

Schriftenreihe **F**achdidaktische **F**orschung

Nr. 10, August 2014

Moritz Krell & Sarah M. Tieben

**Goal-Framing in der
Kompetenzdiagnostik: Effekte einer
theoriegeleiteten Variation der
Testinstruktion bei einem Test zum
biologischen Fachwissen**

Key words:

Goal-Framing, Instruktionseffekt, Durchführungsobjektivität, fachbezogene
Kompetenzdiagnostik

Herausgeber der Schriftenreihe Fachdidaktische Forschung:

Peter Frei, Katrin Hauenschild, Irene Pieper, Barbara Schmidt-Thieme
Forum Fachdidaktische Forschung
Universität Hildesheim

Redaktion: Julia Rößling

ISSN 2193-5912

Verfügbar unter: <http://www.uni-hildesheim.de/sff>

Moritz Krell & Sarah M. Tieben

Goal-Framing in der Kompetenzdiagnostik: Effekte einer theoriegeleiteten Variation der Testinstruktion bei einem Test zum biologischen Fachwissen

Abstract

In dieser Studie wurde untersucht, inwieweit sich unterschiedliche Testinstruktionen auf die Testleistung von Schülerinnen und Schülern auswirken. Basierend auf der Framing-Theorie wurden drei unterschiedliche Instruktionen umgesetzt: neutrales, negatives und positives Goal-Framing. Bei negativem Goal-Framing wird ein mögliches Versagen betont, während positives Goal-Framing einen möglichen Testerfolg hervorhebt. Die neutrale Instruktion diente als Kontrolle. Schülerinnen und Schüler ($N = 312$; 7.-10. Jahrgangsstufe) bearbeiteten 20 Aufgaben zum biologischen Fachwissen und wurden entweder neutral, negativ oder positiv geframet instruiert. Die Befunde zeigen keinen signifikanten Haupteffekt des Framings auf die Testleistung der Schülerinnen und Schüler. Es ergeben sich aber signifikante Interaktionseffekte zwischen Framing und Jahrgangsstufe sowie zwischen Framing und Geschlecht. Schülerinnen und Schüler des zehnten Jahrgangs zeigen bei positivem Framing eine signifikant bessere Testleistung als bei negativem und neutralem. Für die unteren Jahrgangsstufen ergeben sich keine Unterschiede in der Testleistung zwischen den Instruktionen. Schülerinnen erzielten bei neutralem Framing bessere Testleistungen als bei negativem und positivem, für Schüler ist das Muster genau umgekehrt. Die Befunde werden im Zusammenhang des Forschungsstandes über Goal-Framing diskutiert und mögliche Moderatorvariablen (Aufgabenschwierigkeit, akademisches Selbstkonzept) für die Erklärung herangezogen. Es werden Implikationen für die fachbezogene Kompetenzdiagnostik und weitere Studien über Goal-Framing in diesem Bereich diskutiert.

This study investigated to what extent test instruction influences students' test performance. Adapted from the framing theory, three different test instructions were applied: neutral, negative and positive goal framing. Within negative goal framing, a possible failure in the test is emphasized, whereas positive goal framing emphasizes on a possible pass of the test. The neutral instruction served as a control. Students ($N = 312$; 7.-10. grade) answered 20 tasks testing biological conceptual knowledge while being instructed either neutrally, negatively or positively framed. Findings show no significant main effect of framing on the students' test performance. However, there are significant interaction effects between framing and grade level as well as framing and sex. 10th graders performed significantly better with positive framing than with negative or neutral framing. For the other graders, no differences in test performance between the different instructions occurred. Female students showed better test performances with neutral framing than with negative or positive framing, whereas for male students, the effect seems to be reversed. The findings are discussed with regard to the state of research about goal framing. Possible moderator variables (task difficulty, academic self-concept) are consulted in order to explain the findings. Implications for competence assessment as well as for possible further studies within this field are considered.

1 Einleitung

Eine zentrale Voraussetzung für die Gültigkeit von Testergebnissen in der Kompetenzdiagnostik ist die Einhaltung der klassischen Gütekriterien Objektivität, Reliabilität und Validität (Bortz & Döring, 2006; Moosbrugger & Kelava, 2008; Schmidt-Atzert & Amelang, 2012). Bezogen auf die hier vorgestellte Studie ist die Durchführungsobjektivität von besonderer Relevanz. Ein Test kann dann als objektiv betrachtet werden, wenn die von einem Probanden gezeigte Testleistung nur von

dessen Merkmalsausprägung abhängt (Moosbrugger & Kelava, 2008). Für die Gewährleistung einer objektiven Testdurchführung sollten daher mögliche Effekte der Testinstruktion auf das Antwortverhalten der Probanden minimiert werden (Sarris & Reiss, 2005; Schmidt-Atzert & Amelang, 2012). Insbesondere im Kontext Kompetenzdiagnostik sollte ein möglicher Einfluss der Testinstruktion, beispielsweise auf die kognitive Belastung von Probanden, geprüft werden. Dies hilft einzuschätzen, inwieweit die gezeigte Testperformanz einen angemessenen Indikator für die anvisierte Kompetenz darstellt oder die Performanz durch externe, nicht intendierte Stimuli beeinflusst ist. Eine Untersuchung möglicher Effekte der Testinstruktion auf die Performanz der Probanden kann daher wichtige Anhaltspunkte für die Optimierung von Testinstruktionen liefern. Eine solche Untersuchung wurde etwa begleitend zu PISA 2000 durchgeführt (Artelt et al., 2001, S. 8). Schülerinnen und Schülern wurde ein Test mit vier unterschiedlichen Instruktionen präsentiert: als PISA-Test mit der standardisierten PISA-Instruktion, als Leistungsuntersuchung mit einer anschließenden individuellen Rückmeldung, als Test mit einer Benotung durch die eigene Lehrkraft oder als Test, bei dem besonders gut abschneidende Probanden mit einer bestimmten Geldsumme belohnt werden. Die Ergebnisse dieser Begleitstudie ergaben keine bedeutsamen Unterschiede in der Testleistung der Schülerinnen und Schüler zwischen den vier Instruktionsvarianten (Artelt et al., 2001).

Im Gegensatz hierzu heben Locke und Latham (2002, 2013) mit Fokus auf berufsbezogene Kontexte hervor, dass die Betonung unterschiedlicher Ziele im Vorfeld einer Testsituation einen signifikanten Einfluss auf die Testleistung von Probanden ausüben kann. Für den hochschulpädagogischen Bereich zeigen Roney und Lehman (2008) in ähnlicher Weise, dass unterschiedliche Instruktionen sich auf die Testleistung von Studierenden auswirken können. Krell, Reinisch und Krüger (in press) diskutieren im Kontext naturwissenschaftsdidaktischer Diagnostik, dass die Testleistung von Schülerinnen und Schülern auch durch einen Bezug des Tests zu unterschiedlichen naturwissenschaftlichen Disziplinen (bzw. Schulfächern) beeinflusst werden kann (vgl. Gobert et al., 2011).

Insgesamt deuten diese Befunde darauf hin, dass sich die Testinstruktion auf die Testleistung von Schülerinnen und Schülern auswirken *kann*. Vor diesem Hintergrund können Studien mit einem theoriegeleiteten Vergleich des Effekts unterschiedlicher Testinstruktionen dazu beitragen, die Wirkungsfaktoren von Instruktionseffekten zu identifizieren und somit in zukünftigen Studien kontrollierbar zu machen. Für eine theoriegeleitete Konstruktion unterschiedlicher Testinstruktionen und eine darauf aufbauende Analyse möglicher Effekte auf die Testleistung von Schülerinnen und Schülern können Befunde über Framing-Effekte genutzt werden (z. B. Levin, Gaeth, Schreiber, & Lauriola, 2002; Levin, Schneider, & Gaeth, 1998; Roney & Lehman, 2008). Die Framing-Theorie (engl. *frame*: Rahmen) versucht zu erklären, inwieweit die Rahmung einer Problemstellung einen bedeutsamen Einfluss auf das Verhalten von Probanden hat. Kennzeichnendes Charakteristikum ist hierbei, dass der Inhalt der dargebotenen Informationen jeweils identisch ist, weshalb Framing-Effekte auch als „formulation effects“ (Dunegan, 1996) oder „Darstellungseffekte“ (Martin, 2012) bezeichnet werden.

In der vorliegenden Explorationsstudie bearbeiteten Schülerinnen und Schüler ($N = 312$) Multiple Choice-Aufgaben der *American Association for the Advancement of*

Science (AAAS, 1993) zum biologischen Fachwissen. Hierbei wurden basierend auf der Framing-Theorie drei unterschiedliche Testinstruktionen umgesetzt (neutrales, negatives und positives Goal-Framing nach Roney & Lehman, 2008) und deren Effekt auf die Testleistung der Schülerinnen und Schüler analysiert (*between subject-design*). Es wird somit empirisch geprüft, inwieweit die von Roney und Lehman (2008) im Kontext hochschulpädagogischer Diagnostik berichteten Befunde in einem anderen Kontext (fachbezogene Kompetenzdiagnostik im schulischen Bereich) replizierbar und somit möglicherweise verallgemeinerbar sind (Schmidt, 2009).

2 Theoretische Grundlagen

2.1 Klassische Gütekriterien: Durchführungsobjektivität

Von besonderer Bedeutung für die vorliegende Studie ist die Durchführungsobjektivität. Moosbrugger und Kelava (2008) definieren als Bedingung für eine objektive Testdurchführung: „Die ‚Testleistung‘ soll [...] nur von der Merkmalsausprägung des Individuums abhängen“ (S. 9). Sarris und Reiss (2005) heben im Gegensatz hierzu hervor, dass Testinstruktionen die Funktion haben, „etwas Bestimmtes im Sinne der psychologischen Fragestellung zu leisten“, und die Probanden dazu bewegen sollen, „eine bestimmte Haltung während des Versuchs einzunehmen“ (S. 95). Aus dieser Perspektive kann ein Testergebnis nicht nur von der Merkmalsausprägung des Individuums abhängen, sondern eben auch von der Testinstruktion. Entsprechend argumentieren Schmidt-Atzert und Amelang (2012), dass eine hohe Durchführungsobjektivität angenommen werden darf, „wenn alle Bedingungen *festgelegt* [Herv. v. Verf.] sind, die sich erfahrungsgemäß auf das Testverhalten auswirken können“ (S. 134).

Diese kurze Zusammenfassung zeigt bereits, dass für die Gewährleistung einer objektiven Testdurchführung entweder sichergestellt werden muss, dass die gewählte Testinstruktion *keinen* Effekt auf das Antwortverhalten der Probanden hat (Moosbrugger & Kelava, 2008), oder aber geprüft werden muss, inwieweit die genutzte Instruktion den *gewünschten* Effekt hat (Sarris & Reiss, 2005; Schmidt-Atzert & Amelang, 2012). Vor diesem Hintergrund ermöglicht die Forschung über Framing-Effekte eine theoriegeleitete Variation von Testinstruktionen und eine Analyse der Auswirkung ebendieser auf die Testleistung von Schülerinnen und Schülern.

2.2 Framing-Effekte

Die Forschung über Framing-Effekte geht maßgeblich auf die Arbeit Kahnemans und Tverskys zurück (vgl. Kahneman, 2011). In ihrer Studie zum *Asian disease problem* wurde Probanden ein fiktives Szenario präsentiert (Tversky & Kahneman, 1981, S. 453):

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people.

Ein Teil der Probanden sollte zwischen den folgenden Handlungsoptionen A und B wählen:

If Program A is adopted, 200 people will be saved. [...] If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

Der andere Teil der Probanden sollte zwischen den Optionen C und D wählen:

If Program C is adopted 400 people will die. [...] If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

Die beiden Entscheidungssituationen sind inhaltlich identisch, unterscheiden sich aber hinsichtlich ihrer Darstellungsweise. Die Optionen A und B sind durch die Möglichkeit, Menschen zu retten, eingerahmt, während bei den Optionen C und D der Rahmen durch mögliche Verluste gebildet wird. In der ersten Gruppe wählten etwa 70 % der Probanden Option A, in der zweiten Gruppe entschieden sich hingegen nur etwa 20 % für die äquivalente Option C (Tversky & Kahneman, 1981).

Anknüpfend an das *Asian disease problem* sind diverse Studien publiziert, in denen ähnliche Effekte der Informationsdarstellung auf das Verhalten von Probanden dokumentiert sind (vgl. Krishnamurthy, Carter, & Blair, 2001; Levin et al., 1998; Martin, 2012; Stocké, 2001). Da hierbei stets nur die Darstellungsweise (bzw. der Rahmen), nicht aber der eigentliche Inhalt eines Problems variiert wird, werden solche Befunde als Framing-Effekte bezeichnet. Der Framing-Begriff bezieht sich in diesem Zusammenhang somit auf eine besondere Form der Instruktionsvariation, bei der logisch äquivalente aber begrifflich unterschiedlich formulierte Instruktionen umgesetzt werden (oftmals negativ und positiv konnotiert; siehe das oben beschriebene *Asian disease problem*). In der in der Einleitung skizzierten PISA-Begleitstudie (Artelt et al., 2001) wurden daher keine Framing-, sondern allgemein Instruktionseffekte untersucht.

Eine Kernannahme der Framing-Theorie ist, dass die mentale Repräsentation von Informationen stets mit Selektions-, Reorganisations- und Konstruktionsprozessen verbunden ist und diese Prozesse durch die Darstellungsweise einer Information signifikant beeinflusst werden können (Borah, 2011; Martin, 2012). Brandstätter und Mücke (2009) heben hervor, dass Framing-Effekte entweder als Fehler betrachtet werden können oder aber als Ergebnis kognitiver Anker, die für eine ökonomische Informationsverarbeitung unerlässlich sind. Aus letzterer Perspektive kann argumentiert werden, dass Framing-Effekte notwendiger Bestandteil jeder Handlung sind. Die Frage ist dann nicht *ob*, sondern *wie* sich unterschiedliche Frames auf die Handlung von Probanden auswirken (Greeno, 2009). Aufgrund unterschiedlicher methodologischer Ansätze und fachlicher Einbettungen fehlt allerdings eine einheitliche theoretische Erklärung für Framing-Effekte (Borah, 2011; Stocké, 2001).

2.3 Arten des Framings

In Abhängigkeit der vorhandenen Problemstellung wird zwischen Risky Choice-, Attribute- und Goal-Framing unterschieden (Levin et al., 1998, 2002; Martin, 2012). Beim Risky Choice-Framing werden Risiken entweder als erwartete Gewinne oder Verluste dargestellt und die Wirkung dieser Darstellung auf das Risikowahlverhalten von Probanden untersucht (z. B. das oben beschriebene *Asian disease problem*). In

Übereinstimmung mit Tversky und Kahneman (1981) zeigen die meisten Studien eine vergleichsweise hohe Risikobereitschaft bei einem Fokus auf der Vermeidung von Verlusten (Levin et al., 1998).

Attribute-Framing bezieht sich auf die von Probanden vorgenommene Bewertung von Objekten, wenn deren Eigenschaften unterschiedlich dargestellt werden. Zum Beispiel berichtet Dunegan (1996) eine Studie, in der Probanden die Aufgabe hatten, in einem fiktiven Szenario als Mitglieder einer Disziplinarkommission über die Strafzahlung eines Unternehmens zu entscheiden, welches irreführende Werbung genutzt haben soll. Den Probanden wurde das Szenario entweder positiv (z. B.: die Werbung war mit einer Wahrscheinlichkeit von 20 % versehentlich irreführend, etwa 30 % aller Unternehmen nutzen ähnliche Werbestrategien) oder negativ geframet beschrieben (z. B.: die Werbung war mit einer Wahrscheinlichkeit von 80 % absichtlich irreführend, etwa 70 % aller Unternehmen vermeiden solche Werbestrategien). Die von den Probanden vorgeschlagene Strafzahlung war bei negativem Framing annähernd doppelt so hoch wie bei positivem (Dunegan, 1996). In ähnlicher Weise zeigen die meisten Studien, dass eine positive Beschreibung zu einer positiveren Bewertung eines Objekts führt als eine inhaltlich äquivalente Beschreibung mit negativen Begriffen (Krishnamurthy et al., 2001; Levin et al., 1998).

Schließlich beschreibt Goal-Framing die Wirkung der Betonung negativer Konsequenzen bei dem Ausbleiben eines bestimmten Verhaltens (negativer Frame) im Vergleich zur Betonung positiver Konsequenzen beim Zeigen dieses Verhaltens (positiver Frame; Levin et al., 1998, 2002). Charakteristisch für Goal-Framing ist, dass sowohl der negative als auch der positive Frame das jeweilige Verhalten verstärken sollen (Levin et al., 1998). Hierbei sind bei der Konstruktion negativer und positiver Frames unterschiedliche sprachliche Varianten denkbar, die sich aus den möglichen Kombinationen der Beschreibung des jeweiligen Verhaltens X^+ sowie dessen Konsequenz ergeben (Tabelle 1).

Tabelle 1

Mögliche Varianten zur Konstruktion von Goal-Frames

	negativ	positiv
Verhalten	nicht Verhalten X^+	Verhalten X^+
	gegensätzliches Verhalten X^-	nicht gegensätzliches Verhalten X^-
Konsequenz	keine positive Konsequenz	positive Konsequenz
	negative Konsequenz	keine negative Konsequenz

Roney und Lehman (2008) berichten zwei Studien, in denen Probanden vor der Bearbeitung eines Tests mit Buchstabenrätseln (*anagrams*) durch unterschiedliche Frames instruiert wurden. In der ersten Studie wurden die Probanden mit einem negativen (*Try not to miss answering more than 3 of these 15 anagrams.*), positiven (*Try to answer 12 of these 15 anagrams correctly.*) oder neutralen Frame (*Do your best.*) instruiert. In der zweiten Studie wurden diese Frames durch die Betonung von Misserfolg (negativer Frame) beziehungsweise Erfolg (positiver Frame) ergänzt

(Tabelle 2). In beiden Studien erzielten die Probanden bei negativem Framing die schlechteste Testleistung (Roney & Lehman, 2008).

Im nächsten Absatz werden mögliche Erklärungen für Framing-Effekte beschrieben. Entsprechend dem Schwerpunkt der vorliegenden Studie wird hierbei ein Fokus auf Effekte beim Goal-Framing gelegt.

2.4 Goal-Framing: Erklärungsansätze

Anknüpfend an Kahneman und Tversky (1979, 1984) werden Framing-Effekte oft mit Bezug zur *Prospect Theory* erklärt (Levin et al., 1998, 2002; Martin, 2012; Stocké, 2001). Dieser theoretische Ansatz diente ursprünglich der Erklärung von Studien über Risky Choice-Framing und postuliert eine allgemeine Risikoaversion in Bezug auf erwartete Gewinne und eine Risikofreude in Bezug auf erwartete Verluste (Kahneman, 2011). Hieran anknüpfend argumentieren Levin et al. (1998), dass sich negative Informationen grundsätzlich stärker auf das Verhalten von Probanden auswirken als positive (*negativity bias*). Dementsprechend deuten viele Studien darauf hin, dass ein negativer Goal-Frame zu einer stärkeren Verhaltensausrprägung bei Probanden führt als ein positiver (Krishnamurthy et al., 2001; Levin et al., 1998). Dieser Erklärungsansatz wird allerdings gleichzeitig als zu vereinfachend kritisiert, weil beispielsweise die Wirkung individueller Referenzpunkte nicht berücksichtigt ist (Martin, 2012; Stocké, 2001). Darüber hinaus kann mit Rückgriff auf einen *negativity bias* nicht erklärt werden, *wodurch* sich unterschiedliche Frames auf das Verhalten von Probanden auswirken. Schließlich führte in den oben beschriebenen Studien von Roney und Lehman (2008) der negative Frame zu einer vergleichsweise schlechten Testleistung der Probanden. Roney und Lehman (2008) diskutieren ihre Ergebnisse daher mit Bezug zu selbstregulatorischen Prozessen, die bei der Betonung eines möglichen Versagens (negativer Frame; Tabelle 2) kognitiv hemmend wirken. Im Gegensatz hierzu argumentieren Locke und Latham (2002) mit der motivationalen Wirkung anspruchsvoller Ziele und betonen, dass „difficult goals consistently led to higher performance than urging people to do their best“ (S. 706). Dies lässt eine im Vergleich zum neutralen Framing gute Testleistung bei negativem Framing erwarten und steht demnach im Widerspruch zu den Befunden von Roney und Lehman (2008).

Bereits diese kurze Skizze einiger Erklärungsansätze deutet an, worauf auch Krishnamurthy et al. (2001) hinweisen: „results from goal framing are not wholly consistent“ (S. 384; vgl. auch Levin et al., 2002). Dies kann daran liegen, dass die Wirkung von Goal-Frames unter anderem von der Schwierigkeit der zu lösenden Aufgabe (Locke & Latham, 2002), der Art des Ziels (z. B. intrinsisch vs. extrinsisch: Vansteenkiste et al., 2004; selbstgesetzt vs. zugewiesen: Locke & Latham, 2013) oder der sprachlichen Formulierung des Frames (Levin et al., 1998; vgl. Tabelle 1) moderiert werden kann. Bei einer Abhängigkeit von Framing-Effekten von selbstregulatorischen Prozessen (Roney & Lehman, 2008) müssten außerdem weitere Personenmerkmale (z. B. Selbstkonzept, metakognitive Strategien) für die Wirkung von Frames bedeutsam sein. Beispielsweise kann sich ein geringes (akademisches) Selbstkonzept negativ auf die Ausdauer bei der Aufgabenbearbeitung auswirken, wenn die Aufgaben oder das gesetzte Ziel als zu schwer wahrgenommen werden (Lazarus, 2007). Pauschal gültige Aussagen über die Wirkung von Frames wären damit nicht möglich, Framing-Effekte müssten vielmehr spezifisch für unterschiedliche Personengruppen beschrieben

werden (Levin et al., 2002). Darüber hinaus werden zur Untersuchung von Framing-Effekten meist fiktive Szenarien konstruiert und das vermeintliche Verhalten von Probanden erfragt (z. B. Levin et al., 2002; Meyerowitz & Chaiken, 1987), während einige Studien das tatsächliche Verhalten von Probanden bei unterschiedlichen Frames analysieren (z. B. Roney & Lehman, 2008). Framing-Effekte konnten bereits in berufsbezogenen Kontexten (Locke & Latham, 2002, 2013) sowie in der hochschulpädagogischen Diagnostik (Roney & Lehman, 2008) beschrieben werden. Mögliche Auswirkungen von Frames zum Beispiel auf die kognitive Belastung und damit auf die Testperformanz von Probanden (Roney & Lehman, 2008) sind zum Beispiel dann von Relevanz, wenn die jeweilige Testperformanz als Grundlage für individuelle Bewertungen genutzt wird (z. B. bei Personalentscheidungen oder bei Bewertungen in Ausbildungsphasen). Darüber hinaus können personenspezifische Effekte von Frames, etwa moderiert durch das akademische Selbstkonzept oder metakognitive Strategien, zu ungewünschten Differenzeffekten führen. Vor diesem Hintergrund sind weitere Studien notwendig, um eine angemessene theoretische Beschreibung von Goal-Framing-Effekten und der Wirkung wichtiger Moderatorvariablen zu erreichen (Borah, 2011; Levin et al., 1998, 2002).

3 Ziele und Hypothesen der Arbeit

Ein Ziel der vorliegenden Explorationsstudie ist es, die Befundlage über die Wirkung von Goal-Frames zu ergänzen und damit einen Beitrag zur Weiterentwicklung der Framing-Theorie zu leisten. Insbesondere wird geprüft, inwieweit die von Roney und Lehman (2008) berichteten Befunde im Kontext fachbezogener Kompetenzdiagnostik replizierbar und somit möglicherweise verallgemeinerbar sind (Schmidt, 2009). Sollten signifikante Performanzunterschiede der Probanden zwischen den Framing-Manipulationen auftreten, muss hinterfragt werden, inwieweit die im Einzelfall gezeigte Testperformanz ein sinnvoller Indikator für die anvisierte Kompetenz ist. Der Kompetenzbegriff beschreibt das Potenzial von Personen, in definierten Situationen Fähigkeiten und Fertigkeiten umzusetzen und so eine bestimmte Performanz zeigen zu können (Weinert, 2001). Entsprechend könnten Befunde über Framing-Effekte dazu beitragen, Testinstruktionen so zu konstruieren, dass nicht intendierte Stimuli (und damit beispielsweise zusätzliche kognitive Belastungen) vermieden werden und Probanden ihr Potenzial möglichst optimal umsetzen können. Darüber hinaus können Befunde über Framing-Effekte dazu beitragen, ungewünschte Differenzeffekte zu vermeiden und damit eine objektive Testdurchführung zu gewährleisten (Schmidt-Atzert & Amelang, 2012).

Die leitende Fragestellung dieser Studie ist:

Inwieweit wirkt sich die Testinstruktion (neutrales, negatives und positives Goal-Framing; Roney & Lehman, 2008) auf das Antwortverhalten von Schülerinnen und Schülern bei Testaufgaben zum biologischen Fachwissen aus?

Aus den Befunden Roney und Lehman (2008) kann die folgende Hypothese abgeleitet werden:

H.1a: Schülerinnen und Schüler erreichen bei negativem Framing einen geringeren Testscore als bei positivem und neutralem Framing, da sich die Betonung eines möglichen Versagens kognitiv hemmend auswirkt.

Nach Locke und Latham (2002) kann die konkurrierende Alternativhypothese *H.1b* formuliert werden:

H.1b: Schülerinnen und Schüler erreichen bei positivem Framing den höchsten Testscore, gefolgt von dem bei negativem und neutralem Framing, da ein positiver Zusammenhang zwischen der Höhe des gesetzten Ziels und der Motivation zur Aufgabenbearbeitung besteht.

4 Methoden

4.1 Testinstrument

Es wurden 20 Multiple Choice-Aufgaben zum Kompetenzbereich Fachwissen (AAAS, 1993) zu den biologischen Inhaltsbereichen Zellbiologie, Genetik, Evolution und Ökologie übersetzt und als Testinstrument in dieser Studie verwendet. Um Reihenfolgeeffekte zu vermeiden, wurden nach einem *repeated treatment design* 20 verschiedene Testheftvarianten konstruiert, in denen die Aufgaben jeweils in unterschiedlichen Reihenfolgen enthalten sind (Frey, Hartig, & Rupp, 2009).

H.1a vermutet eine kognitiv hemmende Wirkung des negativen Goal-Frames. Um eine mögliche Auswirkung unterschiedlicher Frames auf die kognitive Belastung von Schülerinnen und Schülern erfassen zu können, wurden die Probanden am Ende jedes Testhefts gebeten, ihren Cognitive Load (CL) während der Testbearbeitung anzugeben (Selbsteinschätzung). Hierzu wurden fünf Items mit fünfstufiger Ratingskala nach Nehring, Nowak, Upmeyer zu Belzen und Tiemann (2012) eingesetzt. Als Kontrollvariablen wurden außerdem die Schulnoten in den Fächern Biologie, Chemie und Physik als Indikatoren für die Leistung der Schülerinnen und Schüler in den naturwissenschaftlichen Fächern erfragt.

4.2 Studiendesign und Stichprobe

Die Probanden wurden nach Roney und Lehman (2008) vor der Testbearbeitung mit einem neutralen, negativen oder positiven Frame instruiert (*between subject-design*; Tabelle 2). Der negative Frame betont die negative Konsequenz eines vergleichsweise niedrigen Testscores, während der positive Frame die positive Konsequenz eines hohen Testscores hervorhebt (Tabelle 1). Die Schülerinnen und Schüler mit neutralem Framing dienten als Kontrollgruppe. Mit Hilfe eines schriftlichen Instruktionsleitfadens wurde der jeweilige Frame den Probanden zu Beginn der Testung verbal vorgetragen. Zusätzlich wurde er an die Tafel geschrieben und stand auf dem Deckblatt jedes Testhefts.

Tabelle 2

Realisierte Frames bei Roney und Lehman (2008) und in der vorliegenden Studie

	negativer Frame	positiver Frame	neutraler Frame
Roney & Lehman	You will fail to reach the goal if you do not answer at least 12 of these 15 anagrams correctly.	You will successfully reach the goal if you don't miss answering more than 3 of these 15 anagrams.	Do your best.
diese Studie	Du bist im Test durchgefallen, wenn du mehr als 30 % der Aufgaben falsch beantwortest!	Du hast den Test bestanden, wenn du 70 % der Aufgaben oder mehr richtig beantwortest!	Gib dein Bestes und beantworte die Aufgaben so, wie du es persönlich am sinnvollsten findest!

Schülerinnen und Schüler der Jahrgangsstufen sieben bis zehn haben freiwillig an dieser Studie teilgenommen. Tabelle 3 zeigt die Stichprobe ($N = 312$) aufgeteilt nach Frame, Jahrgangsstufe und Geschlecht.

Tabelle 3

Stichprobe

Jahrgangsstufe	Geschlecht	neutral (0)	negativ (-)	positiv (+)	Summe
7	weiblich	13	13	13	39
	männlich	13	13	13	39
8	weiblich	13	13	13	39
	männlich	13	13	13	39
9	weiblich	13	13	13	39
	männlich	13	13	13	39
10	weiblich	13	13	13	39
	männlich	13	13	13	39
Summe		104	104	104	312

Anmerkung: Die ursprüngliche Stichprobe ($N = 485$) war nicht gleich über die Zellen verteilt. Um in der (Ko-) Varianzanalyse eine stabile Parameterschätzung zu erreichen, wurden daher zufällig Probanden aus größer besetzten Zellen entfernt (Eschweiler, Evanschitzky, & Woisetschläger, 2007).

4.3 Datenauswertung

Für die weitere Auswertung wurden die AAAS-Aufgaben mit dem einparametrischen Rasch-Modell mit der Software ConQuest 3 analysiert (vgl. Wu, Adams, Wilson, & Haldane, 2007) und *weighted likelihood estimates* (WLEs; Warm, 1989) als Kennwerte der Personenfähigkeiten geschätzt. Als Indikator des CL der Probanden wurde ein Mittelwert über die fünf Items berechnet.

Zur Hypothesenprüfung wurde eine Kovarianzanalyse (ANCOVA) durchgeführt (Backhaus, Erichson, Plinke, & Weiber, 2011; Eschweiler et al., 2007; Field, 2009). Hierbei diente der WLE als abhängige und die Art des Framings (neutral, negativ, positiv), das Geschlecht der Probanden sowie deren Jahrgangsstufe als unabhängige

Variablen. Der angegebene CL wurde als Kovariate in das Modell aufgenommen. Durch den Einbezug einer Kovariaten kann die (unerklärte) Fehlervarianz einer Varianzanalyse durch die vorgeschaltete Regressionsanalyse reduziert und der Einfluss der Kovariate auf die abhängige Variable geschätzt werden (Field, 2009). Um eine stabile Parameterschätzung zu erreichen, sollten die Daten eine gruppenweise Normalverteilung sowie Varianzhomogenität aufweisen. Eine robuste Parameterschätzung ist trotz einer Verletzung dieser Annahmen möglich, wenn eine Gleichbesetzung der Zellen vorliegt (Backhaus et al., 2011; Eschweiler et al., 2007). Darüber hinaus sollte ein signifikanter Zusammenhang ($p < .05$) zwischen der abhängigen Variablen und der Kovariaten bestehen und die Kovariate sollte von den erklärenden Variablen unabhängig sein. Schließlich muss geprüft werden, ob die Regressionssteigungen in der vorgeschalteten Regression homogen ausfallen. Hierzu empfiehlt Field (2009) die Prüfung einer möglichen Interaktion zwischen Kovariate und den unabhängigen Variablen.

5 Ergebnisse

5.1 Deskriptive Statistik

Tabelle 4 zeigt deskriptive Statistiken für die Gesamtstichprobe sowie, separat für die drei Framing-Gruppen, die vier Jahrgangsstufen und die beiden Geschlechter. Es liegen keine signifikanten Unterschiede in M_{AAAS} zwischen den Framing-Gruppen vor. Es ist ersichtlich, dass die mittlere Lösungswahrscheinlichkeit mit den Jahrgangsstufen tendenziell steigt, die Aufgaben insgesamt aber verhältnismäßig schwer waren. In einer einfaktoriellen Varianzanalyse (ANOVA) ergibt sich insgesamt ein signifikanter Unterschied in M_{AAAS} zwischen den Jahrgangsstufen ($p < .001$; $\eta_p^2 = 0.12$; mittlerer bis großer Effekt; Fritz, Morris, & Richler, 2012), der im post-hoc Vergleich (Bonferroni) auf die im Vergleich zu den drei anderen Jahrgangsstufen signifikant höhere Lösungswahrscheinlichkeit der Zehntklässler zurückzuführen ist ($p \leq .002$). Außerdem zeigen die Jungen eine signifikant höhere Lösungswahrscheinlichkeit als die Mädchen ($p = .008$; $d = 0.30$; kleiner Effekt).

In einer ANOVA ergeben sich keine signifikanten Unterschiede in M_{CL} zwischen den Framing-Gruppen ($p = .835$) und den Jahrgangsstufen ($p = .413$). Der Unterschied in der kognitiven Belastung zwischen den Geschlechtern fällt marginal signifikant aus ($p = 0.900$; $d = 0.19$; kleiner Effekt).

Tabelle 4

Deskriptive Statistiken der Rohdaten

	Frame			Jahrgangsstufe				Geschlecht		Σ
	0	-	+	7	8	9	10	w	m	
M_{AAAS}	0.42	0.42	0.42	0.35	0.42	0.40	0.51	0.39	0.44	0.42
SD_{AAAS}	0.15	0.16	0.18	0.13	0.17	0.16	0.16	0.15	0.17	0.16
M_{CL}	2.92	2.97	2.91	2.99	2.99	2.92	2.82	3.00	2.86	2.93
SD_{CL}	0.79	0.72	0.66	0.66	0.80	0.67	0.76	0.67	0.77	0.73

Anmerkung: Mittelwert (M) und Standardabweichung (SD) der Lösungswahrscheinlichkeit der Testaufgaben (AAAS) sowie des angegebenen CL.

Es bestehen zwischen den Framing-Gruppen keine signifikanten Unterschiede bezogen auf die durchschnittliche Jahrgangsstufe (Tabelle 3), die Geschlechterverteilung (Tabelle 3) sowie die erfragten Schulnoten ($.172 \leq p \leq .404$; ANOVA).

5.2 Schätzung der Personenfähigkeiten

Die gewichteten und ungewichteten *means of squared standardised residuals* (MNSQs) liegen in einem Bereich von 0.9 bis 1.1 sowie 0.9 bis 1.3 und weisen auf einen ausreichenden *fit* zwischen den Daten und dem einparametrischen Rasch-Modell hin (Wright & Linacre, 1994). Die Reliabilität der Schätzung der Personenfähigkeiten ist ebenfalls ausreichend ($rel_{WLE} = 0.6$). Schließlich zeigt eine sehr hohe Pearson-Korrelation zwischen den WLEs und der mittleren Lösungswahrscheinlichkeit der Testaufgaben ($r = .99$), dass die geschätzten Personenfähigkeiten gute Indikatoren für das Antwortverhalten der Probanden sind.

5.3 Prüfung der Modellannahmen der AN(C)OVA

Es besteht eine signifikante Pearson-Korrelation zwischen M_{CL} und WLE ($r = -.16$; $p = .01$). In separaten ANOVAs zeigen sich keine signifikanten Unterschiede in M_{CL} zwischen den Ausprägungen der Faktoren Frame, Jahrgangsstufe sowie Geschlecht (Unabhängigkeit von Kovariate und unabhängigen Variablen; Field, 2009). Es ergeben sich ebenfalls keine signifikanten Interaktionseffekte zwischen der Kovariaten und den unabhängigen Variablen (Homogenität der Regressionssteigungen; Field, 2009). Die Annahme der zellenweisen Normalverteilung ist nicht erfüllt. In zwei Fällen weicht die Verteilung signifikant von der erwarteten Normalverteilung ab (Shapiro-Wilk-Test; $p \leq .024$).¹ Eine Varianzhomogenität kann angenommen werden (Levene-Test; $p = .274$).

5.4 Hypothesenprüfung

In der ANCOVA ergeben sich signifikante Haupteffekte der Jahrgangsstufe und des Geschlechts sowie signifikante Interaktionseffekte zwischen Frame und Jahrgangsstufe sowie Frame und Geschlecht (Tabelle 5, oben). Die Variable M_{CL} trägt

¹ Für die größere Ursprungsstichprobe (vgl. Tabelle 3) ergaben sich acht nicht normalverteilte Zellen.

in der Analyse nicht signifikant zur Erklärung der WLEs bei ($p = .149$). Aus diesem Grund wurde anschließend eine ANOVA ohne die Kovariate M_{CL} gerechnet. In dieser Analyse ergeben sich vergleichbare Modellparameter (Tabelle 5, unten).

Tabelle 5

Ergebnisse von ANCOVA (oben) und ANOVA (unten)

ANCOVA	SS	df	MS	F	p	η_p^2
M_{CL}	1.225	1	1.225	2.099	.149	.007
Frame	0.015	2	0.007	0.013	.987	.000
Jgst.	24.387	3	8.129	13.921	.000	.127
Geschl.	4.377	1	4.377	7.495	.007	.025
Frame*Jgst.	9.073	6	1.512	2.590	.018	.051
Frame*Geschl.	4.301	2	2.150	3.682	.026	.025
Jgst.*Geschl.	0.753	3	0.251	0.430	.732	.004
Fehler	167.589	287	0.584	---	---	---
Gesamtwert	217.388	311	---	---	---	---

$R^2 = .229$

ANOVA	SS	df	MS	F	p	η_p^2
Frame	0.020	2	0.010	0.017	.983	.000
Jgst.	25.543	3	8.514	14.526	.000	.131
Geschl.	4.908	1	4.908	8.374	.004	.028
Frame*Jgst.	10.133	6	1.689	2.881	.010	.057
Frame*Geschl.	4.971	2	2.485	4.240	.015	.029
Jgst.*Geschl.	0.689	3	0.230	0.392	.759	.004
Fehler	168.815	288	0.586	---	---	---
Gesamtwert	217.388	311	---	---	---	---

$R^2 = .223$

Anmerkung: Jgst.: Jahrgangsstufe; Geschl.: Geschlecht

Aufgrund der vorliegenden Interaktionseffekte sollten die Haupteffekte nicht interpretiert werden (Janssen & Laatz, 2005). Abbildung 1 zeigt das Interaktionsdiagramm Frame*Jahrgangsstufe. Die Analyse einfacher Effekte (Field, 2009) ergibt innerhalb der einzelnen Jahrgangsstufen lediglich für die Schülerinnen und Schüler der zehnten Klasse einen signifikanten Effekt des Framings ($p = .003$). Dieser ist auf signifikant höhere WLEs bei positivem Framing ($M_{WLE} = 0.87$; $SD_{WLE} = 0.76$) als bei negativem ($M_{WLE} = 0.29$; $SD_{WLE} = 0.76$) und neutralem ($M_{WLE} = 0.19$; $SD_{WLE} = 0.66$) zurückzuführen. Zwischen den Jahrgangsstufen treten sowohl bei neutralem ($p = .032$) als auch bei positivem ($p < .001$) Framing signifikante (einfache) Effekte auf. Bei neutralem Framing zeigen die Schülerinnen und Schüler der siebten Jahrgangsstufe ($M_{WLE} = -0.39$; $SD_{WLE} = 0.53$) signifikant niedrigere WLEs als die übrigen Schülerinnen und Schüler, während bei positivem Framing die Schülerinnen und Schüler des zehnten Jahrgangs ($M_{WLE} = 0.87$; $SD_{WLE} = 0.76$) signifikant höhere WLEs aufweisen als der Rest der Stichprobe.

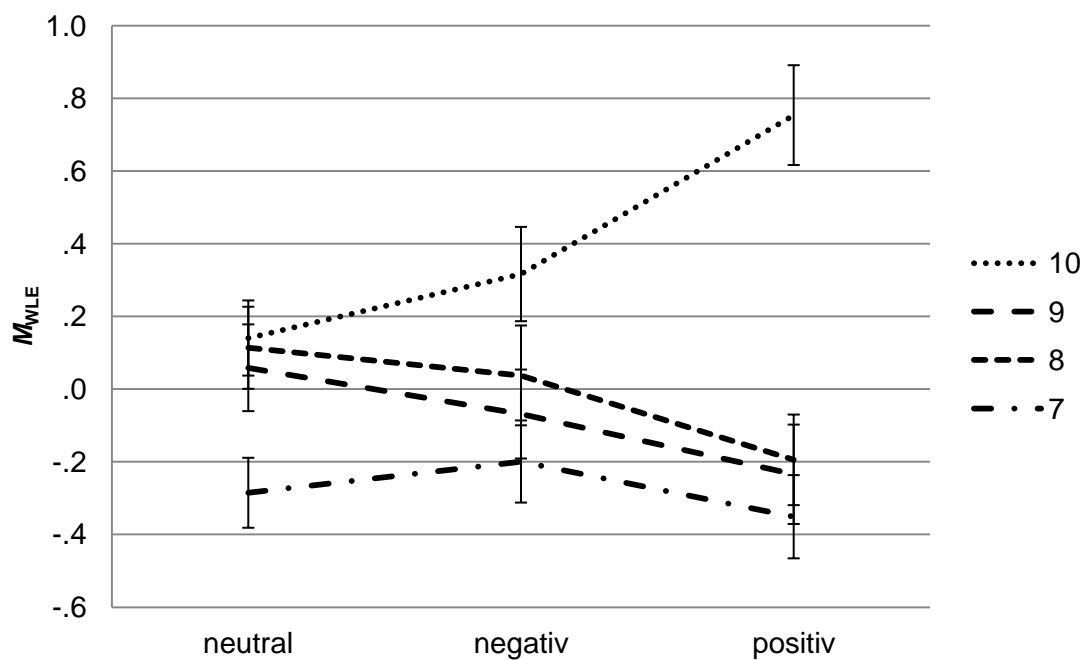


Abbildung 1. Interaktionsdiagramm Frame*Jahrgangsstufe (Fehlerbalken: ± 1 SE)

Das Interaktionsdiagramm Frame*Geschlecht zeigt annähernd entgegengesetzte Muster für Jungen und Mädchen (Abbildung 2). Während die Jungen bei negativem ($M_{WLE} = 0.23$; $SD_{WLE} = 0.77$) und positivem Framing ($M_{WLE} = 0.21$; $SD_{WLE} = 1.00$) besser abschneiden als bei neutralem Framing ($M_{WLE} = -0.01$; $SD_{WLE} = 0.78$), haben die Mädchen bei neutralem Framing ($M_{WLE} = 0.05$; $SD_{WLE} = 0.68$) die beste Testleistung erzielt. Innerhalb der Geschlechter ergeben sich keine signifikanten Effekte des Framings auf die WLEs. Die Jungen erzielen allerdings bei negativem ($p = .002$) und positivem Framing ($p = .011$) signifikant höhere WLEs als die Mädchen (einfache Effekte).

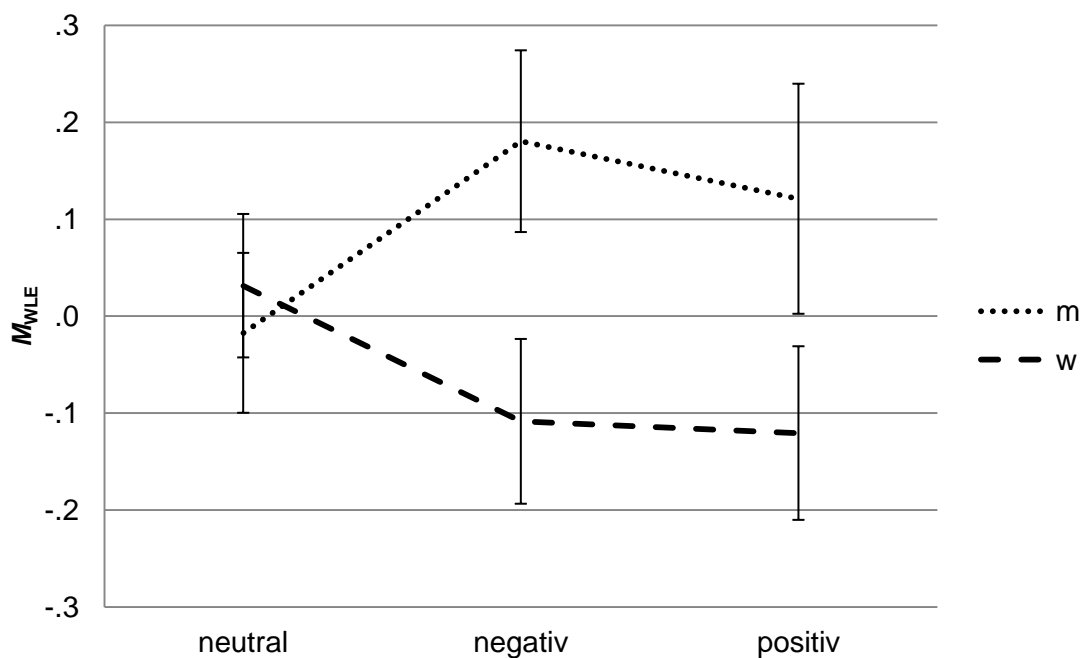


Abbildung 2. Interaktionsdiagramm Frame*Geschlecht (Fehlerbalken: +/- 1 SE)

6 Zusammenfassung und Diskussion

In der vorliegenden Explorationsstudie ist kein signifikanter Haupteffekt des Goal-Framings auf die im einparametrischen Rasch-Modell geschätzten WLEs nachweisbar (Tabelle 5). Für die Gesamtstichprobe müssen daher sowohl *H.1a* als auch *H.1b* verworfen werden. Diese Befunde entsprechen denen der oben beschriebenen PISA-Begleitstudie, in der ebenfalls kein bedeutsamer Effekt der Instruktion auf die Testleistung von Schülerinnen und Schülern nachweisbar war (Artelt et al., 2001).

Es konnten allerdings signifikante Interaktionseffekte zwischen Frame und Jahrgangsstufe sowie zwischen Frame und Geschlecht beschrieben werden. Für die Schülerinnen und Schüler der zehnten Jahrgangsstufe ergaben sich signifikant höhere WLEs bei positivem als bei negativem und neutralem Framing (Abbildung 1). Außerdem zeigten die Jungen bei positivem und negativem Framing signifikant höhere WLEs als die Mädchen (Abbildung 2).

Locke und Latham (2002, 2013) heben hervor, dass Goal-Framing-Effekte von der Aufgabenschwierigkeit moderiert werden können und mit zunehmender Aufgabenschwierigkeit in der Regel geringer ausfallen. Die Aufgaben in der vorliegenden Studie stellten sich als eher schwer heraus (Tabelle 4). Es kann daher angenommen werden, dass die Aufgaben insbesondere für die Schülerinnen und Schüler der Jahrgangsstufen sieben bis neun zu schwer waren, um Framing-Effekte nachweisen zu können. Schmidt (2013) argumentiert im Rahmen einer methodisch vergleichbar angelegten Studie ($N = 176$; Jahrgangsstufen sieben bis zehn) und dem Einsatz von Multiple Choice-Aufgaben zur Modellkompetenz im Biologieunterricht (vgl. Terzer, 2013), dass auch bei sehr leichten Aufgaben keine Framing-Effekte auftreten. Insgesamt spricht dies dafür, dass Goal-Framing-Effekte in der fachbezogenen

Kompetenzdiagnostik bei einem Einsatz sehr schwerer (vorliegende Studie) oder sehr leichter Aufgaben (Schmidt, 2013) nicht nachweisbar sind.

Werden daher nur die Schülerinnen und Schüler der zehnten Jahrgangsstufe betrachtet, ergeben sich bei positivem Framing signifikant höhere WLEs als bei negativem und neutralem, während kein signifikanter Unterschied zwischen neutralem und negativem Framing nachweisbar ist (Abbildung 1). Auf dieser Grundlage muss *H.1a* verworfen werden. Anders als von Roney und Lehman (2008) diskutiert, hat negatives Framing offenbar nicht zu einer kognitiven Hemmung der Schülerinnen und Schüler geführt. Dafür spricht ergänzend, dass der CL in den drei Framing-Manipulationen ähnlich hoch ausfällt (Tabelle 4) und in der ANCOVA nicht signifikant zur Erklärung der WLEs beiträgt (Tabelle 5). Entsprechend kann *H.1b* für die Schülerinnen und Schüler der zehnten Jahrgangsstufe bestätigt werden. Es wird angenommen, dass sich die Nennung eines anspruchsvollen Ziels (positiver Frame) positiv auf die Motivation zur Testbearbeitung bei Schülerinnen und Schülern auswirkt und daher zu einer besseren Testleistung führen könnte (Locke & Latham, 2002).

In der vorliegenden Studie konnte ein signifikanter Interaktionseffekt *Frame*Geschlecht* nachgewiesen werden. Insbesondere haben die Jungen bei negativem und positivem Framing eine signifikant bessere Testleistung erzielt als die Mädchen (Abbildung 2). Levin et al. (2002) konnten keine geschlechterspezifische Wirkung von Goal-Frames beschreiben. Allerdings heben unterschiedliche Autoren die moderierende Rolle individueller Personenmerkmale für die Wirkung von Framing-Effekten hervor (z. B. Levin et al., 2002; Roney & Lehman, 2008; Tversky & Kahneman, 1981). Etwa diskutieren Roney und Lehman (2008), dass defensive Pessimisten (vgl. Norem & Cantor, 1986) durch das Herabsetzen der eigenen Erwartungen vor einer Leistungssituation weniger anfällig gegenüber unterschiedlichen Frames sein könnten. Levin et al. (2002) haben geprüft, inwieweit die Faktoren des *Big Five*-Persönlichkeitsinventars eine personenspezifische Wirkung von Goal-Frames erklären können. Es ergab sich allerdings kein systematischer Zusammenhang zwischen der Ausprägung der Skalen und der Wirkung unterschiedlicher Goal-Frames. In der vorliegenden Studie haben die Jungen in den beiden Frames, die eine Testsituation betonen, besser abgeschnitten als bei neutralem Framing, während für die Mädchen das umgekehrte Muster vorliegt. Eine mögliche Erklärung für diese geschlechterspezifischen Befunde ist das akademische Selbstkonzept. Dieses ist in Bezug auf die naturwissenschaftlichen Fächer bei Jungen in der Regel positiver ausgeprägt als bei Mädchen (Schilling, Sparfeldt, & Rost, 2006). Allerdings sind diese Befunde im Zusammenhang mit dem biologiebezogenen Selbstkonzept nicht eindeutig (vgl. Jansen, Schroeders, & Stanat, 2013). Das akademische Selbstkonzept kann sich auf die Leistungsfähigkeit von Schülerinnen und Schülern in akademischen Testsituationen auswirken. Lazarus (2007) beschreibt, dass sich Schülerinnen und Schüler mit einem geringen akademischen Selbstkonzept in Testsituationen zum Beispiel weniger ausdauernd mit den gestellten Aufgaben beschäftigen und stärker durch aufgabenirrelevante Gedanken abgelenkt sind. Demgegenüber schneiden Schülerinnen und Schüler mit einem ausgeprägten akademischen Selbstkonzept in der Regel vergleichsweise gut in Testsituationen ab. Sollten die in der vorliegenden Studie beteiligten Mädchen von Geschlechterunterschieden im biologischen Fachwissen überzeugt sein, könnten außerdem Mechanismen des *Stereotype Threat* (vgl. Inzlicht

& Schmader, 2012) zu der vergleichsweise geringen Testleistung bei negativem und positivem Framing geführt haben.

Neben der vergleichsweise hohen Aufgabenschwierigkeit ist eine weitere methodische Einschränkung der vorliegenden Studie die vorgenommene Zuweisung des Ziels, das in den Frames beschrieben wird (d. i. das Bestehen des Tests; Tabelle 2). Zwar können sich sowohl selbstgesetzte als auch zugewiesene Ziele auf die Performanz von Probanden auswirken (Locke & Latham, 2013), Locke und Latham (2002) weisen allerdings darauf hin, dass der Effekt zugewiesener Ziele auf die Performanz einer Person durch deren Selbstwirksamkeitserwartung sowie selbstgesetzte Ziele moderiert wird. Darüber hinaus ergaben sich für die Schülerinnen und Schüler durch das Bestehen oder das Durchfallen im eingesetzten Test keine weiteren Konsequenzen. Es kann daher vermutet werden, dass für viele Schülerinnen und Schüler das Ziel eingeschränkte Relevanz und lediglich extrinsischen Charakter hatte. Vansteenkiste et al. (2004) zeigen, dass Framing-Effekte bei einem Bezug zu intrinsischen Zielen (z. B. Selbstverwirklichung) stärker ausfallen können als bei extrinsischen Zielen.

Zusammenfassend kann bei einem Fokus auf die Teilstichprobe der Schülerinnen und Schüler der zehnten Jahrgangsstufe die nach Locke und Latham (2002) formulierte Hypothese *H.1b* im Kontext fachbezogener Kompetenzdiagnostik und einem Einsatz von Testaufgaben zum biologischen Fachwissen bestätigt werden. Aufgrund der spezifischen Stichprobe und des auf biologisches Fachwissen beschränkten Erhebungsinstrumentes müssen diese Befunde in weiteren Untersuchungen repliziert werden, bevor sie als allgemeingültig für die fachbezogene Kompetenzdiagnostik betrachtet werden dürfen.

7 Fazit und Ausblick

Aus den Befunden der vorliegenden Studie können, ungeachtet der oben diskutierten Einschränkungen, Empfehlungen für eine objektive Testdurchführung in der fachbezogenen Kompetenzdiagnostik abgeleitet werden. Bei der Entwicklung eines Instruktionsleitfadens sollte berücksichtigt werden, dass Schülerinnen und Schüler bei neutralem Framing (und entsprechender Aufgabenschwierigkeit) eine geringere Testleistung erzielen könnten als bei Instruktionen mit negativem oder positivem Framing. Im Zusammenhang fachbezogener Kompetenzdiagnostik scheint die Testperformanz bei positivem Goal-Framing somit der beste Indikator für die Kompetenz von Schülerinnen und Schülern zu sein. Darüber hinaus deuten die vorliegenden Befunde darauf hin, dass Instruktionen mit neutralem Framing zu vernachlässigbaren Geschlechterunterschieden führen. Aus dieser Perspektive können neutrale Frames für die Durchführung von Studien empfohlen werden, um ungewünschte (geschlechtsbezogene) Differenzeffekte zu vermeiden. Hier scheint somit eine Abwägung zwischen einer möglicherweise negativen Wirkung auf die Testleistung und einer geschlechtergerechten Testdurchführung notwendig zu sein.

Inwieweit die Ergebnisse der vorliegenden Studie auf Instruktionen im Zusammenhang mit unterrichtsbezogener Diagnostik (d. h. Tests, Klassenarbeiten etc.) übertragbar sind, müssten weitere Studien zeigen. Insbesondere die oben bereits diskutierte Zuschreibung eines Ziels mit vermeintlich extrinsischem Charakter (Locke & Latham,

2002, 2013; Vansteenkiste et al., 2004) könnte eine Übertragbarkeit der Befunde auf Unterrichtssituationen einschränken.

Aus den Befunden der vorliegenden Explorationsstudie ergeben sich außerdem zwei Hypothesen über Goal-Framing in der fachbezogenen Kompetenzdiagnostik, die in weiteren Forschungsarbeiten geprüft werden können.

H.1: In der fachbezogenen Kompetenzdiagnostik steigern positive Goal-Frames bei angemessener (d. h. mittlerer) Aufgabenschwierigkeit aufgrund der expliziten Nennung eines anspruchsvollen Ziels die Motivation zur Aufgabenbearbeitung und führen daher zu einer besseren Testleistung als neutrale oder negative Goal-Frames (vgl. Locke & Latham, 2013).

H.2: Personen mit einem geringen akademischen Selbstkonzept erzielen bei negativem und positivem Goal-Framing aufgrund der Betonung einer akademischen Testsituation schlechtere Testleistungen als Personen mit einem stärker ausgeprägten Selbstkonzept (vgl. Lazarus, 2007; Schilling et al., 2006).

Eine Prüfung dieser Hypothesen in anknüpfenden Studien kann einen Beitrag zur Weiterentwicklung der Framing-Theorie (Borah, 2011) leisten und damit Erkenntnisse liefern, die für die objektive Durchführung von Tests in der fachbezogenen Kompetenzdiagnostik genutzt werden können.

8 Danksagung

Die Autoren bedanken sich bei den anonymen Gutachtern für die hilfreichen Anmerkungen zur Optimierung des ursprünglich eingereichten Manuskripts sowie bei Charlotte Schmidt, die im Rahmen ihrer Masterarbeit bei der Konzeption und Durchführung dieser Studie beteiligt war.

Literatur

- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy, Project 2061*. New York, NY: Oxford U.P.
- Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., ...Weiß, M. (Eds.). (2001). *PISA 2000: Zusammenfassung ausgewählter Befunde*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2011). *Multivariate Analysemethoden*. Berlin: Springer.
- Borah, P. (2011). Conceptual issues in framing theory: A systematic examination of a decade's literature. *Journal of Communication, 61*, 246–263.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Brandstätter, E., & Mücke, R. (2009). Interpreting test results. *Personality and Individual Differences, 46*, 183–186.
- Dunegan, K. (1996). Fines, frames, and images: Examining formulation effects on punishment decisions. *Organizational Behavior and Human Decision Processes, 68*, 58–67.

- Eschweiler, M., Evanschitzky, H., & Woisetschläger, D. (2007). Ein Leitfadens zur Anwendung varianzanalytisch ausgerichteter Laborexperimente. *Wirtschaftswissenschaftliches Studium*, 36, 546–554.
- Field, A. (2009). *Discovering statistics using SPSS*. Los Angeles, CA: Sage.
- Frey, A., Hartig, J., & Rupp, A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educational Measurement: Issues and Practice*, 28, 39–53.
- Fritz, C., Morris, P., & Richler, J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141, 2–18.
- Gobert, J., O'Dwyer, L., Horwitz, P., Buckley, B., Levy, S., & Wilensky, U. (2011). Examining the relationship between students' understanding of the nature of models and conceptual learning in biology, physics, and chemistry. *International Journal of Science Education*, 33, 653–684.
- Greeno, J. (2009). A theory bite on contextualizing, framing, and positioning: A companion to Son and Goldstone. *Cognition and Instruction*, 27, 269–275.
- Inzlicht, M., & Schmader, T. (2012). *Stereotype threat: Theory, process, and application*. New York, NY: Oxford University Press.
- Jansen, M., Schroeders, U., & Stanat, P. (2013). Motivationale Schülermerkmale in Mathematik und den Naturwissenschaften. In H. Pant, P. Stanat, U. Schroeders, A. Roppelt, & T. Siegle (Hrsg.), *IQB-Ländervergleich 2012* (S. 347–365). Münster: Waxmann.
- Janssen, J., & Laatz, W. (2005). *Statistische Datenanalyse mit SPSS für Windows*. Berlin: Springer.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39, 341–350.
- Krell, M., Reinisch, B., & Krüger, D. (in press). Analyzing students' understanding of models and modeling referring to the disciplines biology, chemistry, and physics. *Research in Science Education*. doi: 10.1007/s11165-014-9427-9
- Krishnamurthy, P., Carter, P., & Blair, E. (2001). Attribute framing and goal framing effects in health decisions. *Organizational Behavior and Human Decision Processes*, 85, 382–399.
- Lazarus, S. (2007). *Motivation und Leistung* (Doctoral dissertation). Ludwig-Maximilians-Universität München. Retrieved from http://edoc.ub.uni-muenchen.de/7206/1/Lazarus_Simone.pdf
- Levin, I., Schneider, S., & Gaeth, G. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76, 149–188.
- Levin, I., Gaeth, G., Schreiber, J., & Lauriola, M. (2002). A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes*, 88, 411–429.
- Locke, E., & Latham, G. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57, 705–717.
- Locke, E., & Latham, G. (2013). Goal setting theory: The current state. In E. Locke & G. Latham (Eds.), *New developments in goal setting and task performance* (pp. 623–630). New York, NY: Routledge.

- Martin, A. (2012). *Fehlentscheidungen: Warum wir tun, was wir später bereuen*. Zusatzmaterial online. Darmstadt: WBG. Retrieved from http://download.wbg-wissenverbindet.de/B235490_Zusatzmaterial.pdf
- Maule, J., & Villejoubert, G. (2007). What lies beneath: Reframing framing effects. *Thinking & Reasoning*, 13, 25–44.
- Meyerowitz, B., & Chaiken, S. (1987). The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of Personality and Social Psychology*, 52, 500–510.
- Moosbrugger, H., & Kelava, A. (Eds.). (2008). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.
- Nehring, A., Nowak, K., Upmeyer zu Belzen, A., & Tiemann, R. (2012). Doing inquiry in chemistry and biology: The context's influence on the students' cognitive load. *La Chimica nella Scuola*, XXXIV, 253–258.
- Norem, J., & Cantor, N. (1986). Defensive pessimism: Harnessing anxiety as motivation. *Journal of Personality and Social Psychology*, 51, 1208–1217.
- Roney, C., & Lehman, D. (2008). Self-regulation in goal striving: Individual differences and situational moderators of the goal-framing/performance link. *Journal of Applied Social Psychology*, 38, 2691–2709.
- Sarris, V., & Reiss, S. (2005). *Kurzer Leitfaden der Experimentalpsychologie*. München: Pearson Studium.
- Schilling, S., Sparfeldt, J., & Rost, D. (2006). Facetten schulischen Selbstkonzepts: Welchen Unterschied macht das Geschlecht? *Zeitschrift für Pädagogische Psychologie*, 20, 9–18.
- Schmidt, C. (2013). *Goal-Framing-Effekte in der Kompetenzdiagnostik. Eine Untersuchung bei Tests zu dem Kompetenzbereich Erkenntnisgewinnung* (unpublished Master thesis). Freie Universität Berlin.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13, 90–100.
- Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik*. Berlin: Springer.
- Stocké, V. (2001). Framing ist nicht gleich Framing: Eine Typologie unterschiedlicher Framing-Effekte und Theorien zu ihrer Erklärung. In U. Druwe, V. Kunz, & T. Plümper (Eds.), *Jahrbuch für Handlungs- und Entscheidungstheorie* (pp. 75–105). Opladen: Leske und Budrich.
- Terzer, E. (2013). *Modellkompetenz im Kontext Biologieunterricht* (Doctoral dissertation). Humboldt-Universität zu Berlin. Retrieved from <http://edoc.hu-berlin.de/dissertationen/terzer-eva-2012-12-19/PDF/terzer.pdf>
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.
- Vansteenkiste, M., Simons, J., Lens, W., Soenens, B., Matos, L., & Lacante, M. (2004). Less is sometimes more: Goal content matters. *Journal of Educational Psychology*, 96, 755–764.
- Warm, T. (1989). Weighted likelihood estimation of ability in Item Response Theory. *Psychometrika*, 54, 427–450.
- Weinert, F. (2001). Concept of competence: A conceptual clarification. In D. Rychen & L. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Kirkland, WA: Hogrefe.
- Wright, B., & Linacre, J. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.

Wu, M., Adams, R., Wilson, M., & Haldane, S. (2007). *ACER ConQuest*. Camberwell, Vic: ACER Press.

Zu den Autoren

Dr. Moritz Krell
Freie Universität Berlin
Didaktik der Biologie
Schwendenerstraße 1
14195 Berlin
moritz.krell@fu-berlin.de

Sarah M. Tieben
Lehrerin im Vorbereitungsdienst
An der Weiche 5
49716 Meppen
stieben@zedat.fu-berlin.de

Zitation

Bitte zitieren Sie diesen Beitrag wie folgt:

Krell, Moritz & Tieben, Sarah M. (2014). Goal-Framing in der Kompetenzdiagnostik: Effekte einer theoriegeleiteten Variation der Testinstruktion bei einem Test zum biologischen Fachwissen. In: Schriftenreihe Fachdidaktische Forschung, Nr. 10, August 2014. Online verfügbar: http://www.uni-hildesheim.de/media/fff/Krell_08-2014.pdf.