < .40 sind nicht dargestellt. Negative Faktorladungen deuten auf negativ kodierte Items hin.

Der letzte latente Faktor besteht aus Items zur extrinsischen Motivation und aus einem Item zur Wichtigkeit. Nach Prüfung der Formulierung des Items zur Erfassung der Wichtigkeit, zeigt sich, dass das Item einen starken Bezug zur Leistung im Fach hat. Dadurch kann der Bezug zur extrinsischen Motivation erklärt werden.

Zur Bestätigung des identifizierten Drei-Faktor-Modells wurde eine CFA berechnet. Die CFA spricht für eine unzureichende Modellpassung ($\chi^2(321) = 896.324$, $p < .001$, CFI = .899, TLI = .889, RMSEA = .065, SRMR = .054). Die Modifikationsindizes implizieren korrelierte Messfehler zwischen den Items zur Erfassung des Interesses am Inhaltsfeld saure und alkalische Lösungen. Die Modellierung eines latenten Faktors für das Interesse am Inhaltsfeld saure und alkalische Lösungen verbessert den Modellfit signifikant ($\Delta\chi^2(3) = 138.3$, $p < .001$). Das so spezifizierte Modell zeigt eine ausreichende Passung ($\chi^2(318) = 718.397$, $p < .001$, CFI = .930, TLI = .923, RMSEA = .054, SRMR = .049). Insbesondere die Faktorladungen der Items für die Skala zum individuellen Interesse und zum inhaltsfeldspezifischen Interesse zeigen hohe standardisierte Faktorladungen. Die Interkorrelation zwischen den vier latenten Faktoren liegt zwischen $r = .34$ und $r = .77$. Abgesehen von der Skala zur Erfassung der extrinsischen Motivation weisen alle Skalen Reliabilitäten auf, die über dem geforderten Wert von > .70 liegen (Tab. 20).

Tabelle 20: Itemanzahl, Faktorladungen und Reliabilität der Skalen zum individuellen Interesse an Chemie und zur Motivation im Fach Chemie (Teilstudie I)

Skala	*n*	λ	ω	α
Individuelles Interesse an Chemie	15	.56 - .82	.94	.93
Inhaltsfeldspezifisches Interesse	3	.70 - .83	.82	.82
Themenspezifisches Interesse	5	.42 - .79	.80	.79
Extrinsische Motivation in Chemie	4	.35 - .86	.68	.68

Anmerkung: λ = standardisierte Faktorladung, ω = McDonalds Omega, α = Cronbachs Alpha. Es wurden Antworten von *N* = 467 Personen berücksichtigt.

Freizeitinteresse und -aktivitäten: Das Instrument zur Erfassung spezifischer Freizeitinteressen und -aktivitäten wurde mithilfe einer EFA überprüft, um unterschiedliche Interessensbereiche zu identifizieren. Sowohl der Bartlett-Test ($\chi^2(105) = 1103.68$, $p < .001$) als auch die MSA-Werte ($\overline{MSA} = .76$) zeigen, dass die Durchführung der EFA legitim ist (Bühner, 2021). Der MAP-Test verweist auf eine einfaktorielle Struktur, wohingegen die Parallelanalyse eine vierfaktorielle Struktur anzeigt. Da eine einfaktorielle Struktur nur wenig zur Erklärung der Itemantworten beiträgt, wurde im Folgenden eine Hauptachsenfaktoranalyse mit Promax-Rotation und vier Faktoren durchgeführt. Die Ergebnisse sind in Tabelle 21 dargestellt.

Tabelle 21: Ergebnis der Hauptachsenfaktorenanalyse für das Instrument zur Erfassung des Freizeitinteresse und -aktivitäten

		Faktor			
Skala	Item	1	2	3	4
Interesse an Medien	*frei1*	.46			
	frei2	.45			
	frei3	.74			
	frei4	.57			
	frei15	.53			
Nicht extrahierbar	*frei8*		.72		
	frei10		.71		
Nicht extrahierbar	*frei7*			-.48	
	frei12			.63	
Interesse an kreativen Aktivitäten	*frei5*				.70
	frei6				.44
	frei13				.53

Anmerkung: Items mit Querladungen oder Faktorladungen < .40 sind nicht dargestellt. Negative Faktorladungen deuten auf negativ kodierte Items hin.

Es ist auffällig, dass der zweite und dritte Faktor lediglich aus zwei Items gebildet werden, so dass sie nicht extrahierbar sind, da eine psychometrische Skala mindestens drei Items umfassen sollte (Döring & Bortz, 2016). Der erste extrahierbare Faktor wird aus Items gebildet, die das Interesse an Freizeitaktivitäten erfasst, die im Zusammenhang mit Medien stehen (Musik, Film, soziale Netzwerke). Der zweite Faktor besteht aus drei Items, die sich dem Interesse an kreativen Aktivitäten und dem Lesen zuordnen lassen.

Mithilfe einer CFA wurde ein Zwei-Faktor-Modell spezifiziert und empirisch geprüft. Die Ergebnisse verweisen auf keinen hinreichenden Modellfit ($\chi^2(19) = 59.44$, $p < .001$, CFI = .908, TLI = .865, RMSEA = .070, SRMR = .058). Außerdem liegt die Reliabilität für die Skala zur Erfassung des Interesses an kreativen Aktivitäten in einem inakzeptablen Bereich ($\omega = .59$, $\alpha = .55$). Von der weiteren Verwendung des Instruments wurde deshalb abgesehen.

Chemiebezogenes Selbstkonzept: Zur Untersuchung der faktoriellen Struktur der Skalen zum Selbstkonzept wurde eine CFA durchgeführt. Hierbei wurde zunächst ein General-Faktor-Modell berechnet, welches sich in der Arbeit von Habig (2017) auf Grundlage einer EFA als suffizient erwiesen hat und in dem alle Items auf einen einzigen latenten Faktor laden. Zudem wurde ein theoriebasiertes Zwei-Faktor-Modell berechnet, in dem das chemiebezogene Selbstkonzept und die Selbstwirksamkeitserwartung eigenständige latente Variablen bilden.

Sowohl das eindimensionale Modell ($\chi^2(44) = 235.46$, $p < .001$, CFI = .938, TLI = .923, RMSEA = .097, SRMR = .042) als auch das zweidimensionale Modell (χ^2 (43) = 141.14, $p < .001$, CFI = .968, TLI = .959, RMSEA = .070, SRMR = .033) zeigen einen akzeptablen Modellfit. Ein Likelihood-Ratio-Test zeigt eine bessere Passung für das zweite Modell ($\Delta\chi^2(1) = 94.32$, $p < .001$). Demnach wird das Zwei-Faktor-Modell beibehalten. Alle Items scheinen stark mit den latenten Faktoren zusammenzuhängen, wie die hohen Faktorladungen für beide Skalen zeigen. Die Korrelation zwischen beiden latenten Faktoren liegt dabei in einem hohen Bereich ($r = .86$). Sowohl Cronbachs α als auch McDonalds ω deuten auf exzellente Reliabilität hin (Tab. 22).

Tabelle 22: Itemanzahl, Faktorladungen und Reliabilitäten der Skalen zum Selbstkonzept und zur Selbstwirksamkeitserwartung (Teilstudie I)

Skala	*n*	λ	ω	α
Selbstkonzept	8	.61 - .86	.91	.91
Selbstwirksamkeitserwartung	3	.72 - .84	.83	.83

Anmerkung: λ = standardisierte Faktorladung, ω = McDonalds Omega, α = Cronbachs Alpha. Es wurden Antworten von $N = 467$ Personen berücksichtigt.

Wahlmotive: Das Instrument zur Erfassung der Wahlmotive wurde bereits in einer explorativen Faktorenanalyse von van Vorst und Aydogmus (2021) geprüft. Die dort identifizierte Struktur wird nachfolgend konfirmatorisch überprüft. Es zeigt sich eine nicht ausreichende Passung ($\chi^2(129) = 453.02$, $p < .001$, CFI = .839, TLI = .809, RMSEA = .086, SRMR = .090). Die Betrachtung der Modifikationsindizes zeigt, dass mehrere Fehlerkovarianzen zwischen den Indikatoren vorliegen. In diesem Fall bietet sich die Durchführung einer EFA an, um die faktorielle Struktur des Konstruktes erneut zu prüfen (Gäde et al., 2020). Sowohl der Bartlett-Test ($\chi^2(171) = 2606.21$, $p < .001$) als auch die MSA-Werte ($\overline{MSA} = .86$) zeigen, dass die Durchführung der EFA legitim ist (Bühner, 2021). Die Parallelanalyse und der MAP-Test verweisen auf drei Faktoren statt der theoretisch implizierten vier Faktoren. Die Hauptachsenfaktorenanalyse mit Promax-Rotation und drei Faktoren klärt 45 % der Varianz auf. In Tabelle 23 sind die Ergebnisse dargestellt.

Tabelle 23: Ergebnis der Hauptachsenfaktorenanalyse für die Skalen zur Erfassung der Wahlmotive

		Faktoren		
Skala	Item	1	2	3
Überraschungsmoment	*uber1*	.86		
	uber2	.84		
	uber3	.83		
	uber4	.78		
	uber5	.60		
Interesse & Neugier	*inneu1*		.69	
	inneu2		.91	
	inneu3		.73	
	inneu4		.78	
	inneu5		.43	
Persönlicher Bezug	*per1*			.70
	per3			.75
	per4			-.45
	biub3			.59

Anmerkung: Items mit Querladungen oder Faktorladungen < .40 sind nicht dargestellt. Negative Faktorladungen deuten auf negativ kodierte Items hin.

Wie angenommen, bilden die Items zum Wahlmotiv *Überraschungsmoment* und zum Wahlmotiv *Interesse & Neugier* einen eigenständigen Faktor. Der Faktor zum Wahlmotiv *Persönlicher Bezug* wird neben den drei Items zur Erfassung des persönlichen Bezugs auch von einem Item zum Wahlmotiv *Bild und Überschrift* gebildet. Das Item *biub3* ist möglicherweise aufgrund der Itemformulierung (…weil ich mit der Überschrift des Arbeitsblattes Positives verbinde) ein Indikator für das Wahlmotiv persönlicher Bezug.

Die CFA zeigt, dass die gefundene Struktur nicht auf die Daten passt ($\chi^2(74) = 307.83$, $p < .001$, CFI = .863, TLI = .832, RMSEA = .096, SRMR = .098). Die Modifikationsindizes deuten Doppelladungen der Items *uber5* und *inneu5* an. Daher werden die Items von der weiteren Analyse ausgeschlossen. Es deuten sich daneben Fehlerkovarianzen zwischen den Items *inneu2* und *inneu4* sowie *uber1* und *uber3* an. Bei genauerer Prüfung zeigen sich sehr ähnli-

che Formulierungen der Items, die für Redundanz sprechen. Die Items *inneu2* und *uber3* wurden daraufhin aufgrund der geringeren Faktorladung ausgeschlossen. Das so spezifizierte Modell zeigt einen ausreichenden Fit ($\chi^2(32) = 91.34$, $p < .001$, CFI = .935, TLI = .908, RMSEA = .068, SRMR = .080). Die Korrelation zwischen den latenten Faktoren *Überraschungsmoment* und *Persönlicher Bezug* deutet nur auf einen geringen Zusammenhang hin, $r = -.07$. Die Korrelationen zum Wahlmotiv *Interesse und Neugier* sind höher ($r_{Per} = .26$ und $r_{Über} = .48$). Insgesamt wird damit deutlich, dass die Lernenden eine kontextualisierte Aufgabe entweder aufgrund des Überraschungsmoments oder aufgrund des persönlichen Bezugs wählen. Beide Wahlmotive scheinen in einem gewissen Zusammenhang mit Interesse und Neugier zu stehen. In Tabelle 24 sind die Faktorladungen sowie Cronbachs α und McDonalds ω als Maß für die Reliabilität angegeben.

Tabelle 24: Itemanzahl, Faktorladungen und Reliabilitäten der Skalen zur Erfassung der Wahlmotive (Teilstudie I)

Skala	*n*	λ	ω	α
Interesse und Neugier	3	.56 - .74	.74	.73
Überraschungsmoment	3	.74 - .77	.79	.79
Persönlicher Bezug	4	.36 - .84	.69	.68

Anmerkung: λ = standardisierte Faktorladung, ω = McDonalds Omega, α = Cronbachs Alpha. Es wurden Antworten von *N* = 467 Personen berücksichtigt

Die Reliabilitäten liegen überwiegend in einem ausreichenden Bereich. Für die Skala zur Erfassung des Wahlmotivs persönlicher Bezug weisen einige Items eine geringe Faktorladung auf, allerdings werden diese aus Gründen der Inhaltsvalidität nicht aus dem Instrument entfernt.

Zufriedenheit und situationales Interesse: Die Zufriedenheit wird in der Literatur als übergeordnete affektive Reaktion auf zentrale Aspekte bei der Aufgabenbearbeitung beschrieben, weshalb eine gewisse konzeptionelle Nähe zum situationalen Interesse besteht. Demnach sind mehrere theoretisch plausible Modelle zur Beschreibung der Zufriedenheit und des situationalen Interesses denkbar:

- Ein General-Faktor-Modell, bei dem sowohl die Items zur Zufriedenheit als auch die Items aller Subdimensionen des situationalen Interesses einen gemeinsamen latenten Faktor bilden.

- Ein Zwei-Faktoren-Modell, bei dem die Items zur Zufriedenheit einen gemeinsamen Faktor und die Items der Subdimensionen zum situationalen Interesse einen zweiten Faktor bilden.
- Ein Vier-Faktoren-Modell, bei dem erneut die Items zur Zufriedenheit einen gemeinsamen Faktor bilden und alle Items der Subdimensionen zum situationalen Interesse (emotionale Valenz, wertbezogene Valenz und epistemische Komponente) eigenständige Faktoren bilden.

Die Fit-Statistiken der berechneten Modelle sind in Tabelle 25 dargestellt.

Tabelle 25: Modellgüte des General-Faktoren-, Zwei-Faktoren-, Vier-Faktoren und des modifizierten Vier-Faktoren-Modells des Instruments zur Erfassung des situationalen Interesses und der Zufriedenheit (Teilstudie I)

Modell	χ^2	df	CFI	TLI	SRMR	RMSEA
General-Faktor-Modell	426.41	90	.844	.818	.071	.104
Zwei-Faktoren-Modell	367.89	89	.872	.849	.066	.095
Vier-Faktoren-Modell	317.09	84	.895	.869	.064	.088
Vier-Faktoren-Modell[a]	182.94	71	.942	.925	.049	.067

Anmerkung: [a] Ausschluss des Items *epi4*. Es wurden Antworten von $N = 388$ Personen berücksichtigt.

Das General-Faktor-Modell beschreibt die Daten nur unzureichend. Die Modellierung von zwei latenten Faktoren beschreibt die Daten signifikant besser ($\Delta\chi^2(1) = 42.47$, $p < .001$), zeigt allerdings auch eine nicht ausreichende Passung. Ein Vier-Faktor-Modell zeigt einen signifikanten Anstieg der Anpassungsgüte auf Grundlage des berechneten Likelihood-Ratio-Tests an ($\Delta\chi^2(5) = 41.87$, $p < .001$), wobei die Fit-Indizes nach wie vor keine ausreichende Passung implizieren. Die Modifikationsindizes zeigen Querladungen des Items *epi4* an. Durch Ausschluss des Items *epi4* kann eine hinreichende Passung auf die Daten erzielt werden. Wie zu erwarten, bestehen hohe Interkorrelationen zwischen den vier latenten Faktoren ($.61 \leq r \leq .89$). Die so konzipierten Subskalen (Tab. 26) zeigen gute Reliabilitäten mit überwiegend hohen Faktorladungen ($\lambda > .50$).

Tabelle 26: Itemanzahl, Faktorladungen und Reliabilität der Skala zur Erfassung der Zufriedenheit und der Skalen zur Erfassung des situationalen Interesses (Teilstudie I)

Skala	*n*	λ	ω	α
Zufriedenheit	4	.46 - .81	.80	.79
Emotionale Valenz	3	.64 - .72	.71	.70
Wertbezogene Valenz	3	.55 - .70	.67	.67
Epistemische Komponente	4	.61 - .81	.80	.79

Anmerkung: λ = standardisierte Faktorladung, ω = McDonalds Omega, α = Cronbachs Alpha. Es wurden Antworten von *N* = 388 Personen berücksichtigt.

Merkmalszugehörigkeit: Zur Überprüfung der Merkmalszugehörigkeit der Kontexte wurde zunächst ein einfaktorielles Modell spezifiziert, bei dem alle Items einen gemeinsamen latenten Faktor bilden (Tab. 27). Bei dem zweifaktoriellen Modell bilden die Items zur Unterscheidung der Kontextmerkmale *Alltäglich* und *Besonders* sowie die Items zur Unterscheidung der Kontextmerkmale *Außerfachlich* und *Innerfachlich* einen gemeinsamen Faktor. Sowohl das einfaktorielle ($\chi^2(27) = 112.04$, $p < .001$, CFI = .935, TLI = .913, RMSEA = .11, SRMR = .067) als auch das zweifaktorielle Modell ($\chi^2(26) = 43.37$, $p = .018$, CFI = .987, TLI = .982, RMSEA = .051, SRMR = .033) zeigen einen ausreichenden Fit, wobei das zweite Modell die Daten besser beschreibt ($\Delta\chi^2(1) = 49.98$, $p < .001$). Sowohl die Faktorladungen als auch die Reliabilität sind hinreichend.

Tabelle 27: Itemanzahl, Faktorladungen und Reliabilität der Skala zur Überprüfung der Merkmalszugehörigkeit

Skala	*n*	λ	ω	α
Alltäglich/Besonders	5	.80 - .88	.92	.92
Außerfachlich/Innerfachlich	4	.46 - .82	.73	.73

Anmerkung: λ = standardisierte Faktorladung, ω = McDonalds Omega, α = Cronbachs Alpha. Es wurden Antworten von *N* = 403 Personen berücksichtigt.

8.5.2 Überprüfung der Merkmalszugehörigkeit der Kontextaufgaben

Um die Kontextwahl der Lernenden zu untersuchen, war es notwendig, Kontextaufgaben mit eindeutiger Merkmalszugehörigkeit zu entwickeln. Aus diesem Grund wurden die entwickelten Kontextaufgaben hinsichtlich ihrer Zugehörigkeit zu den Merkmalen *Außerfachlich/Innerfachlich* sowie *Alltäglich/Besonders* untersucht.

In Abbildung 8 ist die mittlere Einschätzung der Merkmale *Außerfachlich* und *Innerfachlich* für jede Kontextaufgabe und für die Kontextaufgaben je Kontextmerkmal dargestellt. Eine hohe Merkmalseinschätzung auf der Skala impliziert eine Zugehörigkeit zum Merkmal *Innerfachlich*. Wie angenommen, wird der innerfachliche Kontext auf der Skala höher bewertet als die besonderen und alltäglichen Kontexte. Eine berechnete ANOVA verweist auf statistisch signifikante Unterschiede, $F(2, 344) = 124.75$, $p < .001$, $\eta^2 = .42$. Alle paarweisen Vergleiche sind statistisch signifikant, wie der Bonferroni-Post-Hoc-Test zeigt ($p < .001$). Ein differenzierter Blick auf die Merkmalseinschätzung der einzelnen Kontextaufgaben zeigt, dass es signifikante Unterschiede zwischen den einzelnen Aufgaben gibt, $F(4, 342) = 66.49$, $p < .001$, $\eta^2 = .44$. Lediglich die Unterschiede zwischen den besonderen Kontexten sind statistisch nicht signifikant. Zwischen den alltäglichen Kontexten *Zähneputzen* und *Unsere Verdauung* bestehen signifikante Unterschiede in der Merkmalseinschätzung.

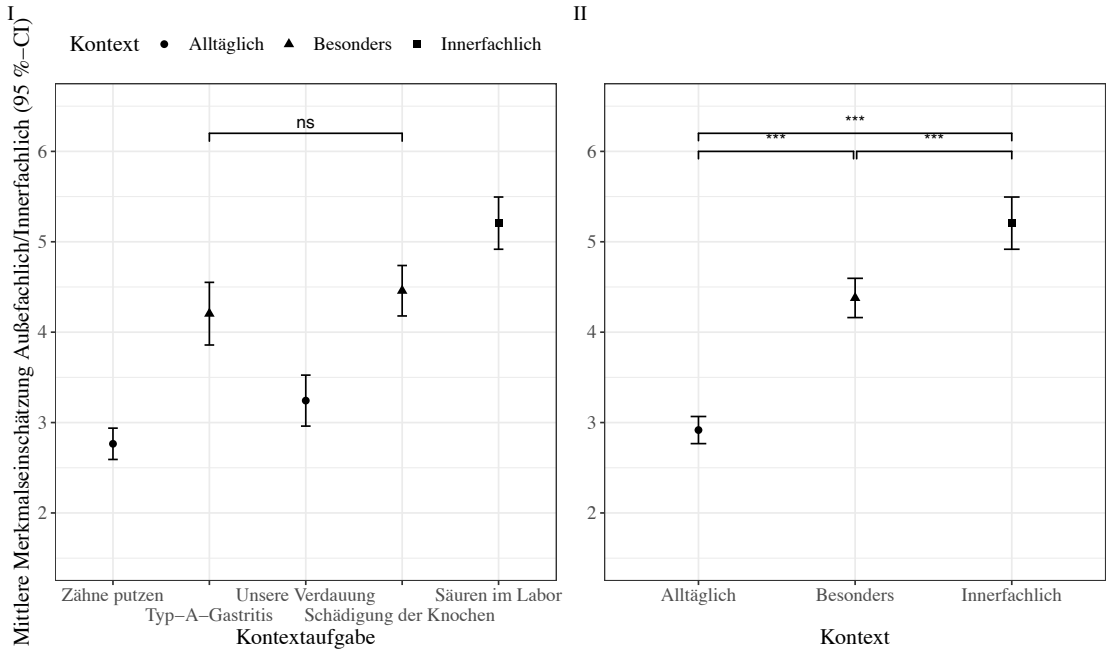


Abbildung 8: Vergleich der Merkmalszugehörigkeit Außerfachlich/Innerfachlich für jede Kontextaufgabe (I) und für die Kontextaufgaben je Kontextmerkmal (II). Aus Gründen der Übersichtlichkeit sind in Abbildung I nur nicht signifikante Unterschiede (ns) markiert

Die mittlere Einschätzung der Zugehörigkeit zu den Merkmalen *Alltäglich/Besonders* ist in Abbildung 9 dargestellt. Kontexte mit einem hohen Wert auf der Skala, sind dem Merkmal *Besonders* zuzuordnen. Die besonderen Kontexte weisen einen höheren Wert als der innerfachliche und die alltäglichen Kontexte auf, $F(2, 103.59) = 227.09$, $p < .001$, $\eta^2 = .59$. Erneut unterscheidet sich die Merkmalseinschätzung der besonderen Kontexte nicht. Es lassen sich lediglich Unterschiede zwischen den alltäglichen Kontexten identifizieren. Der niedrigere

Wert auf der Skala für den Kontext *Zähneputzen* verweist im Vergleich zum Kontext *Unsere Verdauung* auf einen stärkeren Alltagsbezug.

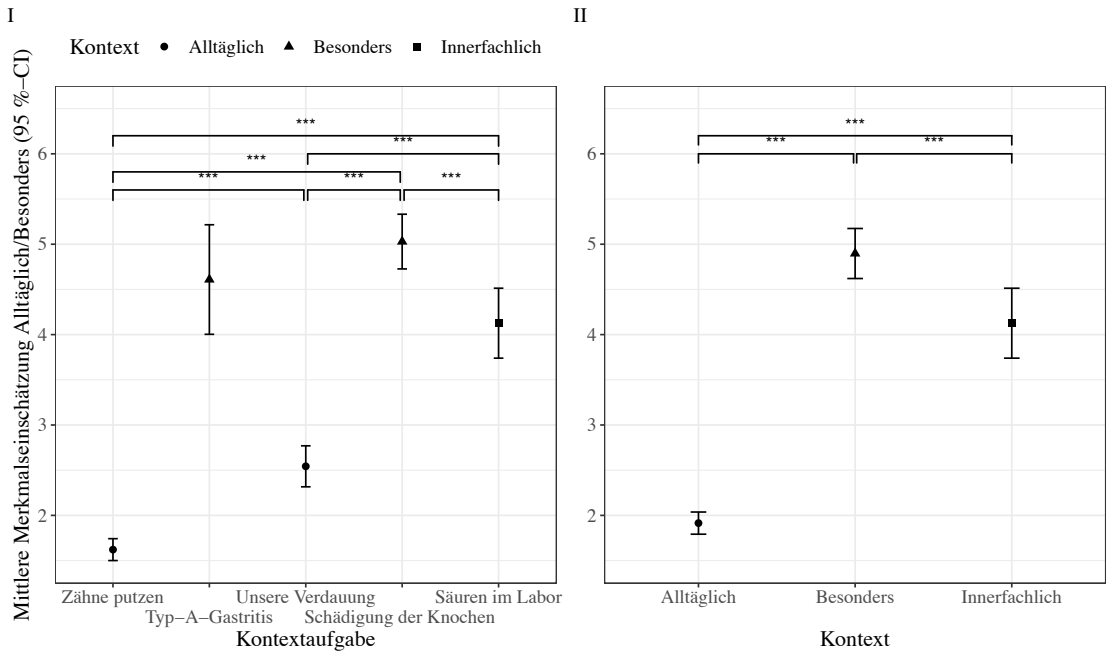


Abbildung 9: Vergleich der Merkmalszugehörigkeit Alltäglich/Besonders für jede Kontextaufgabe (I) und für die Kontextaufgaben je Kontextmerkmal (II)

Zusammenfassend entsprechen die entwickelten Kontexte damit den angenommenen Kontextmerkmalen. Der innerfachliche Kontext wird erwartungskonform auch als innerfachlich von den Lernenden eingeschätzt. Im Vergleich zu den alltäglichen Kontexten wird ein höherer innerfachlicher Bezug der besonderen Kontexte deutlich. Dies ist möglicherweise auf den geringeren Alltagsbezug der besonderen Kontexte zurückzuführen, wodurch die Lernenden einen stärkeren fachlichen Bezug wahrnehmen. Auf der Skala zur Einschätzung der Kontextmerkmale *Alltäglich* und *Besonders* wurden die alltäglichen und besonderen Kontexte tatsächlich am unteren bzw. oberen Ende der Skala eingeordnet. Der innerfachliche Kontext wird von den Lernenden auf dieser Skala eher als besonders eingeschätzt, was nicht überraschend ist. Für die meisten Schülerinnen und Schüler dürfte das chemische Laboratorium nicht Teil der alltäglichen Lebenswelt sein.

8.5.3 Auswahl relevanter Personenmerkmale

Um die Passung von Kontexten mit spezifischen Merkmalen zu unterschiedlichen Lernendengruppen zu untersuchen, mussten die Lernenden zunächst anhand relevanter Personenmerkmale gruppiert werden. Hierfür mussten vorab geeignete Personenmerkmale ausgewählt werden, die mit der Kontextwahl in Verbindung stehen. Dafür wurde der Zusammenhang zwischen den erhobenen Personenmerkmalen und dem gewählten Kontext mithilfe der in Kapitel 7.3 beschriebenen inferenzstatistischen Verfahren untersucht. Von einem multivariaten Mittelwertvergleich mithilfe einer multivariaten Varianzanalyse wurde aufgrund der hohen Variablenanzahl abgesehen (Field et al., 2012).

Geschlecht: Zunächst wurde untersucht, ob es einen Zusammenhang zwischen dem Geschlecht und dem gewählten Kontext gibt. 66.47 % der Schülerinnen haben sich für einen alltäglichen Kontext entschieden. Der Anteil der Schüler, die einen alltäglichen Kontext gewählt haben, ist etwas geringer (58.75 %). Auffällig ist, dass die Schüler (23.13 %) den innerfachlichen Kontext häufiger wählten als die Schülerinnen (10.40 %). Der besondere Kontext wurde ungefähr gleich häufig von den Geschlechtern ausgewählt (m: 18.13 %, w: 23.12 %). Der Zusammenhang zwischen dem Geschlecht und dem gewählten Kontext ist gemäß einem χ^2-Test statistisch signifikant, $\chi^2(2) = 9.93$, $p < .007$, $V = .17$). Es handelt sich allerdings um einen eher schwachen Zusammenhang. Möglicherweise ist dieser Effekt durch dritte Variablen zu erklären.

Leistung im Fach Chemie: Die Chemienote wurde in vorangegangenen Studien als relevantes Merkmal für die Kontextwahl identifiziert (Habig et al., 2018; van Vorst & Aydogmus, 2021). Wie Abbildung 10 zeigt, können auch im Rahmen dieser Analyse deskriptive Unterschiede in der Chemienote nachgewiesen werden. Die alltäglichen Kontexte wurden von Lernenden mit den schlechteren Chemienoten gewählt. Lernende, die sich für einen besonderen Kontext entschieden, hatten bessere Chemienoten. Die besten Chemienoten hatten Lernende, die den innerfachlichen Kontext zum Lernen auswählten. Gemäß dem Ergebnis einer berechneten ANOVA handelt es sich um einen statistisch signifikanten Effekt mit einer kleinen bis mittleren Effektstärke ($F(2, 344) = 9.59$, $p < .001$, $\eta^2 = .05$). Dieser Effekt kann gemäß dem Bonferroni-Test vorwiegend auf die Differenz zwischen Lernenden zurückgeführt werden, die einen alltäglichen oder den innerfachlichen Kontext gewählt haben ($p < .001$).

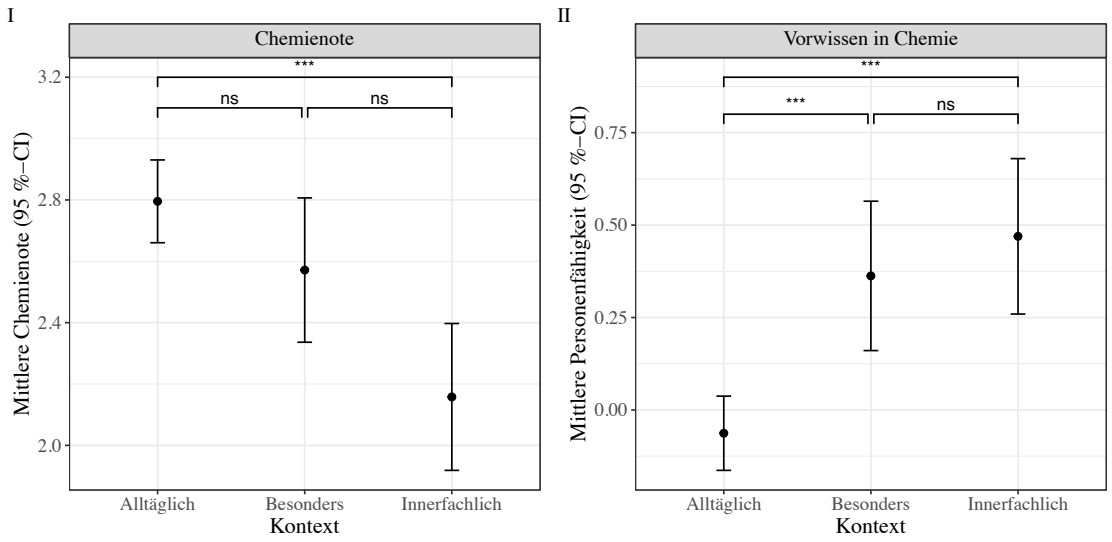


Abbildung 10: Vergleich der mittleren Chemienote (I) und der mittleren Personenfähigkeit (II) der Lernenden mit alltäglicher, besonderer und innerfachlicher Kontextwahl

Ein valideres Maß zur Erfassung des Vorwissens in Chemie ist der genutzte Fachwissenstest (Abb. 10). Beim Vergleich der Lernenden, die einen bestimmten Kontext gewählt haben, zeigt sich ein ähnlicher Trend: Erneut wählten tendenziell Lernende mit einer geringeren Leistung einen der alltäglichen Kontexte, $F(2, 344) = 15.10$, $p < .001$, $\eta^2 = .08$. Die Effektstärke fällt hierbei etwas höher aus und liegt im mittleren bis großen Bereich. Es gibt gemäß dem Bonferroni-Post-Hoc-Test keine signifikanten Unterschiede zwischen den Lernenden, die einen besonderen oder den innerfachlichen Kontext gewählt haben ($p = .72$).

Chemiebezogenes Selbstkonzept: Das Selbstkonzept und die Selbstwirksamkeitserwartung sind eng mit der Leistung und dem Vorwissen verknüpft und lassen sich in diesem Zusammenhang als selbsteingeschätzte Leistungsfähigkeit definieren. Sowohl für das Selbstkonzept ($F(2, 344) = 15.93$, $p < .001$, $\eta^2 = .08$) als auch für die Selbstwirksamkeitserwartung ($F(2, 344) = 11.49$, $p < .001$, $\eta^2 = .06$) zeigt sich ein signifikanter Effekt in Abhängigkeit vom gewählten Kontext in einer mittleren Effektstärke (Abb. 11).

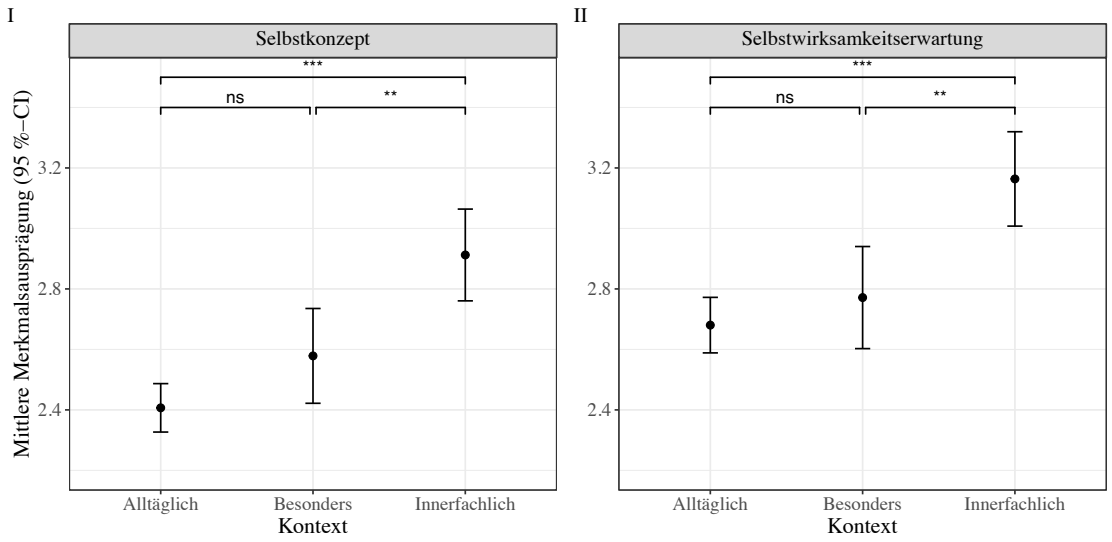


Abbildung 11: Vergleich des mittleren Selbstkonzepts (I) und der mittleren Selbstwirksamkeitserwartung (II) der Lernenden mit alltäglicher, besonderer und innerfachlicher Kontextwahl

Innerfachliche Kontexte wurden vornehmlich von Lernenden gewählt, die ein vergleichsweise hohes Selbstkonzept und eine hohe Selbstwirksamkeitserwartung aufwiesen. Das Selbstkonzept und die Selbstwirksamkeitserwartung von Lernenden sind niedriger, wenn sie einen alltäglichen oder besonderen Kontext gewählt haben. Es lassen sich keine Unterschiede zwischen Lernenden mit einer alltäglichen und besonderen Kontextwahl in Bezug auf das Selbstkonzept (p = .10) oder die Selbstwirksamkeitserwartung (p = .59) ausmachen.

Individuelles Interesse an Chemie und Motivation im Chemieunterricht: In Abbildung 12 sind die Unterschiede in den Skalen zum individuellen Interesse an Chemie und zur Motivation im Chemieunterricht dargestellt.

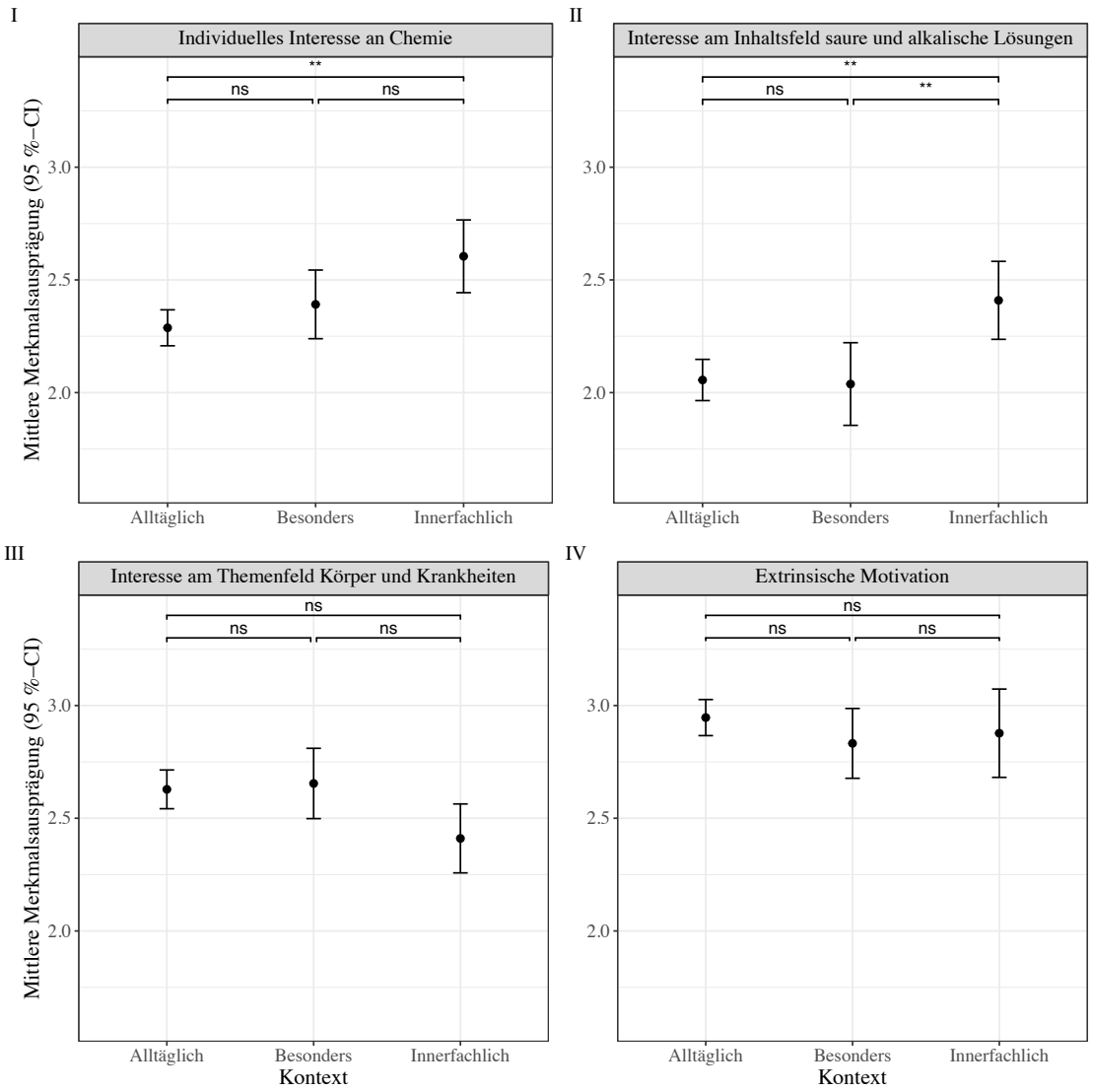


Abbildung 12: Vergleich des mittleren individuellen Interesses an Chemie (I), dem Interesse am Inhaltsfeld saure und alkalische Lösungen (II), dem Interesse am Themenfeld Körper und Krankheiten (III) und der extrinsischen Motivation (IV) der Lernenden mit alltäglicher, besonderer und innerfachlicher Kontextwahl

Es lassen sich signifikante Unterschiede im individuellen Interesse an Chemie ($F(2, 344) = 6.24$, $p = .002$, $\eta^2 = .04$) und im Interesse am Inhaltsfeld saure und alkalischen Lösungen ($F(2, 344) = 6.25$, $p = .002$, $\eta^2 = .04$) ausmachen, die als kleiner bis mittlerer Effekt zu interpretieren sind. Das Interesse an Chemie steigt von Lernenden mit einer alltäglichen Kontextwahl, über Lernende mit einer besonderen Kontextwahl, bis hin zu Lernenden, die den innerfachlichen Kontext ausgewählt haben. Hierbei ist lediglich der Unterschied zwischen den Lernenden signifikant ($p = .002$), die einen alltäglichen bzw. den innerfachlichen Kontext gewählt haben. Die Analyse des inhaltsfeldspezifischen Interesses zeigt, dass innerfachliche Kontexte von Lernenden mit einem höheren Interesse an sauren und alkalischen Lösungen

gewählt wurden. Das Interesse an sauren und alkalischen Lösungen ist geringer bei Lernenden mit einer alltäglichen oder besonderen Kontextwahl ausgeprägt. Demzufolge ist nur der Unterschied zwischen den Lernenden mit einer alltäglichen und innerfachlichen sowie besonderen und innerfachlichen Kontextwahl statistisch bedeutsam. Für das themenspezifische Interesse zeigen sich deskriptive Unterschiede, die gerade nicht signifikant sind, $F(2, 344) = 2.99$, $p = .05$, $\eta^2 = .02$. Die extrinsische Motivation ($F(2, 344) = 0.96$, $p = .38$, $\eta^2 = .00$) unterscheidet sich nicht in Abhängigkeit vom gewählten Kontext und ist demnach für die Kontextwahl weniger entscheidend.

Berufswahlperspektive: In Bezug auf die Berufsaspirationen der Lernenden zeigt sich, dass fast die Hälfte der Lernenden (45.82 %) noch keinen konkreten Berufswunsch hat. 36.89 % der Lernenden mit einem konkreten Berufswunsch wollen einen Beruf ergreifen, der sich dem Bereich *Sonstige* zuordnen lässt (Tabelle 12). Bedeutend weniger Lernende interessieren sich für einen naturwissenschaftlichen (2.88 %) oder medizinischen Beruf (14.41 %). Hierbei ist auffällig, dass der Anteil von Lernenden mit einer innerfachlichen Kontextwahl in der Gruppe mit einem naturwissenschaftlichen Berufswunsch größer ist. Der Zusammenhang zwischen der Kontextwahl und den Berufsaspirationen ist statistisch nicht signifikant, $p = .06$, $V = .14$.

Wahlmotive: Van Vorst und Aydogmus (2021) haben besonders die Wahlmotive als eine für die Kontextwahl relevante Variable herausgearbeitet. Im Rahmen dieser Datenanalyse zeigen sich signifikante Unterschiede hinsichtlich der Wahlmotive *Überraschungsmoment* ($F(2, 344) = 13.17$, $p < .001$, $\eta^2 = .07$) und *Persönlicher Bezug* ($F(2, 344) = 100.88$, $p < .001$, $\eta^2 = .37$). Abbildung 13 zeigt die deskriptiven Unterschiede in der Bewertung der Wahlmotive in Abhängigkeit vom gewählten Kontext.

Insbesondere die Lernenden mit einer besonderen Kontextwahl bewerteten das Wahlmotiv *Überraschungsmoment* wichtiger als die Lernenden mit einer alltäglichen oder innerfachlichen Kontextwahl. Der Post-Hoc-Test verweist darauf, dass sich die alltäglich und innerfachlich Wählenden nicht signifikant unterscheiden ($p = .88$). Dahingegen bewerteten Lernende, die einen alltäglichen Kontext wählten, den persönlichen Bezug höher als Lernende mit einer besonderen oder innerfachlichen Kontextwahl. Der paarweise Vergleich mit Bonferroni-Korrektur verweist auf signifikante Unterschiede zwischen allen durchgeführten Vergleichen ($p < .001$). Die Ausprägung des Wahlmotivs *Interesse und Neugier* unterscheidet sich nicht in Abhängigkeit vom gewählten Kontext, $F(2, 344) = 1.39$, $p < .001$, $\eta^2 = .00$.

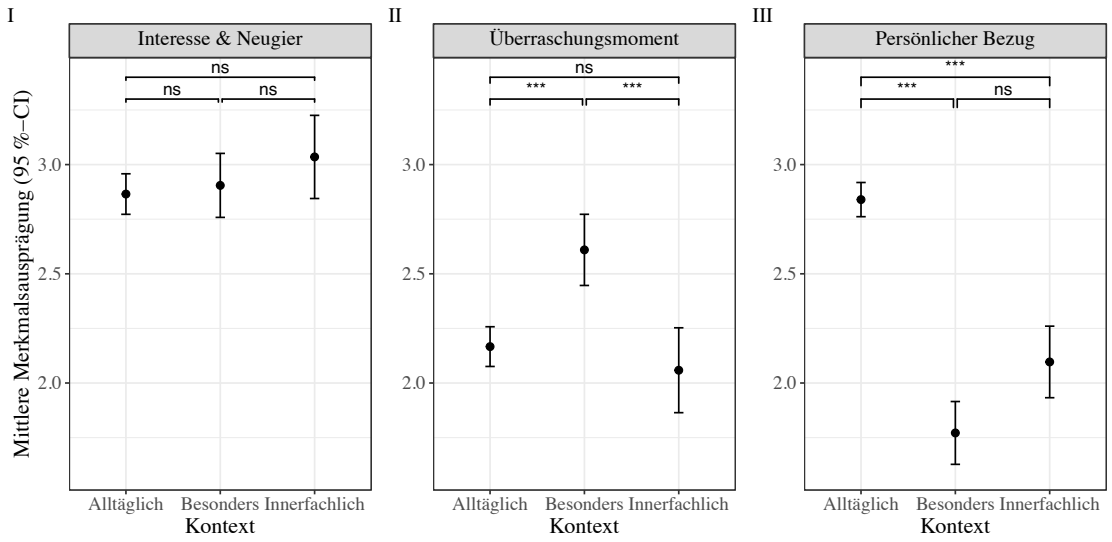


Abbildung 13: Vergleich der mittleren Ausprägung der Wahlmotive Persönlicher Bezug (I), Überraschungsmoment (II) sowie Interesse und Neugier (III) der Lernenden mit alltäglicher, besonderer und innerfachlicher Kontextwahl

Insgesamt zeigt die deskriptive und inferenzstatistische Analyse der Personenmerkmale in Abhängigkeit vom gewählten Kontextmerkmal statistisch signifikante Unterschiede für das Geschlecht, die Chemienote, das Vorwissen in Chemie, das chemiebezogene Selbstkonzept und die Selbstwirksamkeitserwartung, das individuelle Interesse an Chemie und das inhaltsfeldspezifische Interesse sowie für die Wahlmotive *Überraschungsmoment* und *Persönlicher Bezug*. Die Effektstärken liegen überwiegend im kleinen bis mittleren Bereich. Lediglich der Unterschied im Wahlmotiv *Persönlicher Bezug* weist einen großen Effekt auf. Insgesamt kann basierend auf den Ergebnissen davon ausgegangen werden, dass diese Personenmerkmale relevant in Bezug auf die Kontextwahl sind.

8.5.4 Identifizierung der Lernendengruppen

Ausgangspunkt für die Identifizierung der Lernendengruppen ist die Auswahl von relevanten Variablen für den Clusterprozess. Backhaus et al. (2021) sowie Bacher et al. (2010) empfehlen in diesem Zusammenhang nicht alle Variablen in die Clusteranalyse aufzunehmen, um eine implizite Gewichtung bestimmter Aspekte zu vermeiden.

Als Maß für die Leistung im Fach Chemie konnten statistisch signifikante Unterschiede für die letzte Chemienote und für das Vorwissen nachgewiesen werden. Da beide Variablen tendenziell die Leistung im Fach Chemie messen, wurde für die Clusteranalyse nur das Vorwissen ausgewählt, da es sich um das objektivere Maß handelt, welches durch einen Leistungstest gemessen wurde (Marsh et al., 2005). Die selbsteingeschätzte Kompetenz wird durch das

Selbstkonzept und die Selbstwirksamkeitserwartung erfasst. Für die Clusteranalyse wurde nur das Selbstkonzept berücksichtigt, da die Skala die höhere Reliabilität aufweist und eine größere Effektstärke hinsichtlich des Selbstkonzepts nachgewiesen wurde. Außerdem wurde für den Clusterprozess auf die Skala zur Erfassung des inhaltsfeldspezifischen Interesses verzichtet, da auch hier konzeptuelle Überschneidungen zum individuellen Interesse an Chemie bestehen, wie die Korrelationsanalyse zeigt (r = .68, p < .001). Das individuelle Interesse an Chemie erscheint für die Identifizierung der Lernendengruppen besonders relevant, da es auch durch bereits durchgeführte Arbeiten als wichtiges Merkmal herausgearbeitet wurde (Habig et al., 2018; van Vorst & Aydogmus, 2021). Demzufolge wurde lediglich das individuelle Interesse an Chemie für die Gruppierung der Lernenden berücksichtigt. Das Geschlecht blieb bei der Clusteranalyse unberücksichtigt, da der Effekt des Geschlechts möglicherweise durch andere Variablen konfundiert ist. Überdies war die unter Berücksichtigung des Geschlechts erhaltene Clusterlösung nicht zweifelsfrei interpretierbar. Das Geschlecht wurde stattdessen für die inhaltliche Validitätsprüfung der Cluster genutzt (Bacher et al., 2010).

Folglich gingen das Vorwissen, das Selbstkonzept, das individuelle Interesse an Chemie, das Wahlmotiv des Überraschungsmoments sowie das Wahlmotiv des persönlichen Bezugs in die K-Means-Clusteranalyse ein. Alle Variablen wurden vorher z-standardisiert, um die Variablen vergleichbar zu skalieren (Backhaus et al., 2020).

Zunächst wurden Clusterlösungen berechnet, die ein bis acht Cluster umfassten. Zur Bestimmung der Clusteranzahl wurde die Gesamtstreuungsquadratsumme gegen die berechnete Clusterlösung (Abb. 14.I) sowie der PRE-Koeffizient (Abb. 14.II) aufgetragen.

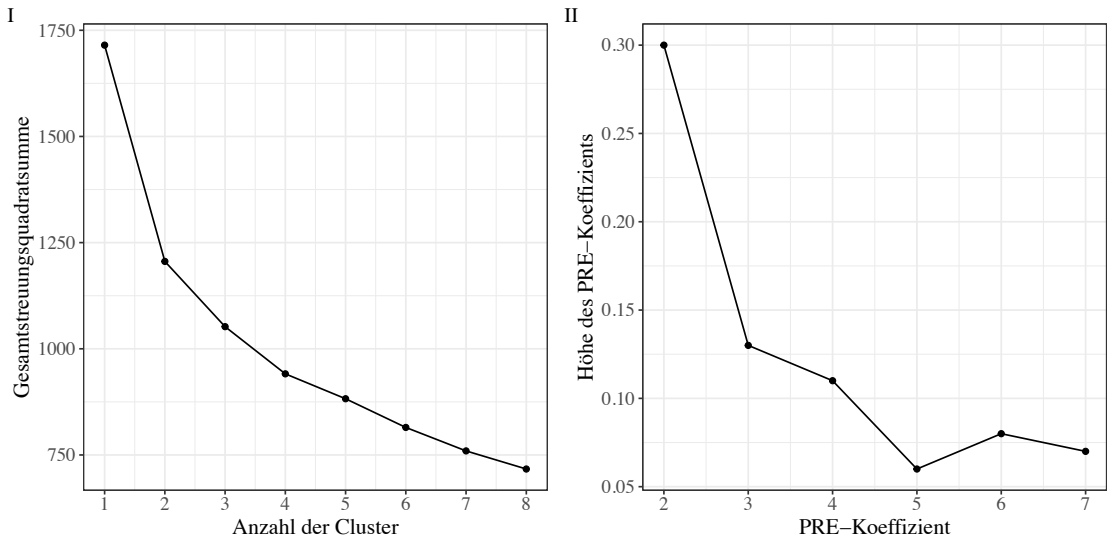


Abbildung 14: Gesamtstreuungsquadratsumme gegen die Clusteranzahl (I) und die Höhe des PRE-Koeffizients (II)

Der PRE-Koeffizient beschreibt die relative Verbesserung in der Streuungsquadratsumme innerhalb der Cluster gegenüber der vorangegangenen Lösung. So beschreibt beispielsweise der PRE_3-Koeffizient die relative Verbesserung der 3- zur 2-Clusterlösung in der Streuungsquadratsumme innerhalb der Cluster. In Abbildung 14.II ist ein deutlicher Abfall zwischen der 2- und 3-Clusterlösung sowie der 4- und 5-Clusterlösung erkennbar. Folglich führt sowohl die 3- als auch 5-Clusterlösung nur noch zu geringen Verbesserungen in der Streuungsquadratsumme in den Clustern. In der Gesamtstreuungsquadratsumme ist nach der 2- und 4-Clusterlösung eine geringere Abnahme erkennbar (Abb. 14.I).

Insgesamt deuten die Ergebnisse damit auf eine 2- oder 4-Clusterlösung hin. Zusätzlich wird die 3-Clusterlösung weiter untersucht, da sie einer theoretisch plausiblen Lösung entspricht, bei der jede Lernendengruppe eine Art von Kontext wählt. Vorab wurden außerdem 10 Ausreißer auf Grundlage der standardisierten Distanz zu den Clusterzentren entfernt, da Ausreißer insbesondere die Homogenität in den Clustern beeinflussen (Bacher et al., 2010). Die Ergebnisse der formalen Gültigkeitsprüfung sind in Tabelle 28 zusammengefasst.

Wie zu erwarten, steigt die erklärte Varianz mit zunehmender Clusteranzahl (Backhaus et al., 2020). In gleichem Maß wie die Homogenität in den Clustern steigt, sinkt die Heterogenität zwischen den Clustern mit steigender Clusteranzahl. Die Stabilitätsprüfung zeigt, dass einige Cluster der 3-Clusterlösung nicht stabil sind ($\bar{\gamma} < .75$), weshalb die 3-Clusterlösung nicht weiter betrachtet wird. Demnach erfüllen lediglich die 2- und 4-Clusterlösung die Stabilitätsanforderungen. Allerdings erklärt die 4-Clusterlösung einen höheren Varianzanteil und zeigt die höchsten Homogenitäts-Kennwerte.

Tabelle 28: Die erklärte Varianz (η^2), der mittlere Homogenitätsindex ($Homo_k$), der mittlere Heterogenitätsindex ($Hetero_k$) und der mittlere Jaccard-Koeffizient ($\bar{\gamma}$) als Maß für die Stabilität der 2-, 3- und 4-Clusterlösung (N_k).

N_k	η^2	$Homo_k$	$Hetero_k$	$\bar{\gamma}$
2	.25	.24	.97	.82 - .87
3	.36	.35	.94	.69 - .80
4	.44	.43	.89	.84 - .91

Anmerkung: Cluster sind stabil, wenn $\bar{\gamma} > .75$ (Hennig, 2007).

Zusammenfassend motivieren die Befunde der Clusteranalyse damit die inhaltliche Interpretation der 4-Clusterlösung. Die *z*-standardisierten Mittelwerte der Personenmerkmale für jedes Cluster der 4-Clusterlösung sind in Abbildung 15 dargestellt.

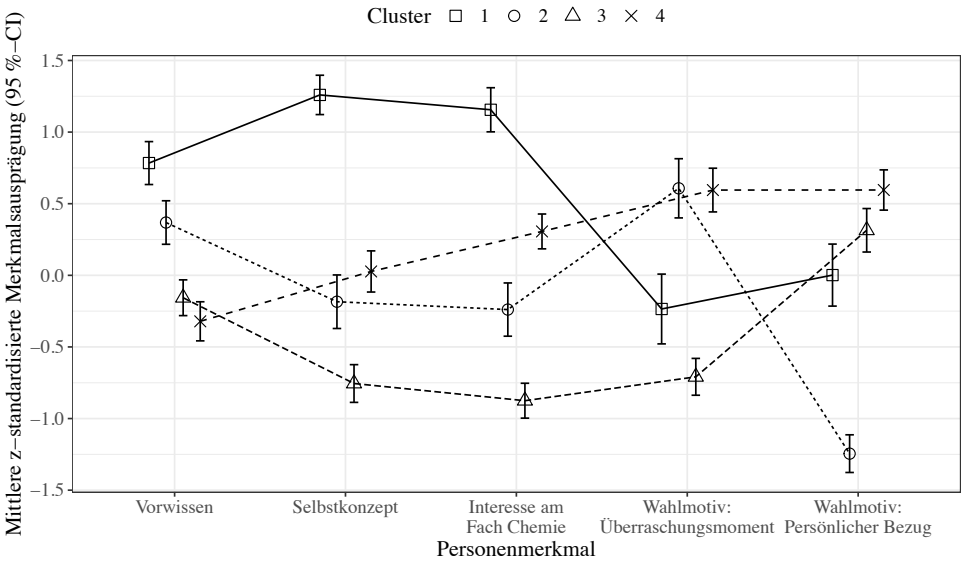


Abbildung 15: Die z-standardisierten Mittelwerte der Personenmerkmale für jedes Cluster

Im ersten Cluster befinden sich 73 Schülerinnen und Schüler. Sie zeichnen sich durch das höchste Vorwissen, chemiebezogene Selbstkonzept und individuelles Interesse am Fach Chemie aus. Die beiden Wahlmotive sind ungefähr gleich wichtig, wobei das Wahlmotiv des persönlichen Bezugs als etwas bedeutender bewertet wird. Im Vergleich zu den Wahlmotiven der Schülerinnen und Schüler aus den anderen Clustern fällt die Bewertung der Wahlmotive insgesamt niedriger aus. Dieses Ergebnis deutet darauf hin, dass die erfassten Wahlmotive eine eher untergeordnete Rolle gespielt haben. Insgesamt scheint damit unklar, aus welchem Motiv diese Gruppe einen Kontext gewählt hat.

Das zweite Cluster besteht aus Lernenden mit einem geringeren Vorwissen. Sowohl das Selbstkonzept als auch das individuelle Interesse sind geringer ausgeprägt als im ersten Cluster. Die Lernenden aus diesem Cluster bewerten das Wahlmotiv des Überraschungsmoments als besonders wichtig. Es ist das kleinste Cluster mit 69 Lernenden.

Schülerinnen und Schüler des dritten Clusters zeigen ein geringeres Vorwissen als die Schülerinnen und Schüler aus Cluster 1 und 2. Zudem zeigen die Lernenden das geringste Selbstkonzept und individuelle Interesse der gesamten Stichprobe. Der Kontext wurde in dieser Gruppe vor allem wegen des persönlichen Bezugs gewählt. Mit 108 Schülerinnen und Schülern ist dies das größte Cluster.

Das letzte Cluster (n = 87) zeigt ein eher ungewöhnliches Profil. In dieser Gruppe befinden sich die Lernenden mit dem geringsten Vorwissen. Trotz dessen ist das Selbstkonzept ausgeprägter als das der Lernenden aus dem zweiten Cluster, was dafür spricht, dass diese Lernenden sich selbst als kompetent wahrnehmen. Das Interesse am Fach Chemie ist im Vergleich zu dem zweiten und dritten Cluster eher hoch ausgeprägt. Außerdem geben die Lernenden

an, den Kontext sowohl aufgrund des persönlichen Bezugs als auch aufgrund des Überraschungsmoments zu wählen.

Zur weiteren inhaltlichen Validitätsprüfung der Cluster können die Variablen genutzt werden, die nicht in die Clusteranalyse eingegangen sind, sich allerdings als bedeutsam für die Kontextwahl herausgestellt haben oder als Kontrollvariable genutzt wurden. Die deskriptiven Statistiken dieser Variablen für jedes Cluster sind in Tabelle 29 dargestellt.

Tabelle 29: Deskriptive Statistiken zum Geschlecht, der letzten Chemienote, der Selbstwirksamkeitserwartung, dem Interesse am Themenfeld saure und alkalische Lösungen und der erreichten Punktzahl im LGVT (Leseverständnis) der vier Cluster

Variablen	Cluster			
	C_1	C_2	C_3	C_4
Geschlecht	w = 57.35 %	w = 55.88 %	w = 51.92 %	w = 46.99 %
	m = 42.65 %	m = 44.12 %	m = 48.08 %	m = 53.01 %
Chemienote	*M* = 1.81	*M* = 2.71	*M* = 3.09	*M* = 2.74
	SD = 0.76	*SD* = 0.99	*SD* = 0.87	*SD* = 1.03
Selbstwirksamkeitserwartung	*M* = 3.43	*M* = 2.69	*M* = 2.41	*M* = 2.72
	SD = 0.45	*SD* = 0.56	*SD* = 0.67	*SD* = 0.59
Inhaltsfeldspezifisches Interesse	*M* = 2.66	*M* = 1.92	*M* = 1.69	*M* = 2.30
	SD = 0.56	*SD* = 0.64	*SD* = 0.58	*SD* = 0.62
Leseverständnis	*M* = 27.16	*M* = 26.58	*M* = 21.09	*M* = 17.51
	SD = 13.5	*SD* = 12.7	*SD* = 13.8	*SD* = 12.6

In Bezug auf das Geschlecht zeigt sich, dass der Anteil der Schülerinnen von Cluster 1 zu Cluster 4 abnimmt. Der Zusammenhang zwischen dem Geschlecht und der Clusterzugehörigkeit ist statistisch allerdings nicht signifikant, $\chi^2(3) = 1.98$, $p = .58$, $V = .08$.

Hinsichtlich der letzten Chemienote kann ein signifikanter Unterschied zwischen den Clustern mit einer großen Effektstärke nachgewiesen werden, $F(3, 174.64) = 38.51$, $p < .001$, $\eta^2 = .21$). Wie in Tabelle 29 ersichtlich, erzielten die Lernenden aus dem ersten Cluster die besten Chemienoten. Die Chemienote des zweiten und vierten Clusters unterscheidet sich gemäß dem paarweise berechneten Bonferroni-Test nicht ($p = .99$). Dies ist überraschend, da das Vorwissen des zweiten Clusters – gemäß dem durchgeführten Leistungstest – höher ist

als das Vorwissen des vierten Clusters. Die Lernenden aus dem dritten Cluster wurden im Chemieunterricht am schlechtesten bewertet.

Für die Selbstwirksamkeitserwartung ergibt sich ein ähnliches Bild: Die Lernenden des ersten Clusters zeigen die höchste Selbstwirksamkeitserwartung. Eine signifikant niedrigere Selbstwirksamkeitserwartung konnte bei den Lernenden des zweiten und vierten Clusters gemessen werden, wobei der Unterschied zwischen den beiden Clustern statistisch nicht bedeutsam ist (p = .99). Die Selbstwirksamkeitserwartung der zum dritten Cluster zugehörigen Schülerinnen und Schüler liegt signifikant unter der Selbstwirksamkeitserwartung der anderen Lernenden, wie der paarweise Vergleich zeigt. Der globale Modelltest verweist auf einen großen Effekt der Clusterzugehörigkeit hinsichtlich der Selbstwirksamkeitserwartung, der zudem statistisch signifikant ist, $F(3, 333) = 45.07$, $p < .001$, $\eta^2 = .29$.

Ferner ist das inhaltsfeldspezifische Interesse bei den Lernenden des ersten Clusters signifikant höher ausgeprägt im Vergleich zu den Lernenden der anderen Cluster, $F(3, 333) = 42.76$, $p < .001$, $\eta^2 = .28$). Hierbei ist lediglich der Unterschied zwischen Cluster 2 und 3 statistisch unbedeutsam (p = .08). Die Ergebnisse ähneln dem Resultat zum individuellen Interesse an Chemie, wobei auch die Korrelationsanalyse einen starken Zusammenhang impliziert (r = .67, $p < .001$).

Vergleicht man die mittlere Punktzahl im Leseverständnistest zwischen den Clustern, so zeigen sich signifikante Unterschiede mit einer mittleren Effektstärke, $F(3,333) = 9.77$, $p < .001$, $\eta^2 = .08$). Die mittlere Punktzahl nimmt vom ersten zum vierten Cluster ab, wobei die Unterschiede zwischen dem ersten und zweiten (p = .99) sowie dritten und vierten Cluster (p = .36) statistisch unbedeutend sind.

8.5.5 Kontextwahl der Lernendengruppen

Gemäß einem berechneten χ^2-Test gibt es einen statistisch signifikanten Zusammenhang zwischen dem gewählten Kontext und der Clusterzugehörigkeit, $\chi^2(6) = 120.73$, $p < .001$, $V = .43$). Es handelt sich um einen moderaten bis starken Zusammenhang. Die relative Häufigkeit der gewählten Kontexte in den Clustern und in der Gesamtstichprobe sind in Abbildung 16 veranschaulicht.

63.80 % aller Lernenden haben sich für einen alltäglichen Kontext entschieden. Insgesamt haben 43.32 % der Lernenden den alltäglichen Kontext *Warum Zähneputzen so wichtig ist* und 20.48 % den Kontext *Unsere Verdauung* gewählt. Dieses Ergebnis deutet daraufhin, dass die meisten Lernenden am liebsten mit einem alltäglichen Kontext lernen. Broman und Kollegen (2018, 2020) kamen in ihren Untersuchungen zu vergleichbaren Ergebnissen. In beiden Untersuchungen wurden Kontexte mit einem hohen Alltagsbezug am häufigsten gewählt. Der hohe Alltagsbezug ist vermutlich für die meisten Lernenden motivierend. Möglicherweise wird der alltägliche Kontext von den Lernenden auch als zugänglicher und fachlich weniger

anspruchsvoll wahrgenommen. Sowohl die besonderen (19.58 %) als auch der innerfachliche Kontext (16.62 %) werden sehr viel seltener von den Lernenden gewählt. Bei den besonderen Kontexten war der Kontext *Schädigung der Knochen durch Übersäuerung* (13.35 %) etwas beliebter als der Kontext *Die Folgen der chronischen Gastritis Typ-A* (6.23 %).

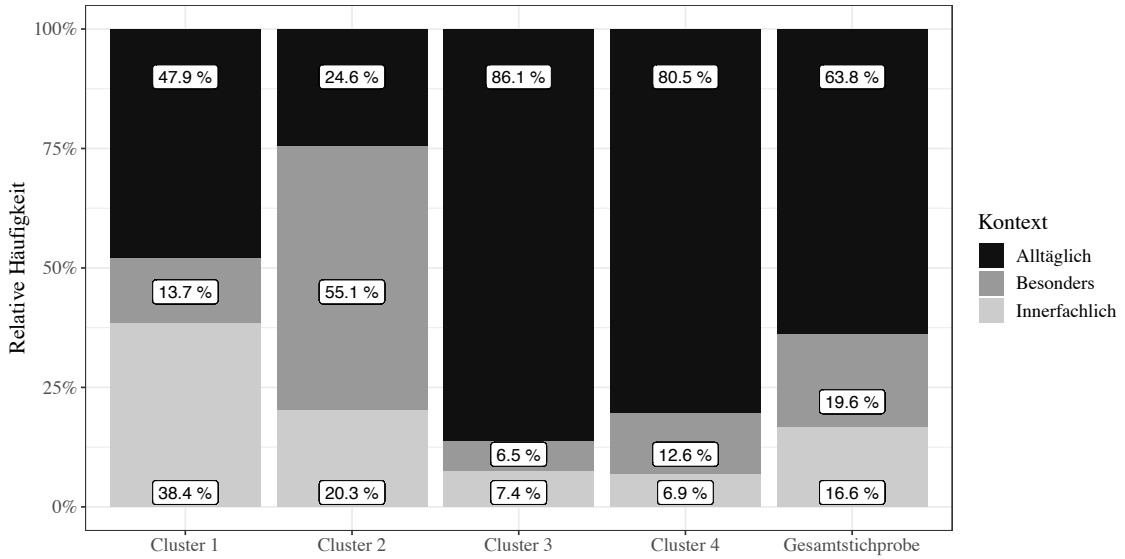


Abbildung 16: Relative Häufigkeit der gewählten Kontexte in den Clustern und der Gesamtstichprobe

Im ersten Cluster wählten weniger Lernende einen alltäglichen Kontext als in der Gesamtstichprobe (63.8 %). Dagegen wurde der innerfachliche Kontext sehr viel häufiger als in der Gesamtstichprobe gewählt. Hier entschieden sich nur 16.62 % der Lernenden für den innerfachlichen Kontext. Ein besonderer Kontext wurde in diesem Cluster von 13.70 % der Lernenden gewählt, was etwas unter dem Anteil in der Gesamtstichprobe liegt. So scheinen die Lernenden mit einem hohen Vorwissen, Selbstkonzept und individuellem Interesse an Chemie häufiger einen innerfachlichen Kontext zum Lernen auszuwählen. Möglicherweise schätzen die Lernenden den innerfachlichen Kontext als fachlich herausfordernder ein, da der fachliche Bezug höher ist. Demnach verspüren lediglich Lernende mit einem hohen Selbstkonzept eine ausreichend hohe Erfolgserwartung hinsichtlich der Aufgabenbearbeitung (z. B. Eccles & Wigfield, 2020), die wiederum zur häufigeren Auswahl anregt. Lernende mit einem hohen individuellen Interesse am Fach erleben den Kontext möglicherweise als störend im Lernprozess, da sie fachliche Inhalte vertiefen und keine für sie nebensächlichen Themen kennenlernen wollen. Beide Wahlmotive wurden in diesem Cluster als ungefähr gleich wichtig bewertet, wodurch kein klares Wahlmotiv identifiziert werden kann. Das Testinstrument zur Erfassung der Wahlmotive deckte die Motive zur Wahl eines innerfachlichen Kontextes nicht ab. Bislang ist auch nur wenig über die Gründe zur Wahl eines innerfachlichen Kontextes bekannt. Der Vergleich des Interesses am Inhaltsfeld *saure und alkalische Lösungen* zeigt in

diesem Zusammenhang, dass das Interesse am Inhaltsfeld in dieser Gruppe im Vergleich zu den anderen Gruppen besonders ausgeprägt ist. Da der innerfachliche Kontext einen stärkeren Bezug zu dem Inhaltsfeld hat, könnte auch dieses Ergebnis eine Erklärung für die häufigere Auswahl des innerfachlichen Kontextes in diesem Cluster sein. Die etwas höhere Bedeutsamkeit des persönlichen Bezugs bei der Kontextwahl ist vermutlich die Ursache für die häufige Auswahl der alltäglichen Kontexte.

Die Lernenden des zweiten Clusters wählten am häufigsten einen besonderen Kontext zum Lernen aus (55.07 %). Der Anteil ist um ein Vielfaches höher als in der Gesamtstichprobe. Hier wählten nur 19.58 % der Lernenden einen besonderen Kontext. Ein etwas höherer Anteil als in der Gesamtstichprobe hat sich für den innerfachlichen Kontext entschieden. Die Schülerinnen und Schüler aus dieser Gruppe wählten sehr viel seltener einen der alltäglichen Kontexte. In diesem Cluster kann die vornehmliche Auswahl der besonderen Kontexte über das von den Lernenden angegebene Wahlmotiv des Überraschungsmoments erklärt werden. Ferner ist in der Literatur beschrieben, dass Schülerinnen und Schüler mit einem hohen Interesse und einer hohen Leistung das Lernen mit besonderen Kontexten präferieren (Habig et al., 2018). In diesem Zusammenhang ist anzunehmen, dass besondere Kontexte den Lernenden ermöglichen, das eigene Wissen durch unbekannte Kontexte zu erweitern.

Das dritte Cluster beinhaltet Lernende, die fast ausschließlich einen alltäglichen Kontext (86.11 %) wählten. Ein geringer Anteil der Lernenden wählten den innerfachlichen (7.41 %) oder einen besonderen Kontext (6.48 %). Die Kontextwahl in dieser Gruppe kann insbesondere durch die hohe Relevanz des persönlichen Bezugs bei der Wahl des Kontextes erklärt werden. Da es sich bei dieser Gruppe um Lernende mit einem geringen Vorwissen, Selbstkonzept und individuellen Interesse an Chemie handelt, könnte der hohe Alltagsbezug zum Chemielernen motivieren. Möglicherweise ist der Alltagskontext auch zugänglicher und verständlicher für schwache Lernende. In anderen Untersuchungen konnte ebenfalls gezeigt werden, dass niedrig interessierte Lernende eine Präferenz für Kontexte mit einem hohen Alltagsbezug haben (Habig et al., 2018; van Vorst & Aydogmus, 2021).

Im letzten Cluster wird der alltägliche Kontext (80.46 %) am häufigsten gewählt. Der Anteil ist dennoch etwas geringer als im dritten Cluster, aber höher als in der Gesamtstichprobe. Gleichwohl die beiden Wahlmotive als etwa gleich wichtig bewertet wurden, wurden besondere Kontexte von den Lernenden aus dieser Gruppe kaum ausgewählt. Aufgrund der geringen Punktzahl im Leseverständnistest ist unklar, inwieweit die Lesekompetenz ausreichend war, um die Kontextbeschreibung hinreichend zu verstehen. Demzufolge kann die Kontextwahl in dieser Gruppe nicht ohne weiteres mit der Kontextwahl der anderen Gruppen verglichen werden.

8.5.6 Evaluation der Kontextwahl

Als Indikatoren für die Passung der gewählten Kontexte wurden die Zufriedenheit, das situationale Interesse sowie die kognitive Belastung als Prozessvariablen nach der Aufgabenbearbeitung erhoben. In diesem Zusammenhang stellt sich insbesondere die Frage, ob sich die prozessbezogenen Variablen in Abhängigkeit vom gewählten Kontext in den identifizierten Lernendengruppen unterscheiden, was als Hinweis auf mangelnde Passung der Kontextaufgabe zu interpretieren wäre.

Zunächst wurde für das erste Cluster eine nicht-parametrische multivariate Varianzanalyse mit den prozessbezogenen Variablen als abhängige und dem gewählten Kontext als unabhängige Variable durchgeführt. Das Ergebnis der MANOVA verweist auf keine signifikanten Unterschiede in einer der prozessbezogenen Variablen in Abhängigkeit vom gewählten Kontext im ersten Cluster, $F(6.42, 162.57) = 0.92$, $p = .49$). In Tabelle 30 sind darüber hinaus die deskriptiven Statistiken für die prozessbezogenen Variablen in Abhängigkeit vom gewählten Kontext für das erste Cluster dargestellt.

Tabelle 30: Deskriptive Statistiken für die Zufriedenheit, die emotionale Valenz, die wertbezogene Valenz, die epistemische Komponente, die wahrgenommene Aufgabenschwierigkeit und die investierte Denkanstrengung, getrennt nach dem gewählten Kontext für das Cluster 1

	Alltäglich		Besonders		Innerfachlich	
	M	*SD*	*M*	*SD*	*M*	*SD*
Zufriedenheit	2.73	0.86	2.38	0.78	2.97	0.81
Emotionale Valenz	2.68	0.83	2.63	0.87	2.81	0.83
Wertbezogene Valenz	2.47	0.70	2.13	0.74	2.42	0.61
Epistemische Komponente	2.29	0.77	2.10	0.70	2.37	0.80
Wahrgenommene Aufgabenschwierigkeit	4.03	1.48	3.80	1.55	3.82	1.61
Investierte Denkanstrengung	4.23	1.65	3.80	1.40	4.46	1.64

Insbesondere Lernende, die einen alltäglichen oder den innerfachlichen Kontext gewählt hatten, gaben eine höhere Zufriedenheit nach der Aufgabenbearbeitung an. Auffällig ist hierbei, dass diese Kontexte in der Lernendengruppe am häufigsten gewählt wurden. Eine univariate Varianzanalyse verweist auf einen kleinen bis mittleren Effekt, der allerdings statistisch nicht signifikant ist, $F(2, 70) = 2.01$, $p = .14$, $\eta^2 = .05$. In Bezug auf die Subdimensionen des situa-

tionalen Interesses zeigen sich ähnliche Unterschiede, die allerdings nicht so deutlich ausgeprägt sind. Dementsprechend lassen sich auch hier keine signifikanten Unterschiede für die emotionale Valenz ($F(2, 70) = 0.26$, $p = .26$, $\eta^2 = .00$), die wertbezogene Valenz ($F(2, 70) = 0.98$, $p = .38$, $\eta^2 = .03$) oder die epistemische Komponente ($F(2, 70) = 0.43$, $p = .65$, $\eta^2 = .02$) des situationalen Interesses identifizieren. Insgesamt fallen die Effektstärken im Vergleich zur Zufriedenheit etwas geringer aus und sind im kleinen Bereich. Die deskriptive Analyse der Items zur Messung der kognitiven Belastung deutet auf eine höhere wahrgenommene Aufgabenschwierigkeit bei Lernenden mit alltäglicher oder innerfachlicher Kontextwahl hin. Ebenso können deskriptive Unterschiede in der investierten Denkanstrengung zwischen den Lernenden nachgewiesen werden, die einen alltäglichen und besonderen Kontext gewählt haben. Sowohl für die wahrgenommene Aufgabenschwierigkeit ($F(2, 70) = 0.17$, $p = .84$, $\eta^2 = .00$) und die investierte Denkanstrengung ($F(2, 70) = 0.64$, $p = .53$, $\eta^2 = .02$) sind die Unterschiede statistisch nicht signifikant und von vernachlässigbarer Effektstärke.

Auf Grundlage der berechneten nicht-parametrischen MANOVA können auch im zweiten Cluster keine Unterschiede in den prozessbezogenen Variablen in Abhängigkeit vom Kontext nachgewiesen werden, $F(6.42, 173.45) = 0.73$, $p = .64$. Auch in diesem Cluster zeigt sich, dass die Lernenden tendenziell zufriedener nach der Aufgabenbearbeitung sind, wenn sie mit dem besonderen Kontext gelernt haben, der in diesem Cluster am häufigsten gewählt worden ist (Tab. 31). Allerdings ist der Unterschied statistisch nicht signifikant, $F(2, 66) = 1.00$, $p = .37$, $\eta^2 = .03$).

Tabelle 31: Deskriptive Statistiken für die Zufriedenheit, die emotionale Valenz, die wertbezogene Valenz, die epistemische Komponente, die wahrgenommene Aufgabenschwierigkeit und die investierte Denkanstrengung, getrennt nach dem gewählten Kontext für das Cluster 2

	Alltäglich		Besonders		Innerfachlich	
	M	*SD*	*M*	*SD*	*M*	*SD*
Zufriedenheit	2.47	0.65	2.78	0.76	2.68	0.79
Emotionale Valenz	2.47	0.66	2.61	0.72	2.69	0.61
Wertbezogene Valenz	2.16	0.75	2.13	0.57	1.98	0.40
Epistemische Komponente	1.62	0.70	1.79	0.60	1.80	0.50
Wahrgenommene Aufgabenschwierigkeit	4.76	1.52	4.39	1.60	4.14	1.75
Investierte Denkanstrengung	3.82	1.67	4.32	1.68	4.00	1.36

Die deskriptive Analyse der Subdimensionen des situationalen Interesse deutet auf kein höheres Interesse der Lernenden hin, die einen besonderen Kontext bearbeitet haben. Bedeutsame Unterschiede bestehen weder im Hinblick auf die emotionale Valenz ($F(2, 66) = 0.44$, $p = .65$, $\eta^2 = .01$) oder wertbezogene Valenz ($F(2, 66) = 0.432$, $p = .65$, $\eta^2 = .01$), noch die epistemische Komponente des situationalen Interesses, $H(2) = 0.09$, $p = .95$, $\eta^2 = .00$. Genauso wenig können Unterschiede in der wahrgenommenen Aufgabenschwierigkeit ($F(2, 66) = 0.60$, $p = .55$, $\eta^2 = .02$) oder der investierten Denkanstrengung ($F(2, 66) = 0.60$, $p = .55$, $\eta^2 = .02$) identifiziert werden, wobei die beiden Items zur kognitiven Belastung hier konträre Ergebnisse hinsichtlich der deskriptiven Statistiken in Abhängigkeit vom bearbeiteten Kontext zeigen. So wurde die Aufgabenschwierigkeit des alltäglichen Kontexts von Lernenden dieses Clusters am höchsten empfunden, wobei die Lernenden bei der Bearbeitung der besonderen Kontexte die höchste Denkanstrengung investierten.

Auch für das dritte Cluster können keine signifikanten Unterschiede in den prozessbezogenen Variablen mithilfe der nicht-parametrischen MANOVA nachgewiesen werden, $F(5.44, 76.36) = 1.55$, $p = .15$). Wie in Tabelle 32 dargestellt, zeigen sich deskriptiv höhere Kennwerte auf der Zufriedenheitsskala und den Skalen zum situationalen Interesse, wenn die Lernenden aus dieser Gruppe mit einem besonderen Kontext gelernt haben.

Tabelle 32: Deskriptive Statistiken für die Zufriedenheit, die emotionale Valenz, die wertbezogene Valenz, die epistemische Komponente, die wahrgenommene Aufgabenschwierigkeit und die investierte Denkanstrengung, getrennt nach dem gewählten Kontext für das Cluster 3

	Alltäglich		Besonders		Innerfachlich	
	M	*SD*	*M*	*SD*	*M*	*SD*
Zufriedenheit	2.55	0.66	2.82	0.66	2.50	0.81
Emotionale Valenz	2.25	0.57	2.62	0.49	2.08	0.75
Wertbezogene Valenz	2.21	0.60	2.38	0.45	1.79	0.64
Epistemische Komponente	1.80	0.63	2.04	0.76	1.34	0.35
Wahrgenommene Aufgabenschwierigkeit	3.94	1.76	4.57	0.98	4.88	1.36
Investierte Denkanstrengung	4.35	1.66	4.43	1.72	4.50	1.60

Angesichts der post-hoc durchgeführten ANOVAs sind die Unterschiede in der Zufriedenheit ($F(2, 105) = 0.55$, $p = .58$, $\eta^2 = .01$), der emotionalen Valenz ($F(2, 105) = 1.72$, $p = .18$,

$\eta^2 = .03$), der wertbezogenen Valenz ($F(2, 105) = 2.20$, $p = .11$, $\eta^2 = .04$) und der epistemischen Komponente ($H(2) = 0.09$, $p = .95$, $\eta^2 = .00$) statistisch vernachlässigbar. Darüber hinaus geben die Lernenden aus diesem Cluster die höchste kognitive Belastung an, wenn sie mit dem innerfachlichen Kontext gelernt haben. Dagegen ist die kognitive Belastung am geringsten, wenn sie mit dem alltäglichen Kontext lernten, der am häufigsten in diesem Cluster gewählt wurde. Allerdings entsprechen die Unterschiede für die wahrgenommene Aufgabenschwierigkeit ($F(2, 105) = 0.03$, $p = .97$, $\eta^2 = .00$) und für die investierte Denkanstrengung ($F(2, 105) = 1.48$, $p = .23$, $\eta^2 = .03$) gemäß der berechneten univariaten Varianzanalyse einem kleinen und nicht signifikanten Effekt.

Die deskriptiven Statistiken der prozessbezogenen Variablen sind für das letzte Cluster in Tabelle 33 aufgeführt. Wie schon bei den anderen Clustern können gemäß der MANOVA alle Unterschiede als unbedeutend angesehen werden, $F(6.82, 97.47) = 0.76$, $p = .62$).

Tabelle 33: Deskriptive Statistiken für die Zufriedenheit, die emotionale Valenz, die wertbezogene Valenz, die epistemische Komponente, die wahrgenommene Aufgabenschwierigkeit und die investierte Denkanstrengung, getrennt nach dem gewählten Kontext für das Cluster 4

	Alltäglich		Besonders		Innerfachlich	
	M	*SD*	*M*	*SD*	*M*	*SD*
Zufriedenheit	2.81	0.63	2.89	0.64	2.92	0.34
Emotionale Valenz	2.80	0.65	2.67	0.56	2.50	0.46
Wertbezogene Valenz	2.66	0.61	2.45	0.40	2.33	0.82
Epistemische Komponente	2.19	0.61	2.36	0.41	1.96	0.64
Wahrgenommene Aufgabenschwierigkeit	3.96	1.35	4.73	1.49	4.33	1.37
Investierte Denkanstrengung	3.86	1.48	4.09	1.51	4.50	0.84

In der Zufriedenheit können nur geringe deskriptive Unterschiede herausgestellt werden, die dementsprechend statistisch nicht signifikant sind, $F(2, 84) = 0.15$, $p = .86$, $\eta^2 = .00$. Hinsichtlich des situationalen Interesses können wiederum gegensätzliche Ergebnisse durch die deskriptive Untersuchung der Kennzahlen aufgedeckt werden: Die Bearbeitung des alltäglichen Kontextes führt zu der höchsten emotionalen und wertbezogenen Valenz in dieser Lernendengruppe, wohingegen die Lernenden nach der Bearbeitung des besonderen Kontextes das höchste epistemische situationale Interesse zeigen. Die Unterschiede in der emotionalen Valenz ($F(2, 84) = 0.77$, $p = .47$, $\eta^2 = .02$), der wertbezogenen Valenz ($F(2, 84) = 1.24$, $p = .29$,

$\eta^2 = .03$) und der epistemischen Komponente ($F(2, 84) = 0.92$, $p = .40$, $\eta^2 = .02$) können aufgrund der univariaten Varianzanalyse statistisch nicht abgesichert werden. Die deskriptiven Ergebnisse zur kognitiven Belastung fallen erneut nicht eindeutig aus. In diesem Cluster kann die höchste investierte Denkanstrengung bei der Bearbeitung des innerfachlichen Kontextes beobachtet werden, wohingegen die Aufgabenschwierigkeit der besonderen Kontexte am höchsten wahrgenommen wird. Erneut sind die Unterschiede für die investierte Denkanstrengung ($F(2, 84) = 1.62$, $p = .20$, $\eta^2 = .04$) und die wahrgenommene Aufgabenschwierigkeit ($F(2, 84) = 0.62$, $p = .54$, $\eta^2 = .01$) statistisch irrelevant.

Durch die Analyse der prozessbezogenen Variablen je Cluster können keine statistisch bedeutsamen Unterschiede identifiziert werden, die auf mangelnde Passung der Kontextaufgaben hinweisen. Es gibt zwar Hinweise darauf, dass der Kontext, der am häufigsten im Cluster gewählt wurde, auch zur höchsten Zufriedenheit nach der Aufgabenbearbeitung führte (z. B. im Cluster 1), allerdings verweist die Signifikanzprüfung auf statistisch unbedeutende Unterschiede. Kleine bis mittlere Effektstärken können hierbei herausgearbeitet werden, vermutlich werden diese aber aufgrund des geringen Stichprobenumfangs und der damit verbundenen fehlenden statistischen Teststärke nicht signifikant. Gemäß einer Poweranalyse wird eine Stichprobe von mindestens 186 Lernenden benötigt, um Effektstärken von $\eta^2 = .05$ bei einem Signifikanzniveau von 5 % und einer Teststärke von .80 als statistisch signifikant auszuweisen. Die Größe der identifizierten Cluster liegt deutlich unter der berechneten Stichprobengröße. Zudem werden insbesondere im dritten und vierten Cluster einige Kontexte nur von sehr wenigen Lernenden gewählt, so dass Gruppenvergleiche hier auf der Basis sehr kleiner Teilstichproben erfolgen.

Ein Vergleich der identifizierten Cluster hinsichtlich der prozessbezogenen Variablen soll weiterführende Informationen über die Passung der gewählten Kontexte liefern. Die mittlere Zufriedenheit nach der Aufgabenbearbeitung für die vier Lernendengruppen ist in Abbildung 17 dargestellt.

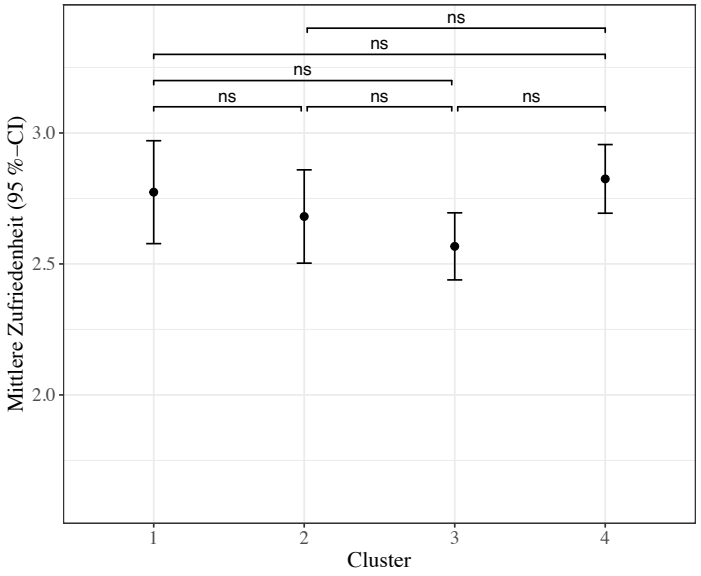


Abbildung 17: Vergleich der mittleren Zufriedenheit nach der Aufgabenbearbeitung zwischen den vier Clustern

Es zeigt sich ein geringer deskriptiver Unterschied in der Zufriedenheit in Abhängigkeit der Lernendengruppe, der durch eine ANOVA als statistisch unbedeutend ausgewiesen wird, $F(3, 333) = 2.42$, $p = .06$, $\eta^2 = .02$). Zudem handelt es sich um einen kleinen Effekt.

Das situationale Interesse nach der Aufgabenbearbeitung kann in Hinblick auf die drei postulierten Subdimensionen untersucht werden (Abb. 18).

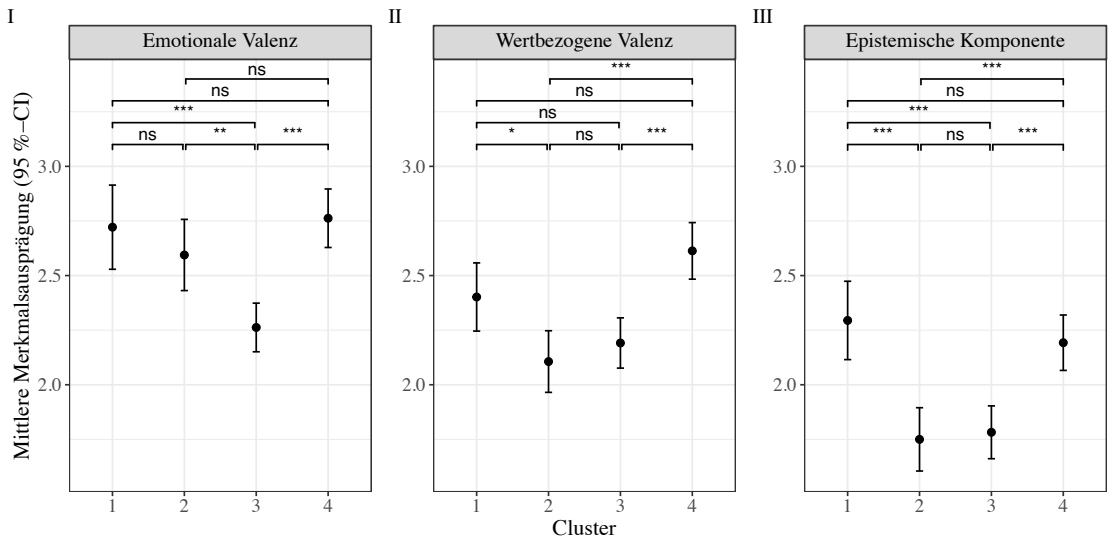


Abbildung 18: Vergleich der mittleren emotionalen Valenz (I), der wertbezogenen Valenz (II) und der epistemischen Komponente (III) des situationalen Interesses zwischen den vier Clustern

Für die drei Subdimensionen des situationalen Interesses zeigen sich signifikante Unterschiede zwischen den Lernendengruppen. Hinsichtlich der emotionalen Valenz besteht ein mittlerer bis großer Effekt zwischen den Clustern, $F(3, 169.36) = 12.94$, $p < .001$, $\eta^2 = .09$. Hier bewerten die Lernenden des dritten Clusters die emotionale Valenz signifikant geringer im Vergleich zu den übrigen Clustern. Die übrigen Cluster unterscheiden sich nicht signifikant voneinander. Sowohl für die wertbezogene Valenz als auch für die epistemische Komponente können signifikante Unterschiede von mittlerer bis großer Effektstärke ausgemacht werden. Die Analyse der deskriptiven Kennwerte zeigt ein ähnliches Muster: Das erste und vierte sowie das zweite und dritte Cluster zeigen eine ähnliche Merkmalsausprägung, wobei das erste und vierte Cluster höhere Werte auf der Skala zeigen als das zweite und dritte Cluster.

Die mittlere investierte Denkanstrengung und wahrgenommene Aufgabenschwierigkeit als Maß für die kognitive Belastung ist für die Lernendengruppen in Abbildung 19 dargestellt. Die Analyse der deskriptiven Kennwerte verweist auf geringe Unterschiede zwischen den Clustern. Diese Unterschiede sind sowohl für die investierte Denkanstrengung ($F(2, 333) = 0.29$, $p = .83$, $\eta^2 = .00$) als auch für die wahrgenommene Aufgabenschwierigkeit ($F(2, 333) = 2.57$, $p = .05$, $\eta^2 = .02$) statistisch nicht signifikant und von kleiner Effektstärke.

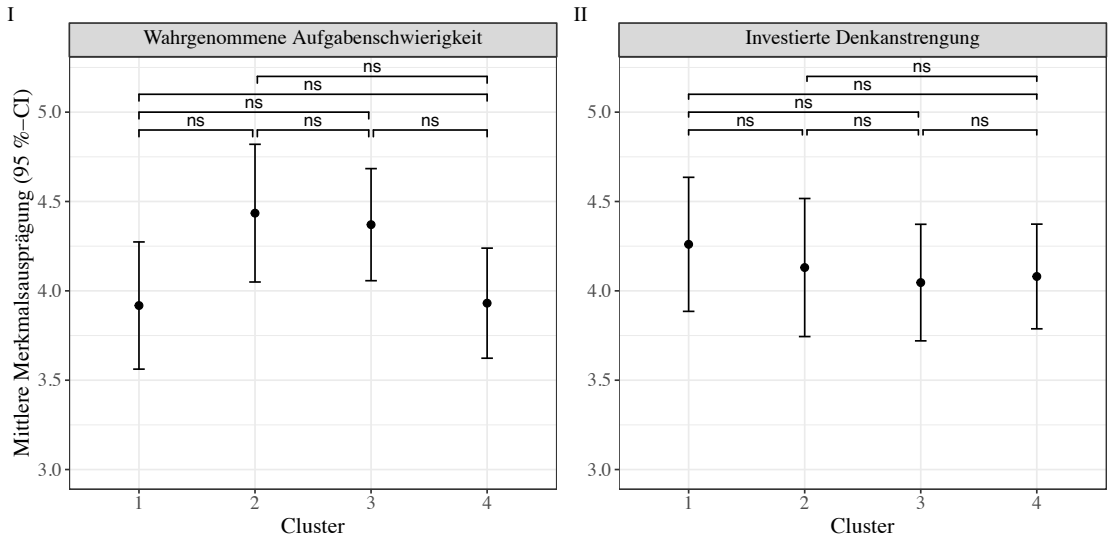


Abbildung 19: Vergleich der mittleren wahrgenommenen Aufgabenschwierigkeit (I) und der investierten Denkanstrengung (II) zwischen den vier Clustern

Insgesamt deutet sich damit an, dass alle Lernendengruppen nach der Aufgabenbearbeitung eine vergleichbare Zufriedenheit und kognitive Belastung angaben. Dennoch können Unterschiede im situationalen Interesse ausgemacht werden. Die Analyse des situationalen Interes-

ses in Abhängigkeit von den Lernendengruppen verweist darauf, dass die Lernenden des ersten und vierten Clusters das höchste situationale Interesse nach der Aufgabenbearbeitung zeigten. Auffällig ist, dass die Schülerinnen und Schüler dieses Clusters zugleich auch das höchste individuelle Interesse aufwiesen. Da sich die Lernendengruppen sowohl hinsichtlich ihrer Personenmerkmale (z. B. individuelles Interesse) als auch in den bearbeiteten Kontextaufgaben unterscheiden, ist unklar, worauf die Unterschiede im situationalen Interesse zurückzuführen sind.

Der Zusammenhang zwischen den Personenmerkmalen, dem gewählten Kontext und den prozessbezogenen Variablen kann durch ein Strukturgleichungsmodell untersucht werden. Hierbei wird lediglich ein Pfadmodell basierend auf den manifesten Variablen geschätzt, die auch schon für die bisherige Analyse genutzt wurden. Prinzipiell könnte zwar auch ein vollständiges Strukturgleichungsmodell geschätzt werden – welches neben dem Strukturmodell auch das korrespondierende Messmodell enthält – allerdings wird die hierfür zur Verfügung stehende Stichprobe als nicht ausreichend angesehen, da in der Literatur hier üblicherweise notwendige Stichprobenumfänge angegeben werden, die fünfmal so groß wie die Anzahl der zu schätzenden Parameter sind (siehe Weiber & Mühlhaus, 2014).

Gemäß der Konzeptualisierung der Zufriedenheit als übergeordnetes Merkmal wird die Zufriedenheit in die höchste Ebene des Pfadmodells integriert. Demgemäß wird angenommen, dass die Zufriedenheit sowohl durch das situationale Interesse und die kognitive Belastung als auch durch die Personenmerkmale sowie den gewählten Kontext beeinflusst wird. Auf der mittleren Ebene befinden sich das situationale Interesse und die kognitive Belastung, die wiederum durch die Merkmale der Person und den gewählten Kontext determiniert wird. Die unterste Ebene wird durch die Merkmale der Person und den gewählten Kontext gebildet. Hier wird auf die Personenmerkmale zurückgegriffen, die sich in der bisherigen Analyse als relevant erwiesen haben. Weiterhin wird die Punktzahl im LGVT als Kontrollvariable berücksichtigt. Da es sich bei Geschlecht und dem gewählten Kontext um nominale Variablen handelt, müssen diese vor der Aufnahme in das Pfadmodell in Dummy-Variablen überführt werden. Die Variable zur Erfassung des Geschlechts wird zu diesem Zweck dichotomisiert (0 = „männlich“, 1 = „weiblich“). Für die Erfassung des gewählten Kontextes werden zwei Indikatorvariablen gebildet. Die erste Dummy-Variable gibt an, ob ein alltäglicher Kontext gewählt oder nicht gewählt wurde (1 = „Alltäglicher Kontext“, 0 = „Kein alltäglicher Kontext“). Durch die zweite Dummy-Variable wird erfasst, ob ein besonderer Kontext gewählt oder nicht gewählt wurde (1 = „Besonderer Kontext“, 0 = „Kein besonderer Kontext“). Die Parameter des so spezifizierten Pfadmodells werden mit dem robusten Maximum-Likelihood-Schätzer berechnet, der auch bei nicht normalverteilten Indikatorvariablen robuste Standardfehler liefert (Gäde et al., 2020). Zur Verbesserung der Modellpassung werden nicht signifikante Pfade schrittweise auf den Wert Null fixiert und mittels χ^2-Test geprüft, ob die Restriktion zulässig ist. Ein signifikantes Ergebnis verweist darauf, dass der restringierte Parameter

ungleich Null ist und die Restriktion aufgehoben werden sollte (Kline, 2005; Weiber & Mühlhaus, 2014). Das so spezifizierte Pfadmodell mit standardisierten Pfadkoeffizienten ist in Abbildung 20 dargestellt und zeigt einen guten Modellfit. Die standardisierten Pfadkoeffizienten sind durch die durchgezogenen Pfeile dargestellt. Auf den gestrichelten Linien ist die standardisierte Kovarianz zwischen den Variablen abgebildet.

Die Zufriedenheit nach der Aufgabenbearbeitung wird durch die drei Subdimensionen des situationalen Interesses, der wahrgenommenen Aufgabenschwierigkeit, dem individuellen Interesse an Chemie und dem Leseverständnis determiniert. Durch diese Prädiktoren kann mehr als die Hälfte der Varianz ($R^2 = .55$) der Zufriedenheit erklärt werden. Den größten Einfluss auf die Zufriedenheit haben die drei Subdimensionen des situationalen Interesses. Die positiven Koeffizienten deuten darauf hin, dass mit steigendem situationalen Interesse an der Aufgabe auch die Zufriedenheit nach der Aufgabenbearbeitung zunimmt. Hier ist insbesondere die emotionale Valenz bedeutsam. Demnach sind vor allem positive Gefühle bei der Aufgabenbearbeitung ursächlich für eine hohe Zufriedenheit nach Bearbeitung der kontextorientierten Aufgabe. Neben dem situationalen Interesse hat auch das zeitlich überdauernde individuelle Interesse an Chemie einen substanziellen Effekt auf die Zufriedenheit nach der Aufgabenbearbeitung. Der negative Pfadkoeffizient deutet hierbei darauf hin, dass die Zufriedenheit mit zunehmendem individuellem Interesse an Chemie abnimmt. Die negative Wirkungsrichtung des individuellen Interesses ist überraschend, könnte aber darauf zurückzuführen sein, dass die Lernenden im Rahmen der Lerneinheit lediglich zwischen kontextualisierten Aufgaben wählen konnten. Möglicherweise bevorzugen die Lernenden mit zunehmendem Interesse Aufgaben, die weder einen innerfachlichen noch außerfachlichen Anwendungsbezug haben, die im Rahmen der Lerneinheit allerdings nicht zur Wahl standen. Ferner führt eine höhere wahrgenommene Aufgabenschwierigkeit zu einer geringeren Zufriedenheit. Den geringsten Effekt hat das Leseverständnis. Hier nimmt die Zufriedenheit erwartungskonform mit zunehmendem Leseverständnis zu. Da die Lernenden in der Lerneinheit ausschließlich mit textbasiertem Lernmaterial gearbeitet haben, ist anzunehmen, dass ein höheres Leseverständnis zu einer höheren Zufriedenheit nach der Bearbeitung der Aufgaben führte.

Für die kognitive Belastung zeigt sich ein sehr unbefriedigendes Ergebnis: Lediglich das Selbstkonzept hat einen Einfluss auf die wahrgenommene Aufgabenschwierigkeit, wodurch nur rund 2 % der Streuung aufgeklärt werden können. Ein höheres Selbstkonzept führte dazu, dass Aufgaben als weniger schwierig wahrgenommen wurden. In diesem Zusammenhang ließe sich eigentlich erwarten, dass das Vorwissen wesentlich für die wahrgenommene Schwierigkeit der Aufgabe ist. Dieser Wirkungszusammenhang kann angesichts der Pfadanalyse jedoch nicht bestätigt werden. Weitere personenbezogene Merkmale beeinflussen die kognitive Belastung nicht.

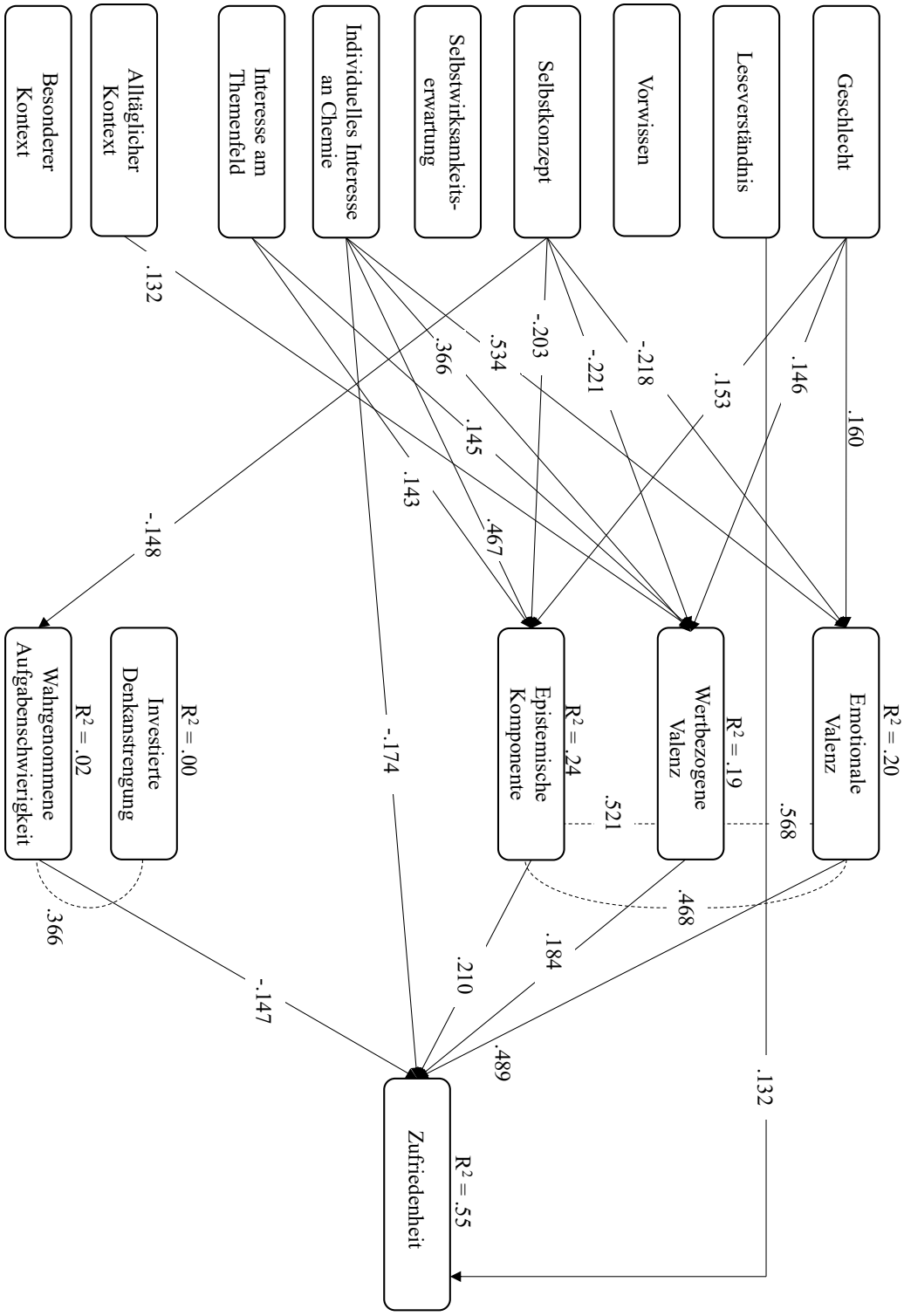


Abbildung 20: Pfadmodell zur Vorhersage der Zufriedenheit, des situationalen Interesse und der kognitiven Belastung nach der Aufgabenbearbeitung

Der gewählte Kontext hat keinen signifikanten Effekt auf die investierte Denkanstrengung oder die wahrgenommene Aufgabenschwierigkeit. Wahrscheinlich ist dieses Ergebnis darauf zurückzuführen, dass die kontextualisierten Aufgaben vollkommen identisch hinsichtlich der fachlichen Inhalte sind. Insgesamt scheinen weitere Faktoren die kognitive Belastung nach der Bearbeitung einer kontextualisierten Aufgabe zu determinieren, die allerdings in dieser Untersuchung unberücksichtigt blieben.

Die Subdimensionen des situationalen Interesses werden vorrangig durch die Merkmale einer Person beeinflusst. Den stärksten Prädiktor für die Ausprägung der drei Subdimensionen des situationalen Interesses stellt das individuelle Interesse an Chemie dar. Angesichts der Konzeptualisierung des situationalen Interesses als psychischer Zustand, der aus der konkreten Situation heraus entsteht, aber durch dispositionale Interessenstrukturen beeinflusst wird, erscheint die Datenlage plausibel. Insgesamt wirkt das individuelle Interesse stärker auf die emotionale Valenz als auf die wertbezogene Valenz und die epistemische Komponente.

Daneben werden die drei Subdimensionen des situationalen Interesses negativ von dem Selbstkonzept beeinflusst. Dieser Effekt wurde bereits von Habig (2017) im Rahmen einer experimentellen kontextbasierten Lerneinheit identifiziert. Dies könnte darauf hindeuten, dass Lernende mit einem hohen Selbstkonzept weniger situationales Interesse bei der Arbeit mit kontextualisierten Lernumgebungen verspüren, da sie bevorzugt mit nicht kontextualisierten Lernumgebungen arbeiten wollen. Zudem deutet der positive Pfadkoeffizient des Geschlechts darauf hin, dass Mädchen über ein höheres situationales Interesse bei der Aufgabenbearbeitung berichten. Möglicherweise ist dies auf den starken medizinischen Bezug der Aufgaben zurückzuführen. Aus der Literatur ist bekannt, dass Geschlechterdifferenzen zu medizinischen Themen zwar geringer sind als in anderen Themenbereichen, Mädchen allerdings an medizinischen Themen etwas interessierter sind als Jungen (Schiepe-Tiska, Simm & Schmidtner, 2016). Für die wertbezogene Valenz und epistemische Komponente kann das inhaltsfeldspezifische Interesse als weiterer Prädiktor identifiziert werden. Sowohl die wertbezogene Valenz als auch die epistemische Komponente nehmen mit steigendem Interesse am Inhaltsfeld saure und alkalische Lösungen zu. Dieser Effekt ist vermutlich darauf zurückzuführen, dass die empfundene Bedeutsamkeit und die epistemische Valenz stärker mit dem chemischen Inhalt korrespondieren als die bei der Aufgabe empfundenen Gefühle. Unter Kontrolle der personenbezogenen Merkmale führt lediglich die Wahl eines alltäglichen Kontextes zu einer höheren wertbezogenen Valenz. Dieses Ergebnis ist vor allem durch den größeren Alltagsbezug des alltäglichen Kontexts zu erklären. Daneben hat der gewählte Kontext keinen Effekt auf die prozessbezogenen Variablen. Insgesamt liegt die aufgeklärte Streuung für die drei Subdimensionen des situationalen Interesses zwischen 19 und 24 %. Nach Cohen (1988) entspricht dies einer moderaten bis hohen Varianzaufklärung.

8.6 Zusammenfassung und Implikationen für die weitere Untersuchung

Die für die entwickelten Kontextaufgaben angenommene Zugehörigkeit zu den Kontextmerkmalen konnte anhand des semantischen Differentials bestätigt werden. Bei der Analyse der Kontextwahl der Lernenden hat sich gezeigt, dass die meisten Lernenden einen außerfachlichen Kontext wählten. Alltägliche Kontexte wurden deutlich häufiger gewählt als besondere Kontexte, wobei besondere Kontexte kaum häufiger gewählt wurden als der innerfachliche Kontext. Bei den Alltagskontexten war der Kontext *Warum Zähneputzen so wichtig ist* beliebter als der Kontext *Unsere Verdauung*. Auch die besonderen Kontexte wurden nicht gleich häufig von den Lernenden ausgewählt. Hier war der Kontext *Schädigung der Knochen durch Übersäuerung* etwas beliebter als der Kontext *Die Folgen der chronischen Gastritis Typ-A*.

Durch die deskriptive und inferenzstatistische Analyse der kognitiven und affektiv-motivationalen Personenmerkmale in Abhängigkeit vom gewählten Kontext konnten für die Kontextwahl relevante Personenmerkmale herausgearbeitet werden. Neben den bereits in der Literatur aufgeführten Merkmalen (Leistung, Interesse und Wahlmotive) scheint insbesondere das Selbstkonzept sowie die Selbstwirksamkeitserwartung für die Kontextwahl bedeutsam zu sein. Der Vergleich der Kontextwahl der identifizierten Lernendengruppen impliziert, dass Lernende in Abhängigkeit von ihren Merkmalen unterschiedliche Kontexte zum Lernen auswählen. Demzufolge scheinen sich Kontexte mit systematisch variierten Merkmalen zur interessenbasierten Differenzierung zu eignen.

Innerhalb der Lernendengruppen zeigen sich keine statistisch bedeutsamen Unterschiede in den prozessbezogenen Variablen in Abhängigkeit vom gewählten Kontext, die auf mangelnde Passung einzelner Kontexte hindeuten. Zwar lassen sich in einigen Gruppen deskriptive Unterschiede nachweisen, diese können allerdings – vermutlich aufgrund der geringen Größe der Teilstichproben – nicht statistisch abgesichert werden. Es zeigt sich jedoch, dass die Lernenden zufriedener nach dem Lernprozess sind, wenn sie in dem Kontext gelernt haben, der in der Lernendengruppe am häufigsten gewählt wurde (z. B. in Cluster 1). Der Vergleich der Lernendengruppen im Hinblick auf die prozessbezogenen Variablen deutet darüber hinaus nicht auf Unterschiede in der aufgabenbezogenen Zufriedenheit oder der kognitiven Belastung hin. Hier konnten lediglich Unterschiede im situationalen Interesse identifiziert werden. Wie die multivariate Analyse mithilfe eines Pfadmodells zeigt, sind diese Unterschiede jedoch im Wesentlichen auf die individuellen Merkmale der Lernenden (z. B. individuelles Interesse an Chemie) und nicht auf den gewählten Kontext zurückzuführen. Für den weiteren Untersuchungsverlauf wird daher angenommen, dass die Lernenden für sie passende Kontexte zum Lernen auswählten.

9 Entwicklung des Vorhersagemodells

Für die Beantwortung der dritten Forschungsfrage war es notwendig, einem Teil der Lernenden einen passenden bzw. nicht passenden Kontext zuzuweisen (siehe Kapitel 6). Ausgangspunkt für diese Zuweisung stellen im Sinne eines Differenzierungsansatzes die Merkmale der Lernenden dar. Aufgrund der Ergebnisse der ersten Teilstudie kann hierbei näherungsweise angenommen werden, dass die Lernenden für sie passende Kontexte auswählen. Demzufolge können die im Rahmen der ersten Teilstudie identifizierten Zusammenhänge zwischen den Personenmerkmalen und der Kontextwahl für die Zuordnung passender Kontexte genutzt werden.

Ein solches Zuordnungssystem kann auf unterschiedlichen Wegen entwickelt werden: Im Rahmen einer traditionellen Programmierung könnten - basierend auf identifizierten Zusammenhängen zwischen den Personenmerkmalen und der Kontextwahl - Regeln formuliert werden, anhand derer das System entscheidet, welcher Kontext einer Person mit vorliegenden Merkmalen zum Lernen zugewiesen wird. Ein solches System würde allerdings hohe Anforderungen an die Identifizierung und Formulierung solcher Zuordnungsregeln stellen (Lanquillon, 2019; Nguyen & Zeigermann, 2021). Anstelle dieser aufwendigen manuellen Zuordnung kann mit Methoden des maschinellen Lernens (ML) ein statisches Modell entwickelt werden, das den Zusammenhang zwischen den Merkmalen einer Person und der Kontextwahl modelliert. Anhand dieses statistischen Modells kann die Zuordnung der Kontexte dann vorgenommen werden.

9.1 Maschinelles Lernen

Das maschinelle Lernen (engl. *machine learning*, ML) ist ein Teilgebiet der künstlichen Intelligenz, welches bislang durch keine allgemeingültige Definition beschrieben werden kann. Beim maschinellen Lernen handelt es sich um verschiedene Verfahren, bei denen Systeme so programmiert werden, dass sie aus Erfahrung lernen, um ein Problem zu lösen. Dieses Problem betrifft häufig die Vorhersage zukünftiger Ereignisse anhand von gesammelten Daten. Aus diesem Grund wird maschinelles Lernen auch häufig synonym mit dem Begriff des *predictive modeling* verwendet (M. Kuhn & Johnson, 2013; Lanquillon, 2019; Nguyen & Zeigermann, 2021). Mit Bezug auf Geisser (1993) definieren M. Kuhn und Johnson (2013, S. 2) predictive modeling wie folgt: „the process of developing a mathematical tool or model that generates an accurate prediction". Durch diese Definition wird zugleich auch deutlich, dass es sich beim predictive modeling bzw. maschinellen Lernen keineswegs um echtes Lernen handelt. Das zentrale Ziel ist viel mehr das Auffinden einer mathematischen Funktion (oder eines mathematischen Modells), um ein Kriterium bestmöglich anhand einer Reihe von Prädiktoren vorherzusagen (James et al., 2021). Es handelt sich folglich um ein Optimierungsproblem, welches durch einen Algorithmus gelöst wird (Géron, 2018; Lanquillon, 2019). Die

Modellierung des Zusammenhangs zwischen einer oder mehreren Input- und einer Output-Variable wird auch als *supervised learning* bezeichnet. In Abgrenzung dazu werden beim *unsupervised learning* unbekannte Strukturen in den Daten identifiziert. Hierfür werden häufig Clusterverfahren genutzt (Géron, 2018; James et al., 2021).

Beim supervised learning kann entsprechend der Skalierung des Kriteriums zwischen Regressions- und Klassifikationsproblemen unterschieden werden. Ein Regressionsproblem umfasst die Vorhersage einer metrischen abhängigen Variable, während Klassifikationsprobleme die Vorhersage einer nominalen abhängigen Variable betreffen. Diese nominale Variable kann sowohl dichotom als auch multinomial[6] sein (James et al., 2021). In dieser Arbeit wird ein multinomiales Klassifikationsproblem gelöst, da die Wahl von drei verschiedenen Kontexttypen vorhergesagt werden soll.

Ausgangspunkt für die Entwicklung eines Vorhersagemodells mithilfe von supervised learning ist die Aufteilung des ursprünglichen Datensatzes in einen Trainings- und Testdatensatz. Der Trainingsdatensatz wird zur eigentlichen Entwicklung des Modells genutzt. Dementgegen wird der Testdatensatz zur Evaluation der Vorhersagegüte einiger ausgewählter Modelle eingesetzt, um eine unverzerrte Schätzung der Performanz der Modelle zu erhalten. Normalerweise wird ein Großteil der Daten (70-80 %) im Trainingsdatensatz für die Entwicklung des Modells genutzt. Zur Festlegung des Trainingsdatensatzes empfiehlt sich eine Zufallsstichprobe, sofern die Klassen ungefähr gleich häufig vorkommen. Andernfalls sollte eine stratifizierte Zufallsstichprobe gezogen werden, damit der Anteil der Klassen in der Trainings- und Teststichprobe identisch ist (M. Kuhn & Johnson, 2020; M. Kuhn & Silge, 2022).

Bevor das Vorhersagemodell mithilfe der Trainingsstichprobe entwickelt wird, sollte eine geeignete Datenvorverarbeitung für den Modellierungsprozess durchgeführt werden. Diese umfasst nach M. Kuhn und Johnson (2013, 2020) die Auswahl relevanter Prädiktoren (*feature selection*) sowie deren Kodierung (*feature engineering*).

Grundsätzlich sollten nur Prädiktoren für die Entwicklung des Modells genutzt werden, die einen Beitrag zur Verbesserung der Vorhersagegüte leisten. Solche Merkmale können beispielsweise über inferenzstatistische Verfahren ausgewählt werden (M. Kuhn & Johnson, 2020). Daneben bieten auch die meisten ML-Algorithmen die Möglichkeit, die Relevanz der Prädiktoren bei der Konzeption des Modells zu determinieren. So wird häufig die Prädiktorenrelevanz mithilfe eines Random Forest-Modells bestimmt (James et al., 2021).

Für die Kodierung der Prädiktoren stehen unterschiedliche Transformationen zur Verfügung, damit die Informationen des Prädiktors effizienter genutzt werden können oder um sie überhaupt für die Modellierung zugänglich zu machen (M. Kuhn & Silge, 2022). Manche ML-

[6] In der englischsprachigen Literatur wird häufig auch von *multiclass classification* gesprochen, sofern die nominale Variable mehr als zwei Ausprägungen aufweist (z. B. M. Kuhn & Silge, 2022).

Algorithmen können nur schlecht mit Prädiktoren umgehen, die unterschiedlich skaliert sind. Metrische Prädiktoren können beispielsweise *z*-standardisiert werden, um sie vergleichbar zu skalieren (Géron, 2018). Überdies können hoch korrelierte Prädiktoren die Performanz vieler ML-Algorithmen beeinträchtigen (M. Kuhn & Silge, 2022). M. Kuhn und Johnson (2013) nennen hier einen exemplarischen Grenzwert von Interkorrelationen größer als $r = .75$ und empfehlen u. a. einen der hoch korrelierten Prädiktoren von der Modellierung auszuschließen.

Im Rahmen der Datenvorverarbeitung können außerdem unterschiedliche Strategien angewandt werden, um mit *class imbalance* umzugehen. Class imbalance besteht, wenn bestimmte Klassen bei der Klassifikation seltener auftreten als andere. Dies kann die Performanz der ML-Algorithmen beeinflussen (M. Kuhn & Johnson, 2013). Wenn es bei der Datenerhebung nicht möglich ist, alle Klassen in ähnlichen Verhältnissen zu erheben, können post-hoc auch sogenannte *sampling methods* angewandt werden. Bei großen Stichproben können zufällig Fälle aus den Klassen mit der höchsten relativen Häufigkeit ausgeschlossen werden (*down-sampling*). Für kleine Stichproben bietet es sich dagegen an, synthetische Daten mithilfe von *over-sampling* (auch *up-sampling*) zu generieren (M. Kuhn & Johnson, 2013). Hierfür kann z. B. die *synthetic minority over-sampling technique* (SMOTE) genutzt werden. Die SMOTE generiert synthetische Daten für Minoritätsklassen basierend auf einem zufälligen Datenpunkt der Minoritätsklasse und seinen nächsten Nachbarn (häufig fünf), die aufgrund der geringsten (euklidischen) Distanz ausgewählt werden. Der neue synthetische Datenpunkt ist eine zufällige Kombination der Prädiktoren des ausgewählten Datenpunkts der Minoritätsklasse und der nächsten Nachbarn. Diese Prozedur wird für die Minoritätsklassen so lange wiederholt bis alle Klassen gleich häufig vorkommen (Chawla et al., 2002). Der borderline SMOTE (BSMOTE)-Algorithmus generiert synthetische Datenpunkte für die Minoritätsklassen nur anhand von Datenpunkten, die sich nah an den Klassengrenzen (der *borderline*) befinden, die wiederum stark zur Abgrenzung der Klassen beitragen (Han et al., 2005). Einen ähnlichen Ansatz verfolgt der adaptive synthetic sampling approach (ADASYN): Hier werden synthetische Datenpunkte für die Minoritätsklasse aufgrund von Datenpunkten aus der Minoritätsklasse generiert, die mehrheitlich von Daten aus der gleichen Klasse umgeben sind (He et al., 2008).

Nachdem die Daten für die Modellierung vorbereitet wurden, können überwachte Lern-algorithmen zur Konstruktion des statistischen Modells genutzt werden. Dieser Prozess wird auch *Model building*, *Model training* oder auch *parameter estimation* genannt; er bezeichnet die Schätzung der Modellparameter anhand der Lernalgorithmen und der Trainingsdaten (M. Kuhn & Johnson, 2013). Gemäß dem *No-Free-Lunch-Theorem* von Wolpert (1996) zeigt kein ML-Algorithmus konsistent bei allen Anwendungen die höchste Performanz. Aus diesem Grund empfiehlt es sich bei der Modellierung eine Reihe von ML-Algorithmen auszuwählen, die Modelle hinsichtlich der Performanz zu untersuchen und einige wenige Modelle weiter zu optimieren (M. Kuhn & Silge, 2022). Mittlerweile stehen unzählige Lernalgorithmen für

das überwachte Lernen zur Verfügung. In Kapitel 9.2 wird ein kurzer Überblick über die ML-Algorithmen gegeben, die im Rahmen dieser Studie genutzt werden.

Die ML-Algorithmen unterscheiden sich insbesondere hinsichtlich der Komplexität der geschätzten Modelle. Komplexe Modelle eignen sich i. d. R besser, um die Realität abzubilden als einfache Modelle. Demnach weisen komplexere Modelle eine geringere Verzerrung (engl. *bias*) auf. Allerdings zeigt ein komplexes Modell eine höhere Varianz, da kleine Änderungen im Trainingsdatensatz zu starken Änderungen der Modellparameter führen. Eine hohe Varianz führt häufig dazu, dass das Modell zu stark an die Daten angepasst ist, wodurch es schlecht auf neue Daten generalisiert. Ein überangepasstes Modell ist zwar in der Lage, das Kriterium im Trainingsdatensatz nahezu perfekt vorherzusagen, allerdings sinkt die Performanz schlagartig bei der Vorhersage eines Datensatzes, der nicht zur Schätzung der Modellparameter genutzt wurde. Man spricht auch von *Overfitting* (James et al., 2021; M. Kuhn & Johnson, 2013). Das zentrale Ziel des Modellierungsprozesses ist es, sowohl die Verzerrung als auch die Varianz zu minimieren. Gemäß dem *bias-variance trade-off* ist es allerdings unmöglich, beides gleichzeitig zu minimieren. Die höchste Performanz zeigt demnach ein Modell mit mittlerer Komplexität (Géron, 2018; James et al., 2021).

Anhand der sogenannten *tuning parameters* (auch Hyperparameter) kann die Modellkomplexität weiter angepasst werden. Im Gegensatz zu den Modellparametern können die Hyperparameter nicht auf Grundlage der Trainingsstichprobe geschätzt werden, sondern müssen im Vorfeld festgelegt werden. Die meisten ML-Algorithmen haben i. d. R. mindestens einen Hyperparameter (M. Kuhn & Johnson, 2013; M. Kuhn & Silge, 2022). Die Optimierung dieser Hyperparameter wird auch als *Model tuning* bezeichnet und lässt sich in zwei Kategorien unterteilen: *grid search* und *iterative search* (M. Kuhn & Silge, 2022).

Beim grid search werden mögliche Werte für die Hyperparameter a priori in einem sogenannten grid angelegt. In einem *regular grid* sind alle Kombinationen der festgelegten Parameterwerte hinterlegt. Bei einer höheren Anzahl von Hyperparametern wird der Tuningprozess für viele Modelle dadurch allerdings sehr ineffizient. Eine Alternative dazu ist ein *random grid,* bei der zufällige Parameterkombinationen von definierter Anzahl festgelegt werden. Bei einer kleinen Anzahl von Parameterkombinationen führt die zufällige Auswahl allerdings häufig zu redundanten Parameterkombinationen. Zur Vermeidung von redundanten Parameterkombinationen können sogenannte *space-filling designs* genutzt werden, die im Parameterraum eine Konfiguration von Parametern identifizieren, die nur mit geringer Wahrscheinlichkeit redundant sind (M. Kuhn & Silge, 2022). Hierfür können beispielsweise *Latin hypercubes* genutzt werden (siehe McKay et al., 1979 für weitere Details)

Ein alternativer Ansatz zur Optimierung der Hyperparameter ist die iterative Suche nach bestmöglichen Hyperparametern, basierend auf vorherigen Ergebnissen. Die Bayes'sche Optimierung nutzt dafür ein Gauß-Prozess-Modell, welches auf Grundlage bereits evaluierter Hyperparameter weitere Parameter zur Untersuchung vorschlägt (Schulz et al., 2018). Dieser

Prozess wird so lange durchgeführt, bis keine Verbesserung mehr erzielt oder eine bestimmte Anzahl von Iterationen erreicht ist. Die iterative Suche kann nach dem grid search eingesetzt werden, um die Hyperparameter weiter zu optimieren (M. Kuhn & Silge, 2022).

Um eine unverzerrte Schätzung der Performanz während des Optimierungsprozesses zu erhalten, sollte möglichst eine weitere unabhängige Stichprobe genutzt werden, die nicht für den Modellierungsprozess zum Einsatz kam. Die Evaluation der Modelle anhand der Stichprobe, die zum Tuning verwendet wurde, führt zu verzerrten Schätzungen der Performanz. Wie bereits erwähnt, sollte die Teststichprobe allerdings bis zum Ende des Modellierungsprozesses nicht genutzt werden, um eine unverzerrte Schätzung der Vorhersagegüte der finalen Modelle zu erhalten. Zu diesem Zweck kann eine zusätzliche Validierungsstichprobe als weitere zufällige Stichprobe gezogen werden. Dieses Vorgehen bietet sich allerdings nur an, wenn die zur Verfügung stehende Gesamtstichprobe sehr groß ist (M. Kuhn & Silge, 2022). Stattdessen können Resampling-Methoden zum effizienten Umgang mit den Daten eingesetzt werden. Eine sehr weit verbreitete Methode im maschinellen Lernen ist die *k-fold Cross-Validation* (Kreuzvalidierung), bei der die Trainingsdaten zufällig in k Substichproben gleicher Größe – die sogenannten folds – aufgeteilt werden. Anschließend werden k Iterationen durchgeführt, bei der jeweils k-1 Substichproben für die Modellentwicklung und die übrige Substichprobe für die Evaluation der Performanz genutzt werden. So kommt bei jeder Iteration eine andere Substichprobe für die Modellevaluation zum Einsatz (James et al., 2021; M. Kuhn & Johnson, 2013, 2020). In der Literatur werden üblicherweise Werte von 5 oder 10 für k angegeben, da der geschätzte Wert der Performanz mit steigendem k eine weniger verzerrte Schätzung der wahren Performanz liefert (James et al., 2021; M. Kuhn & Johnson, 2013). Allerdings führt k-fold cross-validation zu einer hohen Varianz der Schätzung, da jede Iteration zu einer anderen Schätzung der Performanz führt. Die Varianz wird geringer, je häufiger die Kreuzvalidierung wiederholt wird. Deswegen empfehlen M. Kuhn & Johnson (2020) fünf Wiederholungen der 10-fold cross-validation bei kleinen Stichproben. Für jede Iteration steht so eine Schätzung der Vorhersagegüte zur Verfügung, die durch Berechnung des arithmetischen Mittels zusammengeführt wird (M. Kuhn & Johnson, 2020).

Die meisten Klassifikationsmodelle geben als Ergebnisse der Klassifikation für jeden Fall eine Wahrscheinlichkeit der Klassenzugehörigkeit an. Basierend auf der kalkulierten Wahrscheinlichkeit wird zusätzlich eine diskrete Klassenzugehörigkeit angegeben, welche die Klasse mit der höchsten Wahrscheinlichkeit darstellt. Zur Beurteilung der Performanz stehen bei Klassifikationsproblemen eine Reihe von Metriken zur Verfügung, die vornehmlich auf der diskreten Klassenzugehörigkeit basieren (M. Kuhn & Johnson, 2013). Viele dieser Metriken nutzen die Konfusionsmatrix, bei der die tatsächliche Kategorie in den Spalten und die vorhergesagte Kategorie in den Zeilen dargestellt ist oder vice versa. Auf der Diagonalen der Matrix befinden sich die richtigen Vorhersagen durch das Modell. Ein Beispiel für eine Konfusionsmatrix ist in Tabelle 34 dargestellt.

Tabelle 34: Beispiel für eine Konfusionsmatrix

Vorhersage	Tatsächliche Klasse	
	Positiv	Negativ
Positiv	*RP*	*FP*
Negativ	*FN*	*RN*

Anmerkung: *RP* = richtig Positiv, *FP* = falsch Positiv, *FN* = falsche Negativ, *RN* = richtig Negativ

Das einfachste Maß für die Performanz eines Klassifikators ist die Genauigkeit (Accuracy). Die Genauigkeit ist definiert als der Anteil der richtigen Vorhersagen (Géron, 2018):

$$\text{Accuracy} = \frac{\text{RP} + \text{RN}}{\text{RP} + \text{FP} + \text{FN} + \text{RN}} \tag{9}$$

Allerdings liefert die Genauigkeit nur bei balancierten Klassen ein sinnvolles Maß zur Beurteilung der Vorhersagegüte. Tritt eine Kategorie in 90 % der Fälle auf, lässt sich eine hohe Genauigkeit durch die Vorhersage der am häufigsten vorkommenden Klasse erzielen (Géron, 2018; M. Kuhn & Johnson, 2013).

Sowohl die *Precision* als auch der *Recall* gehören zu den wichtigsten Kennzahlen zur Beurteilung eines Klassifikators:

$$\text{Precision} = \frac{\text{RP}}{\text{RP} + \text{FP}} \qquad \text{Recall} = \frac{\text{RP}}{\text{RP} + \text{FN}} \tag{10}$$

Die Precision gibt die Genauigkeit der positiven Vorhersagen an: Welcher Anteil der positiven Vorhersagen war richtig positiv? Der Recall gibt dagegen an, wie hoch der Anteil der positiven Fälle ist, die tatsächlich vom Modell erkannt wurden. Ein hoher Recall impliziert, dass das Modell alle Fälle identifiziert (Géron, 2018; M. Kuhn & Johnson, 2013). Der harmonische Mittelwert zwischen der Precision und dem Recall wird F_1-score genannt (Géron, 2018).

Bei mehr als zwei Klassen erweitert sich die Konfusionsmatrix (Tab. 33) entsprechend der zusätzlichen Klassen. Die gerade angeführten Kennzahlen können bei multinomialen Klassifikationen separat für jede Klasse berechnet werden, indem – ähnlich wie bei der Dummy-Kodierung – lediglich die richtige Vorhersage für eine Klasse *k* oder die falsche Vorhersage für Klasse *k* berücksichtigt wird, wobei nicht weiter im Hinblick auf die falsche Vorhersage differenziert wird (Géron, 2018). Zur Berechnung einer globalen Kennzahl (z. B. Precision) kann der ungewichtete Durchschnitt berechnet werden, was als *macro-averaging* bezeichnet wird (M. Kuhn & Silge, 2022). Alle Klassen werden hierbei in gleichem Maße berücksichtigt:

$$\text{Precision}_{\text{macro}} = \frac{\text{Pre}_1 + \text{Pre}_2 + \ldots + \text{Pre}_k}{k} \tag{11}$$

9.2 Algorithmen zum maschinellen Lernen

Zur Konstruktion eines Vorhersagemodells für ein Klassifikationsproblem stehen eine Vielzahl von überwachten Lernalgorithmen zur Verfügung, die sich vereinfacht in lineare (z. B. Multinomiale logistische Regressionen und Diskriminanzanalysen) und nicht lineare Ansätze (z. B. Support Vector Machines und neuronale Netze) unterteilen lassen. Lineare Klassifikationsansätze nutzen lineare Entscheidungsgrenzen zur Klassifikation. Dementgegen können nicht lineare Ansätze häufig nicht lineare Zusammenhänge in den Daten effizienter modellieren (Hastie et al., 2009; M. Kuhn & Johnson, 2013). Da im Vorfeld häufig unklar ist, welche Art von Zusammenhängen in den Daten vorliegen, empfiehlt es sich, verschiedene ML-Algorithmen für das vorliegende Klassifikationsproblem zu testen.

Manche ML-Algorithmen (z. B. Random Forests oder neuronale Netze) können direkt zur multinomialen Klassifikation genutzt werden; bei anderen müssen eine Reihe von binären Modellen konzipiert werden, die anschließend zur Klassifikation von mehr als zwei Klassen genutzt werden können. Hierbei können zwei Ansätze unterschieden werden: Die *One-versus-All-* und die *One-versus-One*-Strategie (Géron, 2018).

Nach der One-vesus-All-Strategie wird für jede Klasse ein Modell trainiert, welches das Vorliegen oder nicht Vorliegen der Klasse anzeigt. Bei der One-versus-One-Strategie wird pro Klassenpaar ein Modell konstruiert, welches die beiden Klassen voneinander diskriminiert. Für N Klassen muss demnach eine Gesamtzahl von $N \times (N\text{-}1) / 2$ Modellen trainiert werden. Dadurch muss für jedes Modell nur eine kleine Auswahl der Trainingsstichprobe genutzt werden. Aus diesem Grund wird bei der Konstruktion von Support Vector Machines für multinomiale Klassifikationsprobleme meist die One-versus-One-Strategie genutzt, da Support Vector Machines schlecht mit großen Mengen von Trainingsdaten skalieren. Die meisten anderen ML-Algorithmen nutzen die One-versus-All-Strategie (Géron, 2018).

Im Folgenden erfolgt eine kurze Einführung in die ML-Algorithmen, die in dieser Arbeit genutzt wurden. Weitere überwachte Lernalgorithmen sind in M. Kuhn und Johnson (2013), James et al. (2021) sowie Alpaydin (2019) aufgeführt.

Multinomiale Logistische Regression: Bei logistischen Regressionsanalysen wird die Wahrscheinlichkeit modelliert, dass die abhängige Variable einer bestimmten Klasse angehört. Dazu wird eine logistische Funktion genutzt, deren Parameter häufig über die Maximum-Likelihood-Funktion geschätzt werden (James et al., 2021; M. Kuhn & Johnson, 2013):

$$\Pr(Y = 1 \mid X) = \frac{e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}} \tag{12}$$

Die Vorhersage basiert hierbei auf der Linearkombination der Prädiktoren (M. Kuhn & Johnson, 2013).

Naive Bayes: Ein Bayes-Klassifikator berechnet die posteriori Wahrscheinlichkeit der Zugehörigkeit zu einer Klasse *k* mithilfe des Satz von Bayes und anhand einer Reihe von Prädiktoren *X*:

$$\Pr(Y = k \mid X) = \frac{\Pr(Y)\,\Pr(X \mid Y = k)}{\Pr(X)} \tag{13}$$

Wenn bei der Berechnung der Wahrscheinlichkeit der Prädiktorvariablen vereinfacht angenommen wird, dass alle Prädiktoren unabhängig sind, wird der Klassifikator als naiv bezeichnet (M. Kuhn & Johnson, 2013).

***K*-Nearest Neighbors (KNN):** Bei dem *K*-Nearest Neighbors-Ansatz handelt es sich um ein nicht-parametrisches Klassifikationsverfahren, welches sich auch für multinomiale Klassifikationsprobleme eignet. Für die Klassifikation werden *K* Datenpunkte aufgrund der geringsten Distanz (z. B. euklidische Distanz) für einen zufällig ausgewählten Datenpunkt identifiziert. Diese *K* Datenpunkte werden als Neighbors bezeichnet, wobei die Anzahl der Neighbors vorab festgelegt werden muss. Die Wahrscheinlichkeit der Klassenzugehörigkeit für den zufällig ausgewählten Datenpunkt wird durch die Klassenzugehörigkeit der am nächsten befindlichen Datenpunkte bestimmt. Der Datenpunkt wird der Klasse zugeordnet, die am häufigsten unter den *K* Datenpunkten vorliegt (James et al., 2013).

Diskriminanzanalyse: Das Ziel der Diskriminanzanalyse ist es, eine (Diskriminanz-) Funktion zu identifizieren, welche die Wahrscheinlichkeit der Fehlklassifikation[7] minimiert (M. Kuhn & Johnson, 2013). Hierfür wird die posteriori Wahrscheinlichkeit mithilfe des Satzes von Bayes modelliert, so dass ein zu klassifizierender Fall einer bestimmten Klasse angehört (siehe naiver Bayes Klassifikator). Die Berechnung der posteriori Wahrscheinlichkeit erfolgt unter der Annahme, dass die Prädiktoren eine identische Kovarianzmatrix über alle Klassen aufweisen, wodurch die Diskriminanzfunktion eine lineare Form aufweist (lineare Diskriminanzanalyse, LDA). Bei quadratischen Diskriminanzanalysen (QDA) wird dagegeangenommen, dass die Kovarianzmatrix für die Klassen spezifisch ist. Dies führt zu quadratisch gekrümmten Diskriminanzfunktionen (siehe Hastie et al., 2009 für die mathematische Herleitung). Der gemeinsame Anteil der Kovarianzmatrix kann in regularisierten Diskriminanzanalysen (RDA) festgelegt werden, wodurch die Diskriminanzfunktion stärker an die Daten angepasst werden kann. Wenn der gemeinsame Anteil der Kovarianzmatrix gleich 1 ist, entspricht die Diskriminanzfunktion der LDA, wenn der gemeinsame Anteil gleich 0 ist, der QDA (M. Kuhn & Johnson, 2013). Darüber hinaus kann die Kovarianzmatrix in der RDA

[7] Der hier beschriebene Ansatz geht auf Welch (1939) zurück. Ein alternativer Ansatz der linearen Diskriminanzanalyse wird von Fisher (1936) beschrieben.

so modifiziert werden, dass sie sich der Einheitsmatrix nähert, wodurch Interkorrelationen der Prädiktoren modelliert werden können (Hastie et al., 2009). Lineare Diskriminanzanalysen können außerdem durch weitere Regularisierungsansätze flexibler gestaltet werden, was in der Literatur als flexible Diskriminanzanalyse (FDA) beschrieben wird. Details sind in Hastie et al. (2009) zu finden.

Baumbasierte Verfahren: Baumbasierte Verfahren verfolgen das Ziel, einen Entscheidungsbaum (Decision Tree) zu konzipieren, welcher die Daten in homogene Gruppen aufteilt. Die Aufteilung erfolgt anhand eines Entscheidungskriteriums in einer unabhängigen Variable, um ein sogenanntes *purity*-Kriterium wie den Gini-Index (G) zu minimieren (James et al., 2021):

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \tag{14}$$

Hierbei entspricht $\hat{p}_{mk}$ dem Anteil der Trainingsbeobachtungen der Klasse k in der Region m nach der Aufteilung anhand des Entscheidungskriteriums in einer unabhängigen Variable. Der Gini-Index wird minimiert, wenn $\hat{p}_{mk}$ nahe 0 oder 1 ist. Dann bestehen alle Knoten vornehmlich aus einer Klasse (M. Kuhn & Johnson, 2013). Mithilfe des Gini-Index kann außerdem die Relevanz der Prädiktoren bestimmt werden. Die Relevanz eines Prädiktors ist hierbei definiert als die Abnahme des Gini-Index (bei mehreren Bäumen die gemittelte Abnahme über alle Bäume) aufgrund der Aufspaltung des Baumes mithilfe eines bestimmten Prädiktors. Je größer die mittlere Abnahme des Gini-Index, desto wichtiger ist der Prädiktor für den Klassifikationsprozess (James et al., 2021).

Um Overfitting zu vermeiden, kann der Entscheidungsbaum hinsichtlich unterschiedlicher Parameter, wie der Tiefe des Baums und der Anzahl der Datenpunkte pro Knoten, restringiert werden (Géron, 2018). Darüber hinaus kann der Entscheidungsbaum auch zunächst ohne Einschränkungen konzipiert und anschließend hinsichtlich überflüssiger Knoten beschnitten werden (*Pruning*). Überflüssige Knoten liefern keine bedeutsame Verbesserung im purity-Kriterium (Géron, 2018; James et al., 2021). Ein zentraler Nachteil von Entscheidungsbäumen ist, dass sie eine hohe Varianz aufweisen, d. h. kleine Änderungen in den Daten führen zu anderen Modellparametern. Um die Varianz zu minimieren, können Bagging, Boosting oder ein Random Forest genutzt werden (James et al., 2021).

Beim *Bagging* werden mehrere Entscheidungsbäume aufgrund von Bootstrap-Trainingsstichproben konzipiert, wodurch die Varianz minimiert werden kann (M. Kuhn & Johnson, 2013). Für einen zu klassifizierenden Fall wird die Klasse vorhergesagt, die durch die Entscheidungsbäume am häufigsten vorhergesagt wurde (James et al., 2021). Ein *Random Forest* wird durch eine kleine Modifikation erzeugt: Erneut werden eine Reihe von Entscheidungsbäumen

mithilfe von Bootstrap-Trainingsstichproben erzeugt, allerdings wird an jedem Knoten nur eine zufällige Auswahl *m* der Prädiktoren *p* zur weiteren Aufspaltung des Baumes genutzt[8]. Durch die zufällige Auswahl der Prädiktoren wird verhindert, dass sehr starke Prädiktoren immer im ersten Knoten des Baumes genutzt werden, wodurch die Entscheidungsbäume dekorreliert werden (James et al., 2021). Bei einem Random Forest müssen die Anzahl der Prädiktoren *m* und die Anzahl der Bäume vorab festgelegt werden (M. Kuhn & Johnson, 2013). Ähnlich wie beim Bagging und beim Random Forest werden auch beim *Boosting* mehrere Entscheidungsbäume zur Vorhersage genutzt. Diese werden allerdings sequenziell erzeugt, so dass Informationen aus vorangegangenen Bäumen zur Konstruktion neuer Bäume genutzt werden können. Schrittweise werden so neue Bäume konzipiert, welche die Vorhersagegüte weiter verbessern. Die Anpassungsrate in jeder Iteration des Boosting-Algorithmus wird über den sogenannten shrinkage parameter festgelegt. Normalerweise führt eine geringe Anpassungsrate zu Modellen mit einer höheren Vorhersagegüte (M. Kuhn & Johnson, 2013; James et al., 2021). Daneben kann bei jeder Iteration des Boostings nur ein Teil der Trainingsstichprobe zur Konstruktion des nächsten Baumes genutzt werden, wodurch häufig die Performanz des Modells verbessert wird. Dieser Anteil muss vor der Schätzung der Modellparameter festgelegt werden (Hastie, 2009).

Ein weit verbreitetes und performantes Boosting-Verfahren ist XGBoost (eXtreme Gradient Boosting). Für eine tiefergehende mathematische Einführung in XGBoost sei auf T. Chen und Guestrin (2016) verwiesen.

Support Vector Machines: Eine Support Vector Machine (SVM) ist ein Klassifikationsansatz, der eine sogenannte *hyperplane* zur Klassifikation nutzt. In einem *p*-dimensionalen Raum ist eine hyperplane als flacher Unterraum der Dimension *p*-1 definiert. So ist die hyperplane in einem 2-dimensionalen Raum beispielsweise eine einfache Linie und im 3-dimensionalen Raum eine Ebene (M. Kuhn & Johnson, 2021). Das zentrale Ziel der SVM ist es, eine hyperplane zu finden, welche zwei Klassen bestmöglich diskriminiert. Eine optimale hyperplane besitzt den größtmöglichen Abstand zu den Trainingsdatenpunkten; diese Distanz wird als *margin* bezeichnet und die hyperplane demnach als *maximal margin hyperplane* (Géron, 2018; James et al., 2021). Die Datenpunkte, welche die geringste Distanz zu der hyperplane aufweisen, sind die *support vectors* (M. Kuhn & Johnson, 2013; James et al., 2013). In Abbildung 21 sind die maximal margin hyperplane und die support vectors schematisch dargestellt.

Häufig ist es nicht möglich, eine hyperplane zu finden, die beide Klassen exakt trennt, da sie nicht linear separierbar sind. Durch sogenannte *Kernel*-Funktionen kann die SVM allerdings

[8] i. d. R werden $m = \sqrt{p}$ Prädiktoren genutzt (Breimann, 2001).

eine hyperplane konzipieren[9], die sich flexibel an die Klassengrenzen anpassen lässt (M. Kuhn & Johnson, 2013). Zwei häufig genutzte Kernel-Funktionen sind der polynomiale Kernel und der Kernel mit radialer Basisfunktion (Rbf). Die Anpassung der SVM mit polynomialem Kernel erfolgt über den Grad des Polynoms, wohingegen die SVM mit Rbf-Kernel durch den Parameter für die radiale Basisfunktion angepasst werden kann (M. Kuhn & Johnson, 2013). Je stärker die hyperplane und die margin an den Trainingsdatensatz angepasst wird, desto eher kommt es zum overfitting.

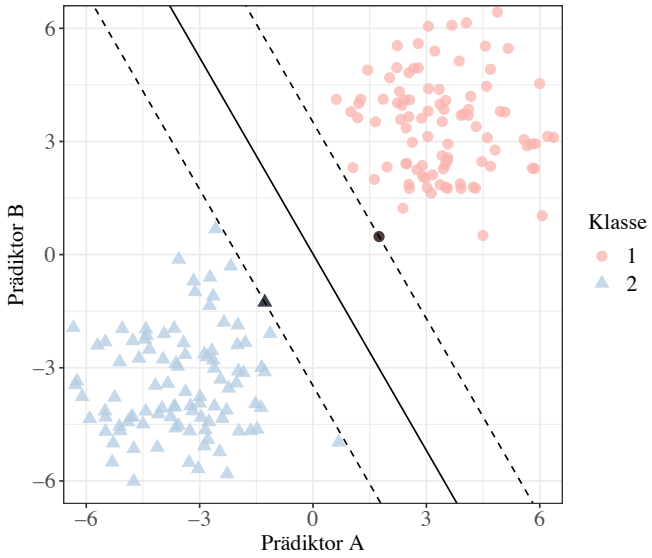


Abbildung 21: Die maximal margin hyperplane (schwarze durchgestrichene Linie), die margin (gestrichelte Linien) und die support vectors (schwarze Punkte) nach M. Kuhn und Johnson (2013)

Zur Vermeidung von overfitting kann eine soft margin eingerichtet werden, welche einen gewissen Anteil von Verletzungen der margin und hyperplane toleriert (James et al., 2021). Verletzungen bestehen, wenn Datenpunkte falsch klassifiziert sind und sich auf der falschen Seite der hyperplane oder margin befinden. Diese Verletzungen können mit einem Kostenparameter *C* belastet werden. Wenn die Kosten *C* sehr hoch sind, wird die margin schmaler, so dass keine Datenpunkte falsch klassifiziert werden. Dadurch besteht allerdings die Gefahr des overfitting (M. Kuhn & Johnson, 2013).

Neuronale Netze: Künstliche neuronale Netze sind nicht-lineare statistische Modelle, die nach dem Vorbild biologischer Neuronen konstruiert wurden. Das am häufigsten genutzte

[9] Tiefergehende Informationen zu den mathematischen Hintergründen sind in Alpaydin (2008) dargestellt.

neuronale Netze ist das feed-forward neuronale Netz mit einem hidden layer, welches schematisch in Abbildung 22 dargestellt ist (Hastie et al., 2009).

Zur Klassifikation nutzen neuronale Netze einen (oder mehrere) hidden layer, welcher aus sogenannten hidden units besteht. Hidden units sind eine Linearkombination der Prädiktoren, die typischerweise durch die Sigmoidfunktion transformiert werden (Hastie et al., 2009; M. Kuhn & Johnson, 2013). In modernen neuronalen Netzen wird statt der Sigmoidfunktion die ReLU-(*rectified linear unit*)-Funktion verwendet (James et al., 2021). Die Anzahl der hidden units muss bei der Konstruktion festgelegt werden, wobei eine höhere Anzahl häufig geeigneter ist, um nicht-lineare Zusammenhänge in den Daten zu modellieren (Hastie et al., 2009). Die Vorhersage für die Klassenzugehörigkeit basiert wiederum auf einer Linearkombination der hidden units, die erneut durch die Sigmoidfunktion transformiert werden. Da die Ausgabe allerdings keiner Wahrscheinlichkeit der Klassenzugehörigkeit entspricht, werden sie durch die softmax-Transformation in eine Metrik überführt, die mathematisch äquivalent zur Wahrscheinlichkeit ist (Géron, 2018; Hastie et al., 2009; M. Kuhn & Johnson, 2013).

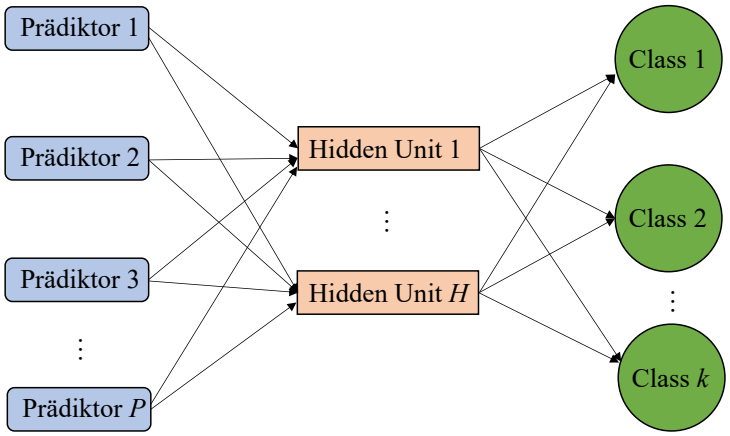


Abbildung 22: Schematische Darstellung eines feed-forward neuronalen Netz mit einem hidden layer zur Klassifikation nach M. Kuhn und Johnson (2013)

Für die Linearkombination der Prädiktoren und der hidden units werden im neuronalen Netz *weights* als unbekannte Parameter durch den back-propagation-Algorithmus geschätzt, um den Vorhersagefehler zu minimieren (siehe Rumelhart et al., 1986 für weitere Informationen). Neuronale Netze neigen u. a. aufgrund der hohen Anzahl der unbekannten Parameter zum overfitting. Aus diesem Grund kann die Größe der weights über den *weight decay* regularisiert werden, wodurch overfitting vermieden wird. Bei dem weight decay handelt es sich um einen Hyperparameter (Hastie et al., 2009; M. Kuhn & Johnson, 2013). Daneben sollte auch die Anzahl der Epochen optimiert werden, um overfitting zu vermeiden (Hastie et al., 2009).

9.3 Modellierung

Das Vorhersagemodell wurde in der statistischen Programmiersprache R (Version 4.2.3) mithilfe von *tidymodels* entwickelt. Bei tidymodels handelt es sich um eine Sammlung von R-Paketen, die verschiedene Funktionen zur Entwicklung von ML-Modellen in R bereitstellen. Eine Einführung in tidymodels geben die Entwickler M. Kuhn und Silge (2022) in ihrer zugehörigen Publikation.

9.3.1 Datenvorverarbeitung

Für die Entwicklung des Vorhersagemodells wurde der Datensatz (N = 347) der ersten Teilstudie zufällig in eine Trainings- (n = 259) und eine Teststichprobe (n = 88) aufgeteilt, womit 75 % der Daten für die Trainingsstichprobe genutzt werden. Zur Berücksichtigung der ungleichen Kontextwahl wurde eine stratifizierte Stichprobe gezogen.

Vorab mussten für die Entwicklung des Vorhersagemodells geeignete Variablen ausgewählt werden, die zur Prädiktion der Kontextwahl genutzt werden können. Hierfür konnte auf die deskriptive und inferenzstatistische Analyse zurückgegriffen werden, die bereits zur Variablenauswahl für die Clusteranalyse durchgeführt wurden. Im Rahmen dieser Analyse haben sich das Geschlecht, das Vorwissen in Chemie, die letzte Chemienote, das individuelle Interesse an Chemie, das inhaltsfeldspezifische Interesse, das Selbstkonzept, die Selbstwirksamkeitserwartung sowie die Wahlmotive persönlicher Bezug und Überraschungsmoment als bedeutsam erwiesen. Wie schon bei der Clusteranalyse, blieb die letzte Chemienote für die Modellierung unberücksichtigt, da das auf Grundlage des Leistungstests erhobene Vorwissen das validere Maß für die Leistung darstellt. Die Selbstwirksamkeitserwartung wurde aufgrund der hohen Korrelation (r = .75, p < .001) mit dem Selbstkonzept nicht für die Entwicklung des Modells genutzt (M. Kuhn & Johnson, 2013). Die etwas geringere Korrelation zwischen dem individuellen Interesse an Chemie und dem Interesse am Inhaltsfeld saure und alkalische Lösungen (r = .68, p < .001) erlaubt die Berücksichtigung beider Facetten des Interesses. Dahingegen wurde das Geschlecht nicht genutzt, da sich in der weiterführenden Analyse gezeigt hat, dass sich die Geschlechter hinsichtlich des themenfeldspezifischen Interesses unterscheiden. Stattdessen wurde untersucht, ob die Berücksichtigung des themenfeldspezifischen Interesses die Performanz des Modells verbessert.

Dazu wurden verschiedene Modelle mithilfe von den in Kapitel 9.2 beschriebenen ML-Algorithmen unter Berücksichtigung des themenfeldspezifischen Interesses trainiert und mit Modellen kontrastiert, die ohne das themenfeldspezifische Interesse konstruiert wurden. Hierbei wurden die Standardeinstellungen der jeweiligen R-Pakete für die Hyperparameter genutzt, die im Anhang aufgeführt sind. In Abbildung 23 ist der kreuzvalidierte F_1-score der Modelle dargestellt, die durch die beiden Variablengruppen entwickelt wurden.

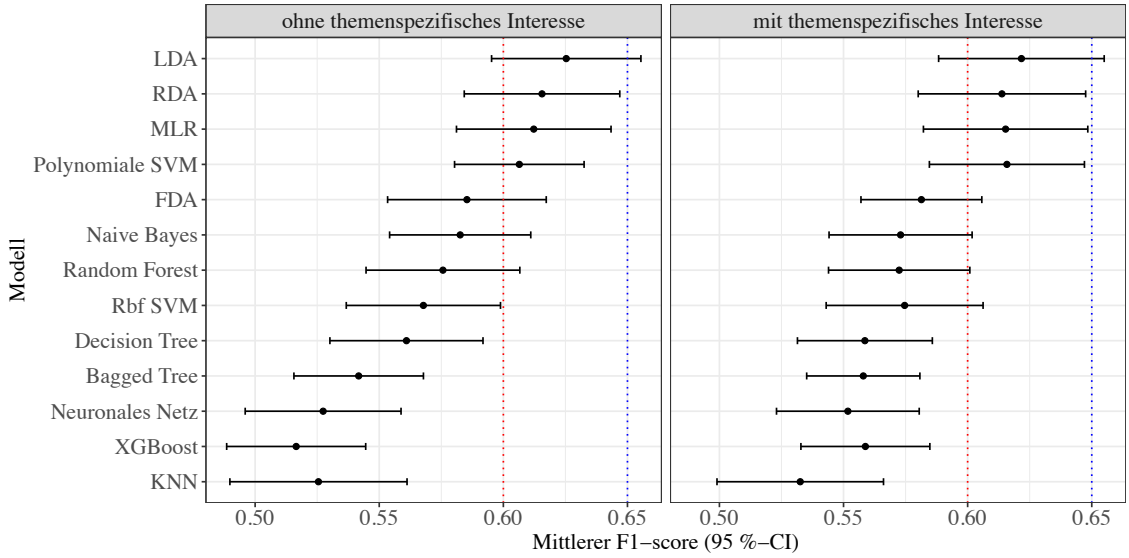


Abbildung 23: Mittlerer F_1-score der trainierten Modelle unter Ausschluss und Berücksichtigung des themenspezifischen Interesses

Unabhängig von der Berücksichtigung des themenbezogenen Interesses führen die weniger komplexen Modelle zu einem höheren F_1-score. Die lineare und regularisierte Diskriminanzanalyse sowie die multinomiale logistische Regression zeigen die höchste Performanz. Etwas schlechter schneiden komplexere Modelle (z. B. XGBoost oder neuronales Netz) ab, deren Hyperparameter noch nicht optimiert wurden. Hier zeigt sich, dass insbesondere die komplexeren Modelle von der Berücksichtigung des themenfeldspezifischen Interesses profitieren. Der über alle Modelle gemittelte F_1-score steigt durch die Berücksichtigung des Interesses am Themenfeld Körper und Krankheiten nur gering (ΔF_1-score = .008).

Zur weiteren Untersuchung der Relevanz der Prädiktoren wurde ein Random Forest-Modell mit 1000 Bäumen berechnet. In Abbildung 24 ist die mittlere prozentuale Abnahme des Gini-Index durch die Prädiktoren relativ zum Prädiktor mit der größten durchschnittlichen Abnahme dargestellt. So bekommt der wichtigste Prädiktor einen Wert von 100 % zugewiesen. Der wichtigste Prädiktor für den Klassifikationsprozess ist das Wahlmotiv des persönlichen Bezugs. Daneben sind das Vorwissen in Chemie und das Selbstkonzept entscheidend. Gemäß dem Random Forest-Modell ist das themenspezifische Interesse bedeutender als das inhaltsfeldspezifische Interesse. Das inhaltsfeldspezifische Interesse ist von allen Prädiktoren am wenigsten relevant. Da sich allerdings in der inferenzstatistischen Analyse ein statistisch bedeutender Effekt des inhaltsspezifischen Interesses gezeigt hat, wird es weiter im Modellierungsprozess genutzt. Überdies wird das themenspezifische Interesse für die Entwicklung des Modells berücksichtigt, da insbesondere die komplexeren Modelle durch die Berücksichtigung des themenspezifischen Interesses profitieren und es sich durch das Random Forest-Modell im Vergleich zum inhaltsfeldspezifischen Interesse sogar als relevanter herausgestellt

hat. Insgesamt werden damit alle in Abbildung 24 dargestellten Prädiktoren zur Vorhersage der Kontextwahl genutzt.

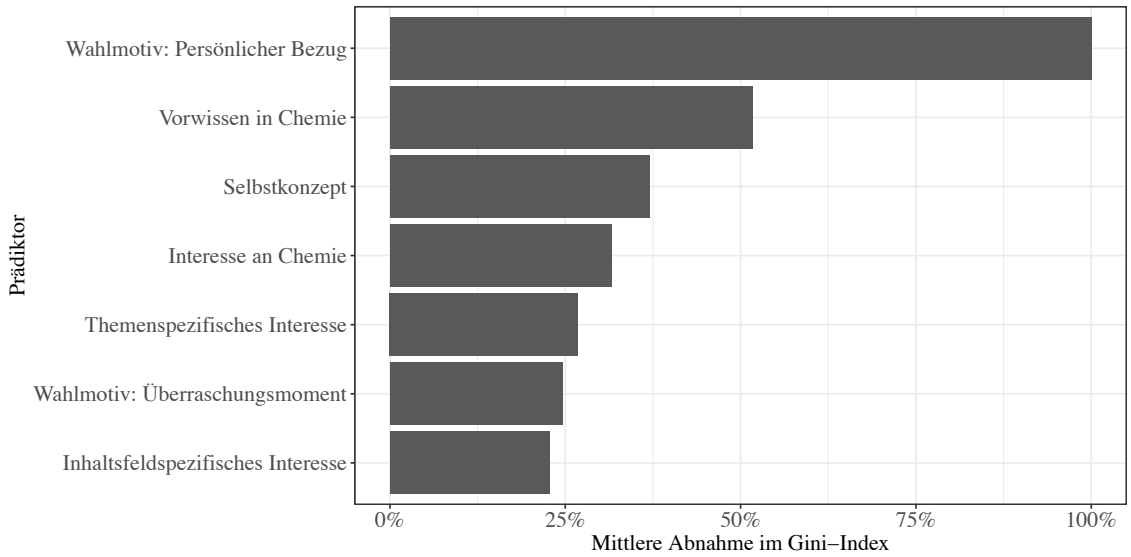


Abbildung 24: Die relative mittlere prozentuale Abnahme im Gini-Index durch die Prädiktoren

Aufgrund der ungleichen Verteilung der Kontextwahl wurden bei der Datenvorverarbeitung mehrere over-sampling Algorithmen im Hinblick auf ihre Performanz verglichen (Abb. 25). Im Vergleich zur Modellierung ohne over-sampling schneidet der ADASYN-Algorithmus über alle Modelle insgesamt etwas schlechter ab (ΔF_1-score = -.018). Die Verwendung des SMOTE-Algorithmus führt im Vergleich zur Modellierung ohne over-sampling nur zu einer minimalen Verbesserung in der Vorhersagegüte (ΔF_1-score = .002), während durch den BSMOTE-Algorithmus ein geringer Zugewinn in der Vorhersagegüte erzielt wird (ΔF_1-score = .011). Insbesondere die komplexeren Modelle (z. B. XGBoost) profitieren geringfügig vom over-sampling mit dem SMOTE und BSMOTE-Algorithmus. Die Vorhersagegüte der einfacheren Modelle (z. B. LDA, RDA, MLR) werden vom over-sampling insgesamt nur wenig beeinflusst. Da der BSMOTE-Algorithmus die Performanz der Modelle geringfügig verbessert, wird dieser Algorithmus zur Modellierung genutzt.

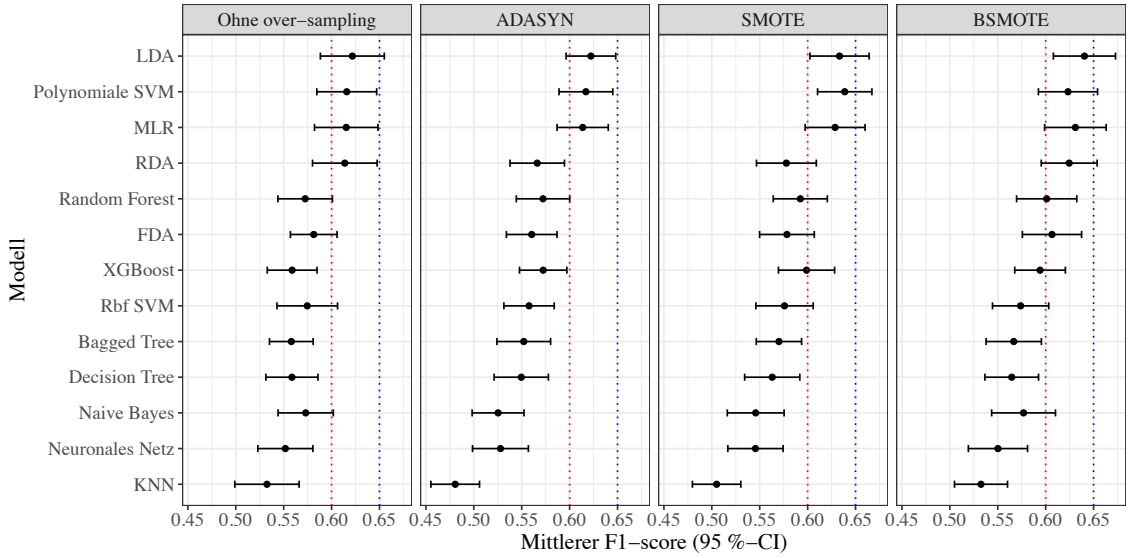


Abbildung 25: Vergleich des mittleren F_1-scores der trainierten Modelle unter Anwendung verschiedener oversampling Algorithmen

9.3.2 Optimierung der Hyperparameter

Die Performanz von komplexeren Modellen wird besonders durch die Wahl der richtigen Hyperparameter (tuning parameter) beeinflusst (M. Kuhn & Silge, 2022). Aus diesem Grund wurde eine Reihe von vielversprechenden Modellen im Tuning-Prozess weiter optimiert. Dazu gehören das Random-Forest-Modell, das XGBoost-Modell, die beiden Support Vector Machines und das neuronale Netz. Die im Rahmen des Tuning-Prozesses weiter optimierten Hyperparameter sind in Tabelle 35 aufgeführt.

Die Hyperparameter der Modelle wurden zunächst im grid search mithilfe eines random grid und 25 zufälligen Parameterkombinationen optimiert, um den F_1-score zu maximieren. Zur Vermeidung von redundanten Parameterkombinationen wurde ein Latin hypercube-Design genutzt (M. Kuhn & Silge, 2022). Das random grid wurde mithilfe von 10-fold cross-validation (fünf Wiederholungen) evaluiert, um die potenteste Parameterkombination auszuwählen. Ausgehend von der so bestimmten Parameterkombination wurden in der Bayes'schen Optimierung in 25 Iterationen weitere geeignete Parameterkombinationen identifiziert und anhand von Kreuzvalidierung evaluiert. Die performanteste Parameterkombination für jedes Modell ist im Anhang aufgelistet.

Tabelle 35: Auflistung der optimierten Hyperparameter im Tuningprozess für jeden ML-Algorithmus

ML-Algorithmus	Hyperparameter
Neuronales Netz	Anzahl der hidden units Anzahl der Epochen Weight decay
Polynomiale SVM	Kosten Grad der Polynomfunktion
Rbf SVM	Kosten Parameter der radialen Basisfunktion
Random Forest	Anzahl der zufällig ausgewählten Prädiktoren Anzahl der Bäume Minimale Knotengröße
RDA	Anteil der gemeinsamen Kovarianzmatrix Anteil der Einheitsmatrix
XGBoost	Baumtiefe Lernrate Anzahl der zufällig ausgewählten Prädiktoren Minimale Knotengröße Anteil der Beobachtungen in der Stichprobe

In Abbildung 26 sind die mittleren F_1-scores basierend auf den Ergebnissen der Kreuzvalidierung für die Modelle mit den performantesten Parameterkombinationen dargestellt, die durch den jeweiligen Tuning-Prozess identifiziert wurden. Zum Vergleich sind außerdem die mittleren F_1-scores der linearen Diskriminanzanalyse und der multinomialen logistischen Regression dargestellt, deren Hyperparameter nicht optimiert wurden.

Insbesondere die Vorhersagegüte des XGBoost-Modells ($t(98) = -3.32$, $p = .001$, $r = .31$), der Support Vector Machine mit Rbf Kernel ($t(98) = -2.89$, $p = .005$, $r = .28$) und des neuronalen Netzes ($t(98) = -2.76$, $p = .007$, $r = .27$) konnte durch die grid search signifikant verbessert werden. Die grid search der übrigen Modelle hat lediglich zu einer geringen deskriptiven Verbesserung des F_1-scores geführt. Durch die daran anknüpfende Bayes'sche Optimierung konnte auch nur eine geringe deskriptive Verbesserung des F_1-scores erzielt werden. Am deutlichsten haben die beiden Support Vector Machines und das neuronale Netz von der Bayes'schen Optimierung profitiert.

Insgesamt zeigt das XGBoost-Modell damit die höchste Vorhersagegüte basierend auf dem Kreuzvalidierungsansatz in der Trainingsstichprobe. Der Vergleich der 95 %-Konfidenzintervalle zeigt allerdings, dass das berechnete Konfidenzintervall für den mittleren F_1-score des XGBoost-Modells die mittleren F_1-scores der übrigen Modelle einschließt, wodurch auf keine statistisch signifikanten Unterschiede zwischen den Modellen zu schließen ist. Demnach zeigen alle entwickelten Modelle nach dem Tuning eine ähnliche Performanz. Um ein Modell für die Vorhersage der Kontexte auszuwählen, wird die Güte der Modelle zunächst anhand des Test-Datensatzes beurteilt, der nicht zur Modellentwicklung genutzt wurde und damit eine unverzerrte Schätzung der Performanz liefert.

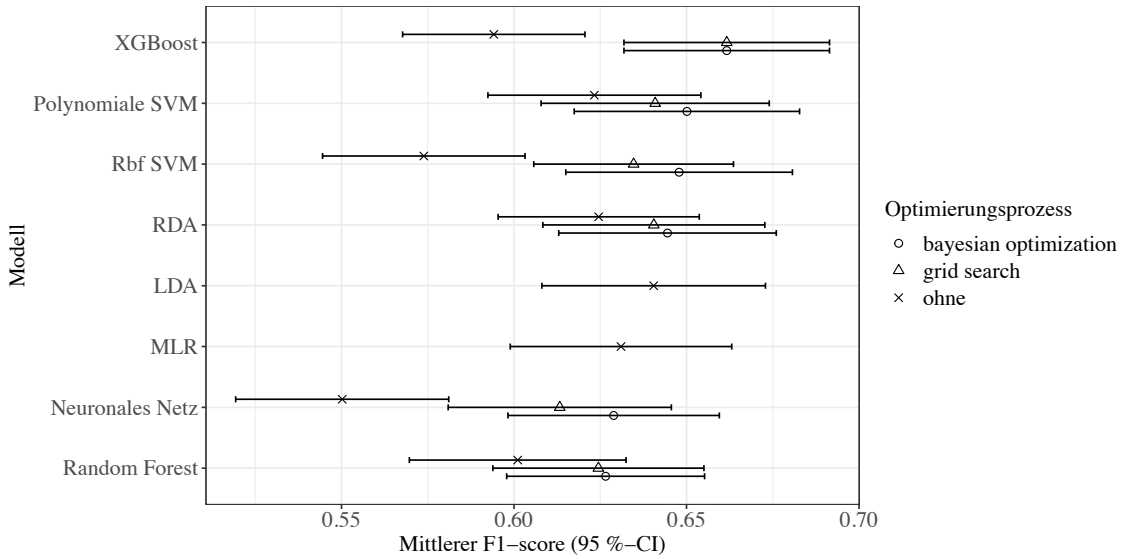


Abbildung 26: Vergleich des mittleren F_1-scores der Modelle nach Optimierung der Hyperparameter

9.3.3 Evaluation im Test-Datensatz

Für die Vorhersage im Test-Datensatz wurden die Modelle genutzt, die mithilfe des Trainingsdatensatzes entwickelt und optimiert wurden. Sofern die Hyperparameter des Modells optimiert wurden, wurden die Parameterkombinationen genutzt (siehe Anhang), die im Trainingsdatensatz die höchste Performanz gezeigt haben. Der auf der Vorhersage im Test-Datensatz basierende F_1-score, Precision und Recall sind in Tabelle 36 aufgeführt. Wie zu erwarten, zeigen alle Modelle im Test-Datensatz eine geringere Performanz im Vergleich zum kreuzvalidierten Ergebnis im Trainingsdatensatz. Einige Modelle (z. B. SVM mit radialer Basisfunktion oder Random Forest) schneiden allerdings deutlich schlechter ab als im Trainingsdatensatz und sind damit vermutlich für die Vorhersage neuer Daten unbrauchbar. Möglicherweise overfitten diese Modelle die Trainingsdaten und können damit schlecht auf neue Daten generalisieren. Den höchsten F_1-score zeigen das neuronale Netz und das XGBoost-

Modell. Hier wird auch deutlich, dass der F_1-score des neuronalen Netzes im Test-Datensatz im 95 %-Konfidenzintervall des F_1-score liegt, der mithilfe der Trainingsdaten berechnet wurde. Folglich scheint das neuronale Netz auch ausreichend auf Daten zu generalisieren, die nicht zur Schätzung der Modellparameter genutzt wurden.

Tabelle 36: F1-score, Precision und Recall der entwickelten Modelle im Test-Datensatz

Modell	F_1-score	Precision	Recall
Neuronales Netz	.614	.647	.597
XGBoost	.611	.639	.597
Multinomiale Logistische Regression	.592	.623	.578
Regularisierte Diskriminanzanalyse	.585	.620	.572
Support Vector Machine – Polynomialer Kernel	.584	.611	.572
Lineare Diskriminanzanalyse	.574	.597	.566
Support Vector Machine – Radiale Basisfunktion	.549	.564	.541
Random Forest	.517	.565	.509

Um genauer zu untersuchen, welche Fehler das neuronale Netz bei der Vorhersage der Kontextwahl im Test-Datensatz begeht, kann die Konfusionsmatrix in Tabelle 37 betrachtet werden.

Tabelle 37: Konfusionsmatrix für die Vorhersage der Kontextwahl im Test-Datensatz mithilfe des neuronalen Netzes

Vorhersage	Kontextwahl der Lernenden		
	Alltäglich	Besonders	Innerfachlich
Alltäglich	52	5	6
Besonders	2	8	3
Innerfachlich	1	5	6

Anmerkung: $N = 88$

Das neuronale Netz zeigt eine hohe Performanz bei der Vorhersage der Wahl von alltäglichen Kontexten. Ein hoher Anteil der Lernenden, die gemäß dem neuronalen Netz einen alltäglichen Kontext wählen, haben auch tatsächlich einen alltäglichen Kontext gewählt ($Precision_{All}$ = .83). Außerdem schlägt das neuronale Netz kaum besondere oder innerfachliche Kontexte vor, obwohl ein alltäglicher Kontext gewählt wurde ($Recall_{All}$ = .95). Lediglich bei 5 % der Lernenden wird fälschlicherweise ein besonderer oder innerfachlicher Kontext vorgeschlagen. Deutlich geringer ist die Performanz bei der Vorhersage von besonderen oder innerfachlichen Kontexten, was darauf hindeutet, dass die class imbalance nach wie vor ein Problem während der Konstruktionsphase des Modells darstellt. Das neuronale Netz liegt bei 61.5 % der Vorhersagen eines besonderen Kontexts richtig ($Precision_{Bes}$ = .615), d. h. 38.5 % der Lernenden haben eigentlich einen anderen als einen der besonderen Kontexte gewählt. Der Recall für die Klassifikation der besonderen Kontexte liegt mit .44 deutlich geringer als für die alltäglichen Kontexte. Weniger als die Hälfte der besonderen Kontexte in der Teststichprobe werden damit vom neuronalen Netz erkannt. Die Vorhersage des innerfachlichen Kontextes ist noch etwas unpräziser: Lediglich die Hälfte der Vorhersagen des innerfachlichen Kontextes sind korrekt ($Precision_{Inn}$ = .50). Hierbei ist auffällig, dass das neuronale Netz häufig einen innerfachlichen Kontext vorhersagt, obwohl tatsächlich einer der besonderen Kontexte gewählt wurde. Dieses Ergebnis deutet darauf hin, dass das neuronale Netz nicht hinreichend zwischen dem besonderen und dem innerfachlichen Kontext diskriminiert. Darüber hinaus schlägt das Modell häufig einen alltäglichen Kontext vor, obwohl tatsächlich der innerfachliche Kontext gewählt. Der Recall für den innerfachlichen Kontext liegt bei .40.

9.4 Überprüfung der Vorhersagegüte

Um die Vorhersagegüte des Modells an einem neuen Datensatz zu überprüfen, wurde eine Zusatzerhebung durchgeführt. Für die Zusatzerhebung wurde ein ähnliches quantitatives Studiendesign wie für die erste Teilstudie gewählt: Die Personenmerkmale der Lernenden wurden zunächst mithilfe von Leistungs- und Persönlichkeitstests erfasst. Hierbei wurden jedoch lediglich die Personenmerkmale erhoben, die für die Vorhersage der Kontextwahl benötigt werden. Anschließend wurden den Lernenden die gleichen Kontextbeschreibungen wie in der ersten Teilstudie vorgelegt, aus denen eine ausgewählt werden sollte. Anders als in der ersten Teilstudie mussten die Lernenden den beschriebenen Kontext allerdings nicht im Rahmen einer kontextualisierten Aufgabe bearbeiten. Das zentrale Ziel der Zusatzerhebung war es, die Vorhersage der Kontextwahl durch das Modell mit der tatsächlichen Kontextwahl der Lernenden zu vergleichen.

9.4.1 Beschreibung der Messinstrumente

Für die Vorhersage durch das Modell mussten das Vorwissen in Chemie, das individuelle Interesse an Chemie, das chemiebezogene Selbstkonzept sowie die Motive zur Wahl eines Kontextes erhoben werden. Zur Erfassung der Personenmerkmale wurden die gleichen Messinstrumente wie schon in der ersten Teilstudie verwendet. Hierbei wurde auf den Einsatz der Skala zur extrinsischen Motivation und zur Selbstwirksamkeitserwartung verzichtet, da sie für das Vorhersagemodell nicht benötigt werden.

Das Instrument zur Erfassung der Wahlmotive musste angepasst werden, um es zur Vorhersage der Kontextwahl nutzen zu können. Das ursprünglich genutzte Instrument hat die Wahlmotive bei der Kontextwahl retrospektiv erfasst. Um die Kontextwahl anhand der Wahlmotive vorherzusagen, ist es allerdings notwendig, die Wahlmotive schon vor der eigentlichen Kontextwahl zu erfassen. Da die Veränderung der Itemformulierung die psychometrische Qualität des Instruments maßgeblich beeinflussen kann, wurde davon abgesehen (z. B. Flake & Fried, 2020). Stattdessen mussten die Probanden und Probandinnen im veränderten Messinstrument zunächst anhand einer kurzen Beschreibung in einer hypothetischen Kontextwahl zwischen einem alltäglichen, besonderen und innerfachlichen Kontext wählen. Hierfür wurden adaptierte Beschreibungen der Kontexte *Unsere Verdauung*, *Die Folgen der chronischen Typ-A-Gastritis* und *Säuren im Labor* genutzt, die sich als merkmalsgetreu erwiesen haben. Anknüpfend an die Wahl zwischen den drei Kontexten, mussten die Items zur Erfassung der Wahlmotive beantwortet werden, die so in ihrer Formulierung unverändert blieben. Das so konzipierte Instrument ist im Anhang einzusehen.

9.4.2 Datenerhebung und Stichprobe

Die Zusatzerhebung wurde im Dezember 2021 im regulären Chemieunterricht durchgeführt. Unabhängig vom Stundenmodell der jeweiligen Schule standen in jeder Unterrichtsstunde 45 Minuten für die Befragung zur Verfügung. Im Gegensatz zur ersten Teilstudie konnte die Befragung vollständig im Präsenzunterricht mithilfe eines anonymisierten Paper-Pencil-Fragebogens durchgeführt werden. Die Testleitung erfolgte bei der gesamten Datenerhebung durch dieselbe Person.

Im ersten Abschnitt des Fragebogens wurden demografische Daten der Lernenden (Alter, Geschlecht, letzte Chemienote) zur Stichprobenbeschreibung erhoben. Darauf folgten die Items zur Erfassung der affektiven Variablen (Individuelles Interesse an Chemie, Selbstkonzept und Motive zur Wahl eines Kontextes). Danach mussten die Lernenden den Leistungstest zur Erfassung des Vorwissens in Chemie bearbeiten. Zuletzt mussten die Lernenden zwischen den fünf Kontextbeschreibungen wählen, die schon in der ersten Teilstudie genutzt wurden. Diese wurden nicht verändert.

In der Zusatzerhebung wurde so eine Gelegenheitsstichprobe von 116 Lernenden aus der Metropolregion Rhein-Ruhr untersucht. Die Schülerinnen und Schüler befanden sich im dritten Lernjahr in Chemie und stammten von zwei Gymnasien und einer Gesamtschule. Der Anteil der Gesamtschüler und Gesamtschülerinnen in der Stichprobe betrug damit 21.55 %. Zudem waren 56.03 % der Lernenden weiblich. Die letzte durchschnittliche Chemienote der Lernenden lag bei $M = 2.24$ ($SD = 0.97$), wobei diese Angabe nur auf Antworten von 113 der 116 Lernenden basiert. Nach einem listenweisen Fallausschluss liegen vollständige Datensätze von 88 Lernenden vor.

9.4.3 Ergebnisse und Diskussion

Bevor die Kontextwahl anhand der gemessenen Personenmerkmale vorhergesagt wurde, wurden die genutzten Messinstrumente hinsichtlich ihrer psychometrischen Qualität beurteilt.

Der Leistungstest zur Erfassung des Vorwissens in Chemie wurde erneut mithilfe eines eindimensionalen Rasch-Modells überprüft. Bislang finden sich in der Literatur keine Empfehlungen zum Umgang mit IRT-skalierten Personenparametern im maschinellen Lernen. Es ist jedoch anzunehmen, dass die Personenparameter für die neue Stichprobe auf Basis fixierter Itemparameter berechnet werden sollten, damit sie direkt vergleichbar zu den Personenparametern sind, die zur Konzeption des Modells genutzt wurden. Zunächst wurde ein Rasch-Modell mit frei geschätzten Itemparametern berechnet (Tab. 38). Das Rasch-Modell mit frei geschätzten Itemparametern verweist lediglich auf eine Aufgabe, deren wMNSQ- und korrespondierenden *t*-Werte nicht zum Rasch-Modell passen. Dieses wird zunächst im Modell belassen. Darüber hinaus liegt die WLE-Reliabilität in einem ähnlichen Bereich wie schon in der ersten Teilstudie. Im nächsten Schritt wurden die Itemparameter vollständig auf die in der ersten Teilstudie bestimmten Itemparameter fixiert. Nach der Fixierung der Itemparameter zeigen 11 der 57 Aufgaben eine unzureichende Passung zum Rasch-Modell. Außerdem sinkt die WLE-Reliabilität auf .57. Aufgrund der unzureichenden Passung der 11 Aufgaben werden die Itemparameter dieser Aufgaben frei geschätzt. Im Vergleich zum Modell mit vollständig fixierten Itemparametern sinkt die Anzahl der Items mit unzureichendem Fit im Rasch-Modell mit teilweisen fixierten Itemparametern auf ein Item. Da dieses Item jedoch sowohl in der ersten Teilstudie als auch in der Arbeit von Celik (2022) eine ausreichende Passung zum Rasch-Modell zeigte, wird es für die weitere Analyse nicht ausgeschlossen. Die WLE-Reliabilität in dem Modell mit teilweise fixierten Itemparametern ist etwas geringer als im frei geschätzten Modell, aber höher als im Modell mit vollständig fixierten Itemparametern.

Tabelle 38: Kennwerte des Rasch-Modells mit frei geschätzten, vollständig fixierten und teilweise fixierten Itemparametern für den Vorwissenstest (Zusatzerhebung)

Modellmerkmal	Kennwert (frei geschätzt)	Kennwert (vollständig fixiert)	Kennwert (teilweise fixiert)
$N_{Personen}$	116	116	116
N_{Items}	57	57	57
Parameterzahl	58	2	13
Deviance	2540.90	2737.37	2673.36
Varianz	.63	.51	.56
WLE-Reliabilität	.62	.57	.59
Itemparameter	-1.71 – 1.46	-1.71 – 1.46	-1.71 – 1.46
Personenparameter	-2.77 – 2.83	-2.08 – 2.81	-2.10 – 3.10
wMNSQ	0.81 – 1.43	0.52 – 1.93	0.54 – 1.45
t-Werte (wMNSQ)	-1.40 – 2.44	-2.64 – 2.56	-1.77 – 2.24

Die Gegenüberstellung der Item- und Personenparameter in der Wright-Map zeigt, dass die Items das Fähigkeitsspektrum der untersuchten Stichprobe im Wesentlichen gut abbilden (Abb. 27). Auch hier ist unklar, worauf die geringe WLE-Reliabilität zurückzuführen ist.

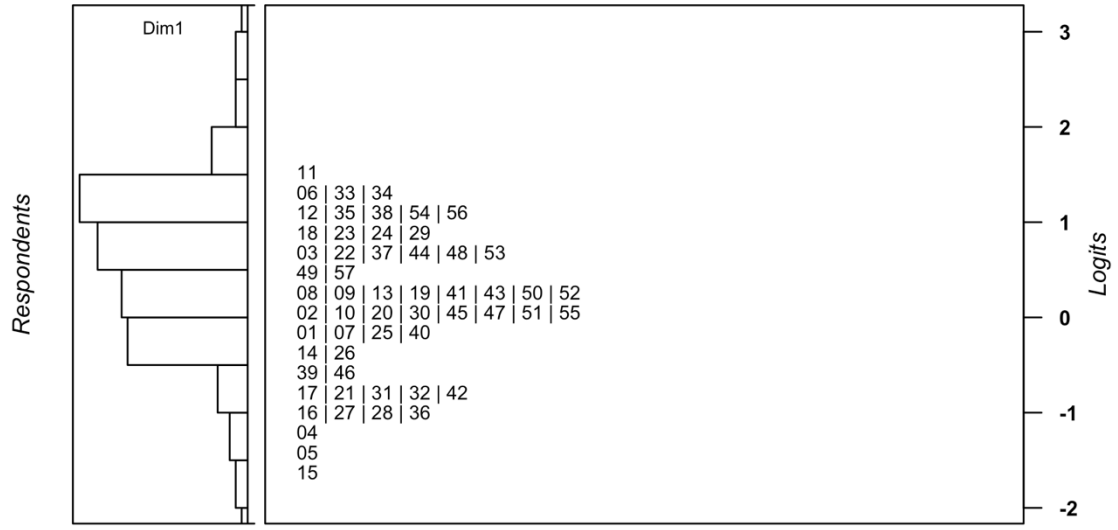


Abbildung 27: Wright-Map zum Vorwissenstest basierend auf dem Rasch-Modell mit teilweise fixierten Itemparametern (Zusatzerhebung)

Die Qualität der übrigen Testinstrumente wurde – analog zur ersten Teilstudie – mithilfe der CFA überprüft. Für das Instrument zur Erfassung des chemiebezogenen Selbstkonzepts wurde ein General-Faktor-Modell spezifiziert, welches die Daten gut beschreibt, $\chi^2(20) = 18.54$, $p = .552$, CFI = 1.00, TLI = 1.00, RMSEA = .000, SRMR = .030. Die Reliabilität der Skala zur Erfassung des Selbstkonzeptes liegt in einem herausragenden Bereich ($\alpha = .92$, $\omega = .92$).

Da auf die Skala zur extrinsischen Motivation verzichtet wurde, wird für das individuelle Interesse an Chemie ein Drei-Faktor-Modell mit korrelierten Faktoren in der CFA definiert. Das Modell beschreibt die Daten adäquat, wie die Modellprüfung zeigt, $\chi^2(227) = 549.12$, $p < .001$, CFI = .952, TLI = .944, RMSEA = .067, SRMR = .070. Wie schon in der ersten Teilstudie, liegen die Reliabilitäten für die Skala zum individuellen Interesse an Chemie ($\alpha = .95$, $\omega = .95$), zum themenspezifischen ($\alpha = .83$, $\omega = .87$) und inhaltsfeldspezifischen Interesse ($\alpha = .88$, $\omega = .89$) in einem guten Bereich. Die marginalen Unterschiede in den beiden Reliabilitätsmaßen α und ω für die Skala zum themenspezifischen Interesse sind vermutlich auf die geringe Faktorladung ($\lambda = .35$) des Items zum Interesse an besonderen Kontexten zurückzuführen (Raykov & Marcoulides, 2015).

Zur Untersuchung der adaptierten Skala zur Erfassung der Wahlmotive wurde das gleiche Modell wie in der ersten Teilstudie spezifiziert und hinsichtlich der Modellpassung überprüft. Der χ^2-Wert und die deskriptiven Statistiken zur Modellpassung verweisen auf eine inakzeptable Modellpassung, $\chi^2(32) = 80.399$, $p < .001$, CFI = .873, TLI = .821, RMSEA = .118, SRMR = .128. Weiterhin zeigt ein Item zur Erfassung des Wahlmotivs des persönlichen Bezugs eine negative Residualvarianz. Eine Inspektion der Modifikationsindizes impliziert die Modellierung der Fehlerkovarianz zwischen den Items *per3* und *per4*. Durch die Prüfung der Itemformulierungen zeigt sich, dass die beiden Items – mit Ausnahme der Verneinung – identisch formuliert sind. Folglich kann angenommen werden, dass es sich um einen Methodeneffekt handelt, der durch die Modellierung der Fehlerkovarianz im Modell berücksichtigt werden kann (Schermelleh-Engel & Gäde, 2020). Die Aufnahme der Fehlerkovarianz zwischen den Items *per3* und *per4* führt zu einer geringfügigen Verbesserung des Modellfits, $\chi^2(31) = 67.469$, $p < .001$, CFI = .908, TLI = .867, RMSEA = .104, SRMR = .102. Dieser kann möglicherweise auf der Grundlage des CFI als gerade noch ausreichend angesehen werden. Die Reliabilitäten für die Skalen zum Wahlmotiv *Interesse und Neugier* ($\alpha = .77$, $\omega = .79$) sowie *Überraschungsmoment* ($\alpha = .85$, $\omega = .85$) liegen in einem guten Bereich. Allerdings zeigt die Skala zur Erfassung des persönlichen Bezugs eine geringere Reliabilität ($\alpha = .61$, $\omega = .59$) als in der ersten Teilstudie.

Zur Vorhersage der Kontextwahl in der Zusatzerhebung wurden die durch das Rasch-Modell skalierten Personenparameter, der Skalenmittelwert als Schätzung der individuellen Merkmalsausprägung der affektiven Merkmale sowie das neuronale Netz genutzt. Die Konfusionsmatrix ist in Tabelle 39 dargestellt.

Tabelle 39: Konfusionsmatrix zur Vorhersage der Kontextwahl in der Zusatzerhebung mithilfe des neuronalen Netzes

Vorhersage	Kontextwahl der Lernenden		
	Alltäglich	Besonders	Innerfachlich
Alltäglich	25	13	4
Besonders	4	11	0
Innerfachlich	7	13	11

Anmerkung: $N = 88$

Das neuronale Netz zeigt im neuen Datensatz (F_1-score = .51, Precision = .56, Recall = .58) eine geringere Performanz als in der Teststichprobe (F_1-score = .614, Precision = .65, Recall = .60). Die Vorhersage der alltäglichen Kontexte gelingt weniger gut als in der Teststichprobe. Für die alltäglichen Kontexte ist die Vorhersage mit $Precision_{All} = .60$ unpräziser. Auffällig ist, dass das neuronale Netz häufig einen alltäglichen Kontext vorschlägt, obwohl tatsächlich einer der besonderen Kontexte gewählt wurde. Dieser Fehler konnte im Testdatensatz nicht beobachtet werden. Außerdem wird häufiger ein besonderer oder innerfachlicher Kontext vorhergesagt, obwohl einer der alltäglichen Kontexte gewählt wurde ($Recall_{All} = .69$). Die Vorhersage der besonderen Kontexte ist etwas präziser als im Testdatensatz ($Precision_{Bes} = .73$). Das neuronale Netz liegt bei der Vorhersage der besonderen Kontexte nur in 26.7 % der Fälle nicht richtig und schlägt fälschlicherweise einen alltäglichen Kontext vor. Allerdings scheint das neuronale Netz kaum noch in der Lage zu sein, die Fälle zu identifizieren, bei denen tatsächlich ein besonderer Kontext gewählt wurde ($Recall_{Bes} = .30$). Hier wird häufig ein alltäglicher oder innerfachlicher Kontext vorgeschlagen. Für die Vorhersage des innerfachlichen Kontexts zeigt sich ein entgegengesetztes Bild: Das neuronale Netz ist sehr unpräzise bei der Vorhersage des innerfachlichen Kontextes. Wenn ein innerfachlicher Kontext vorgeschlagen wird, liegt das Modell nur in 35.5 % der Fälle richtig ($Precision_{Inn} = .36$). Allerdings werden 73.3 % der Lernenden, die einen innerfachlichen Kontext gewählt haben, durch das Modell korrekt identifiziert. Für lediglich 26.7 % der Lernenden wird ein alltäglicher Kontext vorgeschlagen, obwohl tatsächlich ein innerfachlicher Kontext gewählt wurde.

Mögliche Ursachen für die geringere Performanz sind vielfältig: Zum einen unterscheidet sich die Kontextwahl in der vorliegenden Stichprobe von der Stichprobe, die zur Konzeption des neuronalen Netzes genutzt wurde. Offensichtlich haben die Lernenden aus dieser Stichprobe sehr viel seltener einen alltäglichen Kontext (40.90 %) gewählt als die Lernenden der Trainingsstichprobe. Zudem wurde der besondere Kontext mit 42.05 % häufiger gewählt. Der innerfachliche Kontext wurde in einem vergleichbaren Umfang wie in der ersten Teilstudie gewählt (17.05 %). Möglicherweise ist das neuronale Netz zu sehr an die Kontextwahl aus

der Trainingsstichprobe angepasst, so dass es Schwierigkeiten bei der Vorhersage neuer Daten hat. Zum anderen musste das Messinstrument zur Erfassung der Wahlmotive für die Vorhersage der Kontextwahl angepasst werden. Es ist nicht auszuschließen, dass die Anpassung des Messinstruments die Vorhersagegüte des Modells negativ beeinflusst. Insbesondere vor dem Hintergrund fragwürdiger Modellpassung und der geringeren Reliabilität der Skala zur Erfassung des Wahlmotivs des persönlichen Bezugs.

9.5 Implikationen und Limitationen

Bislang existieren kaum Forschungsansätze, bei denen ML-Algorithmen in Verbindung mit psychometrisch gemessenen Merkmalen zur Konstruktion eines Vorhersagemodells genutzt wurden. Die Ergebnisse der vorliegenden Arbeit deuten an, dass sich maschinelles Lernen prinzipiell zur Vorhersage der Kontextwahl anhand von psychometrisch gemessenen Personenmerkmalen eignet. Komplexere ML-Algorithmen liefern hierbei eine geringfügig bessere Performanz als traditionelle Ansätze zur Klassifikation, wie eine multinomiale logistische Regression oder eine Diskriminanzanalyse. Das trainierte neuronale Netz zeigt insbesondere bei der Vorhersage der alltäglichen Kontexte eine hinreichende Vorhersagegüte, wie die Ergebnisse in der Teststichprobe und der Zusatzerhebung zeigen. Im Vergleich zur Vorhersage der besonderen oder des innerfachlichen Kontextes werden hier weniger Fehler durch das Modell begangen. Dieses Ergebnis ist nicht überraschend, da die alltäglichen Kontexte am häufigsten von den Lernenden gewählt wurden und demnach mehr Beispiele zur Modellierung des Zusammenhangs zur Verfügung standen. Für die besonderen und den innerfachlichen Kontext standen vergleichsweise wenige Beispiele zur Verfügung, die zur Entwicklung des Modells nutzbar waren. Hier zeigen die Ergebnisse, dass das neuronale Netz nur eingeschränkt zwischen besonderen und dem innerfachlichen Kontext diskriminiert. Dieser Eindruck kann besonders aufgrund der Ergebnisse der Zusatzerhebung bestätigt werden. Das neuronale Netz schneidet hier schlechter ab als in der Teststichprobe.

Vor dem Hintergrund dieser Ergebnisse ist zwangsläufig zu diskutieren, welche Performanz von einem derartigen Modell überhaupt erwartbar ist. Da das Modell latente Merkmale zur Vorhersage der Kontextwahl nutzt, die mithilfe von psychometrischen Instrumenten gemessen wurden, sind diese stärker messfehlerbehaftet als manifeste Merkmale. Messfehler können dazu führen, dass ML-Algorithmen Zusammenhänge nur unzureichend modellieren (Jacobucci & Grimm, 2020). Weiterhin kann angenommen werden, dass die Wahl einer (kontextualisierten) Aufgabe nicht nur durch personenbezogene Merkmale determiniert wird. Vermutlich ist vielmehr ein reziprokes Verhältnis zwischen personenbezogenen, aufgabenbezogenen und situativen Faktoren entscheidend für die Wahl einer Aufgabe. Da im Modell nur eine begrenzte Auswahl personenbezogener Faktoren berücksichtigt wurde, scheint die Performanz in einem zufriedenstellenden Bereich zu liegen, insbesondere da bislang kaum vergleichbare empirische Ergebnisse vorliegen, die als Vergleichsmaßstab dienen könnten.

Zusammenfassend muss damit trotzdem festgehalten werden, dass die Vorhersage durch das neuronale Netz nicht vollständig mit der tatsächlichen Kontextwahl übereinstimmt. Bei der weiterführenden Untersuchung muss deshalb berücksichtigt werden, dass einige Lernende möglicherweise einen Kontext zugewiesen bekommen, der nicht zu ihren Personenmerkmalen passt.

10 Teilstudie II: Effektivität interessenbasierter Differenzierung

Das zentrale Ziel der zweiten Teilstudie war die Beantwortung der dritten Forschungsfrage mithilfe der in Kapitel 6 beschriebenen experimentellen Studie. Dabei sollte insbesondere untersucht werden, welche Bedeutung die Passung zwischen Kontext und den individuellen Merkmalen der Lernenden sowie die selbstständige Wahl einer kontextualisierten Aufgabe auf den Erfolg einer Lernsituation im Fach Chemie hat.

10.1 Weiterentwicklung des Lernmaterials

Ausgehend von dem für die erste Teilstudie entwickelten Lernmaterial wurden insgesamt drei unterschiedliche Lerneinheiten entwickelt, die hinsichtlich des Kontextmerkmals variiert wurden. Analog zur ersten Teilstudie wurde das Lernmaterial zur digitalen Bearbeitung an Tabletcomputern entwickelt. Da iBooks Author zum Zeitpunkt der Materialentwicklung nicht länger verfügbar war, wurde das Lernmaterial stattdessen mithilfe der Software PowerPoint konstruiert.

Zur Berücksichtigung von Neuigkeitseffekten während der experimentellen Studie sollte das Lernmaterial aus der ersten Teilstudie um je zwei Teilaufgaben erweitert werden. Ziel war die Entwicklung einer kontextorientierten Lerneinheit mit drei aufeinander aufbauenden Teilaufgaben. Basierend auf einer Lehrplananalyse wurden verschiedene Kompetenzerwartungen für das Inhaltsfeld saure und alkalische Lösungen formuliert, die inhaltlich aufeinander aufbauen und einen möglichen Weg der Kompetenzentwicklung im Kompetenzbereich Umgang mit Fachwissen beschreiben (Tab. 40).

Als Einstieg in die kontextorientierte Lerneinheit wurde der Nachweis von sauren und alkalischen Lösungen mithilfe von pH-Indikatoren ausgewählt. Diese Kompetenz bildet durch die Betrachtung auf phänomenologischer Ebene eine zentrale Grundlage für die weitere Konzeptentwicklung. Es kann davon ausgegangen werden, dass ein Großteil der Lernenden im dritten Lernjahr über diese Kompetenz verfügt. Somit soll durch die erste Teilaufgabe im Wesentlichen das Vorwissen der Lernenden aktiviert werden, um einen geeigneten Einstieg in die zweite Teilaufgabe vorzubereiten. Anknüpfend an die erste Teilaufgabe soll die Kompetenzerwartung gefördert werden, die bereits Grundlage für die Konstruktion der Kontextaufgabe aus Teilstudie I war (siehe Kapitel 8.1). Bei der Durchsicht der Antworten im Lernmaterial der ersten Teilstudie hat sich gezeigt, dass manche Lernenden besondere Schwierigkeiten beim Schritt der Hydratisierung der Wasserstoff-Ionen hatten. Um das Anforderungsniveau der zweiten Teilaufgabe zu reduzieren, wurde die Kompetenzerwartung hinsichtlich der Hydratisierung der Ionen angepasst. Anknüpfend an die zweite Kompetenzerwartung sollen in der dritten Teilaufgabe die Eigenschaften alkalischer Lösungen auf submikroskopischer Ebene erklärt werden. Dies ermöglicht anschließend die Einführung der Neutralisationsreaktion als Reaktion zwischen Wasserstoff- und Hydroxid-Ionen.

Tabelle 40: Die festgelegten Kompetenzerwartungen für die drei sequenzierten Teilaufgaben des Lernmaterials für Teilstudie II

Teilaufgabe	Kompetenzerwartung
	Die Lernenden können…
1	…saure und alkalische Lösungen mithilfe von Indikatoren nachweisen
2	…Säuren als Stoffe beschreiben, deren wässrige Lösungen Wasserstoff-Ionen enthalten und die Eigenschaften von sauren Lösungen mit dem Vorhandensein von Wasserstoff-Ionen erklären
3	…Basen als Stoffe beschreiben, deren wässrige Lösungen Hydroxid-Ionen enthalten und die Eigenschaften von alkalischen Lösungen mit dem Vorhandensein von Hydroxid-Ionen erklären
	…die Neutralisationsreaktion als Reaktion zwischen Wasserstoff- und Hydroxid-Ionen beschreiben und erklären.

Für die zweite Teilstudie mussten ein alltäglicher, ein besonderer und ein innerfachlicher Kontext ausgewählt werden. Sowohl das Inhaltsfeld saure und alkalische Lösungen als auch das Themenfeld Körper und Krankheiten haben sich in der ersten Teilstudie als geeigneter Rahmen zur Konstruktion des kontextorientierten Lernmaterials erwiesen. Die Ergebnisse der ersten Teilstudie deuten darauf hin, dass die entwickelten Kontexte den angenommenen Kontextmerkmalen entsprechen und demnach zur systematischen Variation des Kontextmerkmals geeignet sind. Als innerfachlicher Kontext wurde der Kontext *Säuren im Labor* aus der ersten Teilstudie übernommen. Da für die erste Teilstudie zwei alltägliche und zwei besondere Kontexte entwickelt wurden, mussten für die weitere Materialentwicklung ein alltäglicher und ein besonderer Kontext ausgewählt werden. Hierbei wurden jeweils die Kontexte für die weitere Materialentwicklung genutzt, die am häufigsten von den Lernenden ausgewählt wurden und somit vermutlich am ansprechendsten für einen Großteil der Schülerinnen und Schüler sind. Folglich wurden der alltägliche Kontext *Warum Zähne putzen so wichtig ist* und der besondere Kontext *Schädigung der Knochen durch Übersäuerung* für das Lernmaterial der zweiten Teilstudie genutzt.

Dieses Material wurde erneut in Anlehnung an die ChiK-Unterrichtsphasen strukturiert und nach den Prinzipien des CTML gestaltet (siehe Kapitel 8.1). Gemäß dem *Segmenting Principle* wurde das Lernmaterial in kleinere Sinnabschnitte unterteilt, die der Einführung, der Kontextbeschreibung und den drei Teilaufgaben entsprachen. Die drei Teilaufgaben wurden wiederum in eine Aneignungsphase, Übungsphase und Erklärungsphase sowie die abschließende Befragungsphase unterteilt, die jeweils durch ein Piktogramm gekennzeichnet wurden.

Der Name der Teilaufgabe sowie das Piktogramm der Lernphase sind im Lernmaterial in der oberen rechten Ecke dargestellt, um das Lernmaterial stärker zu strukturieren (Abbildung 28).

In der Einführungsphase wurden die Lernenden über die Zielsetzung und die Struktur des Lernmaterials instruiert. Das Ziel jeder Phase wurde hier explizit dargestellt. Darüber hinaus wurden die Lernenden in der Einführungsphase auf den Einsatz des Piktogramms einer Comicfigur hingewiesen, welches wichtige Textstellen markiert (siehe *Signaling Principle*).

Wie in der ersten Teilstudie wurden die Lernenden durch eine Kontextbeschreibung in den Kontext eingeführt (Phase der Begegnung, Parchmann et al., 2008). Die Kontextbeschreibungen wurden nahezu unverändert aus der ersten Teilstudie übernommen. Lediglich auf den letzten Textblock zur Verwendung im Unterricht wurde verzichtet, da sich dieser auf die Teilaufgabe aus Teilstudie I bezog. Die Konstruktion der Kontextbeschreibungen wird in Kapitel 8.1 näher erläutert.

Danach folgten nacheinander die Teilaufgaben 1 bis 3, welche die in Tabelle 37 beschriebenen Kompetenzerwartungen adressierten. Die Teilaufgaben 1 und 3 wurden für die zweite Teilstudie neu konzipiert, wohingegen die zweite Teilaufgabe auf dem Lernmaterial der ersten Teilstudie basiert. In jeder der Teilaufgaben wird ein neuer Aspekt des Kontextes aufgegriffen, so dass der Kontext als storyline die gesamte Lerneinheit umrahmt (Nentwig et al., 2007). Die Bearbeitung dieser drei Teilaufgaben ähnelt damit konzeptionell der Phase der Erarbeitung von ChiK, in der fachliche Inhalte in Kontexten mit steigendem fachlichem Niveau erarbeitet werden. Die wiederholte Dekontextualisierung und Rekontextualisierung werden innerhalb der Teilaufgaben durch die Aneignungs- und Übungsphase noch stärker akzentuiert.

Die Aneignungsphase wurde ähnlich wie in dem Lernmaterial der ersten Teilstudie gestaltet. Zu Beginn jeder Aneignungsphase wird eine aus dem jeweiligen Kontext resultierende fachliche Problemstellung aufgeworfen (Abb. 28).

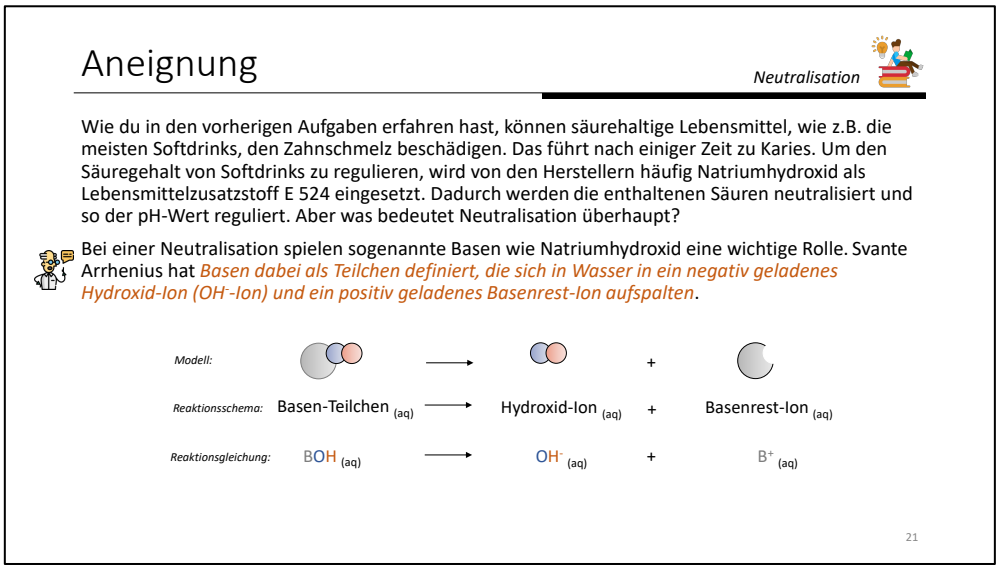


Abbildung 28: Ausschnitt aus der Aneignungsphase der dritten Teilaufgabe mit alltäglichem Kontext

Die fachliche Problemstellung aller Teilaufgaben war identisch, allerdings hat sich die Herleitung der Problemstellung in Abhängigkeit vom Kontext unterschieden. Im Verlauf der Aneignungsphase wurde der fachliche Inhalt zur Bearbeitung des Problems von den Lernenden erarbeitet. Im Gegensatz zum Lernmaterial aus der ersten Teilstudie wurde vor allem das *Signaling Principle* durch ein Icon berücksichtigt, auf dem ein Forscher als Comicfigur dargestellt war. Das Icon wurde zur Markierung wichtiger Textstellen genutzt, die zudem farbig hinterlegt wurden. Im Übrigen wurden alle weiteren Gestaltungsmerkmale der CTML wie im Lernmaterial der Teilstudie I umgesetzt (z. B. der Einsatz von Repräsentationen). Die Aneignungsphase endete mit der Beantwortung der eingangs aufgeworfenen Problemstellung auf Kontextebene.

Anschließend sollte das erarbeitete Fachwissen in der Übungsphase gefestigt werden. Mit Ausnahme der ersten Teilaufgabe wurden in alle Teilaufgaben drei Übungsaufgaben integriert, die basierend auf dem Konstruktionsdesign des ESNaS-Kompetenzstrukturmodells entwickelt wurden (Tabelle 7). Für die erste Teilaufgabe war die Konstruktion einer Aufgabe, in der ein Konzept angewandt werden sollte, nicht zielführend, da in der ersten Teilaufgabe kein Konzept adressiert wird. Demzufolge wurden für die erste Teilaufgabe nur Übungsaufgaben genutzt, bei denen Fakten und Zusammenhänge wiedergegeben werden mussten. Bei den Teilaufgaben 2 und 3 musste jeweils in der letzten Übungsaufgabe das jeweilige Konzept angewandt werden. Die jeweils dritten Übungsaufgaben sind in diesem Zusammenhang als Brückenaufgaben zu verstehen, welche die Dekontextualisierung und Anwendung des erworbenen Wissens fördern sollen (Parchmann et al., 2008).

Alle Übungsaufgaben bestehen aus einem Aufgabenstamm und einer Aufgabenstellung. In dem Aufgabenstamm werden relevante Informationen aus der Aneignungsphase und dem jeweiligen Kontext wiederholt. Die Aufgabenstellung war nahezu identisch und wurde nur im Hinblick auf den zugrundeliegenden Kontext angepasst. Eine beispielhafte Aufgabe ist in Abbildung 29 dargestellt.

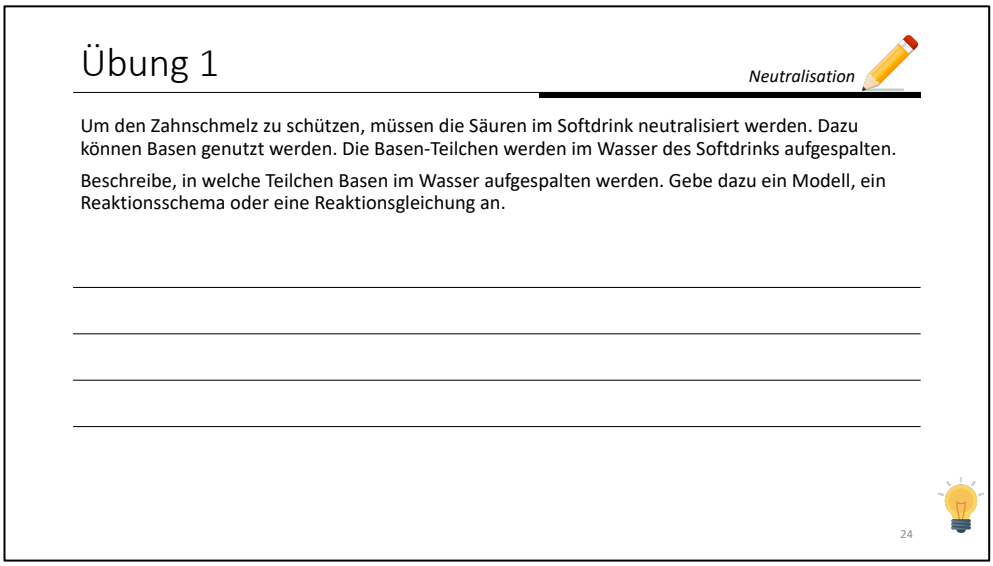


Abbildung 29: Beispiel für eine Übungsaufgabe aus der dritten Teilaufgabe im Lernmaterial mit alltäglichem Kontext

Wie schon im Lernmaterial der ersten Teilstudie konnten die Lernenden eine Erklärung für die Aufgabe einsehen, die im Sinne einer Musterlösung zu interpretieren ist. Dafür musste das Icon mit der Glühbirne angeklickt werden, das auf die korrespondierende Erklärung der Aufgabe verweist.

10.2 Beschreibung der Messinstrumente

10.2.1 Erfassung der Kontrollvariablen

Als relevante Kontrollvariablen wurden die demografischen Daten, das Vorwissen in Chemie, das Interesse an Chemie und die Motivation im Fach Chemie, das chemiebezogene Selbstkonzept sowie die Motive zur Wahl eines Kontextes erfasst. Die Erfassung dieser Variablen diente einerseits zur Kontrolle personenbezogener Variablen in der experimentellen Studie. Andererseits wurden das Vorwissen, das Interesse an Chemie, das chemiebezogene Selbstkonzept und die Wahlmotive zur Vorhersage der Kontextwahl durch das entwickelte Vorhersagemodell genutzt. Abgesehen von dem Messinstrument zur Erfassung der Wahlmotive wurden alle Instrumente unverändert aus der ersten Teilstudie übernommen.

Das Instrument zur Erfassung der Wahlmotive wurde für die zweite Teilstudie angepasst, da die Skala zum Wahlmotiv persönlicher Bezug in der Zusatzerhebung verbesserungswürdige Reliabilitätskennwerte aufwies. Auch hier wurde davon abgesehen, die Itemformulierung zu verändern, da sich das ursprüngliche Instrument von van Vorst und Aydogmus (2021) durch zufriedenstellende psychometrische Kennwerte auszeichnete. Stattdessen wurden für das überarbeitete Instrument adaptierte Kontextbeschreibungen entwickelt, die auf den Kontexten basieren, die auch im Rahmen dieser Studie genutzt wurden. Bei der Adaption wurde der Wortlaut der Kontextbeschreibungen angepasst, ohne den Inhalt zu verändern. Darüber hinaus wurden der Name und das Bild in der Kontextbeschreibung angepasst. So sollte sichergestellt werden, dass das Instrument die im weiteren Verlauf der Studie genutzten Kontexte valide abbildet, aber zugleich die Kontextwahl für einen Teil der Lernenden nicht redundant ist. Das adaptierte Messinstrument ist im Anhang einsehbar.

10.2.2 Erfassung der abhängigen Variablen

Das zentrale Ziel der zweiten Teilstudie war die Identifikation von Veränderungen im Fachwissen, der Zufriedenheit, dem situationalen Interesse und der kognitiven Belastung durch die Manipulation der Kontextwahl und der Passung zwischen Kontext und Lernenden. Geeignete Instrumente zur Erfassung der Zufriedenheit, des situationalen Interesses und der kognitiven Belastung wurden bereits in der ersten Teilstudie genutzt. Die beiden letzten Instrumente wurden unverändert aus Teilstudie I übernommen, weshalb auf deren Beschreibung an dieser Stelle verzichtet wird (siehe Kapitel 8.2.2). Für die Verwendung in der zweiten Teilstudie wurde allerdings das Instrument zur Erfassung der Zufriedenheit angepasst, um das Konstrukt der Zufriedenheit weiter auszudifferenzieren. Ein geeigneter Fachwissenstest zur Erfassung des Fachwissenszuwachses musste für die zweite Teilstudie neu entwickelt bzw. adaptiert werden. Dadurch sollte sichergestellt werden, dass der Test das während der kontextualisierten Lerneinheit erworbene Wissen inhaltsvalide abbildet. Der größere Umfang der kontextualisierten Lerneinheit der zweiten Teilstudie im Vergleich zur ersten Teilstudie erforderte auch die Konstruktion eines umfangreicheren Tests.

Fachwissen im Inhaltsfeld saure und alkalische Lösungen: Ausgangspunkt für die Entwicklung des Fachwissenstests waren die festgelegten Kompetenzerwartungen für die kontextualisierte Lerneinheit aus Tabelle 40. Testitems zum Inhaltsfeld saure und alkalische Lösungen wurden u. a. von Habig (2017) und Fechner (2009) entwickelt, so dass vorliegende Testitems für die Konzeption dieses Tests adaptiert werden konnten. Darüber hinaus wurde der Itempool einer zu diesem Zweck durchgeführten Masterarbeit genutzt (Schwantes, 2021).

Für jede Teilaufgabe wurde eine definierte Anzahl von Testitems erstellt bzw. adaptiert, die den Umfang der jeweiligen Kompetenzerwartungen repräsentierte. So wurden für die erste

Teilaufgabe zwei Testitems, für die zweite Teilaufgabe vier Testitems und für die dritte Teilaufgabe sechs Testitems generiert, welche die jeweiligen Inhalte der Teilaufgabe bzw. die jeweiligen Kompetenzerwartungen valide abbilden sollten. Insgesamt besteht der Test so aus 12 Testitems.

Um das Kompetenzspektrum der Lernenden hinreichend abzubilden, wurden die Test-items gezielt im Hinblick auf die schwierigkeitserzeugenden Merkmale *Komplexität* und *Kognitive Prozesse* des ESNaS-Kompetenzstrukturmodells variiert (Tabelle 7). Analog zur Erstellung der Kontextaufgaben wurde hierbei zwischen folgenden Niveaustufen unterschieden: Fakten wiedergeben (Niveau 1), Zusammenhänge wiedergeben (Niveau 2) und Konzepte anwenden (Niveau 3). In Tabelle 41 ist die Anzahl der entwickelten Testitems pro Teilaufgabe und Niveaustufe dargestellt.

Tabelle 41: Anzahl der entwickelten Testitems pro Teilaufgabe und Niveaustufe für den Fachwissenstest

Teilaufgabe	Niveaustufe		
	Niveau 1	Niveau 2	Niveau 3
1	1	1	
2	1	2	1
3	2	3	1

Anmerkung: Zu der ausgegrauten Zelle ist die Konstruktion von Aufgaben nicht sinnvoll, da Fakten nicht angewandt werden können

Da die meisten Lernenden vermutlich ein mittleres Kompetenzniveau aufweisen, wurden die meisten Testitems auf der zweiten Niveaustufe entwickelt, um das Kompetenzspektrum der Zielpopulation abzubilden.

Die Testitems wurden im Multiple-Choice-Single-Select-Format entwickelt. Jedes Test-item enthält neben dem richtigen Attraktor drei weitere Distraktoren. Zur Formulierung der Distraktoren wurden gängige Schülerfehlvorstellungen genutzt. Die Testitems der dritten Niveaustufe wurden gezielt in einen Kontext eingebettet, um zu überprüfen, wie transferfähig das kontextualisiert erworbene Wissen ist. Hier wurden Kontexte genutzt, die im Verlauf der Untersuchung bisher unberücksichtigt blieben. Der Kontext wurde im Wesentlichen in den Stamm des Items integriert. Die Items sind im Anhang einzusehen.

Zufriedenheit: Durch die Analyse der Daten aus der ersten Teilstudie hat sich gezeigt, dass das Konstrukt der Zufriedenheit vom situationalen Interesse separierbar ist. Um das Konstrukt der Zufriedenheit weiter auszudifferenzieren und weitere Informationen über die Ursache der

aufgabenbezogenen Zufriedenheit zu erhalten, wird das Konstrukt im Rahmen der zweiten Teilstudie durch zwei Skalen operationalisiert. Durch die erste Skala soll die Zufriedenheit mit dem Inhaltsfeld erhoben werden, in welches die Aufgaben eingebettet sind. Möglicherweise ist ein Teil der Lernenden mit der Aufgabe zufrieden, weil das Inhaltsfeld und nicht der Kontext ansprechend ist. Mithilfe der zweiten Skala wird die Zufriedenheit mit dem Kontext der Aufgabe erhoben.

Zur Konstruktion der Skalen wurden die Itemformulierungen aus der ersten Teilstudie unverändert übernommen, da die Skala eine zufriedenstellende psychometrische Qualität zeigte. Um die Zufriedenheit am Inhaltsfeld valide zu erfassen, wurde die Bezeichnung *chemischer Inhalt* zur Umschreibung des Inhaltsfeldes in die Items zur Erfassung der inhaltsfeldspezifischen Zufriedenheit integriert. In der Skala zur kontextbezogenen Zufriedenheit wurde der Kontextbegriff wiederum erneut mit dem Begriff *Thema* umschrieben. Erneut wurde wieder der Titel des bearbeiteten Kontexts hinter der Bezeichnung *Thema* eingeblendet, um Inhaltsvalidität zu gewährleisten. Darüber hinaus waren die Items identisch formuliert (Tab. 42). Alle Items waren auf einer vierstufigen Likert-Skala von *stimmt gar nicht* (1) bis *stimmt völlig* (4) zu beantworten.

Tabelle 42: Überblick über die eingesetzten Skalen zur Erfassung der Zufriedenheit

Skala	Itemanzahl	Beispiel
Zufriedenheit - Inhaltsfeld (*zufi*)	4	Ich war mit dem chemischen Inhalt (saure und alkalische Lösungen) meiner Aufgabe zufrieden.
Zufriedenheit – Kontext (*zufk*)	4	Ich war mit dem Thema (Zähneputzen/Schädigung der Knochen/Chemielabor) meiner Aufgabe zufrieden.

Um weitere Informationen über die Zufriedenheit mit dem Kontext zu erhalten, wurde ein weiteres Item entwickelt, das am Ende der Lerneinheit eingesetzt wurde (vgl. Kapitel 10.3). Hierfür wurden den Lernenden drei verkürzte Beschreibungen der Kontexte vorgelegt, die in dieser Studie genutzt wurden. Daraufhin wurden die Lernenden mit einem geschlossenen Item gefragt, ob sie den Kontext erneut bearbeiten würden, wobei der Kontextbegriff erneut mit dem Begriff Thema umschrieben wurde. Die Antwortoptionen reichten hierbei von *Ja, ich würde das gleiche Thema bearbeiten (1)* über *Nein, ich würde ein anderes Thema bearbeiten (2)* bis *Nein, ich würde am liebsten keines der Themen bearbeiten (3)*.

Wenn die erste Option gewählt wurde, mussten die Lernenden in einem offenen Item beschreiben, was ihnen am bearbeiteten Thema gut gefallen hat. Bei der zweiten Option erschien

ein weiteres geschlossenes Item, bei dem die Lernenden angeben sollten, welchen Kontext sie stattdessen gerne bearbeitet hätten. Die möglichen Antworten entsprachen hierbei den Titeln der Kontexte: *Warum Zähneputzen so wichtig ist (1)*, *Schädigung der Knochen durch Übersäuerung (2)* und *Säuren im Labor (3)*. Sowohl bei der Auswahl der zweiten und dritten Option wurden die Lernenden mithilfe eines weiteren offenen Items hinsichtlich ihrer Gründe für die Wahl eines anderen Kontextes befragt.

Die offenen Antworten der Lernenden wurden mithilfe einer qualitativen Inhaltsanalyse nach Mayring (2022) ausgewertet. Hierfür wurden zunächst deduktiv Kategorien aus der Untersuchung von van Vorst und Aydogmus (2021) abgeleitet. Die Autorinnen haben in ihrer Studie sowohl kognitive als auch affektiv-motivationale Motive für die Zufriedenheit identifiziert. Demzufolge wurden die Kategorien als Ausgangspunkt für die induktive Ausarbeitung des Kategoriensystems anhand der Antworten der Lernenden genutzt. Nach einer ersten Durchsicht des Materials zeigte sich, dass einige Antworten sehr unspezifisch waren, so dass es notwendig erschien, eine dritte Oberkategorie zu bilden. Neben kognitiven und affektiv-motivationalen Motiven für die Zufriedenheit wurde deshalb eine dritte Oberkategorie *unspezifische Motive* festgelegt (Tab. 43). In dieser Kategorie werden Antworten zusammengefasst, die kein spezifisches Motiv für die Antwort nennen oder sowohl affektive als auch kognitive Ursachen ansprechen, so dass insgesamt unklar bleibt, was genau der Grund für die Zufriedenheit ist.

Tabelle 43: Ober- und Unterkategorien für die qualitative Inhaltsanalyse der offenen Items zur Zufriedenheit

Oberkategorie	Unterkategorie
Kognitive Motive	Anforderungsniveau
	neues Lernen
Affektiv-motivationale Motive	Interesse unspezifisch
	Persönlicher Bezug
	Fachspezifisches Interesse
	Themenfeldspezifisches Interesse
	Inhaltsfeldspezifisches Interesse
	Kontextspezifisches Interesse
Unspezifische Motive	-

Anhand des Materials wurden die Oberkategorien *kognitive Motive* und *affektiv-motivationale Motive* weiter ausdifferenziert. Bei den kognitiven Motiven kann zusätzlich zwischen

Antworten unterschieden werden, die sich auf das Anforderungsniveau der Aufgaben oder auf den Erwerb neuen Wissens (*neues Lernen*) beziehen. Die Oberkategorie *affektiv-motivationale Motive* wurde durch sechs weitere Unterkategorien ausgeschärft. In der ersten Unterkategorie wurden alle Antworten gesammelt, die keinen spezifischen Grund für das Interesse an der Aufgabe umfassen. Alle anderen Unterkategorien umfassen jeweils einen konkreten Auslöser für das Interesse: den persönlichen Bezug, das Fach Chemie, das Themenfeld, das Inhaltsfeld oder den spezifischen Kontext.

Die Antworten wurden in zwei Durchgängen kodiert. Antworten, welche die eigentliche Zufriedenheit mit der kontextorientierten Aufgabe nicht betreffen, wurden als nicht kodierbar eingestuft. Das finale Kodiermanual ist im Anhang einzusehen.

10.3 Datenerhebung

Die zweite Teilstudie wurde zwischen März 2022 und Juni 2022 im regulären Chemieunterricht durchgeführt. Trotz der anhaltenden Covid-19-Pandemie konnte die Datenerhebung vollständig im Präsenzunterricht stattfinden. Der Studienablauf ist schematisch in Abbildung 30 dargestellt.

Für die Studienteilnahme wurden allen Schülerinnen und Schülern Tabletcomputer und Tablet-Stifte bereitgestellt. Auf den Tabletcomputern war das Lernmaterial als PDF-Datei hinterlegt. Den Lernenden wurde zudem ein Übersichtsblatt mit dem Studienablauf zur Verfügung gestellt, welches als Strukturierungshilfe dienen sollte (siehe Anhang). Die Lernenden hatten hier außerdem die Möglichkeit, ihren individuellen Identifikationscode zu notieren, welcher für eine eindeutige Zuordnung der Testinstrumente und Lernmaterialien benötigt wurde. Dieser Identifikationscode musste während der Datenerhebung häufiger angegeben werden. Weiterhin waren QR-Codes auf dem Übersichtsblatt hinterlegt, welche die Lernenden zu den entsprechenden LimeSurvey-Befragungen weiterleiteten, in denen die Testinstrumente integriert waren.

In der ersten Schulstunde wurden die Lernenden kurz über den Ablauf der Studie instruiert. Die Lernenden wurden hierbei nicht über den eigentlichen Zweck der Untersuchung informiert, um Erwartungseffekte zu minimieren (Döring & Bortz, 2016). Aus diesem Grund wurde den Lernenden mitgeteilt, dass die Evaluation des Lernmaterials der eigentliche Zweck der Untersuchung war. Nach der Instruktion wurde allen Lernenden das Übersichtsblatt zur Verfügung gestellt, auf dem sie ihren individuellen Identifikationscode formulieren sollten. Danach wurden die Tabletcomputer verteilt, mit denen die Lernenden den QR-Code scannen konnten. Dadurch wurden sie zum Prä-Test geleitet, der in einer LimeSurvey-Umfrage hinterlegt war. Die Lernenden haben die übrige Stunde genutzt, um den Prä-Test zu bearbeiten, mit dessen Hilfe die demografischen Daten, das Vorwissen in Chemie, das Interesse an Chemie, das chemiebezogene Selbstkonzept sowie die Motive zur Wahl eines Kontextes erhoben

wurden. Außerdem wurde das Fachwissen im Inhaltsfeld saure und alkalische Lösungen gemessen.

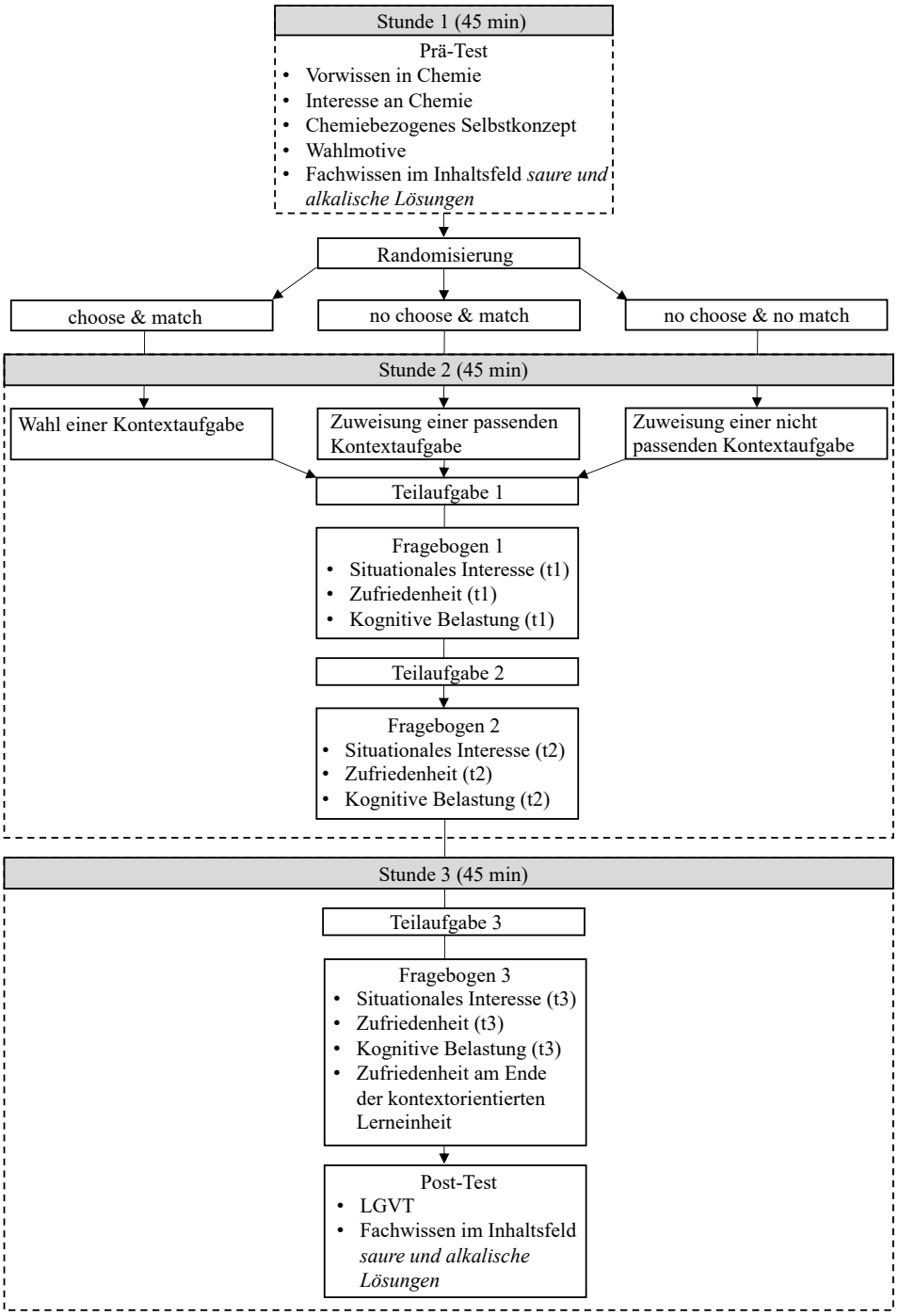


Abbildung 30: Ablauf der Datenerhebung (Teilstudie II)

Die Lernenden wurden nach der ersten Schulstunde randomisiert einer der drei Treatmentgruppen zugeteilt. Zur Vorhersage der Kontextwahl für die no choose & match- und die no choose & no match-Treatmentgruppe wurde das entwickelte Vorhersagemodell genutzt. Hierfür wurden die mithilfe der LimeSurvey-Umfrage erhobenen Daten exportiert, um sie für die Vorhersage der Kontextwahl zu nutzen. Damit die Lernenden nicht darüber instruiert wurden, welcher Treatmentgruppe sie angehören und welchen Kontext sie oder ihre Mitschülerinnen und Mitschüler bearbeiten, wurden pseudonymisierte Versionen des Lernmaterials erstellt, die nachfolgend als Datei bezeichnet wurden. In den Dateien 1 bis 3 war das entwickelte Lernmaterial zu jeweils einem Kontext enthalten. In der vierten Datei waren lediglich die Beschreibungen der drei Kontexte hinterlegt, die für die Kontextwahl der ersten Treatmentgruppe (choose & match) genutzt wurde.

Zur Instruktion der Lernenden wurde eine Liste basierend auf der Randomisierung und der Vorhersage der Kontextwahl erstellt. In dieser Liste waren die Identifikationscodes der Lernenden in der ersten Spalte und der Verweis auf eine zu bearbeitende Datei in der zweiten Spalte dargestellt (siehe Anhang). Die Lernenden der ersten Bearbeitungsgruppe erhielten die Datei 4 mit den Kontextbeschreibungen, aus denen sie einen Kontext zur Bearbeitung auswählen sollten. Nach der Auswahl eines Kontexts wurden die Schülerinnen und Schüler der choose & match-Grupppe direkt zur ausgewählten Aufgabe weitergeleitet. Gemäß der Vorhersage durch das Modell, wurde den Lernenden der no choose & match und no choose & no match-Treatmentgruppe die Datei 1 (Alltäglich), 2 (Besonders) oder 3 (Innerfachlich) zugewiesen.

Zu Beginn der zweiten Schulstunde wurde den Lernenden die Liste auf den Tabletcomputern zur Verfügung gestellt. Alle Probanden und Probandinnen mussten ihren Identifikationscode in der Liste ausfindig machen, der auf eine zu bearbeitende Datei verwies. Die Lernenden aus der ersten Treatmentgruppe mussten zunächst alle Kontextbeschreibungen durchlesen und die für sich ansprechendste auswählen. Nachdem sich die Lernenden der ersten Treatmentgruppe für einen Kontext entschieden hatten, wurden sie auf die Datei mit dem zugehörigen Lernmaterial verwiesen. Lernende der zweiten und dritten Treatmentgruppe konnten entsprechend dem Studiendesign nicht zwischen den Kontexten wählen und wurden stattdessen direkt zum Lernmaterial mit dem zugewiesenen Kontext geführt. Alle Probanden und Probandinnen bearbeiteten zunächst die erste Teilaufgabe des Lernmaterials und anschließend mussten den dazugehörigen Begleitfragebogen, mit dem die Zufriedenheit, das situationale Interesse und die kognitive Belastung erhoben wurden. Daran anknüpfend bearbeiteten die Lernenden die zweite Teilaufgabe sowie den zweiten Begleitfragebogen, der vollkommen identisch zum ersten Begleitfragebogen war.

In der dritten Schulstunde bearbeiteten die Lernenden die dritte Teilaufgabe des gewählten bzw. zugewiesenen Lernmaterial sowie den dazugehörigen Begleitfragebogen, der um das offene Item zur Zufriedenheit ergänzt wurde. Zuletzt wurde der Post-Test ausgefüllt, mit dem

das Fachwissen im Inhaltsfeld saure und alkalische Lösungen erneut gemessen wurde. Hierin wurde zudem der Leseverständnistest integriert.

Insgesamt nahm jede Schulstunde etwa 45 Minuten in Anspruch. Bei einer 60-minütigen Unterrichtsstunde gestaltete die Lehrkraft die verbleibenden 15 Minuten der Unterrichtsstunde, um die Bearbeitungszeit für alle Lernenden zu kontrollieren.

10.4 Stichprobe

Im Zuge der zweiten Teilstudie wurde erneut eine Gelegenheitsstichprobe aus der Metropolregion Rhein-Ruhr in Nordrhein-Westfalen befragt. Abermals konnten Schülerinnen und Schüler von vier Gymnasien und drei Gesamtschulen berücksichtigt werden. Die Lernenden stammten dabei aus insgesamt 20 unterschiedlichen Klassen bzw. Kursen. Zusammenfassend nahm damit eine Stichprobe von 355 Lernenden an der Untersuchung teil.

Da sich die Untersuchung über mehrere Schulstunden verteilte, liegen nicht von allen Lernenden vollständige Datensätze vor. Der Prä-Test wurde von allen Lernenden bearbeitet. Der Anteil der fehlenden Werte für den ersten Begleitfragebogen liegt zwischen 10.99 % und 11.27 %. Für den zweiten Begleitfragen liegt der Anteil der fehlenden Werte zwischen 16.62 % und 17.18 % und im dritten Begleitfragen liegt der Anteil der fehlenden Werte zwischen 18.59 % und 25.07 % im Vergleich zum Prä-Test. Im Post-Test fehlen zwischen 25.92 % und 35.68 % der Daten. Die fehlenden Werte sind hauptsächlich auf die Abwesenheit einzelnen Lernender zurückzuführen. Bedingt durch die Covid-19-Pandemie kam es auch zum Zeitpunkt der Datenerhebung für die zweite Teilstudie zu einer übermäßigen Anzahl von krankheitsbedingten Ausfällen. Somit wird auch für die zweite Teilstudie angenommen, dass die fehlenden Werte unter der Annahme MCAR entstanden. Folglich wird erneut ein listenweiser Fallausschluss durchgeführt (Lüdtke et al., 2007).

Nach diesem Vorgehen liegen vollständige Datensätze von 217 Lernenden vor. Die Lernenden waren zwischen 14 und 18 Jahre alt (M = 15.02, SD = 0.80). 45.66 % der Lernenden waren männlich und 47.03 % waren weiblich, wobei 7.31 % kein Geschlecht angegeben haben. Überdies haben etwa 81.74 % der Lernenden ein Gymnasium besucht.

10.5 Ergebnisse und Diskussion

Die quantitativen Daten der zweiten Teilstudie wurden – wie schon in der ersten Teilstudie – mithilfe von R (Version 4.2.3) in der Entwicklungsumgebung RStudio analysiert (R Core Team, 2022). Die für die Analyse genutzten Pakete sind in Tabelle 44 aufgeführt.

Tabelle 44: Übersicht über die verwendeten R-Pakete zur Auswertung der zweiten Teilstudie

Paket	Funktion	Version	Autor-/innen
tidyverse	Datenaufbereitung	1.3.2	Wickham et al. (2019)
TAM	IRT-Analysen	4.0-16	Robitzsch et al. (2022)
lavaan	CFA	0.6-12	Rosseel (2012)
semTools	Reliabilitätsbestimmung	0.5-6	Jorgenson et al. (2022)
lme4	Mehrebenenmodellierung	1.1-30	Bates et al. (2015)
lmerTest	Tests für Mehrebenenmodelle	3.1-3	Kuznetsova et al. (2017)
emmeans	Estimated marginal means	1.7.5	Lenth (2022)

10.5.1 Analyse der Messinstrumente

Die Messinstrumente wurden erneut hinsichtlich ihrer psychometrischen Qualität untersucht. Wie in Kapitel 7.1 ausgeführt, wurde ein Rasch-Modell für die Leistungstestdaten berechnet. Die Messinstrumente zur Erfassung affektiv-motivationaler Merkmale wurden mithilfe der CFA untersucht. Hier wurde erneut auf die robuste Maximum-Likelihood-Methode zurückgegriffen, um auch bei fehlender Normalverteilung effiziente Schätzungen der Modellparameter zu erhalten (Gäde et al., 2020). Fehlende Werte wurden dabei mithilfe der FIML-Methode berücksichtigt, um alle Informationen zur Beurteilung der Messqualität zu nutzen (Weiber & Mühlhaus, 2014). Da die Messinstrumente zur Erfassung der abhängigen Variablen zu mehreren Messzeitpunkten eingesetzt wurden, wurde zusätzlich auf longitudinale Messinvarianz geprüft. Lediglich die Skalen zur Erfassung der kognitiven Belastung konnten nicht mit den bereits mehrfach beschriebenen Methoden analysiert werden, da es sich um Single-Item-Skalen handelt, die jedoch bereits vielfach erprobt wurden (z. B. Schmeck et al., 2015). Genauso konnte der LGVT aufgrund der Testkonzeption nicht durch die beschriebenen Methoden untersucht werden (W. Schneider et al., 2017).

Vorwissen in Chemie: Analog zur ersten Teilstudie und zur Zusatzerhebung wurde die Qualität des Leistungstests zur Erfassung des Vorwissens in Chemie mithilfe eines eindimensionalen Rasch-Modells untersucht. Da das Vorwissen in Chemie zur Vorhersage der Kontextwahl genutzt wurde, wurden die Personenparameter auf Basis fixierter Itemparameter berechnet, so dass sie direkt vergleichbar zu den Personenparametern sind, die zur Konzeption des Modells genutzt wurden. Auch hier wurde allerdings zunächst ein Rasch-Modell mit frei geschätzten Itemparametern berechnet (Tab. 45).

Tabelle 45: Kennwerte des Rasch-Modells mit frei geschätzten, vollständig fixierten und teilweise fixierten Itemparametern für den Vorwissenstest (Teilstudie II)

Modellmerkmal	Kennwert (frei geschätzt)	Kennwert (vollständig fixiert)	Kennwert (teilweise fixiert)
$N_{Personen}$	355	355	355
N_{Items}	57	57	57
Parameterzahl	58	2	9
Deviance	9011.21	9176.80	9110.43
Varianz	.41	.44	.42
WLE-Reliabilität	.61	.62	.61
Itemparameter	-1.18 – 1.59	-1.71 – 1.46	-1.71 – 1.27
Personenparameter	-2.53 – 3.01	-2.89 – 2.81	-2.85 – 2.77
wMNSQ	0.89 – 1.11	0.88 – 1.69	0.87 – 1.22
t-Werte (wMNSQ)	-1.99 – 1.48	-2.03 – 3.92	-2.18 – 2.39

Die Modellierung der Daten durch das Rasch-Modell mit frei geschätzten Itemparametern zeigt, dass alle Testitems den festgelegten Richtwerten für die wMNSQ- und *t*-Werten entsprechen. Demzufolge weisen alle Items eine ausreichende Passung zum Rasch-Modell auf. Die durch den Vorwissenstest erzeugte Varianz liegt in einem geringen Bereich. Zudem beträgt die WLE-Reliabilität = .61. Insgesamt liegen Varianz und Reliabilität damit in einem ähnlichen Bereich wie in Teilstudie I und der Zusatzerhebung.

Daran anschließend wurde ein zweites Rasch-Modell mit vollständig fixierten Itemparametern berechnet. Die Varianz und die WLE-Reliabilität steigen marginal. Durch die Analyse der wMNSQ- und *t*-Werte zeigt sich allerdings, dass sieben Testitems nicht zum Rasch-Modell passen und damit eine gewisse Messvarianz aufweisen. Die Itemparameter dieser Testitems wurden in einem dritten Rasch-Modell frei geschätzt. In dem Rasch-Modell mit teilweise fixierten Itemparametern zeigen alle Testitems eine ausreichende Passung wie die Analyse der entsprechenden Kennwerte zeigt. Die Varianz und die WLE-Reliabilität sinken etwas, liegen aber dennoch in einem vergleichbaren Bereich wie in der ersten Teilstudie. Die zugehörige Wright-Map zeigt, dass der Test das Fähigkeitsspektrum der Stichprobe gut abbildet (Abb. 31).

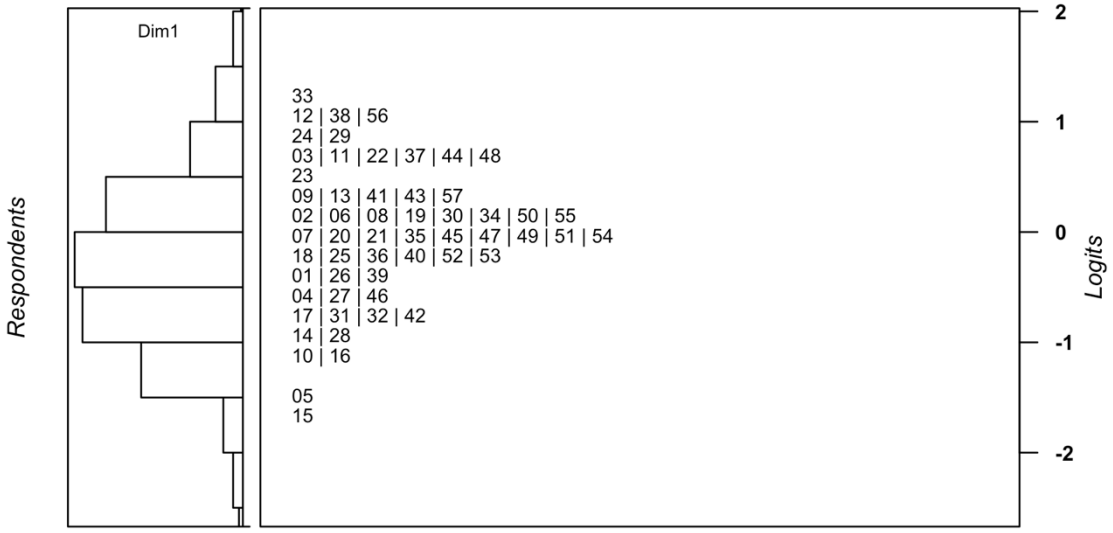


Abbildung 31: Wright-Map zum Vorwissenstest basierend auf dem Rasch-Modell mit teilweise fixierten Itemparametern (Teilstudie II)

Interesse an Chemie und Motivation im Fach Chemie: Das Instrument zur Erfassung des Interesses an Chemie und zur Motivation im Fach Chemie wurde erneut mithilfe der CFA untersucht. Es wurde das gleiche Modell wie in der ersten Teilstudie spezifiziert. Die Fit-Indizes verweisen auf eine ausreichende Passung, ($\chi^2(318) = 709.62$, $p < .001$, CFI = .916, TLI = .908, RMSEA = .061, SRMR = .056). Ein Item (*exm4*) zur Erfassung der extrinsischen Motivation zeigt eine äußerst geringe Faktorladung ($\lambda = .17$). Da die Skala zur extrinsischen Motivation nicht zur Konzeption des Vorhersagemodells genutzt wurde, wird die Veränderung und der Ausschluss des Items zur Verbesserung der Reliabilität als legitim angesehen. Das Modell zeigt ohne das Item *exm4* eine geringfügig bessere Passung auf die Daten, ($\chi^2(293) = 658.65$, $p < .001$, CFI = .921, TLI = .912, RMSEA = .062, SRMR = .054). In Tabelle 46 sind die Anzahl der Items pro Skala, die Faktorladungen und Reliabilitäten dargestellt. Einige Items weisen zwar niedrige Faktorladungen auf ($\lambda < .50$), allerdings liegen die Reliabilitäten für alle Skalen in einem zufriedenstellenden Bereich.

Tabelle 46: Itemanzahl, Faktorladungen und Reliabilität der Skalen zum individuellen Interesse an Chemie und zur Motivation im Fach Chemie (Teilstudie II)

Skala	*n*	λ	ω	α
Individuelles Interesse an Chemie	15	.47 - .81	.94	.94
Inhaltsfeldspezifisches Interesse	3	.72 - .78	.84	.84
Themenspezifisches Interesse	5	.38 - .76	.80	.82
Extrinsische Motivation in Chemie	3	.53 - .67	.73	.71

Anmerkung: λ = standardisierte Faktorladung, ω = McDonalds Omega, α = Cronbachs Alpha. Es wurden Antworten von *N* = 355 Personen berücksichtigt.

Selbstkonzept: Für die Skalen zum Selbstkonzept wurde ein identisches Zwei-Faktoren-Modell berechnet wie in der ersten Teilstudie. Das Modell beschreibt die Daten hinreichend, $\chi^2(43) = 78.76$ $p < .001$, CFI = .983, TLI = .978, RMSEA = .051, SRMR = .029. Sowohl Faktorladungen als auch die Reliabilität der Skalen liegen in einem guten Bereich (Tab. 47).

Tabelle 47: Itemanzahl, Faktorladungen und Reliabilität der Skalen zum chemiebezogenen Selbstkonzept (Teilstudie II)

Skala	*n*	λ	ω	α
Selbstkonzept	8	.55 - .84	.91	.91
Selbstwirksamkeitserwartung	3	.77 - .84	.85	.85

Anmerkung: λ = standardisierte Faktorladung, ω = McDonalds Omega, α = Cronbachs Alpha. Es wurden Antworten von *N* = 355 Personen berücksichtigt.

Wahlmotive: Zur Untersuchung der Konstruktvalidität des Instruments zur Erfassung der Wahlmotive wurde das in Teilstudie I identifizierte Drei-Faktoren-Modell spezifiziert und hinsichtlich der Modellpassung untersucht. Außerdem wurde die Fehlerkovarianz zwischen den Items *per3* und *per4* in das Modell aufgenommen, da das Item *per4* in der Zusatzerhebung eine negative Residualvarianz zeigte. Das beschriebene Modell zeigt eine gute Modellpassung, $\chi^2(31) = 78.57$ $p < .001$, CFI = .941, TLI = .915, RMSEA = .067, SRMR = .061. Sowohl die Items zur Erfassung des Wahlmotivs *Interesse und Neugier* als auch zur Erfassung des Wahlmotivs *Überraschungsmoment* zeigen suffiziente Faktorladungen (Tab. 48). Weiterhin liegen die Reliabilitäten in einem guten Bereich.

Tabelle 48: Itemanzahl, Faktorladungen und Reliabilitäten der Skalen zur Erfassung der Wahlmotive (Teilstudie II)

Skala	*n*	λ	ω	α
Interesse & Neugier	3	.58 - .84	.78	.76
Überraschungsmoment	3	.69 - .79	.79	.79
Persönlicher Bezug	4	.02 - .71	.44	.45

Anmerkung: λ = standardisierte Faktorladung, ω = McDonalds Omega, α = Cronbachs Alpha. Es wurden Antworten von *N* = 355 Personen berücksichtigt

Ein anderes Bild zeigt sich für die Skala zur Erfassung des Wahlmotivs des persönlichen Bezugs: Sowohl Cronbachs α als auch McDonalds ω sind unzureichend. Insbesondere das Item *per4* zeigt mit λ = .02 eine äußerst niedrige Faktorladung. Ferner liegt auch die Faktorladung des Items *biub3* (λ = .24) in keinem zufriedenstellenden Bereich. Damit scheinen die geringen Faktorladungen der beiden Items die Ursache für die geringe Reliabilität der Skala zu sein. So verbessert der Ausschluss der beiden Items zwar die Reliabilität der Skala (α = .70; ω = .71), wobei auch das so spezifizierte Modell zu den Daten passt ($\chi^2(17) = 32.12$, $p < .001$, CFI = .981, TLI = .968, RMSEA = .050, SRMR = .039), allerdings wird von der Anpassung der Skala aus mehreren Gründen abgesehen. Zum einen würde die Skala so nur aus zwei Items bestehen, wobei psychometrische Skalen mindestens drei Items umfassen sollten (Döring & Bortz, 2016). Zum anderen wäre die Skala in Teilstudie nicht mehr vergleichbar zu der Form, die zur Vorhersage der Kontextwahl und zur Konstruktion des Modells genutzt wurde, da sie aus unterschiedlichen Items bestehen würde. Hierbei würden Informationen über das zu messende Konstrukt verloren gehen (Döring & Bortz, 2016), wodurch möglicherweise die Performanz des Modells beeinträchtigt wäre.

Fachwissen im Inhaltsfeld saure und alkalische Lösungen: Die Qualität des konzipierten Leistungstests zur Erfassung des Fachwissens im Inhaltsfeld saure und alkalische Lösungen wurde mithilfe des eindimensionalen Rasch-Modells beurteilt. Zunächst wurde ein Rasch-Modell für die Daten zum Prä- und Post-Zeitpunkt ohne Restriktionen geschätzt (Tab. 48).

Die Varianz der Personenverteilung ist zum Prä-Zeitpunkt äußerst gering, was darauf hindeutet, dass der Test nicht deutlich zwischen den Personen unterscheidet. Demzufolge liegt die WLE-Reliabilität in einem desolaten Bereich. Trotz dessen deuten die wMNSQ- und *t*-Werte auf eine ausreichende Passung der Testitems zum Rasch-Modell hin. Zum Post-Zeitpunkt ist die durch den Fachwissenstest erzeugte Varianz etwas größer, allerdings immer noch in keinem zufriedenstellenden Bereich. Zugleich ist die WLE-Reliabilität deutlich höher als zum Prä-Zeitpunkt, kann aber dennoch als nicht ausreichend eingestuft werden. Wie Tabelle 49

zeigt, weisen die Testitems auch zum Post-Zeitpunkt eine hinreichende Passung zum Rasch-Modell auf.

Tabelle 49: Kennwerte der Rasch-Modelle für den Fachwissenstest zum Prä- und Post-Zeitpunkt

Modellmerkmal	Prä-Zeitpunkt (frei geschätzt)	Post-Zeitpunkt (frei geschätzt)
$N_{Personen}$	355	263
N_{Items}	12	12
Parameterzahl	13	13
Deviance	5168.85	3934.58
Varianz	.16	.46
WLE-Reliabilität	.16	.48
Itemparameter	0.04 – 1.21	-1.14 – 1.01
Personenparameter	-2.45 – 2.30	-2.98 – 3.84
wMNSQ	0.98 – 1.02	0.96 – 1.05
t-Werte (wMNSQ)	-1.13 – 0.38	-0.71 – 0.95

Zur Modellierung des Fachwissenszuwachses werden die Itemparameter des Prä-Zeitpunktes auf die für den Post-Zeitpunkt berechneten Parameter fixiert, da zum Post-Zeitpunkt eine größere Varianz erzielt wird und die WLE-Reliabilität am höchsten ist. Durch die vollständige Fixierung der Itemparameter verbessert sich die Varianz und die WLE-Reliabilität zum Prä-Zeitpunkt marginal (Tab. 50). Sowohl die Varianz als auch die WLE-Reliabilität sind allerdings nach wie vor nicht zufriedenstellend. Durch die Fixierung der Itemparameter erweisen sich drei Testitems als nicht Rasch-konform und damit als messvariant. Deshalb werden die Itemparameter dieser drei Testitems in einem nachfolgenden Rasch-Modell frei geschätzt. Im Rasch-Modell mit teilweise fixierten Itemparametern zeigen diese Aufgaben eine hinreichende Passung, weshalb sie im Datensatz belassen und für die weitere Analyse genutzt werden. Die Varianz und WLE-Reliabilität verschlechtern sich durch die Freisetzung der Itemparameter marginal und sind weiterhin nicht zufriedenstellend.

Tabelle 50: Kennwerte der Rasch-Modelle mit vollständig fixierten und teilweise fixierten Itemparametern für den Fachwissenstest zum Prä-Zeitpunkt

Modellmerkmal	Prä-Zeitpunkt (vollständig fixiert)	Prä-Zeitpunkt (teilweise fixiert)
$N_{Personen}$	355	263
N_{Items}	12	12
Parameterzahl	2	5
Deviance	5282.37	5190.40
Varianz	.22	.18
WLE-Reliabilität	.21	.17
Itemparameter	-1.14 – 1.01	-0.33 – 0.93
Personenparameter	-2.98 – 2.03	-2.83 – 1.96
wMNSQ	0.94 – 1.29	0.94 – 1.16
t-Werte (wMNSQ)	-1.69 – 5.83	-1.83 – 2.23

Die Gegenüberstellung der Item- und Personenparameter mithilfe der Wright-Map deutet auf eine wesentliche Ursache für die wenig zufriedenstellende Testgüte hin: Der Leistungstest bildet das Fähigkeitsspektrum der Stichprobe nur sehr eingeschränkt ab (Abb. 32). Auf der linken Seite der Wright-Map ist die Verteilung der Personenparameter zum Prä- und Post-Zeitpunkt dargestellt, auf der rechten Seite die Verteilung der Itemparameter. Wenn sich die Itemparameter zum Prä- und Post-Zeitpunkt nicht unterscheiden, die Items also messinvariant sind, ist dies durch eine Graufärbung dargestellt. Für messvariante Items ist die Itemschwierigkeit zum Prä- (rot) und Post-Zeitpunkt (blau) durch unterschiedliche Farben gekennzeichnet.

Es zeigt sich deutlich, dass die meisten Lernenden in der unteren Hälfte des Fähigkeitsspektrums angesiedelt sind. Aufgrund der Normierung der Parameterschätzung liegt die mittlere Personenfähigkeit zum Post-Zeitpunkt bei $M = 0.04$ Logits ($SD = 0.97$). Zum Prä-Zeitpunkt liegt die mittlere Personenfähigkeit bei $M = -0.36$ Logits ($SD = 0.76$). Infolge der Restriktionen der Itemparameter auf den Post-Zeitpunkt, sind die Fähigkeitsausprägungen zu beiden Messzeitpunkten direkt vergleichbar, was auf einen Fachwissenszuwachs von 0.40 Logits hindeutet. Demgegenüber liegen die mittleren Itemparameter zum Post-Zeitpunkt bei $M = 0.49$ Logits ($SD = 0.61$) und zum Prä-Zeitpunkt bei $M = 0.48$ Logits ($SD = 0.40$). Dieses Ergebnis deutet darauf hin, dass der Leistungstest zu schwierig für die untersuchte Stichprobe ist. Wie in der Wright-Map ersichtlich, sind die meisten Items in einem mittleren Bereich

zwischen 0 und 1 Logit angesiedelt, wodurch besonders Items im unteren Fähigkeitsspektrum fehlen (Abbildung 32). Ebenso fehlen Items für die wenigen Personen mit einer hohen Merkmalsausprägung. Folglich könnte das Instrument durch weitere Items vermutlich verbessert werden. Insgesamt erscheint das Instrument damit nur wenig geeignet, um das Fachwissen im oberen und unteren Bereich differenziert zu erfassen. Dieser Befund muss bei der nachfolgenden Ergebnisdarstellung berücksichtigt werden.

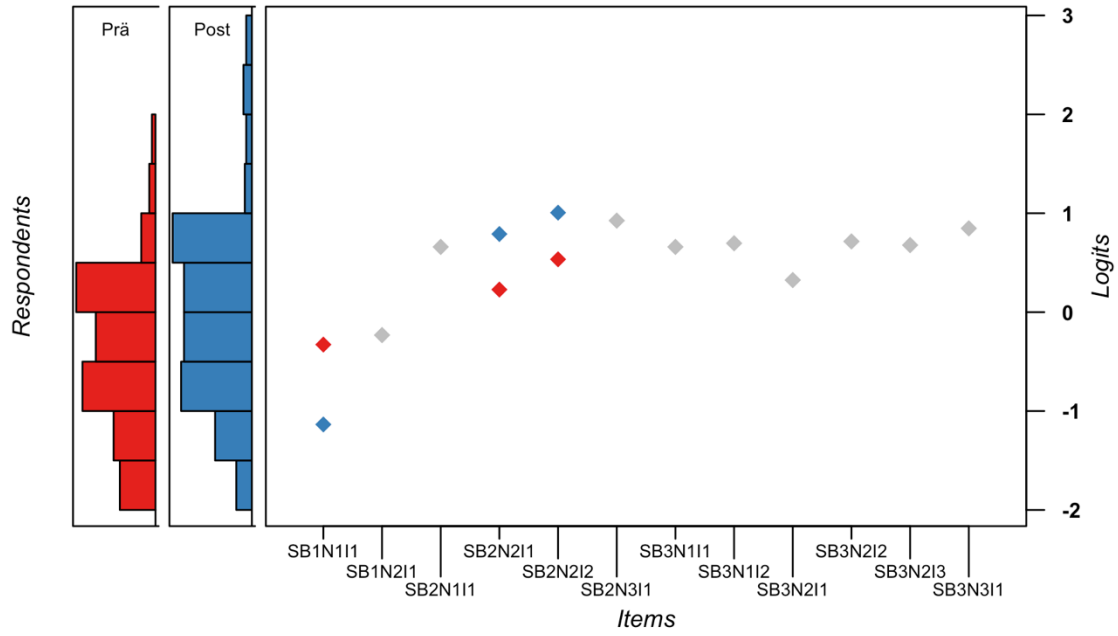


Abbildung 32: Wright-Map zum Fachwissenstest basierend auf den Rasch-Modellen mit frei geschätzten Itemparametern zum Post-Zeitpunkt und teilweise fixierten Itemparametern zum Prä-Zeitpunkt

Zufriedenheit: Im Rahmen der ersten Teilstudie konnte gezeigt werden, dass es sich beim situationalen Interesse und der Zufriedenheit um zwei abgrenzbare Konstrukte handelt. Um das Konstrukt der Zufriedenheit in der zweiten Teilstudie weiter auszudifferenzieren, wurde eine Skala zur Erfassung der Zufriedenheit mit dem Inhaltsfeld und eine Skala zur Zufriedenheit mit dem Kontext eingesetzt. Zur Untersuchung des Instruments wurden für jeden Messzeitpunkt ein Generalfaktor-Modell und ein Zwei-Faktoren-Modell spezifiziert (Tab. 51).

Wie Tabelle 51 zeigt, können die Daten zu keinem Messzeitpunkt hinreichend durch das General- oder Zwei-Faktoren-Modell beschrieben werden. Die Modifikationsindizes verweisen auf die Modellierung der Fehlerkovarianz zwischen den Items zur Zufriedenheit am Inhaltsfeld und zur Zufriedenheit an Kontexten. Weiterhin liegt die latente Korrelation der beiden Skalen im Zwei-Faktoren-Modell über alle Messzeitpunkte in einem sehr hohen Bereich ($.94 \leq r \leq .96$), was auf fehlende diskriminante Validität hindeutet. Die Zufriedenheit am

Inhaltsfeld und die Zufriedenheit am Kontext sind anscheinend nicht separat erfassbar. Die Schülerinnen und Schüler scheinen nicht hinreichend dazu in der Lage zu sein, zwischen dem Inhaltsfeld und dem Kontext zu diskriminieren (Kölbach & Sumfleth, 2013). Deswegen wurden die Items zur Zufriedenheit mit dem Inhaltsfeld aus dem Instrument entfernt.

Tabelle 51: Modellgüte des General- und Zwei-Faktoren-Modells des Instruments zur Erfassung der Zufriedenheit für den ersten, zweiten und dritten Messzeitpunkt

Modell	χ^2	df	CFI	TLI	SRMR	RMSEA
Messzeitpunkt 1						
General-Faktor	107.27	20	.868	.816	.072	.118
Zwei-Faktoren	107.14	19	.872	.812	.071	.139
Messzeitpunkt 2						
General-Faktor	121.73	20	.858	.801	.083	.149
Zwei-Faktoren	128.75	19	.859	.792	.083	.153
Messzeitpunkt 3						
General-Faktor	118.38	20	.865	.812	.083	.156
Zwei-Faktoren	124.47	19	.865	.801	.083	.160

Anmerkung: Es wurden Antworten von $N = 339$ Personen berücksichtigt.

Zur Prüfung auf longitudinale Messinvarianz wurde zunächst ein angepasstes Modell für jeden Messzeitpunkt spezifiziert (Tab. 52). Sowohl der Likelihood-Test als auch die deskriptiven Gütekriterien verweisen auf eine exzellente Modellpassung für jeden Messzeitpunkt. Im nächsten Schritt wurde ein Messmodell mit konfiguraler Messinvarianz berechnet. Da zu jedem Messzeitpunkt dieselben Items eingesetzt wurden, wurde die Residualkovarianz gleicher Items zwischen den jeweiligen Messzeitpunkten modelliert (Mackinnon et al., 2022). Das Messmodell mit konfiguraler Messinvarianz beschreibt die Daten adäquat. Im nächsten Schritt wurden die Faktorladungen gleichgesetzt und die beiden Messmodelle verglichen. Durch die Restriktion der Faktorladungen ändern sich CFI (ΔCFI = -.001), SRMR (ΔSRMR = .006) und RMSEA (ΔRMSEA = .002) nur marginal, wodurch die Annahme metrischer Invarianz beibehalten wird. Die Skala Invarianz des Instruments wurde geprüft, indem die Intercepts der Items restringiert wurden. Das Messmodell mit skalarer Invarianz beschreibt die Daten signifikant schlechter als das Messmodell mit metrischer Invarianz (ΔCFI = -.083, ΔSRMR = .044, ΔRMSEA = .056).

Tabelle 52: Kennwerte der Modelle zur Prüfung auf longitudinale Messinvarianz des Instruments zur Erfassung der Zufriedenheit

Modell	χ^2	df	CFI	TLI	SRMR	RMSEA
MZP1	2.15	2	.999	.998	.014	.015
MZP2	1.48	2	1.000	1.006	.012	.000
MZP3	0.95	2	1.000	1.009	.009	.000
Konfigural	47.74	39	.993	.988	.030	.027
Metrisch	56.71	47	.992	.989	.044	.025
Skalar	166.62	53	.909	.887	.088	.081

Anmerkung: Es wurden Antworten von N = 339 Personen berücksichtigt.

Aus diesem Grund wird im Folgenden auf partielle skalare Messinvarianz geprüft. Für partielle Messinvarianz müssen mindestens zwei Items pro Skala restringiert bleiben (Steenkamp & Baumgartner, 1998). Auch die Freisetzung der Intercepts der Items *zufk1* und *zufk2* führt verglichen mit dem Messmodell metrischer Invarianz zu einer signifikanten Änderung im CFI (ΔCFI = -.046) und RSMEA (ΔRMSEA = .017), so dass die Annahme (partieller) skalarer Invarianz zu verwerfen ist. Dies ist nicht überraschend, da skalare Messinvarianz häufig eine zu strenge Annahme an das Messinstrument ist, die meist nicht bestätigt werden kann (z. B. Dimitrov, 2010). Da sich die Zufriedenheit sehr stark auf die jeweilige Aufgabe bezieht, ist es fraglich, inwieweit überhaupt skalare Messinvarianz erzielt werden kann, da die Lernenden zu jedem Messzeitpunkt eine andere Teilaufgabe bewerten. Möglicherweise handelt es sich hierbei also um ein methodisches Artefakt und um keine systematische Verzerrung im Konstrukt, so dass die Vergleiche der Messzeitpunkte legitimiert sind. Die Reliabilitätsprüfung verweist darüber hinaus auf zufriedenstellende Reliabilitäten zu jedem Messzeitpunkt (MZP1: α = .72, ω = .79; MZP2: α = .70, ω = .76; MZP3: α = .73, ω = .71).

Situationales Interesse: Für das Instrument zur Erfassung des situationalen Interesses wurde ein Drei-Faktor-Modell spezifiziert, welches sich in der ersten Teilstudie als geeignet herausgestellt hat. Zunächst wurde für jeden Messzeitpunkt ein separates Modell geschätzt, welches die Daten zu jedem Messzeitpunkt adäquat beschreibt (Tab. 53).

Anschließend wurde ein Messmodell mit konfiguraler Messinvarianz geprüft. Auch hier wurde die Residualkovarianz gleicher Items zwischen den jeweiligen Messzeitpunkten modelliert (Mackinnon et al., 2022). Die Prüfung der Kovarianzmatrix der latenten Variablen zeigt, dass die Kovarianzmatrix nicht positiv definiert ist. Das Modell ist damit nicht identifizierbar (Gäde et al., 2020). Mögliche Ursache ist die hohe latente Korrelation zwischen der

Skala zu Erfassung der emotionalen und wertbezogenen Valenz am dritten Messzeitpunkt (r = .98). Die hohe latente Korrelation deutet darauf hin, dass die Schülerinnen und Schüler nach der dritten Kontextaufgabe nicht mehr hinreichend zwischen wertbezogener und emotionaler Valenz differenzieren. Daher wurde für die Analyse des Messinstruments ein gemeinsamer Faktor für die wertbezogene und emotionale Valenz spezifiziert, der nachfolgend als affektive Komponente (bzw. affektive Valenz) des situationalen Interesses bezeichnet wird.

Tabelle 53: Modellgüte des Drei-Faktoren-Modells des Instruments zur Erfassung des situationalen Interesse für den ersten, zweiten und dritten Messzeitpunkt

Modell	χ^2	df	CFI	TLI	SRMR	RMSEA
MZP1	50.91	32	.979	.971	.038	.047
MZP2	61.45	32	.970	.958	.043	.061
MZP3	55.19	32	.977	.968	.038	.055

Anmerkung: Es wurden Antworten von N = 339 Personen berücksichtigt.

Das angepasste Modell eignet sich gut zur Beschreibung der Daten am ersten, zweiten und dritten Messzeitpunkt (Tab. 54). Zur Messinvarianzprüfung wurde zunächst ein Messmodell mit konfiguraler Messinvarianz als Vergleichsmodell berechnet. Erneut wurde die Residualkovarianz gleicher Items zwischen den Messzeitpunkten in das Modell aufgenommen. Die Kennwerte des berechneten Messmodells entsprechen den festgesetzten Cut-off Werten und verweisen damit auf hinreichende Modellpassung. Im nächsten Schritt wurden die Faktorladungen der Items restringiert, um auf metrische Invarianz zu überprüfen. Die Änderungen im CFI (ΔCFI = -.007), SRMR (ΔSRMR = .014) und im RMSEA (ΔRMSEA = .001) zeigen metrische Invarianz an (Chen, 2007). Danach wurden die Intercepts gleicher Items über die drei Messzeitpunkte fixiert, was zu einer bedeutenden Abnahme im CFI (ΔCFI = .034) führt. Deshalb wurde auf partielle skalare Messinvarianz geprüft, indem die Intercepts einiger Items wieder freigesetzt wurden. Für partielle Messinvarianz müssen mindestens zwei Items pro Skala restringiert bleiben (Steenkamp & Baumgartner, 1998). Dementsprechend wurden die Intercepts der Items *emo1*, *emo2*, *emo3*, *wert2* und *epi1* freigesetzt, was zu einer vertretbaren Änderung im CFI (ΔCFI = -.006), SRMR (ΔSRMR = .002) und RMSEA (ΔRMSEA = .002) geführt hat. Folglich wird die Annahme partieller skalarer Messinvarianz beibehalten. Im letzten Schritt wurde die Residualvarianz der Items über die Messzeitpunkte gleichgesetzt, um auf strikte Messinvarianz zu prüfen. Hier verweist die Abnahme im CFI (ΔCFI = -.007) sowie die Zunahme im SRMR (ΔSRMR = .001) und RMSEA (ΔRMSEA = .002) auf Gültigkeit des Messmodells mit (partieller) strikter Messinvarianz. Die Reliabilitäten liegen sowohl für die affektive Valenz (MZP1: α = .80, ω = .92; MZP2: α = .84, ω = .75; MZP3: α = .82, ω = .80) als auch für die epistemische Komponente (MZP1: α = .80, ω = .81; MZP2: α = .79,

ω = .90; MZP3: α = .86, ω = .74) des situationalen Interesses in einem zufriedenstellenden Bereich.

Tabelle 54: Kennwerte der Zwei-Faktoren-Modelle zur Prüfung auf longitudinale Messinvarianz des Instruments zur Erfassung des situationalen Interesse

Modell	χ^2	df	CFI	TLI	SRMR	RMSEA
MZP1	76.48	34	.954	.939	.047	.068
MZP2	87.52	34	.946	.928	.050	.080
MZP3	57.23	34	.977	.970	.040	.053
Konfigural	550.16	360	.953	.943	.052	.042
Metrisch	586.40	380	.946	.938	.066	.043
Skalar	728.85	400	.919	.912	.080	.052
part. Skalar	632.72	390	.940	.933	.068	.045
part. Strikt	679.88	410	.933	.929	.069	.047

Anmerkung: Es wurden Antworten von $N = 339$ Personen berücksichtigt.

10.5.2 Vergleichbarkeit der Untersuchungsgruppen

In einem ersten Schritt wurde untersucht, ob die Randomisierung zu vergleichbaren Treatmentgruppen geführt hat. Nur wenn keine signifikanten Unterschiede im Hinblick auf mögliche personenbezogene Störvariablen bestehen, sind Unterschiede in den abhängigen Variablen mit hoher Wahrscheinlichkeit auf die Manipulation der unabhängigen Variable zurückzuführen (Döring & Bortz, 2016). In Tabelle 55 sind die demografischen und die deskriptiven Statistiken der Kontrollvariablen für jede Treatmentgruppe dargestellt.

In wenigen Variablen (z. B. Leseverständnis) zeigen sich geringe deskriptive Unterschiede. Zur Überprüfung der Unterschiedshypothese wurde eine multivariate Varianzanalyse mit den Kontrollvariablen als abhängige Variable und dem Treatment als unabhängige Variable berechnet. Auf Prüfung der Verteilungsannahmen wurde verzichtet, da die Gruppen annähernd gleich groß sind (Field et al., 2012). Die MANOVA mit Pillai's Trace verweist auf keine signifikanten Unterschiede, $V = 0.11$, $F(18, 414) = 1.33$, $p = .19$, $\eta^2_p = .05$. Ebenso zeigen die post-hoc durchgeführten univariaten Varianzanalysen kein signifikantes Ergebnis, wobei auch die Effektstärken auf vernachlässigbare Effekte hindeuten (siehe Anhang). Weiterhin gibt es keinen Zusammenhang zwischen den Treatmentgruppen und dem Geschlecht ($\chi^2(4) = 0.49$, $p = .97$, $V = .03$) oder der besuchten Schulform ($\chi^2(2) = 0.57$, $p = .75$, $V = .05$). Insgesamt

kann die Randomisierung damit als erfolgreich angesehen werden, da die Untersuchungsgruppen vergleichbar im Hinblick auf die personenbezogenen Kontrollvariablen sind.

Tabelle 55: Demografische Daten und Deskriptive Statistiken der Kontrollvariablen für jedes Treatment

	choose & match	no choose & match	no choose & no match
N	76	72	69
Geschlecht (%)			
Männlich	43.42	47.22	46.38
Weiblich	50.00	45.83	46.38
Keine Antwort	6.58	6.94	7.24
Schulform (%)			
Gymnasium	84.21	79.17	81.16
Gesamtschule	15.79	20.83	18.84
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Lesever	17.03 (15.09)	14.00 (13.48)	13.67 (15.97)
Vorwissen	-0.01 (0.87)	0.11 (0.89)	0.20 (0.88)
$Fachwissen_{prä}$	-0.36 (0.79)	-0.22 (0.73)	-0.27 (0.82)
IntCh	2.46 (0.64)	2.43 (0.63)	2.25 (0.69)
IntIn	2.25 (0.87)	2.23 (0.76)	1.99 (0.89)
IntThe	2.55 (0.72)	2.69 (0.62)	2.53 (0.71)
ExMot	2.92 (0.61)	3.00 (0.63)	2.91 (0.66)
Selko	2.56 (0.63)	2.68 (0.59)	2.44 (0.69)
Selwi	2.87 (0.64)	2.92 (0.69)	2.70 (0.70)

Anmerkung: Lesever = Leseverständnis; Vorwissen = Vorwissen in Chemie; $Fachwiss_{prä}$ = Fachwissen im Inhaltsfeld saure und alkalische Lösungen zum Prä-Zeitpunkt; IntCh = Individuelle Interesse an Chemie; IntIn = Inhaltsfeldspezifisches Interesse; IntThe = Themenspezifisches Interesse, ExMot = Extrinsische Motivation; Selko = Selbstkonzept; Selwi = Selbstwirksamkeitserwartung

Daneben wurden die Treatments in Bezug auf die bearbeiteten Kontexte verglichen (Abb. 33). Zunächst ist auffällig, dass die Lernenden aus der choose & match-Gruppe häufiger den besonderen und innerfachlichen Kontext gewählt haben als die Lernenden aus der ersten Teilstudie. Dies ist womöglich darauf zurückzuführen, dass in der zweiten Teilstudie jeweils nur

ein alltäglicher und besonderer Kontext zur Wahl standen. Der relative Anteil der bearbeiteten Kontexte unterscheidet sich in der choose & match und der no choose & match-Treatmentgruppe nur marginal. Da das Vorhersagemodell auf der Grundlage von gewählten Kontexten konzipiert wurde, ist dieses Ergebnis nur wenig überraschend. Allerdings zeigt sich in der no choose & no match-Gruppe eine etwas andere Verteilung: Die Lernenden haben hier häufiger den innerfachlichen Kontext bearbeitet, wohingegen der alltägliche Kontext seltener bearbeitet wurde. Diese Verteilung ist im Wesentlichen darauf zurückzuführen, dass der alltägliche Kontext für einen Großteil der Lernenden passend ist, während der innerfachliche Kontext häufig nicht zu den Personenmerkmalen passt. So wurde einem Großteil der Lernenden in der dritten Treatmentgruppe der innerfachliche Kontext zugewiesen. Gemäß dem berechneten χ^2-Test gibt es einen Zusammenhang zwischen der Treatmentgruppe und der bearbeiteten Kontextaufgabe ($\chi^2(4) = 13.27$, $p = .01$, $V = .17$), wobei es sich um einen eher kleinen Zusammenhang handelt. Folglich muss der bearbeitete Kontext bei der weiteren Analyse berücksichtigt werden, da nicht ausgeschlossen werden kann, dass der bearbeitete Kontext die abhängigen Variablen beeinflusst.

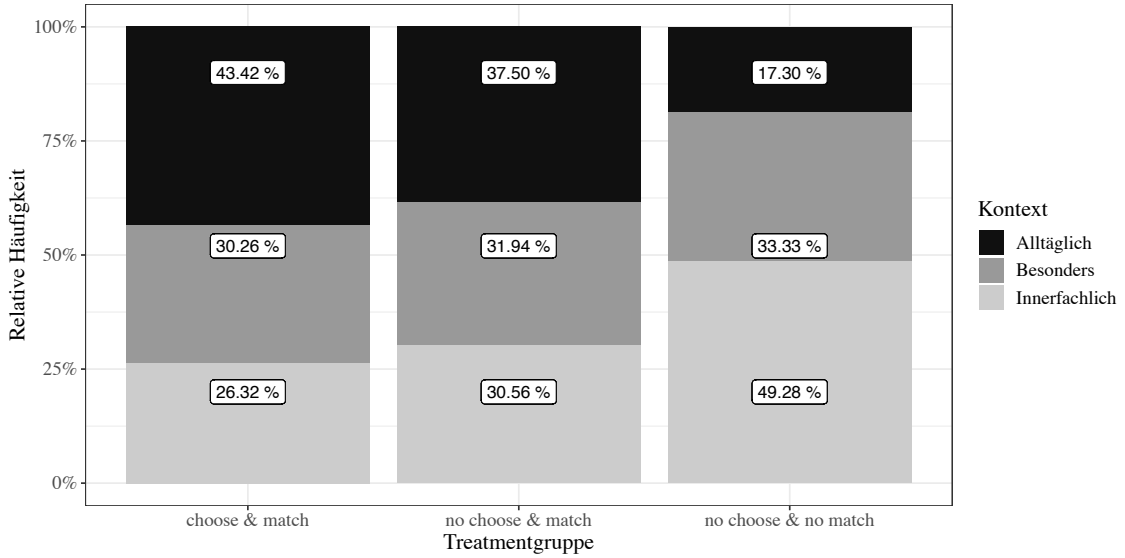


Abbildung 33: Relative Häufigkeit der bearbeiteten Kontexte in den Treatmentgruppe

10.5.3 Überprüfung der Vorhersagegüte

Die Untersuchung der Vorhersagegüte des neuronalen Netzes ist im Rahmen der zweiten Teilstudie unabdingbar, um zu beurteilen, wie valide die Zuordnung der Kontextaufgaben für die no choose & match und no choose & no match-Treatmentgruppe ist. Die Vorhersagegüte des neuronalen Netzes kann im Rahmen der zweiten Teilstudie lediglich mit der choose & match-Treatmentgruppe überprüft werden, da die anderen Treatmentgruppen keinen Kontext

ausgewählt haben. Da die choose & match-Gruppe allerdings eine zufällige Auswahl der Gesamtstichprobe darstellt und sich die Treatmentgruppen in den Kontrollvariablen nicht signifikant voneinander unterscheiden, wird angenommen, dass die Performanz des neuronalen Netzes in der choose & match-Gruppe mit der Performanz in den übrigen Treatmentgruppen übereinstimmt. In Tabelle 56 ist die Konfusionsmatrix für die Vorhersage in der choose & match-Treatmentgruppe dargestellt.

Die Vorhersagegüte hat sich im Vergleich zur Zusatzerhebung erneut verschlechtert (F_1-score = .40, Precision = .42, Recall = .41). Hierbei ist die Zuordnung der alltäglichen Kontexte allerdings etwas präziser als in der Zusatzerhebung ($Precision_{All}$ = .64). Weniger gut gelingt die Identifikation der Lernenden, die einen alltäglichen Kontext gewählt haben. Häufig schlägt das neuronale Netz einen innerfachlichen Kontext vor, obwohl tatsächlich ein alltäglicher Kontext gewählt wurde. Der $Recall_{all}$ liegt bei .43. Das neuronale Netz hat zunehmend Schwierigkeiten bei der Vorhersage der besonderen Kontexte. Sowohl die $Precision_{Bes}$ = .36 als auch der $Recall_{Bes}$ = .39 liegen in einem fragwürdigen Bereich. Für die Vorhersage der innerfachlichen Kontexte zeigt sich mit .28 eine etwas schlechtere Precision als in der Zusatzerhebung. Darüber hinaus ist der Recall für die Zuordnung der innerfachlichen Kontexte auf .40 gesunken.

Tabelle 56: Konfusionsmatrix für die Vorhersage der Kontextwahl in der choose & match-Treatmentgruppe mithilfe des neuronalen Netz

Vorhersage	Kontextwahl der Lernenden		
	Alltäglich	Besonders	Innerfachlich
Alltäglich	14	5	3
Besonders	7	9	9
Innerfachlich	12	9	8

Anmerkung: $N = 76$

Unklar ist in diesem Zusammenhang, wodurch die weitere Abnahme der Performanz des neuronalen Netzes zu erklären ist. Hier werden erneut Faktoren von Bedeutung sein, die bereits in Kapitel 9 diskutiert wurden. Insbesondere die deutlich verringerte Reliabilität der Skala zur Erfassung des Wahlmotivs persönlicher Bezug ist an dieser Stelle zu nennen. Zusammenfassend muss berücksichtigt werden, dass die Vorhersage durch das neuronale Netz nicht vollständig mit der tatsächlichen Kontextwahl übereinstimmt. Diese Limitation wird bei der Datenauswertung explizit in Kapitel 10.5.8 berücksichtigt.

10.5.4 Einfluss auf die Zufriedenheit

Die aufgabenbezogene Zufriedenheit wurde sowohl im Verlauf als auch am Ende der kontextorientierten Lerneinheit gemessen.

Zufriedenheit im Verlauf der kontextorientierten Lerneinheit: Da die Messungen der aufgabenbezogenen Zufriedenheit nach jeder Teilaufgabe (Level 1) hierarchisch in der zugehörigen Person verschachtelt sind, wurde zunächst mithilfe der Intraklassen-Korrelation (ICC) bestimmt, in welchem Ausmaß die Unabhängigkeitsannahme verletzt ist (Eid et al., 2017). Dazu wurde zunächst ein Intercept-Only-Modell berechnet, in dem keine festen Effekte, sondern nur ein Random-Intercept für die Personen modelliert wird. Die Personen sind zwar in Klassen verschachtelt, allerdings wurden die Klassen bei der folgenden Analyse nicht als zufälliger Effekt berücksichtigt, da die Anzahl der Klassen mit $N = 19$ unter der empfohlenen Anzahl von 20 liegt (z. B. Singmann & Kellen, 2019). Die ICC liegt bei $\rho = .534$, was darauf hinweist, dass 53.4 % der Gesamtvarianz in der Zufriedenheit auf die spezifische Person zurückzuführen ist. Demzufolge wurde für jede Person ein Random-Intercept berechnet, um die Abhängigkeit der Messungen zu modellieren.

Bei der weiteren Spezifikation der zufälligen Effektstruktur wurde auf die Integration eines Random-Slopes für die Zeit verzichtet, da die Zeit als kategoriale unabhängige Variable modelliert wird. Dies bietet sich an, wenn nur wenige Messzeitpunkte vorliegen und diese sich in einem experimentellen Design qualitativ voneinander unterscheiden (Eid et al., 2017). In diesem Fall liegt für jeden kategorialen Messzeitpunkt nur eine Beobachtung pro Messzeitpunkt vor, so dass die Random-Slope Varianz vollständig mit der zufälligen Fehlervarianz konfundiert ist. Das Modell ist damit nicht identifizierbar (Barr, 2013; Barr et al., 2013). Da keine weiteren Innersubjektfaktoren auf dem ersten Level erhoben wurden, können keine weiteren Random-Slopes in das Modell integriert werden.

Im nächsten Schritt wurden der Messzeitpunkt, das Treatment, der bearbeitete Kontext und die zugehörigen (Cross-level) Interaktionseffekte als feste Effekte in das Modell aufgenommen. Die über die Kenward-Roger-Approximation berechneten Teststatistiken für die Haupt- und Interaktionseffekte sind in Tabelle 57 dargestellt.

Es kann ein signifikanter Haupteffekt der Zeit identifiziert werden (Tab. 57). Unabhängig von dem Treatment oder dem bearbeiteten Kontext sinkt die aufgabenbezogene Zufriedenheit im Verlauf der kontextorientierten Lerneinheit (Abb. 34). Die Zufriedenheit nimmt vom ersten zum zweiten Messzeitpunkt signifikant ab ($b = 0.31$, $t(415) = 7.76$, $p < .001$, $r = .36$), wobei es sich um einen mittleren Effekt handelt. Vom zweiten zum dritten Messzeitpunkt sinkt die Zufriedenheit allerdings nicht weiter, $b = 0.06$, $t(415) = 1.51$, $p = .44$, $r = .07$.

Tabelle 57: Haupteffekte der Zeit, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die aufgabenbezogene Zufriedenheit

Fester Effekt	df	*F*	*p*
Zeit	2, 415.18	48.23	< .001
Treatment	2, 208.00	23.61	< .001
Kontext	2, 208.00	1.33	.267
Zeit × Treatment	4, 415.19	1.74	.140
Zeit × Kontext	4, 415.18	2.15	.074
Treatment × Kontext	4, 208.00	0.56	.694
Zeit × Treatment × Kontext	8, 415.19	2.15	.030

Anmerkung: Die Teststatistiken wurden über die Kenward-Roger-Approximation berechnet.

Ferner kann ein signifikanter Haupteffekt des Treatments ausgemacht werden (Tab. 57). In Abbildung 34 zeigt sich, dass die geschätzte mittlere Zufriedenheit der Lernenden aus der choose & match- und no choose & match-Gruppe höher ist als die Zufriedenheit der Lernenden aus der no choose & no match-Gruppe.

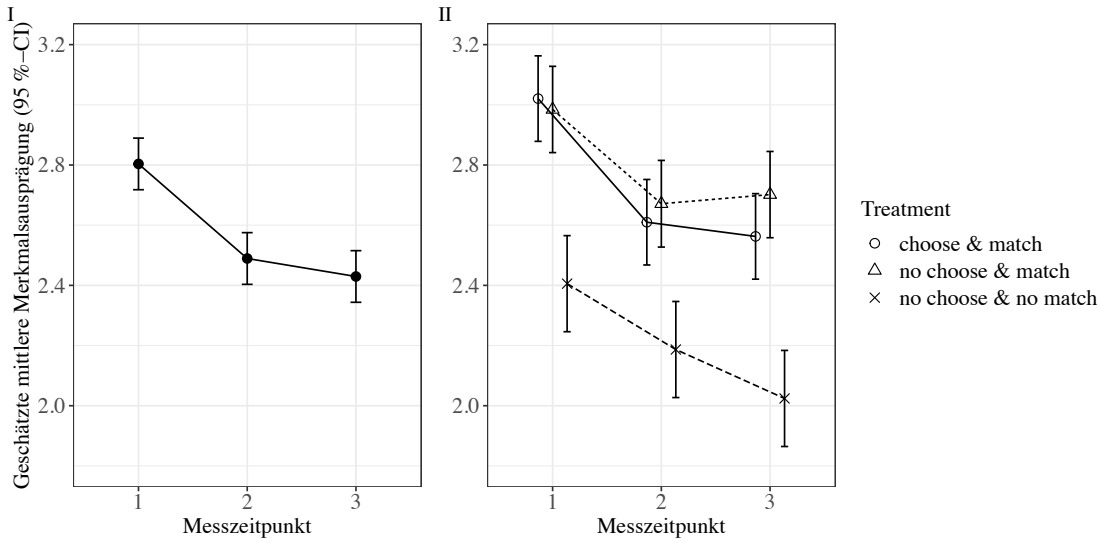


Abbildung 34: Verlauf der geschätzten mittleren Zufriedenheit (I) getrennt nach Treatment (II)[10]

[10] In dieser und den folgenden Abbildungen stellt die y-Achse jeweils nur einen Ausschnitt der gesamten Skala dar, um Unterschiede grafisch hervorzuheben.

Die Kontrastanalyse bestätigt signifikante Unterschiede über alle Messzeitpunkte mit kleinen bis mittleren Effektstärken ($.20 \leq r \leq .30$). Außerdem ist der Unterschied zwischen der choose & match- und no choose & match-Gruppe zu keinem Messzeitpunkt statistisch signifikant, wobei die Unterschiede auch von sehr kleiner Effektstärke sind ($.02 \leq r \leq .07$). Gleichwohl sich die Verläufe der Zufriedenheit über die drei Messzeitpunkte deskriptiv unterscheiden, sind diese Unterschiede statistisch unbedeutend (Tab. 57).

Die signifikante Cross-level Interaktion zwischen der Zeit, dem Treatment und dem Kontext verweist darauf, dass sich der zeitliche Verlauf der Zufriedenheit in Abhängigkeit von Treatment und dem bearbeiteten Kontext unterscheidet (Tab. 57). In Abbildung 35 ist der Verlauf der geschätzten Zufriedenheit über die drei Messzeitpunkte getrennt nach dem Treatment und dem bearbeiteten Kontext dargestellt.

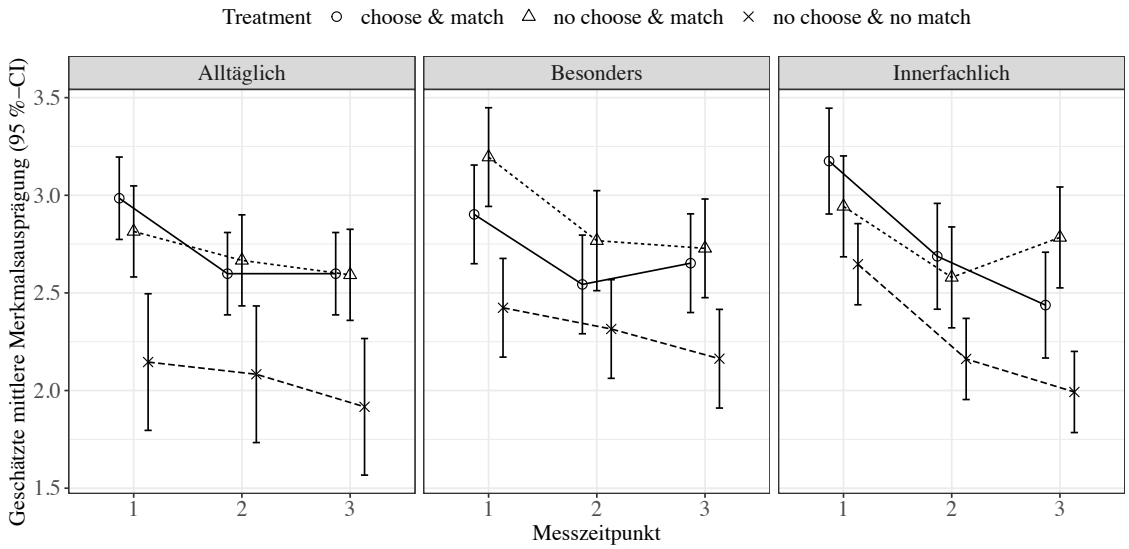


Abbildung 35: Verlauf der geschätzten mittleren Zufriedenheit über die drei Messzeitpunkte getrennt nach Treatment und bearbeiteten Kontext

In der Gruppe der Lernenden, die den alltäglichen Kontext bearbeitet haben, zeigen sich die identifizierten Unterschiede zwischen dem Treatment erneut: Die choose & match- und die no choose & match-Treatmentgruppe zeigen eine höhere Zufriedenheit als die no choose & no match-Gruppe über alle Messzeitpunkte hinweg. Der Verlauf der Zufriedenheit unterscheidet sich nicht in Abhängigkeit vom Treatment.

Ähnliche Ergebnisse zeigen sich auch für die Lernenden, die den besonderen Kontext bearbeitet haben, wobei die Unterschiede nicht ganz so deutlich wie bei den Lernenden mit dem alltäglichen Kontext ausfallen. Der Unterschied zwischen der choose & match und no choose & no match-Gruppe ist zum zweiten Messzeitpunkt statistisch nicht mehr bedeutend, $b = 0.23$,

$t(383) = 1.26$, $p = .63$, $r = .06$. Deskriptiv deutet sich an, dass die Lernenden aus der no choose & match-Gruppe zufriedener sind als die Lernenden aus der choose & match-Gruppe, wobei die Kontraste zu keinem Messzeitpunkt signifikant sind.

Für die Gruppe mit innerfachlichem Kontext sind die identifizierten Unterschiede noch etwas undeutlicher. Zum ersten Messzeitpunkt besteht kein Unterschied zwischen der no choose & match und der no choose & no match-Gruppe, $b = 0.30$, $t(383) = 1.76$, $p = .24$, $r = .09$. Am zweiten und dritten Messzeitpunkt zeigen sich die identifizierten Unterschiede zwischen dem choose & match- und no choose & no match-Treatment sowie dem no choose & match- und no choose & no match-Treatment. Darüber hinaus unterscheidet sich der Verlauf der Zufriedenheit der no choose & match-Gruppe von dem Verlauf der übrigen Treatments. Vom ersten zum dritten Messzeitpunkt steigt die Zufriedenheit der Lernenden, denen ein passender Kontext zugewiesen wurde, während die Zufriedenheit der Lernenden mit der selbstständigen Kontextwahl ($b = 0.58$, $t(415) = 3.24$, $p = .011$, $r = .16$) und nicht passender Kontextzuweisung ($b = -0.50$, $t(415) = -3.15$, $p = .017$, $r = .15$) sinkt.

Um zu überprüfen, inwieweit die Wirksamkeit des Treatments vom individuellen Interesse abhängt, wurden das individuelle Interesse und alle möglichen Interaktionen als fester Faktor in das Modell aufgenommen. Vorab wurde das individuelle Interesse auf den Gesamtmittelwert zentriert (Eid et al., 2017). Sowohl die Zeit ($F(2, 397.69) = 37.21$, $p < .001$) als auch das Treatment ($F(2, 199.54) = 12.28$, $p < .001$) zeigen weiterhin einen signifikanten Haupteffekt. Die Zeit × Treatment × Kontext-Interaktion ist unter Berücksichtigung des individuellen Interesses nicht mehr bedeutsam, $F(8, 398.69) = 1.18$, $p = .31$. Aggregiert über die Zeit, das Treatment und den bearbeiteten Kontext beeinflusst das individuelle Interesse die Zufriedenheit, $F(1, 200.84) = 11.57$, $p = .001$). Ferner wird der Einfluss der Zeit ($F(2, 398.92) = 6.26$, $p = .002$) und des bearbeiteten Kontextes ($F(2, 200.81) = 3.16$, $p = .045$) vom individuellen Interesse beeinflusst. Der Interaktionseffekt des individuellen Interesses mit der Zeit ist in Abbildung 36 dargestellt.

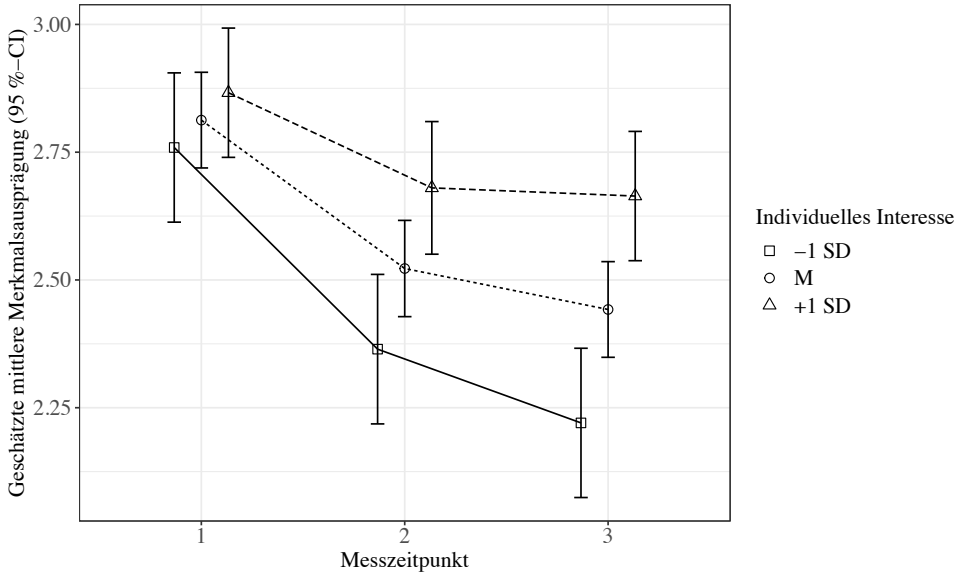


Abbildung 36: Verlauf der geschätzten mittleren aufgabenbezogenen Zufriedenheit über die drei Messzeitpunkte getrennt nach individuellem Interesse

Wenig überraschend unterscheidet sich der Abfall der aufgabenbezogenen Zufriedenheit in Abhängigkeit vom Interesse. Lernende mit einem hohen Interesse (b = 0.20, t(397) = 3.31, p = .003, r = .16) verzeichnen vom ersten zum dritten Messzeitpunkt einen geringeren Abfall als die durchschnittlich (b = 0.37, t(397) = 8.21, p < .001, r = .38) und unterdurchschnittlich Interessierten (b = 0.54, t(397) = 7.65, p < .001, r = .36).

Eine Betrachtung der Zufriedenheit, aggregiert über die Messzeitpunkte und Treatmentgruppen, zeigt, dass die Lernenden in Abhängigkeit von ihrem individuellen Interesse zufriedener mit unterschiedlichen Kontexten sind (Abb. 37). Lernende mit einem hohen Interesse berichten aggregiert über die drei Messzeitpunkte eine höhere Zufriedenheit, wenn sie mit dem besonderen und innerfachlichen Kontext lernen. Sowohl der Kontrast zwischen dem besonderen und alltäglichen (b = -0.33, t(203) = -2.48, p = .042, r = .17) sowie innerfachlichen und alltäglichen Kontext ist statistisch bedeutsam (b = -0.37, t(199) = -2.88, p = .013, r = .20). Der Effekt des Treatments über die Messzeitpunkte wird nicht vom individuellen Interesse beeinflusst, F(4, 398.90) = 0.74, p = .57).

Weiterführend wurde das chemiebezogene Selbstkonzept zentriert und in das Modell integriert. Hierbei wurden sowohl der Haupteffekt als auch mögliche Interaktionen mit der Zeit, dem Treatment und dem bearbeiteten Kontext modelliert. Auch unter Berücksichtigung des chemiebezogenen Selbstkonzeptes ist der Haupteffekt der Zeit (F(2, 397.45) = 41.60, p < .001) und des Treatments (F(2, 199.29) = 15.33, p < .001) wesentlich. Die übrigen Effekte – darunter die Zeit × Treatment × Selbstkonzept-Interaktion – sind dagegen unbedeutend, F(4, 397.87) = 0.60, p = .66.

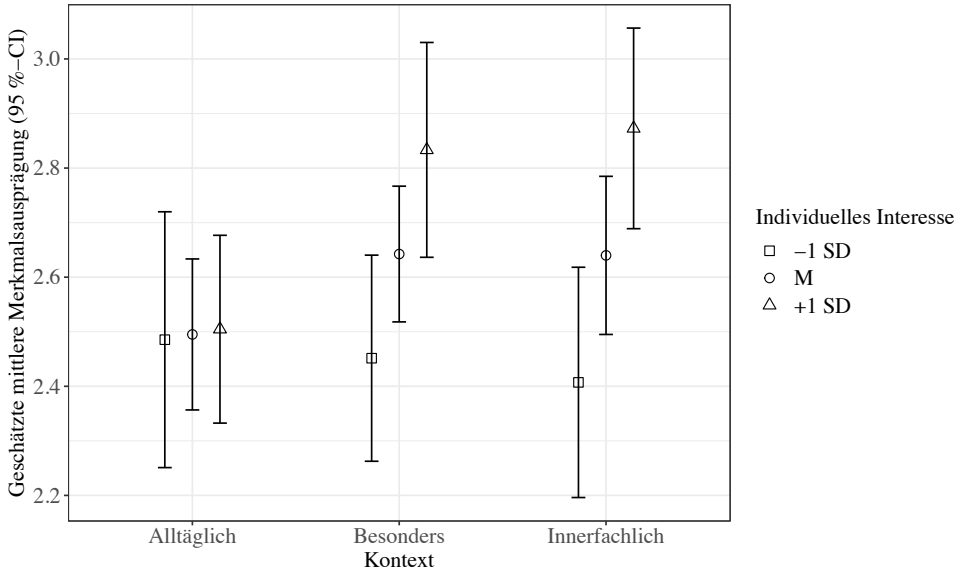


Abbildung 37: Geschätzte mittlere aufgabenbezogene Zufriedenheit, getrennt nach individuellem Interesse und bearbeitetem Kontext, aggregiert über die drei Messzeitpunkte und die Treatmentgruppen

Zufriedenheit am Ende der kontextorientierten Lerneinheit: Am Ende der kontextorientierten Lerneinheit wurden die Lernenden gefragt, ob sie denselben Kontext wieder bearbeiten würden. 40.09 % der Lernenden gaben an, dass sie nach der dritten Teilaufgabe wieder denselben Kontext bearbeiten würden. Ein geringfügig größerer Anteil (43.78 %) gab an, dass sie stattdessen einen anderen Kontext bearbeiten würden. Nur ein kleiner Teil der Lernenden (16.13 %) würde sich mit keinem der Kontexte mehr beschäftigen wollen.

Eine differenzierte Betrachtung der bearbeiteten Kontexte zeigt, dass Lernende, die retrospektiv einen anderen oder keinen der Kontexte bearbeiten würden, im Vergleich zu Lernenden, die denselben Kontext erneut bearbeiten würden, seltener den alltäglichen und häufiger den innerfachlichen Kontext bearbeitet haben (Abb. 38).

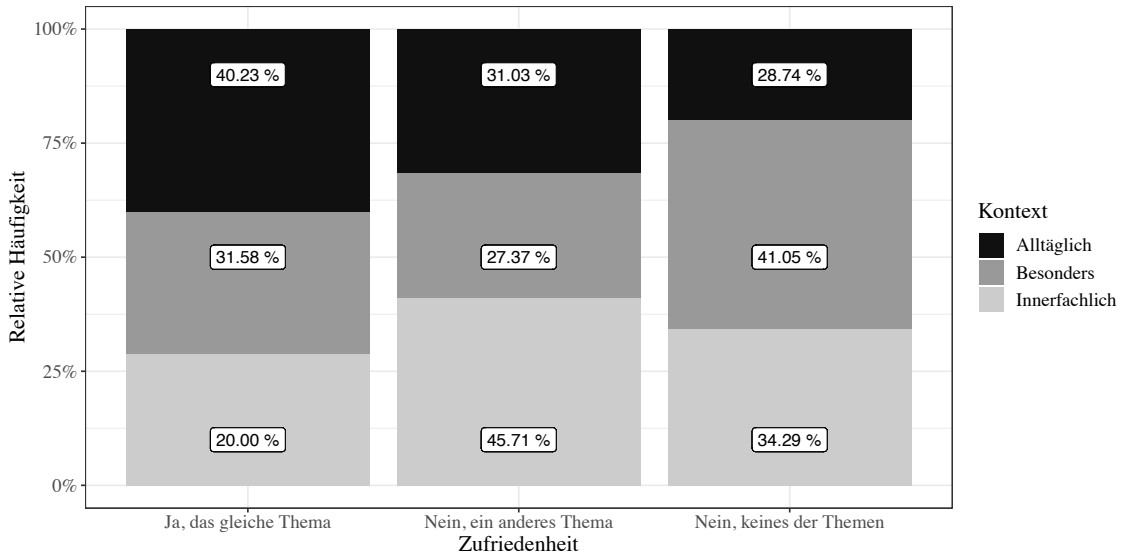


Abbildung 38: Relative Häufigkeit der genannten Antwortalternativen für die Erfassung der Zufriedenheit getrennt nach Kontext

Ein berechneter χ^2-Test zeigt allerdings, dass es keinen Zusammenhang zwischen der erneuten Bearbeitung des Kontextes und dem bearbeiteten Kontext gibt, $\chi^2(4) = 7.92$, $p = .095$, $V = .14$. Demnach handelt es sich nur um einen deskriptiven Unterschied.

Lernende, die rückblickend einen anderen Kontext bearbeiten würden, geben am häufigsten an, dass sie den alltäglichen (37.89 %) oder besonderen Kontext (36.84 %) anstelle des ursprünglich bearbeiteten Kontextes wählen würden. 25.26 % der Lernenden würden stattdessen den innerfachlichen Kontext wählen. Dieses Ergebnis scheint nicht auf die Bearbeitung eines spezifischen Kontextes zurückführbar (Abb. 39). Lernende, die sich den alltäglichen Kontext anstelle des ursprünglich bearbeiteten Kontext wünschen, haben vorher zu ungefähr gleichen Anteilen den besonderen und innerfachlichen bearbeitet. Analoge Ergebnisse finden sich für die Lernenden, die sich den besonderen oder innerfachlichen Kontext für eine erneute Bearbeitung wünschen. Auch hier haben die Lernenden zu ungefähr gleichen Anteilen die jeweils anderen Kontexte bearbeitet.

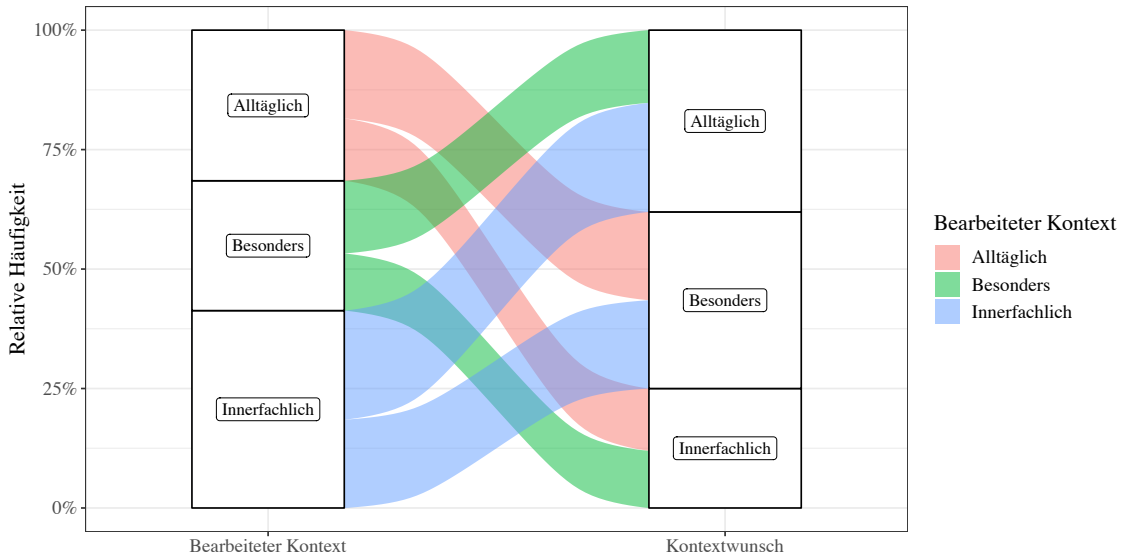


Abbildung 39: Vergleich der relativen Häufigkeit des bearbeiteten Kontext mit der relativen Häufigkeit des retrospektiv gewünschten Kontext für Lernende, die rückblickend einen anderen Kontext bearbeiten würden

Insgesamt hängt die Zufriedenheit am Ende der kontextorientierten Lerneinheit nicht mit der Bearbeitung eines spezifischen Kontextes zusammen. Mögliche Ursachen für die Zufriedenheit am Ende der kontextualisierten Lerneinheit finden sich in den angegebenen Motiven der Lernenden. Nach dem berechneten χ^2-Test besteht ein kleiner bis mittlerer signifikanter Zusammenhang zwischen der Zufriedenheit nach der gesamten Lerneinheit und den angegebenen Zufriedenheitsmotiven, $\chi^2(4) = 25.24$, $p < .001$, $V = .25$. Lernende, die einen anderen oder keinen der Kontexte noch einmal bearbeiten würden, geben zu einem hohen Anteil affektiv-motivationale Motive an (Abb. 40). Im Gegensatz dazu führen Lernende, die denselben Kontext erneut bearbeiten würden, sehr viel häufiger unspezifische Motive (z. B. „alles hat mir gefallen“) statt affektiv-motivationaler Motive für ihre Zufriedenheit mit der Aufgabe an. Ein differenzierter Blick auf die genauen affektiv-motivationalen Motive der Lernenden, die den Kontext noch einmal bearbeiten würden, zeigt ein aufschlussreiches Bild: Jeder vierte Lernende, der einen Kontext aus affektiv-motivationalen Motiven erneut wählen würde, gab kontextspezifisches Interesse am bearbeiteten Kontext an. In der Gruppe der Lernenden, die einen anderen Kontext wählen würden, gaben nur 1.47 % der Lernenden kontextspezifisches Interesse an dem anderen Kontext als Motiv an. Schülerinnen und Schüler begründen die Wahl eines anderen Kontextes dagegen häufiger über unspezifisches Interesse an anderen Kontexten (55.84 %). Lernende, die sich aus affektiv-motivationalen Motiven nicht mit einem der Kontexte beschäftigen möchten, rechtfertigen ihre Entscheidung häufig über unspezifische Interessenmotive (48.00 %). Sie geben auch häufig an, dass ihnen grundsätzlich das Interesse am Fach fehlt (40.00 %), wobei sie die einzige Gruppe von Lernenden sind, die angibt, kein Interesse am Fach zu haben. Kontextspezifisches Interesse spielt hier keine Rolle.

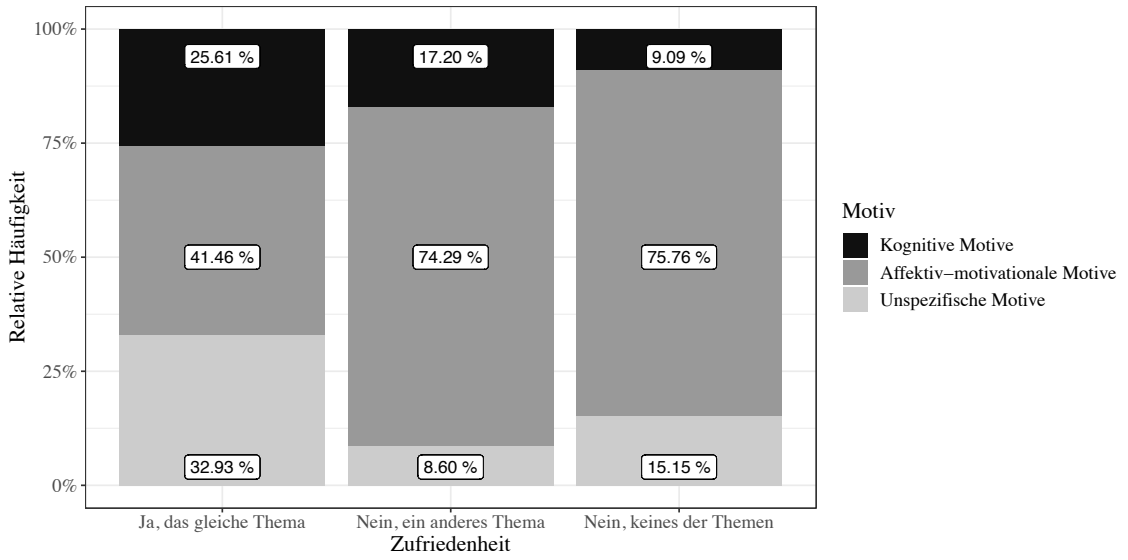


Abbildung 40: Relative Häufigkeit der genannten Motive für die Antwortalternativen zur Erfassung der Zufriedenheit

Eine explorative Analyse hinsichtlich der Merkmale der Lernenden, die den gleichen Kontext erneut, einen anderen Kontext oder keinen der Kontexte bearbeiten würden, zeigt außerdem ein aufschlussreiches Ergebnis: Die Lernenden der drei möglichen Antwortoptionen unterscheiden sich insbesondere im Hinblick auf das individuelle Interesse an Chemie ($F(2, 214) = 3.65$, $p = .028$, $\eta^2 = .03$) sowie das themenspezifische ($F(2, 214) = 7.60$, $p < .001$, $\eta^2 = .07$) und inhaltsfeldspezifische Interesse ($F(2, 214) = 8.29$, $p < .001$, $\eta^2 = .07$). Lernende, die denselben Kontext erneut bearbeiten würden, haben ein höheres individuelles Interesse, ein höheres themenspezifisches Interesse und ein höheres inhaltsfeldspezifisches Interesse als Lernende, die einen anderen oder gar keinen der Kontexte bearbeiten würden. Wie ein Bonferroni-korrigierter Post-hoc-Test zeigt, bestehen diese Unterschiede zwischen allen Gruppen. Es bestehen keine Unterschiede im Hinblick auf das Vorwissen ($F(2, 214) = 1.07$, $p = .35$, $\eta^2 = .00$) oder das chemiebezogene Selbstkonzept ($F(2, 214) = 0.84$, $p = .43$, $\eta^2 = .00$). Demzufolge geben die Lernenden auch zu einem Großteil affektiv-motivationale Gründe (61.54 %) für ihre Antwort an. Kognitive (19.23 %) und unspezifische Motive (19.23 %) wurden seltener genannt.

Vergleicht man die Zufriedenheit nach der kontextualisierten Lerneinheit zwischen den Treatmentgruppen, so verweist ein χ^2-Test auf einen signifikanten Zusammenhang zwischen der Zufriedenheit und der Treatmentgruppe, $\chi^2(4) = 25.24$, $p < .001$, $V = .24$. Schülerinnen und Schüler mit Kontextwahl oder passender Kontextzuordnung geben deutlich häufiger an, denselben Kontext noch einmal bearbeiten zu wollen (Abb. 41). Die Unterschiede zwischen dem choose & match- und no choose & match-Treatment sind dagegen vernachlässigbar.

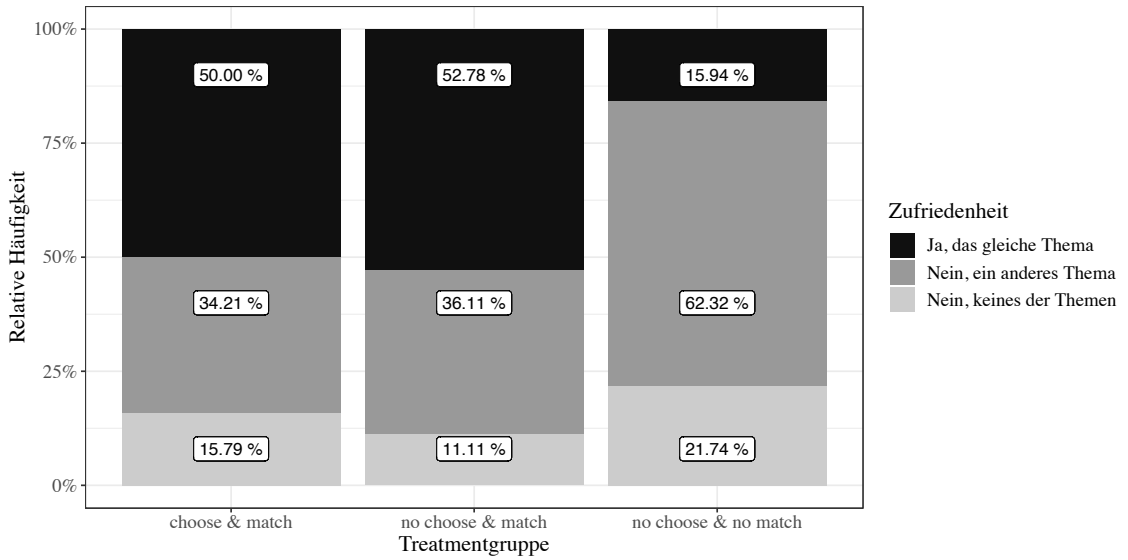


Abbildung 41: Relative Häufigkeit der genannten Antwortalternativen für die Erfassung der Zufriedenheit getrennt nach Treatment

Im Gegensatz dazu würde ein deutlich größerer Anteil der Lernenden mit der unpassenden Kontextzuweisung lieber einen anderen Kontext bearbeiten. Viele dieser Lernenden haben ursprünglich den innerfachlichen Kontext (53.49 %) bearbeitet. Wenige haben den alltäglichen (18.60 %) und besonderen Kontext (27.91 %) bearbeitet. Die meisten Lernenden hätten auch hier einen alltäglichen (41.86 %) oder besonderen Kontext (37.21 %) bevorzugt. Wenige hätten den innerfachlichen Kontext vorgezogen (20.93 %). Ungefähr die Hälfte der Lernenden, die im innerfachlichen Kontext lernten, hätten retrospektiv lieber den alltäglichen Kontext bevorzugt. Die andere Hälfte hätte den besonderen Kontext vorgezogen (Abb. 42).

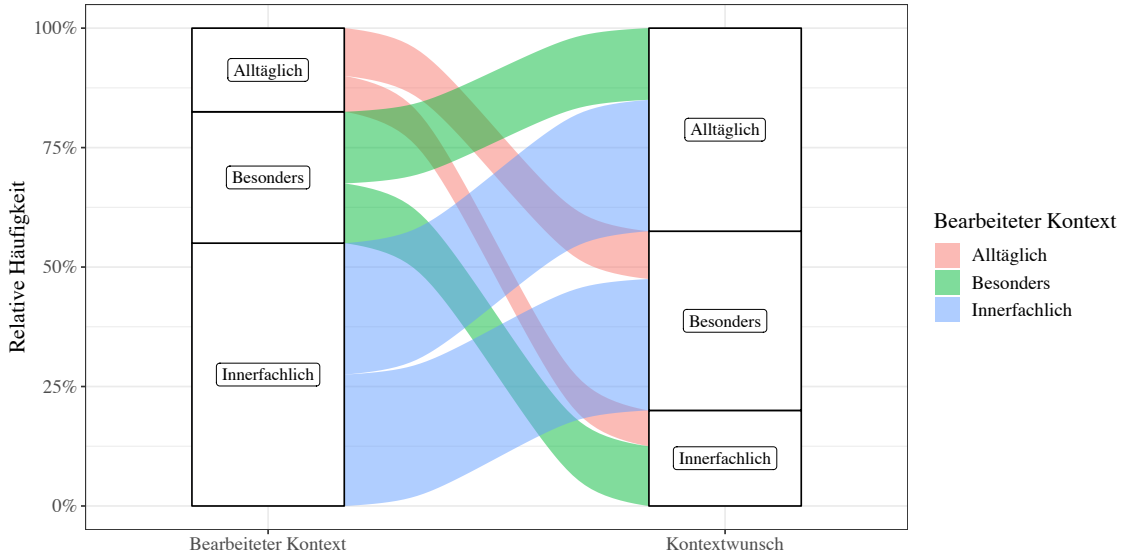


Abbildung 42: Vergleich der relativen Häufigkeit des bearbeiteten Kontext mit der relativen Häufigkeit des retrospektiv gewünschten Kontext für Lernende aus dem no choose & no match-Teatment, die rückblickend einen anderen Kontext bearbeiten würden

Die Ergebnisse verweisen zudem auf einen signifikanten Zusammenhang zwischen dem Treatment und den angegebenen Motiven, $\chi^2(4) = 13.48$, $p = .009$, $V = .18$. Schülerinnen und Schüler, die einen Kontext selbstständig wählen konnten, gaben häufiger kognitive Gründe für ihre Zufriedenheit an (Abb. 43).

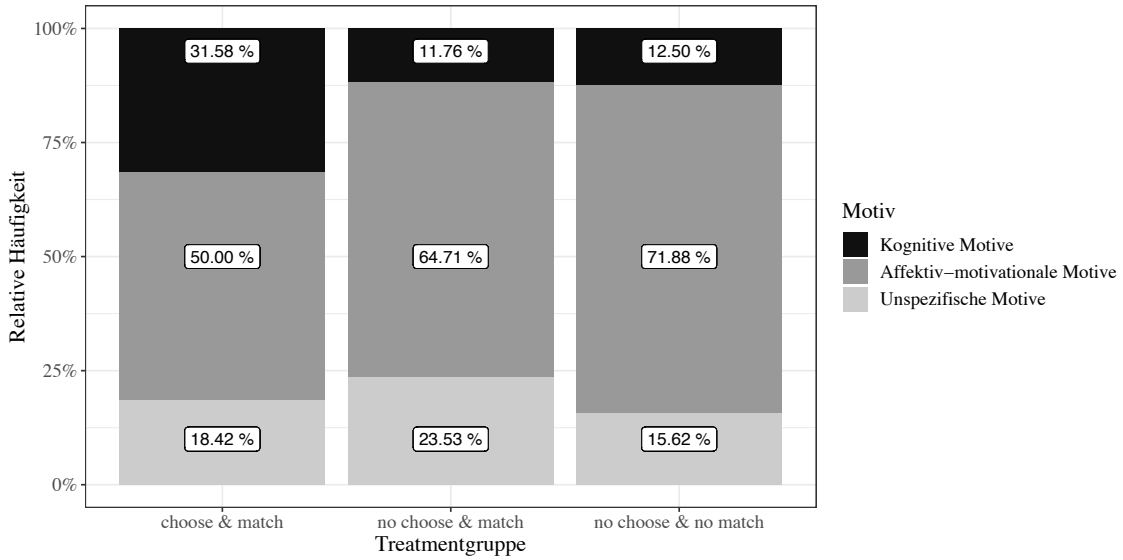


Abbildung 43: Relative Häufigkeit der genannten Motive der Treatmentgruppen

Diese bezogen sich häufiger auf den Erwerb neuen Wissens (68.18 %) als auf das Anforderungsniveau der Aufgaben. In Bezug auf die affektiv-motivationalen Motive finden sich nur geringe Unterschiede zwischen den Treatmentgruppen (Abb. 43).

Dennoch geben 50.00 % der Lernenden aus der choose & match-Gruppe an, dass sie sich rückblickend lieber für einen anderen oder keinen der Kontexte entschieden hätten. Eine explorative Analyse der persönlichen Merkmale zeigt, dass Lernende, die den Kontext erneut wählen würden ($M = 2.50$, $SD = 0.89$), ein höheres inhaltsfeldspezifisches Interesse haben als Lernende, die einen anderen ($M = 2.09$, $SD = 0.86$) oder keinen der Kontexte ($M = 1.81$, $SD = 0.58$) wählen würden, $F(2, 73) = 3.87$, $p = .025$, $\eta^2 = .10$. In Bezug auf weitere Personenmerkmale konnten keine Unterschiede identifiziert werden.

10.5.5 Einfluss auf das situationale Interesse

Gemäß der Ergebnisse der konfirmatorischen Faktorenanalyse wird das situationale Interesse getrennt nach der affektiven und epistemischen Komponente analysiert.

Affektive Valenz: Die durch das Intercept-Only-Modell berechnete ICC liegt bei $\rho = .667$. Dieses Ergebnis deutet darauf hin, dass 66.7 % der Varianz in der affektiven Valenz des situationalen Interesses durch die jeweilige Person zu erklären ist. Aus diesem Grund wurde ein Mehrebenenmodell analog zu dem für die Zufriedenheit mit dem Messzeitpunkt, dem Treatment, dem bearbeiteten Kontext sowie den entsprechenden Interaktionen als feste Effekte und einem Random-Intercept für die Personen als zufälliger Effekt berechnet. Es können signifikante Haupteffekte des Messzeitpunktes und des Treatments identifiziert werden (Tab. 58).

Tabelle 58: Haupteffekte des Messzeitpunktes, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die affektive Valenz

Fester Effekt	df	*F*	*p*
Zeit	2, 415.12	44.95	< .001
Treatment	2, 208.00	14.96	< .001
Kontext	2, 208.00	1.05	.352
Zeit × Treatment	4, 415.13	1.13	.343
Zeit × Kontext	4, 415.13	0.10	.984
Treatment × Kontext	4, 208.00	0.25	.911
Zeit × Treatment × Kontext	8, 415.13	1.12	.346

Anmerkung: Die Teststatistiken wurden über die Kenward-Roger-Approximation berechnet

Die affektive Valenz des situationalen Interesses nimmt vom ersten zum zweiten Messzeitpunkt ab (b = 0.27, $t(415)$ = 7.84, p < .001, r = .36), verändert sich aber vom zweiten zum dritten Messzeitpunkt nicht mehr signifikant, b = 0.02, $t(415)$ = 0.68, p = .98, r = .03 (Abb. 44).

Eine spezifische Analyse der Treatmentgruppen zeigt, dass – wie bei der Zufriedenheit – signifikante Unterschiede in Abhängigkeit vom Treatment bestehen (Abb. 44). Zu allen Messzeitpunkten geben Lernende, die selbst wählen konnten oder denen ein passender Kontext zugewiesen wurde, eine höhere affektive Valenz an als Lernende, denen ein nicht passender Kontext zugewiesen wurde. Die Unterschiede sind von kleiner bis mittlerer Effektstärke ($.16 \leq r \leq .26$). Obwohl sich deskriptiv andeutet, dass die Lernenden der no choose & match-Gruppe eine höhere affektive Valenz zeigen als die Lernenden der choose & match-Gruppe, verweist die Kontrastanalyse zu keinem Messzeitpunkt auf signifikante Unterschiede. Die Effektstärken sind hierbei auch in einem kleinen Bereich ($.03 \leq r \leq .10$). Auch der Interessenabfall unterscheidet sich nicht in Abhängigkeit vom Treatment, was durch die nicht signifikante Cross-level Interaktion zwischen der Zeit und dem Treatment deutlich wird (Tab. 58).

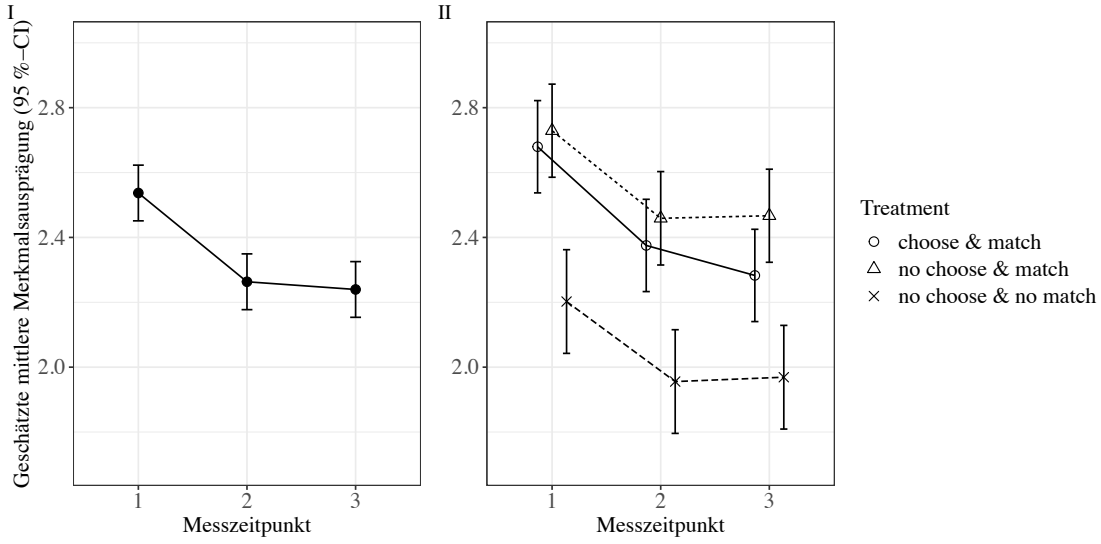


Abbildung 44: Verlauf der geschätzten affektiven Valenz (I), getrennt nach Treatment (II)

Auch unter Berücksichtigung des individuellen Interesses zeigt sich ein signifikanter Haupteffekt der Zeit ($F(2, 397.56)$ = 34.73, p < .001) und des Treatments ($F(2, 199.44)$ = 6.20, p = .002). Weiterhin zeigen sich signifikante Haupteffekte des Interesses ($F(1, 200.50)$ = 37.61, p < .001) sowie signifikante Interaktionseffekte mit der Zeit ($F(2, 398.57)$ = 5.06, p = .007) und dem Kontext ($F(2, 200.48)$ = 6.99, p = .001).

Noch etwas deutlicher als bei der Zufriedenheit zeigen sich unterschiedliche Verläufe der affektiven Valenz in Abhängigkeit vom individuellen Interesse (Abb. 45).

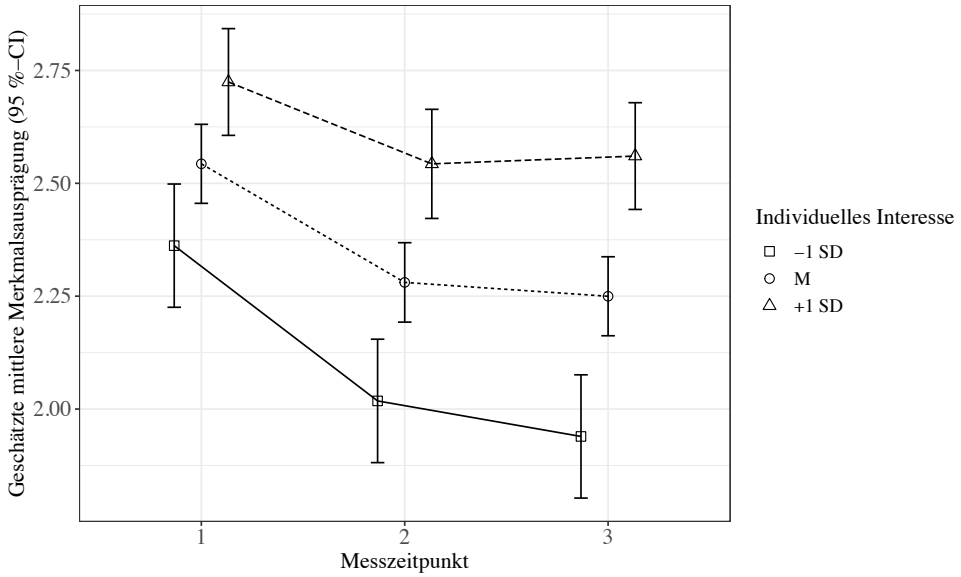


Abbildung 45: Verlauf der geschätzten mittleren affektiven Valenz über die drei Messzeitpunkte, getrennt nach individuellem Interesse

Vom ersten zum dritten Messzeitpunkt nimmt die affektive Valenz der Lernenden mit einem hohen Interesse (b = 0.16, $t(397) = 3.14$, $p < .001$, $r = .16$) weniger stark ab als die affektive Valenz der Lernenden mit einem durchschnittlichen (b = 0.29, $t(397) = 7.59$, $p < .001$, $r = .36$) oder unterdurchschnittlichen Interesse (b = 0.42, $t(397) = 7.01$, $p = .005$, $r = .33$).

Erneut moderiert das individuelle Interesse den Einfluss des Kontexts, wenn die Daten aggregiert über die Messzeitpunkte und die Treatments betrachtet werden (Abb. 46). Bei den durchschnittlich interessierten Schülerinnen und Schülern ist es in Bezug auf das situationale Interesse unerheblich, welcher Kontext bearbeitet wurde. Die Kontexte führen zu einem vergleichbaren Interesse. Für die Hochinteressierten belegt die Kontrastanalyse allerdings, dass die Lernenden mit besonderem Kontext ein höheres situationales Interesse zeigen als Lernende mit alltäglichem Kontext, b = -0.31, $t(203) = -2.45$, $p = .045$, $r = .17$. Der Kontrast zwischen den Lernendengruppen mit innerfachlichem und alltäglichem Kontext ist dagegen vernachlässigbar (b = -0.24, $t(199) = -1.99$, $p = .14$, $r = .14$), obwohl ein deskriptiver Unterschied besteht. Ein entgegengesetztes Bild zeigt sich für die unterdurchschnittlich Interessierten: Diese Lernenden profitieren von dem alltäglichen verglichen mit dem innerfachlichen Kontext (b = 0.47, $t(199) = 3.04$, $p = .008$, $r = .21$). Die unterdurchschnittlich Interessierten

scheinen im Vergleich zum besonderen Kontext vom alltäglichen Kontext zu profitieren. Allerdings handelt es sich dabei lediglich um einen deskriptiven Unterschied, wie die Kontrastanalyse zeigt, $b = 0.27$, $t(199) = 1.82$, $p = .21$, $r = .13$.

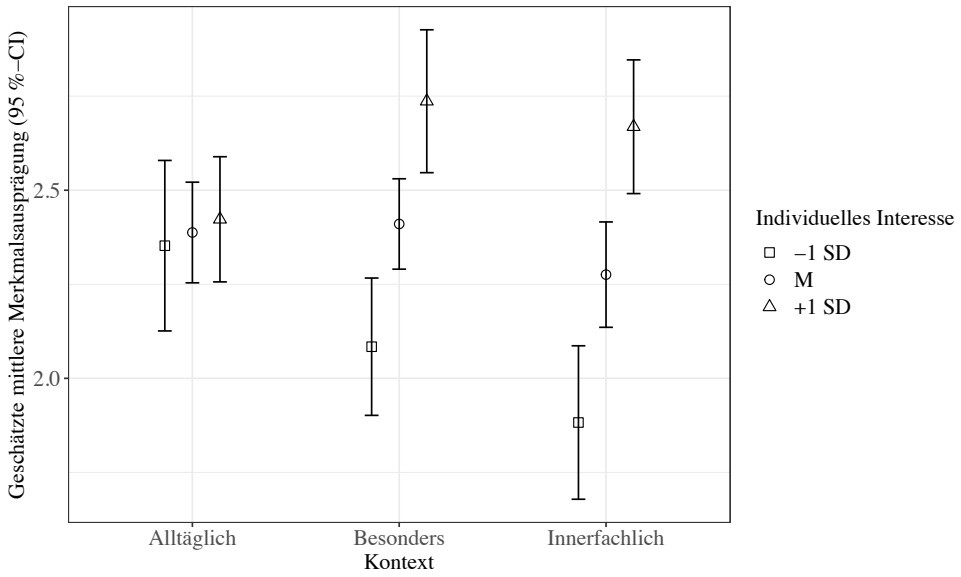


Abbildung 46: Geschätztes mittlere affektive Valenz getrennt nach individuellem Interesse und bearbeiteten Kontext, aggregiert über die drei Messzeitpunkte und die Treatmentgruppen

Werden das chemiebezogene Selbstkonzept und mögliche Interaktionseffekte in das Modell einbezogen, so lassen sich weder ein Haupteffekt des Selbstkonzepts noch Interaktionseffekte mit dem Selbstkonzept nachweisen. Auch unter Berücksichtigung des Selbstkonzepts erweisen sich die Haupteffekte der Zeit ($F(2, 397.31) = 34.45$, $p < .001$) und des Treatments ($F(2, 199.20) = 9.38$, $p < .001$) als substanziell.

Epistemische Komponente: Für die epistemische Komponente des situationalen Interesses wurde eine ICC von $\rho = .747$ ermittelt, die sogar über der ICC für die affektive Valenz des situationalen Interesses liegt. Folglich wurde erneut ein Mehrebenenmodell mit einem Random-Intercept für die Personen berechnet. Die berechneten Teststatistiken für das Mehrebenenmodell zur Analyse der epistemischen Komponente des situationalen Interesses sind in Tabelle 59 dargestellt.

Tabelle 59: Haupteffekte der Zeit, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die epistemische Komponente

Fester Effekt	df	*F*	*p*
Zeit	2, 415.10	12.25	< .001
Treatment	2, 208.00	6.08	.003
Kontext	2, 208.00	0.86	.425
Zeit × Treatment	4, 415.10	0.87	.481
Zeit × Kontext	4, 415.10	1.62	.167
Treatment × Kontext	4, 208.00	0.20	.936
Zeit × Treatment × Kontext	8, 415.10	0.60	.775

Anmerkung: Die Teststatistiken wurden über die Kenward-Roger-Approximation berechnet.

Der signifikante Haupteffekt der Zeit verweist auf Unterschiede zwischen den Messzeitpunkten (Abb. 47). Nach Bearbeitung der ersten Teilaufgabe nimmt die epistemische Komponente signifikant ab ($b = 0.13$, $t(415) = 4.19$, $p < .001$, $r = .20$), wobei nach der zweiten Teilaufgabe keine weitere Abnahme zu verzeichnen ist ($b = 0.01$, $t(415) = 0.17$, $p = .99$, $r = .01$.

Die getrennte Analyse nach Treatmentgruppen zeigt den bereits identifizierten Unterschied zwischen den Treatmentgruppen an (Abb. 47), der auch durch den signifikanten Haupteffekt des Treatments als wesentlich eingestuft werden kann (Tab. 59).

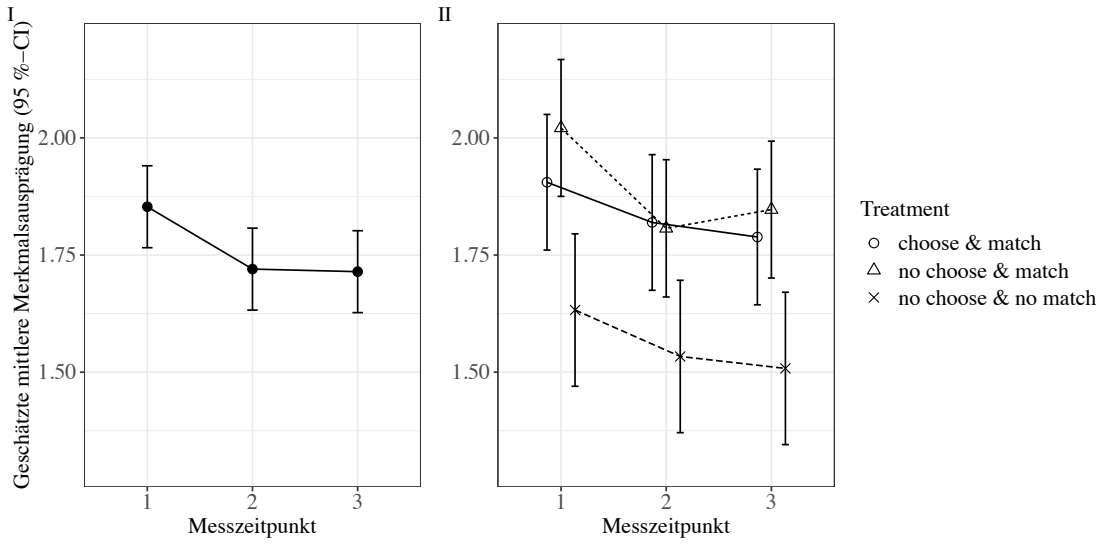


Abbildung 47: Verlauf der geschätzten mittleren epistemischen Komponente (I) getrennt nach Treatment (II)

Die für jeden Messzeitpunkt durchgeführten Kontrastanalysen bestätigen die bereits mehrfach identifizierten Unterschiede zwischen dem choose & match- und no choose & match- sowie dem no choose & match und no choose & no match-Treatment mit überwiegend kleinen Effektstärken ($.14 \leq r \leq .20$).

Durch den Einbezug des individuellen Interesses und entsprechender Interaktionseffekte wird der Haupteffekt des Treatments statistisch vernachlässigbar, $F(2, 199.35) = 2.49$, $p = .085$. Der signifikante Haupteffekt des individuellen Interesses ($F(1, 200.17) = 32.20$, $p < .001$) und die nahezu statistisch bedeutsame Treatment × Interesse-Interaktion ($F(1, 200.22) = 3.04$, $p = .050$) deutet viel mehr an, dass der Effekt des Treatments abhängig vom individuellen Interesse der Lernenden ist (Abb. 48).

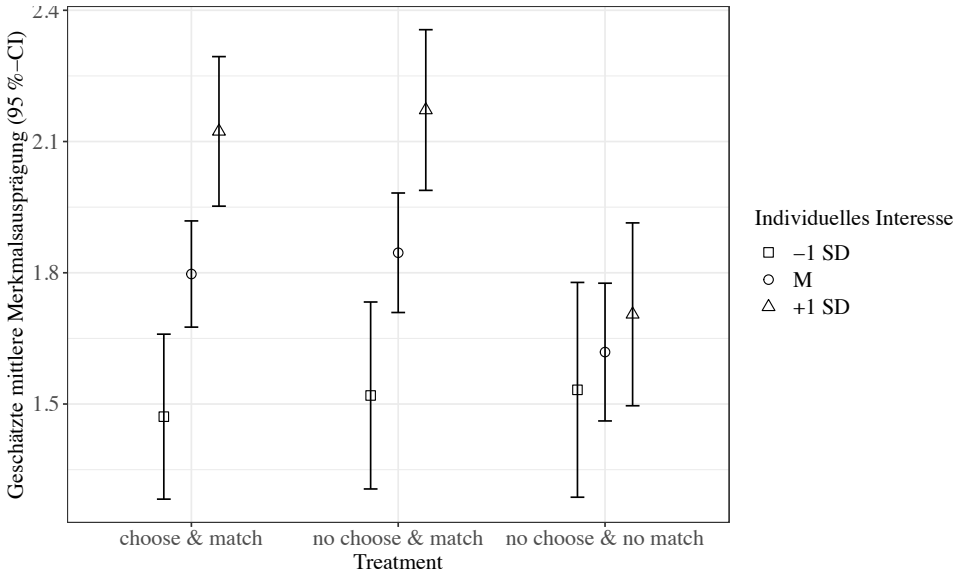


Abbildung 48: Geschätzte mittlere epistemische Komponente, getrennt nach individuellem Interesse und der Treatmentgruppe

Aggregiert man die epistemische Komponente des situationalen Interesses über die drei Messzeitpunkte und die bearbeiteten Kontexte, belegen die Kontrastanalysen, dass lediglich die Hochinteressierten nach der Bearbeitung einer passenden Kontextaufgabe eine höhere epistemische Komponente zeigen. Demzufolge bestehen bei den Hochinteressierten signifikante Unterschiede zwischen dem choose & match- und no choose & no match-Treatment ($b = 0.42$, $t(199) = 3.05$, $p = .008$, $r = 21$) sowie dem no choose & match- und no choose & no match-Treatment, $b = 0.47$, $t(199) = 3.31$, $p = .003$, $r = 23$. In den Gruppen der unterdurchschnittlich und durchschnittlich Interessierten können diese Unterschiede nicht nachgewiesen werden.

Daneben kann in dem spezifizierten Modell mit dem individuellen Interesse ein signifikanter Interaktionseffekt zwischen dem bearbeiteten Kontext und dem individuellen Interesse ausgemacht werden, $F(2, 200.16) = 4.52$, $p = .012$. In der Gruppe der unterdurchschnittlich Interessierten führt die Bearbeitung des alltäglichen Kontexts zu einem höheren epistemischen Interesse als die Bearbeitung des innerfachlichen Kontexts ($b = 0.47$, $t(199) = 2.90$, $p = .012$, $r = .20$). Gleichwohl die Bearbeitung des besonderen Kontexts deskriptiv zu einem geringeren epistemischen Interesse führt (Abb. 49), ist der Unterschied statistisch nicht relevant, $b = 0.33$, $t(199) = 2.12$, $p = .11$, $r = .15$. Demgegenüber zeigen überdurchschnittlich interessierte Schülerinnen und Schüler eine höhere epistemische Komponente des situationalen Interesses, wenn sie mit dem besonderen oder innerfachlichen Kontext lernen. Dies ist jedoch nur ein deskriptiver Effekt (Abb. 49), wie die Kontrastanalyse bestätigt.

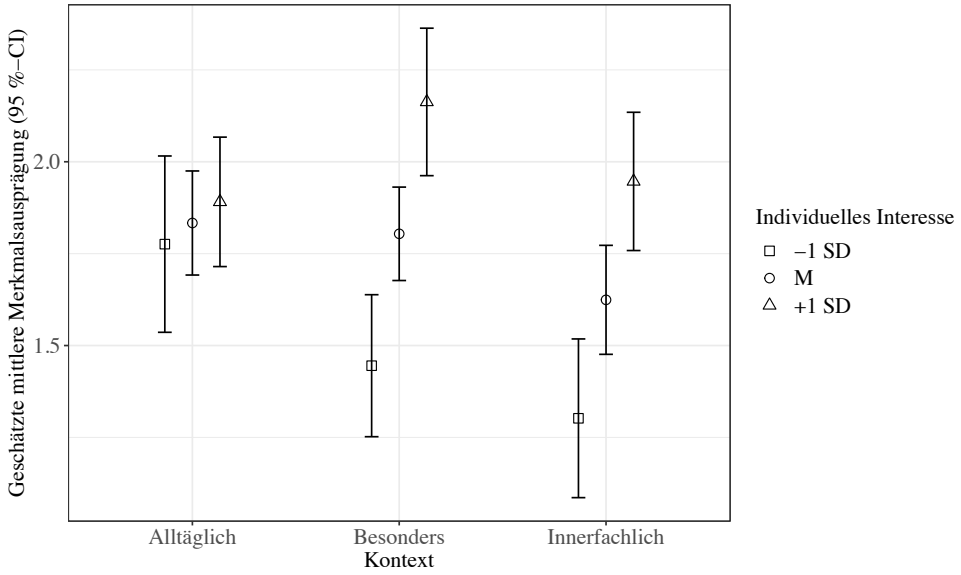


Abbildung 49: Geschätzte mittlere epistemische Komponente, getrennt nach individuellem Interesse und dem bearbeiteten Kontext

Ein weiteres Modell wird unter Berücksichtigung des Selbstkonzepts und möglicher Interaktionseffekte berechnet. In dem Modell zeigt sich weiterhin ein signifikanter Haupteffekt der Zeit ($F(2, 397.26) = 7.45$, $p < .001$), jedoch ist der Haupteffekt des Treatments statistisch irrelevant ($F(2, 199.17) = 2.98$, $p = .053$). Ferner kann ein signifikanter Haupteffekt des Selbstkonzepts ausgemacht werden, $F(1, 199.41) = 5.69$, $p = .018$. Je höher das Selbstkonzept ist, desto höher ist das epistemische situationale Interesse, $b = 0.17$, $t(199.22) = 2.39$, $p = .018$, $r = .17$). Alle übrigen Interaktionseffekte sind statistisch nicht bedeutsam.

10.5.6 Einfluss auf die kognitive Belastung

Die kognitive Belastung wurde durch die investierte Denkanstrengung und die wahrgenommene Aufgabenschwierigkeit operationalisiert. Dementsprechend erfolgt die Analyse der Daten separat für die investierte Denkanstrengung und die wahrgenommene Aufgabenschwierigkeit.

Investierte Denkanstrengung: Für die investierte Denkanstrengung wurde ein ICC von .084 identifiziert, der im Vergleich zu den übrigen Konstrukten geringer ausfällt. Dennoch wurde für die Personen ein Random-Intercept spezifiziert, da die Nichtberücksichtigung bestehender Abhängigkeiten in den Daten die Wahrscheinlichkeit eines Typ-I-Fehlers erhöht (z. B. Singmann & Kellen, 2019). Als feste Effekte wurden abermals die Zeit, das Treatment, der bearbeitete Kontext sowie mögliche Interaktionen berücksichtigt (Tab. 60).

Tabelle 60: Haupteffekte des Messzeitpunktes, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die investierte Denkanstrengung

Fester Effekt	df	*F*	*p*
Zeit	2, 404.64	161.84	< .001
Treatment	2, 207.91	3.22	.042
Kontext	2, 207.92	0.38	.687
Zeit × Treatment	4, 404.66	0.80	.525
Zeit × Kontext	4, 404.67	0.38	.823
Treatment × Kontext	4, 207.85	1.12	.349
Zeit × Treatment × Kontext	8, 404.59	0.92	.497

Anmerkung: Die Teststatistiken wurden über die Kenward-Roger-Approximation berechnet

Wie der signifikante Haupteffekt der Zeit andeutet, verändert sich die investierte Denkanstrengung im Verlauf der kontextorientierten Lerneinheit (Abb. 50). Nach Bearbeitung der zweiten Teilaufgabe liegt die investierte Denkanstrengung weitaus höher als nach Bearbeitung der ersten Teilaufgabe, b = -2.05, $t(400)$ = -15.47, $p < .001$, r = .61. Der Unterschied entspricht einem großen Effekt. Die Bearbeitung der dritten Teilaufgabe führt allerdings zu keinem weiteren Anstieg der investierten Denkanstrengung, b = -0.06, $t(407)$ = -0.45, p = .98, r = .02.

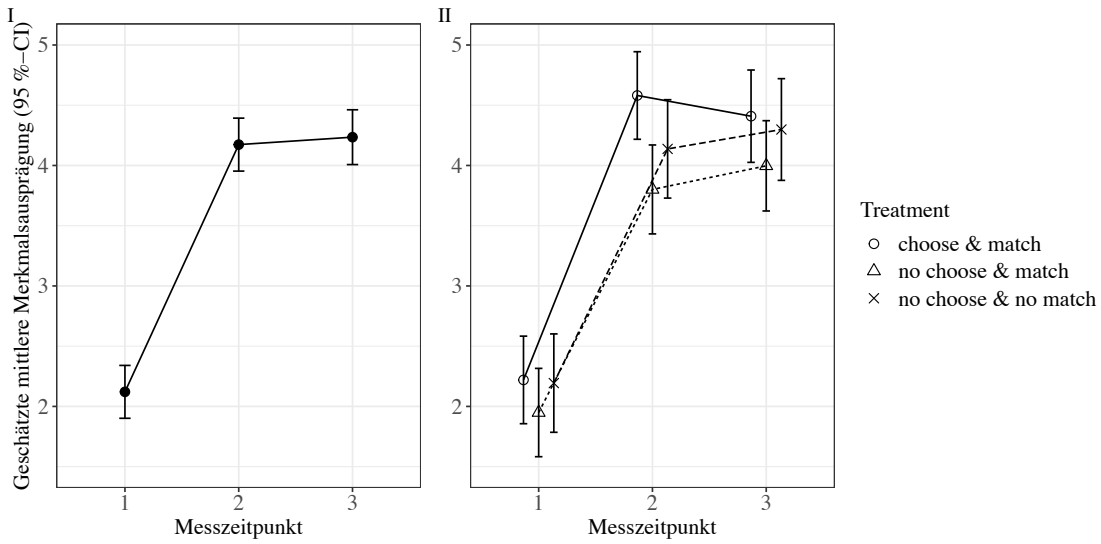


Abbildung 50: Verlauf der geschätzten mittleren investierten Denkanstrengung (I) getrennt nach Treatment (II)

Neben einem signifikanten Haupteffekt der Zeit, zeigt sich ein bedeutender Effekt des Treatments (Tab. 60). Wie die Kontrastanalyse belegt, geben Lernende mit Kontextwahl zum zweiten Messzeitpunkt eine höhere investierte Denkanstrengung an als Lernende mit passender Kontextzuweisung, $b = 0.78$, $t(520) = 2.96$, $p = .010$, $r = .13$.

Um zu überprüfen, ob der in Abbildung 50 dargestellte Treatmenteffekt vom individuellen Interesse an Chemie abhängt, wurden das individuelle Interesse sowie mögliche Interaktionen in das Modell eingeführt. Erneut können lediglich signifikante Haupteffekte der Zeit ($F(2, 388.72) = 123.85$, $p < .001$) und des Treatments ($F(2, 201.14) = 4.73$, $p = .010$) identifiziert werden. Weitere Interaktionseffekte sind statistisch nicht relevant.

In einem weiteren Schritt wurden Haupt- und mögliche Interaktionseffekte des chemiebezogenen Selbstkonzepts überprüft, indem das Selbstkonzept in das Modell integriert wurde. Sowohl der Haupteffekt der Zeit ($F(2, 388.60) = 119.01$, $p < .001$) als auch der Haupteffekt des Treatments bleibt signifikant, $F(2, 200.89) = 3.08$, $p = .048$. Außerdem ist der Haupteffekt des Selbstkonzepts statistisch signifikant ($F(1, 197.70) = 12.11$, $p < .001$). Die detaillierte Analyse der geschätzten durchschnittlichen investierten Denkanstrengung zeigt eine Abnahme der investierten Denkanstrengungen mit steigendem Selbstkonzept, sofern die investierte Denkanstrengung über die Messzeitpunkte, das Treatment und den bearbeiteten Kontext aggregiert wird, $b = -0.51$, $t(197.51) = -3.48$, $p < .001$, $r = .24$.

Wahrgenommene Aufgabenschwierigkeit: Der berechnete ICC von .271 impliziert, dass 27.1 % der Variation in der wahrgenommenen Aufgabenschwierigkeit auf die spezifische Person zurückzuführen sind. Demzufolge wurde ein Random-Intercept für die spezifische

Person in die zufällige Effektstruktur des Modells integriert. Daneben wurden die Haupteffekte der Zeit, des Treatments und des Kontexts sowie deren Interaktionseffekte als feste Effekte modelliert. Die Teststatistiken sind in Tabelle 61 dargestellt.

Tabelle 61: Haupteffekte der Zeit, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die wahrgenommene Aufgabenschwierigkeit

Fester Effekt	df	*F*	*p*
Zeit	2, 404.18	61.38	< .001
Treatment	2, 207.98	0.44	.644
Kontext	2, 207.99	0.72	.486
Zeit × Treatment	4, 404.20	0.17	.953
Zeit × Kontext	4, 404.21	0.41	.803
Treatment × Kontext	4, 207.93	1.48	.210
Zeit × Treatment × Kontext	8, 404.14	0.24	.972

Anmerkung: Die Teststatistiken wurden über die Kenward-Roger-Approximation berechnet

Wie schon bei der investierten Denkanstrengung, kann ein signifikanter Haupteffekt der Zeit identifiziert werden (Tab. 61). Abbildung 51 zeigt den Verlauf der geschätzten mittleren wahrgenommenen Aufgabenschwierigkeit über die drei Teilaufgaben. Die wahrgenommene Aufgabenschwierigkeit steigt – wie schon die investierte Denkanstrengung – signifikant und mit großem Effekt von der ersten zur zweiten Teilaufgabe, $b = -1.25$, $t(400) = -9.55$, $p < .001$, $r = .43$. Die wahrgenommene Aufgabenschwierigkeit ändert sich jedoch nicht von der zweiten zur dritten Teilaufgabe, $b = -0.03$, $t(407) = -0.24$, $p = .99$, $r = .01$. Der Verlauf der geschätzten mittleren Aufgabenschwierigkeit unterscheidet sich nicht in Abhängigkeit des Treatments (Abb. 51).

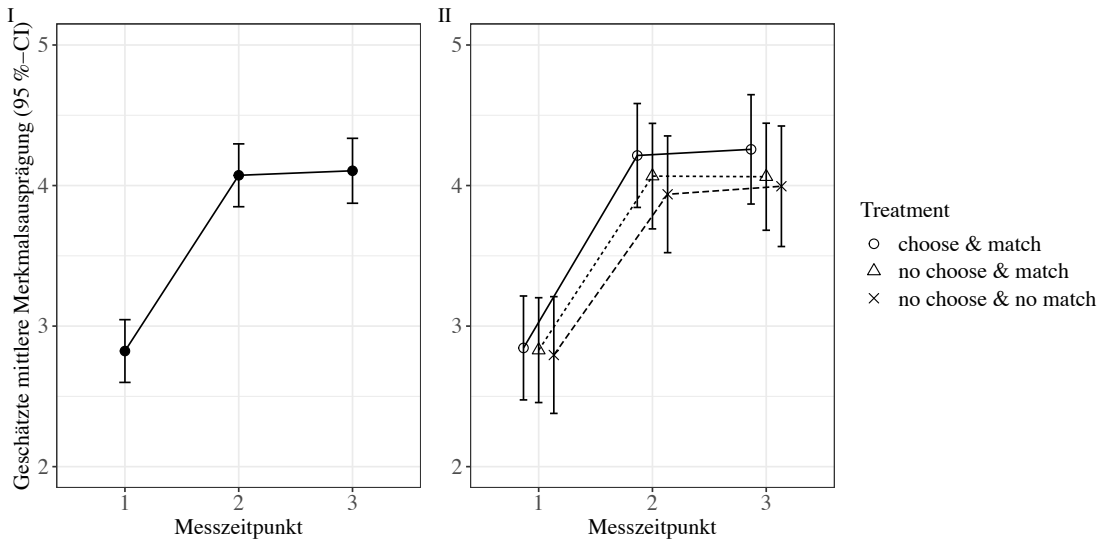


Abbildung 51: Verlauf der geschätzten mittleren wahrgenommenen Aufgabenschwierigkeit (I) getrennt nach Treatment (II)

Im Folgenden wurden außerdem Interaktionseffekte mit dem individuellen Interesse untersucht. Im Modell mit einem Haupteffekt des Interesses und möglichen Interaktionseffekten erweist sich wiederum der Haupteffekt der Zeit als signifikant, $F(2, 388.05) = 57.30$, $p < .001$). Ferner ist der Effekt des Treatments abhängig von dem individuellen Interesse der Lernenden, $F(2, 200.96) = 3.69$, $p = .027$. Hier zeigen die berechneten Kontrastanalysen, dass die Wahl einer Kontextaufgabe bei den unterdurchschnittlich Interessierten zu einer höheren über die Messzeitpunkte aggregierten wahrgenommenen Aufgabenschwierigkeit führt als die passende Zuweisung, $b = 0.89$, $t(201) = 2.65$, $p = .026$, $r = .18$. Es gibt keine Unterschiede zwischen den Lernenden mit selbstständiger Kontextwahl und nicht passenden Kontextzuweisung ($b = 0.63$, $t(200) = 1.73$, $p = .25$, $r = .12$) noch zwischen Lernenden mit einer passenden und nicht passenden Kontextzuweisung, $b = -0.26$, $t(201) = -0.67$, $p = .98$, $r = .05$. Ferner zeigen sich keine signifikanten Unterschiede bei den durchschnittlich und überdurchschnittlich Interessierten in Abhängigkeit vom Treatment (Abb. 52). Insgesamt sind die identifizierten Effekte damit schwierig interpretierbar.

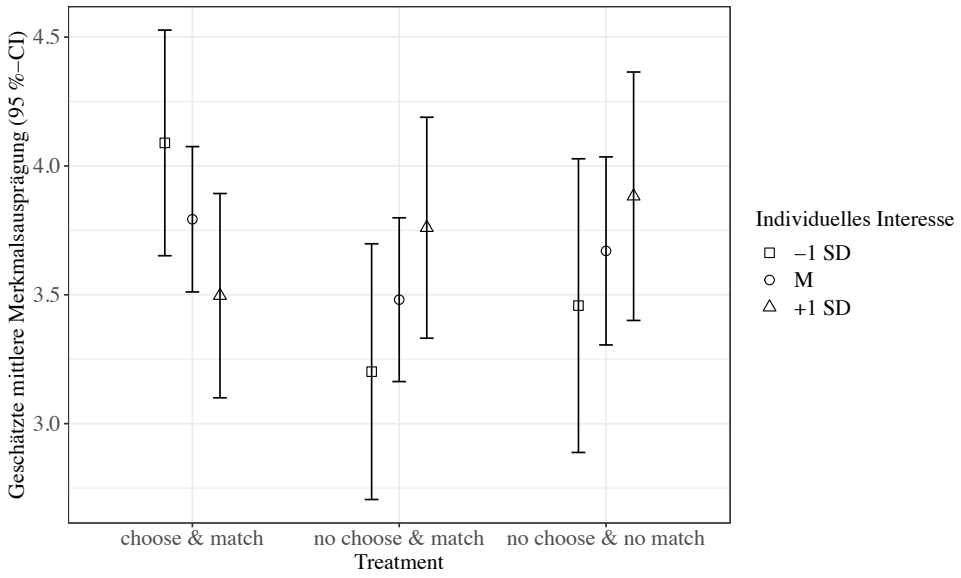


Abbildung 52: Geschätzte mittlere wahrgenommene Aufgabenschwierigkeit getrennt nach individuellem Interesse und der Treatmentgruppe

Die signifikante Zeit × Treatment × Interesse-Interaktion legt außerdem nahe (Abb. 53), dass die Effekte des Treatments und des individuellen Interesses über den zeitlichen Verlauf der Lerneinheit variieren, $F(4, 388.06) = 2.58$, $p = .037$). Auch hier weist die Analyse der Bonferroni-korrigierten Kontraste auf Unterschiede in der Gruppe der unterdurchschnittlich Interessierten hin . Die Lernenden mit Kontextwahl geben zum zweiten Messzeitpunkt eine höhere Aufgabenschwierigkeit als die Lernenden mit nicht passender Kontextzuweisung an, $b = 1.37$, $t(478) = 2.82$, $p = .015$, $r = .13$. Der Unterschied zwischen den Lernenden mit Kontextwahl und passender Kontextzuweisung ist dagegen zum zweiten Messzeitpunkt statistisch nicht bedeutsam, $b = 1.05$, $t(481) = 2.35$, $p = .057$, $r = .11$). Zu den anderen Messzeitpunkten zeigen sich keine signifikanten Unterschiede. Überdies ist der Verlauf der wahrgenommenen Aufgabenschwierigkeit bei den durchschnittlich und überdurchschnittlich Interessierten nicht durch das Treatment beeinflusst. Insgesamt sind die Befunde damit schwierig interpretierbar.

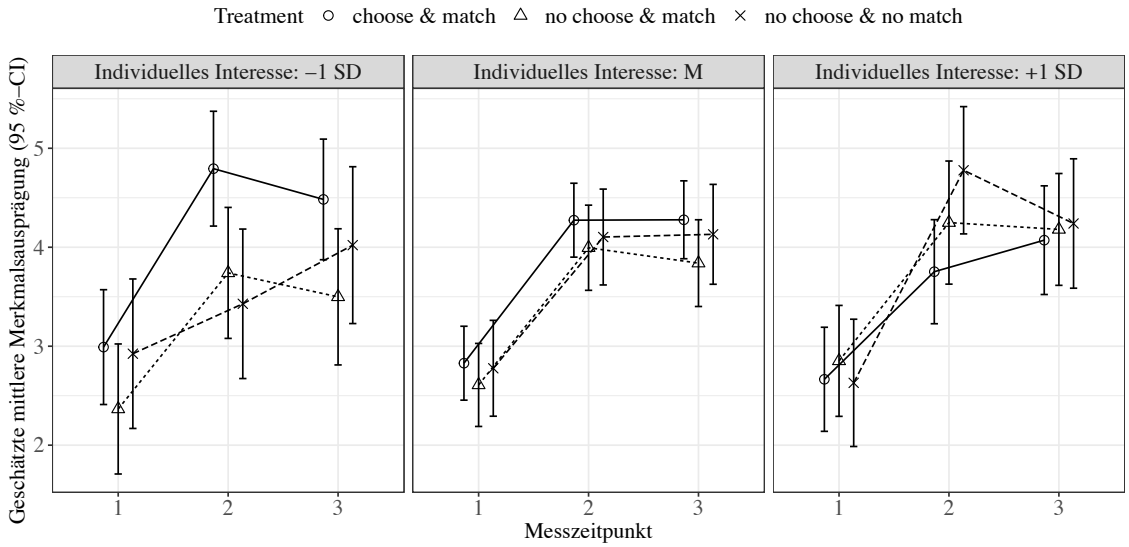


Abbildung 53: Verlauf der geschätzten mittleren wahrgenommenen Aufgabenschwierigkeit getrennt nach Treatment und individuellem Interesse

Die Frage nach dem Einfluss des Selbstkonzeptes auf die Wirksamkeit des Treatments wurde durch die Modellierung eines Interaktionseffekts untersucht. Es zeigt sich ein signifikanter Haupteffekt des Selbstkonzeptes, $F(1, 202.06) = 7.86$, $p = .006$. Auch hier deutet sich an, dass die wahrgenommene Aufgabenschwierigkeit mit steigendem Selbstkonzept abnimmt, $b = -0.39$, $t(198.40) = -2.49$, $p = .014$, $r = .17$. Die übrigen Interaktionseffekte sind dagegen unbedeutend.

10.5.7 Einfluss auf den Fachwissenszuwachs

Der Fachwissenszuwachs ergibt sich aus der Differenz zwischen den IRT-skalierten Personenparametern zum Prä- und Post-Zeitpunkt. 26.5 % der Gesamtvarianz im Fachwissenszuwachs sind auf die spezifische Person zurückzuführen, wie die durch das Intercept-Only-Modell berechnete ICC zeigt. Folglich wird ein Mehrebenenmodell mit einem Random-Intercept für die Personen, mit Haupteffekten für den Messzeitpunkt, das Treatment und den bearbeiteten Kontext sowie möglichen Interaktionseffekten berechnet (Tab. 62).

Wie in Tabelle 62 dargestellt, kann ein signifikanter Haupteffekt der Zeit auf das Fachwissen im Inhaltsfeld *saure und alkalische Lösungen* ausgemacht werden. Die Kontrastanalyse verweist auf einen signifikanten Unterschied zwischen dem Prä- und Post-Zeitpunkt, $b = 0.45$, $t(208) = 6.23$, $p < .001$, $r = .40$. Unabhängig von der Gruppenzugehörigkeit und dem bearbeiteten Kontext hat sich das Fachwissen damit um 0.45 Logits verbessert, was auf einen mittleren bis großen Effekt hindeutet (siehe Abb. 54).

Tabelle 62: Haupteffekte der Zeit, des Treatments und des Kontexts sowie deren Interaktionseffekte auf das Fachwissen im Inhaltsfeld saure und alkalische Lösungen

Fester Effekt	df	*F*	*p*
Zeit	1, 208	38.78	< .001
Treatment	2, 208	0.26	.771
Kontext	2, 208	0.70	.498
Zeit × Treatment	2, 208	0.66	.518
Zeit × Kontext	2, 208	2.58	.078
Treatment × Kontext	4, 208	1.25	.291
Zeit × Treatment × Kontext	4, 208	0.86	.492

Anmerkung: Die Teststatistiken wurden über die Kenward-Roger-Approximation berechnet

Es gibt keine signifikanten Haupteffekte des Treatments und des Kontexts (Tab. 62). Betrachtet man den Fachwissenszuwachs getrennt nach dem Treatment, so scheint es keinen deskriptiven Effekt des Treatments zum Post-Zeitpunkt zu geben (Abb. 54).

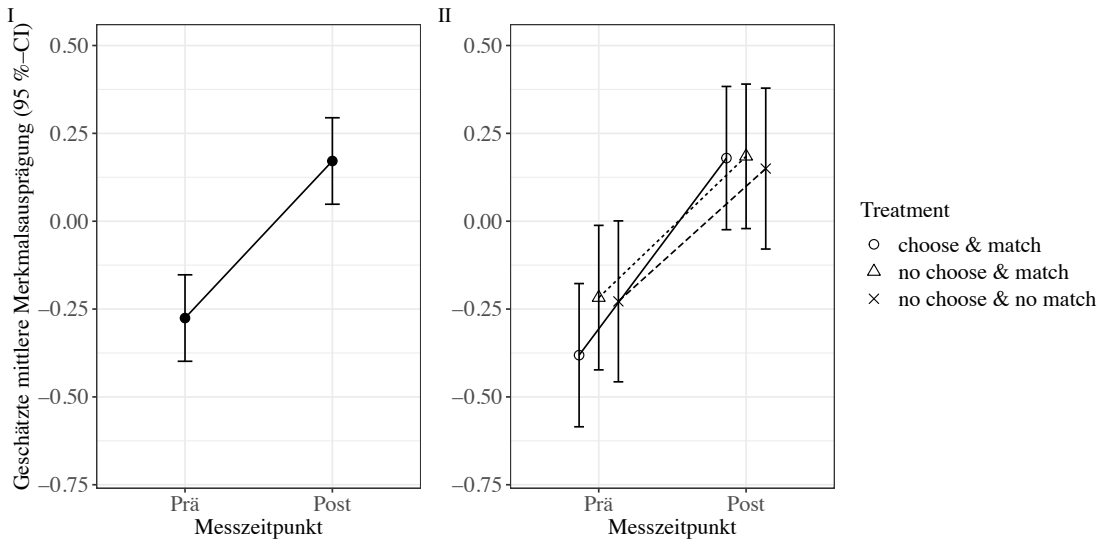


Abbildung 54: Einfluss des Messzeitpunkts (I) und des Treatments (II) auf die geschätzte mittlere Fachwissensausprägung[11]

[11] In dieser und den folgenden Abbildungen zum Fachwissenszuwachs ist auf der y-Achse die geschätzte mittlere Fachwissensausprägung in Logits dargestellt.

Auch die Fachwissenszuwächse unterscheiden sich nicht in Abhängigkeit vom Treatment. Dieser Eindruck wird von der statistisch nicht signifikanten Zeit × Treatment-Cross-level-Interaktion bestätigt (Tab. 62).

Um zu überprüfen, ob die Wirksamkeit des Treatments von dem individuellen Interesse an Chemie abhängig ist. Dafür wurden das individuelle Interesse und mögliche Interaktionseffekte modelliert. Die Effektivität des Treatments hängt von dem individuellen Interesse ab, wie die signifikante Zeit × Treatment × Interesse-Interaktion zeigt, $F(2, 199) = 6.30$, $p = .002$). Wie in Abbildung 55 dargestellt, profitieren die Hochinteressierten stärker von der passenden als von der nicht passenden Zuweisung der Kontextaufgaben, $b = 0.70$, $t(199) = 2.57$, $p = .033$, $r = .18$. Weiterhin zeigt sich für die Hochinteressierten, dass sie beim Lernen stärker von der selbstständigen Wahl im Vergleich zur nicht passenden Zuweisung profitieren, $b = 0.65$, $t(199) = 2.44$, $p = .047$, $r = .17$. Der Fachwissenszuwachs der Hochinteressierten ist unbeeinflusst von der Wahl eines Kontextes, wenn der Kontext zu den individuellen Merkmalen passt, $b = -0.05$, $t(199) = -0.22$, $p = .99$, $r = .02$. Einschränkend muss jedoch erwähnt werden, dass sich dieser Effekt in der Gruppe der durchschnittlich und unterdurchschnittlich Interessierten nicht zeigt. Weitere, statistisch bedeutsame Interaktionseffekte mit dem individuellen Interesse liegen nicht vor.

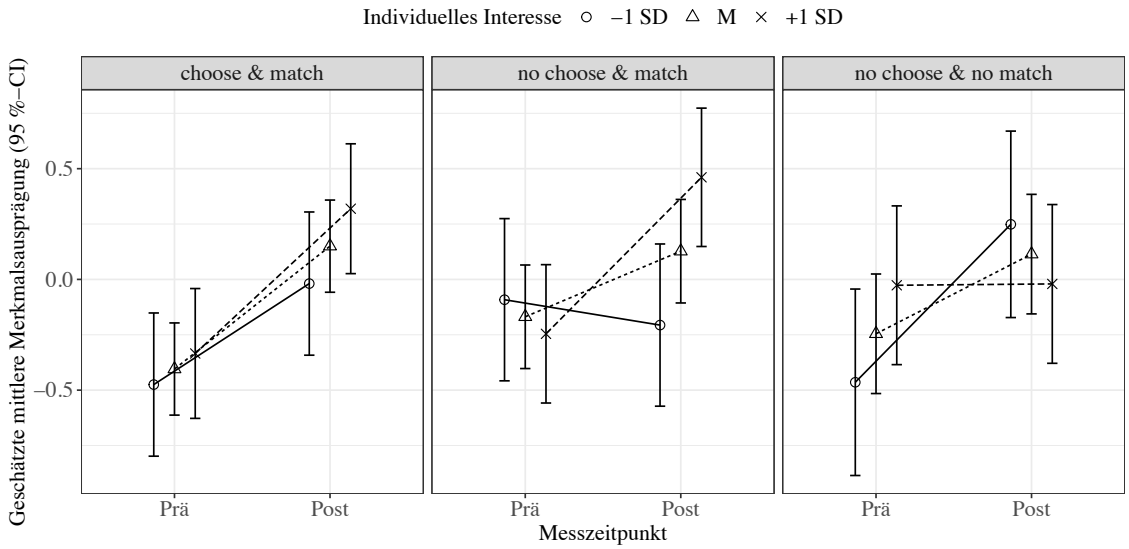


Abbildung 55: Einfluss der Zeit, des individuellen Interesse und des Treatments auf die geschätzte mittlere Fachwissensausprägung

Zuletzt wurde überprüft, ob die Wirksamkeit des Treatments von dem chemiebezogenen Selbstkonzept abhängt. Dafür wurden das Selbstkonzept und die möglichen Interaktionseffekte als feste Faktoren in das Modell integriert. Auch hier wurde das Selbstkonzept auf den

Gesamtmittelwert zentriert. Es kann ein signifikanter Haupteffekt des Selbstkonzeptes ausgemacht werden ($F(1, 199) = 10.90$, $p = .001$), was darauf hindeutet, dass das Selbstkonzept einen signifikanten Einfluss auf das Fachwissen hat. Zudem beeinflusst das Selbstkonzept den Fachwissenszuwachs, $F(1, 199) = 4.95$, $p = .027$.

Die signifikante Zeit × Treatment × Selbstkonzept-Interaktion zeigt, dass die Effektivität des Treatments vom Selbstkonzept beeinflusst wird, $F(2, 199) = 2.83$, $p = .026$). Der Interaktionseffekt zwischen der Zeit und dem Selbstkonzept ist in Abbildung 56 dargestellt.

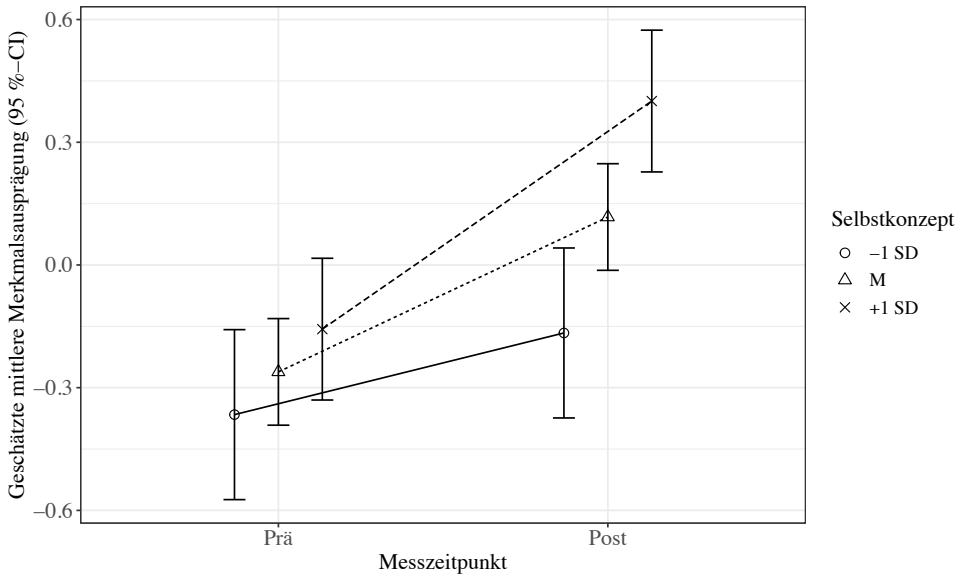


Abbildung 56: Einfluss der Zeit und des Selbstkonzeptes auf die geschätzte mittlere Fachwissensausprägung

Lernende mit einem unterdurchschnittlichen Selbstkonzept haben keinen bedeutsamen Fachwissenszuwachs erzielt, $b = 0.20$, $t(199) = 1.67$, $p = .10$, $r = .12$). Im Gegensatz dazu haben Lernende mit einem durchschnittlichen ($b = 0.37$, $t(199) = 5.06$, $p < .001$, $r = .34$) und einem überdurchschnittlichen Selbstkonzept ($b = 0.56$, $t(199) = 5.60$, $p < .001$, $r = .37$) während der Arbeit mit dem kontextualisierten Lernmaterial dazugelernt.

Der signifikante Interaktionsterm zwischen der Zeit, dem Selbstkonzept und dem Treatment deutet an, dass der Effekt des Treatments auch vom Selbstkonzept abhängt (Abb. 57). Lernende, die sich selbst als wenig kompetent wahrnehmen, profitieren eher von selbst gewählten ($b = 1.03$, $t(199) = 3.78$, $p < .001$, $r = .26$) oder unpassend zugewiesenen Kontexten ($b = -1.03$, $t(199) = -3.43$, $p = .001$, $r = .24$) im Vergleich zu passenden Kontextzuweisungen. In der Gruppe der Lernenden mit einem durchschnittlichen Selbstkonzept hat das Treatment keinen Einfluss, wie bereits herausgestellt wurde. Lernende mit einem hohen Selbstkonzept zeigen einen höheren Fachwissenszuwachs, wenn sie einen passenden statt einen nicht passenden Kontext zugewiesen bekommen, $b = 0.71$, $t(199) = 2.90$, $p = .013$, $r = .20$).

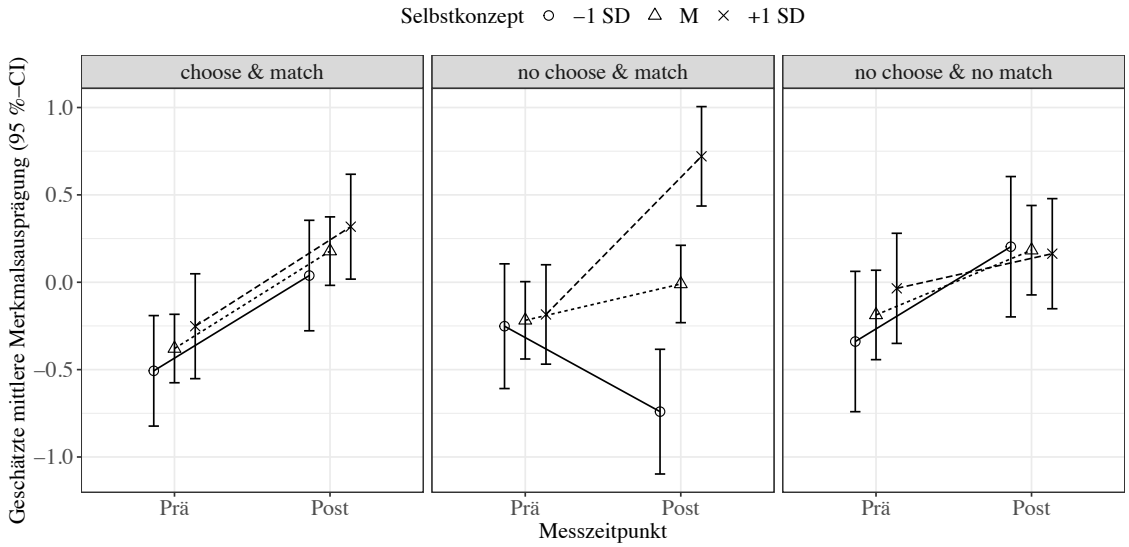


Abbildung 57: Einfluss der Zeit, des Selbstkonzepts und des Treatments auf die geschätzte mittlere Fachwissensausprägung

Überdies wird der Einfluss des Treatments vom bearbeiteten Kontext beeinflusst, $F(4, 199) = 2.83$, $p = .026$). Die Interaktion ist in Abbildung 58 dargestellt. Die Kontrastanalyse zeigt, dass das Treatment keinen Einfluss auf den Fachwissenszuwachs bei Lernenden hat, die im alltäglichen und besonderen Kontext gelernt haben. Das Lernen im innerfachlichen Kontext führt zu einem höheren Fachwissenszuwachs, wenn der innerfachliche Kontext selbstgewählt statt passend zugewiesen wird, $b = 0.99$, $t(199) = 2.97$, $p = .01$, $r = .21$. Die Wahl des innerfachlichen Kontextes führt nicht zu einem höheren Fachwissenszuwachs im Vergleich zur nicht passenden Zuweisung des innerfachlichen Kontextes, $b = 0.60$, $t(199) = 1.98$, $p = .15$, $r = .14$. Überdies unterscheidet sich der Fachwissenszuwachs der Lernenden mit passender und nicht passender innerfachlicher Kontextzuweisung nicht, $b = -0.40$, $t(199) = -1.18$, $p = .71$, $r = .08$.

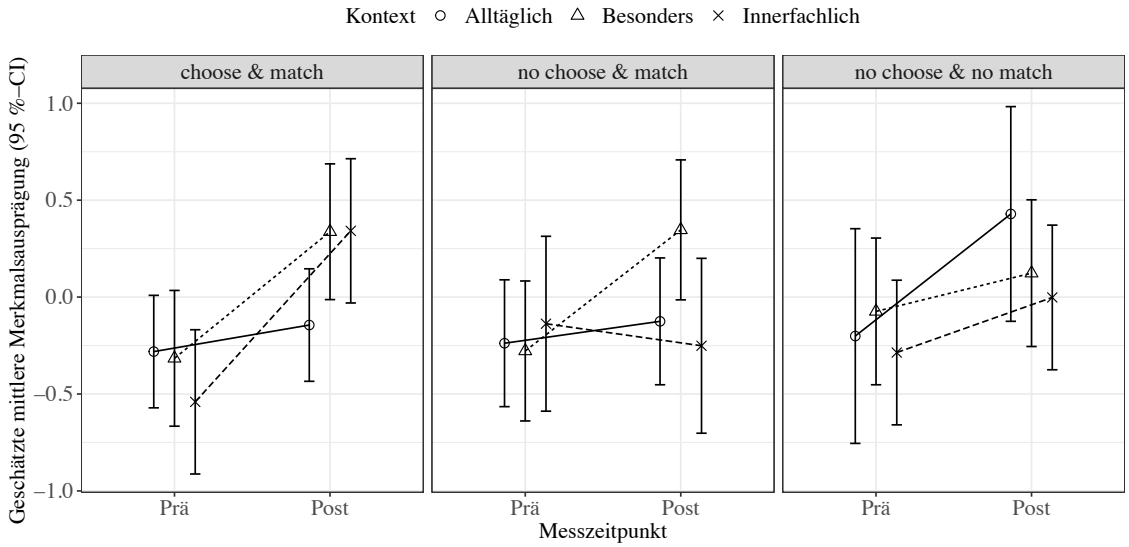


Abbildung 58: Einfluss der Zeit, des Treatments und des bearbeiteten Kontexts auf die geschätzte mittlere Fachwissensausprägung

10.5.8 Konsolidierung der Ergebnisse

Die bisher vorgestellten Ergebnisse sind insbesondere durch die begrenzte Prädiktionsgüte des neuronalen Netzes eingeschränkt. Die am Ende der Lerneinheit durchgeführte Zufriedenheitsbefragung ergab, dass einige Lernende in dem choose & match- und no choose & match-Treatment den Kontext nicht noch einmal wählen würden. Folglich scheint der Kontext nicht passend gewesen zu sein, obwohl angenommen wurde, dass alle Lernenden in der Gruppe einen passenden Kontext bearbeitet haben. Auf Grundlage dieses Ergebnisses können die ursprünglich konzipierten Treatmentgruppen neu zusammengesetzt werden (Tab. 63).

Tabelle 63: Beschreibung und Personenzahl (n) der angepassten Gruppen

Gruppe	*n*	Beschreibung
A (choose & match)	38	Lernende, die einen Kontext wählen konnten und diesen erneut bearbeiten würden.
B (choose & no match)	38	Lernende, die einen Kontext wählen konnten und diesen nicht erneut bearbeiten würden.
C (no choose & match)	49	Lernende, denen ein Kontext zugewiesen wurde und die diesen Kontext erneut bearbeiten würden.
D (no choose & no match)	92	Lernende, denen ein Kontext zugewiesen wurde und die diesen Kontext nicht erneut bearbeiten würden.

Zunächst wurde untersucht, ob die in Tabelle 62 beschriebenen Gruppen hinsichtlich der Kontrollvariablen vergleichbar sind. Eine MANOVA mit Pillai`s Trace verweist auf signifikante Unterschiede zwischen den Gruppen im Hinblick auf die Kontrollvariablen (V = 0.22, $F(27, 621) = 1.84$, $p = .005$, $\eta^2_p = .07$). Die post-hoc berechnete ANOVA zeigt Unterschiede im inhaltsfeldspezifischen Interesse an ($F(3, 213) = 5.20$, $p = .002$, $\eta^2 = .07$). Bei den anderen Kontrollvariablen sind keine Unterschiede festzustellen. Weiterhin besteht kein Zusammenhang zwischen der Gruppe und der Schulform ($\chi^2(3) = 1.97$, $p = .58$, $V = .10$) sowie dem Geschlecht ($\chi^2(3) = 3.42$, $p = .33$, $V = .13$). Darüber hinaus gibt es auch keinen Zusammenhang zwischen der Gruppe und dem bearbeiteten Kontext, $\chi^2(6) = 10.04$, $p = .12$, $V = .15$. Insgesamt zeigt sich damit, dass die neu zusammengesetzten Gruppen nur hinsichtlich des inhaltsfeldspezifischen Interesses nicht vergleichbar sind. Folglich muss das inhaltsfeldspezifische Interesse bei den weiteren Analysen als Kovariate berücksichtigt werden.

Bedeutsame Effekte wurden bisher nur für die Zufriedenheit sowie die affektive Valenz und epistemische Komponente des situationalen Interesses nachgewiesen, so dass sich die folgende Ergebnisdarstellung auf diese Variablen beschränkt. Da bisher keine relevanten Interaktionseffekte des individuellen Interesses und des Selbstkonzepts nachgewiesen werden konnten, werden diese im Folgenden ebenfalls nicht weiter betrachtet. Im Folgenden wird für jede abhängige Variable ein Mehrebenenmodell berechnet. Wie schon zuvor, wurde für jede Person ein Random-Intercept berechnet, um die Abhängigkeit zwischen den Messzeitpunkten zu berücksichtigen. Die feste Effektstruktur setzt sich aus Haupteffekten für die Zeit, der Gruppe und dem inhaltsfeldspezifische Interesse sowie den jeweiligen Interaktionseffekten zusammen. Für die Haupt- und Interaktionseffekte wurden *F*-Statistiken, basierend auf der Kenward-Roger-Approximation, berechnet. Diese sind für jedes Modell in Tabelle 64 aufgeführt.

Erwartungskonform zeigt sich ein signifikanter Haupteffekt der Zeit über alle abhängigen Variablen. Dies ist nicht überraschend, da lediglich die Zusammensetzung der Gruppen angepasst wurde. Demzufolge kann auch hier von einer Abnahme über die Messzeitpunkte – insbesondere vom ersten zum zweiten Messzeitpunkt – gesprochen werden. Wie die MANOVA und die post-hoc ANOVA bereits zeigten, beeinflusst das inhaltsfeldspezifische Interesse die abhängigen Variablen. Die Kontrastanalyse verweist hierbei auf einen positiven Zusammenhang. Der signifikante Interaktionseffekt zwischen der Zeit und dem inhaltsfeldspezifischen Interesse im Modell zur Aufklärung der Zufriedenheit und im Anschluss durchgeführte Kontrastanalysen deuten darauf hin, dass die Zufriedenheit mit der Zeit weniger stark abnimmt, wenn die Lernenden ein höheres inhaltsfeldspezifisches Interesse haben.

Tabelle 64: Haupteffekte der Zeit, der Gruppe und des inhaltsfeldspezifischen Interesses sowie deren Interaktionseffekte auf die Zufriedenheit, die affektive Valenz des situationalen Interesses und die epistemische Komponente des situationalen Interesses

Effekt	Zufriedenheit	Affektive Valenz	Epistemische Komponente
Z	43.21^{***}	43.11^{***}	7.75^{***}
G	16.49^{***}	12.26^{***}	8.96^{***}
I	5.49^{*}	24.32^{***}	30.63^{**}
Z × G	1.69	2.27^{*}	0.89
Z × I	3.49^{*}	2.54	0.47
G × I	0.46	2.26	1.71
Z × G × I	0.88	0.84	0.19

Anmerkung: Z = Zeit, G = Gruppe, I = Inhaltsfeldspezifisches Interesse. In der Tabelle sind die *F*-Statistiken basierend auf der Kenward-Roger-Approximation angegeben. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$

Für die Zufriedenheit, die affektive Valenz und die epistemische Komponente des situationalen Interesses zeigt sich ein signifikanter Haupteffekt der Gruppe, wobei der Interaktionseffekt zwischen der Gruppe und der Zeit für die affektive Komponente des situationalen Interesses signifikant ist. Demzufolge unterscheiden sich die abhängigen Variablen zu jedem Messzeitpunkt in Abhängigkeit der Gruppe, wobei sich für die affektive Valenz des situationalen Interesses sogar andeutet, dass sich der zeitliche Verlauf in Abhängigkeit zur Gruppe verhält. Der zeitliche Verlauf jeder abhängigen Variable ist getrennt nach der Gruppe in Abbildung 59 dargestellt.

Sowohl für die aufgabenbezogene Zufriedenheit (I), die affektive Valenz des situationalen Interesses (II) als auch für die epistemische Komponente des situationalen Interesses (III) zeigen sich die vielfach identifizierten Unterschiede: Lernende mit passendem Kontext weisen zu allen Messzeitpunkten höhere Merkmalsausprägungen auf als Lernende mit unpassendem Kontext, unabhängig davon, ob der Kontext selbst gewählt oder zugewiesen wurde. Für die epistemische Komponente ist der Unterschied zwischen der A- und der C-Gruppe zum ersten Messzeitpunkt allerdings statistisch nicht bedeutsam, $b = .25$, $t(321) = 2.16$, $p = .19$, $r = .12$. Die Effektstärken liegen für die Zufriedenheit ($.17 \leq r \leq .27$), die affektive ($.15 \leq r \leq .29$) und epistemische Komponente ($.18 \leq r \leq .25$) in vergleichbaren Bereichen wie bei der Analyse des ursprünglichen Treatments.

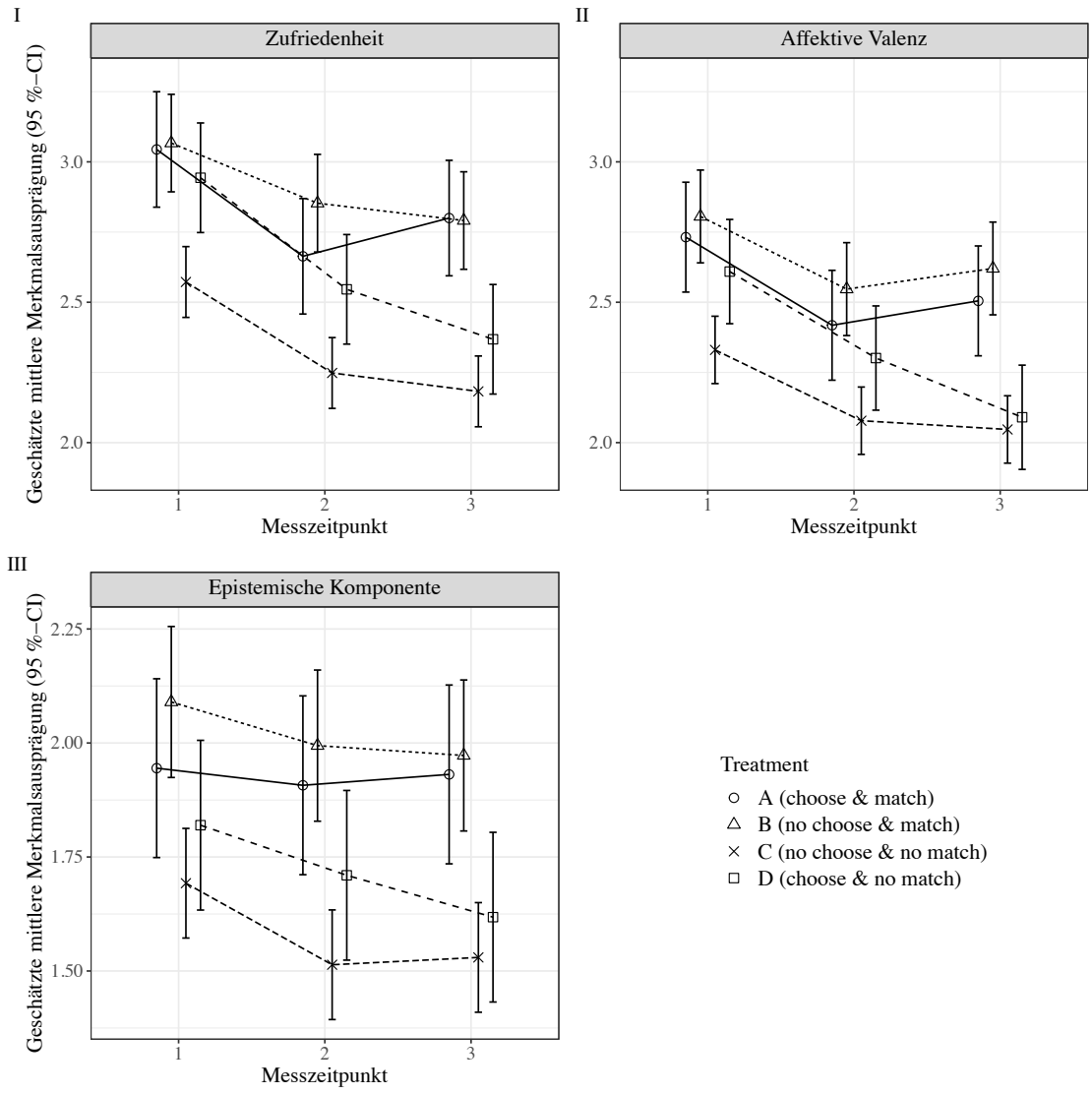


Abbildung 59: Verlauf der aufgabenbezogenen Zufriedenheit (I), der affektiven Valenz des situationalen Interesses (II) und der epistemischen Komponente des situationalen Interesses (III) in Abhängigkeit der neu zusammengesetzten Gruppen

Ein bemerkenswerter Effekt zeigt sich hingegen bei den Lernenden, die eine nicht passende kontextorientierte Aufgabe gewählt haben (Gruppe D, choose & no match). Nach der ersten Teilaufgabe gaben diese Lernenden eine ähnliche aufgabenbezogene Zufriedenheit an wie die Lernenden der A- (b = 0.10, $t(400)$ = 0.80, p = .99, r = .04) und B-Gruppe (b = 0.12, $t(400)$ = 0.93, p = .99, r = .05). Diese war jedoch höher als in der C-Gruppe (b = -0.37, $t(400)$ = -3.14, p = .011, r = .16). Zum zweiten Messzeitpunkt ist der Unterschied zwischen der D- und C-Gruppe allerdings nicht mehr signifikant, b = -0.30, $t(400)$ = -2.52, p = .07, r = .13. Nach der dritten Teilaufgabe liegt die aufgabenbezogene Zufriedenheit der A- (b = 0.43, $t(400)$ = 2.99,

p = .018, r = .15) und B-Gruppe (b = 0.42, t(400) = 3.18, p = .009, r = .16) über der Zufriedenheit der D-Gruppe. Die Zufriedenheit der D-Gruppe hat sich im Verlauf der kontextorientierten Lerneinheit der Zufriedenheit der C-Gruppe angenähert.

Ein ähnlicher Verlauf zeigt sich auch für die affektive Valenz des situationalen Interesses, wobei der Unterschied zwischen der D- und C-Gruppe zum ersten Messzeitpunkt knapp über dem Signifikanzniveau von 5 % liegt, b = -0.29, t(336) = -2.53, p = .071, r = .14. Bei der epistemischen Komponente des situationalen Interesses sind die Unterschiede etwas weniger deutlich.

Die nicht signifikante Interaktion zwischen dem Treatment und dem inhaltsfeldspezifischen Interesse sowie zwischen der Zeit, dem Treatment und dem inhaltsfeldspezifischen Interesse impliziert, dass der Effekt des Treatments nicht durch das inhaltsfeldspezifische Interesse beeinflusst wird.

Die Analyse der im Verlauf der Lerneinheit gemessenen Zufriedenheit zeigte, dass die Wirkung des Treatments im Verlauf der Zeit vom bearbeiteten Kontext abhängt (siehe Kapitel 10.5.4). Aus diesem Grund wurde ein weiteres Mehrebenenmodell zur Analyse der Zufriedenheit berechnet, bei dem neben der Zeit, den angepassten Gruppen und dem inhaltsfeldspezifischen Interesse auch der bearbeitete Kontext berücksichtigt wurde. In dem Modell mit den entsprechenden Haupt- und Interaktionseffekten ist die Zeit × Gruppe × Kontext-Interaktion statistisch nicht mehr bedeutsam, F(12, 385.46) = 1.01, p = .44. Demzufolge scheint die Effektivität des Treatments mehr durch den Kontext moderiert zu werden, wenn die Zuweisung der Kontexte nicht länger mit der verbesserungswürdigen Performanz des neuronalen Netzes konfundiert ist.

10.6 Zusammenfassende Diskussion der Ergebnisse

Die wesentlichen Ergebnisse der zweiten Teilstudie sind im Anhang (Tab. A14) zusammengefasst. Insgesamt belegen die Ergebnisse der zweiten Teilstudie, dass sich sowohl die aufgabenbezogene Zufriedenheit, das situationale Interesse als auch die kognitive Belastung im Verlauf der kontextorientierten Lerneinheit ändern. Während die Zufriedenheit und das situationale Interesse von der ersten zur zweiten Teilaufgabe abnehmen, steigt die kognitive Belastung. Eine mögliche Erklärung für den rapiden Anstieg der kognitiven Belastung von der ersten zur zweiten Teilaufgabe könnte aus den unterschiedlichen Kompetenzerwartungen der Aufgaben resultieren. Während in der ersten Teilaufgaben vornehmlich die makroskopische Ebene adressiert wird, müssen zur Bearbeitung der zweiten Teilaufgabe Phänomene auf submikroskopischer Ebene erklärt werden. Dies stellt die Lernenden vor größere Schwierigkeiten (z. B. Treagust et al., 2003), die sich in einer höheren kognitiven Belastung zeigen. Da von der zweiten zur dritten Teilaufgabe kein erneuter Wechsel der Repräsentationsebene vollzogen wird, bleibt die kognitive Belastung auf einem konstant hohen Niveau. Ursächlich für

den Abfall der Zufriedenheit und des situationalen Interesses vom ersten zum zweiten Messzeitpunkt könnten Neuigkeitseffekte sein, die durch das digitale Aufgabenformat oder die Kontextorientierung zu erklären sind (Theyßen, 2014). Möglicherweise ist der Abfall der Zufriedenheit und des situationalen Interesses vom ersten zum zweiten Messzeitpunkt auf die Zunahme der Aufgabenschwierigkeit zurückführbar. Im Rahmen der ersten Teilstudie hat sich gezeigt (im Pfadmodell; Abbildung 20), dass die wahrgenommene Aufgabenschwierigkeit einen Einfluss auf die Zufriedenheit hat.

Für keine abhängige Variable konnte ein signifikanter Haupteffekt des Kontexts oder eine signifikante Zeit × Kontext-Interaktion nachgewiesen werden. Demzufolge ist kein Kontext grundsätzlich lernförderlicher als ein anderer. Lernende wählen unterschiedliche Kontexte zum Lernen aus (siehe Teilstudie I sowie Broman et al., 2018 und van Vorst & Aydogmus, 2018) und profitieren hinsichtlich des situationalen Interesses von unterschiedlichen Kontexten (Habig et al., 2018). Wie die weiterführende Datenanalyse bestätigt, ist insbesondere die Passung zwischen Person und Kontext entscheidend. Für die aufgabenbezogene Zufriedenheit und das situationale Interesse zeigen sich kleine bis mittlere Treatmenteffekte im Verlauf der kontextorientierten Lerneinheit, wobei diese Effekte für die epistemische Komponente kleiner ausfallen, sich vorrangig für die hochinteressierten Schülerinnen und Schüler zeigen und nach Einbezug des individuellen Interesses und des Selbstkonzept nicht mehr statistisch signifikant sind. Über alle Messzeitpunkte deutet sich an, dass die Lernenden zufriedener mit der kontextorientierten Aufgabe sind und eine höhere affektive Valenz angeben, wenn sie mit einem Kontext lernten, der zu ihren Personenmerkmalen passt. Unerheblich ist allerdings, ob sie diesen Kontext aktiv wählen konnten oder nicht. Ähnliche Effekte wurden u. a. von Wilde et al. (2018) sowie Flowerday und Shell (2015) identifiziert, sodass sich die hier berichteten Ergebnisse nahtlos in den bestehenden Forschungsstand einfügen. Auch die Höhen der Effektstärken sind im Vergleich zu bisherigen Ergebnissen erwartungskonform (z. B. Patall et al., 2008; Wilde et al., 2018). Erklärungen für dieses Ergebnis sind in der Selbstbestimmungstheorie von Deci und Ryan (z. B. 2017) hinreichend beschrieben. Demnach resultiert das für die affektiv-motivationalen Effekte kausale Autonomieerleben nicht aus der Wahl zwischen Wahloptionen, sondern aus der Passung der Wahloptionen zu eigenen Interessen und Werten bzw. Merkmalen einer Person. Hierbei muss jedoch einschränkend erwähnt werden, dass das Autonomieerleben im Rahmen der zweiten Teilstudie nicht erfasst wurde.

Der identifizierte Effekt der Kontextpassung scheint nicht bloß ein *catch*-Faktor zu sein (Mitchell, 1993), der kurzzeitig affektiv-motivationale Merkmale beeinflusst. Viel mehr deuten die Daten über den Verlauf der kontextorientierten Lerneinheit an, dass dieser Effekt über die Zeit stabil ist. Die deskriptiven Daten zur aufgabenbezogenen Zufriedenheit und zur affektiven Valenz des situationalen Interesses zeigen im Vergleich zur Kontextwahl eine höhere Merkmalsausprägung an den Messzeitpunkten und eine geringere Abnahme vom zweiten zum dritten Messzeitpunkt, wenn der passende Kontext zugewiesen und nicht selbst gewählt wurde. Wilde et al. (2018) konnten in ihrer Studie einen ähnlichen Effekt beobachten, der

zudem statistisch signifikant war. Im Rahmen dieser Untersuchung wurde dieser Unterschied allerdings nicht signifikant. Besonders am dritten Messzeitpunkt liegen allerdings Unterschiede mit kleiner Effektstärke vor, die möglicherweise aufgrund der Stichprobengröße nicht als signifikant ausgewiesen werden können. In der Untersuchung von Wilde et al. (2018) war die Stichprobe etwas größer.

Ein signifikanter Interaktionseffekt belegt, dass sich die Effekte des Treatments auf die aufgabenbezogene Zufriedenheit in Abhängigkeit vom bearbeiteten Kontext unterscheiden. Die beschriebenen Unterschiede zwischen den Treatmentgruppen sind zu allen Messzeitpunkten nur in der Gruppen der Lernenden nachweisbar, die in dem alltäglichen Kontext gelernt haben. Schülerinnen und Schüler, die in der besonders und innerfachlich kontextualisierten Lerneinheit gearbeitet haben, weisen die identifizierten Effekte nicht zu allen Messzeitpunkten auf. Maßgeblich für dieses Ergebnis scheint die etwas schlechtere Leistung des neuronalen Netzes bei der Vorhersage des besonderen und innerfachlichen Kontextes zu sein (siehe Kapitel 9 und 10.5.3). Darauf deutet auch die Absicherung der Ergebnisse in Kapitel 10.5.8 hin. Einschränkend muss jedoch erwähnt werden, dass der Interaktionseffekt bei Einbeziehung des individuellen Interesses in das Mehrebenenmodell statistisch nicht mehr signifikant ist (Tab. 65).

In der Literatur beschriebene Interaktionseffekte von Wahlmöglichkeiten mit dem individuellen Interesse und dem Selbstkonzept können im Rahmen dieser Studie nicht repliziert werden (Patall, 2013; Patall et al., 2014). Stattdessen wurde eine Reihe von weiteren Interaktionseffekten aufgedeckt. Das individuelle Interesse beeinflusst den zeitlichen Verlauf der aufgabenbezogenen Zufriedenheit und der affektiven Valenz des situationalen Interesses. Erwartungsgemäß ist der Abfall über die Lerneinheiten geringer, je höher das individuelle Interesse der Lernenden ist. Die Wirkungsrichtung des individuellen Interesses auf die Zufriedenheit ist damit zwar entgegengesetzt zur Wirkungsrichtung aus dem Pfadmodell der ersten Teilstudie (Abbildung 20), erweist sich aber dennoch als theoriekonform. Darüber hinaus zeigen die über die Messzeitpunkte und Treatmentgruppen aggregierten Daten für die Zufriedenheit und das situationalen Interesse differenzielle Effekte des bearbeiteten Kontexts in Abhängigkeit vom individuellen Interesse, die sich im Einklang mit den Befunden der ersten Teilstudie und der Literatur befinden (Habig et al., 2018; van Vorst & Aydogmus, 2021). Hochinteressierte Lernende geben eine höhere Zufriedenheit und ein höheres situationales Interesse an, wenn sie mit einem besonderen oder innerfachlichen Kontext gelernt haben. Diese Schülerinnen und Schüler sind in affektiver und motivationaler Hinsicht durch die alltäglichen Kontexte benachteiligt. Dagegen profitieren die unterdurchschnittlich interessierten Schülerinnen und Schüler gemittelt über die Messzeitpunkte und Treatmentgruppen hinsichtlich des situationalen Interesses vom Alltagskontext. Diese Lernenden werden im situationalen Interesse durch den besonderen und innerfachlichen Kontext benachteiligt. Für die Zufriedenheit zeigt sich dieser Effekt nicht. Hier deutet sich viel mehr an, dass den durchschnittlich und unterdurch-

schnittlich Interessierten gleichgültig ist, welchen der Kontexte sie bearbeiten. Auch die Analyse der Zufriedenheitsbefragung am Ende der Lerneinheit deutet darauf hin, dass Lernende mit grundsätzlich fehlendem Interesse an Chemie nicht durch die Kontexte zum Chemielernen motiviert werden können. Dieses Ergebnis kann aufgrund des engen thematischen und inhaltlichen Fokus der Kontexte nicht für das gesamte kontextorientierte Lernen generalisiert werden. Dennoch ist dieses Ergebnis alarmierend, da doch der Einsatz von Kontexten insbesondere Schülerinnen und Schüler mit einem geringen Interesse an Chemie zum Chemielernen motivieren soll (z. B. Gilbert, 2006). Dass sich die beschriebenen Effekte des bearbeiteten Kontexts aggregiert über die Treatmentgruppen zeigen, ist wahrscheinlich darauf zurückzuführen, dass zwei Drittel der Lernenden im passenden Kontext gelernt haben. Einschränkend muss auch erwähnt werden, dass nicht alle paarweisen Vergleiche zwischen den Kontexten statistisch abgesichert werden können. Das gilt besonders für Ergebnisse zur epistemischen Komponente des situationalen Interesse.

Am Ende der kontextorientierten Lerneinheit zeigt eine Befragung, dass die meisten Lernenden denselben Kontext oder einen anderen der konzipierten Kontexte bearbeiten würden. Lediglich 16.13 % der Lernenden würden keinen der Kontexte mehr bearbeiten. Folglich scheinen die Kontexte einen Großteil der Lernenden angesprochen zu haben. Sowohl der Vergleich der Personenmerkmale als auch der Vergleich der angegebenen Motive deuten auf vornehmlich affektiv-motivationale Ursachen für dieses Ergebnis hin. Die Entscheidung für denselben, einen anderen oder gar keinen der Kontext scheint im Wesentlichen vom Interesse der Lernenden an Chemie generell und in Bezug auf das Themen- oder Inhaltsfeld im Speziellen abzuhängen. Der vorher bearbeitete Kontext ist hier weniger entscheidend. Die qualitativen Daten zeigen darüber hinaus, dass jeder vierte Lernende, der den Kontext erneut bearbeiten würde, spezifische affektiv-motivationale Gründe nannte, die sich auf Merkmale des bearbeiteten Kontexts bezogen. Dieser Befund könnte auf eine tiefere Auseinandersetzung mit dem Kontext und die Wahrnehmung kontextspezifischer Informationen bei zufriedenen Lernenden hindeuten. Andererseits gaben zufriedene Schülerinnen und Schüler auch häufig unspezifische Gründe für ihre Zufriedenheit an, die ganz unterschiedliche Aufgabenmerkmale adressieren. Möglicherweise handelt es sich hierbei jedoch um einen Methodeneffekt, wenn es für die Lernenden insgesamt schwieriger ist, positives als negatives Feedback zu den Aufgaben zu geben.

Schülerinnen und Schüler, die rückblickend einen anderen Kontext bearbeitet hätten, würden eher den alltäglichen oder besonderen Kontext bevorzugen. Diese Lernenden haben vorher nicht ausschließlich den innerfachlichen Kontext bearbeitet. Auch dieses Ergebnis deutet darauf hin, dass die meisten Lernenden außerfachliche gegenüber innerfachlichen Kontexten favorisieren (siehe Teilstudie I sowie Broman et al., 2018 und Broman et al., 2020). Die Ergebnisse der qualitativen Befragung liefern weitere Hinweise auf die Ursachen für dieses Ergebnis. Häufig wurde ein größeres Interesse an einer anderen Aufgabe als Ursache genannt, wobei nicht weiter differenziert wurde, weshalb die andere Aufgabe interessanter war. Wie

die Analyse der qualitativen Motive zeigt, gaben viele Lernende, die keinen der Kontexte mehr bearbeiten würden, eine grundsätzliche Abneigung gegenüber dem Fach Chemie oder ein unspezifisches Desinteresse an den Aufgaben an. Während die Lernenden in der Studie von van Vorst und Aydogmus (2021) noch häufig kognitive Motive – allen voran eine hohe Aufgabenschwierigkeit – für ihre Unzufriedenheit angaben, begründeten die Lernenden in dieser Studie ihre Entscheidung eher mit affektiv-motivationalen Motiven. Hieraus kann abgeleitet werden, dass die Aufgaben für die Mehrheit der Schülerinnen und Schüler nicht zu schwierig waren.

Weiterhin belegen die Daten, dass die Lernenden den Kontext häufiger erneut bearbeiten würden, wenn sie den Kontext selbst wählen konnten oder wenn ihnen der passende Kontext durch das neuronale Netz zugewiesen wurde. Diese Ergebnisse stimmen somit mit den Ergebnissen der aufgabenbezogenen Zufriedenheit und der affektiven Komponente des situationalen Interesses überein, die im Verlauf der Lerneinheit häufiger gemessen wurden. Die qualitativen Daten liefern jedoch nur begrenzte Erklärungsansätze. Lernende mit der Möglichkeit der Kontextwahl gaben häufiger als die anderen Treatmentgruppen kognitive Motive für die erneute Bearbeitung an – vor allem, weil sie etwas Neues gelernt hatten. Affektiv-motivationale Motive wurden in den Treatmentgruppen gleich häufig genannt, wobei sich auch die spezifischen affektiv-motivationalen Motive zwischen den Treatments nur geringfügig unterschieden.

Betrachtet man nur die Lernenden des choose & match-Treatments und vergleicht die Lernenden, die denselben, einen anderen oder keinen Kontext noch einmal bearbeiten würden, ergibt sich ein aufschlussreiches Bild: Die Lernenden unterscheiden sich lediglich im Hinblick auf das inhaltfeldspezifische Interesse und nicht in Bezug auf weitere Personenmerkmale. Demzufolge ist es unwahrscheinlich, dass einige Lernende aufgrund von geringem Vorwissen oder mangelndem Selbstkonzept nicht in der Lage sind, einen passenden Kontext auszuwählen. Von den erhobenen Merkmalen ist nur das Interesse am Inhaltsfeld ausschlaggebend dafür, ob eine Aufgabe wieder gewählt wird oder nicht.

Hinsichtlich der kognitiven Belastung konnten die Hypothesen zu den Effekten der selbstständigen Kontextwahl und der Passung von Kontexten nicht bestätigt werden. Mit Blick auf die investierte Denkanstrengung konnte ein signifikanter Treatmenteffekt nachgewiesen werden, dessen Kontrastanalyse auf eine höhere Denkanstrengung der Lernenden mit Kontextwahl im Vergleich zu Lernenden mit passender Kontextzuweisung nach der zweiten Teilaufgabe hinweist. Analog konnte für die unterdurchschnittlich Interessierten nach der zweiten Teilaufgabe gezeigt werden, dass die Lernenden mit autonomer Kontextwahl eine höhere Aufgabenschwierigkeit wahrnahmen als die Lernenden mit passender Kontextzuweisung. Es wäre denkbar, dass die Auswahl einer Kontextaufgabe zu einer höheren kognitiven Belastung führt als die Zuweisung einer Aufgabe. Dieser Effekt sollte sich dann aber vor allem zum

ersten Messzeitpunkt zeigen, der zeitlich näher an der Aufgabenauswahl lag. Es ist auch möglich, dass dieser Effekt erst dann wirksam wird, wenn die Aufgabe ein bestimmtes Anforderungsniveau überschreitet, das von der ersten Teilaufgabe noch nicht erreicht wurde. Da dieser Effekt jedoch für die investierte Denkanstrengung unter Berücksichtigung des Selbstkonzepts im Modell nicht auftritt und für die wahrgenommene Aufgabenschwierigkeit nur für die unterdurchschnittlich Interessierten vorliegt, kann diese Hypothese nicht ohne weiteres bestätigt werden und bedarf weiterer Untersuchungen. Die weiterführende Datenanalyse verweist außerdem auf die Bedeutsamkeit des Selbstkonzepts für die kognitive Belastung. Wie Habig (2017) bereits zeigen konnte, reduziert ein positives Selbstkonzept sowohl die investierte Denkanstrengung als auch die wahrgenommene Aufgabenschwierigkeit. Ein solcher Effekt auf die wahrgenommene Aufgabenschwierigkeit konnte auch im Rahmen der ersten Teilstudie nachgewiesen werden (siehe auch Abbildung 20).

Alle Lernenden lernten signifikant und mit mittlerem bis großem Effekt dazu, was angesichts der kurzen Dauer der Lerneinheit positiv hervorzuheben ist. Die aufgestellten Hypothesen zur Wirksamkeit der Treatments in Bezug auf den Fachwissenszuwachs können nur eingeschränkt bestätigt werden, wobei zu bedenken ist, dass die Ergebnisse aufgrund der geringen Reliabilität des Fachwissenstests nur bedingt interpretierbar sind.

Insbesondere Lernende mit überdurchschnittlichem Interesse profitieren im Fachwissenszuwachs stärker von passgenauen kontextorientierten Aufgaben im Vergleich zu unpassenden kontextorientierten Aufgaben. Auch hier ist es unerheblich, ob der Kontext selbst gewählt oder zugewiesen wurde. Lernende mit einem hohen Selbstkonzept zeigen einen höheren Fachwissenszuwachs, wenn sie einen passenden Kontext statt einen nicht passenden Kontext zugewiesen bekommen. Diese Effekte können in der Gruppe der Lernenden mit durchschnittlichem und unterdurchschnittlichem Selbstkonzept und Interesse nicht identifiziert werden. Insbesondere die Datenanalyse der Zufriedenheit hat gezeigt, dass Lernenden mit durchschnittlichem und unterdurchschnittlichem Interesse im Hinblick auf ihre Zufriedenheit nach der Aufgabenbearbeitung gleichgültig ist, welchen der Kontexte sie bearbeiten. So könnte für diese Lernenden auch irrelevant sein, ob sie in einem passenden Kontext lernen oder ob sie den Kontext selbst wählen können. Diese Tendenz zeigt sich hier möglicherweise auch im Fachwissenszuwachs.

Fehlende Effekte in der Gesamtkohorte sind möglicherweise auch auf die kurze Dauer der Intervention zurückzuführen. Andererseits deutet die bisherige Forschung zum kontextorientierten Lernen häufig auf fehlende Effekte hinsichtlich kognitiver Variablen hin, so dass auch die Bearbeitung geeigneter Kontexte möglicherweise nicht zu einem erhöhten Fachwissenszuwachs führt. Diese Hypothesen können jedoch auf der Grundlage der vorliegenden Daten nicht zweifelsfrei bestätigt werden.

Überdies offenbart die Berücksichtigung des Selbstkonzepts im Mehrebenenmodell, dass Lernende, die im innerfachlichen Kontext lernten, einen höheren Fachwissenszuwachs erzielten, wenn sie den Kontext wählten, als wenn sie den Kontext passend zugewiesen bekamen. Möglicherweise ist der Vorteil der selbstständigen Wahl im Vergleich zur passenden Zuweisung des innerfachlichen Kontext auf die nicht ausreichend valide Zuweisung des innerfachlichen Kontexts durch das neuronale Netz zurückführbar.

Die identifizierten Effekte sind insgesamt unter Betrachtung der in Kapitel 10.5.3 dargelegten Performanz des neuronalen Netzes zu interpretieren. Die Ergebnisse deuten hierbei an, dass nicht jedem Lernenden aus der no choose & match und no choose & no match ein passender bzw. nicht passender Kontext zugewiesen wurde. Dies wird auch aus der Analyse der Zufriedenheitsbefragung am Abschluss der Lerneinheit deutlich. Zugleich verweist die Zufriedenheitsbefragung am Abschluss der Lerneinheit darauf, dass die Vorhersagegenauigkeit des neuronalen Netzes ausreichend war, um die nachgewiesenen Effekte zwischen den Treatmentgruppen zu induzieren. Die Lernenden des choose & match- und des no choose & match-Treatments geben zu vergleichbaren Anteilen an, dass sie den gleichen Kontext erneut oder einen anderen Kontext bearbeiten würden. Diese These wird auch durch die erneute Analyse der Daten mit den rekonstruierten Gruppen unterstützt, die bereits identifizierte Effekt bestätigt.

Weiterhin zeigt die neu zusammengesetzte D-(choose & no match)-Gruppe[12] im Vergleich zur C-(no choose & no match)-Gruppe höhere Ausprägungen in affektiv-motivationalen Variablen am ersten Messzeitpunkt, was auf die Wahlmöglichkeit zurückzuführen ist. Dieser Unterschied zeigt sich vornehmlich in der Zufriedenheit und ist für die affektive Valenz des situationalen Interesses knapp über dem Signifikanzniveau von 5 %. Im Verlauf der kontextorientierten Lerneinheit nimmt die Zufriedenheit und die affektive Valenz der D-Gruppe immer weiter ab, so dass am Ende der Lerneinheit kein Unterschied mehr zwischen der D-Gruppe und der C-Gruppe besteht. Im Gegensatz dazu weisen Lernende, die in einem geeigneten Kontext gelernt haben, im Verlauf der Lerneinheit höhere Merkmalsausprägungen auf, unabhängig davon, ob sie den Kontext selbst gewählt haben oder nicht. Für die epistemische Komponente des situationalen Interesses zeigen sich die beschriebenen Effekte zwar deskriptiv, können aber inferenzstatistisch nicht eindeutig abgesichert werden.

Insgesamt bestätigen die Ergebnisse somit die bisherigen Interpretationen: Die Möglichkeit, zwischen verschiedenen Kontextaufgaben wählen zu können, stellt lediglich einen *catch*-Faktor dar (Mitchell, 1993), der affektiv-motivationale Merkmale kurzzeitig positiv beeinflussen kann. Wenn der Kontext jedoch nicht zur Person passt, kann der ursprüngliche Effekt der

[12] Siehe Tabelle 63 für die Benennung und Beschreibung der angepassten Gruppen.

Wahl nicht aufrechterhalten werden. Dagegen motiviert ein geeigneter Kontext im Lernprozess sowohl kurzfristig als *catch*-Faktor als auch langfristig als *hold*-Faktor.

11 Zusammenfassung und Limitationen

Die Berücksichtigung individueller Lernvoraussetzungen durch die Anpassung der Lernumgebung stellt im Kontext der individuellen Förderung eine sowohl für die Unterrichtspraxis als auch für die empirische Bildungsforschung relevante Herausforderung dar. Die Forschungslage ist dabei aufgrund unterschiedlicher Konzeptualisierungen der individuellen Förderung und damit einhergehenden diversen Forschungszugängen als kontrovers zu bezeichnen. Studien zeigen häufig konträre Ergebnisse und lassen sich aufgrund nicht klar definierter Maßnahmen nur schwierig in den bisherigen Forschungsstand einordnen. Der Schwerpunkt der bisherigen Bemühungen zur Aufklärung potenziell lernförderlicher Effekte der individuellen Förderung lag dabei zumeist auf der Evaluation leistungsbezogener Differenzierungsmaßnahmen (Smale-Jacobse et al., 2019; Steenbergen-Hu et al., 2016). Gleichwohl auch die Berücksichtigung affektiv-motivationaler Merkmale als Grundlage zur Differenzierung in der Literatur beschrieben wird (z. B. Tomlinson, 1999), fokussieren bisherige Forschungsansätze solche nur vereinzelt. Dies gilt insbesondere für das Unterrichtsfach Chemie, für das die empirisch-didaktische Forschung bislang nur wenige – an leistungsbezogenen Merkmalen orientierte – Konzepte evaluiert hat (Anus, 2015; Kallweit & Melle, 2017; Hauerstein, 2019). Das Ziel des in dieser Arbeit beschriebenen Promotionsvorhabens war deshalb die Erschließung dieser Forschungslücke durch die Entwicklung und Evaluation eines Konzepts zur interessenbasierten Differenzierung im Fach Chemie.

Interessendifferenzierende Maßnahmen werden in der Literatur meist als die Öffnung des Unterrichts im Hinblick auf die Auswahl zwischen bereitgestellten Lernmaterialien beschrieben. So wird vorgeschlagen, die Lernenden im Sprachunterricht zwischen Lesetexten zu unterschiedlichen Themen (z. B. Morgan & Wagner, 2013) oder im Mathematikunterricht zwischen Aufgaben mit unterschiedlichen Anwendungsbezügen (z. B. Gijsbers et al., 2020) wählen zu lassen. Da die hier angeführten Studien aufgrund eines qualitativen Forschungsdesigns oder weiterer methodischer Einschränkungen keine abschließende Beurteilung der Effektivität einer solchen Differenzierungsstrategie erlauben, können lehr-lernpsychologische Theorien herangezogen werden, um potenzielle Effekte zu verorten.

In der Selbstbestimmungstheorie wird das psychologische Grundbedürfnis nach Autonomie als eines von drei Bedürfnissen beschrieben, das für die Internalisierung von Verhalten und die Ausbildung intrinsischer Motivation wesentlich ist (Ryan et al., 2021). Die Bereitstellung von Auswahlmöglichkeiten wird hierbei als eine zentrale Maßnahme zur Befriedigung des Autonomiebedürfnisses angesehen, weshalb positive Effekte auf intrinsische Motivation und Leistung zu erwarten sind. Die Autoren weisen jedoch darauf hin, dass insbesondere die Passung der Wahlmöglichkeiten zu den eigenen Interessen, Zielen und Werten für das Autonomieerleben entscheidend ist (Ryan & Deci, 2002). Lernförderliche Effekte resultieren nicht aus der bloßen Auswahl zwischen Optionen, sondern aus der Übereinstimmung der Optionen mit den eigenen Interessen und Wertvorstellungen (z. B. Patall et al., 2013; Wilde et al.,

2019). Der Wert spezifischer Wahloptionen für die eigene Person wird näher in der situated expectancy-value-theory erläutert (Eccles & Wigfield, 2020).

Für das Fach Chemie ist bis dato noch unklar, wie Aufgaben variiert werden können, um unterschiedliche Interessen der Lernenden aufzugreifen. Weit verbreitet und mittlerweile politisch verordnet ist das Chemielernen in sogenannten Kontexten. Kontexte nutzen außerschulische Anwendungssituationen, um chemischen Inhalten zusätzliche Relevanz zu verleihen und den Lernprozess zu strukturieren (Nentwig et al., 2007; Gilbert, 2006). Die chemiedidaktische Forschung verweist auf differentielle Effekte der Kontextorientierung in Abhängigkeit vom spezifischen Kontext, was die Systematisierung bestehender Kontextmerkmale aus der Literatur zur Einordnung der Effekte notwendig machte. Das so von van Vorst et al. (2015) konzipierte Modell bietet einen geeigneten theoretischen Rahmen, um Kontextmerkmale zu operationalisieren und Kontexte für einen interessenbasierten Differenzierungsansatz zu variieren. Insbesondere die Unterscheidung zwischen alltäglichen und besonderen Kontexten scheint genügend Anknüpfungspunkte zu bieten, um Interessen unterschiedlicher Gruppen von Lernenden anzusprechen (siehe Habig et al., 2018; van Vorst & Aydogmus, 2021). Bisherige Studien sind allerdings limitiert in Bezug auf die Charakterisierung der Chemielernenden, die bestimmte Kontexte zum Lernen auswählen. Weitere Hinweise auf eine geringe Zufriedenheit nach der Bearbeitung selbst gewählter Kontextaufgaben erschweren in diesem Zusammenhang die Herstellung optimaler Passung zwischen Lernenden und Aufgaben für einen interessenbasierten Differenzierungsansatz (van Vorst & Aydogmus, 2021). Da empirisch gewonnene Erkenntnisse über die Adaptivität bestimmter Kontexte zur interessenbasierten Differenzierung durch systematisch variierte Kontexte notwendig sind, wurden in einer ersten Teilstudie folgende Forschungsfragen untersucht:

FF1: Welche Lernendengruppen lassen sich hinsichtlich ihrer Personenmerkmale bei der Kontextwahl unterscheiden und welchen Kontext wählen diese zum Lernen?

FF2: Wie bewerten die Lernenden ihre Wahlentscheidung nach der Bearbeitung der Aufgabe im Hinblick auf Zufriedenheit, situationales Interesse und kognitive Belastung?

Basierend auf bisherigen Forschungsarbeiten wurde vermutet, dass sich Lernende bei der Kontextwahl im Hinblick auf die Leistung in Chemie (H1.1), das Interesse an Chemie (H1.2) und ihre Wahlmotive (H1.3) unterscheiden. Darüber hinaus wurde angenommen, dass die Mehrheit der Lernenden eher einen außerfachlichen als einen innerfachlichen Kontext wählt (H1.4). Lernende mit hoher Leistung und hohem Interesse am Fach Chemie wählen hierbei besondere Kontexte (H1.5) und Lernende mit niedriger Leistung und niedrigem Interesse am Fach Chemie wählen alltägliche Kontexte (H1.6).

Mit Blick auf die zweite Forschungsfrage wurde postuliert, dass sich Lernende in den identifizierten Lernendengruppen nach der Aufgabenbearbeitung im Hinblick auf Zufriedenheit (H2.1), das situationale Interesse (H2.2) und die kognitive Belastung (H2.3) unterscheiden, was als Hinweis auf mangelnde Passung der Kontextaufgabe zu interpretieren ist.

Zur Prüfung der Hypothesen und Beantwortung der Forschungsfragen wurde eine explanative Studie mit quantitativen Erhebungsmethoden mit 495 Chemielernenden im dritten Lernjahr an Gymnasien und Gesamtschulen durchgeführt, wobei nach einem listenweisen Fallausschluss noch 347 vollständige Datensätze für die statistische Analyse zur Verfügung standen. Auf Grundlage der bisherigen Forschung zum kontextorientierten Lernen und der situated expectancy-value-theory von Eccles und Wigfield (2020) wurden die Leistung in Chemie, das Interesse an Chemie, das chemiebezogene Selbstkonzept, die Freizeitinteressen und -aktivitäten, die Berufswahlperspektive sowie die Motive zur Wahl einer kontextorientierten Aufgabe mithilfe von Leistungs- und Persönlichkeitstest erhoben. Im weiteren Verlauf der Studie konnten die Lernenden zwischen fünf unterschiedlichen Kontextaufgaben wählen: zwei alltäglichen Kontexten, zwei besonderen Kontexten und einem innerfachlichen Kontext. Diese waren – mit Ausnahme des variierten Kontextmerkmals – identisch im Hinblick auf weitere externe Aufgabenmerkmale, den Fachinhalt *saure und alkalische Lösungen* sowie das Thema *Körper und Krankheiten*. Die Lernenden bearbeiteten das Lernmaterial selbstständig an Tabletcomputern, wobei sich die Struktur des Lernmaterials am prototypischen Ablauf kontextorientierter Lerneinheiten nach dem Vorbild des deutschsprachigen *Chemie im Kontext*-Konzepts (ChiK) orientierte (Parchmann et al., 2006). Nach Bearbeitung der Aufgabe wurden Zufriedenheit, situationales Interesse und kognitive Belastung als prozessbezogene Variablen und Indikatoren für die Adaptivität der Aufgaben erhoben.

Im Rahmen einer inferenzstatistischen Analyse konnten statistisch bedeutsame Unterschiede in der letzten Chemienote und im Vorwissen als Indikator für die Leistung in Chemie, im individuellen Interesse an Chemie, im inhaltsfeldspezifischen Interesse, im Selbstkonzept, in der Selbstwirksamkeitserwartung und in den Wahlmotiven *Persönlicher Bezug* und *Überraschungsmoment* in Abhängigkeit vom gewählten Kontext nachgewiesen werden, wodurch die Hypothesen H1.1, H1.2 und H1.3 verifiziert werden konnten. Die daran anknüpfende K-Means-Clusteranalyse wurde zur Gruppierung der Lernenden anhand dieser inhaltlich relevanten Merkmale genutzt. Um einzelne Merkmale der Lernenden bei der Gruppierung nicht zu stark zu gewichten, wurden nur das Vorwissen in Chemie, das Interesse an Chemie, das Selbstkonzept und die Wahlmotive *Persönlicher Bezug* und *Überraschungsmoment* bei der Clusteranalyse berücksichtigt. Nach der formalen Gültigkeitsprüfung der Clusterlösung konnten vier unterschiedliche Cluster identifiziert werden, die sich hinsichtlich der Personenmerkmale und der gewählten Kontexte sowohl untereinander als auch im Vergleich zur Gesamtstichprobe unterscheiden.

In der Gesamtstichprobe hat ein Großteil der Lernenden (63.80 %) einen alltäglichen Kontext gewählt. Weniger Schülerinnen und Schüler entschieden sich für einen der besonderen (19.58 %) und den innerfachlichen Kontext (16.62 %). Insgesamt haben so 83.38 % der Lernenden einen außerfachlichen Kontext gewählt, wodurch die Hypothese H1.4 bestätigt werden kann.

Hinsichtlich der im Vorfeld aufgestellten Hypothesen zur Kontextwahl der verschiedenen Lernendengruppen lässt sich Folgendes festhalten: Die Lernenden mit dem höchsten Vorwissen, dem höchstem Interesse und Selbstkonzept in Chemie wählen häufiger innerfachliche Kontexte als die anderen Lernendengruppen und nur selten besondere Kontexte, wodurch die Hypothese H1.5 nicht bestätigt werden konnte. Ein vorrangiges Wahlmotiv kann in dieser Gruppe nicht identifiziert werden. Möglicherweise empfindet ein Teil der Schülerinnen und Schüler das Lernen in außerfachlichen Kontexten als störend, da sie keinen Anwendungsbezug benötigen, um sich für chemische Inhalte zu begeistern. Dabei muss berücksichtigt werden, dass ein größerer Teil der Lernenden in diesem Cluster nach wie vor den alltäglichen Kontext wählt. In diesem ersten Cluster sind daher weitere Untergruppen zu vermuten, die im Rahmen der durchgeführten Clusteranalyse nicht hinreichend getrennt werden konnten. Gegebenenfalls sind hierfür weitere spezifische Personenmerkmale notwendig, die im Rahmen der Clusteranalyse nicht berücksichtigt wurden. Da sich die Wahlmotive als deutliches Unterscheidungskriterium zwischen den Clustern und damit der Kontextwahl andeuten, könnte die Identifikation von Motiven zur Wahl von innerfachlichen Kontexten in weiteren Untersuchungen lohnend sein.

Die Lernenden des zweiten Clusters wählen vor allem aufgrund des Überraschungsmoments besondere Kontexte und verfügen im Vergleich zu den Lernenden der anderen Cluster über ein mittleres Vorwissen, ein mittleres Interesse und ein mittleres Selbstkonzept im Fach Chemie. Diese Lernendengruppe wurde vermutlich auch von van Vorst und Aydogmus (2021) im Rahmen einer Clusteranalyse identifiziert. Habig et al. (2018) konnten in diesem Zusammenhang auch zeigen, dass leistungsstärkere und höher interessierte Schülerinnen und Schüler ein höheres situationales Interesse beim Lernen in besonderen Kontexten empfinden. Besondere Kontexte könnten den Lernenden weitere Möglichkeiten bieten, ihr Wissen durch unbekannte Anwendungssituationen weiter auszudifferenzieren.

Bezüglich der Hypothese H1.6 kann festgestellt werden, dass die Lernenden mit der geringsten Leistung, dem geringsten Interesse und dem geringsten Selbstkonzept im Fach Chemie tatsächlich am häufigsten Alltagskontexte aufgrund des persönlichen Bezugs wählen und somit die Hypothese verifiziert werden kann. Eine Präferenz von Lernenden mit niedrigem Interesse und einer niedrigen Leistung wurde hierfür bereits aufgedeckt (Habig et al., 2018). Der hohe Alltagsbezug der alltäglichen Kontexte ist vermutlich notwendig, um diese Lernendengruppen überhaupt zum Chemielernen zu motivieren.

Die dargestellten Ergebnisse sind einschränkend vor dem Hintergrund der Lernenden des vierten Clusters und ihrer Kontextwahl zu betrachten. Die Lernenden weisen im Vergleich das geringste Vorwissen auf, haben aber dennoch ein höheres Interesse und Selbstkonzept als die Lernenden des zweiten Clusters. Auch hier zeigt sich, dass die Lernenden vorrangig alltägliche Kontexte aufgrund des persönlichen Bezugs wählen. Die Angabe des persönlichen Bezugs als vorrangiges Wahlmotiv in den dritten und vierten Clustern verweist in diesem Zusammenhang auf die Bedeutsamkeit des Wahlmotivs für die Wahl alltäglicher Kontext insgesamt. Weiterführende Analysen zum Leseverständnis zeigen, dass die Ergebnisse dieses Clusters aufgrund der vergleichsweise geringen Lesekompetenz möglicherweise nur eingeschränkt interpretierbar sind.

Der Vergleich der prozessbezogenen Variablen als Indikatoren für die Passung der Kontextaufgaben zeigt in keinem Cluster statistisch signifikante Unterschiede in Abhängigkeit vom gewählten Kontext. Die Hypothesen bezüglich der Unterschiede in der Zufriedenheit (H2.1), dem situationalen Interesse (H2.2) und der kognitiven Belastung (H2.3) in Abhängigkeit vom gewählten Kontext können somit nicht bestätigt werden. Die Analyse der deskriptiven Kennwerte zeigt zwar Unterschiede in Abhängigkeit vom Kontext mit kleiner bis mittlerer Effektstärke, diese können allerdings vermutlich aufgrund der geringen Größe der Teilstichproben nicht als statistisch bedeutsam ausgewiesen werden. Es gibt Hinweise darauf, dass der am häufigsten gewählte Kontext im Cluster auch zur höchsten Zufriedenheit nach der Aufgabenbearbeitung führt (z. B. in Cluster 1). Um weitere Informationen über die Passung der Kontexte zu erhalten, wurden die identifizierten Cluster anschließend hinsichtlich der prozessbezogenen Variablen verglichen. Unterschiede zwischen den Clustern zeigen sich lediglich in den Subfacetten des situationalen Interesses. Da sich die Cluster sowohl in den bearbeiteten Kontexten als auch in den Personenmerkmalen der Lernenden unterscheiden, bleibt unklar, worauf die Unterschiede im situationalen Interesse tatsächlich zurückzuführen sind. Multivariate Beziehungen zwischen den Personenmerkmalen der Lernenden, dem bearbeiteten Kontext und den prozessbezogenen Variablen wurden mittels Pfadmodellierung geprüft. Das Pfadmodell zeigt, dass die aufgabenbezogene Zufriedenheit vorrangig durch das situationale Interesse an der Aufgabe und weniger durch die kognitive Belastung determiniert wird. Das situationale Interesse wird durch die Merkmale einer Person beeinflusst und nicht durch den bearbeiteten Kontext. Ferner beeinflusst der bearbeitete Kontext nicht die Zufriedenheit nach Bearbeitung der Aufgabe. Mögliche Erklärungen für dieses Ergebnis sind, dass die bearbeiteten Kontextaufgaben von allen Lernenden ausgewählt wurden und dass die Gesamtkohorte der Lernenden nicht systematisch einen falschen Kontext wählt. Für den weiteren Verlauf des Projekts wurde daher angenommen, dass die Lernenden in der Lage sind, für sie geeignete Kontexte auszuwählen und der gewählte Kontext zugleich auch der passende ist.

In dieser Annahme besteht eine zentrale Limitation dieser Arbeit. Auch aufgrund der Analyse der prozessbezogenen Variablen in Abhängigkeit vom gewählten Kontext für jedes Cluster kann nicht davon ausgegangen werden, dass jeder Lernende einen für sich passenden Kontext

gewählt hat. Insbesondere die für das erste Cluster durchgeführte Analyse impliziert aber auch, dass die am häufigsten in den Clustern gewählten Kontexte auch zur höchsten Zufriedenheit führen. Demzufolge scheint die Wahl eines unpassenden Kontexts nur einen vergleichsweise kleinen Teil der Lernenden zu betreffen.

Über die Ursachen für die Wahl eines unpassenden Kontexts können nur Vermutungen angestellt werden. Ein Wechselspiel zwischen personenbezogenen, aufgabenbezogenen und situativen Faktoren ist hierbei wahrscheinlich. Für einen Teil der Lernenden stand vermutlich aufgrund des engen thematischen und inhaltlichen Feldes der Kontexte kein geeigneter Kontext zur Verfügung. Ergebnisse der zweiten Teilstudie deuten an, dass für die erneute Bearbeitung nach der Kontextwahl lediglich das inhaltsfeldspezifische Interesse als einziges Merkmal der Lernenden entscheidend war. Ein Teil der Lernenden hat außerdem angegeben, alle Aufgaben aufgrund eines grundsätzlich fehlenden Fachinteresses als uninteressant einzustufen. Mit hoher Wahrscheinlichkeit sind an dieser Stelle auch situativ-soziale Kriterien, wie die Wahl des Sitznachbarn oder der -nachbarin, von Bedeutung.

Aufbauend auf den Ergebnissen der ersten Teilstudie, wurde in einer zweiten Projektphase fokussiert, wie die konkrete Lehr-Lernsituation im Rahmen eines interessenbasierten Differenzierungsansatzes im Chemieunterricht aus instruktionspsychologischer Perspektive gestaltet werden sollte. Die Konzeptualisierung interessenbasierter Differenzierung durch die Wahl von Aufgaben mit unterschiedlichen Kontexten und die Ergebnisse der Selbstbestimmungstheorie werfen in diesem Zusammenhang die Frage nach der Bedeutung der Passung zwischen Kontextmerkmal und den Bedarfen der Lernenden einerseits und der selbstständigen Kontextwahl andererseits auf. In diesem Zusammenhang stellt sich folgende Forschungsfrage:

FF3: Welchen Einfluss hat die Passung zwischen Kontext und Lernenden sowie die selbstständige Kontextwahl auf den Fachwissenszuwachs, das situationale Interesse, die Zufriedenheit und die kognitive Belastung?

Angesichts der Forschung zur Selbstbestimmungstheorie wurde angenommen, dass das Lernen mit einem passenden Kontext den Fachwissenszuwachs (H3.1a), die Zufriedenheit (H3.1b) und das situationale Interesse (H3.1c) steigert sowie die kognitive Belastung (H3.1d) vermindert. Wenn mit einem passenden Kontext gelernt wird, hat die autonome Kontextwahl keinen Einfluss auf den Fachwissenszuwachs (H3.2a), die Zufriedenheit (H3.2b), das situative Interesse (H3.2c) und die kognitive Belastung (H3.2d). Darüber hinaus wurden Moderationseffekte der selbstständigen Kontextwahl und der Passung zwischen Kontextaufgabe und Person mit dem individuellen Interesse an Chemie und dem chemiebezogenen Selbstkonzept auf den Fachwissenszuwachs (H3.3a und H3.4a), die aufgabenbezogene Zufriedenheit

(H3.3b und H3.4b) und das situationale Interesse (H3.3c und H3.4c) erwartet. Theoriegeleitete Hypothesen zu möglichen Moderationseffekten durch das individuelle Interesse oder Selbstkonzept der Lernenden konnten in Bezug auf die kognitive Belastung nicht formuliert werden.

Zur Beantwortung der Forschungsfrage wurde eine experimentelle Studie mit drei Untersuchungsgruppen durchgeführt. Die Lernenden der ersten Untersuchungsgruppe wählten selbstständig eine Kontextaufgabe aus (choose & match), während die Lernenden der zweiten Gruppe eine passende Kontextaufgabe zugewiesen bekamen (no choose & match). In der dritten Gruppe wurde den Lernenden eine unpassende Kontextaufgabe zugewiesen (no choose & no match).

Die Zuordnung der Kontextaufgaben wurde datengestützt durch vorher erfasste Personenmerkmale der Lernenden durchgeführt. Anhand des in Teilstudie I erhobenen Datensatzes konnte der Zusammenhang zwischen den individuellen Personenmerkmalen und dem gewählten Kontext mithilfe von machine learning-Algorithmen vorab modelliert werden, um ein Modell für die Zuweisung der Kontexte anhand neu erhobener Personenmerkmale zu erstellen. Aufgrund der Ergebnisse der ersten Teilstudie wird angenommen, dass die gewählten Kontexte zugleich die passenden Kontexte sind. Dadurch war es nicht notwendig, den Lernenden im Vorfeld Kontexte vorzulegen, um individuell passende Kontexte zu identifizieren, wodurch potenzielle lernmindernde Effekte vermieden werden konnten (z. B. Patall et al., 2008).

Nach geeigneter Datenvorverarbeitung und der Optimierung der Hyperparameter wurden die performantesten Modelle auf einen Teil der Daten – den sogenannten Test-Datensatz – angewandt, der nicht für die Modellbildung verwendet wurde. Beim Vergleich der Modelle hat sich ein neuronales Netz als performantestes Modell herausgestellt, dessen Vorhersagegüte im weiteren Verlauf des Projektes an einer unabhängigen Stichprobe getestet wurde. Hier hat sich gezeigt, dass das Modell häufig Schwierigkeiten mit der zuverlässigen Vorhersage der besonderen und innerfachlichen Kontextwahl hat. Dieses Ergebnis kann insbesondere auf Grundlage der zweiten Teilstudie bestätigt werden, in der sich die Performanz des neuronalen Netzes nochmals verschlechterte. Mögliche Ursachen dafür liegen in der geringen Stichprobengröße des Gesamtdatensatzes sowie in der geringen Anzahl von Lernenden, die einen besonderen oder innerfachlichen Kontext gewählt haben. Ferner scheint die Performanz negativ von der psychometrischen Qualität des Messinstruments zur Erfassung der Wahlmotive beeinflusst worden zu sein, welches im Verlauf des Projektes angepasst werden musste. Insgesamt muss vor dem Hintergrund dieses Ergebnisses davon ausgegangen werden, dass einzelnen Lernenden der jeweiligen Treatmentgruppen ein passender bzw. unpassender Kontext zugewiesen wurde, obwohl das Gegenteil beabsichtigt war. Die verbesserungswürdige Performanz des neuronalen Netzes stellt eine zweite zentrale Limitation dieser Arbeit dar, die jedoch bei der weiteren Datenanalyse berücksichtigt wurde.

Die zweite Teilstudie wurde schließlich im regulären Chemieunterricht des dritten Lernjahres mit Lernenden aus Gymnasien und Gesamtschulen durchgeführt. Nach dem listenweisen Fallausschluss standen vollständige Datensätze von 217 Lernenden für die Analyse zur Verfügung. Die Lernenden haben hier über drei Schulstunden in einem Prä-Post-Design an sequenziellen Lernaufgaben zum Inhaltsfeld saure und alkalische Lösungen gearbeitet, die hinsichtlich des Kontextmerkmals variiert wurden. Demzufolge standen ein alltäglicher, besonderer und innerfachlicher Kontext für die Lerneinheit zur Verfügung. In Abhängigkeit der Treatmentgruppe wurde die kontextorientierte Lernaufgabe selbst gewählt oder passend bzw. nicht passend durch das neuronale Netz zugewiesen. Hierbei wurde für das Leseverstehen, das Vorwissen in Chemie, das Interesse an Chemie und das chemiebezogene Selbstkonzept kontrolliert. Als abhängige Variablen wurden das Fachwissen zu Beginn und am Ende der Lerneinheit sowie die aufgabenbezogene Zufriedenheit, das situationale Interesse und die kognitive Belastung nach jeder der drei Teilaufgaben gemessen. Zusätzlich wurde die aufgabenbezogene Zufriedenheit mit der gesamten Lerneinheit am Ende der Lerneinheit erhoben.

Die Datenanalyse durch ein Mehrebenenmodell zeigt, dass sowohl die aufgabenbezogene Zufriedenheit als auch das situationale Interesse im Verlauf der kontextorientierten Lerneinheit abnimmt. Eine bedeutende Abnahme ist insbesondere von der ersten zur zweiten Teilaufgabe zu verzeichnen, die vermutlich auf einen Neuigkeitseffekt (Theyßen, 2014) sowie die gestiegene kognitive Belastung zurückzuführen ist.

Lernende mit selbst gewählten und passend zugeordneten Kontextaufgaben zeigen zu allen Messzeitpunkten eine vergleichbare, aber höhere aufgabenbezogene Zufriedenheit als Lernende ohne passende Kontextzuordnung. Die Befragung der Lernenden am Ende der Lerneinheit verweist darauf, dass Lernende, die den Kontext selbst gewählt haben oder eine passende Zuordnung erhalten haben, häufiger den gleichen Kontext noch einmal bearbeiten würden als Lernende, die eine unpassende Zuordnung erhalten haben. Statistisch bedeutsame Interaktionseffekte des Treatments mit dem individuellen Interesse oder dem Selbstkonzept konnten nicht beobachtet werden, so dass die Hypothesen H3.3b und H3.4b nicht bestätigt werden.

Für das situationale Interesse zeigen sich ähnliche Befunde wie für die Zufriedenheit. Die Möglichkeit, eine kontextorientierte Aufgabe selbst zu wählen, führt nicht zu einem höheren situationalen Interesse als eine passende Aufgabenzuweisung, während eine unpassende Aufgabenzuweisung das situationale Interesse negativ beeinflusst. Die Effektstärken für die affektive Valenz des situationalen Interesses sind etwas größer als für die epistemische Valenz des situationalen Interesses. Für die epistemische Valenz zeigen sich die identifizierten Effekte außerdem vorrangig für die hochinteressierten Schülerinnen und Schüler.

Durch die Zufriedenheitsbefragung am Ende der Lerneinheit wird deutlich, dass nicht jeder Lernende aus dem choose & match- und no choose & match-Treatment einen passenden Kontext bearbeitet hat. Auf Grundlage dieser Befragung konnten die Treatmentgruppen so rekonstruiert werden, dass sich in jeder match-Gruppe nur Lernende befinden, die den Kontext

erneut bearbeiten würden, und in jeder no match-Gruppe Lernende, die den Kontext nicht erneut bearbeiten würden. Aus der Rekonstruktion der ursprünglichen Treatmentgruppen ergibt sich eine A-(choose & match), B-(no choose & match), C-(no choose & no match) und D-(choose & no match)-Gruppe. Die bereits aufgeführten Befunde konnten auch durch die rekonstruierten Gruppen repliziert werden. Darüber hinaus zeigt sich für die Wahl einer unpassenden Kontextaufgabe ein kurzfristiger Vorteil in den affektiv-motivationalen Variablen, der jedoch im Verlauf der kontextorientierten Lerneinheit immer weiter abnimmt, so dass am Ende der Lerneinheit kein Unterschied mehr zur Zuweisung der unpassenden Kontextaufgabe besteht. Insgesamt können damit die Hypothesen H3.1b und H3.2b für die Zufriedenheit und die Hypothesen 3.1c und H3.2c für die affektive Komponente des situationalen Interesses bestätigt werden.

Die kognitive Belastung der Schülerinnen und Schüler ist von der ersten zur zweiten Teilaufgabe gestiegen, was vermutlich darauf zurückzuführen ist, dass in der ersten Teilaufgabe vorrangig die makroskopische Ebene und in der zweiten Teilaufgabe die submikroskopische Ebene adressiert wird (Treagust et al., 2003). In Bezug auf die kognitive Belastung konnte außerdem ein signifikanter Haupteffekt des Treatments auf die investierte Denkanstrengung identifiziert werden. Zum zweiten Messzeitpunkt zeigen Lernende mit selbstständiger Kontextwahl eine höhere Denkanstrengung als die Lernenden mit passender Zuweisung der Kontextaufgabe. Ebenso zeigen die wenig interessierten Lernenden zum zweiten Messzeitpunkt eine höhere wahrgenommene Aufgabenschwierigkeit, wenn sie selbst einen Kontext auswählen mussten, anstatt einen passenden Kontext zugewiesen zu bekommen. Da diese Unterschiede jedoch erst zum zweiten Messzeitpunkt auftreten und zum Teil nur für die Lernenden mit geringem individuellem Interesse vorliegen, sind weitere Forschungsarbeiten notwendig, um diese Effekte näher zu untersuchen. Weiterhin liegt auch keine Evidenz für einen signifikanten Haupteffekt des Treatments auf die wahrgenommene Aufgabenschwierigkeit vor.

Alle Lernenden haben in der kontextorientierten Lerneinheit einen signifikanten und mittleren bis großen Fachwissenszuwachs erzielt. Die Hypothesen zum Fachwissenszuwachs können nicht ohne weitere Einschränkungen bestätigt werden: Es sind insbesondere die Lernenden mit einem überdurchschnittlichen Interesse und Selbstkonzept, die von der passgenauen Zuweisung kontextbezogener Aufgaben im Vergleich zur unpassenden Zuweisung profitiert haben. Insgesamt sind die Ergebnisse zum Fachwissenszuwachs aufgrund der geringen Reliabilität des Leistungstests nur eingeschränkt interpretierbar.

12 Implikationen für die Schulpraxis und Forschung

Die Ergebnisse der vorliegenden Arbeit liefern wichtige Implikationen für die Umsetzung von kontextorientiertem Lernen und interessenbasierter Differenzierung im Fach Chemie. Besondere Relevanz erhalten diese Implikationen durch die gesetzliche Vorgabe, jeden Lernenden und jede Lernende entsprechend seiner oder ihrer individuellen Bedürfnisse zu fördern und im Kontext zu unterrichten.

Für die Schulpraxis kann auf der Grundlage der Ergebnisse der ersten Teilstudie konstatiert werden, dass ein einziger Kontext nicht ausreicht, um den Bedürfnissen aller Lernenden gerecht zu werden. Schülerinnen und Schüler bevorzugen in Abhängigkeit von ihren Lernvoraussetzungen unterschiedliche Kontexte. Lernende mit geringem chemischem Vorwissen, Interesse an Chemie und chemiebezogenem Selbstkonzept bevorzugen fast ausschließlich Kontexte, die einen deutlichen Bezug zu ihrem Alltag aufweisen. Schülerinnen und Schüler mit durchschnittlichem Vorwissen, Interesse und Selbstkonzept wählen häufiger besondere Kontexte aufgrund überraschender Informationen. Leistungsstärkere, interessiertere und sich selbst als kompetent wahrnehmende Schülerinnen und Schüler wählen zu einem höheren Anteil kontextualisierte Aufgaben mit starkem chemischem Bezug. Damit sollten Lernende mit ungünstigen Lernvoraussetzungen im Unterricht nicht mit Kontexten konfrontiert werden, die einen zu starken fachlichen Bezug haben oder Informationen enthalten, die über das alltägliche Leben hinausgehen. Andererseits sollten Lernende mit günstigen Lernvoraussetzungen im Unterricht nicht in alltäglichen Kontexten lernen, da alltägliche Kontexte diese Gruppe von Lernenden vermutlich affektiv-motivational benachteiligen.

Da die Diagnose der Lernvoraussetzungen im Unterrichtsalltag anspruchsvoll ist, bietet es sich bei einem interessendifferenzierten Unterrichtsangebot an, die Lernenden selbst einen Kontext wählen zu lassen. Die Ergebnisse der ersten Teilstudie deuten darauf hin, dass ein Großteil der Lernenden in der Lage ist, einen individuell geeigneten Kontext zu wählen. Zudem ist die selbstständige Wahl im Hinblick auf affektiv-motivationale Merkmale und den Fachwissenszuwachs gleichwertig zur aufwendigeren passenden Zuweisung durch das entwickelte neuronale Netz. Allerdings sind beide Vorgehensweisen einer nicht passenden Zuweisung überlegen. Vor diesem Hintergrund ist der unreflektierte Einsatz eines einzigen Kontextes im Unterricht, der nicht auf die Lernbedürfnisse der Lernenden abgestimmt ist, *nicht* empfehlenswert. Stattdessen würde es sich anbieten, eine Auswahl an kontextorientierten Aufgaben im Unterricht bereitzustellen, aus denen die Schülerinnen und Schüler auswählen können. Hierfür können die Kontexte im Hinblick auf das Kontextmerkmal variiert werden. Dabei ist zu berücksichtigen, dass neben den Kontextmerkmalen auch das Themenfeld der Kontexte relevant ist. In diesem Zusammenhang sind insbesondere geschlechtsspezifische Effekte bekannt, die berücksichtigt werden sollten. Die im Rahmen dieser Studie entwickelten Lernmaterialien geben erste Hinweise darauf, wie unterschiedliche Kontexte in Lernmaterialien integriert werden können, ohne die fachlichen Inhalte zu verändern.

Wie aber sollen Lehrkräfte mit diesen Ergebnissen vor dem Hintergrund knapper personeller und zeitlicher Ressourcen umgehen? Aus der Forschung ist bekannt, dass die Differenzierung der Lerngruppe hohe Anforderungen an die Lehrkraft stellt. Auch aus organisatorischer Sicht ist es im Unterricht eine Herausforderung, wenn verschiedene Gruppen von Lernenden an unterschiedlichen Kontextaufgaben arbeiten. Auch wenn die Kontextaufgaben die gleichen fachlichen Inhalte fokussieren, wird es im weiteren Unterrichtsverlauf notwendig sein, die Lernenden zum Vergleich der Kontexte zu instruieren, um sie bei der Dekontextualisierung des erworbenen Wissens zu unterstützen.

Anstatt jeden einzelnen Lernenden einen Kontext wählen zu lassen, könnte die Lerngruppe durch Mehrheitsbeschluss einen Kontext für die nächste Unterrichtsstunde oder -reihe festlegen (Wilde et al., 2018). Einzelne werden dabei zwangsläufig die nicht favorisierte Option bearbeiten, allerdings sollte sich die Bearbeitung von nicht favorisierten Optionen nicht negativ auf die Motivation auswirken, wenn die Option per Mehrheitsentscheid von der Klasse festgelegt wurde (A. Meyer et al., 2013). Statt die Lernenden einzeln oder als Lerngruppe einen Kontext wählen zu lassen, könnten im Verlauf einer Unterrichtsreihe unterschiedliche Kontexte von der Lehrkraft festgelegt und eingesetzt werden, um die Bedürfnisse unterschiedlicher Lernendengruppen im Verlauf der Unterrichtsreihe zu fokussieren. Dieses Vorgehen entspricht der ursprünglichen Konzeption von Chemie im Kontext und kann die Transferleistung unterstützen, sofern die Lernenden zum Vergleich der Kontexte instruiert werden (Kehne, 2018). Wenn es nicht möglich ist, mehrere Kontexte in einer Unterrichtsstunde oder Unterrichtsreihe anzubieten, muss ein geeigneter Kontext für die Unterrichtsstunde oder Unterrichtsreihe ausgewählt werden. Bei der Wahl des Kontextes sollten in jedem Fall die spezifischen Bedürfnisse der Lernenden berücksichtigt werden. Dennoch kann an dieser Stelle eine Empfehlung ausgesprochen werden, Alltagskontexte im Unterricht zu verwenden, da die meisten Lernenden alltägliche Kontexte bevorzugen. Solch ein Unterrichtsangebot adressiert allerdings nicht die individuellen Bedürfnisse der Lernenden und ist damit insgesamt als kritisch zu beurteilen. Insbesondere leistungsstarke Schülerinnen und Schüler werden wahrscheinlich im Lernprozess durch alltägliche Kontexte motivational benachteiligt.

Zusammenfassend kann damit für die Schulpraxis festgehalten werden, dass ein differenzierter Blick auf die eingesetzten Kontexte und die Lernvoraussetzungen der Schülerinnen und Schüler notwendig ist, um den Lernprozess zu optimieren.

Wie für die schulische Praxis bleibt für die Forschung die Empfehlung, einen differenzierten Blick auf die verwendeten Kontexte und die Lernvoraussetzungen der untersuchten Stichprobe einzunehmen. Ausbleibende Effekte der Kontextorientierung in bereits durchgeführten Studien können auf die geringe Passung zwischen Kontext und Stichprobe hinweisen. Bei der Planung zukünftiger Untersuchungen sollte berücksichtigt werden, dass Interdependenzen zwischen Kontext und Stichprobe die Ergebnisse beeinflussen können. Insbesondere bei Vergleichsstudien gilt es den Einfluss des Kontexts zu kontrollieren.

Die vorliegende Arbeit bietet darüber hinaus eine Reihe von Anknüpfungspunkten für die zukünftige Forschung. Einerseits könnte es sich lohnen, weitere Anstrengungen in die Untersuchung der Adaptivität von Kontexten und einzelnen Lernendengruppen zu investieren, da die Ergebnisse der zweiten Teilstudie zeigen, wie entscheidend die Herstellung von Passung für den Lernprozess ist. Im Rahmen der ersten Teilstudie war es schwierig, die Lernenden mit innerfachlicher Kontextwahl zu isolieren. Besonders die weitere Charakterisierung von Lernenden mit innerfachlicher Kontextwahl könnte hilfreich sein, um diese Lernendengruppe besser im Lernprozess zu unterstützen. Darüber hinaus deuten sowohl die Daten der ersten als auch der zweiten Teilstudie darauf hin, dass ein Teil der Lernenden mit der Auswahl einer kontextorientierten Aufgabe und der anschließenden Bearbeitung unzufrieden war. Entscheidend für weiterführende Forschungsarbeiten wird es sein, näher zu untersuchen, ob es eine Gruppe von Lernenden gibt, die sich systematisch für falsche Kontextaufgaben entscheidet, um sie entweder bei der Wahl passender Kontexte zu unterstützen oder ihnen geeignete Kontexte zuzuweisen. Hierbei ist insbesondere zu klären, ob für die Wahl unpassender Kontexte ausschließlich personenbezogene Merkmale von Bedeutung sind oder ob nicht viel mehr ein Wechselspiel zwischen personenbezogenen, situativen sowie aufgabenbezogenen Merkmalen zu Fehlentscheidungen führt, die mehr oder weniger systematisch ausfallen. Speziell qualitative Forschungsansätze (z. B. strukturierte Interviews) könnten sich in diesem Zusammenhang als lohnend erweisen. Diese kamen in der vorliegenden Arbeit zu kurz. Allein aufgrund des sehr begrenzten Themen- und Inhaltsfeldes der entwickelten Kontexte ist beispielsweise anzunehmen, dass für einzelne Lernende in dieser Studie kein passender Kontext zur Wahl stand. Hier wäre es wünschenswert zu überprüfen, ob sich die Ergebnisse auch in anderen Themen- oder Inhaltsfeldern replizieren lassen.

Die Zuordnung der Kontexte mithilfe eines Vorhersagemodells, das mit maschinellem Lernen entwickelt wurde, stellt eine neue Methode dar, die bisher nur vereinzelt im Fokus der chemiedidaktischen Forschung stand. Gleichwohl die Prädiktionsgüte verbesserungswürdig ist, liefert die vorliegende Arbeit Einblicke in die Konzeption eines solchen Modells. Insbesondere, da der Einsatz von KI in schulischen Bildungsprozessen zukünftig wohl eine größere Bedeutung einnimmt. Es ist anzunehmen, dass die Vorhersagegüte durch die Erhebung weiterer Trainingsdaten weiter verbessert werden könnte. Bei den Ergebnissen hat sich außerdem angedeutet, dass die Anpassung des Messinstruments zur Erfassung der Wahlmotive im Verlauf des Projekts problematisch im Hinblick auf die Performanz des Modells war, so dass zukünftig davon abzusehen ist. Jacobucci und Grimm (2020) haben zudem gezeigt, dass Messfehler dazu führen können, dass ML-Algorithmen Zusammenhänge nur unzureichend modellieren. Folglich sollten die genutzten Instrumente über eine hohe Reliabilität verfügen. Die Identifizierung weiterer Merkmale könnte zu einer deutlichen Verbesserung der Prädiktionsgüte führen, insbesondere für die Vorhersage der besonderen und innerfachlichen Kontexte (Géron, 2018).

In der vorliegenden Form eignet sich das Vorhersagemodell sicherlich noch nicht für weitere Forschungsvorhaben oder den Einsatz in der Schulpraxis. Dies liegt vorwiegend daran, dass die Messung der Personenmerkmale aufwendig ist und die Vorhersage auf nur drei unterschiedliche Kontexte beschränkt ist. Zukünftige Forschungsarbeiten könnten darauf abzielen, eine größere Anzahl von Kontexten aus unterschiedlichen Themen- und Inhaltsfeldern in das Modell zu integrieren sowie den Aufwand für die Messung der Personenmerkmale zu reduzieren. Dazu könnte ein Empfehlungssystem konzipiert werden, das nicht auf der Basis gemessener Personenmerkmale, sondern auf der Basis bereits bewerteter Kontexte weitere Kontexte zur Bearbeitung vorschlägt.

Durch die zweite Teilstudie konnte die Bedeutsamkeit der Passung zwischen Lernenden und Kontext weiter herausgearbeitet werden. Hier deutet sich aufgrund der deskriptiven Daten an, dass die passende Zuweisung der selbstständigen Kontextwahl im Hinblick auf affektiv-motivationale Merkmale überlegen sein könnte. Diese Effekte deuteten sich bereits in der Studie von Wilde et al. (2018) an. Es wäre daher spannend zu sehen, ob die Ergebnisse repliziert werden können, wenn die Zuweisung passender Kontext noch besser gelingt. Insgesamt zeigt die zweite Teilstudie, dass Lernende von geeigneten Kontexten im Lernprozess profitieren, ohne genau zu fokussieren, wie Lernende mit geeigneten bzw. nicht geeigneten Kontexten im Lernprozess arbeiten. Nähere Einblicke in den Lernprozess wären möglicherweise aufschlussreich, um zu untersuchen, wie Lernende mit passenden bzw. nicht passenden Kontexten lernen. Insgesamt liefert die bisherige Forschung noch keinen Einblick in den Lernprozess mit Kontexten, die hinsichtlich der von van Vorst et al. (2015) abgeleiteten Merkmale variiert wurden, so dass in dieser Hinsicht ein besonderes Desiderat besteht. Auch im Hinblick auf die Entwicklung und Evaluation von Unterrichtskonzepten, die interindividuell unterschiedliche affektiv-motivationale Lernbedürfnisse aufgreifen, liefert die vorliegende Arbeit zwar einen ersten, aber sicherlich nicht erschöpfenden Beitrag. Nicht nur in Bezug auf die Naturwissenschafts- und Chemiedidaktik besteht erheblicher Forschungsbedarf, wie unterschiedliche affektiv-motivationale Lernvoraussetzungen im Sinne der individuellen Förderung aufgegriffen werden können, sondern in der pädagogisch-psychologischen Forschung generell – wie auch Lazarides und Schiepe-Tiska (2022) bemerken. In diesem Zusammenhang ist es außerdem Aufgabe der zukünftigen Forschung zu untersuchen, welche Interdependenzen zwischen Maßnahmen der individuellen Förderung bestehen, die sowohl kognitive als auch affektiv-motivationale Lernbedarfe berücksichtigen.

13 Verzeichnisse

13.1 Literaturverzeichnis

Aikenhead, G. S. (1994). What is STS science teaching? In J. Solomon & G. S. Aikenhead (Hrsg.), *STS Education: International Perspectives on Reform* (S. 47–59). Teacher's College Press.

Albert, M., Hurrelmann, K. & Quenzel, G. (2019). *Jugend 2019: eine Generation meldet sich zu Wort*. Beltz.

Allred, J. B. & Cena, M. E. (2020). Reading motivation in high school: Instructional shifts in student choice and class time. *Journal of Adolescent & Adult Literacy*, *64*(1), 27–35.

Alpaydin, E. (2019). *Maschinelles Lernen*. De Gruyter Oldenbourg.

Altıntaş, E. & Özdemir, A. S. (2015a). The effect of the developed differentiation approach on the achievements of the students. *Eurasian Journal of Educational Research*, *15*(61), 199–216. https://doi.org/10.14689/ejer.2015.61.11

Altıntaş, E. & Özdemir, A. S. (2015b). Evaluating a newly developed differentiation approach in terms of student achievement and teachers' opinions. *Educational Sciences: Theory & Practice*, *15*(4), 1103–1118. https://doi.org/10.12738/estp.2015.4.2540

Anderson, R. C., Shirey, L. L., Wilson, P. T. & Fielding, L. G. (1987). Interestingness of children's reading material. In R. E. Snow & M. J. Farr (Hrsg.), *Aptitude, learning, and instruction* (S. 287–299). Routledge.

Anus, S. (2015). *Evaluation individueller Förderung im Chemieunterricht: Adaptivität von Lerninhalten an das Vorwissen von Lernenden am Beispiel des Basiskonzeptes Chemische Reaktion*. Logos Verlag.

Arnold, K.-H. & Richert, P. (2008). Unterricht und Förderung: Die Perspektive der Didaktik. In K.-H. Arnold, O. Graumann & A. Rakhkochkine (Hrsg.), *Handbuch Förderung. Grundlagen, Bereiche und Methoden der individuellen Förderung von Schülern* (S. 26–35). Beltz.

Artelt, C., McElevany, N., Christmann, U., Richter, T., Groeben, N., Köster, J., Schneider, W., Stanat, P., Ostermeier, C., Schiefele, U., Valtin, R., Ring, K. & Saalbach, H. (2007). *Förderung von Lesekompetenz – Expertise*. Bundesministerium für Bildung und Forschung (BMBF).

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological review*, *64*(6, Pt. 1), 359–372.

Averbeck, D. (2021). *Zum Studienerfolg in der Studieneingangsphase des Chemiestudiums: Einfluss kognitiver und affektiv-motivationaler Variablen*. Logos Verlag.

Bacher, J., Pöge, A. & Wenzig, K. (2010). *Clusteranalyse. Anwendungsorientierte Einführung in Klassifikationsverfahren.* Oldenbourg.

Backhaus, K., Erichson, B., Gensler, S., Weiber, R. & Weiber, T. (2020). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Springer Gabler.

Bal, A. P. (2016). The effect of the differentiated teaching approach in the algebraic learning field on students' academic achievements. *Eurasian Journal of Educational Research*, *16*(63), 185–204. https://doi.org/10.14689/ejer.2016.63.11

Barber, M. (2001). *A comparison of NEAB and Salters A-level Chemistry: Student views and achievements (unveröffentlichte Masterarbeit)*. University of York.

Barnett, S. M. & Ceci, S. J. (2002). When and where do we apply what we learn? A Taxonomy for Far Transfer. *Psychological Bulletin*, *128*(4), 612–637.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Bratslavsky, E., Muraven, M. & Tice, D. M. (1998). Ego depletion: Is the active self a limited Resource? *Journal of Personality and Social Psychology*, *74*(5). https://doi.org/10.1037//0022-3514.74.5.1252

Beckmann, J. & Heckhausen, H. (2018). Motivation durch Erwartung und Anreiz. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 119–162). Springer Verlag.

Bennett, J. (2016). Bringing Science to Life: Research Evidence. In R. Taconis, P. den Brok & A. Pilot (Hrsg.), *Teachers creating context-based learning environments in science* (S. 21–39). SensePublishers.

Bennett, J., Gräsel, C., Parchmann, I. & Waddington, D. (2005). Context-based approaches to the teaching of chemistry. *International Journal of Science Education*, *27*(13), 1521–1547. https://doi.org/10.1080/09500690500153808

Bennett, J. & Holman, J. (2002). Context-based approaches to the teaching of chemistry: What are they and what are their effects? In J. K. Gilbert, O. Jong, R. Justi & D. F. Treagust (Hrsg.), *Chemical education: Towards research-based practice* (S. 165–184). Springer Netherlands.

Bennett, J. & Lubben, F. (2006). Context-based chemistry: The salters approach. *International Journal of Science Education*, *28*(9), 999–1015. https://doi.org/10.1080/09500690600702496

Bennett, J., Lubben, F. & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science Education*, *91*(3), 347–370. https://doi.org/10.1002/sce.20186

Ben-Zvi, R. (1999). Non-science oriented students and the second law of thermodynamics. *International Journal of Science Education*, *21*(12), 1251–1267. https://doi.org/10.1080/095006999290057

Black, P. & Wiliam, D. (2009). Developing a Theory of Formative Assessment, *21*(1), 5–31. https://doi.org/10.4135/9781446250808.n13

Blankenburg, J. S., Höffler, T. N. & Parchmann, I. (2016). Fostering today what is needed tomorrow: Investigating students' interest in science. *Science education*, *100*(2), 364–391. https://doi.org/10.1002/sce.21204

Bohl, T., Batzel, A. & Richey, P. (2012). Öffnung - Differenzierung - Individualisierung - Adaptivität. Charakteristika, didaktische Implikationen und Forschungsbefunde verwandter Unterrichtskonzepte zum Umgang mit Heterogenität. In T. Bohl, M. Bönsch, M. Trautmann & B. Wischer (Hrsg.), *Binnendifferenzierung Teil I: Didaktische Grundlagen und Forschungsergebnisse zur Binnendifferenzierung im Unterricht* (S. 40–69). Prolog Verlag.

Bond, T., Yan, Z. & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge.

Bortz, J. & Schuster, C. (2010). *Statistik: Für Human- und Sozialwissenschaftler*. Springer.

Brandt, H. (2020). Explorative Faktorenanalyse (EFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 575–614). Springer.

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.

Brauer, M. & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

Breimann, L. (2001). Random Forests. *Machine Learning*, 5–32.

Broman, K., Bernholt, S. & Christensson, C. (2020). Relevant or interesting according to upper secondary students? Affective aspects of context-based chemistry problems. *Research in Science & Technological Education*, 1–21. https://doi.org/10.1080/02635143.2020.1824177

Broman, K., Bernholt, S. & Parchmann, I. (2018). Using model-based scaffolds to support students solving context-based chemistry problems. *International Journal of Science Education*, *40*(10), 1176–1197. https://doi.org/10.1080/09500693.2018.1470350

Broman, K. & Parchmann, I. (2014). Students' application of chemical concepts when solving chemistry problems in different contexts. *Chem. Educ. Res. Pract.*, *15*(4), 516–529. https://doi.org/10.1039/C4RP00051J

Broman, K. & Simon, S. (2015). Upper secondary school students` choice and their ideas on how to improve chemistry education. *International Journal of Science and Mathematics Education*, *13*(6), 1255–1278. https://doi.org/10.1007/s10763-014-9550-0

Bruner, J. S. (2009). *The process of education*. Harvard university press.

Brüning, L. & Saum, T. (2010). Individualisierung und Differenzierung – aber wie? Kooperatives Lernen erschließt neue Zugänge. *Pädagogik*, *62*, 12–15.

Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion*. Pearson.

Bundesagentur für Arbeit. (2021). *Klassifikation der Berufe 2010 - überarbeitete Fassung 2020: Systematischer und alphabetischer Teil mit Erläuterungen*. Bundesagentur für Arbeit.

Burchett, W. W., Ellis, A. R., Harrar, S. W. & Bathke, A. C. (2017). Nonparametric inference for multivariate data: The R package npmv. *Journal of Statistical Software*, *76*(4), 1–18. https://doi.org/10.18637/jss.v076.i04

Bybee, R. & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, *33*(1), 7–26. https://doi.org/10.1080/09500693.2010.518644

Byrne, B. (2013). *Structural equation modeling with Mplus*. Routledge.

Byrne, B., Shavelson, R. J. & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Campbell, B. & Lubben, F. (2000). Learning science through contexts: helping pupils make sense of everyday situations. *International Journal of Science Education*, *22*(3), 239–252. https://doi.org/10.1080/095006900289859

Celik, K. N. (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"*. Logos Verlag Berlin.

Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504.

Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (S. 785–794). arXiv.

Chi, M. T. H., Feltovich, P. J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, *43*(3), 161–173. https://doi.org/10.1080/00461520802178466

Corno, L. & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching* (S. 605–629). Macmillan.

Cronbach, L. J. & Snow, R. E. (1981). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychology. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Hrsg.), *Optimal experience:*

Psychological studies of flow in consciousness (S. 15–35). Cambridge University Press.

D'Ailly, H. (2004). The role of choice in children's learning: A distinctive cultural and dender difference in efficacy, interest, and effort. *Canadian Journal of Behavioural Science*, *36*(1), 17–29. https://doi.org/10.1037/h0087212

de Charms, R. (1968). *Personal causation: The internal affective determinants of behavior*. Academic Press.

De Jong, O. (2006). Making chemistry meaningful. Conditions for successful context-based teaching. *Educación Química*, *17*, 215–221. https://doi.org/10.22201/fq.18708404e.2006.4e.66010

Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.*

Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, *39*(2), 223–238.

Deci, E. L. & Ryan, R. M. (2000). The "What" and "Why" of goal pursuits: Human needs and the self-determination of behavior, *11*(4), 227–268.

Dierks, P. O., Höffler, T. N., Blankenburg, J. S., Peters, H. & Parchmann, I. (2016). Interest in science: A RIASEC-based analysis of students' interests. *International Journal of Science Education*, *38*(2), 238–258. https://doi.org/10.1080/09500693.2016.1138337

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121–149. https://doi.org/10.1177/0748175610373459

Dorfner, T., Förtsch, C. & Neuhaus, B. J. (2018). Effects of three basic dimensions of instructional quality on students' situational interest in sixth-grade biology instruction. *Learning and Instruction*, *56*, 42–53. https://doi.org/10.1016/j.learninstruc.2018.03.001

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer.

Dorschu, A. (2013). *Die Wirkung von Kontexten in Physikkompetenztestaufgaben*. Logos Berlin.

Dumont, H. (2019). Neuer Schlauch für alten Wein? Eine konzeptuelle Betrachtung von individueller Förderung im Unterricht. *Zeitschrift für Erziehungswissenschaft*, *22*(2), 249–277. https://doi.org/10.1007/s11618-018-0840-0

Dunn, R., Griggs, S. A., Olson, J., Beasley, M. & Gorman, B. S. (1995). A meta-analytic validation of the Dunn and Dunn model of learning-style preferences. *The Journal of Educational Research*, *88*, 353–362.

Duranti, A. & Goodwin, C. (1992). Rethinking context: an introduction. In A. Duranti & C. Goodwin (Hrsg.), *Rethinking context: language as an interactive phenomenon* (S. 1–42). Cambridge University Press.

Eccles, J. S. (2002). Motivational Beliefs, Values, and Goals. *Annual Reviews Psychology*, *53*, 109–132.

Eccles, J. S. (2005). Subjective task values and the Eccles et al. model of achievement-related choices. In A. J. Elliot & C. S. Dweck (Hrsg.), *Handbook of Competence and Motivation* (S. 105–121). The Guilford Press.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L. & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Hrsg.), *Achievement and achievement motivation* (S. 74–146). W. H. Freeman.

Eccles, J. S. & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Eid, M., Gollwitzer, M. & Schmitt, M. (2017). *Statistik und Forschungsmethoden*. Beltz.

Eid, M. & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Hogrefe.

Elster, D. (2007). In welchen Kontexten sind naturwissenschaftliche Inhalte für Jugendliche interessant? Ergebnisse der ROSE-Erhebung in Österreich und Deutschland. *PLUS LUCIS*, *3*, 2–8.

Embretson, S. E. (2009). *Item Response Theory for Psychologists*. Psychology Press.

Engeln, K. (2004). *Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken*. Logos Verlag.

Everitt, B., Landau, S., Leese, M. & Stahl, D. (2011). *Cluster analysis*. *Wiley series in probability and statistics*. Wiley. https://doi.org/10.1002/9780470977811

Fechner, S. (2009). *Effects of context-oriented learning on student interest and achievement in chemistry education*. Logos.

Fechner, S., van Vorst, H., Kölbach, E. & Sumfleth, E. (2015). It's the Situation That Matters: Affective Involvement in Context-Oriented Learning Tasks. In M. Kahveci & M. Orgill (Hrsg.), *Affective Dimensions in Chemistry Education* (S. 159–176). Springer Verlag.

Fensham, P. J. (1988). *Development and Dilemmas in Science Education*. Falmer.

Ferber, N. (2014). *Entwicklung und Validierung eines Testinstruments zur Erfassung von Kompetenzentwicklung im Fach Chemie in der Sekundarstufe I*. Logos Verlag Berlin.

Field, A., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. Sage.

Fine, G. A. (1987). *With the boys*. University of Chicago Press.

Finkelstein, N. (2005). Learning Physics in Context: A study of student learning about electricity and magnetism. *International Journal of Science Education*, *27*(10), 1187–1209. https://doi.org/10.1080/09500690500069491

Fischer, C. (2014). *Individuelle Förderung als schulische Herausforderung* (1. Aufl.). *Schriftenreihe des Netzwerk Bildung: Bd. 31*. Friedrich-Ebert-Stiftung Abt. Studienförderung.

Fischer, L., Rohm, T., Carstensen, C. H. & Gnambs, T. (2021). Linking of Rasch-Scaled Tests: Consequences of Limited Item Pools and Model Misfit. *Frontiers in psychology*, *12*, 633896. https://doi.org/10.3389/fpsyg.2021.633896

Fisher, R. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, *7*(2), 179–188.

Flake, J. K. & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Flowerday, T. & Schraw, G. (2000). Teacher beliefs about instructional choice: A phenomenological study. *Journal of Educational Psychology*, *92*(4), 634–645. https://doi.org/10.1037/0022-0663.92.4.634

Flowerday, T. & Schraw, G. (2003). Effect of choice on cognitive and affective engagement. *The Journal of Educational Research*, *96*(4), 207–215. https://doi.org/10.1080/00220670309598810

Flowerday, T., Schraw, G. & Stevens, J. (2004). The role of choice and interest in reader engagement. *The Journal of Experimental Education*, *72*(2), 93–114. https://doi.org/10.3200/JEXE.72.2.93-114

Flowerday, T. & Shell, D. F. (2015). Disentangling the effects of interest and choice on learning, engagement, and attitude. *Learning and Individual Differences*, *40*, 134–140. https://doi.org/10.1016/j.lindif.2015.05.003

Freudenthal, H. (1968). Why to teach mathematics so as to be useful. *Educational Studies in Mathematics*, *1*, 3–8.

Frey, A., Hartig, J. & Rupp, A. A. (2009). An NCME Instructional Module on Booklet Designs in Large-Scale Assessments of Student Achievement: Theory and Practice. *Educational Measurement: Issues and Practice*, *28*(3), 39–53.

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.

Gardner, H. (2016). Multiple Intelligences: Prelude, Theory, and Aftermath. In R. J. Sternberg, S. T. Fiske & D. J. Foss (Hrsg.), *Scientists Making a Difference* (S. 167–170). Cambridge University Press.

Geisser, S. (1993). *Predictive Inference: An Introduction*. Chapman and Hall.

George, J. M. & Lubben, F. (2002). Facilitating teachers' professional growth through their involvement in creating context-based materials in science. *International Journal of Educational Development*, *22*(6), 659–672. https://doi.org/10.1016/S0738-0593(01)00033-5

Géron, A. (2018). *Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme*. O'Reilly.

Giese, J. L. & Cote, J. A. (2000). Defining Consumer Satisfaction. *Academy of Marketing Science Review*, *1*(1), 1–24.

Gijsbers, D., Putter-Smits, L. de & Pepin, B. (2020). Changing students' beliefs about the relevance of mathematics in an advanced secondary mathematics class. *International Journal of Mathematical Education in Science and Technology*, *51*(1), 87–102. https://doi.org/10.1080/0020739X.2019.1682698

Gilbert, J. K. (2006). On the nature of "Context" in chemical education. *International Journal of Science Education*, *28*(9), 957–976. https://doi.org/10.1080/09500690600702470

Gilbert, J. K., Bulte, A. M. & Pilot, A. (2011). Concept development and transfer in context-based science cducation. *International Journal of Science Education*, *33*(6), 817–837. https://doi.org/10.1080/09500693.2010.493185

Glynn, S. M. & Koballa, T. R. (2006). Motivation to learn college science. In J. J. Mintzes & W. H. Leonard (Hrsg.), *Handbook of College Science Teaching* (S. 25–32). National Science Teachers Association Press.

Gräber, W. (2011). German high school students' interest in chemistry – A comparison between 1990 and 2008. *Educación Química*, *22*(2), 134–140. https://doi.org/10.1016/S0187-893X(18)30125-3

Gruber, H. & Stamouli, E. (2009). Intelligenz und Vorwissen. In W. Elke & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 27–48). Springer Verlag.

Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M. & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology*, *32*(3), 282–313. https://doi.org/10.1016/j.cedpsych.2006.05.004

Habig, S. (2017). *Systematisch variierte Kontextaufgaben und ihr Einfluss auf kognitive und affektive Schülerfaktoren*. Logos Verlag.

Habig, S., van Vorst, H. & Sumfleth, E. (2018). Merkmale kontextualisierter Lernaufgaben und ihre Wirkung auf das situationale Interesse und die Lernleistung von Schülerinnen und Schülern. *Zeitschrift für Didaktik der Naturwissenschaften*, *24*(1), 99–114. https://doi.org/10.1007/s40573-018-0077-8

Hammerstein, S., König, C., Dreisörner, T. & Frey, A. (2021). Effects of COVID-19-related school closures on student achievement: A Systematic Review. *Frontiers in psychology*, *12*, 1–8. https://doi.org/10.3389/fpsyg.2021.746289

Han, H., Wang, W. Y. & Mao, B. H. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In *International Conference on intelligent computing* (S. 878–887). Springer.

Hardy, I., Descristan, J. & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for educational research online 11*(2), 169–191.

Hardy, I., Hertel, S., Kunter, M., Klieme, E., Warwas, J., Büttner, G. & Lühken, A. (2011). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. *Zeitschrift für Pädagogik*, *57*(6), 819–834. https://doi.org/10.1007/978-3-658-04479-4_2

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer New York. https://doi.org/10.1007/b94608

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge. https://doi.org/10.4324/9780203181522

Hauerstein, M.-T. (2019). *Untersuchung zur Effektivität von Strukturierung und Binnendifferenzierung im Chemieunterricht der Sekundarstufe I. Evaluation der Strukturierungshilfe Lernleiter*. Logos Verlag.

Haugwitz, M. (2009). *Kontextorientiertes Lernen und Concept Mapping im Fach Biologie*. Logos Verlag.

Hayes, A. F. & Coutts, J. J. (2020). Use Omega rather than Cronbach's Alpha for estimating reliability. But… *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

He, H., Bai, Y., Garcia, E. A. & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *Proceedings of IJCNN 2008 (IEEE World Congress on Computational Intelligence)* (S. 1322–1328). IEEE. https://doi.org/10.1109/IJCNN.2008.4633969

Heider, F. (1958). *The psychology of interpersonal relations*. John Wiley & Sons Inc. https://doi.org/10.1037/10628-000

Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Klett Kallmeyer.

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, *52*(1), 258–271. https://doi.org/10.1016/j.csda.2006.11.025

Hennig, C. (2023). *fpc: Flexible Procedures for Clustering. R package version 2.2-10.*

Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: concerns and considerations. *Computational Methods in Social Sciences*, *3*(1), 5–17.

Heymann, H. W. (2010). Binnendifferenzierung – eine Utopie? Pädagogischer Anspruch, didaktisches Handwerk, Realisierungschancen. *Pädagogik*, *62*, 6–10.

Hidi, S. & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127.

Hidi, S., Renninger, K. A. & Krapp, A. (2004). Interest, a motivational variable that combines affective and cognitive functioning. In D. Y. Dai & R. J. Sternberg (Hrsg.), *Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Developmen* (S. 89–115). Routledge. https://doi.org/10.4324/9781410610515-11

Hoffmann, L., Häußler, P. & Lehrke, M. (1998). *Die IPN-Interessenstudie Physik* (Bd. 158). IPN.

Höft, L., Bernholt, S., Blankenburg, J. S. & Winberg, M. (2019). Knowing more about things you care less about: Cross-sectional analysis of the opposing trend and interplay between conceptual understanding and interest in secondary school chemistry.

Journal of Research in Science Teaching, *56*(2), 184–210. https://doi.org/10.1002/tea.21475

Hülsmann, C. (2015). *Kurswahlmotive im Fach Chemie. Eine Studie zum Wahlverhalten und Erfolg von Schülerinnen und Schülern in der gymnasialen Oberstufe*. Logos Verlag.

Hussy, W. (2013). Quantitative Forschungsmethoden. In W. Hussy, M. Schreier & G. Echterhoff (Hrsg.), *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (S. 115–164).

Ivey, G. & Johnston, P. H. (2013). Engagement with young adult literature: Outcomes and processes. *Reading Research Quarterly*, *48*(3), 255–275.

Iyengar, S. S. & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*(6), 995–1006.

Jacobucci, R. & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, *15*(3), 809–816. https://doi.org/10.1177/1745691620902467

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R*. Springer.

Jiang, S., Wang, C. & Weiss, D. J. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in psychology*, *7*, 1–10. https://doi.org/10.3389/fpsyg.2016.00109

Jorgenson, T. D., Pornprasertmanit, S., Schoemann, A. M. & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling. R package version 0.5-6*. https://CRAN.R-project.org/package=semTools

Kallweit, I. & Melle, I. (2017). Selbsteinschätzungsbögen als Instrument zur individuellen Förderung im Chemieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, *23*(1), 143–163. https://doi.org/10.1007/s40573-017-0063-6

Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, *23*(34), 197–209. https://doi.org/10.1024/1010-0652.23.34.197

Kasanda, C., Lubben, F., Gaoseb, N., Kandjeo-Marenga, U., Kapenda, H. & Campbell, B. (2005). The role of everyday contexts in learner-centred teaching: The practice in Namibian secondary schools. *International Journal of Science Education*, *27*(15), 1805–1823. https://doi.org/10.1080/09500690500277854

Katz, I. & Assor, A. (2007). When choice motivates and when it does not. *Educational Psychology Review*, *19*, 429–442. https://doi.org/10.1007/s10648-006-9027-y

Kehne, F. (2019). *Analyse des Transfers von kontextualisiert erworbenem Wissen im Fach Chemie*. Logos Verlag Berlin.

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.

Kelava, A., Noventa, S. & Robitzsch, A. (2020). Überblick über Modelle der Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 425–446). Springer.

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, *10*(3), 2–10. https://doi.org/10.1007/BF02905780

King, D. T. (2012). New perspectives on context-based chemistry education: Using a dialectical sociocultural approach to view teaching and learning. *Studies in Science Education*, *48*(1), 51–87. https://doi.org/10.1080/03057267.2012.655037

King, D. T. & Ritchie, S. M. (2013). Academic success in context-based chemistry: Demonstrating fluid transitions between concepts and context. *International Journal of Science Education*, *35*(7), 1159–1182. https://doi.org/10.1080/09500693.2013.774508

Klafki, W. & Stöcker, H. (1976). Innere Differenzierung des Unterrichts. *Zeitschrift für Pädagogik*, *22*(4), 497–523.

Klieme, E. & Warwas, J. (2011). Konzepte der individuellen Förderung. *Zeitschrift für Pädagogik*, *57*(6), 805–817.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. The Guilford Press.

Kölbach, E. & Sumfleth, E. (2013). Analyse von Kontexteffekten beim Lernen mit Lösungsbeispielen im Fach Chemie. *Zeitschrift für Didaktik der Naturwissenschaften*, *19*, 159-188.

Kortland, K. (2005). Physics in personal, social and scientific contexts: A retrospective view on the Dutch Physics Curriculum Development Project PLON. In P. Nentwig & D. Waddington (Hrsg.), *Making it relevant: Context based learning of science* (S. 67–89). Waxmann.

Krapp, A. (1992a). Interesse, Lernen und Leistung. Neue Forschungsansätze in der Pädagogischen Psychologie. *Zeitschrift für Pädagogik*, *38*(5), 747–770.

Krapp, A. (1992b). Das Interessenkonstrukt. Bestimmungsmerkmale der Interessenhandlung und des individuellen Interesses aus der Sicht einer Person-Gegenstands-Konzeption. In Andreas Krapp & M. Prenzel (Hrsg.), *Interesse, Lernen, Leistung. Neuere Ansätze einer pädagogisch-psychologischen Interessenforschung* (S. 297–329). Aschendorff.

Krapp, A. (1999a). Entwicklung und Förderung von Interessen im Unterricht. *Psychologie in Erziehung und Unterricht*, *44*(3), 185–201.

Krapp, A. (1999b). Intrinsische Lernmotivation und Interesse. Forschungsansätze und konzeptuelle Überlegungen. *Zeitschrift für Pädagogik*, *45*(3), 387–406.

Krapp, A. (2002). Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*(4), 383–409. https://doi.org/10.1016/S0959-4752(01)00011-1

Krapp, A. & Prenzel, M. (2011). Research on Interest in Science: Theories, methods, and findings. *International Journal of Science Education*, *33*(1), 27–50. https://doi.org/10.1080/09500693.2010.518645

Kuhn, J. (2010). *Authentische Aufgaben im theoretischen Rahmen von Instruktions- und Lehr-Lern-Forschung: Optimierung von Ankermedien für eine neue Aufgabenkultur im Physikunterricht*. Vieweg + Teubner.

Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.

Kuhn, M. & Johnson, K. (2020). *Feature engineering and selection: A practical approach for predictive models*. *Data Science Series*. CRC Press.

Kuhn, M. & Silge, J. (2022). *Tidy Modeling with R: A framework for modeling in the Tidyverse*. O'Reilly. https://www.tmwr.org

Kulik, C.-L. C. & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings, *19*(3), 415–428.

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. *Materialien aus der Bildungsforschung: Nr. 72*. Max-Planck-Inst. für Bildungsforschung. http://hdl.handle.net/hdl:11858/00-001M-0000-0023-9987-C

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lange, B. & Parchmann, I. (2003). Untersuchung zur Entwicklung fachspezifischen Wissens bei Schülerinnen und Schülern in einem Unterricht nach Chemie im Kontext. In A. Pitton (Hrsg.), *Außerschulisches Lernen in Physik und Chemie* (S. 269–271). LIT Verlag.

Lanquillon, C. (2019). Grundzüge des maschinellen Lernens. In S. Schacht & C. Lanquillon (Hrsg.), *Blockchain und maschinelles Lernen* (S. 89–142). Springer Berlin Heidelberg.

Lazarides, R. & Schiepe-Tiska, A. (2022). Heterogenität motivationaler Merkmale im Unterrichtskontext. *Zeitschrift für Erziehungswissenschaft*, *25*(2), 249–267. https://doi.org/10.1007/s11618-022-01082-3

Lenth, R. (2022). *emmeans: Estimated marginal means, aka Least-squares means*. https://CRAN.R-project.org/package=emmeans

LimeSurvey Project Team. (2012). *LimeSurvey: An Open Source survey tool*. http://www.limesurvey.org

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B. & d'Apollonia, S. (1996). Within-Class Grouping: A Meta-Analysis. *Review of Educational Research*, *66*(4), 423–458.

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

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior research methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Maas, C. J. M. & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92.

Mackinnon, S., Curtis, R. & O'Connor, R. (2022). Tutorial in longitudinal measurement invariance and cross-lagged panel models using lavaan. *Meta-Psychology*, *6.* https://doi.org/10.15626/MP.2020.2595

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in psychology*, *6*, 2. https://doi.org/10.3389/fpsyg.2015.00002

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: reciprocal effects models of causal ordering. *Child development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mayer, R. E. (2020). *Multimedia learning*. Cambridge University Press.

Maykus, S., Böttcher, W., Liesegang, T. & Altermann, A. (2011). Individuelle Förderung in der Ganztagsschule. Theoretisch-empirische Reflexionen zu Anspruch und Konsequenzen eines (sozial-)pädagogischen Programms im professionellen Handeln. In L. Stecher, H.-H. Krüger & T. Rauschenbach (Hrsg.), *Ganztagsschule – Neue Schule?* (S. 125–142). VS.

Mayring, P. (2022). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz.

McDermott, R. (1993). The acquisition of a child by a learning disability. In S. Chaiklin & J. Lave (Hrsg.), *Understanding practice: Perspectives on activity and context* (S. 269–305). Cambridge University Press.

McKay, M. D., Beckman, R. J. & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239–245. https://doi.org/10.2307/1268522

Merzyn, G. (2013). *Naturwissenschaften, Mathematik, Technik - immer unbeliebter? Die Konkurrenz von Schulfächern um das Interesse der Jugend im Spiegel vielfältiger Untersuchungen*. Schneider Verlag Hohengehren GmbH.

Mestre, J. P. (2002). Probing adults' conceptual understanding and transfer of learning via problem posing. *Journal of Applied Developmental Psychology*, *23*(1), 9–50. https://doi.org/10.1016/S0193-3973(01)00101-0

Meteyard, L. & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 1–22. https://doi.org/10.1016/j.jml.2020.104092

Meyer, A., Meyer-Ahrens, I. & Wilde, M. (2013). The beneficial effects of non-received choice: A study on intrinsic motivation in biology education. *European Journal of Educational Research*, *2*(4), 185–190.

Meyer, H. (2011). *Was ist guter Unterricht?*. Cornelsen Scriptor.

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (2008). *Chemie: Kernlehrplan für das Gymnasium - Sekundarstufe I in Nordrhein-Westfalen.*

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (2013). *Naturwissenschaften Biologie, Chemie, Physik: Kernlehrplan für die Gesamtschule - Sekundarstufe I in Nordrhein-Westfalen.*

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (2019). *Chemie: Kernlehrplan für die Sekundarstufe I Gymnasium in Nordrhein-Westfalen.*

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology*, *85*, 424–436.

Moller, A. C., Deci, E. L. & Ryan, R. M. (2006). Choice and ego-depletion: the moderating role of autonomy. *Personality & social psychology bulletin*, *32*(8), 1024–1036. https://doi.org/10.1177/0146167206288008

Möller, J. & Trautwein, U. (2009). Selbstkonzept. In W. Elke & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 179–204). Springer Verlag.

Moosbrugger, H. & Brandt, H. (2020). Itemkonstruktion und Antwortverhalten. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 67–89). Springer.

Moosbrugger, H., Gäde, J. C., Schermelleh-Engel, K. & Rauch, W. (2020). Klassische Testtheorie (KKT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 275–304). Springer.

Moosbrugger, H., Schermelleh-Engel, K., Gäde, J. C. & Kelava, A. (2020). Testtheorien im Überblick. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 251–271). Springer.

Morgan, D. N. & Wagner, C. W. (2013). “What's the Catch?”: Providing Reading Choice in a High School Classroom. *Journal of Adolescent & Adult Literacy*, *56*(8), 659–667.

Nentwig, P., Demuth, R., Parchmann, I., Ralle, B. & Gräsel, C. (2007). Chemie im Kontext: Situating learning in relevant contexts while systematically developing basic chemical concepts. *Journal of Chemical Education*, *84*(9), 1439–1444. https://doi.org/10.1021/ed084p1439

Neumann, K. (2014). Rasch-Analyse naturwissenschaftsbezogener Leistungstests. In D. Krüger, Parchman & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 350–377). Springer.

Nguyen, C. N. & Zeigermann, O. (2021). *Machine Learning: Kurz & gut* (2. Aufl.). *kurz & gut*. O'Reilly.

OECD. (2016). *PISA 2015 Results (Volume II): Policies and Practices for succesful schools*. OECD. https://doi.org/10.1787/9789264267510-en

O'Neill, T., Peabody, M., Tan, R. & Du, Y. (2013). How much item drift is too much. *Rasch Measurement Transactions*, 1423–1424.

Palmer, D. H. (2009). Student interest generated during an inquiry skills lesson. *Journal of Research in Science Teaching*, *46*(2), 147–165. https://doi.org/10.1002/tea.20263

Parchmann, I., Gräsel, C., Baer, A., Nentwig, P., Demuth, R. & Ralle, B. (2006). "Chemie im Kontext": A symbiotic implementation of a context-based teaching and learning approach. *International Journal of Science Education*, *28*(9), 1041–1062. https://doi.org/10.1080/09500690600702512

Parchmann, I., Ralle, B. & Di Fuccia, D. (2008). Entwicklung und Struktur der Unterrichtskonzeption Chemie im Kontext. In R. Demuth, C. Gräsel, I. Parchmann & B. Ralle (Hrsg.), *Chemie im Kontext: Von der Innovation zur nachhaltigen Verbreitung eines Unterrichtskonzepts* (S. 9–48). Waxmann.

Parchmann, I. & Schmidt, S. (2003). *Students' pre-conceptions as a tool to reflect and to design teach-ing and learning processes: a study from the project Chemie im Kontext*. Beitrag zur ESERA Konferenz.

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., Pierczynski, M. & Allen, M. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research*, *88*(2), 205–242. https://doi.org/10.3102/0034654317743198

Patall, E. A. (2013). Constructing motivation through choice, interest, and interestingness. *Journal of Educational Psychology*, *105*(2), 522–534. https://doi.org/10.1037/a0030307

Patall, E. A., Cooper, H. & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings. *Psychological Bulletin*, *134*(2), 270–300. https://doi.org/10.1037/0033-2909.134.2.270

Patall, E. A., Sylvester, B. J. & Han, C. (2014). The role of competence in the effects of choice on motivation. *Journal of Experimental Social Psychology*, *50*, 27–44. https://doi.org/10.1016/j.jesp.2013.09.002

Pentecost, T. C. & Barbera, J. (2013). Measuring learning gains in chemical education: A comparison of two methods. *Journal of Chemical Education*, *90*(7), 839–845.

Pfeifer, P. & Sommer, K. (2019). Fachliche Grundlagen des Chemieunterrichts: Die Basiskonzepte der Chemie. In K. Sommer & J. Wambach-Laicher (Hrsg.), *Konkrete Fachdidaktik Chemie: Grundlagen für das Lernen und Lehren im Chemieunterricht* (S. 17–44). Aulis Verlag.

Pilot, A. & Bulte, A. M. (2006). The use of "Contexts" as a challenge for the chemistry curriculum: Its successes and the need for further development and understanding. *International Journal of Science Education*, *28*(9), 1087–1112. https://doi.org/10.1080/09500690600730737

Podschuweit, S. & Bernholt, S. (2018). Composition-Effects of Context-based Learning Opportunities on Students' Understanding of Energy. *Research in Science Education*, *48*(4), 717–752. https://doi.org/10.1007/s11165-016-9585-z

Pope, M. & Keen, T. (1981). *Personal construct psychology in education*. Academic Press.

Prins, G. T., Bulte, A. M. & Pilot, A. (2018). Designing context-based teaching materials by transforming authentic scientific modelling practices in chemistry. *International Journal of Science Education*, *40*(10), 1108–1135. https://doi.org/10.1080/09500693.2018.1470347

R Core Team. (2022). *R: A language and environment for statistical computing*. https://www.R-project.org/

Raguse, K. & Weber-Peukert, G. (2019). Unterrichtsprozessplanung und -analyse: Lernaufgaben. In K. Sommer & J. Wambach-Laicher (Hrsg.), *Konkrete Fachdidaktik Chemie: Grundlagen für das Lernen und Lehren im Chemieunterricht* (S. 414–426). Aulis Verlag.

Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2010a). *Quantitative Methoden 1: Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. Springer Verlag.

Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2010b). *Quantitative Methoden 2: Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. Springer Verlag.

Raykov, T. & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.

Raykov, T. & Marcoulides, G. A. (2015). A direct latent variable modeling based method for point and interval estimation of coefficient alpha. *Educational and psychological measurement*, *75*(1), 146–156. https://doi.org/10.1177/0013164414526039

Reeve, J., Nix, G. & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology*, *95*(2), 375–392. https://doi.org/10.1037/0022-0663.95.2.375

Reinecke, J. (2014). *Strukturgleichungsmodelle in den Sozialwissenschaften*. De Gruyter Oldenbourg.

Reise, S. P. & Henson, J. M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of personality assessment*, *81*(2), 93–103. https://doi.org/10.1207/S15327752JPA8102_01

Rheinberg, F. & Vollmeyer, R. (2012). *Motivation*. W. Kohlhammer.

Rheinberg, F. & Wendland, M. (2002). Veränderung der Lernmotivation in Mathematik: Eine Komponentenanalyse. *Zeitschrift für Pädagogik*, *45*(Beiheft), 308–320.

Riedel, A. (2008). *Innere Differenzierung–Herausforderung für modernen Unterricht*. http://riedlpublikationen.userweb.mwn.de/pdf/inneredifferenzierungriedl2008.pdf

Rights, J. D. & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological methods*, *24*(3), 309–338. https://doi.org/10.1037/met0000184

Ritter, J. (1976). *Historisches Wörterbuch der Philosophie. 4. I - K*. Schwabe.

Robitzsch, A., Kiefer, T. & Wu, M. (2022). *TAM: Test Analysis Modules*. https://CRAN.R-project.org/package=TAM

Ropohl, M. (2010). *Modellierung von Schülerkompetenzen im Basiskonzept Chemische Reaktion: Entwicklung und Analyse von Testaufgaben*. *Studien Zum Physik- und Chemielernen Ser*. Logos Verlag Berlin.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Huber.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Ryan, R. M. & Connell, J. P. (1989). *Self-regulation questionnaire (unpublished manuscript)*. University of Rochester.

Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist*, *55*(1), 68–78.

Ryan, R. M. & Deci, E. L. (2002). Overview of self-determination theory: An organismic-dialectical perspective. In E. L. Deci & R. M. Ryan (Hrsg.), *Handbook of self-determination research* (S. 3–33). University of Rochester Press.

Ryan, R. M. & Deci, E. L. (2006). Self-Regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of Personality*, *74*(6), 1557-1586.

Ryan, R. M. & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press.

Ryan, R. M. & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Ryan, R. M., Deci, E. L., Vansteenkiste, M. & Soenens, B. (2021). Building a science of motivated persons: Self-determination theory's empirical approach to human experience and the regulation of behavior. *Motivation Science*, *7*(2), 97–110. https://doi.org/10.1037/mot0000194

Schad, D. J., Vasishth, S., Hohenstein, S. & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 1–40. https://doi.org/10.1016/j.jml.2019.104038

Schendera, C. F. G. (2010). *Clusteranalyse mit SPSS: mit Faktorenanalyse*. Oldenbourg.

Schermelleh-Engel, K. & Gäde, J. C. (2020). Modellbasierte Methoden der Reliabilitätsschätzung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 335–369). Springer.

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, *26*(3-4), 299–323. https://doi.org/10.1080/00461520.1991.9653136

Schiefele, U. & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences*, *8*(2), 141–160.

Schiefele, U., Krapp, A. & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi & A. Krapp (Hrsg.), *The role of interest in learning and development* (S. 183–212). Psychology Press. https://doi.org/10.4324/9781315807430-18

Schiepe-Tiska, A., Rönnebeck, S., Schöps, K., Neumann, K., Schmidtner, S., Parchmann, I. & Prenzel, M. (2016). Naturwissenschaftliche Kompetenz in PISA 2015 - Ergebnisse des internationalen Vergleichs mit einem modifizierten Testansatz. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015: Eine Studie zwischen Kontinuität und Innovation* (S. 45–98). Waxmann.

Schiepe-Tiska, A., Simm, I. & Schmidtner, S. (2016). Motivationale Orientierungen, Selbstbilder und Berufserwartungen in den Naturwissenschaften in PISA 2015. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015: Eine Studie zwischen Kontinuität und Innovation* (S. 99–132). Waxmann.

Schmeck, A., Opfermann, M., van Gog, T., Paas, F. G. W. C. & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instructional Science*, *43*(1), 93–114. https://doi.org/10.1007/s11251-014-9328-3

Schneider, S., Nebel, S., Beege, M. & Rey, G. D. (2018). The autonomy-enhancing effects of choice on cognitive load, motivation and learning with digital media. *Learning and Instruction*, *58*, 161–172. https://doi.org/10.1016/j.learninstruc.2018.06.006

Schneider, W., Schlagmüller, M. & Ennemoser, M. (2017). *LGVT 5-12: Lesegeschwindigkeits- und -verständnistest für die Klassen 5-12: Manual*. Hogrefe.

Schraw, G., Flowerday, T. & Lehman, S. (2001). Increasing situational interest in the classroom. *Educational Psychology Review*, *13*(3), 211–224.

Schraw, G., Flowerday, T. & Reisetter, M. F. (1998). The role of choice in reader engagement. *Journal of Educational Psychology*, *90*(4), 705–714. https://doi.org/10.1037/0022-0663.90.4.705

Schulz, E., Speekenbrink, M. & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, *85*, 1–16. https://doi.org/10.1016/j.jmp.2018.03.001

Schwantes, M. (2021). *Entwicklung und Erprobung eines Fachwissenstests zum Inhaltsfeld „saure und alkalische Lösungen" für die Sekundarstufe I (unveröffentlichte Masterarbeit)*. Universität Duisburg-Essen.

Schwartz, B. (2000). Self-Determination: The Tyranny of Freedom. *American Psychologist*, *55*(1), 79–88.

Sevian, H., Dori, Y. J. & Parchmann, I. (2018). How does STEM context-based learning work: what we know and what we still do not know. *International Journal of Science Education*, *40*(10), 1095–1107. https://doi.org/10.1080/09500693.2018.1470346

Sevian, H., Hugi-Cleary, D., Ngai, C., Wanjiku, F. & Baldoria, J. M. (2018). Comparison of learning in two context-based university chemistry classes. *International Journal of Science Education*, *40*(10), 1239–1262. https://doi.org/10.1080/09500693.2018.1470353

Sevian, H. & Talanquer, V. (2014). Rethinking chemistry: a learning progression on chemical thinking. *Chem. Educ. Res. Pract.*, *15*(1), 10–23. https://doi.org/10.1039/C3RP00111C

Singmann, H. & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In D. H. Spieler & E. Schumacher (Hrsg.), *New methods in cognitive psychology* (S. 4–31). Psychology Press.

Sjøberg, S. & Schreiner, C. (2010). *The ROSE project. An overview and key findings.* https://www.cemf.ca/PDFs/SjobergSchreinerOverview2010.pdf

Smale-Jacobse, A. E., Meijer, A., Helms-Lorenz, M. & Maulana, R. (2019). Differentiated instruction in secondary education: A systematic review of research evidence. *Frontiers in psychology*, *10*, 2366. https://doi.org/10.3389/fpsyg.2019.02366

Snijders, T. (2005). Power and sample size in multilevel modeling. In B. Everitt & D. C. Howell (Hrsg.), *Encyclopedia of Statistics in Behavorial Science* (S. 1570–1573). Wiley.

Snijders, T. & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackermann, R. J. Sternberg & R. Glaser (Hrsg.), *Learning and individual differences: Advances in theory and research* (S. 13–60). W.H. Freeman & Co Ltd.

Stahl, J. A. & Muckle, T. (2007). Investigating drift displacement in Rasch item calibrations. *Rasch Measurement Transactions*, 1126–1127.

Steenbergen-Hu, S., Makel, M. C. & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement. *Review of Educational Research*, *86*(4), 849–899.

Steenkamp, J.-B. E. M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–107. https://doi.org/10.1086/209528

Stuckey, M., Hofstein, A., Mamlok-Naaman, R. & Eilks, I. (2013). The meaning of 'relevance' in science education and its implications for the science curriculum. *Studies in Science Education*, *49*(1), 1–34. https://doi.org/10.1080/03057267.2013.802463

Subban, P. (2006). Differentiated instruction: A research basis. *International Education Journal*, *7*(7), 935–947.

Sumfleth, E. & Wild, E. (2005). *Schulische und familiale Bedingungen des Lernens und der Lernmotivation im Fach Chemie: Evaluation eines integrierten Interventionskonzeptes zur Säure-Base-Thematik.* Abschlussbericht des DFG-Projektes.

Sweller, J., van Merrienboer, J. J. G. & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296.

Taasoobshirazi, G. & Carr, M. (2008). A review and critique of context-based physics instruction and assessment. *Educational Research Review*, *3*(2), 155–167. https://doi.org/10.1016/j.edurev.2008.01.002

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Taconis, R., den Brok, P. & Pilot, A. (2016). Introduction: Context-based learning environments in science. In R. Taconis, P. den Brok & A. Pilot (Hrsg.), *Teachers creating context-based learning environments in science* (S. 1–20). SensePublishers.

Theyßen, H. (2014). Methodik von Vergleichsstudien zur Wirkung von Unterrichtsmedien. In D. Krüger, Parchman & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 67–79). Springer.

Tomlinson, C. A. (1999). *The Differentiated Classroom: Responding to the Needs of All Students*. Association for Supervision and Curriculum Development.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., Conover, L. A. & Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*(2-3), 119–145. https://doi.org/10.1177/016235320302700203

Trautmann, M. & Wischer, B. (2007). Individuell fördern im Unterricht. Was wissen wir über Innere Differenzierung? *Pädagogik*, *12*, 44–48.

Trautmann, M. & Wischer, B. (2009). Das Konzept der Inneren Differenzierung — eine vergleichende Analyse der Diskussion der 1970er Jahre mit dem aktuellen Heterogenitätsdiskurs. In M. A. Meyer, M. Prenzel & S. Hellekamps (Hrsg.), *Zeitschrift für Erziehungswissenschaft - Sonderheft: Vol. 9. Perspektiven der Didaktik* (S. 159–172). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91775-7_11

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G. & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology*, *104*(3), 763–777. https://doi.org/10.1037/a0027470

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

Ullmann-Margalit, E. & Morgenbesser, S. (1977). Picking and choosing. *Social Research*, *44*(4), 757–785.

Ültay, N. & Çalık, M. (2012). A thematic review of studies into the effectiveness of context-based chemistry curricula. *Journal of Science Education and Technology*, *21*(6), 686–701. https://doi.org/10.1007/s10956-011-9357-5

Urhahne, D. (2008). Sieben Arten der Lernmotivation. *Psychologische Rundschau*, *59*(3), 150–166. https://doi.org/10.1026/0012-1924.59.3.150

van Vorst, H. (2013). *Kontextmerkmale und ihr Einfluss auf das Schülerinteresse im Fach Chemie*. Logos Verlag Berlin.

van Vorst, H. & Aydogmus, H. (2021). One context fits all? – analysing students' context choice and their reasons for choosing a context-based task in chemistry education. *International Journal of Science Education*, 1–23. https://doi.org/10.1080/09500693.2021.1908640

van Vorst, H., Dorschu, A., Fechner, S., Kauertz, A., Krabbe, H. & Sumfleth, E. (2015). Charakterisierung und Strukturierung von Kontexten im naturwissenschaftlichen Unterricht – Vorschlag einer theoretischen Modellierung. *Zeitschrift für Didaktik der Naturwissenschaften*, *21*(1), 29–39. https://doi.org/10.1007/s40573-014-0021-5

van Vorst, H., Fechner, S. & Sumfleth, E. (2018). Unterscheidung von Kontexten für den Chemieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, *24*(1), 167–181. https://doi.org/10.1007/s40573-018-0081-z

Vock, M. & Gronostaj, A. (2017). *Umgang mit Heterogenität im Unterricht*. Friedrich-Ebert-Stiftung.

Waddington, D. (2005). Context-based learning in science education: a review. In P. Nentwig & D. Waddington (Hrsg.), *Making it relevant: Context based learning of science* (305-321). Waxmann.

Wagner, L. (2016). Adaptive und evidenzbasierte Förderung im Unterricht: Wozu braucht man das? *Potsdamer Zentrum für empirische Inklusionsforschung (ZEIF)*, *11*.

Walpuski, M., Kauertz, A., Kampa, N., Fischer, H. E., Mayer, J., Sumfleth, E. & Wellnitz, N. (2010). ESNaS: Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In A. Gehrmann & U. Hericks (Hrsg.), *Bildungsstandards und Kompetenzmodelle: Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 171–184). Verlag Julius Klinkhardt.

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: from alpha to omega. *The Clinical neuropsychologist*, *31*(6-7), 1113–1126. https://doi.org/10.1080/13854046.2017.1317364

Weiber, R. & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS. Springer-Lehrbuch*. Springer Gabler.

Weiss, L. & Müller, A. (2015). The notion of authenticity in the PISA units in physical science: An empirical analysis. *Zeitschrift für Didaktik der Naturwissenschaften*, *21*(1), 87–97. https://doi.org/10.1007/s40573-015-0025-9

Welch, B. L. (1939). Note on Discriminant Functions. *Biometrika*, *31*(1/2), 218. https://doi.org/10.2307/2334985

Werner, B. (2011). Didaktik und Förderung unter erschwerten Be. In A. Kaiser, D. Schmetz, P. Wachtel & B. Werner (Hrsg.), *Didaktik und Unterricht* (S. 43–63). Kohlhammer.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wierstra, R. F. & Wubbels, T. (1994). Student perception and appraisal of the learning environment: Core concepts in the evaluation of the plon physics curriculum. *Studies in Educational Evaluation*, *20*(4), 437–455. https://doi.org/10.1016/0191-491X(94)00036-G

Wigfield, A. & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81.

Wigfield, A. & Eccles, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. *Advances in Motivation Science*, *7*, 161–198. https://doi.org/10.1016/bs.adms.2019.05.002

Wild, K.-P. & Krapp, A. (1995). Elternhaus und intrinsische Lernmotivation. *Zeitschrift für Pädagogik*, *41*(4), 579–595.

Wilde, M., Basten, M., Großmann, N., Haunhorst, D., Desch, I., Strüber, M. & Randler, C. (2018). The (non-)benefit of choosing: If you get what you want it is not important that you chose it. *Motivation and Emotion*, *42*(3), 348–359. https://doi.org/10.1007/s11031-018-9675-5

Winthur, A. A. & Volk, T. L. (1994). Comparing achievement of inner-city high School students in traditional versus STS-based chemistry courses. *Journal of Chemical Education*, *71*(6), 501–505. https://doi.org/10.1021/ed071p501

Wirth, R. J. & Edwards, M. C. (2007). Item factor analysis: current approaches and future directions. *Psychological methods*, *12*(1), 58–79. https://doi.org/10.1037/1082-989X.12.1.58

Wischer, B. (2014). Individuelle Förderung als neue Leitidee? Kritische Anmerkungen zu einer aktuellen Reformstrategie. In V. Eisenbraun & S. Uhl (Hrsg.), *Geschlecht und Vielfalt in Schule und Lehrerbildung* (S. 163–175). Waxmann.

Wischer, B. & Trautmann, M. (2012). Innere Differenzierung als reformerischer Hoffnungsträger? Eine einführende Problemskizze zu Leerstellen und ungelösten Fragen. In T. Bohl, M. Bönsch, M. Trautmann & B. Wischer (Hrsg.), *Binnendifferenzierung Teil I: Didaktische Grundlagen und Forschungsergebnisse zur Binnendifferenzierung im Unterricht* (S. 24–39). Prolog Verlag.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*(7), 1341–1390. https://doi.org/10.1162/neco.1996.8.7.1341

Wright, B. D. (2003). Rack and stack: time 1 vs. time 2. *Rasch Measurement Transactions*, *17*(1), 905–906.

Yager, R. E. & Weld, J. D. (1999). Scope, sequence and coordination: The Iowa Project, a national reform effort in the USA. *International Journal of Science Education*, *21*(2), 169–194. https://doi.org/10.1080/095006999290778

13.2 Abbildungsverzeichnis

Abbildung 1: Das Selbstbestimmungs-Kontinuum mit der Art der Motivation, der Regulierung und dem perceived locus of causality (in Anlehnung an Ryan & Deci, 2000)..... 17

Abbildung 2: Das reduzierte Modell der SEVT nach Wigfield und Eccles (2020)..... 20

Abbildung 3: Die relationale Struktur des Interessenkonstrukts nach Krapp (1992a)... 23

Abbildung 4: Modell zur Operationalisierung der Kontextmerkmale (van Vorst et al., 2015) 34

Abbildung 5: Exemplarische Kontextbeschreibung *(*Teilstudie I*)* 77

Abbildung 6: Ablauf der Datenerhebung (Teilstudie I) 90

Abbildung 7: Wright-Map zum Vorwissenstest (Teilstudie I)..... 95

Abbildung 8: Vergleich der Merkmalszugehörigkeit Außerfachlich/Innerfachlich für jede Kontextaufgabe (I) und für die Kontextaufgaben je Kontextmerkmal (II). Aus Gründen der Übersichtlichkeit sind in Abbildung I nur nicht signifikante Unterschiede (ns) markiert..... 105

Abbildung 9: Vergleich der Merkmalszugehörigkeit Alltäglich/Besonders für jede Kontextaufgabe (I) und für die Kontextaufgaben je Kontextmerkmal (II)106

Abbildung 10: Vergleich der mittleren Chemienote (I) und der mittleren Personenfähigkeit (II) der Lernenden mit alltäglicher, besonderer und innerfachlicher Kontextwahl..... 108

Abbildung 11: Vergleich des mittleren Selbstkonzepts (I) und der mittleren Selbstwirksamkeitserwartung (II) der Lernenden mit alltäglicher, besonderer und innerfachlicher Kontextwahl..... 109

Abbildung 12: Vergleich des mittleren individuellen Interesses an Chemie (I), dem Interesse am Inhaltsfeld saure und alkalische Lösungen (II), dem Interesse am Themenfeld Körper und Krankheiten (III) und der extrinsischen Motivation (IV) der Lernenden mit alltäglicher, besonderer und innerfachlicher Kontextwahl..... 110

Abbildung 13: Vergleich der mittleren Ausprägung der Wahlmotive Persönlicher Bezug (I), Überraschungsmoment (II) sowie Interesse und Neugier (III) der Lernenden mit alltäglicher, besonderer und innerfachlicher Kontextwahl..... 112

Abbildung 14: Gesamtstreuungsquadratsumme gegen die Clusteranzahl (I) und die Höhe des PRE-Koeffizients (II) 113

Abbildung 15: Die z-standardisierten Mittelwerte der Personenmerkmale für jedes Cluster 115

Abbildung 16: Relative Häufigkeit der gewählten Kontexte in den Clustern und der Gesamtstichprobe..... 118

Abbildung 17: Vergleich der mittleren Zufriedenheit nach der Aufgabenbearbeitung zwischen den vier Clustern 125

Abbildung 18: Vergleich der mittleren emotionalen Valenz (I), der wertbezogenen Valenz (II) und der epistemischen Komponente (III) des situationalen Interesses zwischen den vier Clustern 125

Abbildung 19: Vergleich der mittleren wahrgenommenen Aufgabenschwierigkeit (I) und der investierten Denkanstrengung (II) zwischen den vier Clustern..... 126

Abbildung 20: Pfadmodell zur Vorhersage der Zufriedenheit, des situationalen Interesse und der kognitiven Belastung nach der Aufgabenbearbeitung 129

Abbildung 21: Die maximal margin hyperplane (schwarze durchgestrichene Linie), die margin (gestrichelte Linien) und die support vectors (schwarze Punkte) nach M. Kuhn und Johnson (2013) 143

Abbildung 22: Schematische Darstellung eines feed-forward neuronalen Netz mit einem hidden layer zur Klassifikation nach M. Kuhn und Johnson (2013) 144

Abbildung 23: Mittlerer F_1-score der trainierten Modelle unter Ausschluss und Berücksichtigung des themenspezifischen Interesses..... 146

Abbildung 24: Die relative mittlere prozentuale Abnahme im Gini-Index durch die Prädiktoren..... 147

Abbildung 25: Vergleich des mittleren F_1-scores der trainierten Modelle unter Anwendung verschiedener over-sampling Algorithmen..... 148

Abbildung 26: Vergleich des mittleren F_1-scores der Modelle nach Optimierung der Hyperparameter 150

Abbildung 27: Wright-Map zum Vorwissenstest basierend auf dem Rasch-Modell mit teilweise fixierten Itemparametern (Zusatzerhebung) 155

Abbildung 28: Ausschnitt aus der Aneignungsphase der dritten Teilaufgabe mit alltäglichem Kontext..... 164

Abbildung 29: Beispiel für eine Übungsaufgabe aus der dritten Teilaufgabe im Lernmaterial mit alltäglichem Kontext 165

Abbildung 30: Ablauf der Datenerhebung (Teilstudie II)..... 171

Abbildung 31: Wright-Map zum Vorwissenstest basierend auf dem Rasch-Modell mit teilweise fixierten Itemparametern (Teilstudie II) 176

Abbildung 32: Wright-Map zum Fachwissenstest basierend auf den Rasch-Modellen mit frei geschätzten Itemparametern zum Post-Zeitpunkt und teilweise fixierten Itemparametern zum Prä-Zeitpunkt 181

Abbildung 33: Relative Häufigkeit der bearbeiteten Kontexte in den Treatmentgruppe187

Abbildung 34: Verlauf der geschätzten mittleren Zufriedenheit (I) getrennt nach Treatment (II) 190

Abbildung 35: Verlauf der geschätzten mittleren Zufriedenheit über die drei Messzeitpunkte getrennt nach Treatment und bearbeiteten Kontext 191

Abbildung 36: Verlauf der geschätzten mittleren aufgabenbezogenen Zufriedenheit über die drei Messzeitpunkte getrennt nach individuellem Interesse 193

Abbildung 37: Geschätzte mittlere aufgabenbezogene Zufriedenheit, getrennt nach individuellem Interesse und bearbeitetem Kontext, aggregiert über die drei Messzeitpunkte und die Treatmentgruppen 194

Abbildung 38: Relative Häufigkeit der genannten Antwortalternativen für die Erfassung der Zufriedenheit getrennt nach Kontext 195

Abbildung 39: Vergleich der relativen Häufigkeit des bearbeiteten Kontext mit der relativen Häufigkeit des retrospektiv gewünschten Kontext für Lernende, die rückblickend einen anderen Kontext bearbeiten würden..... 196

Abbildung 40: Relative Häufigkeit der genannten Motive für die Antwortalternativen zur Erfassung der Zufriedenheit..... 197

Abbildung 41: Relative Häufigkeit der genannten Antwortalternativen für die Erfassung der Zufriedenheit getrennt nach Treatment..... 198

Abbildung 42: Vergleich der relativen Häufigkeit des bearbeiteten Kontext mit der relativen Häufigkeit des retrospektiv gewünschten Kontext für Lernende aus dem no choose & no match-Teatment, die rückblickend einen anderen Kontext bearbeite*n* würden..... 199

Abbildung 43: Relative Häufigkeit der genannten Motive der Treatmentgruppen..... 199

Abbildung 44: Verlauf der geschätzten affektiven Valenz (I), getrennt nach Treatment (II) 201

Abbildung 45: Verlauf der geschätzten mittleren affektiven Valenz über die drei Messzeitpunkte, getrennt nach individuellem Interesse 202

Abbildung 46: Geschätztes mittlere affektive Valenz getrennt nach individuellem Interesse und bearbeiteten Kontext, aggregiert über die drei Messzeitpunkte und die Treatmentgruppen 203

Abbildung 47: Verlauf der geschätzten mittleren epistemischen Valenz (I) getrennt nach Treatment (II) 204

Abbildung 48: Geschätzte mittlere epistemische Valenz, getrennt nach individuellem Interesse und der Treatmentgruppe 205

Abbildung 49: Geschätzte mittlere epistemische Valenz, getrennt nach individuellem Interesse und dem bearbeiteten Kontext 206

Abbildung 50: Verlauf der geschätzten mittleren investierten Denkanstrengung (I) getrennt nach Treatment (II) 208

Abbildung 51: Verlauf der geschätzten mittleren wahrgenommenen Aufgabenschwierigkeit (I) getrennt nach Treatment (II) 210

Abbildung 52: Geschätzte mittlere wahrgenommene Aufgabenschwierigkeit getrennt nach individuellem Interesse und der Treatmentgruppe 211

Abbildung 53: Verlauf der geschätzten mittleren wahrgenommenen Aufgabenschwierigkeit getrennt nach Treatment und individuellem Interesse 212

Abbildung 54: Einfluss des Messzeitpunkts (I) und des Treatments (II) auf die geschätzte mittlere Fachwissensausprägung 213

Abbildung 55: Einfluss der Zeit, des individuellen Interesse und des Treatments auf die geschätzte mittlere Fachwissensausprägung 214

Abbildung 56: Einfluss der Zeit und des Selbstkonzeptes auf die geschätzte mittlere Fachwissensausprägung 215

Abbildung 57: Einfluss der Zeit, des Selbstkonzepts und des Treatments auf die geschätzte mittlere Fachwissensausprägung 216

Abbildung 58: Einfluss der Zeit, des Treatments und des bearbeiteten Kontexts auf die geschätzte mittlere Fachwissensausprägung 217

Abbildung 59: Verlauf der aufgabenbezogenen Zufriedenheit (I), der affektiven Valenz des situationalen Interesses (II) und der epistemischen Komponente des situationalen Interesses (III) in Abhängigkeit der neu zusammengesetzten Gruppen 220

13.3 Tabellenverzeichnis

Tabelle 1: Gütekriterien und Cut-off-Werte zur Beurteilung des Modellfit (Gäde et al., 2020) 58

Tabelle 2: Interpretation von Cramér‘s V gemäß Cohen (1988) 66

Tabelle 3: Interpretation von η^2 gemäß Cohen (1988) 67

Tabelle 4: Interpretation von r gemäß Cohen (1988) 72

Tabelle 5: Titel der entwickelten Kontexte und deren Merkmalszugehörigkeit 75

Tabelle 6: Standardisierungskriterien bei der Entwicklung der Kontextbeschreibung .. 76

Tabelle 7: Konstruktionsdesign auf Basis der kognitiven Prozesse und Komplexität im ESNaS-Kompetenzstrukturmodell 79

Tabelle 8: Anzahl der Kernideen pro Basiskonzept und Lernjahr mit ausgewähltem Beispiel 81

Tabelle 9: Verteilung der Itemblöcke auf die Testhefte 82

Tabelle 10: Überblick über die eingesetzten Skalen zur Erfassung des individuellen und themenspezifisches Interesses an Chemie 84

Tabelle 11: Überblick über die eingesetzten Skalen zur Erfassung der Motivation im Chemieunterricht 85

Tabelle 12: Kodierung der angegeben Berufsaspirationen mit Kategorie, KldB und Beispiel 86

Tabelle 13: Überblick über die eingesetzten Skalen zur Erfassung des chemiebezogenen Selbstkonzepts 86

Tabelle 14: Überblick über die eingesetzten Skalen zur Erfassung der Wahlmotive 87

Tabelle 15: Überblick über die eingesetzten Skalen zur Erfassung des situationalen Interesse 88

Tabelle 16: Überblick über die ausgewählten Schlüsselwortpaare für die Merkmale Alltäglich/Besonders und Außerfachlich/Innerfachlich 90

Tabelle 17: Übersicht über die verwendeten R-Pakete zur Auswertung der ersten Teilstudie 93

Tabelle 18: Kennwerte des Rasch-Modells für den Vorwissenstest (Teilstudie I) 94

Tabelle 19: Ergebnis der Hauptachsenfaktorenanalyse für das individuelle Interesse an Chemie 97

Tabelle 20: Itemanzahl, Faktorladungen und Reliabilität der Skalen zum individuellen Interesse an Chemie und zur Motivation im Fach Chemie (Teilstudie I) ... 98

Tabelle 21: Ergebnis der Hauptachsenfaktorenanalyse für das Instrument zur Erfassung des Freizeitinteresse und -aktivitäten ... 99

Tabelle 22: Itemanzahl, Faktorladungen und Reliabilitäten der Skalen zum Selbstkonzept und zur Selbstwirksamkeitserwartung (Teilstudie I) ... 100

Tabelle 23: Ergebnis der Hauptachsenfaktorenanalyse für die Skalen zur Erfassung der Wahlmotive ... 101

Tabelle 24: Itemanzahl, Faktorladungen und Reliabilitäten der Skalen zur Erfassung der Wahlmotive (Teilstudie I) ... 102

Tabelle 25: Modellgüte des General-Faktoren-, Zwei-Faktoren-, Vier-Faktoren und des modifizierten Vier-Faktoren-Modells des Instruments zur Erfassung des situationalen Interesses und der Zufriedenheit (Teilstudie I) ... 103

Tabelle 26: Itemanzahl, Faktorladungen und Reliabilität der Skala zur Erfassung der Zufriedenheit und der Skalen zur Erfassung des situationalen Interesses (Teilstudie I) ... 104

Tabelle 27: Itemanzahl, Faktorladungen und Reliabilität der Skala zur Überprüfung der Merkmalszugehörigkeit ... 104

Tabelle 28: Die erklärte Varianz (η^2), der mittlere Homogenitätsindex ($Homo_k$), der mittlere Heterogenitätsindex ($Hetero_k$) und der mittlere Jaccard-Koeffizient (γ) als Maß für die Stabilität der 2-, 3- und 4-Clusterlösung (N_k). ... 114

Tabelle 29: Deskriptive Statistiken zum Geschlecht, der letzten Chemienote, der Selbstwirksamkeitserwartung, dem Interesse am Themenfeld saure und alkalische Lösungen und der erreichten Punktzahl im LGVT (Leseverständnis) der vier Cluster ... 116

Tabelle 30: Deskriptive Statistiken für die Zufriedenheit, die emotionale Valenz, die wertbezogene Valenz, die epistemische Komponente, die wahrgenommene Aufgabenschwierigkeit und die investierte Denkanstrengung, getrennt nach dem gewählten Kontext für das Cluster 1 ... 120

Tabelle 31: Deskriptive Statistiken für die Zufriedenheit, die emotionale Valenz, die wertbezogene Valenz, die epistemische Komponente, die wahrgenommene Aufgabenschwierigkeit und die investierte Denkanstrengung, getrennt nach dem gewählten Kontext für das Cluster 2 ... 121

Tabelle 32: Deskriptive Statistiken für die Zufriedenheit, die emotionale Valenz, die wertbezogene Valenz, die epistemische Komponente, die wahrgenommene

Aufgabenschwierigkeit und die investierte Denkanstrengung, getrennt nach dem gewählten Kontext für das Cluster 3 122

Tabelle 33: Deskriptive Statistiken für die Zufriedenheit, die emotionale Valenz, die wertbezogene Valenz, die epistemische Komponente, die wahrgenommene Aufgabenschwierigkeit und die investierte Denkanstrengung, getrennt nach dem gewählten Kontext für das Cluster 4 123

Tabelle 34: Beispiel für eine Konfusionsmatrix 138

Tabelle 35: Auflistung der optimierten Hyperparameter im Tuningprozess für jeden ML-Algorithmus 149

Tabelle 36: F1-score, Precision und Recall der entwickelten Modelle im Test-Datensatz151

Tabelle 37: Konfusionsmatrix für die Vorhersage der Kontextwahl im Test-Datensatz mithilfe des neuronalen Netzes 151

Tabelle 38: Kennwerte des Rasch-Modells mit frei geschätzten, vollständig fixierten und teilweise fixierten Itemparametern für den Vorwissenstest (Zusatzerhebung)155

Tabelle 39: Konfusionsmatrix zur Vorhersage der Kontextwahl in der Zusatzerhebung mithilfe des neuronalen Netzes 157

Tabelle 40: Die festgelegten Kompetenzerwartungen für die drei sequenzierten Teilaufgaben des Lernmaterials für Teilstudie II 162

Tabelle 41: Anzahl der entwickelten Testitems pro Teilaufgabe und Niveaustufe für den Fachwissenstest 167

Tabelle 42: Überblick über die eingesetzten Skalen zur Erfassung der Zufriedenheit 168

Tabelle 43: Ober- und Unterkategorien für die qualitative Inhaltsanalyse der offenen Items zur Zufriedenheit 169

Tabelle 44: Übersicht über die verwendeten R-Pakete zur Auswertung der zweiten Teilstudie 174

Tabelle 45: Kennwerte des Rasch-Modells mit frei geschätzten, vollständig fixierten und teilweise fixierten Itemparametern für den Vorwissenstest (Teilstudie II) 175

Tabelle 46: Itemanzahl, Faktorladungen und Reliabilität der Skalen zum individuellen Interesse an Chemie und zur Motivation im Fach Chemie (Teilstudie II) 177

Tabelle 47: Itemanzahl, Faktorladungen und Reliabilität der Skalen zum chemiebezogenen Selbstkonzept (Teilstudie II) 177

Tabelle 48: Itemanzahl, Faktorladungen und Reliabilitäten der Skalen zur Erfassung der Wahlmotive (Teilstudie II) 178

Tabelle 49: Kennwerte der Rasch-Modelle für den Fachwissenstest zum Prä- und Post-Zeitpunkt 179

Tabelle 50: Kennwerte der Rasch-Modelle mit vollständig fixierten und teilweise fixierten Itemparametern für den Fachwissenstest zum Prä-Zeitpunkt 180

Tabelle 51: Modellgüte des General- und Zwei-Faktoren-Modells des Instruments zur Erfassung der Zufriedenheit für den ersten, zweiten und dritten Messzeitpunkt 182

Tabelle 52: Kennwerte der Modelle zur Prüfung auf longitudinale Messinvarianz des Instruments zur Erfassung der Zufriedenheit 183

Tabelle 53: Modellgüte des Drei-Faktoren-Modells des Instruments zur Erfassung des situationalen Interesse für den ersten, zweiten und dritten Messzeitpunkt 184

Tabelle 54: Kennwerte der Zwei-Faktoren-Modelle zur Prüfung auf longitudinale Messinvarianz des Instruments zur Erfassung des situationalen Interesse 185

Tabelle 55: Demografische Daten und Deskriptive Statistiken der Kontrollvariablen für jedes Treatment 186

Tabelle 56: Konfusionsmatrix für die Vorhersage der Kontextwahl in der choose & match-Treatmentgruppe mithilfe des neuronalen Netz 188

Tabelle 57: Haupteffekte der Zeit, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die aufgabenbezogene Zufriedenheit 190

Tabelle 58: Haupteffekte des Messzeitpunktes, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die affektive Valenz 200

Tabelle 59: Haupteffekte der Zeit, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die epistemische Valenz 204

Tabelle 60: Haupteffekte des Messzeitpunktes, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die investierte Denkanstrengung 207

Tabelle 61: Haupteffekte der Zeit, des Treatments und des Kontextes sowie deren Interaktionseffekte auf die wahrgenommene Aufgabenschwierigkeit 209

Tabelle 62: Haupteffekte der Zeit, des Treatments und des Kontexts sowie deren Interaktionseffekte auf das Fachwissen im Inhaltsfeld saure und alkalische Lösungen 213

Tabelle 63: Beschreibung und Personenzahl (n) der angepassten Gruppen 217

Tabelle 64: Haupteffekte der Zeit, der Gruppe und des inhaltsfeldspezifischen Interesses sowie deren Interaktionseffekte auf die Zufriedenheit, die affektive Valenz des situationalen Interesses und die epistemische Komponente des situationalen Interesses 219

13.4 Abkürzungsverzeichnis

ANOVA:	Analysis of Variance
CFI:	Comparative Fit Index
ChiK:	Chemie im Kontext
CTL:	Cognitive Load Theory
CTML:	Cognitive Theory of Multimedia Learning
EPLOC:	External perceived locus of causality
ESNaS:	Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I
FDA:	Flexible Diskriminanzanalyse
FPC:	Fixed Parameter Calibration
IPLOC:	Internal perceived locus of causality
IPN:	Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik
IRT:	Item-Response-Theorie
KI:	Künstliche Intelligenz
KNN:	*K*-Nearest Neighbor
KTT:	Klassische Testtheorie
LDA:	Lineare Diskriminanzanalyse
LGVT:	Lesegeschwindigkeits- und -verständnistest
MANOVA:	Multivariate Analysis of Variance
MAR:	Missing at Random
MCAR:	Missing Completely at Random
ML:	Maschinelles Lernen
MLR:	Multinomiale Logistische Regression
MNAR:	Missing not at Random
PISA:	Programme for international student assessment
PLOC:	Perceived locus of causality
POI:	Person-object theory of interest

RDA:	Regularisierte Diskriminanzanalyse
RMSEA:	Root Mean Square Error of Approximation
SDT:	Self-determination theory
SEM:	Structural Equation Modeling
SEVT:	Situated expectancy-value theory
SRMR:	Standardized Root Mean Square Residual
SVM:	Support Vector Machine
TLI:	Tucker Lewis Index
WLE:	Weighted Likelihood Estimate
wMNSQ:	Weighted Mean-Square

14 Anhang

A Teilstudie I 278

A.1 Messinstrumente und Lernmaterial 278

A.2 Überprüfung der Voraussetzungen für parametrische Testverfahren 278

B Entwicklung des Vorhersagemodells 284

B.1 Standardeinstellungen der Hyperparameter 284

B.2 Optimierte Hyperparameter nach dem Tuningprozess 286

C Messinstrumente der Zusatzerhebung 287

D Teilstudie II 288

D.1 Messinstrumente und Lernmaterial 288

D.2 Manual zur Kodierung der offenen Antworten zur Zufriedenheit 289

D.3 Ergebnisse der ANOVAs zum Vergleich der Treatmentgruppen 292

D.4 Zusammenfassung der identifizierten Effekte der zweiten Teilstudie 293

E Lebenslauf 294

F Publikationen und Vorträge 295

A Teilstudie I

A.1 Messinstrumente und Lernmaterial

Die Messinstrumente und das Lernmaterial befinden sich im digitalen Anhang und sind unter folgendem Link einsehbar:

https://doi.org/10.5281/zenodo.8394461.

A.2 Überprüfung der Voraussetzungen für parametrische Testverfahren

Tabelle A1: Prüfung auf Homoskedastizität in der Merkmalszugehörigkeit Alltäglich/Besonders und Außerfachlich/Innerfachlich in Abhängigkeit vom gewählten Kontextmerkmal

Abhängige Variable	Levene-Test
Alltäglich - Besonders	$F(2, 344) = 7.89, p < .001$
Außerfachlich - Innerfachlich	$F(2, 344) = 2.41, p = .09$

Tabelle A2: Prüfung auf Normalverteilung und Homoskedastizität in der Merkmalszugehörigkeit Alltäglich/Besonders und Außerfachlich/Innerfachlich in Abhängigkeit vom gewählten Kontext

	Zähne putzen		Typ-A-Gastritis		Unsere Verdauung		Schädigung der Knochen		Säuren im Labor		Levene-Test
	z_{Schiefe}	z_{Kurtosis}	z_{Schiefe}	z_{Kurtosis}	z_{Schiefe}	z_{Kurtosis}	z_{Schiefe}	z_{Kurtosis}	z_{Schiefe}	z_{Kurtosis}	
All-Bes	3.46	1.55	-0.54	-0.39	0.40	-0.93	-0.79	0.17	-0.13	-0.47	$F(4, 342) = 9.78$, $p < .001$
Auß-Inn	-0.29	-0.38	-0.42	-0.57	-0.40	-0.83	0.23	-0.75	-0.65	-0.003	$F(4, 342) = 1.74$, $p = .14$

Tabelle A3: Prüfung auf Homoskedastizität in den Personenmerkmalen in Abhängigkeit vom gewählten Kontextmerkmal

Abhängige Variable	Levene-Test
Chemienote	$F(2, 344) = 0.79$, $p = .46$
Vorwissen in Chemie	$F(2, 344) = 0.55$, $p = .58$
Interesse an Chemie	$F(2, 344) = 0.21$, $p = .81$
Inhaltsfeldspezifisches Interesse	$F(2, 344) = 1.11$, $p = .33$
Themenspezifisches Interesse	$F(2, 344) = 1.82$, $p = .16$
Extrinsische Motivation	$F(2, 344) = 2.40$, $p = .09$
Selbstkonzept	$F(2, 344) = 0.26$, $p = .77$
Selbstwirksamkeitserwartung	$F(2, 344) = 0.53$, $p = .59$
Wahlmotiv: Überraschungsmoment	$F(2, 344) = 0.48$, $p = .63$
Wahlmotiv: Interesse & Neugier	$F(2, 344) = 2.08$, $p = .12$
Wahlmotiv: Persönlicher Bezug	$F(2, 344) = 0.05$, $p = .95$

Tabelle A4: Prüfung auf Homoskedastizität in den Personenmerkmalen in Abhängigkeit von der Clusterzugehörigkeit

Abhängige Variable	Levene-Test
Chemienote	$F(3, 333) = 2.73$, $p = .04$
Selbstwirksamkeitserwartung	$F(3, 333) = 1.76$, $p = .15$
Inhaltsfeldspezifisches Interesse	$F(3, 333) = 0.48$, $p = .70$
LGVT	$F(3, 333) = 0.13$, $p = .94$

Tabelle A5: Prüfung auf multivariate Normalverteilung der prozessbezogenen Variablen in Abhängigkeit des gewählten Kontext für jedes Cluster

Cluster	Multivariater Shapiro-Wilk-Test
Cluster 1	$W = 0.943$, $p = .002$
Cluster 2	$W = 0.901$, $p < .001$
Cluster 3	$W = 0.940$, $p < .001$
Cluster 4	$W = 0.972$, $p = .05$

Tabelle A6: Prüfung auf Normalverteilung und Homoskedastizität in den prozessbezogenen Variablen für das erste Cluster

Abhängige Variable	Alltäglich		Besonders		Innerfachlich		Levene-Test
	$z_{Schiefe}$	$z_{Kurtosis}$	$z_{Schiefe}$	$z_{Kurtosis}$	$z_{Schiefe}$	$z_{Kurtosis}$	
Zuf.	-0.17	-0.64	-0.21	-0.72	-0.76	-0.31	$F(2, 70) = 0.19$, $p = .82$
Emo. Val.	-0.32	-0.54	-0.23	-0.56	-0.69	-0.29	$F(2, 70) = 0.02$, $p = .98$
Wert. Val.	-0.56	-0.06	0.31	0.60	-0.73	-0.34	$F(2, 70) = 0.25$, $p = .78$
Epi. Kom.	0.29	-0.53	0.03	-0.68	-0.04	-0.43	$F(2, 70) = 0.02$, $p = .98$
Aufgabe	-0.32	-0.18	0.45	-0.27	0.67	-0.31	$F(2, 70) = 0.05$, $p = .95$
Denk	-0.16	-0.34	-0.57	-0.33	-0.35	-0.33	$F(2, 70) = 0.38$, $p = .69$

Anmerkung: Zuf. = Zufriedenheit, Emo. Val. = Emotionale Valenz des situationalen Interesses, Wert. Val. = Wertbezogene Valenz des situationalen Interesses, Epi. Kom. = Epistemische Komponente des situationalen Interesses, Aufgabe = Wahrgenommene Aufgabenschwierigkeit (Kognitive Belastung), Denk = Investierte Denkanstrengung (Kognitive Belastung

Tabelle A7: Prüfung auf Normalverteilung und Homoskedastizität in den prozessbezogenen Variablen für das zweite Cluster

Abhängige Variable	Alltäglich		Besonders		Innerfachlich		Levene-Test
	$z_{Schiefe}$	$z_{Kurtosis}$	$z_{Schiefe}$	$z_{Kurtosis}$	$z_{Schiefe}$	$z_{Kurtosis}$	
Zuf.	-0.14	-0.50	-0.22	-0.72	-0.68	-0.18	$F(2, 66) = 0.35$, $p = .70$
Emo. Val.	-0.09	-0.52	-0.19	-0.22	-0.38	-0.16	$F(2, 66) = 0.23$, $p = .80$
Wert. Val.	0.66	0.06	0.19	-0.03	-0.10	-0.52	$F(2, 66) = 1.43$, $p = .25$
Epi. Kom.	1.12	0.37	0.49	-0.65	0.79	-0.33	$F(2, 66) = 0.67$, $p = .51$
Aufgabe	-0.12	-0.63	-0.18	-0.70	-0.03	-0.54	$F(2, 66) = 0.08$, $p = .92$
Denk	-0.32	-0.58	-0.16	-0.64	0.00	-0.64	$F(2, 66) = 0.58$, $p = .56$

Anmerkung: Zuf. = Zufriedenheit, Emo. Val. = Emotionale Valenz des situationalen Interesses, Wert. Val. = Wertbezogene Valenz des situationalen Interesses, Epi. Kom. = Epistemische Komponente des situationalen Interesses, Aufgabe = Wahrgenommene Aufgabenschwierigkeit (Kognitive Belastung), Denk = Investierte Denkanstrengung (Kognitive Belastung

Tabelle A8: Prüfung auf Normalverteilung und Homoskedastizität in den prozessbezogenen Variablen für das dritte Cluster

Abhängige Variable	Alltäglich		Besonders		Innerfachlich		Levene-Test
	$z_{Schiefe}$	$z_{Kurtosis}$	$z_{Schiefe}$	$z_{Kurtosis}$	$z_{Schiefe}$	$z_{Kurtosis}$	
Zuf.	0.42	-0.20	0.19	-0.63	-0.32	-0.38	$F(2, 105) = 0.25$, $p = .78$
Emo. Val.	-0.32	-0.19	0.13	-0.57	-0.35	-0.58	$F(2, 105) = 0.60$, $p = .55$
Wert. Val.	0.68	-0.11	-0.14	-0.45	-0.03	-0.62	$F(2, 105) = 0.50$, $p = .61$
Epi. Kom.	1.59	0.55	0.05	-0.57	0.33	-0.38	$F(2, 105) = 1.57$, $p = .21$
Aufgabe	-0.20	-0.90	0.07	-0.49	0.00	-0.45	$F(2, 105) = 0.11$, $p = .90$
Denk	0.23	-0.72	-0.11	-0.42	0.13	-0.54	$F(2, 105) = 1.41$, $p = .25$

Anmerkung: Zuf. = Zufriedenheit, Emo. Val. = Emotionale Valenz des situationalen Interesses, Wert. Val. = Wertbezogene Valenz des situationalen Interesses, Epi. Kom. = Epistemische Komponente des situationalen Interesses, Aufgabe = Wahrgenommene Aufgabenschwierigkeit (Kognitive Belastung), Denk = Investierte Denkanstrengung (Kognitive Belastung

Tabelle A9: Prüfung auf Normalverteilung und Homoskedastizität in den prozessbezogenen Variablen für das vierte Cluster

Abhängige Variable	Alltäglich		Besonders		Innerfachlich		Levene-Test
	$z_{Schiefe}$	$z_{Kurtosis}$	$z_{Schiefe}$	$z_{Kurtosis}$	$z_{Schiefe}$	$z_{Kurtosis}$	
Zuf.	-0.13	-0.09	-0.45	-0.17	0.29	-0.34	$F(2, 84) = 0.97$, $p = .38$
Emo. Val.	-0.58	0.11	-0.35	-0.50	0.00	-0.59	$F(2, 84) = 0.47$, $p = .62$
Wert. Val.	-0.57	-0.26	0.46	-0.18	-0.20	-0.39	$F(2, 84) = 1.11$, $p = .33$
Epi. Kom.	-0.25	-0.45	-0.51	-0.30	-0.02	-0.58	$F(2, 84) = 1.05$, $p = .34$
Aufgabe	-0.23	-0.62	0.37	-0.42	-0.51	-0.34	$F(2, 84) = 2.14$, $p = .12$
Denk	-0.36	-0.07	0.19	-0.57	0.17	-0.56	$F(2, 84) = 0.22$, $p = .80$

Anmerkung: Zuf. = Zufriedenheit, Emo. Val. = Emotionale Valenz des situationalen Interesses, Wert. Val. = Wertbezogene Valenz des situationalen Interesses, Epi. Kom. = Epistemische Komponente des situationalen Interesses, Aufgabe = Wahrgenommene Aufgabenschwierigkeit (Kognitive Belastung), Denk = Investierte Denkanstrengung (Kognitive Belastung

Tabelle A10: Prüfung auf Homoskedastizität in den prozessbezogenen Variablen in Abhängigkeit von der Clusterzugehörigkeit

Abhängige Variable	Levene-Test
Zufriedenheit	$F(3, 333) = 2.56$, $p = .08$
Emotionale Valenz	$F(3, 333) = 5.01$, $p = .002$
Wertbezogene Valenz	$F(3, 333) = 0.53$, $p = .66$
Epistemische Komponente	$F(3, 333) = 2.76$, $p = .04$
Aufgabenschwierigkeit	$F(3, 333) = 0.94$, $p = .42$
Denkanstrengung	$F(3, 333) = 1.76$, $p = .15$

B Entwicklung des Vorhersagemodells

B.1 Standardeinstellungen der Hyperparameter

Tabelle A11: Standardeinstellungen der Hyperparameter für jeden ML-Algorithmus

ML-Algorithmus	Hyperparameter	Wert
Neuronales Netz	Anzahl der hidden units	= 5
	Anzahl der Epochen	= 100
	Weight decay	= 0.0
KNN	Anzahl der Nachbarn	= 5
	Minkowski-Distanzordnung	= 2.0
FDA	Anzahl der Modellterme	= [a]
	Grad der Interaktion	= 1
	Beschneidungsmethode	= backward
Bagged Tree	Baumtiefe	= 30
	Minimale Knotengröße	= 2
	Kosten-Komplexitätsparameter	= 0.01
Decision Tree	Baumtiefe	= 30
	Minimale Knotengröße	= 2
	Kosten-Komplexitätsparameter	= 0.01
Polynomiale SVM	Kosten	= 1.0
	Grad der Polynomfunktion	= 1
	Skalierungsfaktor	= 1.0
	Margin	= 0.1
Rbf SVM	Kosten	= 1.0
	Parameter der radialen Basisfunktion	= [b]
	Margin	= 0.1

ML-Algorithmus	Hyperparameter	Wert
Random Forest	Anzahl der zufällig ausgewählten Prädiktoren	= 2
	Anzahl der Bäume	= 500
	Minimale Knotengröße	= 10
RDA	Anteil der gemeinsamen Kovarianzmatrix	= [c]
	Anteil der Einheitsmatrix	= [d]
XGBoost	Baumtiefe	= 6
	Anzahl der Bäume	= 1000
	Lernrate	= 0.3
	Anteil der zufällig ausgewählten Prädiktoren	= 2
	Minimale Knotengröße	= 1
	Minimale Loss Reduction	= 0.0
	Anteil der Beobachtungen in der Stichprobe	= 1.0
	Anzahl der Iterationen	= Inf

Anmerkung: Rbf = Radiale Basisfunktion, SVM = Support vector machine, RDA = Regularisierte Diskriminanzanalyse, KNN = *K*-Nearest Neighbors. [a] Die Anzahl der Modellterme wird durch die Anzahl der Prädiktoren berechnet. [b] Der Wert für die radiale Basisfunktion wird aufgrund der Daten bestimmt (siehe Karatzoglou et al., 2023). [c] Der Anteil der gemeinsamen Kovarianzmatrix wird numerisch durch einen Nelder-Mead-(Simplex-)Algorithmus bestimmt (siehe Roever et al., 2023). [d] Der Anteil der Einheitsmatrix wird numerisch durch einen Nelder-Mead-(Simplex-)Algorithmus bestimmt (siehe Roever et al., 2023).

B.2 Optimierte Hyperparameter nach dem Tuningprozess

Tabelle A12: Optimierte Hyperparameter für jeden ML-Algorithmus

ML-Algorithmus	Hyperparameter	Wert
Neuronales Netz	Anzahl der hidden units	= 2
	Anzahl der Epochen	= 42
	Weight decay	= 0.978
Polynomiale SVM	Kosten	= 0.00403
	Grad der Polynomfunktion	= 1
	Skalierungsfaktor	= 1.0
	Margin	= 0.1
Rbf SVM	Kosten	= 0.00104
	Parameter der radialen Basisfunktion	= 0.0579
	Margin	= 0.1
Random Forest	Anzahl der zufällig ausgewählten Prädiktoren	= 6
	Anzahl der Bäume	= 18
	Minimale Knotengröße	= 39
RDA	Anteil der gemeinsamen Kovarianzmatrix	= 0.923
	Anteil der Einheitsmatrix	= 0.0106
XGBoost	Baumtiefe	= 5
	Anzahl der Bäume	= 1000
	Lernrate	= 0.0119
	Anteil der zufällig ausgewählten Prädiktoren	= 3
	Minimale Knotengröße	= 30
	Minimale Loss Reduction	= 0.0
	Anteil der Beobachtungen in der Stichprobe	= 0.824
	Anzahl der Iterationen	= Inf

Anmerkung: Rbf = Radiale Basisfunktion, SVM = Support vector machine, RDA = Regularisierte Diskriminanzanalyse

C Messinstrumente der Zusatzerhebung

Die Dokumentation des adaptierten Messinstruments zur Erfassung der Wahlmotive befinden sich im digitalen Anhang und sind unter folgendem Link einsehbar:

https://doi.org/10.5281/zenodo.8394461

D Teilstudie II

D.1 Messinstrumente und Lernmaterial

Im digitalen Anhang befindet sich:

- Das Lernmaterial
- Das Übersichtsblatt zum Studienablauf
- Die Dokumentation der Messinstrumente

Der digitale Anhang ist unter folgendem Link einsehbar:

https://doi.org/10.5281/zenodo.8394461

D.2 Manual zur Kodierung der offenen Antworten zur Zufriedenheit

1. Allgemeine Informationen

Es werden stets ganze Antworten kodiert. Die Kodierung erfolgt in zwei Durchgängen. Zunächst werden die Oberkategorien Kognitive Gründe, Affektive Gründe und Unspezifische Gründe kodiert. Anschließend werden die Oberkategorien in einem zweiten Durchgang in die Unterkategorien ausdifferenziert. Alle Antworten, welche die eigentlich Frage nicht betreffen (z. B. „Idk“ oder „Kp“) werden als nicht kodierbar (Code: 0) eingestuft.

2. Kategoriensystem

2.1 Kognitive Motive

In dieser Kategorie (Code 1) werden Antworten gesammelt, die sich auf das Verständnis der Aufgabe und den Erwerb neuen Wissens beziehen.

Beispiele:

- „Weil ich davon nichts verstehe“
- „Weil ich die Hoffnung hatte dass ich mit diesen Aufgaben weniger überfordert wäre.“

2.1.1 Anforderungsniveau

Antworten, die sich auf das Anforderungsniveau (die Schwierigkeit) der Aufgaben beziehen (Code 1A).

Beispiele:

- „Weil ich sie schwer verstehe“
- „Ich fand dieses Thema relativ schwer.“

2.1.2 Neues Lernen

Antworten, die sich auf den Erwerb von neuem Wissen beziehen (Code 2A).

Beispiel:

- „Ich habe viel neues gelernt“

2.2 Affektiv-motivationale Motive

Diese Kategorie (Code 2) umfasst Antworten, welche affektiv-motivationale Merkmale bei der Aufgabenbearbeitung betrifft. Darunter gefasst wird Interesse an der Aufgabe generell, ein persönlicher Bezug oder Interesse am bearbeiteten Kontext.

Beispiele:

- „Ich denke das es interessanter ist“
- „Weil es mehr im Alltag vorkommt“
- „Langweilig“

2.2.1 Interesse unspezifisch

Antworten, die keine genaue Ursache für das Interesse umfassen, werden in dieser Kategorie gesammelt (Code 2A).

Beispiele:

- „Weil mich keines besonders anspricht oder sich interessiert anhört.“
- „Weil ich das Thema interessanter finde.“
- „Hätte es spannender gefunden.“

2.2.2 Persönlicher Bezug

Alle Antworten, die einen persönlichen Bezug zur Aufgabe beinhalten (Code 2B). Dabei kann es sich sowohl um einen Alltagsbezug als auch um einen Bezug zur Familie oder zum zukünftigen Beruf handeln.

Beispiele:

- „Interessanter und wichtiger für den Alltag“
- „Weil ich mich mehr für medizinische Chemie interessieren würde, da mein Beruf damit eher zu tuen haben wird.“
- „Weil mich das Problem in der Familie betrifft und ich gerne die Ursachen wissen möchte weshalb dies so ist“

2.2.3 Themenfeldspezifisches Interesse

In dieser Kategorie werden Antworten zusammengefasst, die das Interesse am Themenfeld der Aufgaben („Körper und Krankheiten“) ansprechen (Code 2C).

Beispiele:

- „Ich interessiere mich für den menschlichen Körper“
- „Weil es das thema unsere Gesundheit anspricht und diese mit Chemie kombiniert“

2.2.4 Inhaltsfeldspezifisches Interesse

Antworten, die das Interesse am Inhaltsfeld („saure und alkalische Lösungen“) der Aufgaben umfassen (Code 2D).

Beispiel:

- „Es war sehr interessant mehr über Säuren zu lernen.“

2.2.5 Fachspezifisches Interesse

Unter dieser Kategorie werden alle Antworten gefasst, die das allgemeine Fachinteresse betreffen (Code 2E).

Beispiele:

- „Weil mich Chemie leider überhaupt nicht interessiert und ich es Langweilig finde.“
- „Weil ich kein Chemie mag“

2.2.6 Kontextspezifisches Interesse

In dieser Kategorie werden Antworten zusammenfasst, die spezifische Informationen am Kontext thematisieren und als Ursache für das Interesse angeben.

Beispiel:

- „Ich fand es interessant was passiert um z.B die Knochen zu schützen oder was dem Körper gegeben wird, um bestimmte Sachen zu produzieren.“

3. Unspezifische Motive

Alle Antworten, die keinen spezifischen Motive angeben oder sowohl affektive als auch kognitive Ursachen adressieren, sodass insgesamt unklar bleibt, was das genaue Motiv für die Zufriedenheit ist.

Beispiele:

- „Alles“
- „Weil keins davon mich interessiert und ich da eh nichts verstehe“

D.3 Ergebnisse der ANOVAs zum Vergleich der Treatmentgruppen

Tabelle A13: Ergebnisse der ANOVAs für jede Kontrollvariable in Abhängigkeit des Treatments

Variable	df	*F*	*p*	η^2
Leseverständnis	2, 216	1.29	.28	.01
Vorwissen	2, 216	0.84	.43	.00
$\text{Fachwissen}_{\text{prä}}$	2, 216	0.65	.52	.00
Individuelles Interesse an Chemie	2, 216	2.28	.10	.02
Inhaltsspezifisches Interesse	2, 216	2.21	.11	.02
Themenspezifisches Interesse	2, 216	1.40	.25	.01
Extrinsische Motivation	2, 216	0.43	.65	.00
Selbstkonzept	2, 216	2.20	.11	.02
Selbstwirksamkeitserwartung	2, 216	1.75	.18	.02

D.4 Zusammenfassung der identifizierten Effekte der zweiten Teilstudie

Tabelle A14: Zusammenfassung der identifizierten Effekte der zweiten Teilstudie

	Abhängige Variablen					
Effekt	Zuf	AV	EK	ID	WA	F
Z	✓	✓	✓	✓	✓	✓
T	✓	✓	✓ [a]	✓	---	---
K	---	---	---	---	---	---
Z × T	---	---	---	---	---	---
Z × K	---	---	---	---	---	---
T × K	---	---	---	---	---	---
Z × T × K	✓ [b]	---	---	---	---	✓ [c]
(Interaktions-)Effekte des individuellen Interesses an Chemie						
I	✓	✓	✓	---	---	---
Z × I	✓	✓	---	---	---	---
K × I	✓	✓	✓	---	---	---
T × I	---	---	(✓)	---	✓	---
Z × T × I	---	---	---	---	✓	✓
(Interaktions-)Effekte des chemiebezogenen Selbstkonzeptes						
S	---	---	✓	✓	✓	✓
Z × S	---	---	---	---	---	✓
Z × T × S	---	---	---	---	---	✓

Anmerkung: Zuf = Zufriedenheit, AV = Affektive Valenz des situationalen Interesses, EK = Epistemische Komponente des situationalen Interesses, ID = Investierte Denkanstrengung (Kognitive Belastung), WA = Wahrgenommene Aufgabenschwierigkeit (Kognitive Belastung), F = Fachwissen. Z = Zeit, T = Treatment, K = Kontext, I = Individuelles Interesse, S = Selbstkonzept.

✓ = Statistisch signifikanter Effekt, (✓) = Nahezu statistisch signifikanter Effekt ($p = .05$), --- = Kein statistisch signifikanter Effekt.

[a] = Der Effekt ist unter Berücksichtigung des individuellen Interesses und des Selbstkonzepts statistisch nicht mehr signifikant. [b] = Der Effekt ist unter Berücksichtigung des individuellen Interesses statistisch nicht mehr signifikant. [c] = Der Effekt ist nur unter Berücksichtigung des Selbstkonzepts statistisch signifikant.

E Lebenslauf

Persönliche Daten

Name:	Fabien Louis Güth
Geburtsdatum:	[redacted]
Geburtsort:	[redacted]
Staatsangehörigkeit:	[redacted]
[redacted]	[redacted]

Werdegang

seit 05/2023	Referendariat für das Lehramt an Gymnasien und Gesamtschulen am Heinrich-Heine-Gymnasium Bottrop, Zentrum für schulpraktische Lehrerbildung Gelsenkirchen
09/2019 – 04/2023	Wissenschaftlicher Mitarbeiter (Doktorand) in der Didaktik der Chemie in der Arbeitsgruppe von Frau Prof. Dr. Elke Sumfleth an der Universität Duisburg-Essen
01/2018 – 08/2019	Wissenschaftliche Hilfskraft in der Didaktik der Chemie an der Universität Duisburg-Essen
10/2017 – 07/2019	Master-Studium der Fächer Chemie und Sport für das Lehramt an Gymnasien und Gesamtschulen, Universität Duisburg-Essen Abschluss: Master of Education (M. Ed.)
09/2017 – 12/2017	Studentische Hilfskraft in der Didaktik der Chemie an der Universität Duisburg-Essen
10/2014 – 09/2017	Bachelor-Studium der Fächer Chemie und Sport für das Lehramt an Gymnasien und Gesamtschulen, Universität Duisburg-Essen Abschluss: Bachelor of Arts (B. A.)
10/2013 – 09/2014	Bachelor-Studium der Fächer Biologie und Sport für das Lehramt an Gymnasien und Gesamtschulen, Universität Duisburg-Essen
06/2013	Abitur an der Evangelischen Gesamtschule Gelsenkirchen-Bismarck

F Publikationen und Vorträge

Beiträge in Sammelbänden (mit Peer-Review)

- **Güth, F.**, & van Vorst, H. (2023). Context-based learning as a method for differentiated instruction in chemistry education. In G. S. Carvalho, Z. Anastácio, & A. S. Afonso (Hrsg.), *Fostering Scientific Citizenship in an Uncertain World - Selected Papers from the ESERA 2021 Conference* (S. 153-169). Springer Verlag. https://doi.org/10.1007/978-3-031-32225-9_10

Tagungsbandbeiträge

- **Güth, F.** & van Vorst, H. (2023). Einsatz variierter Kontexte zur interessenbasierten Differenzierung im Fach Chemie. In H. van Vorst (Hrsg.), Lernen, Lehren und Forschen in einer digital geprägten Welt. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Aachen 2022 (S. 254-257). Universität Duisburg-Essen. Verfügbar unter: https://gdcp-ev.de/wp-content/uploads/securepdfs/2023/05/D17_Gueth.pdf
- **Güth, F.** & van Vorst, H. (2022). Context-based learning as a method for differentiated instruction in chemistry education. In G.S. Carvalho, A.S. Afonso & Z. Anastácio (Eds.), Fostering scientific citizenship in an uncertain world (Proceedings of ESERA 2021), Part 2 (co-ed. F. Le Hebel & V. Kind), pp. 129-136. Braga: CIEC, University of Minho. ISBN 978-972-8952-82-2.
- **Güth, F.** & van Vorst, H. (2022). Untersuchung der Kontextwahl unterschiedlicher Schülergruppen im Chemieunterricht. In S. Habig, & H. van Vorst (Hrsg.), *Unsicherheit als Element von naturwissenschaftsbezogenen Bildungsprozessen*. Gesellschaft für Didaktik der Chemie und Physik. Virtuelle Jahrestagung (S. 392-395). Universität Duisburg-Essen. Verfügbar unter https://www.gdcp-ev.de/wp-content/tb2022/TB2022_392_Gueth.pdf
- Behrendt, A., **Güth, F.**, Kaulhausen, S., Steinbach, M., Vogelsang, C., & Weiler, D. C. (2021). Flexible Gestaltung von Datenerhebungen in Schulen und Universitäten. In S. Habig (Hrsg.), *Naturwissenschaftlicher Unterricht und Lehrerbildung im Umbruch?.* Gesellschaft für Didaktik der Chemie und Physik. Virtuelle Jahrestagung (S. 48-50). Universität Duisburg-Essen. Verfügbar unter https://www.gdcp-ev.de/wp-content/tb2021/TB2021_48_Behrendt.pdf
- **Güth, F.** & van Vorst, H. (2021). Interessengestützte Binnendifferenzierung im Chemieunterricht. In S. Habig (Hrsg.), *Naturwissenschaftlicher Unterricht und Lehrerbildung im Umbruch?.* Gesellschaft für Didaktik der Chemie und Physik. Virtuelle Jahrestagung (S. 597-600). Universität Duisburg-Essen. Verfügbar unter https://www.gdcp-ev.de/wp-content/tb2021/TB2021_597_Gueth.pdf

Posterpräsentationen und Vorträge

- **Güth, F.** & van Vorst, H. (2023). *Interest-based differentiated instruction through varied contextual tasks in chemistry education*. Vortrag zur NARST 96th annual international conference. Chicago, USA.
- **Güth, F.** & van Vorst, H. (2023). *To choose or not to choose? Interessengestützte Differenzierung im Fach Chemie.* Vortrag zur Jahrestagung der Gesellschaft für empirische Bildungsforschung (GEBF). Essen, Deutschland.
- **Güth, F.** & van Vorst, H. (2023). *To choose or not to choose? Zum Einfluss von Wahlfreiheit beim Chemielernen mit kontextorientierten Aufgaben*. Vortrag zum Neujahrskolloquium der Fakultät für Chemie (Universität Duisburg-Essen). Essen, Deutschland.
- **Güth, F.** & van Vorst, H. (2022). *Systematisch variierte Kontextaufgaben zur interessenbasierten Differenzierung im Fach Chemie*. Vortrag im Rahmen des Graduiertenkolloquiums „Naturwissenschaftlicher Unterricht" des Interdisziplinären Zentrums für Bildungsforschung (IZfB), Universität-Duisburg-Essen. Essen, Deutschland.
- **Güth, F.** & van Vorst, H. (2022). *Einsatz variierter Kontexte zur interessenbasierten Differenzierung im Fach Chemie*. Vortrag zur Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik (GDCP). Aachen, Deutschland.
- **Güth, F.** & van Vorst, H. (2022). *Investigating students' context choice in chemistry education*. Vortrag zur NARST 95th annual international conference. Vancouver, Kanada.
- **Güth, F.** & van Vorst, H. (2022). *Interessengestützte Differenzierung durch systematisch variierte Kontextaufgaben im Chemieunterricht.* Vortrag zur Jahrestagung der Gesellschaft für empirische Bildungsforschung (GEBF). Virtuelle Jahrestagung.
- **Güth, F.** & van Vorst, H. (2021). *Untersuchung der Kontextwahl unterschiedlicher Schülergruppen im Chemieunterricht*. Vortrag zur Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik (GDCP). Virtuelle Jahrestagung.
- **Güth, F.** & van Vorst, H. (2021). *Context-based learning as a method for differentiated instruction in chemistry education***.** Vortrag zur 14th Conference der European Science Education Research Association (ESERA). Virtuelle Tagung.
- **Güth, F.** & van Vorst, H. (2020). *Kontextualisierungen als Methode der Binnendifferenzierung im Chemieunterricht*. Vortrag zum Doktorierendenkolloquium der Gesellschaft für Didaktik der Chemie und Physik (GDCP). Virtuelles Doktorierendenkolloquium.
- **Güth, F.** & van Vorst, H. (2020). *Interessengestützte Binnendifferenzierung im Chemieunterricht*. Poster zur Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik (GDCP). Virtuelle Jahrestagung.

- **Güth, F.**, van Vorst, H. & Sumfleth, E. (2020). *Untersuchung des Einsatzes unterschiedlicher Kontextualisierungen als Methode der Binnendifferenzierung im Chemieunterricht*. Vortrag im Rahmen des Graduiertenkolloquiums „Naturwissenschaftlicher Unterricht“ des Interdisziplinären Zentrums für Bildungsforschung (IZfB), Universität-Duisburg-Essen.

Bisher erschienene Bände der Reihe *„Studien zum Physik- und Chemielernen“*

ISSN 1614-8967 (vormals *Studien zum Physiklernen* ISSN 1435-5280)

1 Helmut Fischler, Jochen Peuckert (Hrsg.): Concept Mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie
ISBN 978-3-89722-256-4 40.50 EUR

2 Anja Schoster: Bedeutungsentwicklungsprozesse beim Lösen algorithmischer Physikaufgaben. *Eine Fallstudie zu Lernprozessen von Schülern im Physiknachhilfeunterricht während der Bearbeitung algorithmischer Physikaufgaben*
ISBN 978-3-89722-045-4 40.50 EUR

3 Claudia von Aufschnaiter: Bedeutungsentwicklungen, Interaktionen und situatives Erleben beim Bearbeiten physikalischer Aufgaben
ISBN 978-3-89722-143-7 40.50 EUR

4 Susanne Haeberlen: Lernprozesse im Unterricht mit Wasserstromkreisen. *Eine Fallstudie in der Sekundarstufe I*
ISBN 978-3-89722-172-7 40.50 EUR

5 Kerstin Haller: Über den Zusammenhang von Handlungen und Zielen. *Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*
ISBN 978-3-89722-242-7 40.50 EUR

6 Michaela Horstendahl: Motivationale Orientierungen im Physikunterricht
ISBN 978-3-89722-227-4 50.00 EUR

7 Stefan Deylitz: Lernergebnisse in der Quanten-Atomphysik. *Evaluation des Bremer Unterrichtskonzepts*
ISBN 978-3-89722-291-5 40.50 EUR

8 Lorenz Hucke: Handlungsregulation und Wissenserwerb in traditionellen und computergestützten Experimenten des physikalischen Praktikums
ISBN 978-3-89722-316-5 50.00 EUR

9 Heike Theyßen: Ein Physikpraktikum für Studierende der Medizin. *Darstellung der Entwicklung und Evaluation eines adressatenspezifischen Praktikums nach dem Modell der Didaktischen Rekonstruktion*
ISBN 978-3-89722-334-9 40.50 EUR

10 Annette Schick: Der Einfluß von Interesse und anderen selbstbezogenen Kognitionen auf Handlungen im Physikunterricht. *Fallstudien zu Interessenhandlungen im Physikunterricht*
ISBN 978-3-89722-380-6 40.50 EUR

11 Roland Berger: Moderne bildgebende Verfahren der medizinischen Diagnostik. *Ein Weg zu interessanterem Physikunterricht*
ISBN 978-3-89722-445-2 40.50 EUR

12 Johannes Werner: Vom Licht zum Atom. *Ein Unterrichtskonzept zur Quantenphysik unter Nutzung des Zeigermodells*
ISBN 978-3-89722-471-1 40.50 EUR

13 Florian Sander: Verbindung von Theorie und Experiment im physikalischen Praktikum. *Eine empirische Untersuchung zum handlungsbezogenen Vorverständnis und dem Einsatz grafikorientierter Modellbildung im Praktikum*
ISBN 978-3-89722-482-7 40.50 EUR

14 Jörn Gerdes: Der Begriff der physikalischen Kompetenz. *Zur Validierung eines Konstruktes*
ISBN 978-3-89722-510-7 40.50 EUR

15 Malte Meyer-Arndt: Interaktionen im Physikpraktikum zwischen Studierenden und Betreuern. *Feldstudie zu Bedeutungsentwicklungsprozessen im physikalischen Praktikum*
ISBN 978-3-89722-541-1 40.50 EUR

16 Dietmar Höttecke: Die Natur der Naturwissenschaften historisch verstehen. *Fachdidaktische und wissenschaftshistorische Untersuchungen*
ISBN 978-3-89722-607-4 40.50 EUR

17 Gil Gabriel Mavanga: Entwicklung und Evaluation eines experimentell- und phänomenorientierten Optikcurriculums. *Untersuchung zu Schülervorstellungen in der Sekundarstufe I in Mosambik und Deutschland*
ISBN 978-3-89722-721-7 40.50 EUR

18 Meike Ute Zastrow: Interaktive Experimentieranleitungen. *Entwicklung und Evaluation eines Konzeptes zur Vorbereitung auf das Experimentieren mit Messgeräten im Physikalischen Praktikum*
ISBN 978-3-89722-802-3 40.50 EUR

19 Gunnar Friege: Wissen und Problemlösen. *Eine empirische Untersuchung des wissenszentrierten Problemlösens im Gebiet der Elektrizitätslehre auf der Grundlage des Experten-Novizen-Vergleichs*
ISBN 978-3-89722-809-2 40.50 EUR

20 Erich Starauschek: Physikunterricht nach dem Karlsruher Physikkurs. *Ergebnisse einer Evaluationsstudie*
ISBN 978-3-89722-823-8 40.50 EUR

21 Roland Paatz: Charakteristika analogiebasierten Denkens. *Vergleich von Lernprozessen in Basis- und Zielbereich*
ISBN 978-3-89722-944-0 40.50 EUR

22 Silke Mikelskis-Seifert: Die Entwicklung von Metakonzepten zur Teilchenvorstellung bei Schülern. *Untersuchung eines Unterrichts über Modelle mithilfe eines Systems multipler Repräsentationsebenen*
ISBN 978-3-8325-0013-9 40.50 EUR

23 Brunhild Landwehr: Distanzen von Lehrkräften und Studierenden des Sachunterrichts zur Physik. *Eine qualitativ-empirische Studie zu den Ursachen*
ISBN 978-3-8325-0044-3 40.50 EUR

24 Lydia Murmann: Physiklernen zu Licht, Schatten und Sehen. *Eine phänomenografische Untersuchung in der Primarstufe*
ISBN 978-3-8325-0060-3 40.50 EUR

25 Thorsten Bell: Strukturprinzipien der Selbstregulation. *Komplexe Systeme, Elementarisierungen und Lernprozessstudien für den Unterricht der Sekundarstufe II*
ISBN 978-3-8325-0134-1 40.50 EUR

26 Rainer Müller: Quantenphysik in der Schule
ISBN 978-3-8325-0186-0 40.50 EUR

27 Jutta Roth: Bedeutungsentwicklungsprozesse von Physikerinnen und Physikern in den Dimensionen Komplexität, Zeit und Inhalt
ISBN 978-3-8325-0183-9 40.50 EUR

28 Andreas Saniter: Spezifika der Verhaltensmuster fortgeschrittener Studierender der Physik
ISBN 978-3-8325-0292-8 40.50 EUR

29 Thomas Weber: Kumulatives Lernen im Physikunterricht. *Eine vergleichende Untersuchung in Unterrichtsgängen zur geometrischen Optik*
ISBN 978-3-8325-0316-1 40.50 EUR

30 Markus Rehm: Über die Chancen und Grenzen moralischer Erziehung im naturwissenschaftlichen Unterricht
ISBN 978-3-8325-0368-0 40.50 EUR

31 Marion Budde: Lernwirkungen in der Quanten-Atom-Physik. *Fallstudien über Resonanzen zwischen Lernangeboten und SchülerInnen-Vorstellungen*
ISBN 978-3-8325-0483-0 40.50 EUR

32 Thomas Reyer: Oberflächenmerkmale und Tiefenstrukturen im Unterricht. *Exemplarische Analysen im Physikunterricht der gymnasialen Sekundarstufe*
ISBN 978-3-8325-0488-5 40.50 EUR

33 Christoph Thomas Müller: Subjektive Theorien und handlungsleitende Kognitionen von Lehrern als Determinanten schulischer Lehr-Lern-Prozesse im Physikunterricht
ISBN 978-3-8325-0543-1 40.50 EUR

34 Gabriela Jonas-Ahrend: Physiklehrervorstellungen zum Experiment im Physikunterricht
ISBN 978-3-8325-0576-9 40.50 EUR

35 Dimitrios Stavrou: Das Zusammenspiel von Zufall und Gesetzmäßigkeiten in der nichtlinearen Dynamik. *Didaktische Analyse und Lernprozesse*
ISBN 978-3-8325-0609-4 40.50 EUR

36 Katrin Engeln: Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken
ISBN 978-3-8325-0689-6 40.50 EUR

37 Susann Hartmann: Erklärungsvielfalt
ISBN 978-3-8325-0730-5 40.50 EUR

38 Knut Neumann: Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker
ISBN 978-3-8325-0762-6 40.50 EUR

39 Michael Späth: Kontextbedingungen für Physikunterricht an der Hauptschule. *Möglichkeiten und Ansatzpunkte für einen fachübergreifenden, handlungsorientierten und berufsorientierten Unterricht*
ISBN 978-3-8325-0827-2 40.50 EUR

40 Jörg Hirsch: Interesse, Handlungen und situatives Erleben von Schülerinnen und Schülern beim Bearbeiten physikalischer Aufgaben
ISBN 978-3-8325-0875-3 40.50 EUR

41 Monika Hüther: Evaluation einer hypermedialen Lernumgebung zum Thema Gasgesetze. *Eine Studie im Rahmen des Physikpraktikums für Studierende der Medizin*
ISBN 978-3-8325-0911-8 40.50 EUR

42 Maike Tesch: Das Experiment im Physikunterricht. *Didaktische Konzepte und Ergebnisse einer Videostudie*
ISBN 978-3-8325-0975-0 40.50 EUR

43 Nina Nicolai: Skriptgeleitete Eltern-Kind-Interaktion bei Chemiehausaufgaben. *Eine Evaluationsstudie im Themenbereich Säure-Base*
ISBN 978-3-8325-1013-8 40.50 EUR

44 Antje Leisner: Entwicklung von Modellkompetenz im Physikunterricht
ISBN 978-3-8325-1020-6 40.50 EUR

45 Stefan Rumann: Evaluation einer Interventionsstudie zur Säure-Base-Thematik
ISBN 978-3-8325-1027-5 40.50 EUR

46 Thomas Wilhelm: Konzeption und Evaluation eines Kinematik/Dynamik-Lehrgangs zur Veränderung von Schülervorstellungen mit Hilfe dynamisch ikonischer Repräsentationen und graphischer Modellbildung – mit CD-ROM
ISBN 978-3-8325-1046-6 45.50 EUR

47 Andrea Maier-Richter: Computerunterstütztes Lernen mit Lösungsbeispielen in der Chemie. *Eine Evaluationsstudie im Themenbereich Löslichkeit*
ISBN 978-3-8325-1046-6 40.50 EUR

48 Jochen Peuckert: Stabilität und Ausprägung kognitiver Strukturen zum Atombegriff
ISBN 978-3-8325-1104-3 40.50 EUR

49 Maik Walpuski: Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback
ISBN 978-3-8325-1184-5 40.50 EUR

50 Helmut Fischler, Christiane S. Reiners (Hrsg.): Die Teilchenstruktur der Materie im Physik- und Chemieunterricht
ISBN 978-3-8325-1225-5 34.90 EUR

51 Claudia Eysel: Interdisziplinäres Lehren und Lernen in der Lehrerbildung. *Eine empirische Studie zum Kompetenzerwerb in einer komplexen Lernumgebung*
ISBN 978-3-8325-1238-5 40.50 EUR

52 Johannes Günther: Lehrerfortbildung über die Natur der Naturwissenschaften. *Studien über das Wissenschaftsverständnis von Grundschullehrkräften*
ISBN 978-3-8325-1287-3 40.50 EUR

53 Christoph Neugebauer: Lernen mit Simulationen und der Einfluss auf das Problemlösen in der Physik
ISBN 978-3-8325-1300-9 40.50 EUR

54 Andreas Schnirch: Gendergerechte Interessen- und Motivationsförderung im Kontext naturwissenschaftlicher Grundbildung. *Konzeption, Entwicklung und Evaluation einer multimedial unterstützten Lernumgebung*
ISBN 978-3-8325-1334-4 40.50 EUR

55 Hilde Köster: Freies Explorieren und Experimentieren. *Eine Untersuchung zur selbstbestimmten Gewinnung von Erfahrungen mit physikalischen Phänomenen im Sachunterricht*
ISBN 978-3-8325-1348-1 40.50 EUR

56 Eva Heran-Dörr: Entwicklung und Evaluation einer Lehrerfortbildung zur Förderung der physikdidaktischen Kompetenz von Sachunterrichtslehrkräften
ISBN 978-3-8325-1377-1 40.50 EUR

57 Agnes Szabone Varnai: Unterstützung des Problemlösens in Physik durch den Einsatz von Simulationen und die Vorgabe eines strukturierten Kooperationsformats
ISBN 978-3-8325-1403-7 40.50 EUR

58 Johannes Rethfeld: Aufgabenbasierte Lernprozesse in selbstorganisationsoffenem Unterricht der Sekundarstufe I zum Themengebiet ELEKTROSTATIK. *Eine Feldstudie in vier 10. Klassen zu einer kartenbasierten Lernumgebung mit Aufgaben aus der Elektrostatik*
ISBN 978-3-8325-1416-7 40.50 EUR

59 Christian Henke: Experimentell-naturwissenschaftliche Arbeitsweisen in der Oberstufe. *Untersuchung am Beispiel des HIGHSEA-Projekts in Bremerhaven*
ISBN 978-3-8325-1515-7 40.50 EUR

60 Lutz Kasper: Diskursiv-narrative Elemente für den Physikunterricht. *Entwicklung und Evaluation einer multimedialen Lernumgebung zum Erdmagnetismus*
ISBN 978-3-8325-1537-9 40.50 EUR

61 Thorid Rabe: Textgestaltung und Aufforderung zu Selbsterklärungen beim Physiklernen mit Multimedia
ISBN 978-3-8325-1539-3 40.50 EUR

62 Ina Glemnitz: Vertikale Vernetzung im Chemieunterricht. *Ein Vergleich von traditionellem Unterricht mit Unterricht nach Chemie im Kontext*
ISBN 978-3-8325-1628-4 40.50 EUR

63 Erik Einhaus: Schülerkompetenzen im Bereich Wärmelehre. *Entwicklung eines Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells fachbezogener Kompetenzen*
ISBN 978-3-8325-1630-7 40.50 EUR

64 Jasmin Neuroth: Concept Mapping als Lernstrategie. *Eine Interventionsstudie zum Chemielernen aus Texten*
ISBN 978-3-8325-1659-8 40.50 EUR

65 Hans Gerd Hegeler-Burkhart: Zur Kommunikation von Hauptschülerinnen und Hauptschülern in einem handlungsorientierten und fächerübergreifenden Unterricht mit physikalischen und technischen Inhalten
ISBN 978-3-8325-1667-3 40.50 EUR

66 Karsten Rincke: Sprachentwicklung und Fachlernen im Mechanikunterricht. *Sprache und Kommunikation bei der Einführung in den Kraftbegriff*
ISBN 978-3-8325-1699-4 40.50 EUR

67 Nina Strehle: Das Ion im Chemieunterricht. *Alternative Schülervorstellungen und curriculare Konsequenzen*
ISBN 978-3-8325-1710-6 40.50 EUR

68 Martin Hopf: Problemorientierte Schülerexperimente
ISBN 978-3-8325-1711-3 40.50 EUR

69 Anne Beerenwinkel: Fostering conceptual change in chemistry classes using expository texts
ISBN 978-3-8325-1721-2 40.50 EUR

70 Roland Berger: Das Gruppenpuzzle im Physikunterricht der Sekundarstufe II. *Eine empirische Untersuchung auf der Grundlage der Selbstbestimmungstheorie der Motivation*
ISBN 978-3-8325-1732-8 40.50 EUR

71 Giuseppe Colicchia: Physikunterricht im Kontext von Medizin und Biologie. *Entwicklung und Erprobung von Unterrichtseinheiten*
ISBN 978-3-8325-1746-5 40.50 EUR

72 Sandra Winheller: Geschlechtsspezifische Auswirkungen der Lehrer-Schüler-Interaktion im Chemieanfangsunterricht
ISBN 978-3-8325-1757-1 40.50 EUR

73 Isabel Wahser: Training von naturwissenschaftlichen Arbeitsweisen zur Unterstützung experimenteller Kleingruppenarbeit im Fach Chemie
ISBN 978-3-8325-1815-8 40.50 EUR

74 Claus Brell: Lernmedien und Lernerfolg - reale und virtuelle Materialien im Physikunterricht. *Empirische Untersuchungen in achten Klassen an Gymnasien (Laborstudie) zum Computereinsatz mit Simulation und IBE*
ISBN 978-3-8325-1829-5 40.50 EUR

75 Rainer Wackermann: Überprüfung der Wirksamkeit eines Basismodell-Trainings für Physiklehrer
ISBN 978-3-8325-1882-0 40.50 EUR

76 Oliver Tepner: Effektivität von Aufgaben im Chemieunterricht der Sekundarstufe I
ISBN 978-3-8325-1919-3 40.50 EUR

77 Claudia Geyer: Museums- und Science-Center-Besuche im naturwissenschaftlichen Unterricht aus einer motivationalen Perspektive. *Die Sicht von Lehrkräften und Schülerinnen und Schülern*
ISBN 978-3-8325-1922-3 40.50 EUR

78 Tobias Leonhard: Professionalisierung in der Lehrerbildung. *Eine explorative Studie zur Entwicklung professioneller Kompetenzen in der Lehrererstausbildung*
ISBN 978-3-8325-1924-7 40.50 EUR

79 Alexander Kauertz: Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben
ISBN 978-3-8325-1925-4 40.50 EUR

80 Regina Hübinger: Schüler auf Weltreise. *Entwicklung und Evaluation von Lehr-/ Lernmaterialien zur Förderung experimentell-naturwissenschaftlicher Kompetenzen für die Jahrgangsstufen 5 und 6*
ISBN 978-3-8325-1932-2 40.50 EUR

81 Christine Waltner: Physik lernen im Deutschen Museum
ISBN 978-3-8325-1933-9 40.50 EUR

82 Torsten Fischer: Handlungsmuster von Physiklehrkräften beim Einsatz neuer Medien. *Fallstudien zur Unterrichtspraxis*
ISBN 978-3-8325-1948-3 42.00 EUR

83 Corinna Kieren: Chemiehausaufgaben in der Sekundarstufe I des Gymnasiums. *Fragebogenerhebung zur gegenwärtigen Praxis und Entwicklung eines optimierten Hausaufgabendesigns im Themenbereich Säure-Base*
978-3-8325-1975-9 37.00 EUR

84 Marco Thiele: Modelle der Thermohalinen Zirkulation im Unterricht. *Eine empirische Studie zur Förderung des Modellverständnisses*
ISBN 978-3-8325-1982-7 40.50 EUR

85 Bernd Zinn: Physik lernen, um Physik zu lehren. *Eine Möglichkeit für interessanteren Physikunterricht*
ISBN 978-3-8325-1995-7 39.50 EUR

86 Esther Klaes: Außerschulische Lernorte im naturwissenschaftlichen Unterricht. *Die Perspektive der Lehrkraft*
ISBN 978-3-8325-2006-9 43.00 EUR

87 Marita Schmidt: Kompetenzmodellierung und -diagnostik im Themengebiet Energie der Sekundarstufe I. *Entwicklung und Erprobung eines Testinventars*
ISBN 978-3-8325-2024-3 37.00 EUR

88 Gudrun Franke-Braun: Aufgaben mit gestuften Lernhilfen. *Ein Aufgabenformat zur Förderung der sachbezogenen Kommunikation und Lernleistung für den naturwissenschaftlichen Unterricht*
ISBN 978-3-8325-2026-7 38.00 EUR

89 Silke Klos: Kompetenzförderung im naturwissenschaftlichen Anfangsunterricht. *Der Einfluss eines integrierten Unterrichtskonzepts*
ISBN 978-3-8325-2133-2 37.00 EUR

90 Ulrike Elisabeth Burkard: Quantenphysik in der Schule. *Bestandsaufnahme, Perspektiven und Weiterentwicklungsmöglichkeiten durch die Implementation eines Medienservers*
ISBN 978-3-8325-2215-5 43.00 EUR

91 Ulrike Gromadecki: Argumente in physikalischen Kontexten. *Welche Geltungsgründe halten Physikanfänger für überzeugend?*
ISBN 978-3-8325-2250-6 41.50 EUR

92 Jürgen Bruns: Auf dem Weg zur Förderung naturwissenschaftsspezifischer Vorstellungen von zukünftigen Chemie-Lehrenden
ISBN 978-3-8325-2257-5 43.50 EUR

93 Cornelius Marsch: Räumliche Atomvorstellung. *Entwicklung und Erprobung eines Unterrichtskonzeptes mit Hilfe des Computers*
ISBN 978-3-8325-2293-3 82.50 EUR

94 Maja Brückmann: Sachstrukturen im Physikunterricht. *Ergebnisse einer Videostudie*
ISBN 978-3-8325-2272-8 39.50 EUR

95 Sabine Fechner: Effects of Context-oriented Learning on Student Interest and Achievement in Chemistry Education
ISBN 978-3-8325-2343-5 36.50 EUR

96 Clemens Nagel: eLearning im Physikalischen Anfängerpraktikum
ISBN 978-3-8325-2355-8 39.50 EUR

97 Josef Riese: Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften
ISBN 978-3-8325-2376-3 39.00 EUR

98 Sascha Bernholt: Kompetenzmodellierung in der Chemie. *Theoretische und empirische Reflexion am Beispiel des Modells hierarchischer Komplexität*
ISBN 978-3-8325-2447-0 40.00 EUR

99 Holger Christoph Stawitz: Auswirkung unterschiedlicher Aufgabenprofile auf die Schülerleistung. *Vergleich von Naturwissenschafts- und Problemlöseaufgaben der PISA 2003-Studie*
ISBN 978-3-8325-2451-7 37.50 EUR

100 Hans Ernst Fischer, Elke Sumfleth (Hrsg.): nwu-essen – 10 Jahre Essener Forschung zum naturwissenschaftlichen Unterricht
ISBN 978-3-8325-3331-1 40.00 EUR

101 Hendrik Härtig: Sachstrukturen von Physikschulbüchern als Grundlage zur Bestimmung der Inhaltsvalidität eines Tests
ISBN 978-3-8325-2512-5 34.00 EUR

102 Thomas Grüß-Niehaus: Zum Verständnis des Löslichkeitskonzeptes im Chemieunterricht. *Der Effekt von Methoden progressiver und kollaborativer Reflexion*
ISBN 978-3-8325-2537-8 40.50 EUR

103 Patrick Bronner: Quantenoptische Experimente als Grundlage eines Curriculums zur Quantenphysik des Photons
ISBN 978-3-8325-2540-8 36.00 EUR

104 Adrian Voßkühler: Blickbewegungsmessung an Versuchsaufbauten. *Studien zur Wahrnehmung, Verarbeitung und Usability von physikbezogenen Experimenten am Bildschirm und in der Realität*
ISBN 978-3-8325-2548-4 47.50 EUR

105 Verena Tobias: Newton'sche Mechanik im Anfangsunterricht. *Die Wirksamkeit einer Einführung über die zweidimensionale Dynamik auf das Lehren und Lernen*
ISBN 978-3-8325-2558-3 54.00 EUR

106 Christian Rogge: Entwicklung physikalischer Konzepte in aufgabenbasierten Lernumgebungen
ISBN 978-3-8325-2574-3 45.00 EUR

107 Mathias Ropohl: Modellierung von Schülerkompetenzen im Basiskonzept Chemische Reaktion. *Entwicklung und Analyse von Testaufgaben*
ISBN 978-3-8325-2609-2 36.50 EUR

108 Christoph Kulgemeyer: Physikalische Kommunikationskompetenz. *Modellierung und Diagnostik*
ISBN 978-3-8325-2674-0 44.50 EUR

109 Jennifer Olszewski: The Impact of Physics Teachers' Pedagogical Content Knowledge on Teacher Actions and Student Outcomes
ISBN 978-3-8325-2680-1 33.50 EUR

110 Annika Ohle: Primary School Teachers' Content Knowledge in Physics and its Impact on Teaching and Students' Achievement
ISBN 978-3-8325-2684-9 36.50 EUR

111 Susanne Mannel: Assessing scientific inquiry. *Development and evaluation of a test for the low-performing stage*
ISBN 978-3-8325-2761-7 40.00 EUR

112 Michael Plomer: Physik physiologisch passend praktiziert. *Eine Studie zur Lernwirksamkeit von traditionellen und adressatenspezifischen Physikpraktika für die Physiologie*
ISBN 978-3-8325-2804-1 34.50 EUR

113 Alexandra Schulz: Experimentierspezifische Qualitätsmerkmale im Chemieunterricht. *Eine Videostudie*
ISBN 978-3-8325-2817-1 40.00 EUR

114 Franz Boczianowski: Eine empirische Untersuchung zu Vektoren im Physikunterricht der Mittelstufe
ISBN 978-3-8325-2843-0 39.50 EUR

115 Maria Ploog: Internetbasiertes Lernen durch Textproduktion im Fach Physik
ISBN 978-3-8325-2853-9 39.50 EUR

116 Anja Dhein: Lernen in Explorier- und Experimentiersituationen. *Eine explorative Studie zu Bedeutungsentwicklungsprozessen bei Kindern im Alter zwischen 4 und 6 Jahren*
ISBN 978-3-8325-2859-1 45.50 EUR

117 Irene Neumann: Beyond Physics Content Knowledge. *Modeling Competence Regarding Nature of Scientific Inquiry and Nature of Scientific Knowledge*
ISBN 978-3-8325-2880-5 37.00 EUR

118 Markus Emden: Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens. *Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I*
ISBN 978-3-8325-2867-6 38.00 EUR

119 Birgit Hofmann: Analyse von Blickbewegungen von Schülern beim Lesen von physikbezogenen Texten mit Bildern. *Eye Tracking als Methodenwerkzeug in der physikdidaktischen Forschung*
ISBN 978-3-8325-2925-3 59.00 EUR

120 Rebecca Knobloch: Analyse der fachinhaltlichen Qualität von Schüleräußerungen und deren Einfluss auf den Lernerfolg. *Eine Videostudie zu kooperativer Kleingruppenarbeit*
ISBN 978-3-8325-3006-8 36.50 EUR

121 Julia Hostenbach: Entwicklung und Prüfung eines Modells zur Beschreibung der Bewertungskompetenz im Chemieunterricht
ISBN 978-3-8325-3013-6 38.00 EUR

122 Anna Windt: Naturwissenschaftliches Experimentieren im Elementarbereich. *Evaluation verschiedener Lernsituationen*
ISBN 978-3-8325-3020-4 43.50 EUR

123 Eva Kölbach: Kontexteinflüsse beim Lernen mit Lösungsbeispielen
ISBN 978-3-8325-3025-9 38.50 EUR

124 Anna Lau: Passung und vertikale Vernetzung im Chemie- und Physikunterricht
ISBN 978-3-8325-3021-1 36.00 EUR

125 Jan Lamprecht: Ausbildungswege und Komponenten professioneller Handlungskompetenz. *Vergleich von Quereinsteigern mit Lehramtsabsolventen für Gymnasien im Fach Physik*
ISBN 978-3-8325-3035-8 38.50 EUR

126 Ulrike Böhm: Förderung von Verstehensprozessen unter Einsatz von Modellen
ISBN 978-3-8325-3042-6 41.00 EUR

127 Sabrina Dollny: Entwicklung und Evaluation eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Chemielehrkräften
ISBN 978-3-8325-3046-4 37.00 EUR

128 Monika Zimmermann: Naturwissenschaftliche Bildung im Kindergarten. *Eine integrative Längsschnittstudie zur Kompetenzentwicklung von Erzieherinnen*
ISBN 978-3-8325-3053-2 54.00 EUR

129 Ulf Saballus: Über das Schlussfolgern von Schülerinnen und Schülern zu öffentlichen Kontroversen mit naturwissenschaftlichem Hintergrund. *Eine Fallstudie*
ISBN 978-3-8325-3086-0 39.50 EUR

130 Olaf Krey: Zur Rolle der Mathematik in der Physik. *Wissenschaftstheoretische Aspekte und Vorstellungen Physiklernender*
ISBN 978-3-8325-3101-0 46.00 EUR

131 Angelika Wolf: Zusammenhänge zwischen der Eigenständigkeit im Physikunterricht, der Motivation, den Grundbedürfnissen und dem Lernerfolg von Schülern
ISBN 978-3-8325-3161-4 45.00 EUR

132 Johannes Börlin: Das Experiment als Lerngelegenheit. *Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität*
ISBN 978-3-8325-3170-6 45.00 EUR

133 Olaf Uhden: Mathematisches Denken im Physikunterricht. *Theorieentwicklung und Problemanalyse*
ISBN 978-3-8325-3170-6 45.00 EUR

134 Christoph Gut: Modellierung und Messung experimenteller Kompetenz. *Analyse eines large-scale Experimentiertests*
ISBN 978-3-8325-3213-0 40.00 EUR

135 Antonio Rueda: Lernen mit ExploMultimedial in kolumbianischen Schulen. *Analyse von kurzzeitigen Lernprozessen und der Motivation beim länderübergreifenden Einsatz einer deutschen computergestützten multimedialen Lernumgebung für den naturwissenschaftlichen Unterricht*
ISBN 978-3-8325-3218-5 45.50 EUR

136 Krisztina Berger: Bilder, Animationen und Notizen. *Empirische Untersuchung zur Wirkung einfacher visueller Repräsentationen und Notizen auf den Wissenserwerb in der Optik*
ISBN 978-3-8325-3238-3 41.50 EUR

137 Antony Crossley: Untersuchung des Einflusses unterschiedlicher physikalischer Konzepte auf den Wissenserwerb in der Thermodynamik der Sekundarstufe I
ISBN 978-3-8325-3275-8 40.00 EUR

138 Tobias Viering: Entwicklung physikalischer Kompetenz in der Sekundarstufe I. *Validierung eines Kompetenzentwicklungsmodells für das Energiekonzept im Bereich Fachwissen*
ISBN 978-3-8325-3277-2 37.00 EUR

139 Nico Schreiber: Diagnostik experimenteller Kompetenz. *Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells*
ISBN 978-3-8325-3284-0 39.00 EUR

140 Sarah Hundertmark: Einblicke in kollaborative Lernprozesse. *Eine Fallstudie zur reflektierenden Zusammenarbeit unterstützt durch die Methoden Concept Mapping und Lernbegleitbogen*
ISBN 978-3-8325-3251-2 43.00 EUR

141 Ronny Scherer: Analyse der Struktur, Messinvarianz und Ausprägung komplexer Problemlösekompetenz im Fach Chemie. *Eine Querschnittstudie in der Sekundarstufe I und am Übergang zur Sekundarstufe II*
ISBN 978-3-8325-3312-0 43.00 EUR

142 Patricia Heitmann: Bewertungskompetenz im Rahmen naturwissenschaftlicher Problemlöseprozesse. *Modellierung und Diagnose der Kompetenzen Bewertung und analytisches Problemlösen für das Fach Chemie*
ISBN 978-3-8325-3314-4 37.00 EUR

143 Jan Fleischhauer: Wissenschaftliches Argumentieren und Entwicklung von Konzepten beim Lernen von Physik
ISBN 978-3-8325-3325-0 35.00 EUR

144 Nermin Özcan: Zum Einfluss der Fachsprache auf die Leistung im Fach Chemie. *Eine Förderstudie zur Fachsprache im Chemieunterricht*
ISBN 978-3-8325-3328-1 36.50 EUR

145 Helena van Vorst: Kontextmerkmale und ihr Einfluss auf das Schülerinteresse im Fach Chemie
ISBN 978-3-8325-3321-2 38.50 EUR

146 Janine Cappell: Fachspezifische Diagnosekompetenz angehender Physiklehrkräfte in der ersten Ausbildungsphase
ISBN 978-3-8325-3356-4 38.50 EUR

147 Susanne Bley: Förderung von Transferprozessen im Chemieunterricht
ISBN 978-3-8325-3407-3 40.50 EUR

148 Cathrin Blaes: Die übungsgestützte Lehrerpräsentation im Chemieunterricht der Sekundarstufe I. *Evaluation der Effektivität*
ISBN 978-3-8325-3409-7 43.50 EUR

149 Julia Suckut: Die Wirksamkeit von piko-OWL als Lehrerfortbildung. Eine Evaluation zum Projekt *Physik im Kontext* in Fallstudien
ISBN 978-3-8325-3440-0 45.00 EUR

150 Alexandra Dorschu: Die Wirkung von Kontexten in Physikkompetenztestaufgaben
ISBN 978-3-8325-3446-2 37.00 EUR

151 Jochen Scheid: Multiple Repräsentationen, Verständnis physikalischer Experimente und kognitive Aktivierung: *Ein Beitrag zur Entwicklung der Aufgabenkultur*
ISBN 978-3-8325-3449-3 49.00 EUR

152 Tim Plasa: Die Wahrnehmung von Schülerlaboren und Schülerforschungszentren
ISBN 978-3-8325-3483-7 35.50 EUR

153 Felix Schoppmeier: Physikkompetenz in der gymnasialen Oberstufe.*Entwicklung und Validierung eines Kompetenzstrukturmodells für den Kompetenzbereich Umgang mit Fachwissen*
ISBN 978-3-8325-3502-5 36.00 EUR

154 Katharina Groß: Experimente alternativ dokumentieren. *Eine qualitative Studie zur Förderung der Diagnose- und Differenzierungskompetenz in der Chemielehrerbildung*
ISBN 978-3-8325-3508-7 43.50 EUR

155 Barbara Hank: Konzeptwandelprozesse im Anfangsunterricht Chemie. *Eine quasiexperimentelle Längsschnittstudie*
ISBN 978-3-8325-3519-3 38.50 EUR

156 Katja Freyer: Zum Einfluss von Studieneingangsvoraussetzungen auf den Studienerfolg Erstsemesterstudierender im Fach Chemie
ISBN 978-3-8325-3544-5 38.00 EUR

157 Alexander Rachel: Auswirkungen instruktionaler Hilfen bei der Einführung des (Ferro-)Magnetismus. *Eine Vergleichsstudie in der Primar- und Sekundarstufe*
ISBN 978-3-8325-3548-3 43.50 EUR

158 Sebastian Ritter: Einfluss des Lerninhalts Nanogrößeneffekte auf Teilchen- und Teilchenmodellvorstellungen von Schülerinnen und Schülern
ISBN 978-3-8325-3558-2 36.00 EUR

159 Andrea Harbach: Problemorientierung und Vernetzung in kontextbasierten Lernaufgaben
ISBN 978-3-8325-3564-3 39.00 EUR

160 David Obst: Interaktive Tafeln im Physikunterricht. *Entwicklung und Evaluation einer Lehrerfortbildung*
ISBN 978-3-8325-3582-7 40.50 EUR

161 Sophie Kirschner: Modellierung und Analyse des Professionswissens von Physiklehrkräften
ISBN 978-3-8325-3601-5 35.00 EUR

162 Katja Stief: Selbstregulationsprozesse und Hausaufgabenmotivation im Chemieunterricht
ISBN 978-3-8325-3631-2 34.00 EUR

163 Nicola Meschede: Professionelle Wahrnehmung der inhaltlichen Strukturierung im naturwissenschaftlichen Grundschulunterricht. *Theoretische Beschreibung und empirische Erfassung*
ISBN 978-3-8325-3668-8 37.00 EUR

164 Johannes Maximilian Barth: Experimentieren im Physikunterricht der gymnasialen Oberstufe. *Eine Rekonstruktion übergeordneter Einbettungsstrategien*
ISBN 978-3-8325-3681-7 39.00 EUR

165 Sandra Lein: Das Betriebspraktikum in der Lehrerbildung. *Eine Untersuchung zur Förderung der Wissenschafts- und Technikbildung im allgemeinbildenden Unterricht*
ISBN 978-3-8325-3698-5 40.00 EUR

166 Veranika Maiseyenka: Modellbasiertes Experimentieren im Unterricht. *Praxistauglichkeit und Lernwirkungen*
ISBN 978-3-8325-3708-1 38.00 EUR

167 Christoph Stolzenberger: Der Einfluss der didaktischen Lernumgebung auf das Erreichen geforderter Bildungsziele am Beispiel der W- und P-Seminare im Fach Physik
ISBN 978-3-8325-3708-1 38.00 EUR

168 Pia Altenburger: Mehrebenenregressionsanalysen zum Physiklernen im Sachunterricht der Primarstufe. *Ergebnisse einer Evaluationsstudie.*
ISBN 978-3-8325-3717-3 37.50 EUR

169 Nora Ferber: Entwicklung und Validierung eines Testinstruments zur Erfassung von Kompetenzentwicklung im Fach Chemie in der Sekundarstufe I
ISBN 978-3-8325-3727-2 39.50 EUR

170 Anita Stender: Unterrichtsplanung: Vom Wissen zum Handeln.
Theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung
ISBN 978-3-8325-3750-0 41.50 EUR

171 Jenna Koenen: Entwicklung und Evaluation von experimentunterstützten Lösungsbeispielen zur Förderung naturwissenschaftlich-experimenteller Arbeitsweisen
ISBN 978-3-8325-3785-2 43.00 EUR

172 Teresa Henning: Empirische Untersuchung kontextorientierter Lernumgebungen in der Hochschuldidaktik. *Entwicklung und Evaluation kontextorientierter Aufgaben in der Studieneingangsphase für Fach- und Nebenfachstudierende der Physik*
ISBN 978-3-8325-3801-9 43.00 EUR

173 Alexander Pusch: Fachspezifische Instrumente zur Diagnose und individuellen Förderung von Lehramtsstudierenden der Physik
ISBN 978-3-8325-3829-3 38.00 EUR

174 Christoph Vogelsang: Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften. *Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz*
ISBN 978-3-8325-3846-0 50.50 EUR

175 Ingo Brebeck: Selbstreguliertes Lernen in der Studieneingangsphase im Fach Chemie
ISBN 978-3-8325-3859-0 37.00 EUR

176 Axel Eghtessad: Merkmale und Strukturen von Professionalisierungsprozessen in der ersten und zweiten Phase der Chemielehrerbildung. *Eine empirisch-qualitative Studie mit niedersächsischen Fachleiter_innen der Sekundarstufenlehrämter*
ISBN 978-3-8325-3861-3 45.00 EUR

177 Andreas Nehring: Wissenschaftliche Denk- und Arbeitsweisen im Fach Chemie. Eine kompetenzorientierte Modell- und Testentwicklung für den Bereich der Erkenntnisgewinnung
ISBN 978-3-8325-3872-9 39.50 EUR

178 Maike Schmidt: Professionswissen von Sachunterrichtslehrkräften. Zusammenhangsanalyse zur Wirkung von Ausbildungshintergrund und Unterrichtserfahrung auf das fachspezifische Professionswissen im Unterrichtsinhalt „Verbrennung“
ISBN 978-3-8325-3907-8 38.50 EUR

179 Jan Winkelmann: Auswirkungen auf den Fachwissenszuwachs und auf affektive Schülermerkmale durch Schüler- und Demonstrationsexperimente im Physikunterricht
ISBN 978-3-8325-3915-3 41.00 EUR

180 Iwen Kobow: Entwicklung und Validierung eines Testinstrumentes zur Erfassung der Kommunikationskompetenz im Fach Chemie
ISBN 978-3-8325-3927-6 34.50 EUR

181 Yvonne Gramzow: Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion
ISBN 978-3-8325-3931-3 42.50 EUR

182 Evelin Schröter: Entwicklung der Kompetenzerwartung durch Lösen physikalischer Aufgaben einer multimedialen Lernumgebung
ISBN 978-3-8325-3975-7 54.50 EUR

183 Inga Kallweit: Effektivität des Einsatzes von Selbsteinschätzungsbögen im Chemieunterricht der Sekundarstufe I. *Individuelle Förderung durch selbstreguliertes Lernen*
ISBN 978-3-8325-3965-8 44.00 EUR

184 Andrea Schumacher: Paving the way towards authentic chemistry teaching. *A contribution to teachers' professional development*
ISBN 978-3-8325-3976-4 48.50 EUR

185 David Woitkowski: Fachliches Wissen Physik in der Hochschulausbildung. *Konzeptualisierung, Messung, Niveaubildung*
ISBN 978-3-8325-3988-7 53.00 EUR

186 Marianne Korner: Cross-Age Peer Tutoring in Physik. *Evaluation einer Unterrichtsmethode*
ISBN 978-3-8325-3979-5 38.50 EUR

187 Simone Nakoinz: Untersuchung zur Verknüpfung submikroskopischer und makroskopischer Konzepte im Fach Chemie
ISBN 978-3-8325-4057-9 38.50 EUR

188 Sandra Anus: Evaluation individueller Förderung im Chemieunterricht.*Adaptivität von Lerninhalten an das Vorwissen von Lernenden am Beispiel des Basiskonzeptes Chemische Reaktion*
ISBN 978-3-8325-4059-3 43.50 EUR

189 Thomas Roßbegalle: Fachdidaktische Entwicklungsforschung zum besseren Verständnis atmosphärischer Phänomene. *Treibhauseffekt, saurer Regen und stratosphärischer Ozonabbau als Kontexte zur Vermittlung von Basiskonzepten der Chemie*
ISBN 978-3-8325-4059-3 45.50 EUR

190 Kathrin Steckenmesser-Sander: Gemeinsamkeiten und Unterschiede physikbezogener Handlungs-, Denk- und Lernprozesse von Mädchen und Jungen
ISBN 978-3-8325-4066-1 38.50 EUR

191 Cornelia Geller: Lernprozessorientierte Sequenzierung des Physikunterrichts im Zusammenhang mit Fachwissenserwerb. *Eine Videostudie in Finnland, Deutschland und der Schweiz*
ISBN 978-3-8325-4082-1 35.50 EUR

192 Jan Hofmann: Untersuchung des Kompetenzaufbaus von Physiklehrkräften während einer Fortbildungsmaßnahme
ISBN 978-3-8325-4104-0 38.50 EUR

193 Andreas Dickhäuser: Chemiespezifischer Humor. *Theoriebildung, Materialentwicklung, Evaluation*
ISBN 978-3-8325-4108-8 37.00 EUR

194 Stefan Korte: Die Grenzen der Naturwissenschaft als Thema des Physikunterrichts
ISBN 978-3-8325-4112-5 57.50 EUR

195 Carolin Hülsmann: Kurswahlmotive im Fach Chemie. Eine Studie zum Wahlverhalten und Erfolg von Schülerinnen und Schülern in der gymnasialen Oberstufe
ISBN 978-3-8325-4144-6 49.00 EUR

196 Caroline Körbs: Mindeststandards im Fach Chemie am Ende der Pflichtschulzeit
ISBN 978-3-8325-4148-4 34.00 EUR

197 Andreas Vorholzer: Wie lassen sich Kompetenzen des experimentellen Denkens und Arbeitens fördern? *Eine empirische Untersuchung der Wirkung eines expliziten und eines impliziten Instruktionsansatzes*
ISBN 978-3-8325-4194-1 37.50 EUR

198 Anna Katharina Schmitt: Entwicklung und Evaluation einer Chemielehrerfortbildung zum Kompetenzbereich Erkenntnisgewinnung
ISBN 978-3-8325-4228-3 39.50 EUR

199 Christian Maurer: Strukturierung von Lehr-Lern-Sequenzen
ISBN 978-3-8325-4247-4 36.50 EUR

200 Helmut Fischler, Elke Sumfleth (Hrsg.): Professionelle Kompetenz von Lehrkräften der Chemie und Physik
ISBN 978-3-8325-4523-9 34.00 EUR

201 Simon Zander: Lehrerfortbildung zu Basismodellen und Zusammenhänge zum Fachwissen
ISBN 978-3-8325-4248-1 35.00 EUR

202 Kerstin Arndt: Experimentierkompetenz erfassen.
Analyse von Prozessen und Mustern am Beispiel von Lehramtsstudierenden der Chemie
ISBN 978-3-8325-4266-5 45.00 EUR

203 Christian Lang: Kompetenzorientierung im Rahmen experimentalchemischer Praktika
ISBN 978-3-8325-4268-9 42.50 EUR

204 Eva Cauet: Testen wir relevantes Wissen? *Zusammenhang zwischen dem Professionswissen von Physiklehrkräften und gutem und erfolgreichem Unterrichten*
ISBN 978-3-8325-4276-4 39.50 EUR

205 Patrick Löffler: Modellanwendung in Problemlöseaufgaben. *Wie wirkt Kontext?*
ISBN 978-3-8325-4303-7 35.00 EUR

206 Carina Gehlen: Kompetenzstruktur naturwissenschaftlicher Erkenntnisgewinnung im Fach Chemie
ISBN 978-3-8325-4318-1 43.00 EUR

207 Lars Oettinghaus: Lehrerüberzeugungen und physikbezogenes Professionswissen. *Vergleich von Absolventinnen und Absolventen verschiedener Ausbildungswege im Physikreferendariat*
ISBN 978-3-8325-4319-8 38.50 EUR

208 Jennifer Petersen: Zum Einfluss des Merkmals Humor auf die Gesundheitsförderung im Chemieunterricht der Sekundarstufe I.
Eine Interventionsstudie zum Thema Sonnenschutz
ISBN 978-3-8325-4348-8 40.00 EUR

209 Philipp Straube: Modellierung und Erfassung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung bei (Lehramts-) Studierenden im Fach Physik
ISBN 978-3-8325-4351-8 35.50 EUR

210 Martin Dickmann: Messung von Experimentierfähigkeiten. *Validierungsstudien zur Qualität eines computerbasierten Testverfahrens*
ISBN 978-3-8325-4356-3 41.00 EUR

211 Markus Bohlmann: Science Education. Empirie, Kulturen und Mechanismen der Didaktik der Naturwissenschaften
ISBN 978-3-8325-4377-8 44.00 EUR

212 Martin Draude: Die Kompetenz von Physiklehrkräften, Schwierigkeiten von Schülerinnen und Schülern beim eigenständigen Experimentieren zu diagnostizieren
ISBN 978-3-8325-4382-2 37.50 EUR

213 Henning Rode: Prototypen evidenzbasierten Physikunterrichts. *Zwei empirische Studien zum Einsatz von Feedback und Blackboxes in der Sekundarstufe*
ISBN 978-3-8325-4389-1 42.00 EUR

214 Jan-Henrik Kechel: Schülerschwierigkeiten beim eigenständigen Experimentieren. *Eine qualitative Studie am Beispiel einer Experimentieraufgabe zum Hooke'schen Gesetz*
ISBN 978-3-8325-4392-1 55.00 EUR

215 Katharina Fricke: Classroom Management and its Impact on Lesson Outcomes in Physics. *A multi-perspective comparison of teaching practices in primary and secondary schools*
ISBN 978-3-8325-4394-5 40.00 EUR

216 Hannes Sander: Orientierungen von Jugendlichen beim Urteilen und Entscheiden in Kontexten nachhaltiger Entwicklung. *Eine rekonstruktive Perspektive auf Bewertungskompetenz in der Didaktik der Naturwissenschaft*
ISBN 978-3-8325-4434-8 46.00 EUR

217 Inka Haak: Maßnahmen zur Unterstützung kognitiver und metakognitiver Prozesse in der Studieneingangsphase. *Eine Design-Based-Research-Studie zum universitären Lernzentrum Physiktreff*
ISBN 978-3-8325-4437-9 46.50 EUR

218 Martina Brandenburger: Was beeinflusst den Erfolg beim Problemlösen in der Physik? *Eine Untersuchung mit Studierenden*
ISBN 978-3-8325-4409-6 42.50 EUR

219 Corinna Helms: Entwicklung und Evaluation eines Trainings zur Verbesserung der Erklärqualität von Schülerinnen und Schülern im Gruppenpuzzle
ISBN 978-3-8325-4454-6 42.50 EUR

220 Viktoria Rath: Diagnostische Kompetenz von angehenden Physiklehrkräften. *Modellierung, Testinstrumentenentwicklung und Erhebung der Performanz bei der Diagnose von Schülervorstellungen in der Mechanik*
ISBN 978-3-8325-4456-0 42.50 EUR

221 Janne Krüger: Schülerperspektiven auf die zeitliche Entwicklung der Naturwissenschaften
ISBN 978-3-8325-4457-7 45.50 EUR

222 Stefan Mutke: Das Professionswissen von Chemiereferendarinnen und -referendaren in Nordrhein-Westfalen. *Eine Längsschnittstudie*
ISBN 978-3-8325-4458-4 37.50 EUR

223 Sebastian Habig: Systematisch variierte Kontextaufgaben und ihr Einfluss auf kognitive und affektive Schülerfaktoren
ISBN 978-3-8325-4467-6 40.50 EUR

224 Sven Liepertz: Zusammenhang zwischen dem Professionswissen von Physiklehrkräften, dem sachstrukturellen Angebot des Unterrichts und der Schülerleistung
ISBN 978-3-8325-4480-5 34.00 EUR

225 Elina Platova: Optimierung eines Laborpraktikums durch kognitive Aktivierung
ISBN 978-3-8325-4481-2 39.00 EUR

226 Tim Reschke: Lesegeschichten im Chemieunterricht der Sekundarstufe I zur Unterstützung von situationalem Interesse und Lernerfolg
ISBN 978-3-8325-4487-4 41.00 EUR

227 Lena Mareike Walper: Entwicklung der physikbezogenen Interessen und selbstbezogenen Kognitionen von Schülerinnen und Schülern in der Übergangsphase von der Primar- in die Sekundarstufe. *Eine Längsschnittanalyse vom vierten bis zum siebten Schuljahr*
ISBN 978-3-8325-4495-9 43.00 EUR

228 Stefan Anthofer: Förderung des fachspezifischen Professionswissens von Chemielehramtsstudierenden
ISBN 978-3-8325-4498-0 39.50 EUR

229 Marcel Bullinger: Handlungsorientiertes Physiklernen mit instruierten Selbsterklärungen in der Primarstufe. *Eine experimentelle Laborstudie*
ISBN 978-3-8325-4504-8 44.00 EUR

230 Thomas Amenda: Bedeutung fachlicher Elementarisierungen für das Verständnis der Kinematik
ISBN 978-3-8325-4531-4 43.50 EUR

231 Sabrina Milke: Beeinflusst *Priming* das Physiklernen?
Eine empirische Studie zum Dritten Newtonschen Axiom
ISBN 978-3-8325-4549-4 42.00 EUR

232 Corinna Erfmann: Ein anschaulicher Weg zum Verständnis der elektromagnetischen Induktion. *Evaluation eines Unterrichtsvorschlags und Validierung eines Leistungsdiagnoseinstruments*
ISBN 978-3-8325-4550-5 49.50 EUR

233 Hanne Rautenstrauch: Erhebung des (Fach-)Sprachstandes bei Lehramtsstudierenden im Kontext des Faches Chemie
ISBN 978-3-8325-4556-7 40.50 EUR

234 Tobias Klug: Wirkung kontextorientierter physikalischer Praktikumsversuche auf Lernprozesse von Studierenden der Medizin
ISBN 978-3-8325-4558-1 37.00 EUR

235 Mareike Bohrmann: Zur Förderung des Verständnisses der Variablenkontrolle im naturwissenschaftlichen Sachunterricht
ISBN 978-3-8325-4559-8 52.00 EUR

236 Anja Schödl: FALKO-Physik – Fachspezifische Lehrerkompetenzen im Fach Physik. *Entwicklung und Validierung eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Physiklehrkräften*
ISBN 978-3-8325-4553-6 40.50 EUR

237 Hilda Scheuermann: Entwicklung und Evaluation von Unterstützungsmaßnahmen zur Förderung der Variablenkontrollstrategie beim Planen von Experimenten
ISBN 978-3-8325-4568-0 39.00 EUR

238 Christian G. Strippel: Naturwissenschaftliche Erkenntnisgewinnung an chemischen Inhalten vermitteln. *Konzeption und empirische Untersuchung einer Ausstellung mit Experimentierstation*
ISBN 978-3-8325-4577-2 41.50 EUR

239 Sarah Rau: Durchführung von Sachunterricht im Vorbereitungsdienst. *Eine längsschnittliche, videobasierte Unterrichtsanalyse*
ISBN 978-3-8325-4579-6 46.00 EUR

240 Thomas Plotz: Lernprozesse zu nicht-sichtbarer Strahlung. *Empirische Untersuchungen in der Sekundarstufe 2*
ISBN 978-3-8325-4624-3 39.50 EUR

241 Wolfgang Aschauer: Elektrische und magnetische Felder. *Eine empirische Studie zu Lernprozessen in der Sekundarstufe II*
ISBN 978-3-8325-4625-0 50.00 EUR

242 Anna Donhauser: Didaktisch rekonstruierte Materialwissenschaft. *Aufbau und Konzeption eines Schülerlabors für den Exzellenzcluster Engineering of Advanced Materials*
ISBN 978-3-8325-4636-6 39.00 EUR

243 Katrin Schüßler: Lernen mit Lösungsbeispielen im Chemieunterricht. *Einflüsse auf Lernerfolg, kognitive Belastung und Motivation*
ISBN 978-3-8325-4640-3 42.50 EUR

244 Timo Fleischer: Untersuchung der chemischen Fachsprache unter besonderer Berücksichtigung chemischer Repräsentationen
ISBN 978-3-8325-4642-7 46.50 EUR

245 Rosina Steininger: Concept Cartoons als Stimuli für Kleingruppendiskussionen im Chemieunterricht. *Beschreibung und Analyse einer komplexen Lerngelegenheit*
ISBN 978-3-8325-4647-2 39.00 EUR

246 Daniel Rehfeldt: Erfassung der Lehrqualität naturwissenschaftlicher Experimentalpraktika
ISBN 978-3-8325-4590-1 40.00 EUR

247 Sandra Puddu: Implementing Inquiry-based Learning in a Diverse Classroom: Investigating Strategies of Scaffolding and Students' Views of Scientific Inquiry
ISBN 978-3-8325-4591-8 35.50 EUR

248 Markus Bliersbach: Kreativität in der Chemie. *Erhebung und Förderung der Vorstellungen von Chemielehramtsstudierenden*
ISBN 978-3-8325-4593-2 44.00 EUR

249 Lennart Kimpel: Aufgaben in der Allgemeinen Chemie. *Zum Zusammenspiel von chemischem Verständnis und Rechenfähigkeit*
ISBN 978-3-8325-4618-2 36.00 EUR

250 Louise Bindel: Effects of integrated learning: explicating a mathematical concept in inquiry-based science camps
ISBN 978-3-8325-4655-7 37.50 EUR

251 Michael Wenzel: Computereinsatz in Schule und Schülerlabor. *Einstellung von Physiklehrkräften zu Neuen Medien*
ISBN 978-3-8325-4659-5 38.50 EUR

252 Laura Muth: Einfluss der Auswertephase von Experimenten im Physikunterricht. *Ergebnisse einer Interventionsstudie zum Zuwachs von Fachwissen und experimenteller Kompetenz von Schülerinnen und Schülern*
ISBN 978-3-8325-4675-5 36.50 EUR

253 Annika Fricke: Interaktive Skripte im Physikalischen Praktikum. *Entwicklung und Evaluation von Hypermedien für die Nebenfachausbildung*
ISBN 978-3-8325-4676-2 41.00 EUR

254 Julia Haase: Selbstbestimmtes Lernen im naturwissenschaftlichen Sachunterricht. *Eine empirische Interventionsstudie mit Fokus auf Feedback und Kompetenzerleben*
ISBN 978-3-8325-4685-4 38.50 EUR

255 Antje J. Heine: Was ist Theoretische Physik? *Eine wissenschaftstheoretische Betrachtung und Rekonstruktion von Vorstellungen von Studierenden und Dozenten über das Wesen der Theoretischen Physik*
ISBN 978-3-8325-4691-5 46.50 EUR

256 Claudia Meinhardt: Entwicklung und Validierung eines Testinstruments zu Selbstwirksamkeitserwartungen von (angehenden) Physiklehrkräften in physikdidaktischen Handlungsfeldern
ISBN 978-3-8325-4712-7 47.00 EUR

257 Ann-Kathrin Schlüter: Professionalisierung angehender Chemielehrkräfte für einen Gemeinsamen Unterricht
ISBN 978-3-8325-4713-4 53.50 EUR

258 Stefan Richtberg: Elektronenbahnen in Feldern. Konzeption und Evaluation einer webbasierten Lernumgebung
ISBN 978-3-8325-4723-3 49.00 EUR

259 Jan-Philipp Burde: Konzeption und Evaluation eines Unterrichtskonzepts zu einfachen Stromkreisen auf Basis des Elektronengasmodells
ISBN 978-3-8325-4726-4 57.50 EUR

260 Frank Finkenberg: Flipped Classroom im Physikunterricht
ISBN 978-3-8325-4737-4 42.50 EUR

261 Florian Treisch: Die Entwicklung der Professionellen Unterrichtswahrnehmung im Lehr-Lern-Labor Seminar
ISBN 978-3-8325-4741-4 41.50 EUR

262 Desiree Mayr: Strukturiertheit des experimentellen naturwissenschaftlichen Problemlöseprozesses
ISBN 978-3-8325-4757-8 37.00 EUR

263 Katrin Weber: Entwicklung und Validierung einer Learning Progression für das Konzept der chemischen Reaktion in der Sekundarstufe I
ISBN 978-3-8325-4762-2 48.50 EUR

264 Hauke Bartels: Entwicklung und Bewertung eines performanznahen Videovignettentests zur Messung der Erklärfähigkeit von Physiklehrkräften
ISBN 978-3-8325-4804-9 37.00 EUR

265 Karl Marniok: Zum Wesen von Theorien und Gesetzen in der Chemie. *Begriffsanalyse und Förderung der Vorstellungen von Lehramtsstudierenden*
ISBN 978-3-8325-4805-6 42.00 EUR

266 Marisa Holzapfel: Fachspezifischer Humor als Methode in der Gesundheitsbildung im Übergang von der Primarstufe zur Sekundarstufe I
ISBN 978-3-8325-4808-7 50.00 EUR

267 Anna Stolz: Die Auswirkungen von Experimentiersituationen mit unterschiedlichem Öffnungsgrad auf Leistung und Motivation der Schülerinnen und Schüler
ISBN 978-3-8325-4781-3 38.00 EUR

268 Nina Ulrich: Interaktive Lernaufgaben in dem digitalen Schulbuch eChemBook. *Einfluss des Interaktivitätsgrads der Lernaufgaben und des Vorwissens der Lernenden auf den Lernerfolg*
ISBN 978-3-8325-4814-8 43.50 EUR

269 Kim-Alessandro Weber: Quantenoptik in der Lehrerfortbildung. *Ein bedarfsgeprägtes Fortbildungskonzept zum Quantenobjekt „Photon“ mit Realexperimenten*
ISBN 978-3-8325-4792-9 55.00 EUR

270 Nina Skorsetz: Empathisierer und Systematisierer im Vorschulalter. *Eine Fragebogen- und Videostudie zur Motivation, sich mit Naturphänomenen zu beschäftigen*
ISBN 978-3-8325-4825-4 43.50 EUR

271 Franziska Kehne: Analyse des Transfers von kontextualisiert erworbenem Wissen im Fach Chemie
ISBN 978-3-8325-4846-9 45.00 EUR

272 Markus Elsholz: Das akademische Selbstkonzept angehender Physiklehrkräfte als Teil ihrer professionellen Identität. *Dimensionalität und Veränderung während einer zentralen Praxisphase*
ISBN 978-3-8325-4857-5 37.50 EUR

273 Joachim Müller: Studienerfolg in der Physik. *Zusammenhang zwischen Modellierungskompetenz und Studienerfolg*
ISBN 978-3-8325-4859-9 35.00 EUR

274 Jennifer Dörschelln: Organische Leuchtdioden. *Implementation eines innovativen Themas in den Chemieunterricht*
ISBN 978-3-8325-4865-0 59.00 EUR

275 Stephanie Strelow: Beliefs von Studienanfängern des Kombi-Bachelors Physik über die Natur der Naturwissenschaften
ISBN 978-3-8325-4881-0 40.50 EUR

276 Dennis Jaeger: Kognitive Belastung und aufgabenspezifische sowie personenspezifische Einflussfaktoren beim Lösen von Physikaufgaben
ISBN 978-3-8325-4928-2 50.50 EUR

277 Vanessa Fischer: Der Einfluss von Interesse und Motivation auf die Messung von Fach- und Bewertungskompetenz im Fach Chemie
ISBN 978-3-8325-4933-6 39.00 EUR

278 René Dohrmann: Professionsbezogene Wirkungen einer Lehr-Lern-Labor-Veranstaltung. *Eine multimethodische Studie zu den professionsbezogenen Wirkungen einer Lehr-Lern-Labor-Blockveranstaltung auf Studierende der Bachelorstudiengänge Lehramt Physik und Grundschulpädagogik (Sachunterricht)*
ISBN 978-3-8325-4958-9 40.00 EUR

279 Meike Bergs: Can We Make Them Use These Strategies? *Fostering Inquiry-Based Science Learning Skills with Physical and Virtual Experimentation Environments*
ISBN 978-3-8325-4962-6 39.50 EUR

280 Marie-Therese Hauerstein: Untersuchung zur Effektivität von Strukturierung und Binnendifferenzierung im Chemieunterricht der Sekundarstufe I. *Evaluation der Strukturierungshilfe Lernleiter*
ISBN 978-3-8325-4982-4 42.50 EUR

281 Verena Zucker: Erkennen und Beschreiben von formativem Assessment im naturwissenschaftlichen Grundschulunterricht. *Entwicklung eines Instruments zur Erfassung von Teilfähigkeiten der professionellen Wahrnehmung von Lehramtsstudierenden*
ISBN 978-3-8325-4991-6 38.00 EUR

282 Victoria Telser: Erfassung und Förderung experimenteller Kompetenz von Lehrkräften im Fach Chemie
ISBN 978-3-8325-4996-1 50.50 EUR

283 Kristine Tschirschky: Entwicklung und Evaluation eines gedächtnisorientierten Aufgabendesigns für Physikaufgaben
ISBN 978-3-8325-5002-8 42.50 EUR

284 Thomas Elert: Course Success in the Undergraduate General Chemistry Lab
ISBN 978-3-8325-5004-2 41.50 EUR

285 Britta Kalthoff: Explizit oder implizit? *Untersuchung der Lernwirksamkeit verschiedener fachmethodischer Instruktionen im Hinblick auf fachmethodische und fachinhaltliche Fähigkeiten von Sachunterrichtsstudierenden*
ISBN 978-3-8325-5013-4 37.50 EUR

286 Thomas Dickmann: Visuelles Modellverständnis und Studienerfolg in der Chemie. *Zwei Seiten einer Medaille*
ISBN 978-3-8325-5016-5 44.00 EUR

287 Markus Sebastian Feser: Physiklehrkräfte korrigieren Schülertexte. *Eine Explorationsstudie zur fachlich-konzeptuellen und sprachlichen Leistungsfeststellung und -beurteilung im Physikunterricht*
ISBN 978-3-8325-5020-2 49.00 EUR

288 Matylda Dudzinska: Lernen mit Beispielaufgaben und Feedback im Physikunterricht der Sekundarstufe 1. *Energieerhaltung zur Lösung von Aufgaben nutzen*
ISBN 978-3-8325-5025-7 47.00 EUR

289 Ines Sonnenschein: Naturwissenschaftliche Denk- und Arbeitsprozesse Studierender im Labor
ISBN 978-3-8325-5033-2 52.00 EUR

290 Florian Simon: Der Einfluss von Betreuung und Betreuenden auf die Wirksamkeit von Schülerlaborbesuchen. *Eine Zusammenhangsanalyse von Betreuungsqualität, Betreuermerkmalen und Schülerlaborzielen sowie Replikationsstudie zur Wirksamkeit von Schülerlaborbesuchen*
ISBN 978-3-8325-5036-3 49.50 EUR

291 Marie-Annette Geyer: Physikalisch-mathematische Darstellungswechsel funktionaler Zusammenhänge. *Das Vorgehen von SchülerInnen der Sekundarstufe 1 und ihre Schwierigkeiten*
ISBN 978-3-8325-5047-9 46.50 EUR

292 Susanne Digel: Messung von Modellierungskompetenz in Physik. *Theoretische Herleitung und empirische Prüfung eines Kompetenzmodells physikspezifischer Modellierungskompetenz*
ISBN 978-3-8325-5055-4 41.00 EUR

293 Sönke Janssen: Angebots-Nutzungs-Prozesse eines Schülerlabors analysieren und gestalten. *Ein design-based research Projekt*
ISBN 978-3-8325-5065-3 57.50 EUR

294 Knut Wille: Der Productive Failure Ansatz als Beitrag zur Weiterentwicklung der Aufgabenkultur
ISBN 978-3-8325-5074-5 49.00 EUR

295 Lisanne Kraeva: Problemlösestrategien von Schülerinnen und Schülern diagnostizieren
ISBN 978-3-8325-5110-0 59.50 EUR

296 Jenny Lorentzen: Entwicklung und Evaluation eines Lernangebots im Lehramtsstudium Chemie zur Förderung von Vernetzungen innerhalb des fachbezogenen Professionswissens
ISBN 978-3-8325-5120-9 39.50 EUR

297 Micha Winkelmann: Lernprozesse in einem Schülerlabor unter Berücksichtigung individueller naturwissenschaftlicher Interessenstrukturen
ISBN 978-3-8325-5147-6 48.50 EUR

298 Carina Wöhlke: Entwicklung und Validierung eines Instruments zur Erfassung der professionellen Unterrichtswahrnehmung angehender Physiklehrkräfte
ISBN 978-3-8325-5149-0 43.00 EUR

299 Thomas Schubatzky: Das Amalgam Anfangs-Elektrizitätslehreunterricht. *Eine multiperspektivische Betrachtung in Deutschland und Österreich*
ISBN 978-3-8325-5159-9 50.50 EUR

300 Amany Annaggar: A Design Framework for Video Game-Based Gamification Elements to Assess Problem-solving Competence in Chemistry Education
ISBN 978-3-8325-5150-6 52.00 EUR

301 Alexander Engl: Chemie Pur – Unterrichten in der Natur: *Entwicklung und Evaluation eines kontextorientierten Unterrichtskonzepts im Bereich Outdoor Education zur Änderung der Einstellung zu „Chemie und Natur“*
ISBN 978-3-8325-5174-2 59.00 EUR

302 Christin Marie Sajons: Kognitive und motivationale Dynamik in Schülerlaboren. *Kontextualisierung, Problemorientierung und Autonomieunterstützung der didaktischen Struktur analysieren und weiterentwickeln*
ISBN 978-3-8325-5155-1 56.00 EUR

303 Philipp Bitzenbauer: Quantenoptik an Schulen. *Studie im Mixed-Methods Design zur Evaluation des Erlanger Unterrichtskonzepts zur Quantenoptik*
ISBN 978-3-8325-5123-0 59.00 EUR

304 Malte S. Ubben: Typisierung des Verständnisses mentaler Modelle mittels empirischer Datenerhebung am Beispiel der Quantenphysik
ISBN 978-3-8325-5181-0 43.50 EUR

305 Wiebke Kuske-Janßen: Sprachlicher Umgang mit Formeln von LehrerInnen im Physikunterricht am Beispiel des elektrischen Widerstandes in Klassenstufe 8
ISBN 978-3-8325-5183-4 47.50 EUR

306 Kai Bliesmer: Physik der Küste für außerschulische Lernorte. *Eine Didaktische Rekonstruktion*
ISBN 978-3-8325-5190-2 58.00 EUR

307 Nikola Schild: Eignung von domänenspezifischen Studieneingangsvariablen als Prädiktoren für Studienerfolg im Fach und Lehramt Physik
ISBN 978-3-8325-5226-8 42.00 EUR

308 Daniel Averbeck: Zum Studienerfolg in der Studieneingangsphase des Chemiestudiums. *Der Einfluss kognitiver und affektiv-motivationaler Variablen*
ISBN 978-3-8325-5227-5 51.00 EUR

309 Martina Strübe: Modelle und Experimente im Chemieunterricht. *Eine Videostudie zum fachspezifischen Lehrerwissen und -handeln*
ISBN 978-3-8325-5245-9 45.50 EUR

310 Wolfgang Becker: Auswirkungen unterschiedlicher experimenteller Repräsentationen auf den Kenntnisstand bei Grundschulkindern
ISBN 978-3-8325-5255-8 50.00 EUR

311 Marvin Rost: 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

312 Christina Kobl: Förderung und Erfassung der Reflexionskompetenz im Fach Chemie
ISBN 978-3-8325-5259-6 41.00 EUR

313 Ann-Kathrin Beretz: Diagnostische Prozesse von Studierenden des Lehramts – *eine Videostudie in den Fächern Physik und Mathematik*
ISBN 978-3-8325-5288-6 45.00 EUR

314 Judith Breuer: 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

315 Michaela Oettle: Modellierung des Fachwissens von Lehrkräften in der Teilchenphysik. *Eine Delphi-Studie*
ISBN 978-3-8325-5305-0 57.50 EUR

316 Volker Brüggemann: Entwicklung und Pilotierung eines adaptiven Multistage-Tests zur Kompetenzerfassung im Bereich naturwissenschaftlichen Denkens
ISBN 978-3-8325-5331-9 40.00 EUR

317 Stefan Müller: 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: Alltagsentscheidungen für den Chemieunterricht erkennen und Entscheidungsprozesse explorativ begleiten
ISBN 978-3-8325-5379-1 59.00 EUR

319 Lars Ehlert: Entwicklung und Evaluation einer Lehrkräftefortbildung zur Planung von selbstgesteuerten Experimenten
ISBN 978-3-8325-5393-71 41.50 EUR

320 Florian Seiler: Entwicklung und Evaluation eines Seminarkonzepts zur Förderung der experimentellen Planungskompetenz von Lehramtsstudierenden im Fach Chemie
ISBN 978-3-8325-5397-5 47.50 EUR

321 Nadine Boele: 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: Entwicklung und Evaluation digitalisierungsbezogener Kompetenzen von angehenden Chemielehrkräften
ISBN 978-3-8325-5410-1 49.50 EUR

323 Lars-Frederik Weiß: 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: Kumulatives Lehren und Lernen im Lehramtsstudium Physik. *Theorie und Evaluation eines Lehrkonzepts*
ISBN 978-3-8325-5421-7 51.00 EUR

325 Kübra Nur Celik: 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: 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: 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: Schülerlabore als interessefördernde authentische Lernorte für den naturwissenschaftlichen Unterricht nutzen
ISBN 978-3-8325-5454-5 50.50 EUR

329 Christopher Kurth: Die Kompetenz von Studierenden, Schülerschwierigkeiten beim eigenständigen Experimentieren zu diagnostizieren
ISBN 978-3-8325-5457-6 58.50 EUR

330 Dagmar Michna: 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: 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: 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

333 Erik Heine: 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

334 Simon Goertz: 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: 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

336 Alina Behrendt: 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

337 Manuel Daiber: Entwicklung eines Lehrkonzepts für eine elementare Quantenmechanik *Formuliert mit In-Out Symbolen*
ISBN 978-3-8325-5507-8 48.50 EUR

338 Felix Pawlak: Das Gemeinsame Experimentieren (an-)leiten *Eine qualitative Studie zum chemiespezifischen Classroom-Management*
ISBN 978-3-8325-5508-5 46.50 EUR

339 Liza Dopatka: Konzeption und Evaluation eines kontextstrukturierten Unterrichtskonzeptes für den Anfangs-Elektrizitätslehreunterricht
ISBN 978-3-8325-5514-6 69.50 EUR

340 Arne Bewersdorff: Untersuchung der Effektivität zweier Fortbildungsformate zum Experimentieren mit dem Fokus auf das Unterrichtshandeln
ISBN 978-3-8325-5522-1 39.00 EUR

341 Thomas Christoph Münster: Wie diagnostizieren Studierende des Lehramtes physikbezogene Lernprozesse von Schüler*innen? Eine Videostudie zur Mechanik
ISBN 978-3-8325-5534-4 44.50 EUR

342 Ines Komor: Förderung des symbolisch-mathematischen Modellverständnisses in der Physikalischen Chemie
ISBN 978-3-8325-5546-7 46.50 EUR

343 Verena Petermann: 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

344 Jana Heinze: Einfluss der sprachlichen Konzeption auf die Einschätzung der Qualität instruktionaler Unterrichtserklärungen im Fach Physik
ISBN 978-3-8325-5545-0 47.00 EUR

345 Jannis Weber: Mathematische Modellbildung und Videoanalyse zum Lernen der Newtonschen Dynamik im Vergleich
ISBN 978-3-8325-5566-5 68.00 EUR

346 Fabian Sterzing: 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

347 Lars Greitemann: 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: Diagnose experimenteller Kompetenzen in der laborpraktischen Chemielehrer*innenbildung
ISBN 978-3-8325-5587-0 48.00 EUR

349 William Lindlahr: 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: Teilnahmemotivation und situationales Interesse von Kindern und Eltern im experimentellen Lernsetting KEMIE
ISBN 978-3-8325-5598-6 43.00 EUR

351 Katharina Nave: Charakterisierung situativer mentaler Modellkomponenten in der Chemie und die Bildung von Hypothesen *Eine qualitative Studie zur Operationalisierung mentaler Modell-komponenten für den Fachbereich Chemie*
ISBN 978-3-8325-5599-3 43.00 EUR

352 Anna B. Bauer: Experimentelle Kompetenz Physikstudierender *Entwicklung und erste Erprobung eines performanzorientierten Kompetenzstrukturmodells unter Nutzung qualitativer Methoden*
ISBN 978-3-8325-5625-9 47.00 EUR

353 Jan Schröder: Entwicklung eines Performanztests zur Messung der Fähigkeit zur Unterrichtsplanung bei Lehramtsstudierenden im Fach Physik
ISBN 978-3-8325-5655-9 46.50 EUR

354 Susanne Gerlach: Aspekte einer Fachdidaktik Körperpflege *Ein Beitrag zur Standardentwicklung*
ISBN 978-3-8325-5659-4 45.00 EUR

355 Livia Murer: Diagnose experimenteller Kompetenzen beim praktisch-naturwissenschaftlichen Arbeiten *Vergleich verschiedener Methoden und kognitive Validierung eines Testverfahrens*
ISBN 978-3-8325-5657-0 41.50 EUR

356 Andrea Maria Schmid: Authentische Kontexte für MINT-Lernumgebungen *Eine zweiteilige Interventionsstudie in den Fachdidaktiken Physik und Technik*
ISBN 978-3-8325-5605-1 57.00 EUR

357 Julia Ortmann: Bedeutung und Förderung von Kompetenzen zum naturwissenschaftlichen Denken und Arbeiten in universitären Praktika
ISBN 978-3-8325-5670-9 37.00 EUR

358 Axel-Thilo Prokop: Entwicklung eines Lehr-Lern-Labors zum Thema Radioaktivität *Eine didaktische Rekonstruktion*
ISBN 978-3-8325-5671-6 49.50 EUR

359 Timo Hackemann: Textverständlichkeit sprachlich variierter physikbezogener Sachtexte
ISBN 978-3-8325-5675-4 41.50 EUR

360 Dennis Dietz: 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: Vielfalt im Physikunterricht *Zur Wirkung von Lehrkräftefortbildungen unter Diversitätsaspekten*
ISBN 978-3-8325-5672-3 65.50 EUR

362 Simon Kaulhausen: Strukturelle Ursachen für Klausurmisserfolg in Allgemeiner Chemie an der Universität
ISBN 978-3-8325-5699-0 37.50 EUR

363 Julia Eckoldt: 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

364 Albert Teichrew: Physikalische Modellbildung mit dynamischen Modellen
ISBN 978-3-8325-5710-2 58.50 EUR

365 Sascha Neff: Transfer digitaler Innovationen in die Schulpraxis *Eine explorative Untersuchung zur Förderung der Implementation*
ISBN 978-3-8325-5687-7 59.00 EUR

366 Rahel Schmid: 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

367 Dennis Kirstein: Individuelle Bedingungs- und Risikofaktoren für erfolgreiche Lernprozesse mit kooperativen Experimemtieraufgaben im Chemieunterrricht *Eine Untersuchung zum Zusammenhang von Lernvoraussetzungen, Lerntätigkeiten, Schwierigkeiten und Lernerfolg beim Experimentieren in Kleingruppen der Sekundarstufe I*
ISBN 978-3-8325-5729-4 52.50 EUR

368 Frauke Düwel: Argumentationslinien in Lehr-Lernkontexten *Potenziale englischer Fachtexte zur Chromatografie und deren hochschuldidaktische Einbindung*
ISBN 978-3-8325-5731-7 62.50 EUR

369 Fabien Güth: Interessenbasierte Differenzierung mithilfe systematisch variierter Kontextaufgaben im Fach Chemie
ISBN 978-3-8325-5737-9 48.00 EUR

370 Oliver Grewe: 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

371 Anna Nowak: Untersuchung der Qualität von Selbstreflexionstexten zum Physikunterricht *Entwicklung des Reflexionsmodells REIZ*
ISBN 978-3-8325-5739-3 59.00 EUR

372 Dominique Angela Holland: 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

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.

Die Berücksichtigung individueller Lernvoraussetzungen gilt als Qualitätsmerkmal von Unterricht, wobei die chemiedidaktische Forschung bislang vor allem leistungsbezogene Differenzierung fokussiert. Interessenbasierte Differenzierung wird in der Literatur als die Öffnung des Unterrichts im Hinblick auf die Auswahl zwischen bereitgestellten Lernmaterialien, die in unterschiedliche Kontexte eingebettet sind, beschrieben – aber bisher nur wenig evaluiert.

Im Rahmen dieser Arbeit wurden zwei quantitative Studien mit Chemielernenden der 9. und 10. Jahrgangsstufe durchgeführt, um die Effektivität eines interessenbasierten Differenzierungsansatzes im Fach Chemie zu untersuchen. Die Kontextwahl der Lernenden wurde mithilfe einer explanativen Studie mit $N = 347$ Lernenden ermittelt. Daran anknüpfend wurde in einer experimentellen Studie mit $N = 217$ Lernenden untersucht, welchen Einfluss selbstgewählte und systematisch zugewiesene Kontextaufgaben auf die Effektivität einer Lernsituation im Fach Chemie haben.

Die Ergebnisse der Untersuchung zeigen, dass Lernende in Abhängigkeit von ihren Personenmerkmalen alltägliche, besondere oder innerfachliche Kontexte wählen und vor allem die Passung zwischen den Merkmalen der Lernenden und dem Kontext zu lernförderlichen Effekten hinsichtlich affektiv-motivationaler Merkmale führt. Die differenzierte Betrachtung der Merkmale der Lernenden und der Merkmale der verwendeten Kontexte ist notwendig, um den Lernprozess für alle Lernenden zu optimieren.

Logos Verlag Berlin

ISBN 978-3-8325-5737-9

DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

ub | universitäts bibliothek

Dieser Text wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt. Die hier veröffentlichte Version der E-Publikation kann von einer eventuell ebenfalls veröffentlichten Verlagsversion abweichen.

DOI: 10.30819/5737
URN: urn:nbn:de:hbz:465-20260119-162009-1

Dieses Werk kann unter einer Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 Lizenz (CC BY-NC-ND 4.0) genutzt werden.