

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Veränderung von Präkonzepten und Designfehlern bei der
Variablenkontrollstrategie durch Lernen aus Fehlern.

verfasst von | submitted by

Julia Maria Hrabe BEd

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Education (MEd)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on the
student record sheet:

UA 199 502 506 02

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Lehramt Sek (AB) Unterrichtsfach
Biologie und Umweltbildung Unterrichtsfach
Deutsch

Betreut von | Supervisor:

Univ.-Prof. Dr. Andrea Möller

Abstract

Experimentieren als grundlegende wissenschaftliche Arbeitsweise findet in Lehrplänen, insbesondere den neuen Lehrplänen der Sekundarstufe I große Bedeutung. Die Variablenkontrollstrategie (VKS) ist dabei eine wichtige experimentelle Strategie. Viele Schüler*innen treffen jedoch auf Herausforderungen bei der Anwendung der VKS, die mit zugrundeliegenden Schüler*innenvorstellungen einhergehen. Bisher konnten acht Schüler*innenvorstellungen identifiziert werden, die zu fehlerhaften Experimentdesigns, den sogenannten Designfehlern, führen. Die vorliegende Masterarbeit untersucht diese Designfehler einerseits hinsichtlich der vier Subkompetenzen der VKS und andererseits wird der Lerneffekt durch eine Intervention zwischen zwei Testzeitpunkten und die Auswirkungen von Lernen aus eigenen und Lernen aus fremden Fehlern auf die Designfehler untersucht. Die VKS-Kompetenz von Sekundarschüler*innen wurde erstmals Subkompetenz-spezifisch sowie domänenspezifisch untersucht. Die meisten Schwierigkeiten haben Schüler*innen bei der Erkennung von fehlender Aussagekraft eines konfundierten Experiments sowie in der Domäne Chemie bei der Planung von Experimenten und der Identifikation der unabhängigen Variable. Weiters konnte ein Lerneffekt durch die Intervention festgestellt werden, jedoch kein Gruppeneffekt. Es werden mögliche Implikationen für den Unterricht und weitere Forschungsmöglichkeiten diskutiert.

Experimentation as a fundamental scientific method is of great importance in curricula, especially the new curricula for lower secondary level. The variable control strategy (CVS) is an important experimental strategy. However, many students encounter challenges when applying the VKS that are associated with underlying student conceptions. So far, eight student preconceptions have been identified that lead to incorrect experiment designs and design errors. This master's thesis examines these design errors with regard to the four sub-competencies of CVS and also investigates the learning effect of an intervention between two test times and the effects of learning from one's own and learning from others' errors on the design errors. The CVS competence of secondary school students was examined for the first time on a subcompetence-specific and domain-specific basis. Students have the most difficulties in recognizing the lack of significance of a confounded experiment and in the domain of chemistry when planning experiments and identifying the independent variable. Furthermore, a learning effect was found as a result of the intervention, but no group effect. Possible implications for teaching and further research opportunities are discussed.

Inhaltsverzeichnis

1	Einleitung	6
2	Experimentierkompetenz und die Variablenkontrollstrategie	7
2.1	<i>Experimentierkompetenz in den neuen Lehrplänen der Sekundarstufe I</i>	7
2.2	<i>Variablenkontrollstrategie</i>	10
2.2.1	<i>Theoretischer Hintergrund</i>	10
2.2.2	<i>Herausforderungen beim Experimentieren und der Variablenkontrollstrategie</i> 10	
3	Schüler*innenvorstellungen	12
3.1	<i>Theoretischer Hintergrund</i>	12
3.2	<i>Bekannte Schüler*innenvorstellungen beim Experimentieren</i>	14
4	Lernen aus Fehlern	20
4.1	<i>Theoretischer Hintergrund</i>	20
4.2	<i>Aktuelle Forschung zu Lernen aus Fehlern</i>	22
5	Lernen aus Fehlern im Kontext des Experimentierens	24
6	Fragestellungen und Hypothesen	26
7	Material & Methoden	27
7.1	<i>Teilnehmer*innen</i>	27
7.2	<i>Studiendesign</i>	28
7.3	<i>Ablauf der Intervention</i>	29
7.4	<i>Testinstrument</i>	30
7.4.1	<i>VKS-Testheft</i>	30
7.4.2	<i>Messung der Designfehler</i>	35
7.5	<i>Datenbereinigung und Kodierung</i>	36
7.6	<i>Statistische Analyse</i>	37
8	Ergebnisse	38
8.1	<i>Häufigkeit von Designfehlern</i>	38
8.2	<i>Muster von Designfehlern</i>	42
8.2.1	<i>Fehlerwahrscheinlichkeit in Abhängigkeit der Subkompetenzen</i>	42
8.2.2	<i>Fehlerwahrscheinlichkeit in Abhängigkeit der Fachkontexte</i>	45

8.2.3 Fehlerwahrscheinlichkeit in Abhängigkeit der Wechselwirkung der Fachkontexte und Subkompetenzen	48
8.3 Veränderung der Fehlerwahrscheinlichkeit durch Lernen aus Fehlern	51
9 Diskussion	54
9.1 Designfehlerhäufigkeit in den Subkompetenzen ID, PL, IN und UN	54
9.2 Muster von Designfehlern	56
9.2.1 Fehlerwahrscheinlichkeit in Abhängigkeit der Subkompetenzen.....	56
9.2.2 Fehlerwahrscheinlichkeit in Abhängigkeit der Fachkontexte	57
9.2.3 Fehlerwahrscheinlichkeit in Abhängigkeit der Wechselwirkung der Fachkontexte und Subkompetenzen	57
9.3 Veränderung der Fehlerwahrscheinlichkeit durch Lernen aus Fehlern	60
10 Implikationen für den Unterricht	61
11 Limitationen.....	62
12 Zusammenfassung.....	63
13 Literaturverzeichnis.....	64
14 Danksagung	68

Tabellenverzeichnis

<i>Tabelle 1 Übersicht über die Schüler*innenvorstellungen, Designfehler und die Auswirkungen auf die Anwendungen der VKS</i>	16
<i>Tabelle 2 2x2 - Faktoren Versuchsdesign nach Hämmerle (in prep.)</i>	28
<i>Tabelle 3 Übersicht über die Themen je Testheft und Fachkontext</i>	35
<i>Tabelle 4 Fachkontexte (Biologie, Physik, Chemie, Alltagswissen) in den drei Testheften</i>	35
<i>Tabelle 5 mögliche Designfehler je Subkompetenz</i>	36
<i>Tabelle 6 Fehlerhäufigkeit in der Subkompetenz ID</i>	38
<i>Tabelle 7 Fehlerhäufigkeit in der Subkompetenz IN</i>	39
<i>Tabelle 8 Fehlerhäufigkeit in der Subkompetenz UN</i>	40
<i>Tabelle 9 Fehlerhäufigkeit in der Subkompetenz PL</i>	41
<i>Tabelle 10 Ergebnisse des gemischten Modells für die Fehlerwahrscheinlichkeit</i>	42
<i>Tabelle 11 Post hoc Analyse mit paarweisen Vergleichen</i>	43
<i>Tabelle 12 Sensitivitätsanalyse mit Annahme alle fehlenden Werte falsch</i>	44
<i>Tabelle 13 Sensitivitätsanalyse mit 20% Zufallswahrscheinlichkeit</i>	44

<i>Tabelle 14 Ergebnisse des gemischten Modells für die Fehlerwahrscheinlichkeit</i>	45
<i>Tabelle 15 Post hoc Analysen mit paarweisen Vergleichen</i>	46
<i>Tabelle 16 Post hoc Analysen mit paarweisen Vergleichen (alle auf null)</i>	47
<i>Tabelle 17 Post hoc Analysen mit paarweisen Vergleichen (inkl. Zufall)</i>	47
<i>Tabelle 18 Gemischtes lineares Modell für die Fehlerwahrscheinlichkeit unter Hinzunahme der Wechselwirkung von Fachkontexte und Subkompetenz</i>	48
<i>Tabelle 19 Post hoc Analysen mit paarweisen Vergleichen</i>	48
<i>Tabelle 20 Ergebnisse des gemischten Modells für die Fehlerwahrscheinlichkeit je Zeitpunkt</i>	52
<i>Tabelle 21 Lerneffekt je Subkompetenz</i>	53

Abbildungsverzeichnis

<i>Abbildung 1 Rahmenmodell für wissenschaftsmethodische Kompetenzen, verändert nach Mayer 2007, 178.</i>	7
<i>Abbildung 2 Struktur des neuen Lehrplans für die Sekundarstufe I</i>	9
<i>Abbildung 3 ID-Aufgabe des gekürzten Testheftes (Hämmerle in prep.)</i>	31
<i>Abbildung 4 PL-Aufgabe des gekürzten Testheftes (Hämmerle in prep.)</i>	32
<i>Abbildung 5 IN-Aufgabe des gekürzten Testheftes (Hämmerle in prep.)</i>	33
<i>Abbildung 6 UN-Aufgabe des gekürzten Testheftes (Hämmerle in prep.)</i>	34
<i>Abbildung 7 Fehlerwahrscheinlichkeit in Abhängigkeit der Subkompetenz</i>	43
<i>Abbildung 8 Fehlerwahrscheinlichkeit in Abhängigkeit des Fachkontexts</i>	46
<i>Abbildung 9 Fehlerwahrscheinlichkeit je Fachkontext in Abhängigkeit der Subkompetenz</i>	50
<i>Abbildung 10 fehlende Werte je Fachkontext</i>	50
<i>Abbildung 11 fehlende Werte je Subkompetenz</i>	51
<i>Abbildung 12 Veränderung der Fehlerwahrscheinlichkeit zwischen Prä- und Posttest</i>	52
<i>Abbildung 13 Veränderung der Fehlerwahrscheinlichkeit in Abhängigkeit der Subkompetenz zwischen Prä- und Posttest</i>	54

1 Einleitung

Experimentieren ist eine grundlegende wissenschaftliche Arbeitsweise. Die Kompetenzvermittlung der Erkenntnisgewinnung stellt ein zentrales Ziel naturwissenschaftlicher Grundbildung (scientific literacy) dar (Mostafa et al. 2018). Die Fähigkeit, wissenschaftliche Experimente zu entwerfen und durchzuführen, ermöglicht es den Lernenden, die Welt kritisch zu hinterfragen und zu verstehen. In der Sekundarstufe I ist die Durchführung von einfachen Experimenten daher fester Bestandteil des Lehrplans von der fünften bis zur achten Schulstufe (Lehrplan 2023).

Die Variablenkontrollstrategie (VKS) stellt dabei die wichtigste Grundlage der experimentellen Erkenntnisgewinnung dar (Chen & Klahr 1999). Schüler*innen treffen jedoch auf Herausforderungen bezüglich der Anwendung der VKS (Kranz et al. 2022). Die Schwierigkeiten der Schüler*innen sind einerseits durch zugrundeliegende Schüler*innenvorstellungen hinsichtlich der VKS bedingt und andererseits sind die Schwierigkeiten auch abhängig von den vier Subkompetenzen (Identification, Interpretation, Understanding & Planning) der VKS (Schwichow 2022). Schüler*innenvorstellungen führen zu Designfehlern beim Experimentieren, die ein konfundiertes Experiment zur Folge haben. Fehler werden als Abweichung von der Norm definiert (Oser & Spychiger 2005), wobei die Norm in diesem Kontext die korrekte Anwendung der VKS ist. Fehler sind wichtiger Bestandteil des Lernprozesses und führen zum Aufbau von negativem Wissen – dem Wissen darüber, was in Zukunft zu vermeiden ist (Gartmeier et al. 2008). Die Herausforderungen mit der VKS und die damit einhergehenden (Design)fehler können als Lernchancen gesehen werden (Oser & Spychiger 2005). Darüber hinaus sind Fehler für Diagnosezwecke gut geeignet (vgl. Baur 2018). Die vier Subkompetenzen der VKS sind bislang wenig differenziert erforscht, jedoch scheinen die Schwierigkeiten mit der VKS vor allem durch die Subkompetenzen bedingt und weniger abhängig von domänenspezifischem Fachwissen zu sein. Fachwissen scheint demnach eine untergeordnete Rolle im Hinblick auf die VKS-Fähigkeit zu spielen, wobei dies in der Forschung bislang wenig Aufmerksamkeit fand und daher keine eindeutige Aussage getroffen werden kann (Schwichow 2022).

In dieser Masterarbeit wird die Bedeutung von Designfehlern und Präkonzepten im Kontext der Variablenkontrollstrategie eingehend hinsichtlich der vier Subkompetenzen untersucht. Es wird analysiert, wie solche Fehler nicht nur die Experimentierkompetenz von Schüler*innen beeinflussen, sondern auch als Diagnoseinstrument für Lehrkräfte dienen können, um den Lernfortschritt zu unterstützen. Darüber hinaus wird der Lerneffekt zwischen zwei Zeitpunkten im Zuge einer Intervention untersucht, um den Einfluss von Interventionen auf die Entwicklung der Experimentierkompetenz zu bewerten.

Das Ziel dieser Arbeit ist es, ein tieferes Verständnis für die Rolle von Designfehlern und Schüler*innenvorstellungen bei der Anwendung der VKS zu entwickeln. Durch die differenzierte Betrachtung von domänenspezifischem Wissen und dessen potenziellen

Einfluss auf Designfehler soll ein Beitrag zur gezielteren Förderung der naturwissenschaftlichen Bildung geleistet werden. Die Analyse des Lernzuwachses durch Interventionen, insbesondere durch das Lernen aus eigenen bzw. fremden Fehlern, soll neue Einblicke in effektive Strategien zur Förderung der Experimentierkompetenz von Schüler*innen bieten.

2 Experimentierkompetenz und die Variablenkontrollstrategie

2.1 Experimentierkompetenz in den neuen Lehrplänen der Sekundarstufe I

Die naturwissenschaftliche Grundbildung (scientific literacy) wird unter anderem durch die Kompetenzvermittlung der Erkenntnisgewinnung ausgezeichnet. Mayer (2007) beschreibt in seinem Rahmenmodell für die naturwissenschaftliche Erkenntnisgewinnung (vgl. Abbildung 1) drei Dimensionen: praktische Arbeitstechniken (practical work), wissenschaftliche Erkenntnismethoden (scientific inquiry) und Charakteristika der Naturwissenschaften (nature of science). Manuelle Fertigkeiten (practical skills), wissenschaftliches Denken (scientific reasoning) und Wissenschaftsverständnis (epistemological beliefs) können als kognitionspsychologische Konstrukte mit den Dimensionen der Erkenntnisgewinnung in Verbindung gesetzt werden (Mayer 2007).

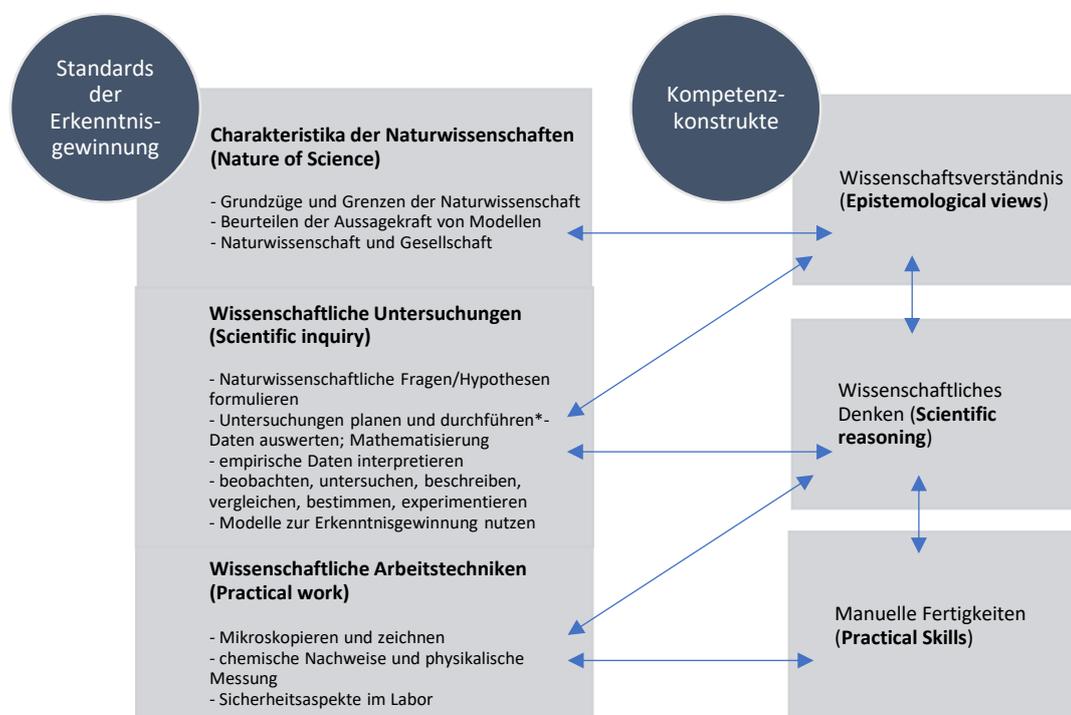


Abbildung 1 Rahmenmodell für wissenschaftsmethodische Kompetenzen, verändert nach Mayer 2007, 178.

Wissenschaftliches Denken umfasst mehrere Schlüsselkomponenten, wie etwa das Verständnis für experimentelle Methoden, die Interpretation von Ergebnissen und den allgemeinen Erkenntnisprozess (Koerber et al. 2011). Der Prozess der Erkenntnisgewinnung wird als anspruchsvoller Vorgang beschrieben, der auf Wissen basiert und durch bestimmte Verfahrensweisen geprägt ist (Mayer 2007). Die Teilkompetenzen der Erkenntnisgewinnung, nämlich Formulieren von Fragen, Generieren von Hypothesen, das Planen eines Experiments oder die Interpretation der Ergebnisse stellen dabei zentrale Punkte dar (Mayer 2007).

In Österreich treten in der Sekundarstufe I mit dem Schuljahr 2023/24 stufenweise, das heißt zunächst ausschließlich beginnend in der 5. Schulstufe, neue Lehrpläne in Kraft. Diese sind deutlich kompetenzorientiert und zielen von der 5. bis zur 8. Schulstufe darauf ab, Schüler*innen unter anderem im Fach „Biologie und Umweltbildung“ im Kompetenzbereich „Erkenntnisse gewinnen“ zu stärken. Der Fokus wird somit in diesem Kompetenzbereich vor allem auf die Experimentierfähigkeit von Schüler*innen gelegt. In der internationalen naturwissenschaftlichen Forschung wird das Konzept der Erkenntnisgewinnung häufig als „Scientific Inquiry“ bezeichnet (Mayer 2007).

Experimentieren war zwar bereits zuvor fester Bestandteil der Lehrpläne, durch die neuen Lehrpläne, die gänzlich im Zeichen der Kompetenzorientierung stehen, wurde jedoch unter anderem die Experimentierkompetenz noch deutlich weiter in den Fokus gerückt. Im Lehrplan (2023) wird bei der Kompetenzorientierung, die im Zentrum der pädagogischen Überlegungen steht, in „fachliche, überfachliche und fächerübergreifende Kompetenzen unterschieden“ (Lehrplan 2023, S. 5). Die fachlichen Kompetenzen stehen dabei direkt mit dem jeweiligen Unterrichtsgegenstand in Verbindung. Die überfachlichen Kompetenzen schließen „insbesondere Motivation, Selbstwahrnehmung und Vertrauen in die eigene Person, soziale Kompetenzen und lernmethodische Kompetenzen“ (Lehrplan 2023, S. 5) mit ein. Mittels fächerübergreifender Kompetenzen soll die Fähigkeit der Auseinandersetzung mit übergreifenden Themen erworben werden. Eine Übersicht über die Struktur des neuen Lehrplans ist auch in Abbildung 2 zu finden.

Bevor die „neuen“ Kompetenzen im Lehrplan diskutiert werden, ist es essenziell, den Begriff „Kompetenz“ und die Struktur der neuen Fachlehrpläne zu klären. Kompetenz bezeichnet die Fähigkeit, situationsgerecht zu handeln (Forst-Heinzer & Oser 2015) und dient der Definition und Differenzierung von Bildungszielen in Lehrplänen. Der Begriff gewann durch PISA-Tests an Bedeutung, da er zur Leistungsmessung von Schüler*innen herangezogen wird (Hartig & Klieme 2006). Weinert (2014) erweitert dies um kognitive, motivationale, volitionale und soziale Fähigkeiten für problemorientiertes Handeln in diversen Kontexten. Schulische Kompetenzen umfassen fachliche, fachübergreifende und Handlungskonzepte, welche auch für PISA-Studien relevant sind (Weinert 2014).

Die neuen spezifischen Fachlehrpläne sind inhärent aufgebaut. Jeder Fachlehrplan besteht aus der jeweiligen Bildungs- und Lehraufgabe, den didaktischen Grundsätzen, den zentralen fachlichen Konzepten, einem fachspezifischen Kompetenzmodell inklusive Kompetenzbereiche sowie geeigneten Kompetenzbeschreibungen (Lehrplan 2023).

Die Kompetenzbeschreibungen der neuen Lehrpläne der Mittelschule setzen sich aus den drei Kompetenzbereichen W (Wissen aneignen, anwenden und kommunizieren), E (Selbstständiger Erkenntnisgewinn) und S (Begründung von Standpunkten) zusammen. Da das Experimentieren als grundlegende wissenschaftliche Arbeitsweise in den neuen Lehrplänen dem Kompetenzbereich E (Erkenntnisse gewinnen) zugeordnet werden kann, ist vor allem dieser Bereich für diese Masterarbeit relevant. Erkenntnisse sollen laut Lehrplan durch eigene Beobachtungen oder durch Experimentieren gewonnen werden (Lehrplan 2023). Konkret sollen Schüler*innen im Kompetenzbereich E „Lebewesen und biologische Phänomene betrachten, beobachten, bestimmen, kriteriengeleitet vergleichen und ordnen, mikroskopieren, zeichnen und messen [können] (Lehrplan 2023)“.

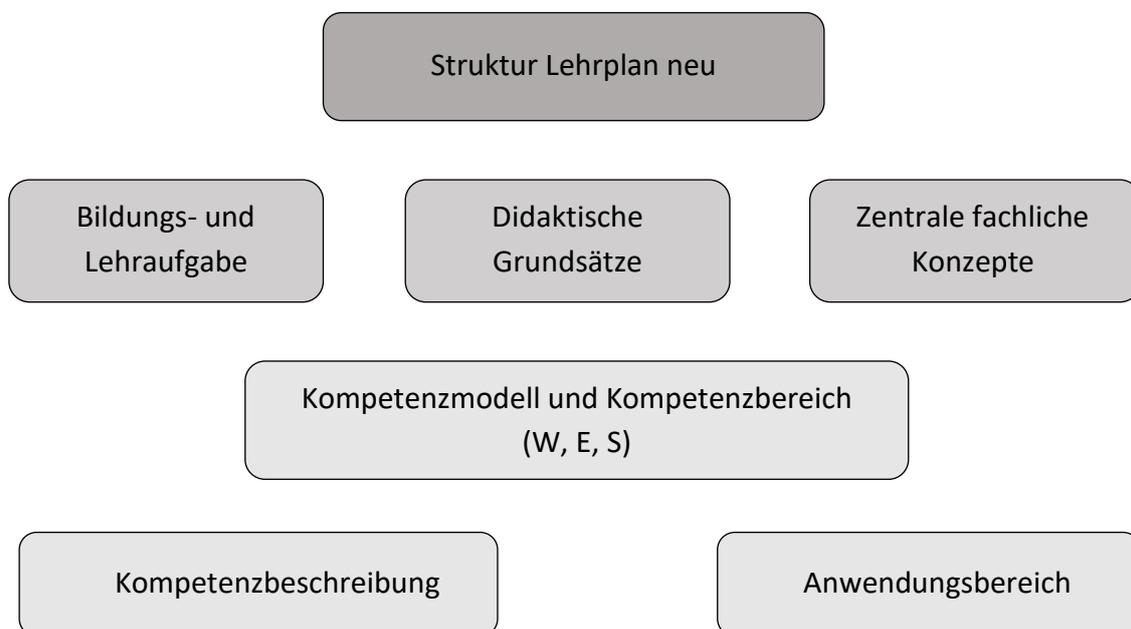


Abbildung 2 Struktur des neuen Lehrplans für die Sekundarstufe I

Die Kompetenzbereiche können im Kompetenzmodell von Mayer (2007) ebenso verankert werden. Die Operatoren sind der Dimension wissenschaftliche Arbeitstechniken nach Mayer (2007) zuzuordnen. Die Kompetenzen, „zu biologischen Vorgängen und Phänomene naturwissenschaftliche Fragen stellen sowie Hypothesen entwickeln und formulieren [zu können] (Lehrplan 2023)“ sowie „Beobachtungen, Versuche, Untersuchungen und Experimente zu naturwissenschaftlichen Fragestellungen planen, durchführen und protokollieren [zu können] (Lehrplan 2023)“, fallen nach Mayer (2007) unter die Dimension Scientific Inquiry, ebenso wie die Fähigkeit von Schüler*innen, „Daten und Ergebnisse von Untersuchungen, Beobachtungen und Experimenten darstellen, analysieren und interpretieren [zu können] (Lehrplan 2023)“.

2.2 Variablenkontrollstrategie

2.2.1 Theoretischer Hintergrund

Experimentieren dient dazu, kausale Zusammenhänge zu untersuchen und sowohl erwartete als auch unerwartete Daten zu sammeln, um entweder eine neue Hypothese auf Basis der gewonnenen Daten zu entwickeln oder die Stichhaltigkeit einer bereits existierenden Hypothese zu überprüfen. Eine Möglichkeit, ein kontrolliertes Experiment durchzuführen, stellt die Variablenkontrollstrategie (VKS) dar. Zudem spielt die Logik kontrollierter Experimente eine wichtige Rolle in der Argumentation und der Betrachtung von Kausalzusammenhängen, in der Wissenschaft ebenso wie im täglichen Leben. Dies umfasst das Verständnis für die fehlende Beweiskraft bei konfundierten Experimenten und gleichermaßen die Wichtigkeit von kontrollierten Experimenten (Kuhn 2005).

Um ein kontrolliertes Experiment durchführen zu können, aus welchem eindeutige Schlussfolgerungen über die Ursache-Wirkungs-Beziehung gezogen werden können, darf nur eine Variable verändert werden, während alle anderen konstant gehalten werden. Die abhängige Variable dient dabei als Messgröße und die unabhängige Variable als Einflussgröße. Durch die Konstanthaltung aller übrigen Variablen kann der Einfluss der unabhängigen auf die abhängige Variable systematisch untersucht werden (Chen & Klahr 1999; Schwichow et al. 2016). Folgende vier Schritte beschreiben das Design eines kontrollierten Experiments: (1) Bestimmung einer Variable, deren kausale Wirkung auf eine abhängige Variable ermittelt werden soll, (2) Festlegung von mindestens zwei Versuchsbedingungen, die sich hinsichtlich dieser Variable unterscheiden, (3) konstant halten von allen anderen Variablen und (4) Durchführen des Experiments bei Beobachtung der Unterschiede zwischen der abhängigen Variable und den Versuchsbedingungen (Siler & Klahr 2012).

Ein vollständiges Verständnis der Variablenkontrollstrategie (VKS) umfasst in weiterer Folge vier Subkompetenzen: Planen und Erstellen von kontrollierten Experimenten (PL), Identifizieren von kontrollierten und konfundierten Experimenten (ID), Interpretation von Ergebnissen kontrollierter Experimente (IN) sowie das Verständnis über die fehlende Aussagekraft eines konfundierten Experiments (UN) (Chen & Klahr 1999). Ein Experiment gilt dann als „konfundiert“, wenn mehrere Variablen variiert werden, wodurch kein Rückschluss über den Einfluss von einzelnen Variablen auf die abhängige Variable getroffen werden kann (Chen & Klahr 1999). Die detaillierte Betrachtung der Lehrpläne und ihrer Ausrichtung auf Kompetenzorientierung hebt die Bedeutung hervor, ein umfassendes Spektrum an Fähigkeiten und Wissen zu vermitteln, um Schüler*innen auf die vielfältigen Herausforderungen der modernen Welt vorzubereiten. Schüler*innen stoßen jedoch beim Experimentieren auf zahlreiche Herausforderungen (Kranz et al. 2022).

2.2.2 Herausforderungen beim Experimentieren und der Variablenkontrollstrategie

Schwierigkeiten, die Schüler*innen bei der Anwendung der VKS erleben, können auf viele Faktoren zurückzuführen sein: (1) Schwierigkeiten beim Verständnis des Aufgabenziels, was möglicherweise mit Problemen beim Erfassen abstrakter Konzepte und Verfahren

zusammenhängt, (2) Vernachlässigung der Variablenebene bei der Kausalitätsprüfung, (3) mangelndes Verständnis für die Notwendigkeit, Variablen über verschiedene Bedingungen hinweg zu vergleichen und abzugrenzen, (4) fehlendes Bewusstsein darüber, dass in einem Experiment nur die Kausalität einer einzigen Variable überprüft werden kann (Siler & Klahr 2012).

Baur (2018) gibt im Zuge seiner Untersuchung eine Übersicht über bereits erforschte und beschriebene Schwierigkeiten von Schüler*innen beim Experimentieren, gegliedert in die Teilkompetenzen des Experimentierens, bestehend aus (1) Naturwissenschaftliche Fragen formulieren, (2) Hypothesen generieren, (3) Untersuchungen planen und durchführen sowie (4) Daten analysieren und interpretieren. Die Variablenkontrollstrategie, die den Haupt-Untersuchungsgegenstand der Masterarbeit bildet, fällt unter die Teilkompetenz (3) Untersuchungen planen und durchführen.

Viele Schüler*innen arbeiten beim Experimentieren nur mit einem Versuch, verwenden also keinen Kontroll- oder Vergleichsansatz (Baur 2018, Hammann et al 2006, Kirchner & Priemer 2007, Chen und Klahr 1999). Einen weiteren limitierenden Faktor der Experimentierfähigkeit stellt die Variation von mehreren Variablen gleichzeitig dar (Baur 2018, Hammann et al 2006, Kuhn & Dean 2005, Chen und Klahr 1999). Manche Schüler*innen arbeiten nicht nach wissenschaftlichen Prinzipien, sondern probieren verschiedene Ansätze aus, in der Hoffnung, das gewünschte Ergebnis zu erhalten (Meier & Mayer 2012; Hamman et al. 2008). Besteht ein Experiment aus mehreren unabhängigen Variablen, so beschränken sich Schüler*innen oftmals auf einige wenige (Baur 2018, Hammann et al 2008). Ebenso können manche Schüler*innen nicht auf Messungenauigkeiten reagieren, da sie keine Messwiederholungen durchführen (Arnold et al. 2013). Manche Schüler*innen weisen zudem Schwierigkeiten in der Verwendung von Laborgeräten auf (Kechel 2016).

Ein ausführliches Review zu den Schwierigkeiten bei den Experimentierphasen ist bei Kranz et al. (2022) zu finden. Insgesamt konnten 43 Schwierigkeiten gefunden werden. In dem Review von Kranz et al. (2022) konnten vier Überkategorien von Problemen hinsichtlich des Experimentierens und der Variablenkontrollstrategie gebildet werden. (1) Oft fehlt Lernenden ein konzeptionelles Wissen über die wissenschaftlichen Grundlagen von Experimenten, was zu Herausforderungen wie dem Nichtverstehen von Hypothesen oder dem Verwechseln von Beobachtung und Interpretation führt. (2) Ein Mangel an epistemischem Wissen, also dem Verständnis, wie wissenschaftliches Wissen generiert wird, führt zu Schwierigkeiten im Verstehen von Experimenten und der Entwicklung von Erklärungen. Herausforderungen umfassen das Fehlen von Wiederholungen bei Messungen oder das Nichtverknüpfen von Hypothesen mit früheren Ergebnissen. (3) Einige Herausforderungen zeigen, dass Schüler*innen nicht wissen, wie man wissenschaftliche Untersuchungen durchführt, was auf einen Mangel an prozeduralen Fähigkeiten hinweist. Dies betrifft den praktischen Umgang mit Experimenten, wobei das Fehlen bestimmter prozeduraler Fertigkeiten beobachtet wird. (4) Kognitive Vorurteile und Präkonzepte beeinflussen alle drei Bereiche und führen oft dazu, dass mehrere Phasen beim Experimentieren betroffen sind (Kranz et al. 2022). Manche schüler*innenspezifischen Schwierigkeiten stehen miteinander in Verbindung. Dies führt

dazu, dass Schüler*innen, die eine der identifizierten Schwierigkeiten erleben, oftmals auch andere Probleme mit der VKS haben. Diese Beziehung ist jedoch nicht kausal, es kann also nicht davon ausgegangen werden, dass Schüler*innen, die eine Schwierigkeit überwunden haben, automatisch auch die restlichen lösen (Baur 2023). In einer Longitudinalstudie von Bullock et al. (2009) wurde bekannt, dass nicht nur Schüler*innen Schwierigkeiten mit der Variablenkontrollstrategie haben, sondern, dass sich jene Schwierigkeiten auch von der Kindheit bis ins Erwachsenenalter ziehen (Bullock et al. 2009).

Während das Erreichen der Lehrziele und somit die Förderung der Experimentierkompetenz bei Schüler*innen im Vordergrund steht, ist es unerlässlich zu erkennen, dass Schüler*innen nicht als unbeschriebene Blätter in den Bildungsprozess eintreten. Sie bringen ihre eigenen Vorstellungen und Erfahrungen mit, die einen erheblichen Einfluss darauf haben, wie sie neues Wissen aufnehmen und verarbeiten. Diese Vorstellungen gilt es zu kennen und erkennen, um einen zielführenden Unterricht gewährleisten zu können.

3 Schüler*innenvorstellungen

3.1 Theoretischer Hintergrund

In den Anfängen der Vorstellungsforschung wurden Schüler*innenvorstellungen, die von etablierten wissenschaftlichen Konzepten abwichen, als Fehl- oder Falschvorstellungen betitelt. Jene galt es zu korrigieren und zu ersetzen. In den 1980er Jahren geriet diese reduktionistische Perspektive jedoch in Kritik. Viele Vorstellungen basieren auf erfahrungsgestützten Interpretationen, die es Schüler*innen ermöglichen, natürliche Phänomene zu konzeptualisieren. Der Begriff wurde schlussendlich durch Schüler*innenvorstellungen oder Präkonzepte ersetzt (Gropengießer & Marohn, 2018). Wandersee et al. (1994) prägten den Begriff der Schüler*innenvorstellungen. Demnach würden Schüler*innen mit Ideen und Vorstellungen in den naturwissenschaftlichen Unterricht kommen, die nicht mit anerkannten wissenschaftlichen Erkenntnissen übereinstimmen (Wandersee et al. 1994).

Hinsichtlich der Zusammensetzung von Schüler*innenvorstellungen herrschen drei Ansichten vor. Der klassische Ansatz geht davon aus, dass Schüler*innenvorstellungen oftmals älteren, überholten bzw. widerlegten wissenschaftlichen Ansichten ähneln. Der zweite Ansatz „Lückenhaftes Wissen“ (original: „knowledge in pieces“) besagt, dass sich Schüler*innenwissen aus einer Vielzahl von Prinzipien zusammensetzt, die allerdings von Erfahrungen abgeleitet wurden und nicht auf wissenschaftlichen Theorien basieren. Schließlich geht die „framework theory“ davon aus, dass Schüler*innenvorstellungen aus einer Sammlung von Überzeugungen und Voraussetzungen bestehen, die in lose Rahmenwerktheorien organisiert sind (Vosniadou 2019). Schüler*innenvorstellungen drücken deren subjektive und themenspezifischen Gedankenkonstrukte aus. Zentral hierfür ist der gedankliche Bereich, der den referenziellen und sprachlichen Bereich verbindet

(Gropengießer 2007). Wandersee (1994) unternahm den Versuch, die Erkenntnisse der Schüler*innenvorstellungen im naturwissenschaftlichen Bereich zusammenzufassen. Dabei gliederten sich acht Thesen heraus: (1) Schüler*innen bringen vielfältige alternative Vorstellungen in den naturwissenschaftlichen Unterricht, (2) Schüler*innen haben alternative Vorstellungen ganz ungeachtet ihres Alters, Geschlechts, Kenntnissen oder Kultur, (3) Schüler*innenvorstellungen sind gegenüber üblichen Lehrstrategien resistent, (4) Schüler*innenvorstellungen sind oft äquivalent zu Vorstellungen, die in früheren Generationen geprägt wurden, (5) Schüler*innenvorstellungen haben ihren Ursprung oft in persönlichen Erfahrungen, (6) Lehrpersonen haben oftmals dieselben Vorstellungen wie ihre Schüler*innen, (7) Die Vorstellungen von Lernenden stehen im Gegensatz zum tatsächlichen unterrichteten Wissen. Diese Auseinandersetzung führt zu einem unbewussten Lernfortschritt, (8) Lehransätze, die zu einer Vorstellungsänderung führen, können potenziell effektive Lehrmethoden sein (Wandersee 1994).

Schüler*innenvorstellungen können grundlegend in zwei verschiedene Bereiche unterteilt werden. Einerseits in domänenspezifische Vorstellungen und andererseits in Vorstellungen, bezogen auf domänenübergreifende wissenschaftliche Verfahren (Siler & Klahr 2012). Schüler*innen können in ihren experimentellen Erfahrungen zwei Arten von Zielen verfolgen, nämlich entweder „Wissenschaftsziele“ oder „Ingenieurs-Ziele“. Um ein gültiges Experiment zu planen, aus welchem kausale Schlussfolgerungen gezogen werden können, müssen die Lernenden ein Wissenschaftsziel verfolgen, was bedeutet, dass sie die kausale Abhängigkeit einer Variable herausfinden wollen. Dem gegenüber stehen die Ingenieurs-Ziele, bei welchen die Schüler*innen versuchen, einen gewünschten Experimentausgang herzustellen (Schauble et al. 1991).

Um Schüler*innenvorstellungen verändern zu können, ist es zunächst notwendig, dass Lehrer*innen die zugrundeliegenden Vorstellungen kennen und einschätzen können (Morrison & Lederman 2003). Diese Annahme baut auf dem Prinzip von Ausubel (1968), welches besagt, das Wichtigste, um Schüler*innenvorstellungen beeinflussen zu können, ist, sich selbst jenes Wissen anzueignen, über welches die Lernenden bereits verfügen.

Die Veränderung der Schüler*innenvorstellungen wird unter dem Begriff Conceptual Change zusammengefasst. Conceptual Change selbst erfordert, dass die eigenen Theorien der Schüler*innen durch wissenschaftlich korrekte Theorien ersetzt werden (Posner et al. 1982). Jene Veränderungen entstehen insbesondere dann, wenn Lerner*innen neue Erfahrungen erlangen oder neue Informationen lernen, die nicht mit ihren bisherigen Vorstellungen übereinstimmen (Bransford et al. 2000). In der Forschung herrscht jedoch bislang kein gemeinsamer Konsens darüber, wodurch eine solche Vorstellungsänderung gekennzeichnet ist (Lin et al. 2016, Gropengießer & Marohn, 2018).

Gropengießer und Marohn (2018) unterschieden die verschiedenen Forschungsansätze hinsichtlich Conceptual Change in vier Ansätze, bestehend aus der „theoretischen Beschreibung eines Conceptual Changes“, den „Einflussfaktoren auf einen Conceptual Change“, der „Initiierung von Konzeptentwicklung“ und der „Entwicklung von Lernangeboten“. (Gropengießer & Marohn 2018). Conceptual Change kann auf verschiedene

Faktoren zurückgeführt werden, was dazu führt, dass die Veränderungen auch unterschiedlich schwer zu erreichen sein können. Krajcik und Shin (2023) beschreiben Conceptual Change als teilweise natürlich auftretende Konsequenz von Alltagserfahrungen, wie beispielsweise das Vermeiden einer heißen Herdplatte. Andere Vorstellungen hingegen verlangen direkten Einsatz, welcher oftmals in der Schule gegeben wird. Conceptual Change tritt auch auf, wenn sich Lernende neues Wissen zu dem bereits bestehenden aneignen müssen, um ein neues Konzept verstehen zu können. Dies kann den größten Wissenszuwachs mit sich bringen.

Conceptual Change kann durch mehrere Faktoren beeinflusst werden. Das Vorwissen von Schüler*innen spielt beispielsweise eine große Rolle in diesem Prozess. Lernen passiert dann, wenn Lernende neues Wissen mit bereits vorhandenem Wissen verbinden (Krajcik & Shin 2023). Posner et al. (1982) formulierten vier Bedingungen, die zentral sind, um den Prozess des Conceptual Change bei Schüler*innen hervorzurufen: *Unzufriedenheit* mit eigenen Vorstellungen, die Schüler*innen dazu zwingt, ihre Vorstellungen zu hinterfragen, *Verständlichkeit* von fachwissenschaftlichen Konzepten, *Plausibilität* und *Fruchtbarkeit* der neuen Vorstellungen in konkreten Situationen (Posner et al. 1982). Die Vorstellungsänderung erfordert demnach nicht nur Maßnahmen auf Schüler*innenseite, sondern erfordert auch Einsatz vonseiten der Lehrpersonen. In seinem Modell der didaktischen Rekonstruktion unternimmt Kattmann (2007) einen Versuch, das Zusammenspiel von Theorie und Praxis auf einer konstruktiven Ebene zu lösen. Bestehend aus den drei Teilen „Fachliche Klärung“, „Erfassen von Einzelperspektiven“ und „Didaktische Strukturierung“, soll das Lernen von Schüler*innen gefördert werden (Kattmann 2007).

3.2 Bekannte Schüler*innenvorstellungen beim Experimentieren

Folgende Schüler*innenvorstellungen konnten bislang von Tschirgi (1980), Siler & Klahr (2012) hinsichtlich der VKS identifiziert werden:

- 1) *Kontrolle der falschen Variable*: Schüler*innen haben zwar das grundsätzliche Verständnis dafür, dass eine Variable variiert werden muss, während die übrigen konstant gehalten werden, jedoch scheitern sie daran, die korrekte unabhängige Variable (Variable, die variiert werden muss) zu identifizieren (Siler & Klahr 2012).
- 2) *VKS für nicht kausale Variable*: Schüler*innen halten nur die abhängige Variable zwischen den Bedingungen konstant, während die übrigen Variablen variiert werden (Siler & Klahr 2012).
- 3) *Testung mehrerer Variablen*: Schüler*innen variieren mehrere Variablen gleichzeitig (Siler & Klahr 2012).
- 4) *Missachtung konfundierter Variablen*: Schüler*innen variieren mehrere Variablen, zusätzlich zur unabhängigen Variable (Siler & Klahr 2012).
- 5) *Kontrast von Bedingungen*: Schüler*innen planen zwei Experimente, bei denen die unabhängige Variable und weitere Variablen kontrastierend variiert werden (Siler & Klahr 2012).

- 6) *Nicht-contrastive Experimente*: Schüler*innen planen zwei exakt gleiche Versuchsbedingungen (Siler & Klahr 2012).
- 7) *Versuch mit nur einer Bedingung*: Schüler*innen wollen versuchen, mit nur einem Versuch, ein Experiment durchzuführen (Siler & Klahr 2012).
- 8) *Hold-one-thing-at-a-time*: Schüler*innen entwerfen Experimente, bei denen die unabhängige Variable nicht variiert, während sich die übrigen Variablen verändern. Die zugrunde liegende Vorstellung dabei ist, dass eine Variable gefunden werden muss, die konstant gehalten werden kann, um unter wechselnden Versuchsbedingungen identische Ergebnisse zu erhalten (Tschirgi 1980).
- 9) *Korrektes VKS Verständnis*: Schüler*innen verstehen die Prinzipien eines kontrollierten Experiments gänzlich und können die korrekte unabhängige Variable bestimmen und verändern, während die übrigen Variablen konstant gehalten werden (Siler und Klahr 2012).

Schwichow et al. (2022) gehen davon aus, dass die unterschiedlichen Vorstellungen darin begründet liegen, dass im klassischen VKS-Unterricht kaum auf das gesamte Spektrum an Schüler*innenvorstellungen eingegangen wird. Schwichow et al. (2016) untersuchten erstmals differenziert die VKS-Subkompetenzen. Demnach würden die Schwierigkeiten, die Schüler*innen bei der Anwendung der VKS haben, eher auf die benötigten Subkompetenzen zurückzuführen sein als auf den Aufgabeninhalt selbst. Ob die VKS-Kompetenz domänenspezifisch oder domänenübergreifend ist, wurde bislang jedoch noch nicht ausführlich genug erforscht. Demnach kann noch keine sichere Aussage darüber getroffen werden (Schwichow et al. 2022, Schwichow et al. In prep).

Haben Schüler*innen Vorstellungen von der VKS, die zu einem konfundierten Experiment führen, wird hinsichtlich der Variablenkontrollstrategie von Designfehlern gesprochen (Siler & Klahr 2012). Eine Übersicht ist in Tabelle 1 zu finden. Unter Designfehlern wird ein fehlerhaftes Versuchsdesign, wie etwa dem Verändern von mehreren Variablen oder konstant Halten aller Variablen, verstanden (Siler & Klahr 2012), beziehungsweise ist ein Experiment dann konfundiert, wenn abhängige Variablen nicht kontrolliert werden (Masnick & Klahr, 2003).

Tabelle 1 Übersicht über die Schüler*innenvorstellungen, Designfehler und die Auswirkungen auf die Anwendungen der VKS

Fehler und Regelmässigkeit bei der Planung von konfundierten Experimenten								
<i>Schüler*innenvorstellung</i>	Designfehler	Fokus auf Variablenausprägung	Testung einer V.	Bestimmung korrekter unabhängiger V.	Vgl. über Bedingungen	Veränderung der untersuchten V.	Kontrolle der übrigen V.	Verständnis des Zshg. Zwischen Experiment und kausaler Beziehung
<i>(1) Kontrolle der falschen Variable</i>	Planung kontrollierter Experimente für die falsche Variable			x				
<i>(2) VKS für nicht kausale Variable</i>	Planung (mehrerer) konfundierter Experimente						x	
<i>(3) Testung mehrerer Variablen</i>	Planung (mehrerer) konfundierter Experimente		x	x			x	x
<i>(4) Missachtung konfundierter Variablen</i>	Planung (mehrerer) konfundierter Experimente						x	x

Fehler und Regelmissachtung bei der Planung von konfundierten Experimenten

<i>Schüler*innenvorstellung</i>	Designfehler	Fokus auf Variablen- ausprägung	Testung einer V.	Bestimmung korrekter unabhängige V.	Vgl. über Bedingungen	Veränderung der untersuchten V.	Kontrol le der übrigen V.	Verständnis des Zshg. Zwischen Experiment und kausaler Beziehung
<i>(5) Kontrast von Bedingungen</i>	Planung (mehrerer) konfundierter Experimente	x	x	x			x	x
<i>(6) Nicht-kontrastive Experimente</i>	Nicht vergleichbare Experimente					x	x	x
<i>(7) Versuch mit nur einer Bedingung</i>	Planung nur einer Versuchsbedin- gung				x	x	x	x
<i>(8) Hold-one-thing-at-a- time</i>	Hold-one- thing-at-a-time					x	x	x
<i>(9) Korrektes VKS- Verständnis</i>	Planung kontrollierter Experimente							

Sechs der Schüler*innenvorstellungen (*Testung mehrerer Variablen, Missachtung konfundierter Variablen, Kontrast von Bedingungen, nicht-kontrastive Experimente, Versuch mit nur einer Bedingung, Hold-one-thing-at-a-time*) resultieren demnach daraus, dass die Lernenden kein Verständnis für das Ziehen kausaler Schlussfolgerungen haben. Ebenso stellt die „Kontrolle der übrigen Variablen“ Probleme für die Schüler*innen (*VKS für nicht kausale Variable, Testung mehrerer Variablen, Missachtung konfundierter Variablen, Kontrast von Bedingungen, nicht-kontrastive Experimente, Versuch mit nur einer Bedingung*) dar, die sich in den Vorstellungen *nicht-kontrastive Variable, Testung mehrerer Variablen, Missachtung konfundierter Experimente, Kontrast von Bedingungen, nicht-kontrastive Experimente sowie Experimente mit nur einer Bedingung* äußern.

In der Praxis bedeutet dies, dass die Schüler*innen zu viele Variablen variieren, wenngleich sie nur die zu untersuchende Variable verändern sollten. Dies geht ebenso mit dem Aspekt „Bestimmung der korrekten unabhängigen Variable“ einher, der dazu führt, dass Schüler*innen entweder zu viele Variablen variieren (*Testung mehrerer Variablen, Kontrast von Bedingungen, Hold-one-thing-at-a-time*) oder die falsche unabhängige Variable bestimmen (*Kontrolle der falschen Variable, Testung mehrerer Variablen, Kontrast von Bedingungen*). Die Schüler*innenvorstellungen *Testung mehrerer Variablen, Kontrast von Bedingungen, nicht kontrastive Experimente, Experimente mit nur einer Bedingung* sowie *hold-one-thing-at-a-time* basieren auf komplexeren Verständnisschwierigkeiten, bei denen drei oder mehr Aspekte der Variablenkontrollstrategie nicht verstanden bzw. missachtet wurden. Demnach bedingen die Schüler*innenvorstellungen *VKS nicht kausal, Testung mehrerer Variablen, Ignoranz konfundierter Experimente* und *Vergleich ganzer Bedingungen* dieselben Designfehler. Obwohl sie initial unterschiedlich erscheinen, führen sie dennoch zum selben Ergebnis – dem Planen von einem oder mehreren konfundierten Experimenten. Die Gründe, die schlussendlich zu einem konfundierten Versuchsdesign führen, sind, dass die Schüler*innen nicht wissen, dass eine zusätzliche veränderte Variable keinen Einfluss auf das Experiment hat, die konfundierten Experimente ignoriert werden, verschiedene Variablen in nur einem Versuchsdesign untersucht werden, oder außerordentlich stark ausgeprägte Auswirkungen auf die abhängige Variable erzeugt werden wollen (Schwichow 2022). Manche Schüler*innen hingegen verstehen das Prinzip eines kontrollierten Experiments, scheitern jedoch daran, die passende unabhängige Variable zu identifizieren. Somit wird das Experiment zwar kontrolliert, jedoch für die falsche Variable, resultierend in dem Designfehler *Planung kontrollierter Experimente, jedoch für die falsche Variable*. Ebenso kann, wie in Designfehler 6 (*nicht vergleichbare Experimente*) beschrieben, ein Verständnis von kontrollierten Experimenten überinterpretiert werden und dazu führen, dass Schüler*innen ausschließlich in ihren Augen „faire“ Versuchsdesigns gestalten, die sich nicht voneinander unterscheiden und somit keine Aussagekraft haben. Fehlt Schüler*innen hingegen das Verständnis dafür, weshalb Bedingungen verglichen werden müssen, planen sie nur ein Experiment ohne Vergleichsbedingung, siehe *Planung nur einer Versuchsbedingung* (Schwichow et al. 2022). *Hold-one-thing-at-a-time* wurde von Tschirgi (1980) beschrieben und von Schwichow et al.

(2022) wieder aufgegriffen. Insgesamt bedingen also die acht bekannten Präkonzepte fünf verschiedene Designfehler (Schwchow et al. 2022). Aufgrund dieser Gegenüberstellung von Designfehlern und deren Manifestation in den Präkonzepten je Subkompetenz, kann davon ausgegangen werden, dass Schüler*innenvorstellungen als über verschiedene Situationen hinweg stabile Konzepte angesehen werden können und nicht nur als fehlerhafte Ansichten in kontextspezifischen Aufgaben. Dementsprechend sollen die Designfehler auch durchgängig in verschiedenen Aufgaben mit verschiedenen Subkompetenzen auftreten (Schwchow et al. 2022).

Im Folgenden wird eine Übersicht nach Schwchow et al. (2022) über alle bekannten Designfehler und deren Ausprägung in den einzelnen VKS-Subkompetenzen Identification (ID), Interpretation (IN) sowie Understanding (UN) gegeben.

Die folgende Auflistung ist adaptiert an Tab. 2 von Schwchow et al. (2022), S. 97. Die Subkompetenz Planing (PL) wurde von Schwchow et al. (2022) nicht untersucht. Für die vorliegende Masterarbeit ist die Subkompetenz jedoch von Bedeutung, weshalb die Auflistung darum ergänzt wurde.

- Korrektes VKS-Verständnis
 - ID: Nur die unabhängige Variable wird verändert, alle anderen Variablen bleiben gleich.
 - IN: Die unabhängige Variable wird verändert, die übrigen Variablen bleiben konstant
 - UN: Mindestens eine weitere Variable hat sich verändert.
 - PL: Unterschiede in der unabhängigen Variable, alle übrigen Variablen sind konstant.
- Kontrolliertes Experiment für die falsche Variable
 - ID: Kontrolliertes Experiment, aber die falsche Variable variiert.
 - IN: Eine abhängige Variable variiert, mindestens eine weitere Variable ist unverändert.
 - UN: Die falsche Variable variiert, eine weitere Variable bleibt konstant.
 - PL: Eine Variable variiert (jedoch die falsche, hinsichtlich der Hypothese), alle anderen werden konstant gehalten.
- Missachtung konfundierter Variablen
 - ID: Zwei oder mehrere Variablen werden zwischen den Versuchen verändert.
 - IN: Die unabhängige Variable variiert, zusätzlich haben sich noch eine oder zwei weitere Variablen verändert.
 - UN: Die unabhängige Variable variiert, andere Variablen variieren ebenso.
 - PL: Die unabhängige Variable variiert, zusätzlich variiert noch mindestens eine weitere Variable.
- Nicht-kontrastive Experimente
 - ID: Vergleich zweier identer Versuchsbedingungen.
 - IN: Alle Variablen (inklusive der unabhängigen) sind unverändert.

- UN: Die unabhängige Variable variiert, ebenso wie andere Variablen.
- PL: Alle Ausprägungen von Variablen beider Telexperimente sind identisch.
- Planung nur einer Versuchsbedingung
 - ID: Experiment hat nur eine Bedingung.
 - IN: /
 - UN: /
 - PL: Planung nur eines Experiments.
- Hold-one-thing-at-a-time
 - ID: Die unabhängige Variable bleibt konstant, alle anderen Variablen werden variiert.
 - IN: /
 - UN: Eine zusätzliche Variable verändert sich.
 - PL: Die unabhängige Variable bleibt unverändert. Alle anderen Variablen variieren.

Die umfassende Forschung zu Schüler*innenvorstellungen hinsichtlich der VKS und den damit einhergehenden Designfehlern, bilden ein breites Forschungsgebiet. Designfehler, die zu fehlerhaften Versuchsdesigns führen, bieten aber nicht nur wertvolle Informationen über die zugrundeliegenden Vorstellungen, sondern stellen auch wirkungsvolle Lerngelegenheiten dar. Fehler sind im Kontext des Lernens als natürliches Nebenprodukt des Lernprozesses zu betrachten und stellen wichtige Lerngelegenheiten dar (Oser & Spychiger 2005, Tulis et al. 2015, Zhang & Fiorella 2023). Dass aus Fehlern gelernt werden kann, wurde zudem in verschiedenen empirischen Studien in anderen Fachbereichen, wie beispielsweise in Mathematik (Kapur 2010 2014).

4 Lernen aus Fehlern

4.1 Theoretischer Hintergrund

Negatives Wissen dient unter anderem zur Abgrenzung, was eine Sache nicht ist, und kann in negativ deklaratives Wissen, negativ prozedurales Wissen, negativ strategisches Wissen und negativ schema-orientiertes Wissen unterteilt werden. Negatives Wissen bezieht sich auf Erinnerungen, welche als falsch, unangebracht oder ineffektiv abgespeichert sind und wird unter anderem durch das Begehen von Fehlern aufgebaut. Fehler allein führen allerdings noch nicht dazu, dass sich neues (negatives) Wissen aufbaut. Vielmehr erfordert dieser Prozess als ersten Schritt die Einsicht darüber, dass ein Fehler begangen wurde. Dabei genügt es nicht, nur zu wissen, dass etwas falsch ist, vielmehr muss verstanden werden, warum es falsch ist und wie es besser gemacht werden kann. Fehler können in weiterer Folge als notwendige Voraussetzung angesehen werden, um überhaupt lernen zu können (Oser & Spychiger 2005). Oser et al. (2012) fassen die Bereiche, in denen negatives Wissen von Vorteil sein kann,

folgendermaßen zusammen: Negatives Wissen kann eine Schutzfunktion haben, die uns davor bewahrt, Fehler zu wiederholen, ebenso wie es helfen kann, zwischen gut und schlecht zu unterscheiden, Unterschiede zwischen konträren Merkmalen zu erkennen oder Sicherheit zu schaffen, indem ein Problem mit einem spezifischen Verfahren behandelt wird (Oser et al. 2012).

Fehler sind oft negativ behaftet, jedoch lässt sich ein unterschiedlicher Umgang mit Fehlern und dem weiteren selbstregulierten Lernverlauf feststellen. Zusammengefasst wird dies von Boekaerts (2006) in dem „dualen Verarbeitungsmodell zur Selbstregulation“ (original „*dual processing self-regulation model*“). Dabei wird zwischen zwei Zielsetzungen der Lernenden differenziert. Einerseits dem „Wachstums-Pfad“ und andererseits dem "Wohlbefinden-Pfad". Lernende, die ein bestimmtes Ziel verfolgen, wie beispielsweise das Erkennen von eigenen Fehlern, bewegen sich auf dem Wachstums-Pfad. Wohingegen sich Lernende, die vor allem auf die negativen Folgen eines Fehlers fokussiert sind, den Wohlbefinden-Pfad bevorzugen. Es wird jedoch vermutet, dass Lernende mithilfe geeigneter Strategien zwischen den Pfaden navigieren können (Boekaerts 2006). Dresel et al. (2013) formulierten zwei Hauptreaktionen, die auf Fehler folgen. Einerseits die Handlungsadaptivität, die aussagt, wie aktiv ein*e Lernende*r darauf reagiert, den Fehler zu erkennen, zu analysieren und das zugrundeliegende Missverständnis zu korrigieren. Andererseits die affektiv-motivationale Adaptivität, die Auskunft darüber gibt, wie gut ein*e Lernende*r die Motivation aufrechterhält, negative Gefühle, die durch den Fehler entstehen können, bewältigt.

Fehler werden grundsätzlich durch eine ungewollte Abweichung zwischen dem aktuellen und dem gewünschten Zustand, oder als eine Abweichung von einem festgelegten Standard beschrieben (Frese & Zapf 1994). Das Erkennen des Fehlers, sei es selbstständig oder durch externes Feedback, ist essenziell, und stellt den Ausgangspunkt dar, um aus Fehlern lernen zu können. Diese Erkenntnisse basieren auf Konzepten des selbstregulierten Lernens, die betonen, wie kognitive, metakognitive und emotionale Prozesse zusammenspielen (Tulis et al. 2015).

Um effektiv aus Fehlern zu lernen, ist es unerlässlich, sich weiterhin intensiv mit der jeweiligen Aufgabe auseinanderzusetzen. Dies ermöglicht nicht nur die Korrektur des Fehlers, sondern auch eine tiefgehende Reflexion über die zugrundeliegenden Annahmen (Tulis et al. 2015). In ihrem Prozessmodell betonen Tulis et al. (2016) die entscheidende Rolle von Reflexion und Selbstbewertung beim Lernen aus Fehlern. Das Modell hebt zudem hervor, dass sowohl emotionale als auch motivationale Faktoren im Lernprozess von zentraler Bedeutung sind (Tulis et al. 2016). Zhang und Fiorella (2023) beschreiben zwei Phasen des Lernens aus Fehlern, spezifisch zum Lernen aus eigenen Fehlern. In der Fehlergenerierungsphase wird zunächst Vorwissen aktiviert, während in der Fehlererkennung- und -korrekturphase durch Selbst-Erklärung gelernt werden kann. Dies setzt jedoch voraus, dass Schüler*innen ihre Antworten mit vorgegebenen Lösungen abgleichen können (Zhang & Fiorella 2023).

Verschiedene Unterrichtsmethoden integrieren Fehler als Lernmöglichkeiten. Zwei konträre Ansätze sind der PS-I (Problem-solving-prior-to-instruction) sowie der I-PS (Instruction-prior-to-Problem-Solving) Ansatz. Während beim I-PS Ansatz zunächst neues Wissen mittels

Instruktionen seitens der Lehrkraft vermittelt wird, um das neue Wissen anschließend durch vertiefende Aufgaben zum Problemlösen gefestigt wird (Kirschner et al. 2006), folgt der PS-I Ansatz der umgekehrten Logik. Dabei wird die Problemlöse-Phase vor der Instruktion eingeführt und die Schüler*innen zuerst mit Aufgaben konfrontiert, die ihren momentanen Wissensstand überschreiten (Kapur & Bielaczyc 2012). Die zugrundeliegenden Mechanismen des PS-I Ansatzes sind (1) Vorwissen, mit dem Schüler*innen an die Aufgabe herantreten, (2) Lenkung der Aufmerksamkeit auf das neue Thema sowie (3) Hinweise auf die Wissenslücken (Loibl et al. 2017). Schüler*innen durchlaufen dabei zwei aufeinanderfolgende Phasen: Zuerst versuchen sie, unbekannte Aufgaben eigenständig zu bewältigen, bevor ihnen anschließend gezielte Anleitungen und Erklärungen geboten werden (Kapur 2010).

Productive Failure (PF) stellt im Rahmen des PS-I Ansatzes ein Konzept dar, das unter das große Design-Paradigma der Vorbereitung auf zukünftiges Lernen fällt (Sinha & Kapur 2021). Productive Failure besteht aus den vier Hauptmechanismen: (1) Aktivierung des Vorwissens, (2) Aufmerksamkeit auf kritische konzeptionelle Merkmale, (3) Erklärung und Ausarbeitung dieser Merkmale sowie (4) Organisation und Zusammenstellung kritischer konzeptioneller Merkmale zu gezielten Konzepten (Kapur & Bielaczyc 2012). PF ist darauf ausgelegt, dass die Problemlösungsphase so gestaltet ist, dass sie zu Fehlern führt. Dadurch, dass Schüler*innen mit einem Problem konfrontiert werden, das ihr Wissen übersteigt, besteht ein hohes Fehlerpotenzial. In der darauffolgenden Phase nutzen Lehrkräfte die von den Schüler*innen erarbeiteten Lösungen, um gezielt Konzepte zu vermitteln. Das Ziel ist, durch initiale Fehler, Fehler auf langfristige Sicht zu minimieren (Sinha & Kapur 2021).

4.2 Aktuelle Forschung zu Lernen aus Fehlern

Im Bildungsbereich hat das Lernen aus Fehlern immer mehr an Bedeutung gewonnen, da es als eine effektive Methode angesehen wird, Schüler*innen dabei zu unterstützen, Konzepte tiefgreifender zu verstehen. In diesem Kontext haben zahlreiche Studien verschiedene Ansätze und Strategien untersucht, um zu ermitteln, wie Fehler am besten in den Lernprozess integriert werden können. Durkin und Rittle-Johnson (2012) untersuchten beispielsweise in der Domäne Mathematik, ob Lernen durch Vergleichen von falschen und richtigen Beispielen effektiver ist, als Lernen mit zwei korrekten Beispielen, mit dem Ergebnis, dass jene Gruppe, die falsche und richtige Beispiele erhielt, Methoden und zentrale Konzepte besser zu verstehen schien als die Gruppe, die sich nur mit korrekten Beispielen beschäftigte. Vorwissen hatte in diesem Kontext keinen Einfluss auf die Ergebnisse (Durkin & Rittle-Johnson 2012).

Um zu untersuchen, ob Lernen aus Fehlern nachhaltiger mittels des PS-I oder I-PS Ansatzes in den Unterricht integriert werden kann, wurden die beiden Methoden miteinander verglichen. Der PS-I Ansatz, bei dem Schüler*innen zuerst Aufgaben lösen und dann Anweisungen erhalten, scheint in den untersuchten Kontexten Mathematik oder Statistik, effektiver zu sein als der I-PS Ansatz, bei dem Schüler*innen zuerst Anweisungen bekommen und dann Aufgaben bearbeiten (Loibl et al. 2017, Loibl & Rummel 2014, Schwartz & Martin 2004, Sinha & Kapur 2021). Die Überlegenheit des PS-I Ansatzes gegenüber dem I-PS Ansatz kann jedoch

nur dann festgestellt werden, wenn die Schüler*innen während der Problemlösung mit kontrastierenden Beispielen arbeiten und zugleich die Schüler*innenlösungen in der „Instruction Phase“ thematisiert wurden. Andernfalls scheint es, als ob der I-PS Ansatz überlegen wäre. (Sinha & Kapur 2019) PS-I kann das Lernen effektiv unterstützen, indem es kognitive Mechanismen wie die Aktivierung von Vorwissen, Bewusstwerdung von Wissenslücken und die Auseinandersetzung mit tiefergehenden Merkmalen des Lernstoffes fördert (Loibl et al. 2017). Ein weiteres effektives Kriterium, um die Überlegenheit des PS-I Ansatzes zu erreichen ist, die Schüler*innen anzuweisen, die fehlerhaften Versuche mit den richtigen Lösungen zu vergleichen (Loibl & Leuders 2018).

Jackson et al. (2021) untersuchten in einer systematischen Literaturanalyse, wie Fehler im MINT-Kontext als Lerngelegenheit genutzt werden können. Dabei konnten fünf Hauptkenntnisse gesammelt werden, die den unterschiedlichen Umgang mit Fehlern in bisherigen Forschungen thematisieren: (1) Unterschiedliche Bedeutung von Fehlern, die entweder aus prozessorientierten oder aus lösungsorientierten Definitionen in der bisherigen Literatur bestanden. (2) Reaktionen von Lernenden auf Fehler, wobei positive Reaktionen vor allem bei weiterem Erfolg verzeichnet werden konnten. (3) Fehler als Mechanismus zur Aufdeckung zentraler Konzepte, insbesondere der Aktivierung von Vorwissen. (4) Fehler fördern die generelle Problemlösekompetenz und (5) Klassenklima und Kommunikation über Fehler als bedeutende Einflussgröße auf den Effekt von Lernen aus Fehlern (Jackson et al. 2021).

Neben den verschiedenen Methoden, die Fehler in das Lernen integrieren, kann auch zwischen Lernen aus eigenen und lernen aus fremden Fehlern unterschieden werden. Metcalfe (2017) legte dar, dass fehlerhaftes Lernen, gefolgt von korrigierendem Feedback vor allem dann sehr lernfördernd ist, wenn Schüler*innen zuvor davon überzeugt sind, dass ihre fehlerhafte Antwort korrekt ist. Sobald sie durch das nachfolgende Feedback verstehen, dass ihre ursprüngliche Antwort nicht korrekt war, tritt ein Lerneffekt ein. Metcalfe und Xu (2018) untersuchten in einer Folgestudie gezielt den Nutzen von eigenen und fremden Fehlern in der Domäne Allgemeinwissen. In der Studie wurden drei Experimente durchgeführt, mit dem Ergebnis, dass die Proband*innen aus eigenen Fehlern besser zu lernen scheinen im Vergleich zum Lernen aus den Fehlern anderer. Das Generieren eigener Antworten, seien es korrekte oder fehlerhafte, im Vergleich zum bloßen Hören der Antworten anderer, verbesserte das spätere Erinnern der korrekten Antworten. Diese Ergebnisse unterstützen die Theorie der Rekonsolidierung, die besagt, dass der Abrufungsprozess Gedächtnisspuren veränderbar macht und somit Korrekturen ermöglicht, wenn die Person selbst aktiv am Lernprozess teilnimmt (Metcalfe & Xu 2018). Zhang und Fiorella (2023) beschäftigten sich mit der Entwicklung und Untersuchung eines integrierten Modells, das den Lernprozess aus Fehlern heraus abbildet. Die Grundannahme des Modells liegt in der Überlegenheit von selbstgenerierten Fehlern gegenüber fremden Fehlern. Lernen aus Fehlern erfolgt in zwei Phasen, zunächst die Generierung von Fehlern, gefolgt von der direkten Korrekturerfahrung. Die Korrekturerfahrung dient dabei dem Lernprozess, indem Lernende ihre Antworten

verarbeiten und vergleichen und sich den Prozess währenddessen selbst erklären (Zhang & Fiorella 2023).

Gezielt mit Schüler*innenvorstellungen in Physik beschäftigten sich Siantuba et al. (2023). Jene Schüler*innen, die in einem forschend-lernenden Ansatz gezielt mit Schüler*innenvorstellungen konfrontiert wurden, hatten gegenüber der Kontrollgruppe einen höheren Lernzuwachs (Siantuba et al. 2023). Eine Analyse der Einstellungen von Lehrenden und Lernenden gegenüber Fehlern zeigte, dass viele Lernende Fehler beim Lernen lieber vermeiden möchten, sich deren potenziellen Lerneffekten aber durchaus bewusst sind. Lehrende hingegen diskutieren Fehlern öfter und betonen die Vorteile des Lernens aus Fehlern, bieten aber unterschiedlich häufig Ressourcen an, die Lernen aus Fehlern fördern (Pan et al. 2020).

Sowohl die Experimentierkompetenz als auch die Forschungsfelder Schüler*innenvorstellungen und Lernen aus Fehlern fanden bislang umfassende Beachtung in der Forschung. Übergreifende Forschungen, die Lernen aus Fehlern hinsichtlich der Variablenkontrollstrategie untersuchen, gibt es bislang jedoch wenige. Ebenso wurden Schüler*innenvorstellungen und Designfehler innerhalb der vier Subkompetenzen der VKS bislang wenig differenziert betrachtet. Damit einhergehend gibt es auch noch keine publizierte differenzierte Betrachtung von domänenspezifischem Fachwissen innerhalb der Subkompetenzen (Schwichow et al. in prep.).

5 Lernen aus Fehlern im Kontext des Experimentierens

In einem Review von Zimmermann (2007) wird die wissenschaftliche Erkenntnisgewinnung von Primar- und Sekundarschüler*innen untersucht. Kinder würden demnach bezüglich ihrer Entwicklung von wissenschaftlichem Denken oft unterschätzt werden. Jedoch erfordert wissenschaftliches Denken auch ein komplexes System an kognitiven und metakognitiven Fähigkeiten, welches durch eine erhebliche Menge an Übung und Praxiserfahrung gestärkt werden kann. Dieser komplexe Prozess hat, je nach Individuum, eine individuelle Entwicklungszeit. Fachwissen und strategisches Handeln beeinflussen sich gegenseitig, was Maßnahmen erfordert, um beide Strategien zu fördern (Zimmermann 2007).

Wissenschaftliche Grundbildung und Experimentierkompetenz sind eng miteinander verknüpft, wobei die Förderung von Experimentierfähigkeiten als zentraler Aspekt der naturwissenschaftlichen Bildung hervorgehoben wird (Lehrplan 2023). Diese Kompetenz umfasst das Stellen wissenschaftlicher Fragen, das Generieren von Hypothesen und die systematische Durchführung von Untersuchungen, was als Fundament für tiefgreifende naturwissenschaftliche Grundbildung angesehen wird (Mayer 2007). Ein entscheidender Faktor in diesem Bildungsprozess sind Schüler*innenvorstellungen, die häufig durch Alltagserfahrungen geprägt sind. Diese Vorstellungen bieten einen wertvollen Ausgangspunkt, um ein wissenschaftliches Verständnis aufzubauen (Schrenk et al. 2019) und ihre

Anerkennung und Bearbeitung sind entscheidend, um Schüler*innen effektiv beim Erwerb wissenschaftlicher Konzepte zu unterstützen (Morrison & Lederman 2003).

Um die Experimentierfähigkeit zu stärken, ist gezieltes Training essenziell (Schwichow et al. 2016). Laut der Metastudie von Schwichow et al. (2016) sind vor allem Interventionen, die sich explizit mit der Vermittlung der VKS-Kompetenz oder auf die VKS und zusätzlichen Inhalt fokussieren, am effektivsten. Es macht demnach keinen Unterschied, ob Lehrende sich ausschließlich auf die Vermittlung der VKS konzentrieren oder noch einen zusätzlichen Fokus auf Fachwissen legen (Schwichow et al. 2016). Ein ausschlaggebender Faktor für den Erfolg von gezieltem VKS-Training ist das Einbeziehen von konfundierten Experimenten, um einen kognitiven Konflikt bei den Schüler*innen auszulösen und so das Lernen zu fördern (Schwichow et al. 2016). Lernen aus Fehlern wird als wesentlicher Bestandteil des Lernprozesses betrachtet, wobei Fehler wertvolle Lerngelegenheiten bieten (Oser & Spychiger 2005), durch die Schüler*innen ihr Verständnis vertiefen und ihre Experimentierfähigkeiten verbessern können, was wiederum einen zentralen Punkt in den neuen Lehrplänen der Mittelschule darstellt (Lehrplan 2023). Die Reflexion über Fehler und die darauffolgende Korrektur tragen wesentlich zur Entwicklung einer wissenschaftlichen Grundbildung bei.

Lernen aus Fehlern sowie Inquiry based research oder der PS-I Ansatz wurden bislang noch nicht umfassend im Zusammenhang mit der Variablenkontrollstrategie untersucht.

Chase & Klahr (2017) betonen, dass bei der Anwendung der Variablenkontrollstrategie, während der Problemlösungsphase bei domänenübergreifenden Fähigkeiten selten implizites Feedback über angemessene Ziele während der Inquiry Phase gegeben wird, was jedoch wesentlich für das Lernen ist. Chase & Klahr (2017) verwiesen auf einen Gleichstand der beiden Ansätze mit leichter Überlegenheit der I-PS Methode. Vor allem das eindeutige Feedback über die Wirksamkeit von Problemlösungsansätzen erwies sich als ausschlaggebend (Chase & Klahr 2017). Am überlegensten erwies sich der PS-I Ansatz laut einer Metaanalyse von Sinhar & Kapur (2021), die zuvor definierten Merkmale von „Productive Failure (PF)“ berücksichtigt wurden. Eine aktuelle Studie von Zhang und Sweller (2024) zeigte, dass Schüler*innen, die grundsätzlich ein höheres Fachwissen (Untersucht wurde die Dichte von Stoffen in der Domäne Physik) aufweisen, hinsichtlich der VKS mehr von einem PS-I Ansatz profitieren als Schüler*innen mit geringerem Vorwissen (Zhang & Sweller 2024).

Diese drei Schlüsselkomponenten: Experimentierfähigkeit, Schüler*innenvorstellungen und das Lernen aus Fehlern, bilden zusammen ein dynamisches Gerüst, das Schüler*innen nicht nur auf die Herausforderungen der modernen Welt vorbereitet, sondern auch eine kritische Denkweise und ein tiefgreifendes Verständnis für naturwissenschaftliche Prozesse fördert. Die Integration dieser Aspekte schafft eine solide Basis für die nachfolgende Analyse und Diskussion der Ergebnisse, indem sie die Bedeutung einer umfassenden naturwissenschaftlichen Bildung unterstreicht, die über das bloße Reproduzieren von Fakten hinausgeht und auf das Verständnis, die Anwendung und die kritische Bewertung von Wissenschaft in realen Kontexten abzielt.

6 Fragestellungen und Hypothesen

Schüler*innenvorstellungen und Designfehler bei der VKS wurden bislang in vielen Studien untersucht. Auch der PS-I Ansatz im Vergleich zum I-PS Ansatz erhielt Aufmerksamkeit in der Forschung. Vor allem die Darlegungen der Präkonzepte und Designfehler bezogen sich jedoch auf Primarschüler*innen. Sekundarschüler*innen fanden wenig Beachtung.

Nach momentanem Wissensstand wurde bislang noch nicht erforscht, ob eine Beschäftigung mit Fehlern im Rahmen eines PS-I Ansatzes zu einer Veränderung der Schüler*innenvorstellungen und Designfehler zwischen zwei Testzeitpunkten führen kann – insbesondere bei Schüler*innen der Sekundarstufe I. Ebenso wurde die Subkompetenz PL bislang noch nicht in die differenzierte Betrachtung der Subkompetenzen miteinbezogen, da sie sich für Studien mit Primarschüler*innen als zu komplex herausgestellt hat. Domänenspezifisches Fachwissen fand bislang eine untergeordnete Rolle in der Erforschung der VKS, insbesondere in Verbindungen mit den vier Subkompetenzen.

Da die Experimentierkompetenz nicht nur eine grundlegende wissenschaftliche Kompetenz darstellt, sondern auch in den neuen Lehrplänen der Sekundarstufe I vermehrt Beachtung findet, sollen in der Masterarbeit folgende Fragen geklärt werden:

1. Welche Designfehler hinsichtlich der Variablenkontrollstrategie treten bei Schüler*innen der Sekundarstufe I auf?
2. Welche Muster von Designfehlern lassen sich bei den einzelnen Subkompetenzen der Variablenkontrollstrategie erkennen?
3. Führt die Beschäftigung mit eigenen Fehlern zwischen Prätest und Posttest zu einer Veränderung von Designfehlern bzw. Präkonzepten und inwiefern unterscheiden sich die Lernstrategien?

Folgende Hypothesen gilt es in der Masterarbeit zu überprüfen.

H1: Alle definierten Designfehler (Siler & Klahr 2012, Schwichow et al. 2022) lassen sich auch bei Schüler*innen der Sekundarstufe I identifizieren.

Schüler*innen der Sekundarstufe I weisen grundsätzlich ein höheres Fachwissen auf als Primarschüler*innen, somit kann angenommen werden, dass ihnen dies auch zu einer höheren VKS-Kompetenz verhilft (Schwichow & Nehring 2018; Edelsbrunner et al. 2018). Jedoch wird vermutet, dass Schüler*innen der Sekundarstufe I auch noch keine fachlich gänzlich korrekten Vorstellungen mitbringen und deshalb dieselben Designfehler vorliegen werden wie bei Primarschüler*innen.

H2.1.: Die Subkompetenz UN erweist sich als die schwierigste Subkompetenz, in welcher die meisten Designfehler gemacht werden.

Zurückgeführt auf Schwichow et al. (2016, 2022) kann davon ausgegangen werden, dass in der Sekundarstufe I genauso wie in der Primarstufe die Subkompetenzen (PL, ID, IN, UN) ein ausschlaggebendes Kriterium für die wahrgenommene Schwierigkeit einer Aufgabe ist. Wie

von Schwichow (2016) beschrieben, ist auch in dieser Untersuchung davon auszugehen, dass die Subkompetenz UN, also das Verständnis über die fehlende Aussagekraft eines konfundierten Experiments, die meisten Schwierigkeiten für die Schüler*innen birgt.

H2.2.: Im Fachbereich Allgemeinwissen treten über alle Subkompetenzen hinweg am wenigsten Designfehler auf.

Es wird davon ausgegangen, dass die VKS domänenspezifisch ist. Da vermutet wird, dass im Bereich Allgemeinwissen das meiste Vorwissen vorhanden ist, wird die Hypothese aufgestellt, dass in diesem Fachkontext am wenigsten Designfehler auftreten.

H3: Schüler*innen, die sich mit eigenen Fehlern beschäftigen, machen im Posttest weniger Designfehler als im Prätest.

Fehler können, wie bereits erläutert, als Lerngelegenheiten dienen. Auch wenn die Überlegenheit des PS-I Ansatzes im Kontext der Experimentierkompetenz noch nicht eingehend untersucht wurde und manche Studien auf einen Gleichstand der beiden Ansätze hindeuten (Chase & Klahr 2017), so kann dennoch argumentiert werden, dass der PS-I Ansatz zu höherem Wissen nach der Intervention führt. Schüler*innen haben demnach die Möglichkeit, aus ihren Fehlern zu lernen und weniger Designfehler im Posttest zu machen. Einhergehend mit der Verringerung der Designfehler ist auch anzunehmen, dass sich die Präkonzepte der Schüler*innen entsprechend verändern.

7 Material & Methoden

7.1 Teilnehmer*innen

Die Daten wurden 2021 im Zuge der Dissertation von Hämmerle in Zusammenarbeit mit drei Masterstudierenden der Biologie erhoben. Fokus jener Interventionsstudie ist es, die Experimentierfähigkeit mittels der Variablenkontrolle von Schüler*innen der Sekundarstufe I zu erfassen, wobei die einzelnen Subkompetenzen der VKS sowie die Fachkontexte Biologie, Chemie, Allgemeinwissen und Physik miteinbezogen wurden. Diese Masterarbeit wurde im Zuge der Interventionsstudie von Hämmerle (in prep.) verfasst.

Die Intervention wurde an n=292 Schüler*innen der 7. und 8. Schulstufe durchgeführt. Teilgenommen haben zwölf Klassen aus insgesamt fünf Schulen der allgemeinbildenden höheren Schule in Wien. Abgesehen vom Geschlecht und Alter der Schüler*innen sowie die letzten Zeugnisnoten in den Fächern Physik, Biologie, Chemie, Deutsch und Mathematik, wurden keine weiteren demographischen Daten erhoben.

Für die Datenanalyse mussten 26 Schüler*innen ausgeschlossen werden, weil sie beim Prätest nicht anwesend waren. Daraus ergab sich eine Teilnehmer*innenzahl von n=266, (männlich: 55,5%, weiblich: 42,7%, divers: 1,8%) die zum Zeitpunkt der Studie ein Durchschnittsalter von 12,79 Jahren (SD = 0,7) hatten.

7.2 Studiendesign

Die Daten wurden im Zuge einer Interventionsstudie, bestehend aus Prä-, Post- und Follow-Up Test mittels eines paper pencil Tests erhoben. Demnach gab es die drei Testzeitpunkte Prä- Post- und Follow-Up Test. Verwendet wurden jedes Mal die gleichen Testhefte, jedoch in geänderter Reihenfolge, so dass kein*e Schüler*in zweimal dasselbe Testheft bearbeitete. Eine Woche vor der Intervention erhielten die Schüler*innen von einer Studienassistentin des SSC Lebenswissenschaften in der Schule das VKS-Testheft und absolvierten den Prätest.

Die Intervention erfolgte zwei Wochen später vor Ort im Lehr-Lern-Labor des AECC Biologie der Universität Wien. Dabei wurde eine halbtägige forschend-lernende Unterrichtsintervention zum Thema Wärmeisolierung durchgeführt. Die Lernmaterialien, mit denen die Schüler*innen während der Intervention arbeiteten, basierten auf den Erkenntnissen der Dissertation von Schmidt (2016) und zwei Masterarbeiten von Lui (2021) und Scholtes (2021) und sind nach dem PS-I Ansatz gegliedert.

In einem randomisierten 2x2-faktoriellen Design (vgl. Tabelle 2) wurden die Faktoren „eigenes Experiment“ und „vorgegebenes Experiment“ variiert, was zu den vier Gruppen (1) Lernen aus fremden Fehlern, (2) Lernen aus eigenen Fehlern (3) Lernen aus eigenen und fremden Fehlern und (4) Keine Fehler führte. Die teilnehmenden Schüler*innen wurden den Gruppen randomisiert zugeteilt.

Tabelle 2 2x2 - Faktoren Versuchsdesign nach Hämmerle (in prep.)

	vorgegebener Versuch	
<i>Eigener Versuch</i>	Keine fremden Fehler	Fremde Fehler
<i>Keine eigenen Fehler</i>	keine Fehler (4)	Lernen aus fremden Fehlern (1)
<i>Eigene Fehler</i>	Lernen aus eigenen Fehlern (2)	Lernen aus eigenen und fremden Fehlern (3)

Die Schüler*innen in der Gruppe 1 „Lernen aus fremden Fehlern“ arbeiteten ausschließlich mit einem fehlerhaften vorgegebenen Experiment. Die Schüler*innen aus der Gruppe 2 „Lernen aus eigenen Fehlern“ experimentieren ausschließlich selbst. Schüler*innen der Gruppe 3 „Lernen aus eigenen und fremden Fehlern“ experimentierten selbst und arbeiteten mit einem vorgegebenen, fehlerhaften Experimentdesign, während die Gruppe „keine Fehler“ keinen Fokus auf Lernen aus Fehlern hatte und somit theoretisch anhand einer Musterlösung experimentiert. Gruppen 1 – 3 erhielten nach fertiger Bearbeitung eine Musterlösung, mit welcher sie ihre Ergebnisse und das Experimentdesign vergleichen konnten. Direkt anschließend an die Intervention erfolgte vor Ort der Posttest. Vier bis sechs Wochen nach der Intervention wurde der Follow-Up Test in der Schule durchgeführt. Die einzelnen Schüler*innen erhielten jedoch keinen Test zweimal.

In der Analyse der Masterarbeit werden im Prätest alle vier Gruppen miteinbezogen. Im Posttest jedoch nur die Gruppen „Lernen aus eigenen Fehlern“ und „Lernen aus fremden Fehlern“, wodurch sich die Schüler*innenanzahl für den zweiten Teil der Analyse (Forschungsfrage 3) auf n= 142 reduziert.

7.3 Ablauf der Intervention

Der Aufbau der Materialien für die Intervention gestaltete sich bei den Gruppen „Lernen aus eigenen Fehlern“ und „Lernen aus fremden Fehlern“ gleich, jedoch mit dem Unterschied, dass die Gruppe „fremde Fehler“ mit einem bereits vorgegebenen Versuchsdesign arbeitete und im Anschluss daran die Fehler des Experiments aufzeigen musste, während die Gruppe „eigene Fehler“ selbst ein Experiment planen musste. Durch Selbstkontrolle mittels eines weiteren Arbeitsblattes konnten alle Gruppen ihre Ergebnisse kontrollieren. Die weiterführenden Arbeitsblätter boten dabei einerseits die Möglichkeit, sich erneut mit dem Experimentdesign auseinanderzusetzen und andererseits konnten die Fehler durch Vergleichen der korrekten Lösungen mit dem eigenen Experimentaufbau gefunden und genau bestimmt werden.

Die Intervention war einerseits die Mechanismen von Productive Failure (Kapur & Bielaczyc 2012) angelehnt und andererseits wurden auch die Ergebnisse von Sinha und Kapur (2019) darüber, warum produktives Scheitern auch scheitern kann, bei der Planung berücksichtigt. Bei der Intervention wurde das Vorwissen der Schüler*innen aktiviert (Kapur & Bielaczyc 2012), indem eine kurze Einführung und Erklärung des theoretischen Hintergrunds des Experiments gegeben wurde. So wurde beispielsweise erklärt, weshalb Vögel bei Kälte mehr Daunenfedern wachsen oder dass Wolle die Schafe vor der Kälte schützt. Indem den Schüler*innen anschließend eine Zusammenstellung eines möglichen Experimentaufbaus gegeben wurde, wurde die vierte Qualität von Productive Failure nach Kapur & Bielaczyc (2012) berücksichtigt. Ein Parameter, welcher bei der Intervention berücksichtigt wurde, war dass Schüler*innen die Möglichkeit erhalten haben, ihre Ergebnisse zu vergleichen und zu paraphrasieren (Sinha & Kapur 2021). Indem sie ihre Hypothesen ausformulieren und begründen mussten, konnte dies gewährleistet werden.

Ebenso konnten einige Parameter vermieden werden, die laut Sinha und Kapur (2019) dazu führen können, dass kein positiver Lerneffekt von Lernen aus eigenen Fehlern gegeben ist. Hierbei ist es wichtig, dass kein zu großer zeitlicher Abstand zwischen der Problemlösung und dem Unterricht, bzw. der Erklärung liegt. Im Falle der Intervention wurde direkt anschließend an die Problemlösungsphase die korrekte Lösung präsentiert. Ebenso wurde darauf geachtet, dass keine negativen emotionalen Reaktionen aufgrund von wiederholtem Scheitern (Sinha & Kapur 2019) bei Schüler*innen auftreten können. Dies konnte gewährleistet werden, indem Schüler*innen einerseits direkt anschließend die korrekten Lösungen erhielten und andererseits indem sie sich nur einmal (und nicht wiederholt) mit einer überschaubaren Anzahl an Beispielen beschäftigen mussten. Auch wurde die Intervention so konzipiert, dass die Problemlöseaktivitäten genaue Anweisungen enthalten, die spezifische Kontexte bereitstellen, die von Schüler*innen gelöst werden können (Sinha & Kapur 2019).

7.4 Testinstrument

7.4.1 VKS-Testheft

Die Betrachtung bestehender VKS-Testinstrumente zeigte einerseits, dass viele Instrumente nur einige wenige Subkompetenzen der VKS miteinbeziehen und andererseits, dass manche Messformate vorherrschen, um bestimmte Subkompetenzen abzubilden. Da dies darauf hindeutet, dass die bislang ermittelten Schüler*innenergebnisse hinsichtlich der VKS womöglich nur eine begrenzte Aussagekraft haben könnten, wurde von Schwichow et al. (2016) ein multiple-choice Instrument (CVSI) entwickelt, welches die einzelnen Subkompetenzen von Schüler*innen erfassen soll. Jenes Format wurde in leicht abgewandelter Form in einer Folgestudie von Schwichow et al. (2022) verwendet.

Um die VKS-Fähigkeit von Schüler*innen zu bewerten, wurde das CVSI-Testinstrument in weiterer Folge von Schwichow et al. (in prep.) adaptiert. In Zusammenarbeit wurde ein neues, erweitertes Testinstrument D-CVSI, bestehend aus insgesamt 36 Items, erstellt. Die D-CVSI Testhefte bestehen aus vier Domänen (Biologie, Chemie, Physik, Allgemeinwissen) und jeweils 2 Kontexten pro Domäne, wobei es pro Domäne jeweils eine Aufgabe für jede Subkompetenz gab. Um die Testzeit für die Schüler*innen überschaubar zu halten, wurden die Testhefte für die Interventionsstudie von Hämmerle (in prep.) mittels Anker Items (Boone, 2016) gekürzt. Das Testheft enthielt anschließend vier Anker Items pro Kontext. Um eine Überschneidung bzw. Wiederholung zu vermeiden, wurden zusätzlich zwei neue Kontexte, für Chemie und für Physik, erstellt. Schlussendlich wurde das Testheft dahingehend adaptiert, dass es drei verschiedene Testhefte (A, B, C) gab, die jeweils einen Kontext pro Domäne enthielten. Insgesamt bestehen die gekürzten Testhefte aus 16 Items

Die gekürzten Testinstrumente beinhalten die Subkompetenzen ID, IN UN und PL sowie jeweils pro Subkompetenzen die Fachkontexte Biologie, Chemie, Physik und Allgemeinwissen (vgl. Tabelle 3). In den Planungs- (PL) und Identifikationsaufgaben (ID) analysierten die Schüler*innen den Einfluss einer unabhängigen Variable auf eine abhängige Variable. Bei ID-Aufgaben mussten sie aus vier dargestellten Experimenten das passendste zur Hypothesenprüfung auswählen, wobei die Experimente sowohl grafisch als auch schriftlich beschrieben waren. Die Auswahl enthält verschiedene Arten von Experimenten, einschließlich solcher mit konfundierten Variablen und solcher, die die falsche unabhängige Variable variieren. (siehe Abbildung 3).

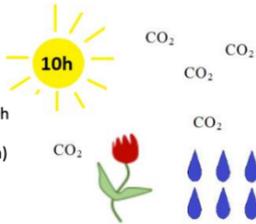
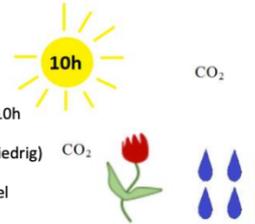
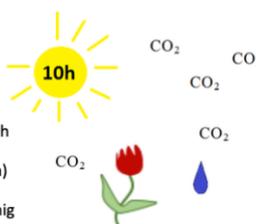
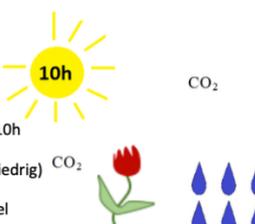
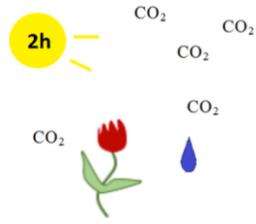
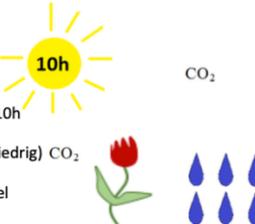
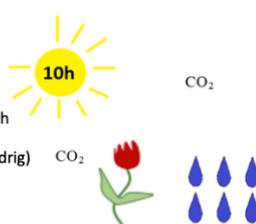
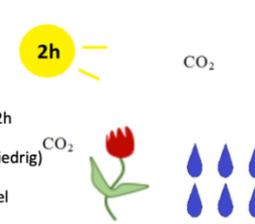
Photosynthese	ID-BIO-2
<p>Bianca möchte überprüfen, ob die Konzentration von Kohlenstoffdioxid (CO₂) einen Einfluss auf die Produktmengen von Sauerstoff (O₂) und Traubenzucker (C₆H₁₂O₆) besitzt. Sie vermutet, dass mehr Sauerstoff (O₂) und Traubenzucker (C₆H₁₂O₆) gebildet werden, wenn die Tulpe einer hohen Konzentration von Kohlenstoffdioxid (CO₂) ausgesetzt ist. Mit welchem Experiment kann sie ihre Vermutung überprüfen?</p>	
Experiment 1	
<div style="border: 1px solid black; padding: 5px; width: fit-content; margin-bottom: 10px;">1a</div> <input type="checkbox"/> <p>Sonnenstunden: 10h c[CO₂] = 0,1% (hoch) Wassermenge: viel</p> 	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin-bottom: 10px;">1b</div> <input type="checkbox"/> <p>Sonnenstunden: 10h c[CO₂] = 0,03% (niedrig) Wassermenge: viel</p> 
Experiment 2	
<div style="border: 1px solid black; padding: 5px; width: fit-content; margin-bottom: 10px;">2a</div> <input type="checkbox"/> <p>Sonnenstunden: 10h c[CO₂] = 0,1% (hoch) Wassermenge: wenig</p> 	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin-bottom: 10px;">2b</div> <input type="checkbox"/> <p>Sonnenstunden: 10h c[CO₂] = 0,03% (niedrig) Wassermenge: viel</p> 
Experiment 3	
<div style="border: 1px solid black; padding: 5px; width: fit-content; margin-bottom: 10px;">3a</div> <input type="checkbox"/> <p>Sonnenstunden: 2h c[CO₂] = 0,1% (hoch) Wassermenge: wenig</p> 	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin-bottom: 10px;">3b</div> <input type="checkbox"/> <p>Sonnenstunden: 10h c[CO₂] = 0,03% (niedrig) Wassermenge: viel</p> 
Experiment 4	
<div style="border: 1px solid black; padding: 5px; width: fit-content; margin-bottom: 10px;">4a</div> <input type="checkbox"/> <p>Sonnenstunden: 10h c[CO₂] = 0,03% (niedrig) Wassermenge: viel</p> 	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin-bottom: 10px;">4b</div> <input type="checkbox"/> <p>Sonnenstunden: 2h c[CO₂] = 0,03% (niedrig) Wassermenge: viel</p> 

Abbildung 3 ID-Aufgabe des gekürzten Testheftes (Hämmerle in prep.)

Bei den PL-Aufgaben planten die Schüler*innen selbst ein Experiment, indem sie die Variablen in zwei Telexperimenten festlegten. Insgesamt mussten dabei sechs Entscheidungen

getroffen werden, wobei es mehrere richtige Lösungen gab, je nachdem, wie die abhängigen Variablen variiert wurden (siehe Abbildung 4).

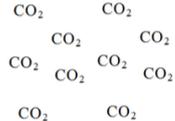
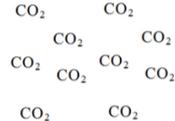
Photosynthese		PL-BIO-2	
Plane ein Experiment, bei dem sich prüfen lässt, ob die Anzahl an Sonnenstunden die Menge von Sauerstoff (O ₂) und Traubenzucker (C ₆ H ₁₂ O ₆) beeinflusst.			
Kreuze hierfür bei „Umgebung 1“ und „Umgebung 2“ an, wie du die Anzahl an Sonnenstunden, die Menge an Wasser (H ₂ O) und die Konzentration von Kohlenstoffdioxid (CO ₂) wählen würdest.			
Umgebung 1		Umgebung 2	
<input type="checkbox"/>  <input type="checkbox"/>  Zehn Sonnenstunden	<input type="checkbox"/>  Zwei Sonnenstunden	<input type="checkbox"/>  <input type="checkbox"/>  Zehn Sonnenstunden	<input type="checkbox"/>  Zwei Sonnenstunden
<input type="checkbox"/>  Hohe Konzentration c[CO ₂] = 0,1%	<input type="checkbox"/>  Niedrige Konzentration c[CO ₂] = 0,03%	<input type="checkbox"/>  Hohe Konzentration c[CO ₂] = 0,1%	<input type="checkbox"/>  Niedrige Konzentration c[CO ₂] = 0,3%
<input type="checkbox"/>  Viel Wasser	<input type="checkbox"/>  Wenig Wasser	<input type="checkbox"/>  Viel Wasser	<input type="checkbox"/>  Wenig Wasser

Abbildung 4 PL-Aufgabe des gekürzten Testheftes (Hämmerle in prep.)

Die IN und UN Aufgaben präsentierten den Schüler*innen das Ergebnis eines Experiments, das sie interpretieren mussten, wobei IN-Aufgaben auf kontrollierten (siehe Abbildung 5) und UN-Aufgaben (siehe Abbildung 6) auf konfundierten Experimenten basierten. Die UN-Aufgaben

erforderten demnach, dass Schüler*innen die fehlende Aussagekraft des gezeigten Experiments erkennen.

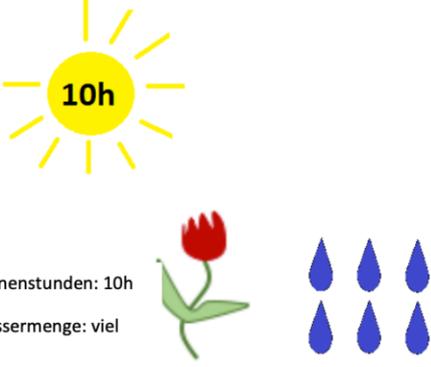
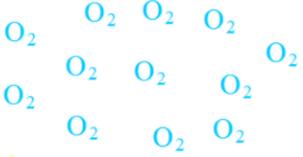
Photosynthese	IN-BIO-2
Bianca hat folgendes Experiment durchgeführt:	
 <p>Sonnenstunden: 10h Wassermenge: wenig</p>	 <p>Sonnenstunden: 10h Wassermenge: viel</p>
Sie hat folgende Beobachtung gemacht:	
 <p>$O_2 < 10g \text{ pro } m^2$ (niedrig)</p>	 <p>$O_2 > 10g \text{ pro } m^2$ (hoch)</p>
Was zeigt dieses Experiment?	
<input type="checkbox"/> Die Menge an Sonnenstunden hat einen Einfluss auf die Menge von Sauerstoff (O_2).	
<input type="checkbox"/> Die Menge von Wasser (H_2O) hat einen Einfluss auf die Menge von Sauerstoff (O_2).	
<input type="checkbox"/> Sowohl die Menge an Sonnenstunden als auch die Menge von Wasser (H_2O) haben einen Einfluss auf die Menge von Sauerstoff (O_2).	
<input type="checkbox"/> Das Experiment lässt keine sichere Schlussfolgerung zu.	

Abbildung 5 IN-Aufgabe des gekürzten Testheftes (Hämmerle in prep.)

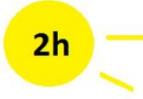
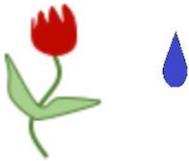
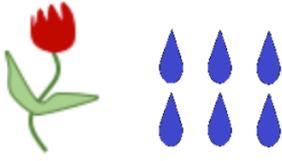
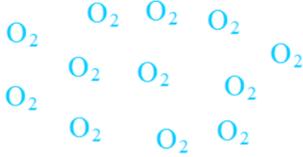
Photosynthese	UN-BIO-2
Bianca hat folgendes Experiment durchgeführt:	
 Sonnenstunden: 2h Wassermenge: wenig 	 Sonnenstunden: 10h Wassermenge: viel 
Sie hat folgende Beobachtung gemacht:	
 $O_2 < 10g \text{ pro } m^2$ (niedrig)	 $O_2 > 10g \text{ pro } m^2$ (hoch)
Was zeigt dieses Experiment?	
<input type="checkbox"/> Die Menge an Sonnenstunden hat einen Einfluss auf die Menge von Sauerstoff (O_2).	
<input type="checkbox"/> Die Menge von Wasser (H_2O) hat einen Einfluss auf die Menge von Sauerstoff (O_2).	
<input type="checkbox"/> Sowohl die Menge an Sonnenstunden als auch die Menge von Wasser (H_2O) haben einen Einfluss auf die Menge von Sauerstoff (O_2).	
<input type="checkbox"/> Das Experiment lässt keine sichere Schlussfolgerung zu.	

Abbildung 6 UN-Aufgabe des gekürzten Testheftes (Hämmerle in prep.)

Die Aufgaben werden in einer festgelegten Reihenfolge (vgl. Tabelle 3) präsentiert, wobei die Fächerreihenfolge in den einzelnen Testheften variiert wurde, um Reihenfolgeeffekte zu minimieren. Die verwendeten Kontexte wurden an den Wissensstand der Sekundarstufe I angepasst. Für eine detaillierte Übersicht über die Fachkontexte und die Testhefte vgl. Tabelle 4.

Die exakten Testhefte, die für diese Studie verwendet wurden, können auf Nachfrage ausgegeben werden.

Tabelle 3 Übersicht über die Themen je Testheft und Fachkontext

Testheft A	Testheft B	Testheft C
Photosynthese (Biologie)	Reaktionsgeschwindigkeit	Pendel (Physik)
Wasser kühlen (Physik)	(Chemie)	Muffins (Allgemeinwissen)
Besen (Allgemeinwissen)	Asseln (Biologie)	Asseln (Biologie)
pH-Wert (Chemie)	Besen (Allgemeinwissen)	Leitfähigkeit (Chemie)
	Starker Strom (Physik)	

Tabelle 4 Fachkontexte (Biologie, Physik, Chemie, Alltagswissen) in den drei Testheften
Die Nummer gibt die Position im Testheft an. Tabelle verändert nach Hämmerle, L. (in prep).

Kontext		Variablen	Test A	Test B	Test C
<i>Photosynthese</i>	(bio)	Bio2	1	-	-
<i>Lebensraum v. Isopoden</i>	(bio)	Bio1	-	2	3
<i>Wasserkühlung</i>	(phy)	Phy1	2	-	-
<i>Starkstrom</i>	(phy)	Phy2	-	4	-
<i>Pendulum</i>	(phy)	Phy4	-	-	1
<i>PH-Wert</i>	(che)	Che3	4	-	-
<i>Reaktionsgeschwindigkeit</i>	(che)	Che1	-	1	-
<i>Leitfähigkeit</i>	(che)	Che2	-	-	4
<i>Besen</i>	(all)	All1	3	3	-
<i>Muffins backen</i>	(all)	All2	-	-	2

7.4.2 Messung der Designfehler

Tabelle 5 zeigt eine Übersicht über die möglichen Designfehler je nach Subkompetenz. Der Designfehler *hold-one-thing-at-a-time* wurde nicht als eigenständiger Fehler berücksichtigt. Es wäre grundsätzlich möglich gewesen, die zugrunde liegende Schüler*innenvorstellung in den PL Aufgaben zu identifizieren. Dabei wären jene Schüler*innen, die die unabhängige Variable konstant gehalten hätten, während alle anderen Variablen variiert worden wären, in dieses Vorstellungsmuster gefallen. Das Ergebnis, im weiteren Sinne der Designfehler, ist ein konfundiertes Experiment und somit gleichzusetzen mit Missachtung konfundierter Variablen (einfach und zweifach). Tschirgi (1980) beschrieb jenen Designfehler als Sonderform konfundierter Experimente, weil die zugrundeliegende Schüler*innenvorstellung ist, dass jene einen Variable gefunden werden muss, die konstant gehalten werden muss, um dieselben Ergebnisse unter veränderten Bedingungen zu erhalten.

Die PL Aufgaben in den Testheften sind jedoch schon so konzipiert, dass Schüler*innen wissen, welche Variable sie untersuchen müssen, indem beispielsweise die Hypothese aufgestellt

wird, dass Muffins bei Unterhitze schneller backen, als bei Umlufthitze. Da die Schüler*innen im Anschluss daran zwei Experimente mit insgesamt drei Bedingungen planen müssen, fällt die Suche nach der zu konstant haltenden Variable bereits weg.

Ebenso konnte der Designfehler „Versuch mit nur einer Bedingung“ nicht identifiziert werden, da es das Antwortformat der Testhefte nicht zugelassen hätte. In jeder Subkompetenz erhielten die Schüler*innen jeweils Experimente mit zwei vorgegebenen Bedingungen zur Auswahl. Die PL Aufgaben waren so konstruiert, dass die Schüler*innen auch darüber instruiert wurden, ein Experiment mit zwei Bedingungen zu planen. Somit wurde dieser Fehler aufgrund des Aufgabenformats ausgeschlossen und kann nicht in die Analyse miteinbezogen werden.

Tabelle 5 mögliche Designfehler je Subkompetenz

Designfehler	Abkürzung	ID	PL	IN	UN
<i>Korrektes Verständnis der VKS</i>	cvS	x	x	x	x
<i>Kontrolliertes Experiment für falsche Variable</i>	cwv	x	x		
<i>Konfundiertes Experiment – zwei Variablen verändert</i>	dcv1	x	x	x	x
<i>Konfundiertes Experiment – drei Variablen verändert</i>	dcv2	x	x	x	x
<i>Nicht vergleichbare Experimente</i>	nce	x	x	x	
<i>Abhängige Variable konstant</i>	ncv	x	x		

Die Schüler*innenvorstellung *Versuch mit nur einer Bedingung* bzw. der Designfehler *Planung nur einer Versuchsbedingung* konnten mit dieser Studie nicht überprüft werden, da die Testhefte so aufgebaut waren, dass in jeder Subkompetenz zwei Versuchsbedingungen gefordert waren.

7.5 Datenbereinigung und Kodierung

Zunächst wurden die einzelnen Fachkontexte zusammengefasst. Die in Tabelle 4 beschriebenen Variablen: Bio1 und Bio2 wurden zu Bio, Phy1, Phy2, Phy 4 zu Phy, Che1, Che2, Che3 zu Che und All1, All2 zu All zusammengefasst, sodass insgesamt nur noch die vier Variablen (Bio, Phy, Che, All) bestehen blieben. Dies diente der Vereinfachung und übersichtlichen Darstellung der Daten und hat keinerlei Einfluss auf die weiteren Berechnungen. Die Kodierung erfolgte mittels einer Kodieranleitung von Schwichow et al. (2022), die von Hämmerle (in prep.) um die neu hinzugefügten Items der Fachkontexte Biologie und Allgemeinwissen ergänzt wurde. In der Kodieranleitung wurden für die Subkompetenzen IN, UN und ID eine eigene Übersichtstabelle erstellt, die die Antwortoptionen der Schüler*innen den damit einhergehenden Schüler*innenvorstellungen zuordnete. Bei den PL Aufgaben wurden zunächst die möglichen Kombinationen je Schüler*innenvorstellung ermittelt, um die Kombination anschließend den

Schüler*innenantworten zuordnen und Rückschlüsse auf die zugrundeliegende Vorstellung ziehen zu können. Die verwendete Kodieranleitung kann auf Nachfrage zur Verfügung gestellt werden. Die Kodierung erfolgte in Excel und SPSS. Nachfolgend wurden die Daten in SPSS von einem Wide-Format in ein Long-Format transformiert, um die weiterfolgenden Berechnungen durchführen zu können.

7.6 Statistische Analyse

Um Häufigkeiten von Designfehlern pro Subkompetenz zu ermitteln, wurden als ersten Schritt alle Designfehler für die vier Subkompetenzen ermittelt und anteilmäßig zur Gesamtzahl der Designfehler in Excel berechnet. Zu den Fehlerhäufigkeiten wurden jeweils absolute Häufigkeiten und Prozentwerte berechnet. Die Ergebnisse wurden tabellarisch dargestellt. Die weiteren Daten wurden mit IBM® SPSS® Statistics, Version 27 für iOS ausgewertet. Für jede Person lagen wiederholte Messungen vor, daher handelte es sich um einen Datensatz mit 266 Clustern, deren Werte von der befragten Person abhängig waren. Um der Abhängigkeitsstruktur zu entsprechen, wurde ein verallgemeinertes gemischtes lineares Regressionsmodell gerechnet, indem der Schüler*innen-Hauptcode als Zufallseffekt berücksichtigt wurde. Im Hinblick auf den Umgang mit fehlenden Werten wurde entschieden, diese im Modell zu belassen. Jedoch wurde im Anschluss eine Sensitivitätsanalyse durchgeführt, die eine Ratewahrscheinlichkeit von 20% annimmt. Die 20% wurden deshalb angenommen, weil Schüler*innen bei vier Antwortmöglichkeiten pro Item in ID, IN und UN jeweils eine 25%ige Wahrscheinlichkeit haben, das korrekte Ergebnis anzukreuzen (Schwichow et al. In prep.). Da bei PL Aufgaben sechs Antwortmöglichkeiten bestehen wurden die 25% auf 20% abgerundet.

Die Zielvariable war binär kodiert mit den Ausprägungen „Fehler“ und „kein Fehler (cvs)“. Infolge wurde als Verknüpfung zwischen dieser binären Zielvariable und dem linearen Modell eine logistische Funktion gewählt. Das gemischte Modell wurde zum einen für die Modellierung der Fehlerwahrscheinlichkeit vor der Intervention gerechnet. Dabei wurden die Fachkontexte und Subkompetenzen als kategorial skalierte Variablen als Prädiktoren (unabhängige Variable) eingesetzt. Zum anderen wurde ein weiteres verallgemeinertes gemischtes Modell für die Modellierung der Veränderung der Fehlerwahrscheinlichkeit nach einer Intervention geschätzt. In diesem Fall wurden als Prädiktoren der Zeitpunkt (Prä - Post) sowie der Zusammenhang von Interventionsgruppen (eigene und fremde Fehler) und Zeitpunkt verwendet.

Das verallgemeinerte gemischte lineare Modell wurde auch eingesetzt, um Muster von fehlenden Werten feststellen zu können. Zufälliges Auftreten von fehlenden Werten stellen keine Verzerrungen der Ergebnisse dar (MCAR, missing completely at random). Fehlende Werte, die auf Grundlage von Prädiktoren erklärbar sind, stellen ebenso keine Verzerrungen dar, wenn die betreffenden Prädiktoren in das Modell aufgenommen werden (MAR, missing at random). Eine Verzerrungen stellen systematische Muster für fehlende Werte dar, die nicht mit Hilfe der Prädiktoren erklärbar sind (MNAR, missing not at random).

Eine Irrtumswahrscheinlichkeit von 5% wurde verwendet, um die Signifikanz der Ergebnisse zu interpretieren. Bei post-hoc Analysen wurde das Signifikanzniveau nach der Methode von Bonferroni korrigiert, um eine Inflation der Irrtumswahrscheinlichkeit zu verhindern. Als Effektstärken wurden Odds Ratio errechnet und interpretiert. Eine Odds Ratio von 1 bedeutet, dass kein Unterschied beziehungsweise Zusammenhang festgestellt werden konnte.

8 Ergebnisse

8.1 Häufigkeit von Designfehlern

Nicht alle von Siler und Klahr (2012) identifizierten Designfehler traten dabei über alle Subkompetenzen in signifikanter Ausprägung (> 4%) auf. Die Signifikanz von 4% wurde von Schwichow et al. (2022) übernommen, da die Vorgehensweise hinsichtlich der Forschungsfragen ident war. Zwar konnten alle von Siler und Klahr (2012) beschriebenen Designfehler identifiziert werden, jedoch trat der Designfehler ncv in allen Subkompetenzen mit einer Häufigkeit von unter 3% auf. Der Designfehler ncv trat nur in der Subkompetenz ID in einer Häufigkeit von 7,2% auf und belief sich in der Subkompetenz PL auf unter 1%.

Nachfolgend wird eine detaillierte Übersicht der Designfehler pro Subkompetenz gegeben.

Identification (ID)

Rund 50% der Schüler*innen wandten die VKS in der Subkompetenz ID korrekt an. Im Durchschnitt traten bei 38,5% der Schüler*innen Designfehler bei der Subkompetenz ID auf (vgl. Tabelle 6). Der häufigste Designfehler bestand mit 14,8% aus einem *konfundierten Experiment*, begründet mit mehreren variierten Variablen. *Nicht kontrastive Experimente* wurden in der Subkompetenz nur von 1% begangen, daher gilt der Fehler als nicht signifikant. Ungefähr gleich viel Schüler*innen planten entweder *kontrollierte Experimente für die falsche Variable* (7,1%) oder identifizierten eine *nicht kausale Variable* (7,2%). Der zweithäufigste Designfehler bestand aus einem *einfach konfundierten Experiment* (8,3%).

Tabelle 6 Fehlerhäufigkeit in der Subkompetenz ID

Beschrieben wurden Häufigkeiten der korrekten VKS-Anwendung und der Designfehler des Prätests in der Subkompetenz ID, je Fachkontext (Biologie, Physik, Allgemeinwissen, Chemie). Die Häufigkeiten sind in absoluter und prozentueller Anzahl angegeben. Die Zahl in der Klammer stellt den prozentualen Anteil in Relation zu den Gesamtwerten (Designfehler innerhalb der Subkompetenz + korrekte Anwendung der VKS) dar.

Fehlertyp	ID_bio	ID_phy	ID_all	ID_che	Durchschnitt
Cvs (kein Fehler)	134 (50,4)	161 (60,5)	161 (60,5)	69 (25,9)	143 (49,3)
Fehler gesamt	103 (38,7)	74 (27,8)	78 (29,3)	155 (58,3)	103 (38,5)
cwv	8 (3,0)	26 (9,8)	0 (0,0)	42 (15,8)	19 (7,1)
dcv1	39 (14,7)	14 (5,3)	28 (10,5)	7 (2,6)	22 (8,3)
dcv2	45 (16,9)	15 (5,6)	34 (12,8)	64 (24,1)	40 (14,8)
ncv	0 (0,0)	19 (7,1)	16 (6,0)	42 (15,8)	19 (7,2)
nce	11 (4,1)	0 (0,0)	0 (0,0)	0 (0,0)	3 (1,0)
fehlend	29 (10,9)	31 (11,7)	27 (10,2)	42 (15,8)	32 (12,1)
Total	266 (100)	266 (100)	266 (100)	266 (100)	

Interpretation (IN)

Der häufigste Designfehler (10,7%) der Subkompetenz IN (vgl. Tabelle 7) bestand aus *mehreren variierten Variablen*, wobei hier aufgrund der Aufgabenstellung keine Unterscheidung in einfache oder doppelte Variation getroffen werden konnte. *Nicht kontrastive Experimente* wurden von nur 1,6% der Schüler*innen interpretiert. Als Error wurden jene Antworten kodiert, die aussagten, dass die Variable zwei einen Einfluss auf die unabhängige Variable hätte, siehe Abbildung 5 zur Veranschaulichung. Aufgrund von nicht sinnvoll möglicher Interpretationsmöglichkeiten, wurde diese Antwort laut Kodieranleitung als „falsch“ angenommen.

Tabelle 7 Fehlerhäufigkeit in der Subkompetenz IN

Beschrieben wurden Häufigkeiten der korrekten VKS-Anwendung und der Designfehler des Prätests in der Subkompetenz IN, je Fachkontext (Biologie, Physik, Allgemeinwissen, Chemie) Die Häufigkeiten sind in absoluter und prozentueller Anzahl angegeben. Die Zahl in der Klammer stellt den prozentualen Anteil in Relation zu den Gesamtwerten (Designfehler innerhalb der Subkompetenz + korrekte Anwendung der VKS) dar.

Fehlertyp	IN_bio	IN_phy	IN_all	IN_che	Durchschnitt
Cvs (kein Fehler)	211 (79,3)	196 (73,7)	198 (74,4)	181 (68)	197 (73,9)
Fehler gesamt	38 (14,3%)	48 (18,0)	54 (20,3)	52 (19,5)	48 (18,0)

Fehlertyp	IN_bio	IN_phy	IN_all	IN_che	Durchschnitt
error	9 (3,4)	13 (4,9)	17 (6,4)	22 (18,3)	22 (5,7)
nce	2 (0,8)	5 (1,9)	2 (0,8)	8 (3,0)	8 (1,6)
tmv	27 (10,2)	30 (11,3)	35 (13,2)	22 (8,3)	22 (10,7)
fehlend	17 (6,4)	22 (8,3)	14 (5,3)	33 (12,4)	33 (8,1)
Total	266 (100)	266 (100)	266 (100)	266 (100)	

Understanding (UN)

Die Subkompetenz UN wies von allen Subkompetenzen die meisten Designfehler (86,7%) auf (vgl. Tabelle 8). Da in dieser Subkompetenz das Verständnis der fehlenden Aussagekraft eines konfundierten Experiments untersucht wurde, war nur die Antwort „Das Experiment lässt keine sichere Schlussfolgerung zu“ als korrektes VKS Verständnis zu werten. Der Designfehler dcv1 (*einfach kondundiertes Experiment*) bedeutet in diesem Fall, dass Schüler*innen entweder Variable 1 oder Variable 2 als Einflussfaktor auf die unabhängige Variable sehen und wurde von 24,8% der Schüler*innen begangen. Hier wurde nicht zwischen den Variablen differenziert. *Testung mehrerer Variablen* gründet auf dem Verständnis, dass sowohl Variable 1 als auch Variable 2 einen Einfluss auf die unabhängige Variable haben und haben somit ein *zweifach kondundiertes Experiment* zur Folge. Demnach glaubten die meisten Schüler*innen (51,8%), dass beide Variablen einen Einfluss auf die unabhängige Variable haben.

Tabelle 8 Fehlerhäufigkeit in der Subkompetenz UN

Beschrieben wurden Häufigkeiten der korrekten VKS-Anwendung und der Designfehler des Prätests in der Subkompetenz UN, je Fachkontext (Biologie, Physik, Allgemeinwissen, Chemie). Die Häufigkeiten sind in absoluter und prozentueller Anzahl angegeben. Die Zahl in der Klammer stellt den prozentualen Anteil in Relation zu den Gesamtwerten (Designfehler innerhalb der Subkompetenz + korrekte Anwendung der VKS) dar.

Fehlertyp	UN_bio	UN_phy	UN_all	UN_che	Durchschnitt
Cvs (kein Fehler)	53 (19,9)	42 (15,8)	25 (7,9)	21 (7,9)	35 (13,3)
Fehler gesamt	188 (80,1)	200 (84,2)	219 (90,6)	208 (92,1)	204 (86,7)
dcv1	59 (22,2)	64 (24,1)	70 (26,7)	71 (26,7)	66 (24,8)
tmv	129 (48,5)	136 (51,1)	149 (56,0)	137 (51,5)	138 (51,8)
fehlend	25 (9,4)	24 (9,0)	22 (13,9)	37 (13,9)	27 (10,2)

Fehlertyp	UN_bio	UN_phy	UN_all	UN_che	Durchschnitt
Total	266 (100)	266 (100)	266 (100)	266 (100)	

Planing (PL)

Bei 48,6% der Schüler*innen traten Designfehler in der Subkompetenz PL auf (vgl. Tabelle 9) Mit Abstand der häufigste Designfehler (28,5%) bestand aus der Variation aller gegebenen Variablen. Nur eine zusätzliche Variable wurde von 9% der Schüler*innen variiert, während 8,5% der Schüler*innen ein kontrolliertes Experiment für die falsche Variable planten. Kaum ein*e Schüler*in plante ein nicht kontrastives Experiment oder plante für eine nicht kontrastive Zielvariable (2,4% bzw. 0,2%).

Tabelle 9 Fehlerhäufigkeit in der Subkompetenz PL

Beschrieben wurden Häufigkeiten der korrekten VKS-Anwendung und der Designfehler des Prätests in der Subkompetenz PL, je Fachkontext (Biologie, Physik, Allgemeinwissen, Chemie) Die Häufigkeiten sind in absoluter und prozentueller Anzahl angegeben. Die Zahl in der Klammer stellt den prozentualen Anteil in Relation zu den Gesamtwerten (Designfehler innerhalb der Subkompetenz + korrekte Anwendung der VKS) dar.

Fehlertyp	PL_bio	PL_phy	PL_all	PL_che	Durchschnitt
Cvs (kein Fehler)	119 (44,7)	117 (44,3)	147 (55,3)	91 (34,2)	119 (44,6)
Fehler gesamt	125 (48,5)	120 (48,1)	105 (42,1)	139 (55,6)	122 (48,6)
cwv	9 (3,4)	35 (13,3)	14 (5,3)	32 (12,0)	23 (8,5)
dcv1	16 (6,0)	22 (8,3)	29 (10,9)	29 (10,9)	24 (9,0)
dcv2	100 (37,6)	63 (23,9)	62 (23,3)	78 (29,3)	76 (28,5)
nce	4 (1,5)	7 (2,7)	6 (2,3)	8 (3,0)	6 (2,4)
ncv	0 (0,0)	0 (0,0)	1 (0,4)	1 (0,4)	1 (0,2)
fehlend	18 (6,8)	20 (7,6)	7 (2,6)	27 (10,2)	18 (6,8)
Total	266 (100)	264 (100)	266 (100)	266 (100)	

8.2 Muster von Designfehlern

8.2.1 Fehlerwahrscheinlichkeit in Abhängigkeit der Subkompetenzen

Um die Ausprägung der Designfehler zwischen den Subkompetenzen zu ermitteln, wurde ein gemischtes binär logistisches Regressionsmodell angewandt (vgl. Tabelle 10). Die Subkompetenz UN erwies sich dabei als jene mit den meisten Designfehlern. UN wurde somit als Vergleichswert genommen, um die Wahrscheinlichkeit, einen Fehler zu machen in den restlichen Subkompetenzen zu ermitteln. Es gab einen signifikanten Einfluss der Subkompetenz hinsichtlich der Wahrscheinlichkeit einen Fehler zu machen, $F(3, 3794) = 275,61$, $p < 0,001$. In den Subkompetenzen ID, PL und IN wurden signifikant weniger Fehler begangen im Vergleich zu UN ($p < 0,001$). Die odds ratios lagen zwischen 0,01 und 0,12; die Wahrscheinlichkeit einen Fehler zu begehen, gegenüber der Wahrscheinlichkeit keinen Fehler zu begehen, war bei ID um 92%, bei PL um 88% und bei IN um 99% gegenüber UN reduziert. UN war demnach jene Subkompetenz mit der höchsten Fehlerwahrscheinlichkeit in Abhängigkeit der Subkompetenzen (vgl. Abbildung 7). Post hoc Analysen (vgl. Tabelle 11) ergaben, dass sich alle Subkompetenzen untereinander unterschieden, $p < 0,001$.

Tabelle 10 Ergebnisse des gemischten Modells für die Fehlerwahrscheinlichkeit

Anmerkung: b = Regressionskoeffizient, SE(b) = Standardfehler von b, OR = odds ratio, 95%-KI = Konfidenzintervall; UN = Referenzkategorie, Schüler*innen-Hauptcode als Zufallseffekt

Prädiktor	b	SE(b)	Test	p-Wert	OR	95%-KI	
<i>Konstanter Term</i>	2,29	0,14	16,76	< 0,001			
<i>Subkompetenz</i>			275,61	<0,001			
<i>ID</i>	-2,52	0,13	-19,08	< 0,001	0,08	0,06	0,11
<i>PL</i>	-2,16	0,13	-16,68	< 0,001	0,12	0,09	0,15
<i>IN</i>	-4,42	0,15	-28,55	< 0,001	0,01	0,01	0,02
<i>UN</i>	0	.	.	.			

Tabelle 11 Post hoc Analyse mit paarweisen Vergleichen

* p-Wert korrigiert nach Bonferroni, df = Freiheitsgrade, SE = Standardfehler

Vergleich	Unterschied	SE	Test	df	p-Wert*
<i>PL – ID</i>	0,09	0,03	3,37	3794	<0,001
<i>IN – ID</i>	-0,34	0,03	-13,29	3794	<0,001
<i>IN – PL</i>	-0,43	0,03	-17,05	3794	<0,001
<i>UN – ID</i>	0,46	0,03	18,17	3794	<0,001
<i>UN – PL</i>	0,38	0,03	14,95	3794	<0,001
<i>UN – IN</i>	0,80	0,01	57,85	3794	<0,001

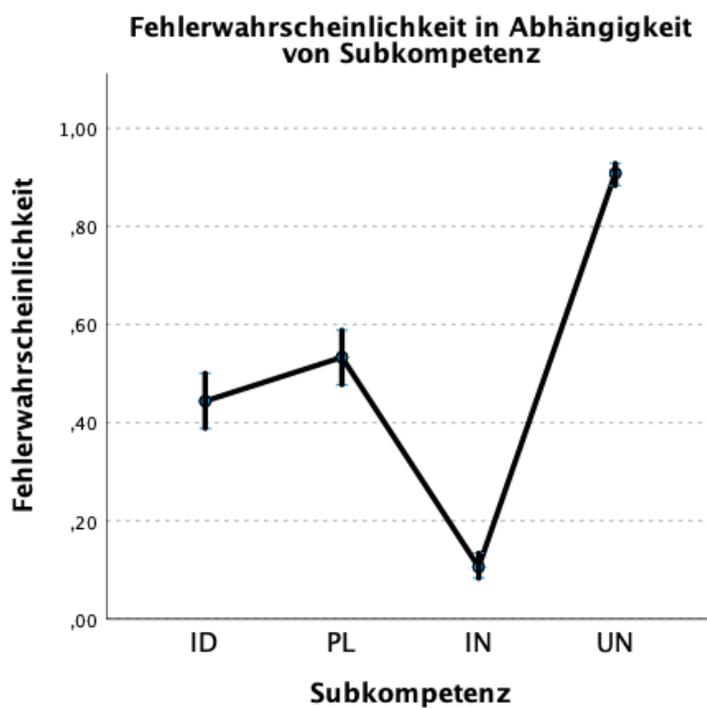


Abbildung 7 Fehlerwahrscheinlichkeit in Abhängigkeit der Subkompetenz

Eine Sensitivitätsanalyse (vgl. Tabelle 12) in der alle fehlenden Werte als falsch angenommen wurden, zeigte, dass die Unterschiede zwischen den Subkompetenzen weniger stark aber immer noch signifikant $p < 0,001$ ausgeprägt waren.

Tabelle 12 Sensitivitätsanalyse mit Annahme alle fehlenden Werte falsch

* p-Wert korrigiert nach Bonferroni, df = Freiheitsgrade, SE = Standardfehler

Vergleich	Unterschied	SE	Test	df	p-Wert*
<i>PL – ID</i>	0,06	0,02	2,62	4668	<0,001
<i>IN – ID</i>	-0,33	0,02	-13,63	4668	<0,001
<i>IN – PL</i>	-0,39	0,02	-16,84	4668	<0,001
<i>UN – ID</i>	0,36	0,03	13,07	4668	<0,001
<i>UN – PL</i>	0,30	0,03	11,50	4668	<0,001
<i>UN – IN</i>	0,69	0,02	31,45	4668	<0,001

Da eine gewisse Zufallswahrscheinlichkeit für richtige Antworten besteht, die Frage auch durch Raten richtig zu beantworten, wurde in einer zweiten Sensitivitätsanalyse eine 20%ige Zufallswahrscheinlichkeit angenommen. Dies zeigte, dass die Unterschiede zwischen den Subkompetenzen noch weniger stark sind als in der Ursprungsanalyse (vgl. Tabelle 13) UN und ID unterschieden sich auch in dieser Sensitivitätsanalyse am stärksten.

Tabelle 13 Sensitivitätsanalyse mit 20% Zufallswahrscheinlichkeit

* p-Wert korrigiert nach Bonferroni, df = Freiheitsgrade, SE = Standardfehler

Vergleich	Unterschied	SE	Test	df	p-Wert*
<i>PL – ID</i>	0,07	0,02	2,96	4668	<0,001
<i>IN – ID</i>	-0,27	0,02	-12,29	4668	<0,001
<i>IN – PL</i>	-0,34	0,02	-15,55	4668	<0,001
<i>UN – ID</i>	0,38	0,02	17,35	4668	<0,001
<i>UN – PL</i>	0,31	0,02	14,53	4668	<0,001
<i>UN – IN</i>	0,65	0,02	37,10	4668	<0,001

8.2.2 Fehlerwahrscheinlichkeit in Abhängigkeit der Fachkontexte

Für die Betrachtung der einzelnen Fachkontexte wurde ebenfalls ein binär logistisches Regressionsmodell herangezogen, um zu untersuchen, ob Fachkontexte, über alle Subkompetenzen hinweg, zu unterschiedlichen Schwierigkeiten führen.

Es gab einen signifikanten Einfluss der Fachkontexte hinsichtlich der Wahrscheinlichkeit, einen Fehler zu machen, $F(3, 3794) = 21,14$, $p < 0,001$ (vgl. Tabelle 14). In den Fachkompetenzen Biologie, Physik und Allgemeinwissen wurden im Vergleich zu Chemie signifikant weniger Fehler begangen ($p < 0,001$). Die odds ratios lagen zwischen 0,50 und 0,53; dies bedeutet, dass die Wahrscheinlichkeit einen Fehler zu begehen, gegenüber der Wahrscheinlichkeit, keinen Fehler zu begehen, bei Biologie, Physik und Chemie um etwa 50% reduziert war.

Chemie stellte demnach den Fachkontext mit der höchsten Fehlerwahrscheinlichkeit dar (vgl. Abbildung 8). Post hoc Analysen ergaben, dass es keinen signifikanten Unterschied zwischen Biologie, Physik und Allgemeinwissen gab (vgl. Tabelle 15).

Tabelle 14 Ergebnisse des gemischten Modells für die Fehlerwahrscheinlichkeit

b = Regressionskoeffizient, $SE(b)$ = Standardfehler von b , OR = odds ratio, 95%-KI= Konfidenzintervall; che = Referenzkategorie, Schüler*innen-Hauptcode als Zufallseffekt.

Prädiktor	b	SE(b)	Test	p-Wert	OR	95%-KI	
Konstanter Term	0,53	0,09	5,59	<0,001			
<i>Fachkontext</i>			21,14	<0,001			
bio	-0,64	0,10	-6,32	<0,001	0,53	0,43	0,64
phy	-0,67	0,10	-6,56	<0,001	0,51	0,42	0,63
all	-0,69	0,10	-6,85	<0,001	0,50	0,41	0,61
che	0	.	.	.			

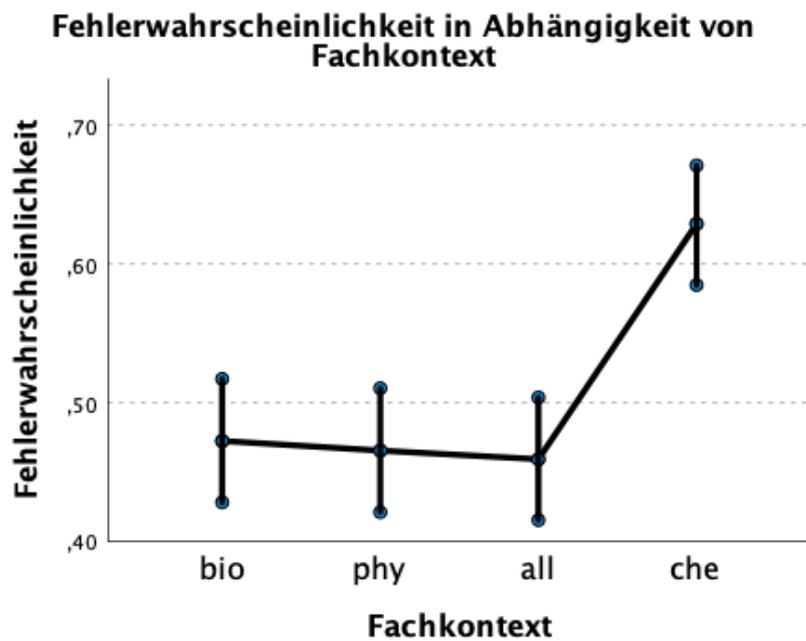


Abbildung 8 Fehlerwahrscheinlichkeit in Abhängigkeit des Fachkontexts

Tabelle 15 Post hoc Analysen mit paarweisen Vergleichen

* p-Wert korrigiert nach Bonferroni, df = Freiheitsgrade, SE = Standardfehler

Vergleich	Unterschied	SE	Test	df	p-Wert*
<i>phy - bio</i>	-0,01	0,03	-0,28	3794	1,000
<i>all - bio</i>	-0,01	0,02	-0,54	3794	1,000
<i>all - phy</i>	-0,01	0,02	-0,25	3794	1,000
<i>che - bio</i>	0,16	0,02	6,43	3794	<0,001
<i>che - phy</i>	0,16	0,02	6,68	3794	<0,001
<i>che - all</i>	0,17	0,02	6,99	3794	<0,001

Eine Sensitivitätsanalyse mit der Annahme, dass alle fehlenden Werte falsch beantwortet worden wären, zeigte, dass die Unterschiede zwischen den Fachkontexten auch unter dieser Annahme annähernd gleich waren, wie in der Post hoc Analyse mit paarweisen Vergleichen (vgl. Tabelle 16).

Tabelle 16 Post hoc Analysen mit paarweisen Vergleichen (alle auf null)

* p-Wert korrigiert nach Bonferroni, df = Freiheitsgrade, SE = Standardfehler

Vergleich	Unterschied	SE	Test	df	p-Wert*
<i>phy - bio</i>	0,001	0,02	0,05	4668	1,000
<i>all - bio</i>	-0,02	0,02	-0,66	4668	1,000
<i>all - phy</i>	-0,02	0,02	-0,71	4668	1,000
<i>che - bio</i>	0,16	0,02	7,42	4668	<0,001
<i>che - phy</i>	0,16	0,02	7,37	4668	<0,001
<i>che - all</i>	0,18	0,02	8,07	4668	<0,001

Eine weitere Sensitivitätsanalyse, in der eine Zufallswahrscheinlichkeit von 20% angenommen wurde, bestätigte das Ergebnis ein weiteres Mal (vgl. Tabelle 17).

Tabelle 17 Post hoc Analysen mit paarweisen Vergleichen (inkl. Zufall)

* p-Wert korrigiert nach Bonferroni, df = Freiheitsgrade, SE = Standardfehler

Vergleich	Unterschied	SE	Test	df	p-Wert*
<i>phy - bio</i>	0,001	0,02	0,05	4668	1,000
<i>all - bio</i>	-0,01	0,02	-0,58	4668	1,000
<i>all - phy</i>	-0,01	0,02	-0,63	4668	1,000
<i>che - bio</i>	0,15	0,02	7,12	4668	<0,001
<i>che - phy</i>	0,15	0,02	7,07	4668	<0,001
<i>che - all</i>	0,17	0,02	7,71	4668	<0,001

8.2.3 Fehlerwahrscheinlichkeit in Abhängigkeit der Wechselwirkung der Fachkontexte und Subkompetenzen

In einem dritten Modell wurde die Wechselwirkung von Fachkontexten und Subkompetenzen auf die Fehlerwahrscheinlichkeit betrachtet. Es gab eine signifikante Wechselwirkung, $F(9, 3782) = 8,64, p = 0,001$ (vgl. Tabelle 18).

Tabelle 18 Gemischtes lineares Modell für die Fehlerwahrscheinlichkeit unter Hinzunahme der Wechselwirkung von Fachkontexte und Subkompetenz

Quelle	F	df1	df2	p-Werte
<i>Korrigiertes Modell</i>	60,43	15	3782	0,001
<i>Fachkontexte</i>	20,18	3	3782	0,001
<i>Subkompetenzen</i>	272,34	3	3782	0,001
<i>Subkompetenzen * Fachkontexte</i>	8,64	9	3782	0,001

Chemie führte zu signifikant mehr Fehlern als alle anderen Fachkontexte, wenn die Subkompetenz ID und PL abgefragt wurde (vgl. Abbildung 9) Einen signifikanten Unterschied zwischen Physik und Biologie gab es lediglich bei ID. Einen signifikanten Unterschied zwischen Allgemeinwissen und Physik zeigte sich nur bei PL. Bei der Subkompetenz IN war die Fehlerwahrscheinlichkeit von Chemie vergleichbar mit allen anderen Kontexten und bei UN gab es keinen Unterschied zwischen Chemie und Allgemeinwissen, jedoch zwischen Chemie und Biologie sowie Chemie und Physik. Allgemeinwissen und Biologie unterschieden sich signifikant in allen Subkompetenzen außer IN (vgl. Tabelle 19).

Tabelle 19 Post hoc Analysen mit paarweisen Vergleichen

SE = Standardfehler, df = Freiheitsgrade

Subkompetenz	Fachkontexte	Unterschied	SE	Test	df	p-Wert
ID	<i>phy - bio</i>	-0,169	0,051	3,34	3782	0,003
	<i>all - bio</i>	-0,158	0,051	3,11	3782	0,004
	<i>all - phy</i>	0,011	0,046	-6,89	3782	0,810
	<i>che - bio</i>	0,336	0,049	-3,34	3782	<0,001

Subkompetenz	Fachkontexte	Unterschied	SE	Test	df	p-Wert
	<i>che - phy</i>	0,505	0,044	-0,24	3782	<0,001
	<i>che - all</i>	0,494	0,045	-11,39	3782	<0,001
<i>PL</i>	<i>phy - bio</i>	0,000	0,053	0,009	3782	0,993
	<i>all - bio</i>	-0,133	0,052	-2,563	3782	0,031
	<i>all - phy</i>	-0,133	0,052	-2,560	3782	0,031
	<i>che - bio</i>	0,140	0,051	2,741	3782	0,031
	<i>che - phy</i>	0,140	0,051	2,717	3782	0,031
	<i>che - all</i>	0,273	0,050	5,474	3782	<0,001
<i>IN</i>	<i>phy - all</i>	-0,002	0,028	-0,059	3782	1,000
	<i>all - bio</i>	0,032	0,025	1,275	3782	1,000
	<i>all - phy</i>	0,002	0,028	0,059	3782	1,000
	<i>che - bio</i>	0,024	0,025	0,930	3782	1,000
	<i>che - phy</i>	-0,007	0,028	-0,243	3782	1,000
	<i>che - all</i>	-0,008	0,028	-0,303	3782	1,000
<i>UN</i>	<i>phy - bio</i>	0,043	0,028	1,516	3782	0,259
	<i>all - bio</i>	0,092	0,026	3,525	3782	0,002
	<i>all - phy</i>	0,049	0,022	2,263	3782	0,071
	<i>che - bio</i>	0,100	0,026	3,879	3782	0,001
	<i>che - phy</i>	0,057	0,021	2,696	3782	0,028
	<i>che - all</i>	0,008	0,016	0,525	3782	0,600

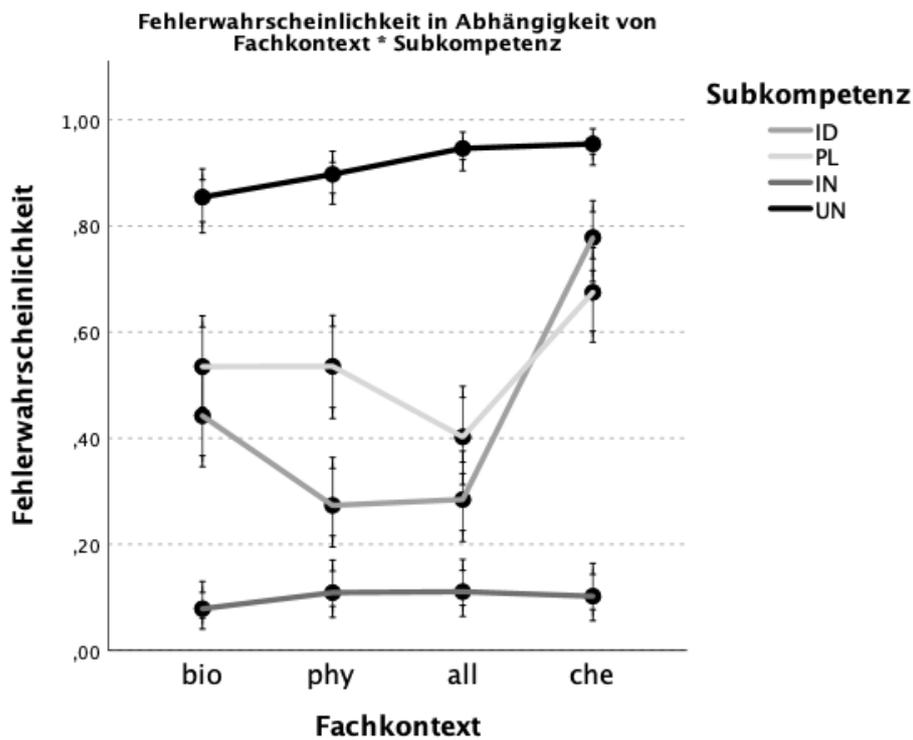


Abbildung 9 Fehlerwahrscheinlichkeit je Fachkontext in Abhängigkeit der Subkompetenz

Eine Analyse der fehlenden Werte ergab, dass Fachkontexte einen Einfluss darauf hatten, ob Schüler*innen keine Antwort gaben, $F(3,4668) = 12,141$, $p < 0,001$. Aus Abbildung 10 ist ersichtlich, dass bei Chemie die meisten fehlenden Werte vorhanden waren.

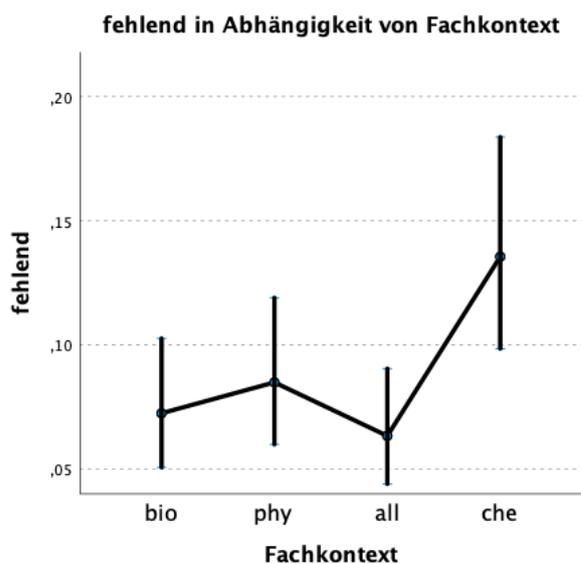


Abbildung 10 fehlende Werte je Fachkontext

Ebenso hatte die Subkompetenz einen signifikanten Einfluss auf fehlende Werte, $F(3, 4668) = 11,192$, $p < 0,001$. Aus Abbildung 11 ist ersichtlich, dass PL am wenigsten fehlende Werte aufwies, IN wies hingegen die meisten fehlenden Werte auf.

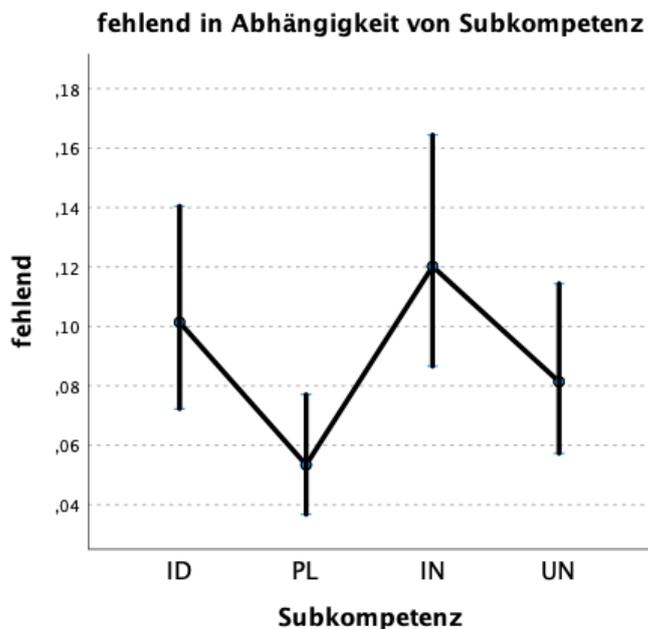


Abbildung 11 fehlende Werte je Subkompetenz

8.3 Veränderung der Fehlerwahrscheinlichkeit durch Lernen aus Fehlern

Ein gemischtes lineares Regressionsmodell zeigte, dass die Fehlerwahrscheinlichkeit nach der Intervention grundsätzlich niedriger war als vor der Intervention (vgl. Abbildung 12), $F(1, 4085) = 35,337$, $p < 0,001$. Gemäß odds ratio ist die Wahrscheinlichkeit, vor der Intervention Designfehler zu machen, um 52% höher als nach der Intervention (vgl. Tabelle 20).

Ob Schüler*innen besser aus eigenen Fehlern lernen und dementsprechend weniger Designfehler im Posttest machen, wurde geprüft, indem eine Wechselwirkung von Zeitpunkt und Gruppenzuteilung (eigene und fremde Fehler) in das binär logistische Regressionsmodell aufgenommen wurde. Dabei konnte kein signifikanter Wechselwirkungseffekt festgestellt werden, $F(2, 4083) = 2,014$, $p = 0,134$.

Tabelle 20 Ergebnisse des gemischten Modells für die Fehlerwahrscheinlichkeit je Zeitpunkt
 b = Regressionskoeffizient, SE(b) = Standardfehler von b, OR = odds ratio, 95%-KI= Konfidenzintervall; Schüler*innen-Hauptcode als Zufallseffekt.

Prädiktor	b	SE(b)	Test	p-Wert	OR	95%-KI
Konstanter Term	-0,38	0,10	-3,83	<0,001		
Zeitpunkt=1	0,42	0,07	5,94	<0,001	1,52	1,32 1,75
Zeitpunkt=2	0					

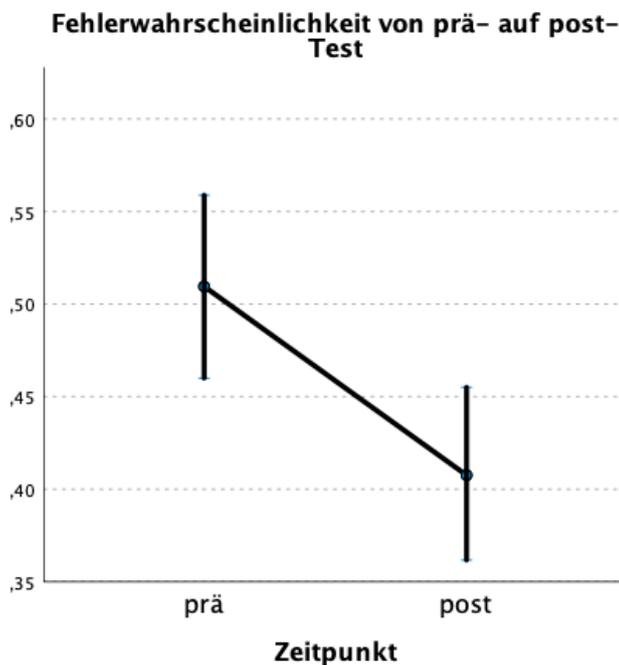


Abbildung 12 Veränderung der Fehlerwahrscheinlichkeit zwischen Prä- und Posttest

Nachdem kein spezifischer Lerneffekt hinsichtlich der Gruppen „Lernen aus eigenen Fehlern“ „Lernen aus fremden Fehlern“ gefunden werden konnte, wurde der Lerneffekt hinsichtlich der Subkompetenzen untersucht, um zu prüfen, ob ein signifikanter Lerneffekt zwischen Prä- und Posttest innerhalb der einzelnen Subkompetenzen vorliegt. Das gemischte lineare Regressionsmodell zeigte, dass bei allen Subkompetenzen ein signifikanter Unterschied zwischen Prä- und Posttest vorliegt. PL wies den größten Lerneffekt auf. Die Fehlerwahrscheinlichkeit war nach der Intervention im Posttest um 62% verringert, OR 0,38 (vgl. Tabelle 21).

Tabelle 21 Lerneffekt je Subkompetenz

b = Regressionskoeffizient, SE(b) = Standardfehler von b, OR = odds ratio, 95%-KI= Konfidenzintervall; UN und Prä = Referenzkategorie, Schüler*innen-Hauptcode als Zufallseffekt.

Modell: cvs = Subkompetenz + Zeitpunkt * Subkompetenz

	b	SE(b)	t	p-Wert	OR	95%-KI	
<i>Konstanter Term</i>	2,08	0,18					
<i>ID prä vs. UN prä</i>	-2,87	0,17	-16,77	<0,001	0,06	0,04	0,08
<i>PL prä vs. UN prä</i>	-2,80	0,17	-16,47	<0,001	0,06	0,04	0,09
<i>IN prä vs. UN prä</i>	-4,91	0,21	-23,02	<0,001	0,01	0,01	0,01
<i>ID post vs prä</i>	-0,52	0,16	-3,31	0,001	0,59	0,44	0,81
<i>PL post vs prä</i>	-0,97	0,15	-6,28	<0,001	0,38	0,28	0,51
<i>IN post vs prä</i>	-0,63	0,22	-2,90	0,004	0,53	0,35	0,82
<i>UN post vs prä</i>	-0,38	0,19	-2,03	0,042	0,68	0,47	0,99

Die Subkompetenz PL näherte sich hinsichtlich ihrer Fehlerwahrscheinlichkeit der Subkompetenz ID an (vgl. Abbildung 13) und erwies sich somit nach der Intervention als ungefähr gleich schwierig für die Schüler*innen wie ID Aufgaben, die zuvor als einfacher empfunden wurden.

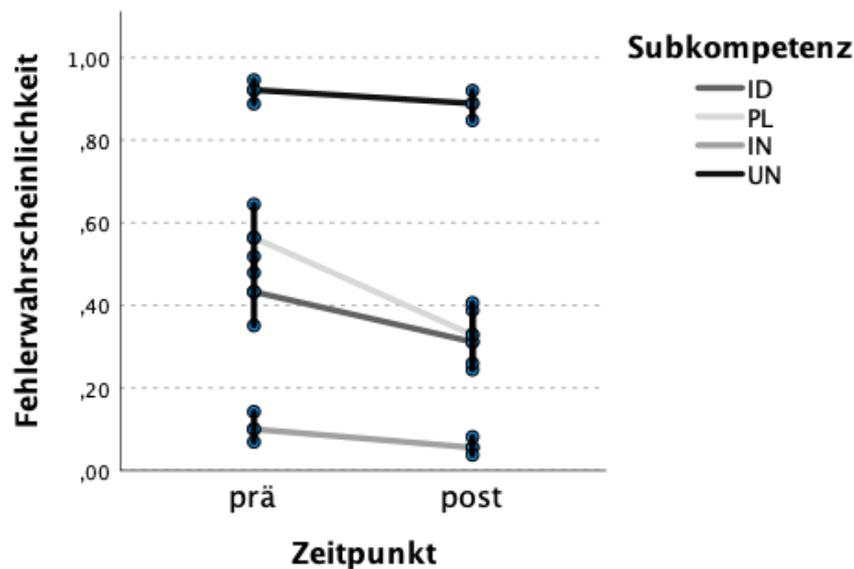


Abbildung 13 Veränderung der Fehlerwahrscheinlichkeit in Abhängigkeit der Subkompetenz zwischen Prä- und Posttest

9 Diskussion

Das Ziel dieser Masterarbeit war es, das Auftreten von Schüler*innenvorstellungen und die damit einhergehenden Designfehler beim Experimentieren in den jeweiligen Subkompetenzen in der Sekundarstufe I zu untersuchen. Weiters sollte der Zusammenhang von Designfehlern über die Subkompetenzen hinweg betrachtet werden. Der Lerneffekt aus eigenen und fremden Fehlern stellte den dritten Schwerpunkt dar.

9.1 Designfehlerhäufigkeit in den Subkompetenzen ID, PL, IN und UN

Hinsichtlich der ersten Forschungsfrage konnten nicht alle Designfehler, die bisher von Tschirgi (1980), Siler und Klahr (2012) bzw. Schwichow et al. (2022) beschrieben wurden, gefunden werden. Die Häufigkeit der identifizierten Designfehler betrug zwischen 0,2% (*nicht kontrastive Variable*) und 51,8% (*konfundiertes Experiment – drei variierte Variablen*). Dies unterschied sich von den Ergebnissen von Schwichow et al. (2022). In jener Studie war die Schüler*innenvorstellung bzw. der Designfehler „Experiment mit nur einer Bedingung“ jener, der am wenigsten häufig gemacht wurde, wobei der Fehler, obwohl er der seltenste war, in einer größeren Häufigkeit (4%) auftrat.

Bei Betrachtung von nur jenen Vorstellungen und Designfehlern, die in einer Häufigkeit von mindestens 4% vorkommen, ist „kontrolliertes Experiment für die falsche Variable“ der seltenste Designfehler in der vorliegenden Analyse.

Der häufigste Designfehler (51%), konfundiertes Experiment mit zwei variierten Variablen, deckt sich mit den Ergebnissen von Schwichow et al. (2022), wo er mit 33% ebenso der häufigste war. Die Verteilung der Häufigkeiten je Subkompetenz unterschieden sich hierbei jedoch von den Ergebnissen von Schwichow et al. (2022). Dies ist darauf zurückzuführen, dass die Testhefte vor allem hinsichtlich der UN Aufgaben unterschiedlich gestaltet waren. In der vorliegenden Masterarbeit wurde mit Testheften gearbeitet, in denen nur zwei Designfehler vorkommen konnten, während in den Testheften von Schwichow et al. (2022) fünf Designfehler möglich waren. Der Unterschied in den UN Aufgaben wird nachfolgend erläutert. H1 (Alle definierten Designfehler von Siler & Klahr 2012 und Schwichow et al. 2022 lassen sich auch bei Schüler*innen der Sekundarstufe I identifizieren) konnte demnach nicht bestätigt werden. Wird die Gestaltung der Testhefte außer Acht gelassen, die das Vorkommen von manchen Designfehlern von vornherein verhinderte, traten dennoch nicht alle Designfehler in signifikanter Ausprägung auf. Eine mögliche Erklärung kann sein, dass Schüler*innen der Sekundarstufe I ein höheres Fachwissen aufweisen, wodurch angenommen werden kann, dass sie auch über höhere VKS-Kompetenzen verfügen und manche Vorstellungen bereits abgelegt haben (Schwichow & Nehring 2018; Edelsbrunner et al. 2018). Die Ergebnisse deuten darauf hin, dass Schüler*innen der Sekundarstufe I die zweite Regel der VKS „Festlegung von mindestens zwei Versuchsbedingungen, die sich hinsichtlich dieser Variable unterscheiden“ verstanden haben und daher die Vorstellung, dass zwei gleiche Bedingungen ein vergleichbares Experiment darstellen können, abgelegt haben.

Auch korrekt gelöste Aufgaben deckten sich nicht mit den Ergebnissen von Schwichow et al. (2022). Während die Subkompetenz ID bei Schwichow et al. (2022) mit 73% am meisten richtig beantwortete Items aufwies, wurde dieselbe Subkompetenz in dieser Studie nur von 49,3% richtig beantwortet, während IN mit 73,9% die Subkompetenz mit den meisten korrekt gelösten Aufgaben war. Eine mögliche Erklärung für den Unterschied bei den IN Aufgaben kann sein, dass das Testheft von Schwichow et al. (2022) auf das Niveau von Volksschüler*innen angepasst war. Im Gegensatz zum aktuellen Testheft, das den Schüler*innen ein kontrolliertes Experiment zeigt und fragt, was getestet wird, wurden die Schüler*innen bei Schwichow (2022) im Voraus darüber informiert, dass es sich um ein kontrolliertes Experiment handelt. Anschließend mussten sie aus mehreren Antwortmöglichkeiten auswählen und begründen, warum es sich um ein kontrolliertes und aussagekräftiges Experiment handelt. Möglicherweise fällt es Schüler*innen schwerer, mehrere Aussagen zu einem Experiment auf einmal zu bewerten, als die korrekte Antwort aus mehreren zu identifizieren. Zusätzlich mussten bei Schwichow (2022) drei Kreuze richtig gesetzt werden, damit die Aufgabe als richtig gelöst erscheint, diese mentale Herausforderung entfällt bei dem jetzt verwendeten Testheft. Möglicherweise führen mehrere Antworten und Aussagen zu einer Verwirrung von Schüler*innen.

UN wies mit nur 13,3% richtig beantworteter Items die meisten Fehler auf, während die Aufgaben bei Schwichow et al. (2022) von 38% der Schüler*innen korrekt beantwortet wurden. Die Subkompetenz PL wurde bislang noch in keinem der CVSI Testhefte

implementiert. Hinsichtlich der korrekt gelösten Aufgaben reichte es sich mit 44,6% unmittelbar hinter ID ein.

9.2 Muster von Designfehlern

9.2.1 Fehlerwahrscheinlichkeit in Abhängigkeit der Subkompetenzen

Die Hypothese 2.1., die UN als die herausforderndste Subkompetenz annimmt, wird durch die Analyseergebnisse bestätigt und findet Unterstützung in den Ergebnissen von Schwichow et al. (2016), die ebenso zu der Erkenntnis kamen, dass die Subkompetenz UN die meisten Schwierigkeiten für Schüler*innen beinhaltet. Schüler*innen haben demnach Probleme, ein konfundiertes Experiment zu erkennen und festzustellen, dass es keine Aussagekraft hat. Die vorhin genannten unterschiedlichen Ergebnisse von Schwichow et al. (2022) zu Schwichow et al. (2016) sowie vorliegender Studie ist durch den Unterschied in der Aufgabenstellung erklärbar. Schwichow et al. (2022) wiesen die Schüler*innen darauf hin, dass das Experiment keine Schlussfolgerung zulässt, und fragten nach dem Grund, während Schwichow et al. (2016) sowie die aktuell verwendeten Testhefte ein konfundiertes Experiment präsentierten und nach Schlussfolgerungen fragten. Die korrekte Antwort wäre hierbei, dass das Experiment keine Schlussfolgerung zulässt. Die Aufgaben in dem für die Masterarbeit verwendeten Testheft waren auf jene von Schwichow et al. (2016) aufgebaut, was die Ergebnisse vergleichbar macht.

Die Kompetenzunterschiede in der Subkompetenz UN könnten durch verschiedene benötigte Erfahrungen für die Entwicklung dieser Kompetenz erklärt werden (Schwichow et al. 2020). Während PL, IN und ID darauf basieren, dass Schüler*innen über Beweise und Schlussfolgerungen nachdenken, setzt UN eine tiefere Reflexion über die Prozesse der Interpretation und Erstellung von Beweisen voraus (Schwichow 2016). Statt sich zu fragen, welche Schlussfolgerungen aus den Beweisen gezogen werden können (PL, IN und ID), müssen die Schüler*innen die Frage beantworten, wieso keine bestimmten Schlüsse aus den Beweisen gezogen werden können. Die Herausforderung liegt zusätzlich darin, zu verstehen, dass die Veränderung von mehreren Variablen gleichzeitig die Identifizierung kausaler Beziehungen verhindert. Dies deutet darauf hin, dass die hierarchische Anordnung der VKS-Subkompetenzen daraus resultiert, dass die anderen drei Aspekte grundlegend für das Erlangen einer vollständigen Kenntnis der VKS sind (Schwichow et al. 2016). Das Verständnis über die fehlende Aussagekraft eines konfundierten Experiments scheint kein kognitiver Aspekt zu sein, sondern fehlendes metakognitives Wissen darüber, wann nach Störfaktoren gesucht werden soll (Zohar & Peled, 2008).

Weiters stellt sich die Frage, ob die Schwierigkeit der UN-Aufgaben auch damit zusammenhängen könnte, dass Schüler*innen selten in die Lage gebracht werden, in der die Antwortoption, dass keine gültige Schlussfolgerung möglich ist, die Korrekte ist. Ohne den Schüler*innen einen eindeutigen Hinweis auf die fehlende Aussagekraft zu geben, gehen sie nicht selbstständig auf die Suche nach konfundierten Variablen (Schwichow et al. 2022).

Das von Schwichow et al. (2016) vorgeschlagene und 2022 umgesetzte UN-Aufgabenformat, kann in weiteren Untersuchungen mit Sekundarschüler*innen eingesetzt werden. Dabei erhalten Schüler*innen vorab die Information, dass das Experiment konfundiert ist, und müssen eine Begründung dafür darlegen. Diese Vorwegnahme könnte zudem dazu beitragen, die potenziellen Schwierigkeiten zu überwinden, die sich aus der Unvertrautheit der Schüler*innen mit Antworten ohne valide Aussagekraft ergeben. Indem diese Hürde genommen wird, könnten Schüler*innenvorstellungen gezielter erhoben und in weiterer Folge im Unterricht besser aufgegriffen werden können. Die Subkompetenz UN setzt voraus, dass Schüler*innen in der Lage sind, zu verstehen, dass bzw. warum, konfundierte Experimente keine Aussagekraft haben (Schwichow et al. 2016) – jedoch nicht daraus, zuerst selbst zu erkennen, dass es sich um ein konfundiertes Experiment handelt.

9.2.2 Fehlerwahrscheinlichkeit in Abhängigkeit der Fachkontexte

Um mögliche Muster über die Subkompetenzen hinweg zu identifizieren (Forschungsfrage 2), wurden die Designfehler zusätzlich abhängig von den Fachkontexten untersucht.

Die Hypothese H2.2. (Im Fachbereich Allgemeinwissen treten über alle Subkompetenzen hinweg am wenigsten Designfehler auf) kann nicht bestätigt werden, da im Fachkontext Allgemeinwissen nicht wie angenommen, über alle Subkompetenzen hinweg am wenigsten Designfehler gemacht wurden. Es wurde angenommen, dass Schüler*innen die Inhalte des Allgemeinwissens leichter zugänglich und verständlich sind als spezifische Fachinhalte, weshalb die Fehlerwahrscheinlichkeit für diesen Fachkontext am geringsten eingeschätzt wurde. Es konnten keine signifikanten Unterschiede in der Fehlerwahrscheinlichkeit über alle Subkompetenzen hinweg beobachtet werden. Jedoch konnte überall außer bei IN Aufgaben festgestellt werden, dass im Fachkontext Allgemeinwissen signifikant weniger Designfehler auftraten als in Biologie. Lediglich bei der Subkompetenz PL konnte zusätzlich auch noch festgestellt werden, dass bei All-Aufgaben weniger Fehler gemacht wurden als in Physik.

Es kann nicht geklärt werden, weshalb die Hypothese nicht bestätigt werden kann und die Schüler*innen nicht durchgehend weniger Fehler im Bereich Allgemeinwissen gemacht haben. Eine mögliche Erklärung könnte sein, dass Schüler*innen verstärkt auf ihre Vorstellungen zurückgreifen, die sich möglicherweise nicht mit der Realität decken. Die Inhalte der Allgemeinwissen-Beispiele bestanden aus dem Einfluss der Länge eines Besenstiels auf dessen Putzleistung sowie der Backzeit von Muffins. Bei Schüler*innen der 7. und 8. Schulstufe kann davon ausgegangen werden, dass sie bereits Erfahrungen in den beiden Bereichen gesammelt haben, auf die sie zurückgreifen können. Dies könnte dazu führen, dass sie die Fragen nicht nach wissenschaftlichen Kriterien beantworten.

9.2.3 Fehlerwahrscheinlichkeit in Abhängigkeit der Wechselwirkung der Fachkontexte und Subkompetenzen

Alle Items wurden grundsätzlich gleich anspruchsvoll konzipiert und sollten ohne fachliches Vorwissen lösbar sein (Schwichow et al. In prep.). Jedoch konnte bei der Analyse festgestellt werden, dass in Chemie in den Subkompetenzen ID und PL die meisten Designfehler begangen wurden. Bei UN und IN erwies sich die Fehleranzahl hingegen als relativ stabil. Die

Unterschiede hinsichtlich der Schwierigkeiten von Schüler*innen in den Subkompetenzen ID und PL im Fachkontext Chemie, können durch eine Kombination von möglichen Faktoren erklärt werden, die sowohl die spezifischen kognitiven Anforderungen dieser Subkompetenzen als auch die Struktur der Lehrpläne betreffen.

Ein ausschlaggebender Punkt ist, dass Schüler*innen erst ab der 8. Schulstufe Chemieunterricht im Ausmaß von zwei Wochenstunden erhalten (Lehrplan 2023). In der Studie wurden jedoch Schüler*innen der 7. und 8. Schulstufe miteinbezogen, was bedeutet, dass einige der Kinder noch keine schulische Bildung in Chemie erhalten haben und es sein kann, dass vor allem den jüngeren Schüler*innen fachspezifische Vorerfahrung fehlt, die sie vermeintlich an der Beantwortung der Fragen hinderte. Dadurch, dass die Studie im Wintersemester durchgeführt wurde, hatten selbst die Schüler*innen der 8. Schulstufe noch keinen umfangreichen Chemieunterricht und ein dementsprechend niedriges Fachwissen. Auch wenn kein spezifisches Fachwissen zur Beantwortung der Fragen notwendig war, kann die Tatsache, dass ein Teil der Schüler*innen noch gar keinen und der andere Teil der Schüler*innen nur sehr begrenzten Chemieunterricht hatte, zu einer Überforderung geführt haben. Bisherige Studien besagen, dass ein Zusammenhang zwischen domänenspezifischem Wissen und der VKS besteht und Schüler*innen mit höherem fachspezifischen Wissen auch eine höhere VKS-Kompetenz aufweisen (Nehring & Schwichow 2020, Schalk et al. 2019, Schwichow et al. 2020). Somit könnte das fehlende in diesem Bereich eine Erklärung für die Ergebnisse hinsichtlich der Chemie Items sein. Dass Schüler*innen bei Chemie-Aufgaben die meisten Schwierigkeiten hatten, ließ sich auch durch eine Analyse der fehlenden Werte bestätigen. Chemie wies hierbei die meisten fehlenden Werte auf, was die mögliche Erklärung einer hohen kognitiven Belastung durch die Wissensdefizite bestärkt.

Angesichts der Erkenntnis von Schwichow et al. (in prep), dass die Domänenspezifität der Leistung von Schüler*innen mit zunehmender VKS-Expertise abnimmt, lässt sich in Bezug auf diese Studie vermuten, dass die domänenspezifischen Ergebnisse hinsichtlich der Chemie die generell niedrigere VKS-Kompetenz von den Schüler*innen widerspiegelt.

Warum Chemie-Items nur in den Subkompetenzen ID und PL zu einer deutlich höheren Fehlerwahrscheinlichkeit geführt haben, kann nicht eindeutig beantwortet werden.

Es kann jedoch argumentiert werden, dass zumindest die Subkompetenz PL ein tieferes Verständnis der experimentellen Methode erfordert, um die Variablen korrekt identifizieren, und um ein kontrolliertes Experiment planen zu können. Die Komplexität und Abstraktheit chemischer Konzepte könnte damit einhergehend eine Ursache für die erhöhte Fehlerquote sein (Sirhan 2007). Ohne die Möglichkeit, durch praktische Erfahrung und intuitives Verständnis für die Materie zu entwickeln, müssen sich Schüler*innen auf ihr theoretisches Wissen verlassen, um Experimente auf dem Papier zu planen oder zu verstehen. Dies kann herausfordernd sein, wenn es um abstrakte Konzepte der Chemie geht. Chemie stellt für Schüler*innen oft aufgrund der Abstraktheit beziehungsweise auch aufgrund fehlender Relevanz chemischer Prozesse im Leben der Schüler*innen eine Herausforderung dar (Gilbert 2007). Außerdem stehen chemische Prozesse oftmals nicht im Einklang mit Schüler*innenvorstellungen, was zu Verständnisschwierigkeiten führen kann (Treagust 2018).

Beispielsweise wie in den verwendeten Testheften, wo es um die Leitfähigkeit des Wassers oder die Auswirkungen von Reaktionsbedingungen auf die Geschwindigkeit chemischer Reaktionen ging. Die Schwierigkeit, die Auswirkungen von Variablenänderungen, ohne die direkte Beobachtung von Experimenten vorherzusagen, könnte dazu beigetragen haben, dass jene Beispiele in PL und ID die meisten Fehler aufgewiesen haben.

Bei genauer Betrachtung der IN und PL-Aufgabenstellungen fallen jedoch einige Unterschiede zu den übrigen Subkompetenzen auf. Im Gegensatz zu den UN und IN Aufgaben, bei denen Schüler*innen ein einzelnes Experiment vorgelegt wurde, aus dem sie die korrekte Antwort aus vier vorgegebenen Optionen auswählen mussten, erforderten ID und PL Aufgaben eine intensivere Auseinandersetzung mit mehreren bildhaft dargestellten Experimentdesigns, was zu einer höheren Belastung geführt haben könnte. Warum dies allerdings nur bei Chemie zu Problemen führt, bleibt ungeklärt. Die Betrachtung der Fachkontexte innerhalb der Subkompetenzen könnte in Folgestudien genauer untersucht werden, damit ausgeschlossen werden kann, dass es sich hierbei um ein zufälliges Ergebnis handelt, weil beispielsweise die Aufgabenschwierigkeit doch nicht, wie angenommen, über alle Fachkontexte und Items hinweg gleich ist.

Die bisherige Forschung zur Variablenkontrollstrategie hat sich primär auf die Rolle der Subkompetenzen konzentriert (Schwichow et al. 2016, 2022), mit der Annahme, dass die Schwierigkeit der Aufgaben hauptsächlich von der jeweiligen Subkompetenz und nicht vom Inhalt der Fachkontexte abhängt. In früheren Untersuchungen wurde der Fokus entweder auf nur einen Fachkontext, z.B. Physik bei Schwichow et al. (2016) oder Allgemeinwissen bei Brandenburger et al. (2022) gelegt. Jedoch hat eine aktuelle Studie von Schwichow et al. (in prep) die Zusammenhänge zwischen Subkompetenzen und Fachkontexten näher beleuchtet. Die Ergebnisse deuten ebenso darauf hin, dass die Subkompetenzen und nicht die Fachkontexte der ausschlaggebende Faktor darüber sind, ob Schüler*innen die Aufgabenstellungen richtig beantworten, wenngleich signifikante Unterschiede zwischen Chemie und Biologie sowie Chemie und Physik-Aufgabenstellungen gefunden werden konnten (Schwichow et al. In prep.). Dies steht im Einklang mit den vorliegenden Analyseergebnissen, die ebenso signifikante Unterschiede zwischen Chemie und anderen Fachkontexten festgestellt haben. Es besteht jedoch auch die Annahme, dass Schüler*innen in der Lage sein können, ihre VKS-Kompetenz domänenübergreifend und unabhängig von spezifischem Fachwissen auf andere Bereiche zu übertragen (Schwichow et al. In prep.). Brandenburger & Mikelskis-Seifert (2019) arbeiteten mit einem ähnlichen Testheft, ebenfalls mit der Erkenntnis, dass die Schwierigkeit bezüglich der VKS von den Subkompetenzen abhängt und nicht von den Fachkontexten (Brandenburger & Mikelskis-Seifert 2019). Die Ergebnisse der vorliegenden Masterarbeit bieten jedoch mögliche Hinweise darauf, dass die Fachkontexte ebenso einen Einfluss auf die VKS-Kompetenz haben können. Insbesondere die Ergebnisse hinsichtlich der ID und PL-Aufgaben in Chemie wirft Fragen bezüglich der Rolle des Fachkontexts auf. Schüler*innen waren offensichtlich nicht in der Lage, ihr VKS-Wissen gleichermaßen auf alle Subkompetenzen zu übertragen. Somit ergeben sich weitere Impulse für die Diskussion über die Bedeutung des Fachkontextes für die VKS und die Notwendigkeit,

in zukünftigen Forschungen die Wechselwirkungen zwischen Subkompetenzen und Fachkontexten zu untersuchen.

Die vorliegende Meisterarbeit unterstreicht die Komplexität der VKS und die potenzielle Relevanz des Fachkontexts. Dies könnte zukünftige Untersuchungen anregen, die spezifischen Herausforderungen innerhalb der Fachkontexte nochmals zu beleuchten. Es sollte untersucht werden, wie Schüler*innen ihre VKS-Kompetenz auch ohne fachspezifisches Wissen domänenübergreifend anwenden können.

9.3 Veränderung der Fehlerwahrscheinlichkeit durch Lernen aus Fehlern

Der dritte Forschungsschwerpunkt war auf Lernen aus Fehlern bezogen. Insbesondere sollte untersucht werden, ob die Schüler*innen, die sich in der Intervention mit eigenen Fehlern beschäftigt haben, einen höheren Lernzuwachs haben als jene, die sich mit fremden Fehlern beschäftigt haben. Die Hypothese, dass Schüler*innen aus eigenen Fehlern besser lernen, kann nicht bestätigt werden. Es gab zwar insgesamt einen Lernzuwachs durch die Intervention und es konnte festgestellt werden, dass nach der Intervention insgesamt weniger Designfehler gemacht wurden, jedoch gab es keinen nachweisbaren Gruppeneffekt.

Somit kann generell angenommen werden, dass das Lernen aus Fehlern, ungeachtet dessen, ob aus eigenen oder aus fremden Fehlern, einen positiven Effekt auf das VKS-Verständnis von Schüler*innen hat. Es konnte jedenfalls nicht festgestellt werden, dass Lernen aus eigenen Fehlern gescheitert ist, was unter anderem auf das Interventionsdesign zurückzuführen ist.

Wie zuvor in der Methodik beschrieben, war die Intervention an Kapur und Bielaczyc (2012) sowie an Sinha und Kapur (2021) angelehnt. Die Gestaltung der Intervention kann unter anderem als ausschlaggebender Faktor hinsichtlich des Lernzuwachses gesehen werden. Als besonders hilfreich könnte gewesen sein, dass die Schüler*innen unmittelbar im Anschluss an die Problemlöseaktivität die Erklärungen erhalten haben (Sinha & Kapur 2019). Ebenso wurde wiederholtes Scheitern von Schüler*innen vermieden (Sinha & Kapur 2019), was ebenso beträchtlich zum Erfolg der Intervention beigetragen haben könnte. Da Sinha und Kapur (2021) festgestellt haben, dass vor allem Schüler*innen einer höheren Schulstufe besser auf den Lernansatz ansprechen, können für zukünftige Forschungen vermehrt höhere Schulstufe herangezogen werden.

Die Analyse hinsichtlich des Lerneffekts innerhalb der Subkompetenzen zeigte eine deutliche Verbesserung bei allen Subkompetenzen. Am größten war der Lerneffekt bei den PL-Aufgaben. Dies kann darauf zurückgeführt werden, dass sich die Schüler*innen während der Intervention vor allem mit der Planung von Experimenten beschäftigt haben und somit mit diesem Teilaspekt der VKS am meisten vertraut waren. Unterstützt wird die These von den Erkenntnissen von Schwichow et al. (2016), die in ihrer Studie den positiven Effekt von Interventionen, die sich explizit mit der Vermittlung einer VKS-(Sub)kompetenz beschäftigt, hervorgehoben haben (Schwichow et al. 2016). Ungeachtet dessen, ob sich die Schüler*innen mit eigenen oder mit fremden Fehlern beschäftigt haben, trat also durch die Intervention ein

deutlicher Lerneffekt auf, wobei Lerneffekt so definiert wird, dass nach der Intervention weniger Designfehler gemacht wurden als vor der Intervention.

Jedoch konnte auch festgestellt werden, dass Schüler*innen hinsichtlich UN-Aufgaben den geringsten Lernzuwachs hatten. Somit blieb UN auch nach der Intervention die schwierigste Subkompetenz für Schüler*innen. Dies eröffnet die Möglichkeit, oder zeigt die Notwendigkeit für weitere Interventionsstudien, wobei sich die Schüler*innen in der Intervention vorrangig mit konfundierten Experimenten und deren fehlender Aussagekraft beschäftigen. Somit könnte einerseits die VKS-Kompetenz im Allgemeinen, als auch die bekannten Schwierigkeiten von Schüler*innen im Speziellen gefördert werden.

10 Implikationen für den Unterricht

Im Rahmen der Masterarbeit wurden die verschiedenen Subkompetenzen der Variablenkontrollstrategie untersucht. Die Ergebnisse dieser Arbeit unterstreichen die Bedeutung von Fehlern als integralen Bestandteil des Lernprozesses. Es hat sich gezeigt, dass Schüler*innen durch die Beschäftigung mit Fehlern, ungeachtet, ob es eigene oder fremde Fehler waren, einen deutlichen Lernzuwachs verzeichnen konnten. Im Unterricht könnten Lehrkräfte daher einen Unterrichtsansatz verfolgen, der Fehler als Lerngelegenheiten und nicht als Misserfolge erkennt. Dies erfordert die Schaffung einer Lernumgebung, in der Schüler*innen sich frei fühlen, Hypothesen zu testen und experimentelle Designs zu entwickeln, ohne Angst vor negativen Bewertungen zu haben. Eine mögliche Implementierung in den Unterricht könnten demnach Aufgaben sein, in denen Schüler*innen bewusst Fehler in experimentellen Designs identifizieren und korrigieren können. Auch wenn selbstständiges Experimentieren nach wie vor einen hohen Stellenwert im Lehrplan hat und auch in der Schule praktiziert werden sollte (Lehrplan 2023), kann der fehlende Gruppeneffekt, der in der Analyse gefunden wurde, dazu genutzt werden, um Schüler*innen auch öfter mit vorgegebenen Experimenten arbeiten zu lassen. Dies bedeutet, dass es nicht zwingend notwendig sein muss, Schüler*innen selbst experimentieren zu lassen. Naturwissenschaftlicher Unterricht ist auf wenige Wochenstunden begrenzt (Lehrplan 2023), was dazu führen kann, dass sich Lehrkräfte oft gezwungen sehen, aus Zeitgründen keine Experimente durchzuführen. Die Ergebnisse der Masterarbeit könnten Möglichkeiten dafür bieten, Schüler*innen entweder vorgegebene Experimente, in Anlehnung an die Materialien der Intervention, zur Verfügung zu stellen. Ebenso könnten Lehrkräfte die Experimente selbst durchführen und für die Schüler*innen veranschaulichen. Da in der Analyse keine Unterschiede zwischen Lernen aus eigenen und Lernen aus fremden Fehlern gefunden werden konnte, kann davon ausgegangen werden, dass Schüler*innen auch von den vorgeschlagenen Experimentdesigns profitieren können.

Um die unterschiedlichen Leistungsniveaus hinsichtlich der Subkompetenzen im Unterricht aufgreifen zu können, ist ein differenzierter Unterrichtsansatz notwendig, um Schüler*innen vor allem zu unterstützen, damit sie die fehlende Aussagekraft von konfundierten

Experimenten verstehen. Um dies gewährleisten zu können, ist es vor allem notwendig, dass Schüler*innen die Voraussetzungen für ein kontrolliertes Experiment kennen und verstehen. Durch die Kenntnis eines kontrollierten Experiments kann ein konfundiertes erkannt werden. Die Analyse der Schüler*innenvorstellungen kann Lehrkräften einen tieferen Einblick in die Verständnisschwierigkeiten von Schüler*innen geben. Die meisten Designfehler sind auf die Veränderung von mehreren Variablen gleichzeitig zurückzuführen. Dies deutet darauf hin, dass Schüler*innen die Grundregel eines kontrollierten Experiments nicht verstanden haben und gibt Lehrkräften die Möglichkeit, gezielt auf diese Vorstellung zu reagieren. Lehrkräfte können über die Gestaltung ihres Unterrichts einen Conceptual Change bei Schüler*innen hervorrufen. In weiterer Folge können dadurch die Designfehler der Schüler*innen verringert werden. Dafür sollten die Materialien progressiv aufeinander aufbauend gestaltet sind und die Lernenden somit stufenweise mehr gefördert werden (Fortus & Krajcik 2012). Einhergehend mit einem kohärenten Lehrplan sind auch nachvollziehbare Kontexte von Bedeutung, um Schüler*innen zu einem besseren Verständnis von neuen Lerninhalten zu verhelfen. Indem Schüler*innen nicht nur ein neues Phänomen vorgestellt wird, sondern das Phänomen auch in den Kontext des alltäglichen Lebens und möglichen Problemlösungen integriert wird, kann ein großer Lernzuwachs entstehen (Bransford et al. 2000). Lernen braucht darüber hinaus Zeit. Schüler*innen müssen sich über einen längeren Zeitraum mit neuen Ideen und Konzepten beschäftigen können, um sie in ihr Wissen zu implementieren (Krajcik & Shin 2023). Durch gezielte Übungen im Unterricht kann somit ein Conceptual Change bei Schüler*innen bewirkt werden, der dazu führt, ihre Vorstellungen eines kontrollierten Experiments abzulegen zugunsten der korrekten wissenschaftlichen Bedeutung.

Die Fähigkeit von Schüler*innen, ihre VKS-Kompetenz auf verschiedene Kontexte übertragen zu können, erweist sich von genereller Wichtigkeit, indem Schüler*innen sehr wahrscheinlich in ihrem Leben außerhalb der Schule vorrangig mit Situationen konfrontiert werden, die sich von den gelernten Szenarien unterscheiden (Kuhn 2005). Weiters sind scientific-reasoning Fähigkeiten besonders dann wertvoll, wenn sie zum Erschließen neuen Wissens genutzt werden können (Schwchow et al. In prep).

Die Ergebnisse des Posttests zeigten, dass Schüler*innen durch die intensive Beschäftigung mit der Planung von Experimenten während der Intervention signifikante Lernfortschritte in der Subkompetenz PL erzielt haben. Daher empfiehlt es sich, im Unterricht verstärkt die Subkompetenzen einzeln zu fördern, anstatt die Variablenkontrollstrategie ausschließlich in ihrer Gesamtheit zu unterrichten.

11 Limitationen

In der Masterarbeit wurden die Vorstellungen von Schüler*innen und damit einhergehende Designfehler hinsichtlich der Variablenkontrollstrategie untersucht, mit einem Fokus auf Lernen aus Fehlern. Eine potenzielle Limitation, die bereits angesprochen wurde, könnte der

fehlende Chemieunterricht in der 7. Schulstufe und erst seit kurzem vorhandene in der 8. Schulstufe sein. Dieser Umstand hat möglicherweise dazu geführt, dass Schüler*innen vermehrt Fragen im Fachkontext Chemie nicht beantwortet haben. Es kann auch nicht ausgeschlossen werden, dass damit ein Motivationsverlust einhergegangen ist und eventuell Auswirkungen auf das Beantworten des weiteren Testhefts hatte.

Die Ergebnisse können jedenfalls als weitere Forschungsgrundlage dienen, indem verstärkt auf die einzelnen Schüler*innenvorstellungen und Designfehler bei der Variablenkontrollstrategie fokussiert wird. Insbesondere die Subkompetenz UN erfordert vermehrte Aufmerksamkeit, um besser auf die Schwierigkeiten, die Schüler*innen damit haben, die fehlende Aussagekraft eines konfundierten Experiments zu verstehen, eingehen zu können.

12 Zusammenfassung

Die Variablenkontrollstrategie als wichtiges Instrument der Experimentierkompetenz stellt nach wie vor Schwierigkeiten für Schüler*innen dar. In der vorliegenden Masterarbeit wurde die Rolle von Schüler*innenvorstellungen und deren Einfluss auf die Experimentierkompetenz hinsichtlich der Variablenkontrollstrategie im naturwissenschaftlichen Bildungsprozess behandelt. Erstmals wurde die Experimentierkompetenz von Schüler*innen der Sekundarstufe I bezüglich der Subkompetenzen der Variablenkontrollstrategie sowie fachkontextspezifisch untersucht. Die Analysen gestalten sich einerseits deskriptiv, um die Fehlerhäufigkeiten in den vier Subkompetenzen zu beschreiben und andererseits wurden Zusammenhänge zwischen Subkompetenzen, Fachkontexten und Testzeitpunkten mittels eines verallgemeinerten gemischten Regressionsmodells untersucht. Dabei konnte abermals festgestellt werden, dass die Subkompetenz UN die meisten Schwierigkeiten für Schüler*innen birgt, und zwar konstant über alle Fachkontexte hinweg. Ein Zusammenhang zwischen Fachkontext und Subkompetenz konnte nur hinsichtlich der Chemie-Aufgaben gefunden werden, wobei dieser teilweise auf den fehlenden Chemieunterricht der Schüler*innen zurückgeführt werden kann.

Auch wenn kein Vorteil von Lernen aus eigenen Fehlern gegenüber dem Lernen aus fremden Fehlern gefunden werden konnte, konnte dennoch festgestellt werden, dass ein deutlicher Lernzuwachs durch die Intervention stattgefunden hat. Es scheint demnach in Bezug auf die Variablenkontrollstrategie in diesem Fall keinen Unterschied zu machen, mit welcher Art des Fehlers sich die Schüler*innen beschäftigen, sondern der ausschlaggebende Faktor scheint die Intervention an sich und der Fokus auf Fehler (sowohl eigene als auch fremde) zu sein.

Die Intervention und die Auswirkungen auf die Schüler*innenvorstellungen bzw. Designfehler wurde ebenso erstmals spezifisch hinsichtlich der Subkompetenzen untersucht. Dabei konnte festgestellt werden, dass die Subkompetenz PL am meisten von der Intervention profitieren konnte und der Lernzuwachs in jener Subkompetenz am höchsten war. Die Beschäftigung mit Planungsaufgaben während der Intervention scheint zu einem deutlich besseren Verständnis

dahingehend geführt zu haben. Für den VKS-Unterricht bedeutet dies, dass Subkompetenzspezifisches VKS-Training eine effektive Methode für den Unterricht bedeuten kann.

Als Folgerung kann nahegelegt werden, dass ein bewusster Umgang mit Fehlern im Unterricht eine wichtige Rolle spielt, um die naturwissenschaftliche Grundbildung und vor allem den Kompetenzbereich Erkenntnisse gewinnen (E) zu fördern.

13 Literaturverzeichnis

Ausubel, D. (1968): Educational psychology: A cognitive viewpoint. New York: Holt, Rinehart & Winston.

Arnold, J.; Kremer, K.; Mayer, J. (2013): Wissenschaftliches Denken beim Experimentieren – Kompetenzdiagnose in der Sekundarstufe II. *Erkenntnisweg Biologiedidaktik* 11, 7–20.

Baur, A. (2018): Fehler, Fehlkonzepte und spezifische Vorgehensweisen von Schülerinnen und Schülern beim Experimentieren. Ergebnisse einer videogestützten Beobachtung. *ZfDN* 24, 115-129.

Baur, A. (2023): Which student problems in experimentation are related to one another? *International Journal of Science Education* 45(10), 781-805.

BGBI. II – ausgegeben am 2. Jänner 2023 – Nr. 1.

Boekarts, M. (2006): Self-regulation and effort investment. In: Sigel, E.; Menninger, K. (Hrsg.): Handbook of child psychology 4, New Jersey, Wiley, 345-377.

Boekaerts, M. (2007): Self-Regulation and Effort Investment. Handbook of child psychology 4.

Bransford, J.; Brown, A.; Cocking, R. (2000): How people learn. *National Academy Press* 11.

Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & W. Schneider (Eds.), Individual development from 3 to 12: Findings from the Munich longitudinal study. Cambridge University Press, 38–54.

Bullock, M.; Sodian B.; Koerber, S. (2008): Doing experiments and understanding science: development of scientific reasoning from childhood to adulthood. In Schneider, W.; Bullock, M. (Hrsg.): Human development from childhood to early adulthood. Findings from a 20 year longitudinal study. Psychology Press, New York.

Brandenburger, M., & Mikelskis-Seifert, S. (2019): Facetten experimenteller Kompetenz in den Naturwissenschaften. In: Maurer, C. (Hrsg.): Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Kiel 2018, 77-80.

Brandenburger, M.; Salim, C.; Schwichow, M.; Wilbers, J., Mikelskis-Seifert, S. (2022): Modellierung der Struktur der Variablenkontrollstrategie und Abbildung von Veränderungen in der Grundschule. *Zeitschrift für Didaktik der Naturwissenschaften* 28(5), 1-20.

Chase, C.; Klahr, D. (2017): Invention versus direct instruction: For some content, it's a tie. *Journal of Science Education and Technology* 26, 582-596.

Chen, Z.; Klahr, D. (1999): All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70(5), 1098-1120.

Durkin, K.; Rittle-Johnson, B. (2012): The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction* 22, 206-214.

- Forster-Heinzer, S.; Oser, F. (2015): Wer setzt das Maß? Eine kritische Auseinandersetzung mit dem Advokatorischen Ansatz. *Zeitschrift für Pädagogik* 61(3), 361-376.
- Fortus, D.; Krajcik, J. (2012): Curriculum coherence and learning progressions. In: Second international handbook of science education. Springer, 783-798.
- Frese, M.; Zapf, D. (1994): Action as the core of work psychology: A German approach. In: Traindis, H.; Dunnette, M.; Hough, L. (Hrsg.): Handbook of industrial and organizational psychology (4), Consulting Psychologists, Palo Alto, 271-340.
- Gartmeier, M.; Bauer, J.; Gruber, H.; Heid, H. (2008): Negative Knowledge: Understanding Professional Learning and Expertise. *Vocations and Learning*, 1, 87-103.
- Gilbert, J. (2007): On the nature of „context“ in chemical education. *International Journal of Science Education* 28(9), 957-976.
- Gropengießer, H. (2007): Didaktische Rekonstruktion des „Sehens“. Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung. In: Kattmann, U. (Hrsg.) BzDR Beiträge zur Didaktischen Rekonstruktion 1. Oldenburg, Didaktisches Zentrum, Carl-von-Ossietzky-Universität.
- Großengießer, H.; Marohn, A. (2018): Schülervorstellungen und Conceptual Change. In: Krüger, D.; Parchmann, I.; Schecker, H. (Hrsg.): Theorien in der naturwissenschaftsdidaktischen Forschung. Springer, Berlin, Bremen, Kiel, 49-68.
- Hammann, M.; Phan, T.; Ehmer, M.; Bayrhuber, H. (2006): Fehlerfrei Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht*, 59(5), 292–299.
- Hämmerle, L., Kuhlmann, S., Krause-Wichmann, T., Möller, A. (in prep.). Learning the control-of-variables strategy from self-generated and vicarious errors.
- Jackson, A.; Godwin, A.; Bartholomew, S.; Mentzer, N. (2021): Learning from failure. A systematized review. *International Journal of Technology and Design Education* 32, 1853-1873.
- Kattmann U., Duit R., Gropengießer H., Komorek M. (1997) Das Modell der Didaktischen Rekonstruktion. *ZfDN* 3(3):3–18.
- Kattmann, U. (2007): Didaktische Rekonstruktion - eine praktische Theorie. In: Krüger, D.; Vogt, H. (Hrsg.): Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudierende und Doktoranden. Springer, Berlin, Heidelberg, New York, 105 - 116.
- Kapur, M. (2010): Productive failure in mathematical problem solving. *Instructional Science*, 38(6), 523-550.
- Kapur, M. (2014): Productive failure in learning math. *Cognitive Science* 38(5), 1008-1022.
- Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, 21(1), 45–83.
- Kechel, J. (2016): Schülerschwierigkeiten beim eigenständigen Experimentieren. Eine qualitative Studie am Beispiel einer Experimentieraufgabe zum Hooke'schen Gesetz. Berlin: Logos.
- Kirchner, S., & Priemer, B. (2007). Probleme von Schülern mit offenen Experimentieraufgaben in Physik. In D. Höttecke (Hrsg.), *Naturwissenschaftlicher Unterricht im internationalen Vergleich*, LIT, Berlin, 346–348.

- Kirschner, P.A., Sweller, J., & Clark, R.E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, 75-86.
- Klahr, D. (2000): Exploring science: The cognition and development of discovery processes. Cambridge: MIT Press.
- Koerber, S.; Sodian, B.; Kropf, N.; Mayer, D.; Schwippert, K. (2011): Die Entwicklung des wissenschaftlichen Denkens im Grundschulalter. Theorieverständnis, Experimentierstrategien, Dateninterpretation. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 43(1), 16-21.
- Krajcik, J.; Shin, N. (2023): Student conceptions, conceptual change, and learning progressions. In: Lederman, N.; Zeidler, D.; Lederman, J. (Hrsg.): Handbook of Research on Science Education. Volume III, New York, Taylor & Francis, 121-157.
- Kryeziu, L. (2015): Learning from Errors. *Elyria International Review* 1, Holzkirchen, Felix Verlag.
- Kuhn, D. (2005). Education for thinking. Harvard University Press.
- Kuhn, D., & Dean, D. (2005): Is developing scientific thinking all about learning to control variables? *Psychological Science*, 16(11), 866–870.
- Lin, J.; Yen, M. Liang, J. Chiu, M.; Guo, C. (2016): Examining the factors that influence students' science learning processes and their learning outcomes. 30 years of conceptual change research. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(9), 2617-2646.
- Loibl, K.; Roll, I.; Rummel, N. (2017): Towards a Theory of when and how problem solving followed by instruction supports learning. *Education Psychology Rev.* 29, 693-715.
- Loibl, K.; Leuders, T. (2018): Errors during exploration and consolidation. The effectiveness of productive failure as sequentially guided discovery learning. *J Math Didaktische* 39, 69-96.
- Lui, H. (2021): Lernen aus Fehlern beim Experimentieren im naturwissenschaftlichen Unterricht mit Fokus auf der Variablenkontrollstrategie. Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes.
- Manik, A.; Klahr, D. (2003): Error Matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of cognition and development*, 4(1), 67-98.
- Mayer, J. (2007): Erkenntnisgewinnung als wissenschaftliches Problemlösen. In: Krüger, D. (Hrsg.): Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden. Springer, Berlin, 177-186.
- Meier, M., & Mayer, M. (2012): Experimentierkompetenz praktisch erfassen. Entwicklung und Validierung eines anwendungsbezogenen Aufgabendesigns. In: Harms, U.; Bogner, F. (Hrsg.): Lehr- und Lernforschung in der Biologiedidaktik, Innsbruck, Studien Verlag, 81–98.
- Metcalf, J. (2017): Learning from Errors. *Annual Reviews Psychology* 68, 465-489.
- Metcalf, J., Xu, J. (2018): Learning from one's own errors and those of others. *Psychonomic Society* 25, 402 - 408.
- Morrison, J.A.; Lederman, N.G. (2003): Science Teachers' Diagnosis and Understanding of Students' Preconceptions. *Science Education*, 87(6), 849-867.

- Mostafa, T.; Echazarra, A.; Guillou, H. (2018): The science of teaching science: An exploration of science teaching practices in PISA 2015. OECD Education Working Papers 188.
- Nieminen, P.; Häikiöniemi, M.; Viiri, J. (2021): Forms and functions of on the fly formative assessment conversations in physics inquiry lessons. *International Journal of Education*, 43(3), 362-384.
- Oser, F.; Spychiger, M. (2005): Lernen ist schmerzhaft. Zur Theorie des Negativen Wissens und zur Praxis der Fehlerkultur. Beltz, Weinheim & Basel.
- Oser, F.; Näpflin, C.; Hofer, C.; Aerni, P. (2012): Towards a Theory of Negative Knowledge (NK): Almost-Mistakes as Drivers of Episodic Memory Amplification. In: Bauer, J.; Hareis, C. (Hrsg.): *Human Fallibility: The Ambiguity of Errors for Work and Learning*. Professional and Practice-based Learning 6. Springer.
- Pan, S.; Sana, F.; Samani, J.; Cooke, J.; Kim, J. (2020): Learning from errors: students' and instructors' practices attitudes, and beliefs. *Memory* 28(9), 1105-1122.
- Peteranderl, S., & Edelsbrunner, P. A. (2020). The predictive value of the understanding of inconclusiveness and confounding for later mastery of the control-of-variables strategy. *Frontiers in Psychology*, 11.
- Posner, G.; Strike, K.; Hewson, P. Gertzog, W. (1982): Accommodation of a scientific conception: towards a theory of conceptual change. *Science education* 66(2), 211-227.
- Schauble, L.; Klopfer, L.; Raghavan, K. (1991): Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching* 28, 859-882.
- Schalk, L.; Edelsbrunner, P.; Deiglmayr, A.; Schumacher, R., Stern, E (2019): Improved application of the control-of-variables strategy as a collateral benefit of inquiry-based physics education in elementary school. *Learning and Instruction* 59, 34-45.
- Schmidt, D. (2016): Modellierung experimenteller Kompetenzen sowie ihre Diagnostik und Förderung im Biologieunterricht. Dissertation im Fachbereich Raum- und Umweltwissenschaften der Universität Trier.
- Scholtes, K. (2021): Lernen aus Fehlern beim forschenden Experimentieren im Chemieunterricht. Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes.
- Schrenk, M.; Gropengießer, H.; Groß, J.; Hammann, M.; Weitzel, H.; Zabel, J. (2018): Schülervorstellungen im Biologieunterricht. In: Groß, J.; Hammann, M.; Schmiemann, P.; Zabel, J. (Hrsg.): *Biologiedidaktische Forschung. Erträge für die Praxis*. Springer Spektrum, Berlin, 3-20.
- Schwartz, D.; Martin, T. (2004): Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction* 22(2), 129-184.
- Schwichow, M.; Croker, S. Zimmermann, C.; Höffler, T; Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review*, 39, 37-63.
- Schwichow, M.; Christoph, S.; Boone, W.J.; Härtig, H. (2016): The impact of sub-skills and item content on students' skills with regard to the control-of-variables-strategy. *International Journal of Science Education*, 38(2), 216-237.
- Schwichow, M., Osterhaus, C., & Edelsbrunner, P. A. (2020). The relation between the control-of-variables strategy and content knowledge in physics in secondary school. *Contemporary Educational Psychology*, 63.
- Schwichow, M.; Brandenburger, M; Brandenburger & Wilbers, J. (2022): Analysis of experimental design errors in elementary school: how do students identify, interpret, and justify controlled and confounded experiments? *International Journal of Science Education*, 44(1), 91-114.

- Siantuba, J.; Nkhata, L.; Jong, T. (2023): The impact of online inquiry-based learning environment addressing misconceptions on students' performance. *Smart Learning Environments* 10(22).
- Siler, S.A.; Klahr, D. (2012): Detecting, classifying and remediating: Childrens's explicit and implicit misconceptions about experimental design. In: R.W. Proctor; E.J. Capaldi (Hrsg.), *Psychology of science: Implicit and explicit processes* (137-180). Oxford University Press.
- Sinha, T.; Kapur, M. (2019): When productive failure fails. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Sinha, T.; Kapur, M. (2021): When Problem Solving Followed by Instruction works: evidence for productive failure. *Review of Educational Research* 91(5), 761-798.
- Sirhan, Ghassan (2007): Learning Difficulties in Chemistry. *Journal of Turkish Science Education* 4(2), 2-20.
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51 (11), 1–10.
- Treagust, D. F.; Duit, R.; Nieswandt, M. (2018): Sources of students' difficulties in learning Chemistry. *Education Química* 11(2), 228-235.
- Tulis, M.; Dresel, M.; Steuer, G. (2015): Learning from errors: Process and contextual conditions towards a model of individual processes within contexts. In: Gartmeier, M.; Gruber, H.; Hascher, T.; Heid, H. (Hrsg.): *Fehler: Ihre Funktionen im Kontext individueller und gesellschaftlicher Entwicklung*. Waxmann, Münster. 53-70.
- Vosniadou S. (2019): The development of students' understanding of science. *Frontiers in Education*, 4(32).
- Wandersee, J.H.; Mintzes, J.J.; Novak, J.D. (1994): Research on alternative conceptions in science. In: Gabel, D.L. (Hrsg.): *Handbook of research in science teaching and learning*. New York, Macmillan, 177-210.
- Zhang, Q., Fiorella L. (2023): An integrated model of learning from errors. *Educational Psychologist* 58 (1), 18-34.
- Zhang, L.; Sweller, J. (2024): Instructional sequences in science teaching: considering element interactivity when sequencing inquiry-based investigation activities and explicit instruction. *European Journal of Psychology of Education*, 1-11.
- Zimmermann, C. (2007): The development of scientific thinking skills in elementary and middle school. *Developmental Review* 27, 172-233.
- Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low-and high-achieving students. *Learning and Instruction*, 18(4), 337–353.

14 Danksagung

An dieser Stelle möchte ich die Gelegenheit nutzen, um von ganzem Herzen all jenen zu danken, die mich auf dem Weg zur Fertigstellung dieser Masterarbeit und während meines gesamten Studiums unterstützt und begleitet haben.

An erster Stelle gebührt mein tiefster Dank meinen Eltern, Melitta und Andreas. Ohne ihre finanzielle und emotionale Unterstützung, wäre dieser Meilenstein in meinem Leben nicht möglich gewesen. Sie haben es mir ermöglicht, mich frei zu entfalten, zu entwickeln und

meinen eigenen Weg zu gehen. Ihre Liebe, Geduld und ihr Glaube an mich waren meine stetige Quelle der Inspiration und vor allem Motivation.

Ein besonderer Dank geht auch an Linda, die mich im Schreibprozess meiner Masterarbeit ungemein unterstützt hat. Ihre prompte Antwort auf meine zahlreichen E-Mails und die Bereitschaft, sich trotz ihres dichten Terminplans immer Zeit für meine Anliegen zu nehmen, haben mir bei der raschen Fertigstellung meiner Arbeit geholfen.

Meinem Partner Denis möchte ich ebenfalls von Herzen danken. Seine unermüdliche Unterstützung, sein Verständnis und seine Ermutigung waren für mich während des Studiums und insbesondere in der Phase meines Abschlusses eine große Stütze. Selbst in Phasen, in denen ich das Gefühl hatte, keinen Ausweg zu haben, ist er mir mit Rat und Tat zur Seite gestanden, um auch diese Schwierigkeiten zu überwinden.

Ein Dankeschön geht auch an Mark, dessen Expertise und Rat mir wertvolle Orientierung geboten haben.

Sandra, Laura und Isabella, sehr gute Freundinnen, die ich in den ersten Jahren des Studiums kennengelernt habe, gebührt ebenfalls mein Dank. Die vielen Stunden, die wir gemeinsam in Seminaren verbracht, Herausforderungen gemeistert und unzählige Momente des Lachens geteilt haben, sind unvergesslich und haben maßgeblich zu meiner Freude am Studium beigetragen. Sie sind aus meinem Leben nicht mehr wegzudenken und ich bin dankbar für ihre Freundschaft.

Zuletzt möchte ich Frau Univ.-Prof. Dr. Möller meinen Dank aussprechen. Ihre fachliche Anleitung und Feedback waren entscheidend für meinen Studienabschluss.