

Analyse des Kompetenzaufbaus zur Variablenkontrollstrategie mithilfe von Sankey-Diagrammen

Tobias Winkens*, Simon Goertz*, Heidrun Heinke*

*RWTH Aachen University

winkens@physik.rwth-aachen.de, goertz@physik.rwth-aachen.de, heinke@physik.rwth-aachen.de

Kurzfassung

Die Variablenkontrollstrategie (VKS) ist eine wichtige experimentelle Kompetenz im naturwissenschaftlichen Unterricht, die Schülerinnen und Schüler (SuS) in verschiedenen Lernsettings jedoch häufig falsch anwenden. Im Rahmen einer Design-Based Research Studie im Prä-Post-Design wurde unter anderem evaluiert, wie die Kompetenzen der SuS hinsichtlich der VKS im Schulunterricht verbessert werden können. Mit einem Testinstrument zur Kompetenzerfassung der VKS wurden entsprechende Daten aufgenommen. Als Intervention kamen zwei verschiedene Lernzirkel mit je 5 Stationen (= Modulen) zum Einsatz. Der eine Lernzirkel besteht aus Modulen zu verschiedenen experimentellen Kompetenzen, wobei nur eines den Schwerpunkt auf die VKS legt, während beim anderen Lernzirkel alle Module die Schwerpunktkompetenz VKS behandeln. Eine detaillierte Auswertung der erhobenen Daten kann durch die Nutzung von Sankey-Diagrammen realisiert werden. Mit ihrer Hilfe kann das Antwortverhalten der SuS beim Testinstrument im Prä-Post-Vergleich insbesondere im Hinblick auf die beiden Lernzirkeltypen visualisiert und analysiert werden. Die resultierenden Erkenntnisse werden in diesem Beitrag vorgestellt.

1. Motivation

Konzeptionelle Probleme beim Verständnis der Variablenkontrollstrategie (VKS¹) sind keine Seltenheit. Trotz intensiver Beschäftigung mit der VKS zeigen SuS und zum Teil auch Erwachsene noch vielfältige Probleme (u.a. beim Ziehen von Schlussfolgerungen auf Basis konfundierter Experimente), die einen Bedarf an effektiveren Wegen zur Förderung der VKS bedingen (vgl. Barzel et al., 2012, S. 114; vgl. Chen und Klahr, 1999, S. 1098). Die Bedeutung der VKS als wichtige experimentelle Kompetenz für den gesamten experimentellen Prozess in den Naturwissenschaften begründet die Notwendigkeit, den SuS diese Kompetenz als Bestandteil einer naturwissenschaftlichen Grundbildung (Scientific Literacy) (vgl. Nerdel, 2017, S. 14–15) mithilfe effektiver Methoden zu vermitteln und durch gezielte Lerngelegenheiten deren (selbstständige) Aneignung zu fördern und zu fördern.

Der vorliegende Beitrag baut auf einer Studie von Goertz (vgl. Goertz, 2022) und den dort aufgenommenen Daten auf. Eine ergänzende Analyse der Daten des eingesetzten VKS-Testinstruments mithilfe von Sankey-Diagrammen bietet die Möglichkeit, ein tieferes Verständnis über den Kompetenzaufbau der VKS bei SuS zu erlangen. Aufbauend auf den gewonnenen Erkenntnissen können entsprechende Maßnahmen zur Förderung der VKS entwickelt und wiederum evaluiert werden, um die Vermittlung der VKS weiter zu verbessern.

2. Variablenkontrollstrategie

2.1. Experimentelle Kompetenz

Die VKS ist als „[d]ie experimentelle Erkenntnisermittlungsmethode“ (Hamann und Asshoff, 2015, S. 79) eine unverzichtbare Arbeitsweise in den Naturwissenschaften, um kausale Zusammenhänge bzw. Ursachen-Wirkungs-Beziehungen zu untersuchen und feststellen zu können (vgl. Schwichow und Nehring, 2018, S. 219). Die Anwendung der VKS ist dabei als Untersuchung eines prinzipiell möglichen Effekts der Messgröße (abhängige Variable) in Folge einer aktiven Manipulation eines ursächlichen Einflussfaktors (unabhängige Variable) charakterisiert (vgl. Schulz et al., 2012, S. 18). Im Gegensatz zu einem kontrollierten Experiment, das der VKS-Methode entspricht, ist ein konfundiertes Experiment dadurch charakterisiert, dass sich zwei Experimentalansätze in mehr als einer potenziell unabhängigen Variable unterscheiden (vgl. Schulz et al., 2012, S. 18).

Die Strategie der Variablenkontrolle muss dabei in Kombination mit einem Kontrollansatz eingesetzt werden. Denn nur bei Unterscheidung zweier Ansätze in lediglich einer Variablen oder einem Faktor, der sog. Testvariablen oder dem Testfaktor, können der Einfluss der Testvariablen beurteilt und Schlussfolgerungen zu dessen Bedeutung gezogen werden. Alle weiteren potenziellen Variablen müssen konstant gewählt werden (vgl. Ehmer, 2008, S. 11; vgl. Schulz et al., 2012, S. 18). Die VKS ist dabei bei der Planung von Experimenten relevant und notwendig,

¹ Teilweise wird in der Literatur auch die englische Abkürzung „CVS“ (für Control-of-Variables Strategy) verwendet.

aber auch für deren Durchführung und Auswertung (vgl. Ehmer, 2008, S. 12). Dieser Aspekt wird bei einer Unterteilung der VKS in die vier folgenden Teilfähigkeiten berücksichtigt (vgl. Schwichow und Nehring, 2018, S. 219; vgl. Schwichow et al., 2016, S. 217):

Dies sind die Fähigkeiten ...

- ... zur Planung kontrollierter Experimente (PL).
- ... zur gezielten Identifizierung kontrollierter Experimente aus einer Auswahl an kontrollierten und konfundierten Experimenten (ID).
- ... zur Interpretation der Befunde kontrollierter Experimente (IN).
- ... zum Verständnis der fehlenden Aussagekraft konfundierter Experimente (UN).

Diese Fähigkeiten sind Basis für die VKS-Test-Items, die in der Datenerhebung eingesetzt wurden und somit auch für die Auswertung mithilfe der Sankey-Diagramme. Beherrschen SuS diese Fähigkeiten und sind sie in der Lage auf ihrer Grundlage korrekte Schlussfolgerungen zu ziehen, besitzen sie nach Chen und Klahr „an essential skill in scientific reasoning“ (Chen und Klahr, 1999, S. 1098).

2.2. Schülervorstellungen und Schwierigkeiten

Beim Anwenden der VKS treten bei SuS unterschiedliche Schwierigkeiten auf, die von verschiedenen Autoren umfassend beschrieben wurden. SuS variieren häufig die Testvariablen nicht unter der Prämisse deren Zusammenhang zu untersuchen, sondern „den untersuchten Effekt herbeizuführen“ (vgl. Ehmer, 2008, S. 19). Das Phänomen oder den Vorgang zu erzeugen ist aus Sicht der SuS wichtiger, als den Einfluss einzelner Variablen zu untersuchen (vgl. Carey et al., 1989, S. 518). Bei dieser Vorstellung, dem sog. Ingenieurmodus (vgl. Hammann und Asshoff, 2015, S. 81), ergeben sich einige Probleme in experimentellen Situationen:

Eine Fehlvorstellung im Bereich der Planung von Experimenten liegt im fehlenden Kontrollansatz, da die SuS häufig dessen Notwendigkeit nicht erkennen und daher nur Testansätze machen, ohne dass ein Abgleich mit dem Kontrollansatz erfolgt (vgl. Carey et al., 1989, S. 518, vgl. Hammann et al., 2006, S. 292–293; vgl. Ehmer, 2008, S. 26). Gemäß diesem Prinzip können zwar Effekte etc. experimentell erzeugt und abgebildet werden, jedoch können keine kausalen Zusammenhänge im Sinne der VKS abgeleitet werden (vgl. Hammann et al., 2006, S. 292–293; vgl. Ehmer, 2008, S. 26). Ein weiteres Indiz für diese Arbeitsweise ist ein unsystematischer Umgang mit Variablen (vgl. Hammann und Asshoff, 2015, S. 81). Sie beachten und kontrollieren nicht, welche Variable die Testvariable ist und welche Variablen konstant gehalten werden müssen (vgl. Hammann et al., 2006, S. 293; vgl. Carey et al., 1989, S. 518; vgl. Ehmer, 2008, S. 26). Nach Carey et al. (1989) ist das unsystematische Vorgehen ein Indiz für ein „limited

understanding of the nature and purpose of experiments“ (S. 518). Das unsystematische Vorgehen kann dabei in drei verschiedene (kognitive) Strategien untergliedert werden. Bei der change all-Strategie (CA-Strategie) werden beim Experimentalansatz alle Variablen – sowohl die Testvariable als auch die weiteren unabhängigen Variablen – gegenüber dem Kontrollansatz verändert. (vgl. Tschirgi, 1980, S. 3; vgl. Ehmer, 2008, S. 26–27; vgl. Hammann et al., 2006, S. 294). Die HOTAT-Strategie (hold one thing at a time) ist eine zweite Variante des unsystematischen Umgangs mit Variablen. Bei der HOTAT-Strategie wird zwischen Kontroll- und Experimentalansatz nur die Testvariable konstant gehalten und alle anderen Variablen werden verändert (vgl. Tschirgi, 1980, S. 2; vgl. Ehmer, 2008, S. 27). Die dritte Strategie von SuS ist für die Planung und den Aufbau konfundierter Experimente relevant (vgl. Ehmer, 2008, S. 27–28; vgl. Schwichow, 2015, S. 5). Die Planung konfundierter Experimente ist darauf zurückzuführen, dass SuS nicht zwischen den Testvariablen unterscheiden oder „unterschiedliche Variablenausprägungen nicht wahrnehmen“ (Schwichow, 2015, S. 5). Außerdem folgern sie aus diesen konfundierten Settings nicht, dass aus solchen experimentellen Ansätzen keine verlässlichen Schlussfolgerungen zu Abhängigkeiten einzelner Variablen gezogen werden können (vgl. Hammann et al., 2006, S. 295–296).

3. Methodik der Analyse

3.1. Eingesetztes VKS-Testinstrument

Grundlage des eingesetzten VKS-Testinstruments war der von Schwichow entwickelte CVSI-Test (engl.: Control-of-Variables Strategy Inventory). Dieser ist ein paper-and-pencil Multiple-Choice Test, wobei bei jedem Item aus vier Antwortmöglichkeiten eine korrekte Antwort gewählt wird (vgl. Schwichow et al., 2016, S. 8). Die Items sind so konstruiert, dass sie dabei immer eine der vier Teilfähigkeiten der VKS (IN, ID, UN oder PL) erfassen. Hinsichtlich der Itemschwierigkeiten ist bezogen auf die VKS-Teilfähigkeiten nach Evaluation der Entwickler sowie nach weiteren Untersuchungen festzuhalten, dass Items der Typen ID und IN signifikant leichter sind als die Itemtypen PL und UN (vgl. Schwichow et al., 2016, S. 13–14; vgl. Brandenburger et al., 2020, S. 131–132; vgl. Brandenburger und Mikelskis-Seifert, 2019, S. 78–79). Für die hier durchgeführte Studie wurde ein Testinstrument aus insgesamt 11 Items (inhaltliche Kontexte: Elektromagnetismus, Wärmelehre und Leitung des Stroms) erstellt, die aus dem Pool des CVSI (in teils adaptierter Form) entnommen wurden (vgl. Goertz, 2022, S. 219–223). Dabei handelt es sich um je drei Items zum Typ IN, ID und UN sowie zwei zum Typ PL. Für einen zielgerichteten Einsatz in der Studie wurden die Testitems so „technisch zu Multiple-Choice- und Drag’N’Drop-Items adaptiert“ (vgl. Goertz, 2022, S. 219), dass das Testinstrument mit einem digitalen Endgerät bearbeitet werden kann.

3.2. Datenbasis

Die im Rahmen dieses Beitrags verwendeten Daten entstammen einer Design-Based Research Studie, in welcher der Untersuchungsschwerpunkt auf der Evaluation der Plattform FLexKom² (Fördern und Lernen experimenteller Kompetenzen) sowie der Wirksamkeit der entwickelten Module in Form zweier Lernzirkel lag (vgl. Goertz, 2022, S. 3–4 und S. 195). Diese zwei verschiedenen Lernzirkeltypen, jeweils bestehend aus 5 Modulen der Plattform FLexKom, wurden als Interventionen in der Studie eingesetzt. Während der eine Lernzirkel (nachfolgend mit Typ-A bezeichnet) verschiedene experimentelle Kompetenzschwerpunkte aus einem Inhaltsfeld (Elektrizitätslehre) besitzt, sind beim anderen Lernzirkel (nachfolgend mit Typ-B bezeichnet) bei immer dem gleichen experimentellen Kompetenzschwerpunkt (nämlich der VKS) die Inhaltsfelder verschieden.

An der Studie haben, wie Tab. 1 zu entnehmen ist, insgesamt N=443 SuS teilgenommen, die sich auf 15 Klassen aus 4 Schulen des Aachener Umlands aufteilen.

	Kl. 7	Kl. 8	Kl. 9	Gesamt
Anzahl Klassen	6	6	3	15
Anzahl SuS	146	216	81	443

Tab. 1: Übersicht über Stichprobengröße und Klassenaufteilung in der Studie. Das durchschnittliche Alter der Probanden lag bei 12,9 Jahren (52% weiblich, 40% männlich und 8% divers bzw. keine Angabe).

Alle Probanden haben in einem Prä-Test Teile des MeK-LSA-Tests zur Überprüfung allgemeiner experimenteller Kompetenzen (vgl. Theyßen et al., 2016) und das VKS-Testinstrument bearbeitet. Im Anschluss haben die Probanden im Klassenverbund in der Interventionsphase entweder den Typ-A-Lernzirkel oder den Typ-B-Lernzirkel bearbeitet. Gerahmt wurde die Durchführung der Lernzirkel durch eine Phase zur Motivation vor der Bearbeitung sowie eine Sicherungsphase nach der Durchführung (vgl. Goertz, 2022, S. 205–207). Im abschließenden Post-Test sind wiederum die beiden oben genannten Testinstrumente eingesetzt worden. Aus dieser Erhebung werden im Wesentlichen die für diesen Beitrag relevanten Daten zum Antwortverhalten der SuS aus dem VKS-Testinstrument für die Items vom Typ ID, IN und UN jeweils für die Zeitpunkte des Prä- und des Post-Tests verwendet. Die Daten für die Teilfähigkeit Planung kontrollierter Experimente (PL) konnten nicht im selben Format verarbeitet werden und wurden daher hier nicht berücksichtigt.

² Diese setzt sich als Ziel, experimentelle Kompetenzen im Physikunterricht explizit zu fördern. Genauere Informationen über die Plattform FLexKom sowie alle bisher entwickelten Module sind auf der Homepage des Schülerlabors Physik (SClphyLAB)

3.3. Sankey-Diagramme

Während das Antwortverhalten der SuS in Tests häufig nur dichotom als Zuordnung von „falsch“ oder „richtig“ betrachtet wird, ist das Kernelement der hier vorgestellten erweiterten Datenauswertung die detaillierte Analyse des Antwortwechselverhaltens der SuS bei den VKS-Test-Items unter Berücksichtigung aller verfügbaren Antwortoptionen. Hierbei kommen Sankey-Diagramme zum Einsatz, um u.a. die Antworthäufigkeiten und Antwortwechselströme darstellen zu können.

Die Namensbezeichnung des Sankey-Diagramms ist auf H. Riiall Sankey zurückzuführen, der eine solche Darstellung erstmalig 1898 nutzte, um den Wärmefluss einer realen und einer idealisierten Dampfmaschine zu vergleichen (vgl. Schmidt, 2006, S. 7–9; vgl. Kennedy und Sankey, 1898, S. 313; vgl. Schmidt, 2008, S. 84). Infolgedessen finden sie vor allem in ingenieurwissenschaftlichen Bereichen, wie zur Darstellung des Energieflusses, Anwendung (vgl. Schmidt, 2006; vgl. Schmidt, 2008). Im Alltag werden Sankey-Diagramme bei der Analyse von Wahlergebnissen verwendet, um Wählerwanderungen zwischen den Parteien zu visualisieren (vgl. tagesschau, 2021).

Zur Erstellung eines Sankey-Diagramms gibt es keine festgelegten Regeln, jedoch nennt Schmidt sog. implizite Annahmen zur Erstellung von Sankey-Diagrammen (vgl. Schmidt, 2006, S. 25). Die folgenden sind auch im Zusammenhang mit dem Einsatz von Sankey-Diagrammen in einem fachdidaktischen Kontext relevant und sinnvoll:

- Es handelt sich um Mengengrößen, die auf eine Zeitperiode bezogen sind.
- Die Mengenskala mittels der Pfeilbreite ist proportional, d.h. die doppelte Menge wird durch einen doppelt so breiten Pfeil dargestellt.
- Es werden i.d.R. keine Bestandsgrößen berücksichtigt, sondern Ströme.

Im fachdidaktischen Bereich der Bildungs- bzw. Lehr-Lern-Forschung ist die Nutzung solcher Diagramme z.B. zur Analyse des Antwortwechselverhaltens bei Wissens- bzw. Kompetenztests wenig verbreitet. Es finden sich aber Beispiele zur Nutzung von Sankey-Diagrammen zur Visualisierung experimenteller Prozesse bei größeren Probandengruppen (vgl. Büsch, 2021, S. 145–154; vgl. Büsch und Heinke, 2018; vgl. Fraß und Heinke, 2017).

Im Rahmen dieses Beitrags wurden Sankey-Diagramme eingesetzt, um für die Daten der VKS-Test-Items mittels einer polytomen Codierung³ (gemäß Antwort A, B, C und D) der Antwortmöglichkeiten das Antwortwechselverhalten der SuS zwischen den

der RWTH Aachen University unter der folgenden Adresse zu finden: <https://www.sciphylab.de/flexkom> (Stand: 18.05.2022)

³ In der Studie von Goertz (2022) wurden die Daten des VKS-Testinstruments dichotom ausgewertet.

beiden Testzeitpunkten (Prä- und Post-Test) (persönlich) zu verfolgen und für die gesamte Probandenkohorte visuell darzustellen. Technisch umgesetzt wurde dies mit dem Programm R unter Nutzung der Funktion „sankeyNetwork“ aus dem Package „networkD3“.

3.4. Design der Test-Items

Wie bereits im Abschnitt 3.1 erwähnt, besteht jedes Item (der Typen ID, IN und UN) aus vier Antwortmöglichkeiten. Bevor die jeweils drei Fragen eines jeden Itemtyps zusammengefasst in Sankey-Diagrammen dargestellt werden, wurden alle Antwortoptionen hinsichtlich ihrer Struktur untersucht. Im Fokus stand dabei insbesondere die Frage, welcher (fehlerhaften) VKS-Strategie (s. Abschnitt 2.2) die angebotenen falschen Antwortoptionen entsprechen und die damit einhergehenden potenzielle Attraktivität der Antwortoptionen für die Lernenden. In dieser Analyse ist deutlich geworden, dass bei allen drei eingesetzten Items vom Typ IN die falschen Antwortmöglichkeiten als strukturgleich anzusehen sind. Deshalb können die Ergebnisse für alle verwendeten Items vom IN-Typ in einem Sankey-Diagramm zusammenfassend betrachtet werden. Dies gilt nicht für die jeweils drei eingesetzten Items des Typs ID und UN, deren Ergebnisse deshalb jeweils in getrennten Sankey-Diagrammen erfasst werden.

3.5. Verwendete Datenfilterungen

Für die Darstellung in Sankey-Diagrammen war es notwendig, die Daten zu verarbeiten und zu filtern. Konkret wurden dabei die folgenden Kriterien genutzt:

1. Teilnahme an Prä- und Post-Test: Da für die Erstellung eines Sankey-Diagramms die Daten von (mindestens) zwei Testzeitpunkten benötigt werden, wurden die Daten von SuS, die nur an einer der beiden Erhebungen teilgenommen haben, entfernt.
2. Teilnahme am Lernzirkel: Um auf die Forschungsfragen zum Einfluss der Intervention Bezug nehmen zu können, sind nur die SuS berücksichtigt worden, die am Lernzirkel teilgenommen haben.
3. Fehlende Antwort/technischer Fehler bei einzelnen Items: Einige SuS haben einzelne Fragen zu einem der beiden Testzeitpunkte nicht beantwortet. In Einzelfällen konnte aufgrund technischer Fehler die Antwort nicht gespeichert werden. Die betroffenen Probanden wurden ausschließlich für das betroffene Item, aus dem unter Punkt 1 genannten Grund, nicht berücksichtigt.

Die beschriebenen Filterroutinen führen dazu, dass die Stichprobengrößen bei den einzelnen Sankey-Diagrammen zu den Test-Items einerseits geringer sind als in Tab. 1 beschrieben und andererseits bei den verschiedenen Itemtypen variieren. Darüber hinaus

wurden die Daten in Bezug auf inhaltliche Fragestellungen weiteren Filterungen unterzogen. Dazu gehören die Unterscheidung der absolvierten Lernzirkeltypen Typ-A und Typ-B. Innerhalb dieser Aufteilung wurden wiederum das allgemeine sowie gesondert das VKS-spezifische Vorwissen berücksichtigt. Die Einordnung des Vorwissens erfolgt dabei auf der Basis der Prä-Test Ergebnisse. Beim VKS-spezifischen Vorwissen wird als Score die Anzahl der richtigen Antworten der Items des VKS-Tests herangezogen. Beim allgemeinen Vorwissen wird ein Gesamtscore aus der Summe des VKS-Scores und den Ergebnissen des MeK-LSA Testteils gebildet. Die Probanden aus dem unteren Drittel des VKS-Scores bzw. des Gesamtscores werden den jeweiligen Gruppen mit niedrigem Vorwissen und diejenigen aus dem oberen Drittel den Gruppen mit hohem Vorwissen zugeordnet (vgl. Goertz, 2022, S. 323). Einen weiteren Aspekt stellt die Berücksichtigung der verschiedenen Jahrgangsstufen 7 bis 9 dar.

4. Forschungsfragen

Mit der Analyse des Antwortwechselverhaltens bei den Items des VKS-Testinstruments sollen folgende Forschungsfragen beantwortet werden:

FF1: Inwiefern haben die Rahmenbedingungen/Voraussetzungen (allgemeines und VKS-spezifisches Vorwissen sowie Jahrgangsstufe) und Lernsettings (Lernzirkeltyp) einen Einfluss auf das Antwortwechselverhalten beim VKS-Testinstrument?

FF2: Inwiefern sind Unterschiede im Antwortwechselverhalten zwischen den einzelnen Teilfähigkeiten (Itemtypen IN, ID und UN) zu beobachten? Forschungsfrage 1 ist itemübergreifend formuliert und adressiert mögliche Einflüsse verschiedener Rahmenbedingungen und Lernsettings im Hinblick darauf, ob sie Veränderungen im Kompetenzaufbau der VKS auslösen oder unterstützen. Etwas spezifischer formuliert ist dagegen Forschungsfrage 2, da sie stärker auf die einzelnen Teilfähigkeiten der VKS (IN, ID und UN) fokussiert und ihren möglichen Einfluss thematisiert.

5. Analyse des Antwortwechselverhaltens mithilfe von Sankey-Diagrammen

Für die vier genannten Differenzierungsaspekte (Lernzirkeltyp, allgemeines und VKS-spezifisches Vorwissen sowie Jahrgangsstufe) wurden jeweils Sankey-Diagramme erstellt, sodass insgesamt eine Vielzahl an Diagrammen aus den Daten erzeugt wurden. Die drei IN-Items wurden bei jeder Filterung zusammenfassend dargestellt, während die drei ID- und UN-Items jeweils in eigenen Sankey-Diagrammen abgebildet wurden. Für eine bessere Übersichtlichkeit der Diagramme ist die jeweils richtige Antwortmöglichkeit in Form grüner Balken markiert, während die falschen Antwortmöglichkeiten grau gefärbt sind. Ähnlich verhält es sich hinsichtlich der Antwortwechselströme. Ein Antwortwechsel zwischen Prä-

und Post-Test von einer falschen hin zu einer richtigen Antwort sowie das Beibehalten der richtigen Antwort wird ebenfalls durch einen grünen Pfeil visualisiert. Ein nicht korrektes Wechselverhalten wird gräulich dargestellt.

In Abb. 1 sind beispielhaft zwei Sankey-Diagramme für das Beispiel eines UN-Items gegenübergestellt; einmal beim Typ-A-Lernzirkel (zur E-Lehre, linke Seite) und einmal beim Typ-B-Lernzirkel (zur VKS, rechte Seite). Dieser beispielhafte Vergleich zeigt zwei Aspekte deutlich, die generell beim Vergleich der Ergebnisse für die beiden Lernzirkel zu beobachten waren. Zum einen konnten bei ähnlichen Prä-Test-Ergebnissen im Post-Test beim Typ-B-Lernzirkel mehr richtige Antworten und damit insgesamt auch mehr richtige Antwortwechsel beobachtet werden. Dies bestätigt die grundsätzlich bessere Lernförderung des Lernzirkels vom Typ-B- gegenüber dem Typ-A-Lernzirkel, die bereits von Goertz nachgewiesen wurde (vgl. Goertz, 2022, S. 293-347). Zum anderen konnten mithilfe der Sankey-Diagramme aber auch zum Teil deutliche Unterschiede im Antwortwechselverhalten zwischen den beiden Lernzirkeltypen beobachtet werden, welche bei den UN-Fragen deutlich stärker ausfielen als bei den leichteren Items IN und ID (Genauerer bei der Diskussion der zweiten Forschungsfrage). Bei ähnlichen Prä-Test-Ergebnissen zeigt das Wechselverhalten für das UN-Item in Abb. 1, dass die falsche Antwort C im Post-Test für die SuS beim Typ-A-Lernzirkel deutlich attraktiver war als bei den SuS des Typ-B-Lernzirkels. Dabei handelt es sich um die Aussage, dass der Einfluss zweier Variablen gleichzeitig überprüfbar ist. Die unterschiedlich ausgeprägte Zustimmung der SuS zu dieser Aussage nach Durchlaufen der beiden Lernzirkel zeigt unmittelbar die deutlich höhere Lernwirksamkeit der Intervention mit dem Lernzirkel vom Typ-B beim Ausräumen der Verständnisschwierigkeiten zur fehlenden Aussagekraft konfundierter Experimente.

Hinsichtlich des Einflusses vom allgemeinen Vorwissen kann insbesondere bei den leichteren Items IN

und ID keine verlässliche Aussage darüber getroffen werden, ob ein niedriges oder hohes Vorwissen mit einer positiveren Kompetenzentwicklung korreliert. Dies liegt vor allem daran, dass die Ergebnisse von den Probanden mit hohem Vorwissen schon beim Prä-Test Deckeneffekte zeigen, die sich im Post-Test verstärken, so dass eine potentielle Verbesserung mithilfe des Testinstruments unabhängig vom Lernzirkeltyp nicht mehr zuverlässig registriert werden konnte. Beim schweren Itemtyp UN sind beim Typ-A-Lernzirkel kleine Unterschiede für Probanden mit niedrigem und hohem Vorwissen festgestellt worden, die auf eine bessere Förderung dieser VKS-Teilfähigkeit bei SuS mit niedrigem Vorwissen hindeuten. Die Unterschiede sind jedoch nicht statistisch signifikant, weshalb hierzu keine abschließende Aussage getroffen werden kann. Beim Typ-B-Lernzirkel, bei dem in allen fünf Stationen explizit VKS-Teilfähigkeiten gefördert werden, ist hingegen zu beobachten, dass statistisch signifikant mehr SuS mit einem hohen Vorwissen von falschen zu richtigen Antworten wechseln als SuS mit einem niedrigen Vorwissen. Konkret ist der Zuwachs bei der richtigen Antwort D zwischen den Testzeitpunkten bei zwei der drei analysierten UN-Items für die SuS mit hohem Vorwissen höher. Zudem kann die Vorstellung, den Einfluss zweier Variablen gleichzeitig überprüfen zu können (Antwortmöglichkeit C), bei SuS mit hohem Vorwissen im Vergleich deutlicher abgebaut werden.

Auch das spezifische Vorwissen im Bereich der VKS wurde als potentielle Einflussgröße auf das Antwortwechselverhalten untersucht. Bei der Analyse der Sankey-Diagramme wurden vergleichbare Ergebnisse wie beim allgemeinen Vorwissen gefunden. Die Daten zeigen, dass sich die Anzahl der Antwortwechsel von falschen zu richtigen Antworten zwischen den Gruppen mit hohem bzw. niedrigem allgemeinen und spezifischen Vorwissen bei beiden Lernzirkeltypen nicht statistisch signifikant unterscheidet.

Eine potentielle Jahrgangsstufenabhängigkeit des Antwortwechselverhaltens ist der letzte Untersu-

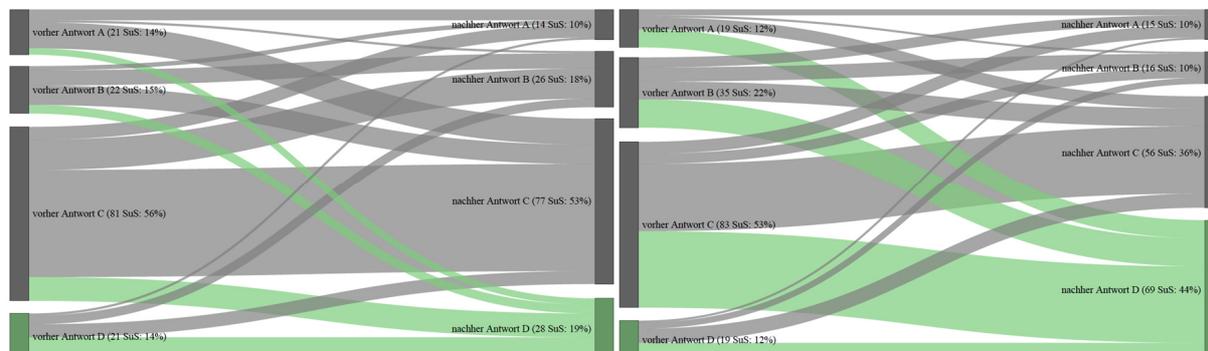


Abb. 1: Sankey-Diagramme zum Antwortverhalten der SuS im Vergleich des Prä- und Post-Tests zu einer UN-Frage im Kontext der Wärmelehre: links zum Typ-A-Lernzirkel (N=145) und rechts zum Typ-B-Lernzirkel (N=156). Bei dem Item sollte aus einem Experiment (bestehend aus zwei Experimentalansätzen á drei Variablen, die sich in zwei Variablen unterscheiden) geschlossen werden, welche Variable einen Einfluss hat. Die Antwortmöglichkeiten lauteten: Je eine einzelne Variable hat einen Einfluss (Antwort A und B), beide Variablen haben einen Einfluss (Antwort C) und es ist keine Aussage möglich (korrekte Antwort D).

chungsaspekt zur Beantwortung der Forschungsfrage 1. Während beim Itemtyp IN unabhängig vom Lernzirkeltyp kein jahrgangsstufenabhängiger Effekt beobachtet werden kann, weisen die Itemtypen ID und UN auf jeweils komplementäre Abhängigkeiten hin. Beim leichteren Typ ID sind Unterschiede, die auf eine bessere bzw. effektivere Förderung in niedrigeren Jahrgangsstufen hindeuten, bei beiden Lernzirkeltypen zu erkennen. Die Unterschiede hinsichtlich der Wechsel von falschen zu richtigen Antworten zwischen den einzelnen Jahrgangsstufen sind aber nicht statistisch signifikant. Dagegen wird beim Typ-B-Lernzirkel bei dem schwierigeren Item vom Typ UN teilweise eine Abhängigkeit von der Jahrgangsstufe beobachtet. Konkret steigt bei einem UN-Item (Frage 8) die Anzahl der Wechsel von falschen zu richtigen Antworten mit zunehmender Jahrgangsstufe. Die Unterschiede sind dort auch statistisch signifikant. Dass das richtige Wechselverhalten mit steigender Jahrgangsstufe ausgeprägter ist, scheint somit nur teilweise für einzelne Itemtypen, hier einzelne Items vom Typ UN, jedoch nicht im Allgemeinen belegt werden zu können.

Insgesamt kann die erste Forschungsfrage damit beantwortet werden. Der positive Effekt des Typ-B-Lernzirkel kann dabei allgemein hervorgehoben werden. Einflüsse auf das Antwortwechselverhalten vom Vorwissen sowie von der Jahrgangsstufe zeigen je nach Itemtyp unterschiedliche Abhängigkeiten, die jedoch keinem allgemeinen Trend folgen. Eine zusammenfassende Übersicht der diskutierten Einflüsse ist in Tab. 2 dargestellt.

Bereits bei der Beantwortung der Forschungsfrage 1 ist an verschiedenen Stellen Bezug auf ein unterschiedliches Antwortwechselverhalten in Abhängigkeit vom Itemtyp genommen worden, was in der Beantwortung der zweiten Forschungsfrage aufgegriffen wird. Grundlage der Argumentation sind vor allem die Sankey-Diagramme der Lernzirkelfilterung (Typ-A und Typ-B) und deren Vergleich für die Gesamtkohorte, da bei den anderen Differenzierungen potentielle Einflüsse auf das Antwortwechselverhalten auftreten können, die nicht direkt auf den Itemtyp zurückzuführen sind.

Grundsätzlich unterscheiden sich jeweils alle drei Itemtypen in dem beobachteten Antwortwechselverhalten. Die Wechselverhalten zwischen den Itemtypen IN und UN besitzen dabei Ähnlichkeiten, was auf das ähnliche Aufgabendesign der Items zurückzuführen

ren ist. Bei beiden Itemtypen lautet die Antwortmöglichkeit C, dass der Einfluss zweier Variablen gleichzeitig überprüft werden kann, und entspricht damit typischen Schwierigkeiten bei der Interpretation konfundierter Experimente und beim Verständnis ihrer mangelnden Aussagekraft. Bei beiden Itemtypen ist diese Auswahloption für die SuS im Vergleich zu den anderen Distraktoren vor allem im Post-Test deutlich attraktiver. Wie groß die Auswahlwahrscheinlichkeit der falschen Option C ausgeprägt ist, unterscheidet sich aber für beide Itemtypen. Die falsche Antwortoption C wird beim Typ UN im Vergleich zum Typ IN häufiger gewählt. Beim Typ ID sind die Auswahlwahrscheinlichkeiten der falschen Antworten hingegen relativ gleichmäßig verteilt.

Ähnlichkeiten beim Antwortwechselverhalten zwischen den Typen IN und ID sind vor allem bei den Sankey-Diagrammen beim Typ-B-Lernzirkel festzustellen. Hier werden grundsätzlich höhere Lösungswahrscheinlichkeiten bereits im Prä-Test gefunden, die mit Deckeneffekten verbunden sind. Folgerichtig stellt die richtige Antwort hier durchgängig auch die beliebteste Antwortoption dar. Im Gegensatz dazu zeigen die Sankey-Diagramme für die UN-Items ein auffallend ausgeprägteres Antwortwechselverhalten. Durch den höheren Schwierigkeitsgrad des Itemtyps gibt es neben einer vergleichsweise geringeren Zahl an Wechseln zur richtigen Antwort auch mehr Wechsel zu falschen Antworten als bei den IN- und ID-Items. Dies ist vermutlich auf eine höhere Ratewahrscheinlichkeit beim schwierigeren Itemtyp UN zurückzuführen. Dass bei allen drei UN-Items eine signifikant höhere Anzahl an Wechseln von falschen zu richtigen Antworten beim Typ-B- gegenüber dem Typ-A-Lernzirkel festgestellt wurde, zeigt, dass der Typ-B-Lernzirkel SuS bei der Überwindung von typischen Fehlvorstellungen zur VKS effektiv unterstützen kann.

Zusammenfassend kann die Forschungsfrage 2 damit beantwortet werden, dass sich das Antwortwechselverhalten für die verschiedenen Itemtypen unterscheidet. Die Unterschiede treten vor allem im Hinblick auf die Attraktivität der Antwortmöglichkeiten auf, die allerdings itemübergreifend von typischen Fehlvorstellungen zur VKS geprägt wird. Zusätzlich ist aber auch festzuhalten, dass SuS zum Erlernen der Teilfähigkeiten IN und ID keine ausschließliche Förderung der VKS benötigen, da sie diese Teilfähigkeiten auch beim Absolvieren des Typ-A-Lernzirkels weiterentwickeln konnten. Die Teilfähigkeit UN hin-

Teilfähigkeit der VKS	Typ-A-Lernzirkel	Typ-B-Lernzirkel			
	Gesamtkohorte	Gesamtkohorte	höhere Jgst.	hohes allgemeines VW	hohes VKS-spezifisches VW
ID	x	x			
IN	x	x			
UN		x	x	x	x

Tab. 2: Übersicht über die wesentlichen Einflüsse auf eine positive Kompetenzentwicklung der VKS, die bei der Analyse mithilfe von Sankey-Diagrammen festgestellt werden konnten. VW steht für Vorwissen und Jgst. für Jahrgangsstufe.

gegen benötigt eine umfassende und fokussierte Förderung wie beim Typ-B-Lernzirkel, um von SuS erlernt bzw. weiterentwickelt werden zu können.

6. Fazit zum Einsatz von Sankey-Diagrammen

An dieser Stelle soll ein kurzes Fazit mit Bezug auf Vor- und Nachteile gezogen werden, die der Einsatz von Sankey-Diagrammen zur Datenauswertung mit sich bringt.

Ein großer Vorteil ist, dass ein einzelnes Diagramm einen sehr hohen Informationsgehalt besitzt. In Bezug auf diesen Beitrag können so nicht nur die Häufigkeiten der einzelnen Antwortmöglichkeiten zu beiden Testzeitpunkten gleichzeitig dargestellt werden, sondern darüber hinaus auch das Wechselverhalten. Dadurch ist schnell ersichtlich, zwischen welchen Antwortmöglichkeiten die SuS hin und her wechseln. Die daraus resultierenden Schlussfolgerungen geben einen besseren Einblick, was z.B. durch eine Intervention ausgelöst wurde, als dies bei einer dichotomen Auswertung der Daten möglich ist. Dies gilt insbesondere, weil Sankey-Diagramme eine personenscharfe Rückverfolgung ermöglichen und somit z.B. auch Wechsel von richtigen Antworten im Prä-Test zu falschen Antwortoptionen im Post-Test sichtbar machen. Einen weiteren Vorteil des Sankey-Diagramms bietet die grafische Darstellungsform an. Trotz vieler Informationen ist sie sehr übersichtlich und anschaulich. Begründet ist dies durch die Tatsache, dass die Pfeilbreiten die Änderungsverhältnisse (hier die Antwortwechsel) visuell direkt abbilden können (vgl. Millar, 2005, S. 15).

Der hohe Informationsgehalt jedes einzelnen Sankey-Diagramms bedingt, dass die Analyse bzw. der Vergleich mehrerer Diagramme nur eingeschränkt möglich ist. Dies liegt daran, dass die aus den Diagrammen zu entnehmenden Informationen für einen Vergleich eines Untersuchungsaspekts zum Teil nochmal separat aufbereitet (z.B. in Tabellen) oder in einer anderen Form visualisiert werden müssen, um eine tiefergehende Analyse zu ermöglichen.

7. Zusammenfassung und Ausblick

In diesem Beitrag wurden Daten eines VKS-Testinstruments, die in einer Studie mit 443 SuS der Jahrgangsstufen 7 bis 9 (vgl. Goertz, 2022) erhoben und dichotom ausgewertet wurden, polytom analysiert. Um insbesondere das Antwortwechselverhalten erfassen und darstellen zu können, wurden dazu Sankey-Diagramme eingesetzt. Dabei konnte bei der Beantwortung der Forschungsfragen gezeigt werden, dass verschiedene Rahmenbedingungen und Voraussetzungen zum Teil unterschiedliche Einflüsse auf den Kompetenzaufbau und die Entwicklung der VKS bei SuS in der Mittelstufe haben. Ein VKS-Schwerpunktlernzirkel zeigt eine positive Entwicklung für leichte und schwere Teilfähigkeiten der VKS, während ein Lernzirkel, der verschiedene experimentelle Kompetenzen adressiert, lediglich bei

leichteren VKS-Teilfähigkeiten eine Kompetenzentwicklung ermöglicht. Das Vorwissen der SuS korreliert nur beim schwereren Itemtyp UN mit der Kompetenzentwicklung, da bei diesem Typ mehr SuS mit einem hohen Vorwissen im Prä-Post-Vergleich zur richtigen Antwort wechseln als unter SuS mit niedrigem Vorwissen. Unterschiede zwischen dem allgemeinen und VKS-spezifischem Vorwissen konnten nicht beobachtet werden. Eine jahrgangsstufenabhängige Entwicklung zugunsten der älteren SuS konnte ebenfalls nur im Rahmen der schwereren Teilfähigkeit UN beobachtet werden. Dazu muss aber erwähnt werden, dass die Daten der leichteren Items (IN und ID) durch Deckeneffekte nur eingeschränkte Aussagekraft haben.

Der Einsatz von Sankey-Diagrammen zur Analyse der Daten hat einerseits gezeigt, dass diese einen gezielteren Aufschluss darüber geben können, was während einer Lern-/Erarbeitungsphase im Hinblick auf eine konzeptionelle Veränderung passiert. Jedoch hat sich auch gezeigt, dass ein Vergleich vieler Sankey-Diagramme nur bedingt möglich ist.

Der Beitrag zeigt, dass es im Bereich der VKS noch Potential gibt, den Kompetenzerwerb genauer zu verstehen, was mit Sankey-Diagrammen grundsätzlich visuell gut umsetzbar ist. Im Rahmen weiterer Untersuchungen können verschiedenen Möglichkeiten verfolgt werden, um das Verständnis hinsichtlich der Kompetenzentwicklung der VKS bei SuS zu verbessern. Diese umfassen u.a. eine Weiterentwicklung des VKS-Testinstruments im Hinblick auf ein vergleichbares Aufgabendesign der einzelnen Items, um zusammenfassende Darstellungen aller Itemtypen in Sankey-Diagrammen zu ermöglichen. Aufbauend darauf kann das Testinstrument hin zu einem adaptiven bzw. interaktiven Design erweitert werden.

8. Literaturverzeichnis

- Barzel, Bärbel, Bernd Reinoffer und Marcus Schrenk (2012). „Das Experimentieren im Unterricht“. In: Experimentieren im mathematisch-naturwissenschaftlichen Unterricht – Schüler lernen wissenschaftlich denken und arbeiten. Hrsg. von Werner Rieß, Markus A. Wirtz, Bärbel Barzel und Andreas Schulz. Münster: Waxmann Verlag, S. 103–127.
- Brandenburger, Martina und Silke Mikelskis-Seifert (2019). „Facetten experimenteller Kompetenz in den Naturwissenschaften“. In: Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Kiel 2018. Hrsg. von Christian Maurer, S. 77–80.
- Brandenburger, Martina, Silke Mikelskis-Seifert, Martin Schwichow und Jens Wilbers (2020). „Variablenkontrollstrategien in der Grundschule“. In: Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen. Gesellschaft

- für Didaktik der Chemie und Physik, Jahrestagung in Wien 2019. Hrsg. von Sebastian Habig, S. 130–133.
- Büsch, Leonard und Heinke, Heidrun (2018). Experimentelle Handlungsabläufe sichtbar machen! Methoden & Ergebnisse. In: C. Maurer (Hrsg.), *Qualitätsvoller Chemie- und Physikunterricht – normative und empirische Dimensionen*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Regensburg 2017, S. 364–367.
- Büsch, Leonard (April 2021). „Entwicklung und Einsatz von neuartigen Methoden zur Erfassung und Analyse experimenteller Prozesse“. Diss. Aachen: RWTH Aachen University.
- Carey, Susan, Risa Evans, Maya Honda, Eileen Jay und Christopher Unger (1989). „An experiment is when you try it and see if it works: A study of grade 7 students understanding of the construction of scientific knowledge“. In: *International Journal of Science Education* 11 (special issue), S. 514–529.
- Chen, Zhe und David Klahr (1999). „All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy“. In: *Child Development* 70(5), S. 1098–1120.
- Ehmer, Maïke (Mai 2008). „Förderung von kognitiven Fähigkeiten beim Experimentieren im Biologieunterricht der 6. Klasse: Eine Untersuchung zur Wirksamkeit von methodischem, epistemologischem und negativem Wissen“. Diss. Kiel: Christian-Albrecht-Universität.
- Fraß, Stephan und Heinke, Heidrun (2017). Auf der Suche nach Strategien bei der Manipulation von Experimenten. In: C. Maurer (Hrsg.), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Zürich 2016, S. 312–315.
- Goertz, Simon (2022, in Druck). *Module und Lernzirkel der Plattform FLexKom zur Förderung experimenteller Kompetenzen in der Schulpraxis - Verlauf und Ergebnisse einer Design-Based Research Studie*. Berlin: Logos Verlag Berlin GmbH.
- Hammann, Marcus und Roman Asshoff (2015). *Schülervorstellungen im Biologieunterricht - Ursachen für Lernschwierigkeiten*. 2. Auflage. Seelze: Klett Kallmeyer.
- Hammann, Marcus, T. T. H. Phan, M. Ehmer und H. Bayrhuber (2006). „Fehlerfrei Experimentieren“. In: *Der mathematische und naturwissenschaftliche Unterricht (MNU)* 59(5), S. 292–299.
- Kennedy, Alex B. W. und H. Riall Sankey (1898). „The Thermal Efficiency of Steam Engines. Report of the Committee Appointed to the Council Upon the Subject of the Definition of a Standard or Standards of Thermal Efficiency for Steam Engines: With an Introductory Note“. In: *Minutes of the Proceedings of the Institution of Civil Engineers* 134(4), S. 278–312.
- Millar, Robin (2005). *Teaching about energy*. York: University of York, Department of Educational Studies.
- Nerdel, Claudia (2017). *Grundlagen der Naturwissenschaftsdidaktik – Kompetenzorientiert und aufgabenbasiert für Schule und Hochschule*. Berlin: Springer-Verlag.
- Schmidt, Mario (2006). „Der Einsatz von Sankey-Diagrammen im Stoffstrommanagement“. In: Hrsg. von Ansgar Häfner, Karl-Heinz Rau und Christa Wehner. *Beiträge der Hochschule Pforzheim* (No. 124). Pforzheim: Hochschule Pforzheim.
- Schmidt, Mario (2008). „The Sankey Diagram in Energy and Material Flow Management: Part I: History“. In: *Journal of Industrial Ecology* 12(1), S. 82–94.
- Schulz, Andreas, Markus Wirtz und Erich Staraschek (2012). „Das Experiment in den Naturwissenschaften“. In: *Experimentieren im mathematisch-naturwissenschaftlichen Unterricht - Schüler lernen wissenschaftlich denken und arbeiten*. Hrsg. von Werner Rieß, Markus A. Wirtz, Bärbel Barzel und Andreas Schulz. Münster: Waxmann Verlag, S. 15–38.
- Schwichow, Martin (Oktober 2015). „Förderung der Variablen-Kontroll-Strategie im Physikunterricht“. Diss. Kiel: Christian-Albrecht-Universität.
- Schwichow, Martin, Simon Christoph, William J. Boone und Hendrik Härtig (2016). „The impact of sub-skills and item content on students' skills with regard to the control-of variables strategy“. In: *International Journal of Science Education* 38(2), S. 216–237.
- Schwichow, Martin und Andreas Nehring (2018). „Variablenkontrolle beim Experimentieren in Biologie, Chemie und Physik: Höhere Kompetenzausprägungen bei der Anwendung der Variablenkontrollstrategie durch höheres Fachwissen? Empirische Belege aus zwei Studien“. In: *Zeitschrift für Didaktik der Naturwissenschaften* 24, S. 217–233.
- tagesschau (27. September 2021). *Wie die Wähler wanderten*. <https://www.tagesschau.de/inland/btw21/waehlerwanderung-bundestagswahl-103.html>. Abgerufen am 20.01.2022.
- Theyßen, Heike, Horst Schecker, Knut Neumann, Bodo Eickhorst und Martin Dickmann (2016). „Messung experimenteller Kompetenz - ein computergestützter Experimentierertest“. In: *Physik und Didaktik in Schule und Hochschule (PhyDid-A)* 15(1), S. 26–48.
- Tschirgi, Judith E. (1980). „Sensible reasoning: A hypothesis about hypotheses“. In: *Child Development* 51(1), S. 1–10.