

Auswerten von Versuchsdaten Wo benötigen Oberstufenschüler Unterstützung?

Theoretischer Hintergrund und Stand der Forschung

Mit dem Konstrukt des Scientific Reasoning bzw. Wissenschaftlichen Denkens werden in der naturwissenschaftsdidaktischen Forschung die kognitiven Komponenten naturwissenschaftlichen Problemlösens referenziert. Üblich ist dabei eine Dreiteilung in die Komponenten Hypothesen generieren, Untersuchungen planen sowie Daten auswerten (Klahr, 2000; Opitz, Heene & Fischer, 2017). Das Scientific Reasoning wird als wissensbasiert konzeptualisiert. Relevante Wissensbereiche sind dabei das prozedurale Wissen, das Fähigkeiten und Strategien naturwissenschaftlicher Denk- und Arbeitsweisen umfasst; das epistemische Wissen, welches deklarative Wissensbestände über Konzepte und Prinzipien der Erkenntnisgewinnung, wie z.B. Gütekriterien, und ihre Begründungen enthält; sowie domänenspezifisches Fachwissen (Arnold, 2015; Kind, 2013).

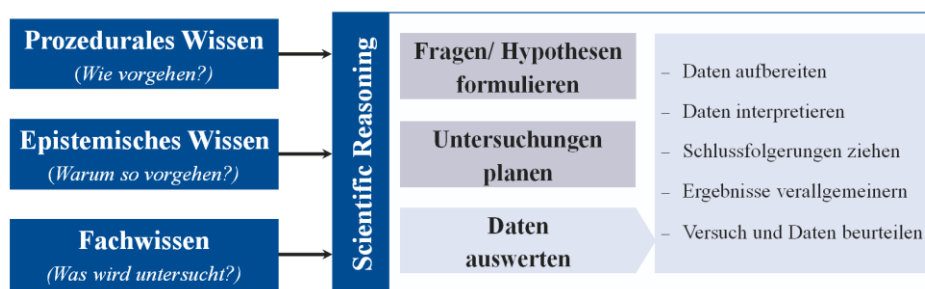


Abb. 1 Scientific Reasoning und relevante Wissensbereiche. Teilkompetenzen sind hier für die Komponente Daten auswerten ausdifferenziert dargestellt.

Die oben beschriebenen Komponenten des Scientific Reasoning lassen sich je nach intendierter Ausschärfung in Teilkompetenzen einteilen. So lässt ein Überblick über vorliegende Konzeptualisierungen der Komponente des Auswertens von Daten (bspw. Chinn & Malhotra, 2002; Wellnitz & Mayer, 2013) eine Ausdifferenzierung in fünf Teilkompetenzen plausibel erscheinen, die in Abbildung 1 dargestellt sind. Es konnte außerdem gezeigt werden, dass Schülerinnen und Schüler beim Auswerten von Daten mit Schwierigkeiten konfrontiert sind, die sich vom Ableiten von Trends aus Daten, über das Belegen von Schlussfolgerungen mit Versuchsergebnissen, bis zum Berücksichtigen von Gütekriterien beim Beurteilen von Daten erstrecken (Jeong, Songer & Lee, 2007; Lubben & Millar, 1996; Sandoval & Millwood, 2005).

Forschungsfragen

Da diese Schwierigkeiten meist in der Physik bzw. in domänenunspezifischen Kontexten bei jüngeren Lernenden untersucht wurden, ist allerdings bislang wenig darüber bekannt, welche Schwierigkeiten Lernenden beim Auswerten prozedural anspruchsvollerer Versuche im

Chemieunterricht der Oberstufe begegnen. Auch ist bislang kaum überprüft worden, ob und inwiefern das Auftreten solcher Schwierigkeiten entsprechend der oben beschriebenen Konzeptualisierung des Scientific Reasoning mit dem prozeduralen, epistemischen und fachlichen Vorwissen der Schülerinnen und Schüler zusammenhängt. Das hier beschriebene Dissertationsprojekt begegnet diesen Desideraten mit den folgenden Forschungsfragen:

FF1. Welche Schwierigkeiten bei der Datenauswertung bestehen bei Lernenden der Oberstufe im Fach Chemie beim Auswerten von aus Experimenten gewonnenen Daten?

FF2. Welche Zusammenhänge bestehen zwischen dem Auftreten dieser Schwierigkeiten und dem prozeduralen und epistemischen Wissen sowie dem Fachwissen der Lernenden?

Studiendesign

Zur Beantwortung der ersten Forschungsfrage wurde ein qualitatives Design gewählt, das auf der Videographie von Lernenden der Oberstufe während der Auswertung selbsterhobener Daten basiert. Schülerinnen und Schüler führen in Zweiergruppen Versuche durch, bei denen Sie mithilfe digitaler Sensoren Datensätze zu Titrationen generieren. Die Lernenden werden auf Grundlage dreier Testinstrumente (s.u.) Zweiergruppen mit ähnlichen Voraussetzungen eingeteilt. Ziel der Versuche ist jeweils das hypothesenprüfende Untersuchen von Salzlösungen hinsichtlich ihrer Säure-Base-Eigenschaften. Ein Versuch befasst sich mit dem Untersuchen der Pufferwirkung eines Phosphatpuffers, der andere basiert auf dem Vergleichen der Säurestärken zweier Salzlösungen. Beide Versuche sind durch Vorgabe von Hypothese und Versuchsdesign für die Lernenden vorstrukturiert, um eine möglichst hohe Vergleichbarkeit zwischen den Gruppen hinsichtlich der entsprechenden Phasen des Experimentierprozesses sicherzustellen. Die Auswertung selbst erfolgt mithilfe von offenen Aufgabenstellungen zu den oben genannten fünf Komponenten der Datenauswertung, die die Lernenden innerhalb ihrer Gruppe gemeinsam bearbeiten. Eine Erhebung von $N = 50$ videografierten Schülergruppen wird angestrebt.

Die Analyse der resultierenden Videodaten zielt auf die Kommunikation der Lernenden während des Auswertens der in ihren Versuchen erhobenen Daten ab. Ziel ist es dabei, durch Anwendung einer induktiven qualitativen Inhaltsanalyse (Mayring, 2015) Schwierigkeitskategorien zu identifizieren. Um hierzu ein Codesystem zu entwickeln, wird zunächst bei $N = 20$ Schülergruppen eine turnbasierte Analyse durchgeführt. Als Selektionskriterium für das Kodieren eines Turns wird dabei auf die Definition einer „unmittelbar beobachtbaren Schülerschwierigkeit“ von Kechel (2016, S. 57ff.) zurückgegriffen. Diese nutzt das Verfehlen eines Teilziels der Aufgabe, unerwünschte Handlungen und Ausdrücke von Mühe und/oder negativem Erleben durch die Lernenden zur Identifikation einer Schwierigkeit. Das resultierende Codesystem wird anschließend deduktiv auf die Videos der restlichen Gruppen angewendet.

Die Ergebnisse dieser qualitativen Analyse sollen als Basis einer Clusteranalyse dienen, die überprüft, ob die Lernenden verschiedenen Schwierigkeitsprofilen zugeordnet werden können. Das Auftreten der Schwierigkeiten sowie die resultierenden Schwierigkeitsprofile sollen in der Folge mithilfe von regressions- und varianzanalytischen Verfahren hinsichtlich ihres Zusammenhangs mit dem Vorwissen der Lernenden in den oben beschriebenen Wissensbereichen hin untersucht werden. Die Ergebnisse dieser Analysen werden zur Beantwortung der zweiten Forschungsfrage herangezogen.

Ergebnisse der Pilotierung

Zum Zwecke der Erhebung der o.g. Wissensbereiche wurden Testinstrumente adaptiert, die in einer Pilotstudie mit Lernenden der Qualifikationsstufe an Gymnasien in NRW ($N = 94$) erprobt wurden. Das Instrument zum prozeduralen Wissen besteht aus 15 Multiple-Choice-Items, die aus verschiedenen bestehenden Instrumenten (Henke, 2007; Nehring, 2015; Vorholzer, von Aufschnaiter & Kirschner, 2016) adaptiert wurden. Das Instrument zum epistemischen Wissen wurde von Arnold (2015) entwickelt und um drei Items auf insgesamt 13 Single-Choice-Items gekürzt. Beim Instrument zum Fachwissen handelt es sich um 24 Single-Choice-Items, welche sich mit dem Themenbereich Säuren und Basen befassen und aus einem für den Einsatz in der Oberstufe entwickelten Itempool von Hülsmann (2015) stammen. Die Ergebnisse der Pilotierung sind in Tabelle 1 ersichtlich.

Tab. 1: Kennwerte der pilotierten Testinstrumente

Skala	Item-anzahl	Mittelwert		Reliabilität (α)	Trennschärfe (r_{ii})	Schwierigkeit (p)
		M	SD			
Prozedurales Wissen	15	41.55 (von 56)	6.25	.68	.07 - .48	.34 - .88
Epistemisches Wissen	13	5.18 (von 13)	2.20	.52	-.06 - .40	.07 - .81
Fachwissen	24	10.97 (von 24)	4.23	.74	-.07 - .58	.06 - .74

Die Ergebnisse für die Skala zum prozeduralen Wissen wurden zum Anlass für eine inhaltliche Überarbeitung des Tests genommen. Hier wurde insbesondere ein Augenmerk auf den hohen Mittelwert gelegt, der anzeigt, dass Varianz durch zu leicht zu erreichende Punkte verloren geht. Außerdem weisen niedrige Trennschärfen von Items am Ende des Tests auf eine zu geringe Testzeit hin. Der Test zum Fachwissen wird im Zuge einer ohnehin intendierten Verringerung der Itemzahl um die Items mit ungünstigen Kennwerten gekürzt. Während die beiden vorgenannten Instrumente bereits in ihrer pilotierten Form ausreichende Reliabilität zeigen, bleibt der Test zum epistemischen Wissen hinter den Anforderungen zurück. Eine weitreichend überarbeitete Version wird gegenwärtig pilotiert.

Eine weitere Pilotstudie widmete sich der Erprobung des Experimentiersettings zum Thema Phosphatpuffer. Insgesamt war in dieser Studie eine Stichprobe von $N = 70$ Lernenden beteiligt, die in 35 Zweiergruppen aufgeteilt wurden. Elf dieser Gruppen wurden beim Auswerten des Versuchs videographiert. Eine qualitative Überprüfung der von den Lernenden generierten Datensätze zeigt an, dass ein großer Teil in der Lage war, mithilfe der für sie neuartigen digitalen Sensoren Datensätze zu generieren, die für die Bearbeitung der offenen Auswertungsaufgaben geeignet waren. Wo dies nicht der Fall war, konnten als Ursache häufig Bedienungsfehler identifiziert werden, denen in der Hauptstudie mit genaueren Instruktionen begegnet wird. Des Weiteren wurden die Antworten der Lernenden auf die offenen Aufgaben untersucht. Hier zeigt sich, dass nur zwölf der 35 Gruppen mehr als zwei der insgesamt neun Aufgaben bearbeitet haben. Da sich die ersten beiden Aufgaben auf das Aufbereiten der Daten in Graphen beziehen, lässt sich erschließen, dass das Durchführen des Versuchs und das anschließende Aufbereiten der Daten sich für viele Gruppen als zu zeitaufwendig darstellt. Das Setting wurde daher unter diesem Aspekt überarbeitet. Die Videodaten der videographierten Zweiergruppen wurden zur Überprüfung der Eignung des aus Kechel (2016) adaptierten Kodiermanuals induktiv auf Schwierigkeiten bei der Datenauswertung hin untersucht. Als Artefakt der oben beschriebenen häufig unvollständigen Bearbeitungen wurden dabei hauptsächlich Schwierigkeiten für die Komponente *Daten aufbereiten* identifiziert. Es konnten allerdings auch für alle anderen Komponenten Schwierigkeiten aufgezeigt werden, sodass von einer ausreichenden Sensitivität des Kodiermanuals und somit einer Eignung zur Verwendung in der Hauptstudie ausgegangen werden kann. Ein zweiter Kodierdurchgang wird zur summativen Prüfung der Intercoder-Reliabilität genutzt.

Literatur

- Arnold, J., Kremer, K. & Mayer, J. (2014). Understanding Students' Experiments – What kind of support do they need in inquiry tasks? *International Journal of Science Education*, 36 (16), 2719–2749. <https://doi.org/10.1080/09500693.2014.930209>
- Arnold, J. (2015). *Die Wirksamkeit von Lernunterstützungen beim Forschenden Lernen*. Berlin: Logos.
- Chinn, C. A. & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools. A theoretical framework for evaluating inquiry tasks. *Science Education*, 86 (2), 175–218. <https://doi.org/10.1002/sce.10001>
- Henke, C. (2007). *Experimentell-naturwissenschaftliche Arbeitsweisen in der Oberstufe. Untersuchung am Beispiel des HIGHSEA-Projekts in Bremerhaven*. Berlin: Logos.
- Hülsmann, C. (2015). *Kurswahlmotive im Fach Chemie. Eine Studie zum Wahlverhalten und Erfolg von Schülerinnen und Schülern in der gymnasialen Oberstufe*. Berlin: Logos.
- Jeong, Heisawn; Songer, Nancy B.; Lee, Soo-Young (2007): Evidentiary Competence. Sixth Graders' Understanding for Gathering and Interpreting Evidence in Scientific Investigations. In: *Research in Science Education* 37 (1), S. 75–97. 10.1007/s11165-006-9014-9.
- Kechel, J.-H. (2016). *Schülerschwierigkeiten beim eigenständigen Experimentieren. Eine qualitative Studie am Beispiel einer Experimentieraufgabe zum Hooke'schen Gesetz*. Berlin: Logos.
- Kind, P. M. (2013). Establishing Assessment Scales Using a Novel Disciplinary Rationale for Scientific Reasoning. *Journal of Research in Science Teaching*, 50 (5), 530–560. <https://doi.org/10.1002/tea.21086>
- Klahr, D. (2000). *Exploring Science*. Cambridge/London: MIT Press.
- Lubben, F. & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education*, 18 (8), 955–968. <https://doi.org/10.1080/0950069960180807>
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12., überarb. Aufl.). Weinheim: Beltz.
- MSW NRW. (2014). *Kernlehrplan für die Sekundarstufe II Gymnasium/Gesamtschule in Nordrhein-Westfalen: Chemie*. Verfügbar unter https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/151/KLP_GOST_Chemie.pdf
- Nehring, A. (2015): *Wissenschaftliche Denk- und Arbeitsweisen im Fach Chemie. Eine kompetenzorientierte Modell- und Testentwicklung für den Bereich der Erkenntnisgewinnung*, Berlin: Logos.
- Opitz, A., Heene, M. & Fischer, F. (2017). Measuring scientific reasoning – a review of test instruments. *Educational Research and Evaluation*, 23 (3-4), 78–101. <https://doi.org/10.1080/13803611.2017.1338586>
- Roberts, R. & Gott, R. (2004). A written test for procedural understanding. A way forward for assessment in the UK science curriculum? *Research in Science & Technological Education*, 22 (1), 5–21. <https://doi.org/10.1080/0263514042000187511>
- Sandoval, W. A. & Millwood, K. A. (2005). The Quality of Students' Use of Evidence in Written Scientific Explanations. *Cognition and Instruction*, 23 (1), 23–55. https://doi.org/10.1207/s1532690xci2301_2
- Toplis, R. (2007). Evaluating Science Investigations at Ages 14–16. Dealing with anomalous results. *International Journal of Science Education*, 29 (2), 127–150. <https://doi.org/10.1080/09500690500498278>
- Vorholzer, A.; Aufschnaiter, C. von; Kirschner, S. (2016): Entwicklung und Erprobung eines Tests zur Erfassung des Verständnisses experimenteller Denk- und Arbeitsweisen. In: *Zeitschrift für Didaktik der Naturwissenschaften* 22 (1), S. 25–41. <https://doi.org/10.1007/s40573-015-0039-3>
- Wellnitz, N. & Mayer, J. (2013). Erkenntnismethoden in der Biologie – Entwicklung und Evaluation eines Kompetenzmodells. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 315–345.