

## Fachspezifische, komplexe Abbildungen verarbeiten – (auch) eine Frage der Überzeugung

Lara Magnus<sup>1,2</sup>, Kerstin Schütte<sup>1</sup> & Julia Schwanewedel<sup>1,2</sup>

<sup>1</sup> IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, Kiel

<sup>2</sup> Humboldt-Universität zu Berlin

*Zusammenfassung:* Abbildungen in den Naturwissenschaften verwenden spezielle grafische Konventionen und können große Datenmengen visualisieren. In der Kommunikation in den Naturwissenschaften sind sie – meist kombiniert mit Texten – omnipräsent. Kompetentes Arbeiten mit ihnen ist daher unverzichtbar, wird aber während der Schullaufbahn nur unzureichend erlernt. In dieser Studie wurde der Effekt einer Selbstwirksamkeitserwartung bezüglich des Arbeitens mit fachspezifischen, komplexen Abbildungen in der Biologie auf Testleistung untersucht. Dazu bearbeiteten  $N=205$  Bachelorstudierende der Biologie 2 auf Abituraufgaben basierende Biologieaufgaben, wobei das Testmaterial zwischen Gruppen variiert wurde. Multiple Regressionsanalysen zeigten bei der Biologieaufgabe mit fachspezifischen, komplexen Abbildungen einen statistisch signifikanten positiven Effekt der Selbstwirksamkeitserwartung auf die Testleistung, der durch die Anforderungen des Materials moderiert wurde. Es werden Implikationen für das Lehren von Biologie diskutiert.

**Schlüsselbegriffe:** Selbstwirksamkeitserwartung, komplexe Abbildungen, Kommunikation in Naturwissenschaften, Testleistung, Sekundarbildung

### *Processing subject-specific, complex pictures – (also) a question of one's belief*

*Summary:* Pictures in science use special graphical conventions, visualize a large amount of data points or variables. They are omnipresent in science communication – usually combined with texts. Therefore, working competently with these pictures is indispensable, but is insufficiently learned during the school career. This study examined the effect of a self-efficacy belief regarding working with subject-specific, complex pictures on test performance. For this,  $N=205$  undergraduate students of biology worked on 2 biology tasks based on German Abitur exams, whereby the test material varied between groups. Multiple regression analyses showed a statistically significant positive effect of the self-efficacy belief on test performance for the task with subject-specific, complex pictures, which was moderated by the requirements of the test material. Implications for teaching biology are discussed.

**Keywords:** Self-efficacy, complex representations, science communication, test performance, secondary education

In den Naturwissenschaften sind Abbildungen allgegenwärtig (Lemke, 1998; Roth & Pozzer-Ardenghi, 2013). Sie werden beispielsweise genutzt, um schwer fassbare, komplexe Phänomene abzubilden (Schönborn & Anderson, 2006) oder Daten zu visualisieren und dabei Erkenntnisse zu gewinnen und zu kommunizieren (Heßler, 2006). Auch während der

COVID-19-Pandemie wurde die Vielfältigkeit der Abbildungen im naturwissenschaftlichen Bereich sehr deutlich, als regelmäßig von Wissenschaftler:innen erstellte Abbildungen zu neuesten Forschungsergebnissen (z. B. Viruskonzentrationen in Proben von Patientengruppen) in die Öffentlichkeit getragen wurden. Kommentare unter solchen Abbildungen

*Author Note:* Lara Magnus ist inzwischen Mitarbeiterin der Universität Hamburg. Julia Schwanewedel ist nun Professorin für Biologiedidaktik an der Universität Hamburg.

Psychologie in Erziehung und Unterricht, 2021, 68, Open Access  
DOI 10.2378/peu2021.art29d  
© Ernst Reinhardt Verlag München Basel

in den sozialen Medien dokumentieren, dass mitunter große Schwierigkeiten bei deren Interpretation auftraten und sogar zum Teil Annahmen resultierten, die zu den eigentlichen Forschungsergebnissen konträr waren. Dies deutet darauf hin, dass Abbildungen ihr gewinnbringendes Potenzial nicht ausschöpfen können, wenn die Rezipierenden nicht wissen, wie sie die Abbildungen „lesen“ müssen. Anders als im Alltag allgegenwärtige Abbildungen (z. B. Diagramme mit wenigen Datenpunkten in Zeitungen) verwenden fachspezifische Abbildungen in den Naturwissenschaften häufig eine Vielzahl spezieller grafischer Konventionen oder visualisieren eine große Anzahl von Datenpunkten oder Variablen, was ihnen eine verhältnismäßig hohe Komplexität verleiht (Lowe, 2000; Trafton, Marshall, Mintz & Trickett, 2002). Verschiedene empirische Studien implizieren, dass gerade diese fachspezifischen, komplexen Abbildungen Probleme bei der Interpretation oder dem Erfassen der dargestellten Informationen hervorrufen können (z. B. Magnus, Schütte & Schwanewedel, 2020; Novick, Stull & Catley, 2012), dies, obwohl nationale Bildungsstandards für die Naturwissenschaften in vielen Ländern (z. B. NGSS Lead States, 2013; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005, 2020) das kompetente Arbeiten mit verschiedenen Abbildungsarten von Schüler:innen explizit fordern. Zudem gibt es Hinweise darauf, dass Abbildungen mitunter nur überflogen oder übersprungen werden, wenn sie in Kombination mit Texten präsentiert werden (Cromley, Snyder-Hogan & Luciw-Dubas, 2010).

Ein auch von Cromley et al. (2010) in Betracht gezogener Erklärungsansatz dafür, dass Abbildungen zum Teil nicht hinreichend beachtet werden, wodurch wertvolle Informationen verloren gehen, könnte die persönliche Überzeugung von den eigenen Fähigkeiten sein, sich Informationen aus komplexen Abbildungen erschließen zu können. Diese spezifische Selbstwirksamkeitserwartung könnte mit einer

bestimmten, auf das Arbeiten mit komplexen Abbildungen bezogenen Motivation einhergehen. Wird sie nicht zum Beispiel durch positive Erfahrungen mit solchen Abbildungen gestärkt, könnte dies negative Auswirkungen darauf haben, wie Personen an komplexe Abbildungen herangehen und wie sie die Informationsverarbeitung bewältigen. In dieser Studie haben wir untersucht, ob sich ein Effekt der Selbstwirksamkeitserwartung Biologiestudierender bezüglich des Arbeitens mit komplexen Abbildungen auf ihre Testleistung in Aufgaben nachweisen lässt, die unter anderem komplexe Abbildungen enthielten.

### *Die Rolle und Effekte von Abbildungen in den Naturwissenschaften*

Die Naturwissenschaften bedienen sich einer Vielzahl von Abbildungen, um erhobene Daten aufzubereiten und Ergebnisse zu vermitteln (Arsenault, Smith & Beauchamp, 2006; Heßler, 2006). Zudem lassen sich Phänomene auf makro-, mikro- und submikroskopischer Organisationsebene und auch abstrakte Konzepte durch Abbildungen wie beispielsweise Schemazeichnungen oder Fotografien sichtbar machen (Schönborn & Anderson, 2006). Abbildungen werden in naturwissenschaftlichen Fachpublikationen und in Lehrbüchern zumeist mit Text und teilweise mit weiteren Abbildungen kombiniert (z. B. Lemke, 1998). Der Grad an Redundanz, also inwieweit sich Informationen in den einzelnen Informationsmedien (Texte, Abbildungen) wiederholen, variiert. Wenn Informationen aus Abbildungen nicht im Text wiederholt werden, müssen Rezipierende zwingend die wesentlichen Elemente in Text und Bild identifizieren und anschließend zueinander in Beziehung setzen, ein Prozess, der als Kohärenzbildung bezeichnet wird (Seufert, 2003).

Zwar konnten empirische Studien positive Effekte von Abbildungen in Lern- und Testmaterialien auf die zu erbringende Leistung nachweisen (z. B. Eilam & Poyas, 2008; Lindner, Ihme, Saß & Köller, 2018), andere Studien

zeigten jedoch, dass gerade fachspezifische, komplexe Abbildungen Probleme bei der Aufgabenbearbeitung bereiten können (z. B. Magnus et al., 2020; Novick et al., 2012). Im Gegensatz zu alltäglichen Abbildungen (z. B. Beschilderungen an Flughäfen, einfache Graphen mit wenigen Variablen und Datenpunkten) zeichnen sich fachspezifische Abbildungen in den Naturwissenschaften durch folgende Eigenschaften aus: Sie sind häufig abstrakt, verwenden Farbcodierungen und eine Vielzahl spezieller grafischer Konventionen, bilden zum Teil eine große Anzahl von Datenpunkten oder Variablen ab (z. B. spezielle Wetterkarten in der Meteorologie) und sind somit komplexer als Abbildungen aus dem Alltag (Lowe, 2000; Traflet et al., 2002). Die Ergebnisse bisheriger Studien legen nahe, dass Schüler:innen nach Beendigung der Schullaufbahn nicht automatisch in der Lage sind, mit jeder Art von Abbildung in den Naturwissenschaften umzugehen (Magnus et al., 2020; Novick et al., 2012). Cromley et al. (2010) beobachteten, dass Biologiestudierende bei der Bearbeitung von Biologieaufgaben mit Text-Bild-Kombinationen die Bilder häufig lediglich überflogen oder sogar übersprungen haben. Wenn sie aber mit den Abbildungen arbeiteten, erreichten sie einen höheren Anteil kognitiver Aktivitäten auf erhöhtem Level im Vergleich zum Arbeiten mit reinem Text. Es stellt sich also die Frage, warum das gewinnbringende Potenzial von Abbildungen häufig nicht ausgeschöpft wird, diese mitunter nicht einmal beachtet werden. Über alle Schulfächer betrachtet werden vornehmlich Texte als Informationsmedium genutzt und der Umgang mit diesen wird – zumindest im Deutschunterricht – explizit geübt (Rosebrock & Nix, 2010). Dagegen wird das Lesen von Abbildungen häufig nicht explizit eingeübt, obwohl es gerade für komplexe, fachspezifische Abbildungen förderlich wäre (Lowe, 2000; Schönborn & Anderson, 2006). Dies könnte sich potenziell über die Schulzeit hinaus auf die Selbstwirksamkeitserwartung der Schüler:innen in Bezug auf das Arbeiten mit solchen Abbildungen auswirken.

### *Die Beziehung zwischen Selbstwirksamkeitserwartung und akademischer Leistung*

Die Selbstwirksamkeitserwartung, verankert in Banduras sozial-kognitiver Lerntheorie, bezeichnet die Überzeugung, durch eigenes Handeln gewünschte Ergebnisse erzielen zu können (Bandura, 2001). Sie ist somit aufgaben- und domänenspezifisch (Pajares, 1996), aber zwischen Aufgaben und zwischen Domänen positiv assoziiert, umso mehr, je ähnlicher diese sich jeweils sind (Bong, 1997). Es wurden verschiedene Quellen der Selbstwirksamkeitserwartung identifiziert: Erfolgserfahrungen, stellvertretende Erfahrung, verbale Überzeugungsarbeit und emotionale Erregung (Bandura, 1977). Das Erfahren von Erfolgen, die Lernende auf ihre eigene Fähigkeit und Anstrengung zurückführen können, wird dabei als einflussreichste Quelle betrachtet (z. B. Schwarzer & Jerusalem, 2002). Wird eine positive Selbstwirksamkeitserwartung in einem Gebiet aufgebaut, ist sie relativ stabil (Schwarzer & Jerusalem, 2002). Die spezifische Selbstwirksamkeitserwartung kann sich auf die Anstrengung, das Durchhaltevermögen und die Widerstandsfähigkeit einer Person bezüglich einer Tätigkeit in genau diesem Gebiet auswirken und somit letztendlich deren Leistung beeinflussen (Pajares, 1996). Zahlreiche empirische Studien belegen den Zusammenhang von Selbstwirksamkeitserwartung und Leistungen im naturwissenschaftlichen Bereich (z. B. Ardura & Pérez-Bitrián, 2019; Schütte, Frenzel, Asseburg & Pekrun, 2007). Dabei wurde jeweils eine spezifische Selbstwirksamkeitserwartung von Schüler:innen oder Studierenden erfragt und deren Effekt auf die Leistung, die in Form von Noten oder Leistungstests erfasst wurde, untersucht.

### *Fragestellung und Hypothese*

Auf das Arbeiten mit komplexen Abbildungen in den Naturwissenschaften und spezifisch in der Biologie übertragen, ist erwartbar, dass die während der Schullaufbahn und des bisherigen Studiums gesammelten Erfahrungen mit solchen

Abbildungen dazu führen, dass Biologiestudierende eine spezifische, diese Anforderung betreffende Selbstwirksamkeitserwartung ausgebildet haben. Eine solche Selbstwirksamkeitserwartung würde sich darauf auswirken, wie Personen an die Bearbeitung von Aufgaben herangehen, die unter anderem fachspezifische, komplexe Abbildungen enthalten, und welche Leistung sie letztendlich bei der Bearbeitung dieser Aufgaben erzielen.

Unter Kontrolle experimentell variiertes Anforderungen an die Arbeit mit komplexen Abbildungen auf die Testleistung (deren Effekte Gegenstand von Magnus et al., 2020, sind) wurde der Einfluss der Selbstwirksamkeitserwartung auf die Testleistung untersucht. Konkret wurde folgende Hypothese geprüft:

Die Selbstwirksamkeitserwartung bezüglich des Arbeitens mit komplexen Abbildungen in der Biologie hat einen positiven Effekt auf die Testleistung der Studierenden, dies auch unter Kontrolle von kognitiven Leistungsmaßen (sowohl fachspezifisch als auch allgemein), der durch die Art des zur Verfügung gestellten Aufgabenmaterials, also die Experimentalgruppenzugehörigkeit, moderiert wird. Bei Studierenden, die zum Lösen der Aufgaben Informationen aus Abbildungen entnehmen mussten (im Folgenden als Gruppe *keine Redundanz* bezeichnet), sollte ein positiver Zusammenhang von Selbstwirksamkeitserwartung und Testleistung vorliegen. Bei Studierenden, die zum Lösen derselben Aufgaben keine Informationen aus Abbildungen entnehmen mussten, wird dieser Zusammenhang grundsätzlich nicht erwartet (im Folgenden als Gruppe *Redundanz* bzw. Gruppe *nur Text* bezeichnet). In dem Maße, wie Studierende aber auch dann Abbildungen zum Lösen der Aufgaben hinzuziehen, wenn der Text alle dafür notwendigen Informationen enthält (Gruppe *Redundanz*), wird wiederum ein positiver Zusammenhang zwischen Selbstwirksamkeitserwartung und der Testleistung erwartet. Es ist dabei unerheblich, ob sich die Studierenden, die über die verschiedenen Versionen des Materials nicht informiert worden waren, der Art des Materials als redundant oder nicht redundant bewusst geworden sind.

## Methode

### *Stichprobe und Studiendurchführung*

An der Studie, die 2016/2017 durchgeführt wurde, nahmen  $N=240$  Biologiestudierende im Bachelorstudium teil. Studierende wurden aus der Stichprobe ausgeschlossen, wenn sie die Items zu Selbstwirksamkeitserwartung oder kognitiven Fähigkeiten nicht vollständig bearbeitet oder ihre letzte Biologienote nicht angegeben hatten. Es ergab sich eine finale Stichprobe von  $N=205$  ( $M=21.7$  Jahre,  $SD=3.3$ ) Studierenden, von denen 71,2% weiblich waren.

Die Studierenden wurden zufällig einer von drei Experimentalgruppen zugeordnet und aufgefordert, zwei Biologieaufgaben mit offenem Antwortformat in insgesamt 45 Minuten zu bearbeiten. Die Erstautorin führte alle Datenerhebungen durch und informierte die Teilnehmenden darüber, dass die Teilnahme freiwillig und anonym erfolgte und die Testleistung keinerlei Einfluss auf ihre Kursnote haben würde. Um Effekte der Aufgabenreihenfolge zu vermeiden, wurden die Biologieaufgaben in einem balancierten Design dargeboten, d. h., in jeder Gruppe gab es zwei Testheftversionen, in denen entweder zuerst Aufgabe A und dann B dargestellt war oder zuerst Aufgabe B und dann A. Das Aufgabenmaterial variierte zwischen den Gruppen: Die Gruppe *keine Redundanz* ( $n=68$ ) erhielt Material mit Abbildungen und Text und war zum Lösen der Aufgaben darauf angewiesen, Informationen sowohl aus Texten als auch aus den Abbildungen zu entnehmen. Im Material der Gruppe *Redundanz* ( $n=71$ ) wurden hingegen lösungsrelevante Informationen aus den Abbildungen zusätzlich im Text wiederholt, sodass die Aufgaben selbst dann gelöst werden konnten, wenn die Abbildungen nicht hinzugezogen wurden. Die Gruppe *nur Text* ( $n=66$ ) arbeitete mit Material, das ausschließlich aus Text bestand. Die Aufgabenstellungen waren in allen Gruppen identisch und die Teilnehmenden wurden nicht auf die unterschiedlichen Experimentalgruppen hingewiesen. Nach Bearbeitung der Biologieaufgaben hatten die Studierenden wiederum bis zu 45 Minuten Zeit, einen Fragebogen mit unter anderem den Skalen zur Erfassung ihrer Selbstwirksamkeitserwartung und kognitiven Fähigkeiten zu bearbeiten und ihre letzte Biologienote sowie einige demografische Daten anzugeben. Am Ende der Studie wurden die Teilnehmenden umfassend über die Ergebnisse unterrichtet. Die Teilnahme wurde mit Süßigkeiten honoriert und zusätzlich bestand die Möglichkeit, an einer Verlosung von Gutscheinen im Wert von € 5,- bis € 50,- teilzunehmen.

## Messinstrumente

### Selbstwirksamkeitserwartung

Zur Erfassung der Selbstwirksamkeitserwartung in Bezug auf das Arbeiten mit komplexen Abbildungen in der Biologie wurde eine Skala mit 4 Items entwickelt. Bei der Formulierung diente die gut etablierte Skala zur Erfassung der schulbezogenen Selbstwirksamkeitserwartung von Jerusalem und Satow (1999) als Orientierung. Die Items lauten wie folgt: „Selbst wenn ich eine komplexe biologische Abbildung nicht sofort verstehe, versuche ich weiterhin, sie mir zu erschließen.“, „Ich kann mir biologische Inhalte auch dann erschließen, wenn sie in für mich ungewohnten, neuen Abbildungen dargestellt sind.“, „Wenn Informationen in Biologie in komplexen Abbildungen dargestellt sind, macht es mir Spaß, die Informationen herauszuarbeiten.“, „Auch wenn eine biologische Abbildung auf den ersten Blick komplex erscheint, fühle ich mich in der Lage, mit ihr zu arbeiten.“ Die Studierenden schätzten sich auf einer 4-stufigen Likert-Skala ein (1 = „Stimme ganz zu“, 4 = „Stimme gar nicht zu“). Im Nachgang wurden die Items recodiert, sodass hohe Werte eine hohe Selbstwirksamkeitserwartung bedeuten. Die interne Konsistenz der Skala war akzeptabel (Cronbachs  $\alpha = .76$ ). Für weitere Analysen wurden Mittelwerte gebildet und  $z$ -transformiert.

### Kognitive Fähigkeiten

Zur Erfassung der kognitiven Fähigkeiten der Studierenden wurde die nonverbale Skala N2 des Kognitiven Fähigkeitstests (KFT) 4–12 + R eingesetzt (Heller & Perleth, 2000). Diese Skala besteht aus 25 Items, die jeweils zwei sinnvoll zusammenpassende Figuren oder Zeichnungen sowie eine dritte Figur oder Zeichnung zeigen. Aus weiteren fünf dargestellten Figuren oder Zeichnungen muss dann die eine ausgewählt werden, die auf die Art und Weise zur dritten Figur oder Zeichnung passt, wie die zweite zur ersten. Es wurde jeweils die Gesamtpunktzahl berechnet und einer  $z$ -Transformation unterzogen.

### Biologienote

Die von den Studierenden in unterschiedlichen Notensystemen angegebenen letzten Schulnoten im Fach Biologie wurden zur Vereinheitlichung dem Notensystem mit Werten von 1 (sehr gut) bis 6 (ungenügend) ohne Dezimalstellen zugeordnet, wobei nur Noten beobachtet wurden, die dem Bereich von 1 bis 4 entsprachen. Anschließend wurden die Werte  $z$ -transformiert.

## Material und Aufgabenbewertung

Das Aufgabenmaterial basierte auf Biologie-Abituraufgaben aus Nordrhein-Westfalen aus den Jahren 2008 (Aufgabe A) und 2009 (Aufgabe B) (Brixius, Jannan & Kunze, 2012, S. 2008-13, 2008-15–2008-17 bzw. S. 2009-1–2009-3). Aufgabe A war eine Genetikaufgabe und thematisierte Ursache und Symptome einer Muskelerkrankung, die genaue Aufgabenstellung lautete: „Erläutern Sie unter Bezug auf die im Material dargestellten Erkenntnisse die molekulargenetische Ursache des MERRF-Krankheitsbildes. Erläutern Sie darüber hinaus, wie es zu dem für das MERRF-Syndrom typischen Symptom der Muskelschwäche kommt.“<sup>1</sup> In Abhängigkeit von der Experimentalgruppe standen den Studierenden für Aufgabe A zwei biologiespezifische schematische Abbildungen aus dem Bereich der Molekulargenetik (eine Karte der mitochondrialen DNA sowie der Ausschnitt der Basenabfolge der mitochondrialen DNA einer gesunden und einer erkrankten Person), ein Graph mit zehn beschrifteten Datenpunkten und einer angepassten Kurve (zu Proteinen in den Mitochondrien Erkrankter) und textliches Material (3 Textblöcke) zur Verfügung (Gruppen *keine Redundanz* und *Redundanz*) oder rein textliches Material (Gruppe *nur Text*). Beispielhaft soll die Variation des Testmaterials der drei Experimentalgruppen anhand der Karte der mitochondrialen DNA erläutert werden. Dies ist eine schematische Abbildung eines ringförmigen DNA-Doppelstrangs, in dem mit den zugehörigen Basenpaarpositionen Regionen eingezeichnet sind, die für bestimmte Gene codieren. Zum Lösen der Aufgabe musste in dieser Abbildung abgelesen werden, welcher Bereich dieser DNA für das Gen einer bestimmten tRNA codiert. Diese Information konnte im Material der Gruppe *keine Redundanz* ausschließlich aus der Abbildung entnommen werden. Im Material der Gruppe *Redundanz* wurde der spezifische Bereich der DNA zusätzlich im Text genannt, sodass diese Information aus der Abbildung oder dem Text entnommen werden konnte. Da die Gruppe *nur Text* keine Abbildungen erhielt, wurde die Information aus der Abbildung wie bei Gruppe *Redundanz* in den Text eingefügt und konnte anders als im Material der Gruppe *Redundanz* lediglich aus diesem entnommen werden. In dem zu dieser Abbildung gehörenden Textblock waren weitere zum Lösen der Aufgabe relevante Informationen enthalten, die aber in allen Gruppen nur in textlicher Form dargeboten

wurden.<sup>2</sup> Da das eingesetzte Material nahezu dem der Original-Abituraufgaben entsprach, kann angenommen werden, dass der Schwierigkeitsgrad sowohl des bildlichen Materials als auch der Texte für Personen, die das Abitur erfolgreich absolviert haben, angemessen war. Es wurden nur Fachbegriffe verwendet, von denen erwartet werden kann, dass Schüler:innen sie bis zum Abitur (im Schulunterricht) erworben haben. Dies gilt für beide Aufgaben. Da in der Gruppe *Redundanz* Informationen aus bildlichen Repräsentationen im Text wiederholt wurden, wurde, um die Textlänge und damit die Leselast zwischen den Gruppen konstant zu halten, der Text der Gruppe *keine Redundanz* durch moderate Umformulierungen unter weitgehendem Erhalt der Textschwierigkeit verlängert.<sup>3</sup> Der Text des Aufgabenmaterials der Gruppe *nur Text* entsprach jeweils weitgehend dem der Gruppe *Redundanz*, wobei kleine Abweichungen dadurch begründet sind, dass im Material der Gruppe *Redundanz* (wie auch der Gruppe *keine Redundanz*) Verweise auf die Abbildungen enthalten waren, die im Material für Gruppe *nur Text* durch Umformulierungen entfernt wurden.

Die Aufgabenstellung von Aufgabe B lautete: „[E]rklären Sie die spezifischen Anpasstheiten der in stehenden Gewässern lebenden Feuersalamanderlarven im Kottenforst“ (Brixius et al., 2012, S. 2009-1). Die bildlichen Darstellungen in Aufgabe B umfassten eine Tabelle mit vier Werten (bzgl. des Gewichts der Larven in verschiedenen Gewässertypen), ein Balkendiagramm mit zwei Balken (zum Energiegehalt von Nahrungsorganismen) und einen Graphen mit 13 Datenpunkten (in dem Metamorphosegewicht gegen Dauer der Larvenphase aufgetragen wurde). Sie wiesen eine geringere Informationsdichte auf als die Abbildungen von Aufgabe A und waren insgesamt weniger biologiespezifisch, da alle genutzten Abbildungsarten auch im Alltag und in anderen Disziplinen präsent sind. Der Text von Aufgabe B wurde in zwei Textblöcken präsentiert.

Anders als in den beiden anderen Gruppen mussten Studierende der Gruppe *keine Redundanz* zwingend relevante Informationen in Abbildungen identifizieren und mit den textlich dargebotenen Informationen in Beziehung setzen, um die Aufgaben beantworten zu können, da diese Informationen nicht im Text standen. Im Aufgabenmaterial der Gruppen, denen Text- und Bildmaterial zur Verfügung stand, konnten jeweils fünf für das Lösen der Aufgabe relevante Informationseinheiten aus den

bildlichen Darstellungen entnommen werden, vier dem textlichen Material (in der Gruppe *Redundanz* zusätzlich der fünf im Text wiederholten ursprünglich bildlichen Informationseinheiten). Pro Aufgabe konnten die Studierenden somit Testleistungen im Wertebereich von 0 bis 9 erzielen; die Punktzahl zeigt an, wie viele der relevanten Informationen aus dem Material korrekt in die offene Antwort eingebunden wurden. Die Bewertung erfolgte für jede Aufgabe anhand eines dafür entwickelten Kategoriensystems. Die Objektivität der Bewertung wurde durch ein Interrating mit je Aufgabe 25 Prozent zufällig ausgewählten Antworttexten der Ursprungsstichprobe sichergestellt (vgl. Magnus et al., 2020). Für das Interrating berechnete Cohens-Kappa-Werte lagen über beide Aufgaben mit Werten zwischen  $\kappa = .62$  und  $\kappa = 1.0$  in einem nach Landis und Koch (1977) als substanzial bis fast perfekt eingeordneten Bereich der Übereinstimmung. Für Aufgabe B trat für eine der neun Informationseinheiten ein niedrigerer Kappa-Wert ( $\kappa = .42$ ) trotz hoher Übereinstimmung (89,1 %) auf, was darauf zurückgeht, dass sich die Häufigkeiten der Ausprägungen der Kategorie erheblich unterschieden. Der Kappa-Wert ist in solchen Fällen nicht sinnvoll interpretierbar.

### Regressionsanalysen

Die Hypothesen wurden anhand multipler Regressionsanalysen unter Nutzung der Software IBM SPSS statistics (Version 24) geprüft. Da die Variable Experimentalgruppe kategorial ist, wurden zwei Dummy-Variablen erstellt; die Gruppe *nur Text* fungierte als Referenzgruppe. Zur Überprüfung der Moderationshypothese wurden zwei Interaktionsterme gebildet, indem jede Dummy-Variable mit dem  $z$ -standardisierten Prädiktor Selbstwirksamkeitserwartung multipliziert wurde; die fünf Terme wurden gleichzeitig in das Modell eingefügt (Aiken & West, 1991). In zwei weiteren Modellen wurden die letzte in der Schule erhaltene Biologienote bzw. die kognitiven Fähigkeiten als zusätzliche Prädiktorvariablen für die Testleistung einbezogen. Die Benjamini-Hochberg-Prozedur wurde mit einem  $q$ -Wert von .05 auf die Gesamtheit der Tests angewendet, um den multiplen Tests Rechnung zu tragen und so den erwarteten Anteil falschpositiver Aussagen zu kontrollieren. Die adjustierten  $p$ -Werte werden für die Fälle berichtet, in welchen die Fehlerkontrolle zu einer anderen Entscheidung führt als der zugrunde liegende Test.

Tab. 1: Mittelwerte und Standardabweichungen nach Experimentalgruppe.

Variable	Gruppe <i>keine Redundanz</i>		Gruppe <i>Redundanz</i>		Gruppe <i>nur Text</i>	
	M	SD	M	SD	M	SD
Testleistung Aufgabe A gesamt	3.58	2.49	4.76	2.90	6.31	2.41
Testleistung Aufgabe A Score_Abb	1.88	1.54	2.97	1.56	3.60	1.34
Testleistung Aufgabe A Score_Text	1.70	1.23	1.79	1.53	2.71	1.35
Testleistung Aufgabe B gesamt	5.91	2.13	6.23	1.64	5.77	1.62
Selbstwirksamkeitserwartung	3.02	0.51	3.13	0.57	3.20	0.48
Letzte Biologienote	1.68	0.72	1.77	0.68	1.77	0.63
Kognitive Fähigkeiten	14.65	4.16	15.15	4.62	14.76	4.51

*Anmerkung:* Für Aufgabe A erfolgte zusätzlich zur Betrachtung der Testleistung der gesamten Aufgabe eine separate Betrachtung der Testleistung für Informationseinheiten, die aus Abbildungen entnommen werden mussten bzw. konnten (in Gruppe *keine Redundanz* und *Redundanz*, in Gruppe *nur Text* wurden sie hingegen rein textlich dargeboten; Score\_Abb), und Informationseinheiten, die in allen drei Gruppen ausschließlich im Text gegeben waren (Score\_Text). Für Aufgabe B erfolgte diese separate Betrachtung nicht.

## Ergebnisse

In Tabelle 1 sind die Mittelwerte und Standardabweichungen aller in den Regressionsmodellen berücksichtigten Variablen für die drei Experimentalgruppen angegeben. Mittels einfaktorieller Varianzanalyse wurde überprüft, ob sich die Selbstwirksamkeitserwartung zwischen den Gruppen unterschied. Es wurde kein statistisch signifikanter Unterschied beobachtet,  $F(2, 202) = 1.89$ ,  $p = .153$ .

### Multiple Regression für Aufgabe A

Die Ergebnisse der multiplen Regression mit Experimentalgruppe als Moderator (Gruppe *nur Text* als Referenzgruppe) und Selbstwirksamkeitserwartung bezüglich des Arbeitens mit komplexen Abbildungen in der Biologie (SWE) als Prädiktor zeigten, dass dieses Modell einen statistisch signifikanten Erklärungsbeitrag zur Varianz der Testleistung in Aufgabe A leistete (korrigiertes  $R^2 = .192$ ;  $R^2 = .212$ ,  $F(5, 199) = 10.69$ ,  $p < .001$ ). Der für Dummy 1 geschätzte Regressionskoeffizient repräsentiert den Unterschied in der mittleren Testleistung zwischen Gruppe *keine Redundanz* und Gruppe *nur Text* (Referenzgruppe) zwischen Studierenden bei gleichermaßen dem Mittelwert entsprechend ausgeprägter SWE. Dieser wies einen statistisch signifikanten Effekt von  $B = -2.437$ ,  $p < .001$  auf die Testleistung aus, d. h., die Testleistung war um ca. 2.4

(von 9) Punkte geringer, wenn Informationen aus Text und Bild entnommen werden mussten und nicht rein textlich dargeboten wurden. Auch der den Unterschied zwischen Gruppe *Redundanz* und Gruppe *nur Text* repräsentierende geschätzte Regressionskoeffizient für Dummy 2 zeigte einen statistisch signifikanten negativen Effekt auf die Testleistung,  $B = -1.378$ ,  $p = .002$  – auch die Gruppe *Redundanz* erzielte eine geringere Punktzahl als die Gruppe *nur Text* (die Effekte des Testmaterials auf die Testleistung sind Gegenstand von Magnus et al., 2020). Darüber hinaus wurde für die SWE ein statistisch signifikanter positiver Effekt auf die Testleistung beobachtet,  $B = 1.087$ ,  $p = .002$ , welcher den Effekt (simple slope) in der Referenzgruppe *nur Text* repräsentiert: Eine höhere SWE ging mit einer besseren Testleistung einher. Diese Beziehung zwischen SWE und Testleistung in Aufgabe A fand sich nicht unabhängig von der Experimentalgruppenzugehörigkeit, was durch den statistisch signifikanten Effekt des Interaktionsterms  $\text{Dummy } 2 \times \text{SWE}$  ersichtlich wurde,  $B = -1.169$ ,  $p = .009$ . Für Studierende, die mit redundantem Material arbeiteten, wurde ein statistisch signifikant vom Effekt für die Referenzgruppe *nur Text* abweichender Zusammenhang von SWE und Testleistung beobachtet ( $1.087 - 1.169 = -0.082$ ). Der Interaktionsterm  $\text{Dummy } 1 \times \text{SWE}$  war nicht statistisch signifikant,  $B = -0.386$ ,  $p = .415$ ; das heißt, für die

Gruppe *keine Redundanz* wurde ein Effekt der SWE auf die Testleistung beobachtet, welcher nicht vom Effekt in der Gruppe *nur Text* verschieden ist.

Um den beobachteten Moderationseffekt genauer zu untersuchen, wurden analog zum Ausgangsmodell mit Referenzgruppe *nur Text* die bedingten einfachen Regressionsgewichte (simple slopes) für den Effekt von SWE auf die Testleistung in Aufgabe A für die beiden Experimentalgruppen auf statistische Signifikanz überprüft, in welchen zusätzlich zum Text fachspezifische, komplexe Abbildungen dargeboten wurden (Abbildung 1). Dazu wurden neue Dummy-Variablen und Interaktionsterme erstellt, wobei einmal Gruppe *keine Redundanz* und einmal Gruppe *Redundanz* als Referenzgruppe fungierte. Für Gruppe *keine Redundanz* ergab sich ein signifikanter Effekt von  $B=0.701$ ,  $p=.030$ , für Gruppe *Redundanz* erwies sich, wie oben beschrieben, kein bedeutsamer Zusammenhang von SWE und Testleistung mit  $B=-0.082$ ,  $p=.769$ . Für die Gruppen *keine Redundanz* und *nur Text* (zu Gruppe *nur Text* als Referenzgruppe s. oben:  $B=1.087$ ,  $p=.002$ ) besteht somit,

anders als für Gruppe *Redundanz*, eine statistisch signifikante positive Beziehung zwischen SWE und Testleistung (Abbildung 1).

#### Berücksichtigung des Repräsentationstyps

Im Folgenden wurde die Testleistung in Aufgabe A zerlegt in die Informationseinheiten, die in der Gruppe *keine Redundanz* und *Redundanz* aus Abbildungen entnommen werden mussten bzw. konnten (diese wurden in der Gruppe *nur Text* rein textlich dargeboten; Score\_Abb), und die Informationseinheiten, die in allen drei Gruppen nur im Text gegeben waren (Score\_Text). Die oben dargestellten Regressionsanalysen wurden mit diesen neuen Kriteriumsvariablen wiederholt. So konnte unter anderem untersucht werden, ob sich der Effekt der SWE auf die Testleistung in Gruppe *keine Redundanz* in erster Linie auf die Testleistung bezüglich der Informationsentnahme aus Abbildungen bezog. Wie für die gesamte Aufgabe A zeigte sich, dass dieses Modell einen statistisch signifikanten Anteil der Varianz in der Testleistung aufklärt, korrigiertes  $R^2=.217$ ,  $R^2=.236$ ,  $F(5, 199)=12.31$ ,

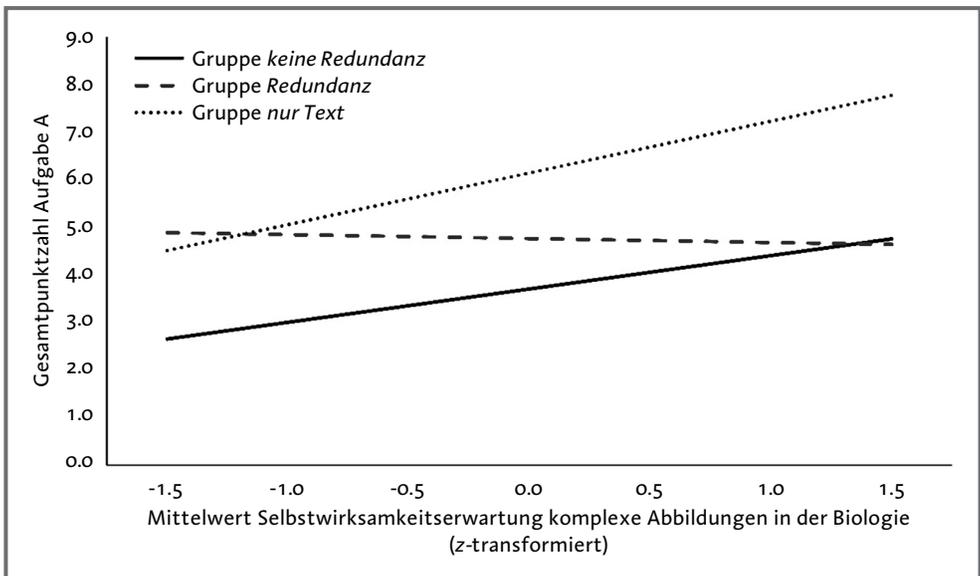


Abb. 1: Testleistung in Aufgabe A als Funktion von Selbstwirksamkeitserwartung und Experimentalgruppe (simple slopes).

$p < .001$  für Score\_Abb und korrigiertes  $R^2 = .127$ ,  $R^2 = .148$ ,  $F(5, 199) = 6.94$ ,  $p < .001$  für Score\_Text. Der den Unterschied in der Testleistung zwischen Gruppe *keine Redundanz* und Gruppe *nur Text* repräsentierende geschätzte Regressionskoeffizient für Dummy 1 zeigte wie für den Gesamtscore von Aufgabe A einen signifikanten Effekt auf Score\_Abb und Score\_Text,  $B = -1.555$ ,  $p < .001$  bzw.  $B = -0.881$ ,  $p < .001$ . Auch der den Unterschied zwischen Gruppe *Redundanz* und Gruppe *nur Text* repräsentierende geschätzte Regressionskoeffizient für Dummy 2 wies einen statistisch signifikanten Effekt auf die Testleistung auf,  $B = -0.543$ ,  $p = .031$  für Score\_Abb und  $B = -0.836$ ,  $p < .001$  für Score\_Text. Der Regressionskoeffizient für die SWE betrug  $B = 0.532$ ,  $p = .008$  (für Score\_Abb) und  $B = 0.555$ ,  $p = .003$  (für Score\_Text) – der statistisch signifikante Einfluss der SWE auf die Subscores war in der Gruppe *nur Text* nahezu gleich groß. Eine höhere SWE ging mit einer besseren Testleistung einher. Wie für den Gesamtscore der Aufgabe zeigte sich ein statistisch signifikanter Effekt des Interaktionsterms Dummy 2  $\times$  SWE,  $B = -0.572$ ,  $p = .025$  für Score\_Abb bzw.  $B = -0.597$ ,  $p = .012$  für Score\_Text. Auch bei einzelner Betrachtung der Informationseinheiten aus Abbildungen oder Texten wurde die Beziehung zwischen SWE und der Testleistung in Aufgabe A also durch die Experimentalgruppenzugehörigkeit der Studierenden moderiert; in der Gruppe *Redundanz* war die Testleistung unabhängig von der SWE. Der Interaktionsterm Dummy 1  $\times$  SWE hatte keinen statistisch signifikanten Effekt auf die Testleistung,  $B = -0.090$ ,  $p = .739$  für Score\_Abb bzw.  $B = -0.296$ ,  $p = .238$  für Score\_Text.

Die bedingten einfachen Regressionsgewichte zeigten wiederum lediglich für Gruppe *keine Redundanz* einen statistisch signifikanten Effekt von  $B = 0.442$ ,  $p = .016$  auf den Score\_Abb, für Gruppe *Redundanz* erwies sich dieser Zusammenhang als nicht statistisch signifikant,  $B = -0.040$ ,  $p = .802$ . Für die Gruppen *keine Redundanz* und *nur Text* besteht somit eine statistisch signifikante positive Beziehung zwischen der SWE bezüglich des Arbeitens mit komplexen Abbildungen in der Biologie und der Entnahme von Infor-

mationen aus Abbildungen (Score\_Abb), während dieser Zusammenhang für die Gruppe *Redundanz* nicht vorliegt. Für die Entnahme von Informationen aus dem Text gibt es lediglich für die Gruppe *nur Text* einen statistisch signifikanten Zusammenhang zwischen SWE und Testleistung (Score\_Text), beide bedingten einfachen Regressionsgewichte für den Effekt der SWE auf Score\_Text für die Gruppen *keine Redundanz* bzw. *Redundanz* waren nicht statistisch signifikant,  $B = 0.259$ ,  $p = .130$  bzw.  $B = -0.042$ ,  $p = .776$ .

### Berücksichtigung objektiver Leistungsmaße

Um Effekte objektiver Leistungsmaße auf die Testleistung zu untersuchen und um zu prüfen, ob die SWE nach Berücksichtigung dieser auf die Testleistung weiterhin einen eigenständigen Beitrag zur Varianzaufklärung leistet, wurden die letzte in der Schule erhaltene Biologienote und allgemeine kognitive Fähigkeiten nacheinander ins Modell eingefügt (Modell 2 und 3, Tabelle 2). Die Biologienote klärte mit 2,1 % ( $F(1, 198) = 5.36$ ,  $p = .022$ ) einen statistisch signifikanten zusätzlichen Varianzanteil am Gesamtscore von Aufgabe A auf (korrigiertes  $R^2 = .209$ ;  $R^2 = .233$ ,  $F(6, 198) = 10.00$ ,  $p < .001$ ). Für den Score\_Abb führte das Einfügen der Biologienote ins Modell nicht zu einer statistisch signifikanten Änderung des  $R^2$ ,  $\Delta R^2 = .010$  ( $F(1, 198) = 2.69$ ,  $p = .102$ ). Für den Score\_Text zeigte sich hingegen wie für den Gesamtscore eine statistisch signifikante Änderung des  $R^2$  um  $.028$  ( $F(1, 198) = 6.76$ ,  $p = .010$ ). Dieses Modell klärte insgesamt einen statistisch signifikanten Anteil der Testleistung für Score\_Text auf, korrigiertes  $R^2 = .152$ ,  $R^2 = .177$ ,  $F(6, 198) = 7.08$ ,  $p < .001$ . Wurde zusätzlich der Prädiktor allgemeine kognitive Fähigkeiten ins Modell eingefügt (korrigiertes  $R^2 = .238$ ;  $R^2 = .264$ ,  $F(7, 197) = 10.08$ ,  $p < .001$ ), wurden weitere 3,1 % ( $F(1, 197) = 8.33$ ,  $p = .004$ ) der Varianz der Testleistung (Gesamtscore) in Aufgabe A erklärt. Auch für Score\_Abb und Score\_Text trugen die kognitiven Fähigkeiten statistisch signifikant zur Varianzaufklärung bei,  $\Delta R^2 = .016$  ( $F(1, 197) = 4.32$ ,  $p = .039^4$ ) für Score\_Abb (korrigiertes  $R^2 = .236$ ;  $R^2 = .263$ ,

$F(7, 197) = 10.03, p < .001$  und  $\Delta R^2 = .041$  ( $F(1, 197) = 10.26, p = .002$ ) für *Score\_Text* (korrigiertes  $R^2 = .190; R^2 = .217, F(7, 197) = 7.82, p < .001$ ). Tabelle 2 zeigt den Einfluss sämtlicher Prädiktoren in den Modellen 2 und 3 auf die Testleistung (Gesamtscore, *Score\_Abb* und *Score\_Text*) in Aufgabe A. Im Vergleich zu Modell 1 änderte sich jeweils durch Hinzufügen der Prädiktoren Biologienote und kognitive Fähigkeiten der Einfluss der SWE und des Moderators Experimentalgruppe nur geringfügig. Erwartungsgemäß war der Zusammenhang von Biologienote und Testleistung positiv. Der Betrag des Regressionskoeffizienten war dabei jeweils nahezu halb so groß wie der Wert für die SWE (Tabelle 2). Für *Score\_Abb* wurde der Effekt der Biologienote auf die Testleistung nicht statistisch signifikant. Wurden zusätzlich allgemeine kognitive Fähigkeiten im Modell berücksichtigt, zeigte der statistisch signifikante Effekt für alle drei Scores an, dass bessere kognitive Fähigkeiten zu einer besseren Testleistung führten, während der Effekt der Biologienote nur noch für *Score\_Text* statistisch signifikant war (Tabelle 2). Aber auch der Effekt der kogni-

tiven Fähigkeiten auf die Testleistung war bei den drei betrachteten (Sub-)Scores jeweils deutlich geringer als der Effekt der SWE (Tabelle 2).

### Multiple Regression für Aufgabe B

Das Regressionsmodell mit Experimentalgruppen und SWE sowie deren Interaktionstermen erklärte, anders als bei Aufgabe A, keinen statistisch signifikanten Varianzanteil in der Testleistung in Aufgabe B (korrigiertes  $R^2 = -.010; R^2 = .015, F(5, 199) = 0.59, p = .708$ ). Weder der Unterschied zwischen Gruppe *keine Redundanz* und Gruppe *nur Text* (Dummy 1), noch der Unterschied zwischen Gruppe *Redundanz* und Gruppe *nur Text* (Dummy 2) hatten einen statistisch signifikanten Effekt auf die Testleistung,  $B = 0.175, p = .584$  bzw.  $B = 0.483, p = .125$ . Ebenso zeigten sowohl die SWE ( $B = 0.201, p = .418$ ) als auch die beiden Interaktionsterme keinen statistisch signifikanten Effekt auf die Testleistung. Für die Interaktion Dummy 1  $\times$  SWE ergab sich  $B = -0.174, p = .606$  und für die Interaktion Dummy 2  $\times$  SWE war  $B = -0.162$  und  $p = .612$ .

Tab. 2: Regressionsmodelle 2 und 3 zur Vorhersage der Testleistung von Aufgabe A.

	Modell 2			Modell 3		
	B (SE)			B (SE)		
	gesamt	Abbildungen	Texte	gesamt	Abbildungen	Texte
Dummy 1 (Gr. <i>nur Text</i> vs. Gr. <i>keine Redundanz</i> )	-2.547*** (.444)	-1.600*** (.255)	-0.947*** (.235)	-2.534*** (.436)	-1.595*** (.252)	-0.940*** (.229)
Dummy 2 (Gr. <i>nur Text</i> vs. Gr. <i>Redundanz</i> )	-1.400** (.434)	-0.551* (.249)	-0.849*** (.229)	-1.455*** (.426)	-0.574* (.247)	-0.881*** (.224)
SWE <sup>a</sup>	0.931** (.349)	0.469* (.200)	0.462* (.185)	0.870* (.344)	0.443* (.199)	0.427* (.181)
Interaktion 1 (Dummy 1 $\times$ SWE <sup>a</sup> )	-0.372 (.467)	-0.084 (.268)	-0.288 (.247)	-0.325 (.459)	-0.064 (.266)	-0.261 (.241)
Interaktion 2 (Dummy 2 $\times$ SWE <sup>a</sup> )	-1.066* (.442)	-0.530 <sup>†</sup> (.253)	-0.536* (.234)	-1.030* (.434)	-0.515 <sup>†</sup> (.251)	-0.515* (.228)
Biologienote <sup>a</sup>	-0.425* (.184)	-0.173 (.105)	-0.253* (.097)	-0.346 (.183)	-0.140 (.106)	-0.206* (.096)
Kognitive Fähigkeiten <sup>a</sup>				0.509** (.176)	0.212 <sup>†</sup> (.102)	0.297** (.093)

Anmerkungen: Gr. = Experimentalgruppe; SWE = Selbstwirksamkeitserwartung bezüglich des Arbeitens mit komplexen Abbildungen in der Biologie. <sup>a</sup> (z-standardisiert). \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ . <sup>†</sup> Nach Korrektur mittels Benjamini-Hochberg-Prozedur nicht mehr statistisch signifikant.

## Diskussion

Die Ergebnisse dieser Studie legen nahe, dass das Lösen von Aufgaben mit fachspezifischen, komplexen Abbildungen auch eine Frage der persönlichen Überzeugung ist hinsichtlich der eigenen Fähigkeit, die dadurch bedingten spezifischen Anforderungen zu bewältigen. Die Selbstwirksamkeitserwartung bezüglich des Arbeitens mit komplexen Abbildungen in der Biologie wies für Aufgabe A einen positiven Effekt auf die Testleistung auf. Diese Beziehung wurde durch die Experimentalgruppenzugehörigkeit der Studierenden moderiert. Der positive Effekt der Selbstwirksamkeitserwartung auf die Testleistung konnte, wie erwartet, bei Studierenden beobachtet werden, die die Abbildungen zum Lösen der Aufgabe heranziehen mussten (Gruppe *keine Redundanz*), nicht aber bei Studierenden, denen dieselben Abbildungen dargeboten wurden, die aber alle lösungsrelevanten Informationen aus dem Text entnehmen konnten (Gruppe *Redundanz*). Überraschend war die Beobachtung eines statistisch signifikanten Effekts für Gruppe *nur Text*. Die Ergebnisse zeigten zudem, dass die Selbsteinschätzung, mit der Personen komplexen Abbildungen begegnen, selbst dann einen Effekt auf die Testleistung hatte, wenn objektive Leistungsmaße wie die Biologienote oder allgemeine kognitive Fähigkeiten im Modell berücksichtigt wurden. Wie erwartet, hatten diese beiden Variablen einen statistisch signifikanten Effekt auf die Testleistung (wenngleich diese bei der gegebenen Stichprobengröße nach Korrektur für multiple Tests nicht mehr durchgängig statistisch signifikant waren), der Effekt der Selbstwirksamkeitserwartung blieb dabei aber nahezu unverändert groß. Als objektive Maße der kognitiven Fähigkeiten teilen sich KFT und Biologienote Varianzanteile, weshalb der Effekt der Biologienote nicht mehr statistisch signifikant war, als der KFT ins Modell eingefügt wurde.

Die Ergebnisse von Aufgabe B unterstützen die Annahme, dass sich die hier betrachtete Selbstwirksamkeitserwartung spezifisch auf das

Arbeiten mit fachspezifischen, komplexen Abbildungen auswirkt. Es gab in Aufgabe B keinen statistisch signifikanten Effekt der Selbstwirksamkeitserwartung auf die Testleistung, was vermutlich daran lag, dass die Abbildungen der Aufgabe weniger fachspezifisch waren und wahrscheinlich als weniger komplex wahrgenommen wurden. Im Folgenden beschränken wir uns auf die Interpretation der Ergebnisse von Aufgabe A.

### Moderatoreffekt

Für verschiedene Leistungen im naturwissenschaftlichen Bereich wurde eine entsprechende Selbstwirksamkeitserwartung als Einflussfaktor identifiziert (z. B. Ardura & Pérez-Bitrián, 2019; Schütte et al., 2007). In dieser Studie konnten wir zeigen, dass sich eine Selbstwirksamkeitserwartung bezüglich des Arbeitens mit fachspezifischen, komplexen Abbildungen in der Biologie auf die Testleistung in Biologie auswirkt, wenn derartige Abbildungen zum Lösen der Aufgabe herangezogen werden müssen. Wie beschrieben, werden Strategien zur Texterschließung in der Schule im Deutschunterricht geübt, während Strategien zum Erschließen von Informationen in Abbildungen möglicherweise nicht explizit gelehrt werden (Schönborn & Anderson, 2006; Rosebrock & Nix, 2010; Ziepprecht, 2016). Diese sollten aber gerade im naturwissenschaftlichen Fachunterricht gezielt gefördert werden, da naturwissenschaftliche Erkenntnisse über diverse fachspezifische, komplexe Abbildungen kommuniziert werden, die nicht selbsterklärend sind, sondern spezieller Lesefähigkeiten bedürfen (Lowe, 2000). Ähnlich wie Ziepprecht (2016) für Schüler:innen der Sekundarstufe I zeigte auch unsere Studie mit Studierenden, dass es ihnen leichter fiel, Informationen aus Texten als aus mit Text kombinierten fachspezifischen Abbildungen zu entnehmen (Hauptgegenstand der Studie Magnus et al., 2020). Wie erwähnt, sind Erfolgserfahrungen die wichtigste Quelle der Selbstwirksamkeitserwartung (Schwarzer & Jerusalem, 2002). Bleiben explizite Erfolgserfahrungen beim Erschlie-

ßen von Informationen aus fachspezifischen, komplexen Abbildungen in Ermangelung einer entsprechenden Förderung im Fachunterricht aus, wird die Selbstwirksamkeitserwartung in diesem Bereich vermutlich nicht ausreichend gestärkt. Eine mögliche Erklärung für unsere Befunde wäre also, dass Studierende, die in der Vergangenheit Erfolg beim Arbeiten mit komplexen Abbildungen erlebt haben, das notwendige Vertrauen in sich selbst und die nötige Motivation hatten, Aufgaben mit fachspezifischen, komplexen Abbildungen zu bearbeiten, und deshalb besser abschnitten als Studierende mit geringeren Erfolgserfahrungen. Dass sich der positive Effekt der Selbstwirksamkeitserwartung auf die Testleistung in Gruppe *keine Redundanz* ausschließlich auf die Testleistung bezüglich der aus Abbildungen und nicht der aus Text zu entnehmenden Informationen bezog, stützt unsere Annahme, dass die Selbstwirksamkeitserwartung bezüglich fachspezifischer, komplexer Abbildungen als ein eigenständiges Konstrukt behandelt werden sollte.

Die Ergebnisse von Gruppe *Redundanz* legen – wie bereits in Magnus et al. (2020) diskutiert – nahe, dass die Studierenden vor allem mit dem Text gearbeitet haben, da diese Gruppe gegenüber der Gruppe *nur Text* keine bessere Testleistung erzielte. Die Abbildungen wurden also nicht zwingend hinzugezogen, was auch nicht notwendig war, da in dieser Experimentalbedingung alle lösungsrelevanten Informationen im Text gegeben waren. Deshalb ist es schlüssig, dass bei Gruppe *Redundanz* keine Beziehung zwischen Selbstwirksamkeitserwartung und Testleistung vorlag.

Anders als erwartet, wirkte sich bei Gruppe *nur Text*, obwohl die Studierenden nicht mit Abbildungen arbeiteten, die Selbstwirksamkeitserwartung positiv auf die Testleistung aus. Eine Erklärung könnte sein, dass für diese Studierenden, anders als für Studierende der Gruppen, die Abbildungen im Testmaterial vorfanden, kein Anker für die Vorstellung von fachspezifischen, komplexen Abbildungen gesetzt worden war. Sie waren für die Bearbei-

tung der SWE-Skala allein auf ihren persönlichen Erfahrungshintergrund angewiesen. Möglicherweise liegen deren Urteilen allgemeinere Urteile über ihre Leistungen im Fach Biologie zugrunde.

### *Limitationen*

Die Ergebnisse stützen die Annahme, dass die untersuchte Selbstwirksamkeitserwartung ein eigenständiges Konstrukt ist, das von einer Selbstwirksamkeitserwartung bezüglich allgemeiner Fähigkeiten in der Biologie (und bezogen auf andere fachspezifische Anforderungen) abgrenzbar ist. Die empirische Überprüfung dieser Annahme anhand entsprechender Selbstwirksamkeitsmaße steht jedoch noch aus. Des Weiteren ist unklar, welche Vorstellungen von fachspezifischen, komplexen Abbildungen in der Biologie Studierende der Gruppe *nur Text* hatten. Die Selbstwirksamkeitserwartung wurde gemeinsam mit Kontrollvariablen und demographischen Daten nach Bearbeitung der Biologieaufgaben abgefragt; die Gruppen, die mit fachspezifischen, komplexen Abbildungen gearbeitet haben, konnten also eine konkrete Vorstellung von solchen Abbildungen haben. Abgesehen davon ist aber nicht davon auszugehen, dass die Aufgabenbearbeitung die Einschätzung der Selbstwirksamkeitserwartung beeinflusst hat. Zum einen gilt die Selbstwirksamkeitserwartung, wenn sie einmal gebildet ist, als relativ stabil (Schwarzer & Jerusalem, 2002), zum anderen hatten die Studierenden zum Zeitpunkt der Bearbeitung der Selbstwirksamkeitserwartungs-Items keine Rückmeldung zu ihrem Abschneiden in den Testaufgaben bekommen.

### *Implikationen*

Die Selbstwirksamkeitserwartung bezüglich des Arbeitens mit komplexen Abbildungen in der Biologie zeigte einen Effekt auf die Testleistung von Studierenden, wenn diese fachspezifische, komplexe Abbildungen zum Lösen der Aufgabe heranziehen mussten. Die Ergebnisse

zeigten also, dass unter Kontrolle objektiver Leistungsmaße selbst Personen, die sich für ein Studium der Biologie entschieden haben, nicht zwangsläufig mit komplexen Abbildungen in der Biologie arbeiten konnten. Darauf, die entsprechende Selbstwirksamkeitserwartung bei Schüler:innen möglichst frühzeitig zu stärken und somit einen Zugang zum Lesen solcher Abbildungen zu erleichtern, sollten Lehrkräfte schon während der Schulzeit im naturwissenschaftlichen Unterricht hinarbeiten. Dazu sollten die Quellen der Selbstwirksamkeitserwartung in den Blick genommen werden und unter ihnen besonders die Erfolgserfahrungen. Zum Vermitteln von Erfolgen gehört das Unterstützen von Bewältigungsstrategien (Schwarzer & Jerusalem, 2002). Mehrere Studien haben sich bereits mit speziellen Anforderungen von fachspezifischen, komplexen Abbildungen an Lernende beschäftigt und Vorschläge ausgearbeitet, wie zur Bewältigung dieser Anforderungen beigetragen werden kann, um Lernende beim Lesen und Verstehen von Abbildungen zu unterstützen (Cromley et al., 2013; Schönborn & Anderson, 2006). Schönborn und Anderson (2006) erstellten zum Beispiel zehn Richtlinien für das Lehren und Lernen mit Abbildungen in der Biochemie. So empfehlen sie unter anderem, die spezielle Symbolik in diesen Abbildungen explizit zu unterrichten und Lernende mit verschiedenen Repräsentationstypen zu demselben Phänomen zu konfrontieren, damit sie sich mit den einzelnen Repräsentationstypen kritisch auseinandersetzen und sich Gemeinsamkeiten und Unterschiede bewusst machen (Schönborn & Anderson, 2006). Auch das Konstruieren von Abbildungen durch Lernende kann zu einem routinierteren Arbeiten mit diesen führen. Scherb und Nitz (2020) zeigten in einer Interviewstudie mit (angehenden) Biologielehrkräften allerdings, dass diese im Unterricht den Fokus eher auf das Konstruieren realistischer anstatt abstrakter Abbildungen legen. Dazu passt, wie von Kotzebue, Gerstl und Nerdel (2015) beobachtet haben, dass selbst Studierende Schwierigkeiten bei der Konstruktion von Diagrammen in biologi-

schen Kontexten haben. In der Studie wurden Aspekte benannt, die Lehrkräfte beachten sollten, wenn sie Lernende im Arbeiten mit Diagrammen unterrichten. Zudem entwickelten Nitz, Prechtl und Nerdel (2014) ein Instrument, mit dem unter anderem die Wahrnehmung von Schüler:innen bezüglich der Verwendung bildlicher Repräsentationen im Biologieunterricht erfasst werden kann und das Lehrenden zur Reflexion dienen könnte. Es ist an Lehrenden, solche Richtlinien und Empfehlungen im Unterricht zu berücksichtigen, und an Forschenden, weitere Richtlinien für Abbildungen in anderen Themengebieten zu entwickeln, sofern die Richtlinien nicht übertragbar sind. Im Sinne einer Weiterentwicklung der von diSessa und Sherin (2000) beschriebenen metarepräsentationalen Kompetenz sollte zudem das Wissen über Abbildungen in den Blick genommen werden, also zum Beispiel das Wissen über Funktionen und Grenzen bestimmter Abbildungsarten oder auch das eigenständige Entwickeln neuer Abbildungsarten durch Lernende.

Eine gestärkte Selbstwirksamkeitserwartung kann dazu führen, dass sich Personen eher auf das Arbeiten mit komplexen Abbildungen einlassen und mehr Mühe in die Bearbeitung investieren als Personen mit geringerer Selbstwirksamkeitserwartung. Wichtig ist, bei Lehrenden in den Naturwissenschaften in Schule und Universität ein Verständnis darüber zu erreichen, dass fachspezifische, komplexe Abbildungen für Lernende nicht unbedingt selbsterklärend sind und das Arbeiten mit solchen Abbildungen explizit unterrichtet werden sollte, was die entsprechende Selbstwirksamkeitserwartung von Lernenden stärkt.

## Danksagung

Wir danken Safa Dadkhah Aseman und Christina Kelting für ihre Hilfe bei Interrating und Dateneingabe, allen Dozierenden, die uns freundlicherweise die Datenerhebung in ihren Kursen ermöglicht haben, sowie allen Studierenden, die bereit waren, an der Studie teilzunehmen.

## Anmerkungen

- <sup>1</sup> Die Aufgabenstellung ist stark an die Aufgabenstellung der Unteraufgaben der entsprechenden Abituraufgabe angelehnt. Der erste Satz ist fast wortgleich übernommen (Brixius et al., 2012, S. 2008-13).
- <sup>2</sup> Das eingesetzte Material ist urheberrechtlich geschützt und darf daher hier nicht reproduziert werden. Eine grafische Demonstration des Materials, die die Variation zwischen den Gruppen zeigt, ist in Magnus et al. (2020) enthalten.
- <sup>3</sup> Der Lesbarkeitsindex LIX zeigt keine bedeutsamen Unterschiede der Textschwierigkeit zwischen den Experimentalgruppen an. Für Aufgabe B wurden für den LIX nach Lenhard und Lenhard (2014–2017) für Gruppe *keine Redundanz*, *Redundanz* und *nur Text* die Werte 57.6, 57.7 und 57.7 ermittelt; damit wird die Komplexität als hoch eingestuft. Für Aufgabe A ergaben sich in gleicher Reihenfolge die LIX-Werte 57.0, 54.7 und 53.2. Den beiden letztgenannten Werten wird von Lenhard und Lenhard (2014–2017) noch mittlere Komplexität zugeschrieben, einem LIX-Wert von 57 wiederum hohe Komplexität. Diese geringfügig unterschiedlichen LIX-Werte zwischen den Experimentalgruppen gehen darauf zurück, dass in die LIX-Formel die Satzanzahl eingeht und diese in Aufgabe A nicht konstant gehalten wurde: In der Gruppe *keine Redundanz* ist ein langer Satz enthalten, während an gleicher Stelle im Text der anderen beiden Gruppen bei identischer Zeichenzahl zwei kurze Sätze enthalten sind. Dieser Unterschied erscheint insofern nicht als Beeinträchtigung der Validität, als der Satz in der Gruppe *keine Redundanz* keine lösungsrelevante Information enthält und zudem kritisch zu hinterfragen ist, ob ein längerer Satz mit weniger Information einen Text tatsächlich in jedem Fall schwieriger werden lässt als zwei kürzere Sätze mit mehr Information.
- <sup>4</sup> Die Änderung des  $R^2$  durch Aufnahme des KFT in das Vorhersagemodell für das Kriterium Score\_Abb war nach Korrektur mittels Benjamini-Hochberg-Prozedur nicht mehr statistisch signifikant.

## Literatur

- Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.
- Ardura, D. & Pérez-Bitrián, A. (2019). Motivational pathways towards academic achievement in physics & chemistry: a comparison between students who opt out and those who persist. *Chemistry Education Research and Practice*, 20, 618–632. <http://dx.doi.org/10.1039/c9rp00073a>
- Arsenault, D. J., Smith, L. D. & Beauchamp, E. A. (2006). Visual inscriptions in the scientific hierarchy: Mapping the “treasures of science”. *Science Communication*, 27, 376–428. <http://dx.doi.org/10.1177/1075547005285030>
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215. <http://dx.doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (2001). Social cognitive theory: an agentic perspective. *Annual Review of Psychology*, 52, 1–26. <http://dx.doi.org/10.1146/annurev.psych.52.1.1>
- Bong, M. (1997). Generality of academic self-efficacy judgments: Evidence of hierarchical relations. *Journal of Educational Psychology*, 89, 696–709. <http://dx.doi.org/10.1037/0022-0663.89.4.696>
- Brixius, R., Jannan, M. & Kunze, H. (2012). *Abitur; Prüfungsaufgaben mit Lösungen 2013; NRW; Biologie LK (6. Aufl.)*. Hallbergmoos: Stark.
- Cromley, J. G., Perez, T. C., Fitzhugh, S. L., Newcombe, N. S., Wills, T. W. & Tanaka, J. C. (2013). Improving students' diagram comprehension with classroom instruction. *The Journal of Experimental Education*, 81, 511–537. <http://dx.doi.org/10.1080/00220973.2012.745465>
- Cromley, J. G., Snyder-Hogan, L. E. & Luciw-Dubas, U. A. (2010). Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology*, 35, 59–74. <http://dx.doi.org/10.1016/j.cedpsych.2009.10.002>
- diSessa, A. A. & Sherin, B. L. (2000). Meta-representation: an introduction. *The Journal of Mathematical Behavior*, 19, 385–398. [http://dx.doi.org/10.1016/S0732-3123\(01\)00051-7](http://dx.doi.org/10.1016/S0732-3123(01)00051-7)
- Eilam, B. & Poyas, Y. (2008). Learning with multiple representations: Extending multimedia learning beyond the lab. *Learning and Instruction*, 18, 368–378. <http://dx.doi.org/10.1016/j.learninstruc.2007.07.003>
- Heller, K. A. & Perleth, C. (2000). *KFT 4-12+R – Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Beltz Test GmbH, Hogrefe.
- Heßler, M. (2006). Einleitung. Annäherungen an Wissenschaftsbilder. In M. Heßler (Hrsg.), *Konstruierte Sichtbarkeiten. Wissenschafts- und Technikbilder seit der Frühen Neuzeit* (S. 11–37). München: Fink.
- Jerusalem, M. & Satow, L. (1999). Schulbezogene Selbstwirksamkeitserwartung. In R. Schwarzer & M. Jerusalem (Hrsg.), *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen* (S. 15–16). Berlin: Freie Universität Berlin.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. <http://dx.doi.org/10.2307/2529310>
- Lemke, J. L. (1998). *Teaching all the languages of science: words, symbols, images, and actions*. Paper presented at the International Conference on Ideas for a Scientific Culture, Barcelona, Spanien.
- Lenhard, W. & Lenhard, A. (2014–2017). *Berechnung des Lesbarkeitsindex LIX nach Björnson*. Zugriff am 7. 7. 2021 unter <http://www.psychometrica.de/lix.html>. Bibergau: Psychometrica. <http://dx.doi.org/10.13140/RG.2.1.1512.3447>
- Lindner, M. A., Ihme, J. M., Saß, S. & Köller, O. (2018). How representational pictures enhance students' performance and test-taking pleasure in low-stakes assessment. *European Journal of Psychological Assessment*, 34, 376–385. <http://dx.doi.org/10.1027/1015-5759/a000351>
- Lowe, R. (2000). *Visual literacy and learning in science*. Zugriff am 27. 7. 2020 unter <http://eric.ed.gov/?id=ED463945>
- Magnus, L., Schütte, K. & Schwanewedel, J. (2020). Challenges solving science tasks with text–picture combinations persist beyond secondary school. *Journal of Research on Educational Effectiveness*, 13, 759–783. <http://dx.doi.org/10.1080/19345747.2020.1750744>

- NGSS Lead States (2013). *Next generation science standards: for states, by states*. Washington, DC: The National Academies Press.
- Nitz, S., Prechtel, H. & Nerdel, C. (2014). Survey of classroom use of representations: development, field test and multilevel analysis. *Learning Environments Research*, 17, 401–422. <http://dx.doi.org/10.1007/s10984-014-9166-x>
- Novick, L. R., Stull, A. T. & Catley, K. M. (2012). Reading phylogenetic trees: The effects of tree orientation and text processing on comprehension. *BioScience*, 62, 757–764. <http://dx.doi.org/10.1525/bio.2012.62.8.8>
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66, 543–578. <http://dx.doi.org/10.3102/00346543066004543>
- Rosebrock, C. & Nix, D. (2010). *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung* (3. Aufl.). Baltmannsweiler: Schneider.
- Roth, W.-M. & Pozzer-Ardenghi, L. (2013). Pictures in biology education. In D. F. Treagust & C.-Y. Tsui (Eds.), *Multiple representations in biological education* (Vol. 7, pp. 39–53). Dordrecht: Springer. [http://dx.doi.org/10.1007/978-94-007-4192-8\\_3](http://dx.doi.org/10.1007/978-94-007-4192-8_3)
- Scherb, C. A. & Nitz, S. (2020). Attitudes of biology teachers towards learner-generated external visual representations. *Research in Science Education*, 50, 2533–2558. <http://dx.doi.org/10.1007/s11165-018-9792-x>
- Schönborn, K. J. & Anderson, T. R. (2006). The importance of visual literacy in the education of biochemists. *Biochemistry and Molecular Biology Education*, 34, 94–102. <http://dx.doi.org/10.1002/bmb.2006.49403402094>
- Schütte, K., Frenzel, A. C., Asseburg, R. & Pekrun, R. (2007). Schülermerkmale, naturwissenschaftliche Kompetenz und Berufserwartung. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 125–146). Münster: Waxmann.
- Schwarzer, R. & Jerusalem, M. (2002). Das Konzept Selbstwirksamkeit. In M. Jerusalem & D. Hopf (Hrsg.), *Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen* (S. 28–53), Zeitschrift für Pädagogik, 44. Beiheft. Weinheim: Beltz.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16. 12. 2004*. München: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife. Beschluss der Kultusministerkonferenz vom 18. 06. 2020*. Zugriff am 19. 3. 2021 unter [http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2020/2020\\_06\\_18-Bildungsstandards\\_AHR\\_Biologie.pdf](http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-Bildungsstandards_AHR_Biologie.pdf)
- Seufert, T. (2003). Kohärenzbildung beim Wissenserwerb mit multiplen Repräsentationen. In K. Sachs-Hombach (Hrsg.), *Was ist Bildkompetenz? Studien zur Bildwissenschaft* (S. 117–129). Wiesbaden: Deutscher Universitätsverlag. [http://dx.doi.org/10.1007/978-3-663-11814-5\\_8](http://dx.doi.org/10.1007/978-3-663-11814-5_8)
- Trafton, J. G., Marshall, S., Mintz, F. & Trickett, S. B. (2002). Extracting explicit and implicit information from complex visualizations. In M. Hegarty, B. Meyer & N. H. Narayanan (Eds.), *Diagrammatic representation and inference. Diagrams 2002. Lecture notes in computer science* (Vol. 2317, pp. 206–220). Berlin: Springer. [http://dx.doi.org/10.1007/3-540-46037-3\\_22](http://dx.doi.org/10.1007/3-540-46037-3_22)
- von Kotzebue, L., Gerstl, M. & Nerdel, C. (2015). Common mistakes in the construction of diagrams in biological contexts. *Research in Science Education*, 45, 193–213. <http://dx.doi.org/10.1007/s11165-014-9419-9>
- Ziepprecht, K. (2016). *Strategien und Kompetenzen von Lernenden beim Erschließen von biologischen Informationen aus unterschiedlichen Repräsentationen*. Berlin: Logos.

**Lara Magnus**

Universität Hamburg  
 Von-Melle-Park 8  
 20146 Hamburg  
 E-Mail: [lara.magnus@uni-hamburg.de](mailto:lara.magnus@uni-hamburg.de)

**Kerstin Schütte**

IPN – Leibniz-Institut für die Pädagogik  
 der Naturwissenschaften und Mathematik  
 24098 Kiel  
 E-Mail: [schuette@leibniz-ipn.de](mailto:schuette@leibniz-ipn.de)  
 ORCID iD: <http://orcid.org/0000-0003-2472-8074>

**Julia Schwanewedel**

Universität Hamburg  
 Von-Melle-Park 8  
 20146 Hamburg  
 E-Mail: [julia.schwanewedel@uni-hamburg.de](mailto:julia.schwanewedel@uni-hamburg.de)  
 ORCID iD: <http://orcid.org/0000-0001-5896-3238>