

Wie bearbeiten Schülerinnen und Schüler Rohdaten aus Experimenten?

Einleitung

In einem Forschungsprojekt wird untersucht, mit welchen Argumentationen Schülerinnen und Schüler eine zuvor aufgestellte Hypothese verwerfen bzw. stützen, wenn experimentelle Daten mit unterschiedlicher Genauigkeit vorliegen. Genauigkeit wird durch die Anzahl an (Nachkomma-)stellen operationalisiert. Dazu wird zusätzlich zu einer Erfassung von bestimmten Argumentationen mit einem geschlossenen Fragebogen (Ludwig & Priemer, 2014) ein Instrument entwickelt, das den Prozess der Be- bzw. Verarbeitung von experimentellen Rohdaten zur Beurteilung der Richtigkeit einer Hypothese erfasst. Es wurden bei einer Vorstudie zur Erprobung des Instruments ca. 60 Schülerinnen und Schüler der 8. Klasse gebeten, in einem strukturierten offenen Antwortformat ihren Datenauswertungsprozess anhand von Messdaten zum Vergleich der Fallzeit beim freien Fall und waagrechtem Wurf schriftlich zu dokumentieren. Zusätzlich wurden Teilgruppen direkt nach der schriftlichen Bearbeitung des Testinstruments interviewt. Dieser Beitrag stellt das Instrument sowie die Ergebnisse der Entwicklungs- und Validierungsstudie vor.

Zielsetzung

Das Instrument zur schriftlichen Erfassung der Bearbeitung von Rohdaten wurde entwickelt, um den wichtigen Schritt von den experimentellen Rohdaten zur Beurteilung der Eingangshypothese genauer aufzulösen. Da keine Testverfahren bekannt waren, die dies leisten können, sollte dies anhand von (teilweise) offenen Antwortformaten explorativ geschehen (vergleiche Masnick & Morris 2002, Engl 2014).

Forschungsfragen

In der Studie wurde untersucht, ob

- der Prozess der Datenbearbeitung mit den vorgegebenen schriftlich auszufüllenden Fragen ausreichend gut erfasst werden kann und, wenn dies der Fall ist,
- welche Tätigkeiten Schülerinnen und Schüler beim Bearbeiten von Rohdaten ausführen und welche Schlüsse sie aus den Ergebnissen ziehen.

Methode

Testkonstruktion Die Schülerinnen und Schüler erhalten eine Form eines Versuchsprotokolls, das teilweise (Fragestellung, Versuchsdurchführung, ...) vorausgefüllt ist und zwei Messreihen enthält, die verglichen werden sollen. Die in den ausgefüllten Bereichen gegebenen Informationen stellen sicher, dass die Schülerinnen und Schüler die inhaltlichen Voraussetzungen besitzen, um eine sinnvolle Eingangshypothese abzugeben und diese anhand der experimentellen Messdaten zu beurteilen. Die Items des Instruments bestehen aus einer offenen Frage zur Bearbeitung der Daten („Schreibe für einen Mitschüler eine ausführliche Anleitung, wie er beim Auswerten der Daten vorgehen soll, um die Vermutung überprüfen zu können. Führe das beschriebene Vorgehen mit den Messdaten durch.“) und einer Ankreuzmöglichkeit mit angegliederter offener Begründung zur Bewertung der Daten („Bewertest du die Messdaten als genau gleich, ungefähr gleich, deutlich unterschiedlich oder nicht vergleichbar? Begründe.“).

Ablauf Nach Aufwerfen der Fragestellung zu den beiden Fallbewegungen erhalten die Schülerinnen und Schüler drei Eingangshypothesen zur Wahl (Fallzeit bei senkrechtem Fall bzw. waagerechter Wurf größer, kleiner oder gleich), anschließend wird das Experiment

beschrieben, vorgeführt und Messreihen gegeben. Danach folgen die beiden beschriebenen Items und abschließend wird die Eingangshypothese beurteilt (Wechseln oder Beibehalten der Eingangshypothese mit Begründung). Zur Beantwortung der ersten Forschungsfrage wurden zusätzlich zum Protokoll mit $n=30$ Probanden Interviews geführt, deren Leitfaden sich an den schriftlichen Items orientierte. Die zweite Forschungsfrage wurde durch Kategorienbildung und -auszählung der Schülerantworten zu den schriftlichen Items beantwortet.

Probanden Die Studie fand im Mai 2015 mit zwei Klassen der 8. Klassenstufe ($n=60$) eines Berliner Gymnasiums statt.

Auswertung Die Auswertung erfolgte gemäß der zwei Fragestellungen: Es wurden zum einen die Items von $n=30$ Probanden inhaltlich ausgewertet und kategorisiert (Mayring, 2010) und die schriftlichen Antworten mit denen der zugehörigen Interviews verglichen. Zum anderen wurden mit einem aus 30 Fragebögen erstellten Kodiermanual die restlichen 30 Fragebögen durch zwei Beurteiler kodiert. Die Beurteilerübereinstimmung wurde anhand Cohens Kappa berechnet. Dann wurde die Verteilung der gefundenen Kategorien über die gesamten 60 Fragebögen erfasst.

Ergebnisse

Es mussten keine Fragebögen aufgrund offensichtlicher Falschantworten aus der Stichprobe entfernt werden. Der Anteil an unbeantworteten Items war sehr gering, so dass alle 60 Datensätze für den Fragebogen bzw. 30 für das Interview Verwendung finden konnten.

Es zeigte sich im Vergleich von schriftlichen Schülerantworten und den zugehörigen Interviews bei allen 30 Datensätzen eine sehr gute inhaltliche Übereinstimmung. Insbesondere wurden keine relevanten Zusatzinformationen aus dem Interview gewonnen. Des Weiteren zeigte sich, dass der Großteil der Probanden die Items nicht nur verstanden hatte, sondern auch in der Lage war, adäquat schriftlich zu antworten.

Zur Beantwortung der Frage, welche Tätigkeiten Schülerinnen und Schüler beim Bearbeitung der Daten ausführen und welche Schlüsse sie aus den Daten ziehen, werden im Folgenden die Kategorien zu den beiden Items und deren Verteilung vorgestellt.

Zu Item 1 (Bearbeiten der Rohdaten zum Vergleich der beiden Messreihen) konnten zunächst fünf Kategorien gefunden werden. Die Namen der Kategorien beschreiben das von den Schülerinnen und Schülern angewendete Verfahren zur Bearbeitung der Daten. Dabei verteilen sich die Antworten aller 60 Fragebögen folgendermaßen:

Kategorie (Art der Bearbeitung der Messreihen)	Verteilung (von 60 Probanden)
Arithmetisches Mittel bilden	31
Paarweises Vergleichen	6
Unspezifisches Vergleichen	11
Summenvergleich (gesamt)	1
Differenzenvergleich (paarweise)	4
Nicht zuzuordnen	7
Beurteilerübereinstimmung: Cohens Kappa = 0,79	

Tab. 1: Kategorien und deren Verteilung zum Item 1 zur Bearbeitung der Rohdaten

Zur offenen Begründung der Wahl zu den geschlossenen Antwortoptionen bei Item zwei (Bewertung der Daten als genau gleich, ungefähr gleich, deutlich unterschiedlich und nicht zu vergleichen, jeweils mit Begründung) ergaben sich die folgenden vier Kategorien und die entsprechende Verteilung:

Kategorie (Begründung für die Bewertung der Daten bezüglich Gleichheit der Messreihen)	Verteilung (von 60 Probanden)
Kleiner Unterschied (nicht quantifiziert)	19

Kleine Zeitdifferenz (quantifiziert)	15
Messunsicherheiten (Grund für Unterschied)	8
Offensichtlich (z.B. „Sieht man doch!“)	14
Nicht zuzuordnen	4
Beurteilerübereinstimmung: Cohens Kappa = 0,91	

Tab. 2: Kategorien und deren Verteilung zum Item 2 zur Bewertung der Messdaten

Diskussion

Das schriftliche Protokollverfahren lieferte für die Erfassung des Umgangs mit Messdaten sehr klare Ergebnisse: Die Schülertätigkeiten waren sehr gut zu kategorisieren und lassen vermuten, dass sich diese gut in geschlossenen Items umsetzen lassen. Einige Kategorien (z.B. Summenvergleich) tauchten nur sehr selten auf. Da das Ziel aber war, explorativ zu erfassen, welche Bearbeitungsweisen Schülerinnen und Schüler überhaupt anwenden, ist es sinnvoll, alle Kategorien beizubehalten. Ferner wurden die Kategorien für die Überarbeitung des Begleitinstrumentes nach theoretischen Überlegungen weiterhin um die beiden folgenden Kategorien ergänzt: Median und Modus.

Die Erfassung der Bewertung der Daten lieferte zwar ebenfalls gut einzuordnende Kategorien, diese waren allerdings z. T. inhaltlich voneinander abhängig. Des Weiteren nutzten die Probanden nur zwei der vier Antwortoptionen (Messwerte sind genau gleich, bzw. ungefähr gleich). In den Begründungen wurde deutlich, dass vielen Probanden der Unterschied zwischen der Bewertung der eigentlichen Messdaten und dem Rückschluss auf die zu untersuchenden theoretischen Fallzeiten nicht klar war. Der Schritt von den bearbeiteten Rohdaten hin zu einer Beurteilung bezüglich der Eingangshypothese ist jedoch nicht trivial und viele Schülerinnen und Schüler schlussfolgern mit den verschiedensten Begründungen nicht konform mit ihren aufgearbeiteten Rohdaten (Chinn & Brewer, 1998; Kanari & Millar, 2004; Lin, 2007). Zur Verbesserung der Erfassung muss klarer zwischen der Bewertung der Messdaten und den Rückschlüssen auf theoretische Zusammenhänge unterschieden werden.

Zusammenfassung

Eine schriftliche Erfassung des Prozesses der Datenbearbeitung mit dem vorgestellten Instrument ist gut möglich und eine Umsetzung der gefundenen Kategorien in geschlossene Antwortoptionen voraussichtlich durchführbar. Die Ergebnisse zur Bewertung der Messreihen durch die Schülerinnen und Schüler machen deutlich, dass bei dem vorgestellten Item noch Bearbeitungsbedarf besteht.

Literatur

- Chinn, C. A. & Brewer, W. F. (1998). An Empirical Test of a Taxonomy of Responses to Anomalous Data in Science. *Journal of Research in Science Teaching*, 35(6), 623-654
- Engl, L. et al (2014). Entwicklung eines Messinstrumentes zur Erfassung der Protokollierfähigkeit – initiiert durch Video-Items. *Zeitschrift für Didaktik der Naturwissenschaften*, DOI 10.1007/s40573-014-0023-3
- Kanari, Z. & Millar, R. (2004). Reasoning from Data: How Students Collect and Interpret Data in Science Investigations. *Journal of Research in Science Teaching*, 41(7), 748-769
- Lin, J.-Y. (2007). Responses to Anomalous Data Obtained From Repeatable Experiments in the Laboratory. *Journal of Research in Science Teaching*, 44(3), 506-528
- Ludwig, T. & Priemer, B. (2014). Ein Instrument zur Erfassung von Argumentationen beim Experimentieren. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in München 2013 (S. 273 - 275). Kiel: IPN.
- Masnack, A. M. & Morris, B. J. (2002). Reasoning from Data: The Effect of Sample Size and Variability on Children's and Adults' Conclusions. *Proceedings of the 24th annual conference of the cognitive science society* (S. 634-638). Austin, TX: Cognitive Science Society.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (11. Auflage, erste Auflage 1983). Weinheim und Basel: Beltz-Verlag