

Der Einfluss von Nachkommastellen auf die Hypothesenänderung aufgrund von Daten

Einleitung

Das Evaluieren von Daten stellt eine Kernkompetenz von Schülerinnen und Schülern (SuS) im naturwissenschaftlichen Unterricht dar (Chinn & Malhotra, 2002). McNeill & Krajcik (2007) haben gezeigt, dass „data literacy“ fundamental für naturwissenschaftliches Argumentieren ist, und viele Länder haben die Evaluation von Daten in ihre Lehrpläne eingebettet. SuS sollen fähig sein, Daten als Evidenz für einen wissenschaftlichen Claim oder eine Hypothese zu nutzen. Dafür müssen SuS in der Lage sein, Daten zu analysieren und interpretieren (Klahr & Dunbar, 1988).

Bezüglich des Umgangs mit Daten von SuS konstatieren Masnick & Morris (2008), dass SuS über ein gutes Verständnis der Konzepte Mittelwert, signifikante Stellen und Anzahl von Messwerten verfügen, komplexere Konzepte wie der Varianz jedoch mehr Aufmerksamkeit gewidmet werden muss. Das Konzept der Varianz zeigt sich im naturwissenschaftlichen Unterricht besonders anhand von Messwerten in physikalischen Experimenten sowie deren Messunsicherheiten. Jedoch weisen SuS viele Schwierigkeiten im Umgang mit diesen Messunsicherheiten auf (Priemer & Hellwig, 2016). So behaupten z. B. viele SuS, dass bei wiederholtem Messen der Messwert derselbe bleibt und eine „perfekte Messung“ existiert, in der keine Fehler auftreten (Allie, Buffler, Lubben, & Campbell, 2002). Viele SuS betrachten ihre Messergebnisse in der Folge als Einzelfälle und beziehen sich nicht auf das Gesamtbild des Datensatzes (Lubben & Millar, 1996). Wenn die Daten im Gegensatz zu ihrer vorherigen Aussage stehen, ignorieren SuS die Resultate darüber hinaus häufig (Kanari & Millar, 2004; Klahr & Dunbar, 1988).

Hier gibt es ein Spannungsfeld. Einerseits verstehen SuS, dass signifikantere Messwerte zu besseren, genaueren Daten führen. Andererseits treten Messunsicherheiten deutlicher hervor, wenn die Resultate signifikanter werden. Ziel dieser Studie ist deshalb, den Einfluss der Anzahl von Nachkommastellen auf den Hypothesenwechsel zu untersuchen, nachdem SuS Messwerte eines Experiments analysiert haben.

Methode

In dieser Studie haben 153 Testpersonen der gymnasialen Klassenstufen 8 bis 10 (Durchschnittsalter 14 Jahre) in Berlin teilgenommen. Der Kontext des Experiments war so ausgewählt, dass bei den Testpersonen keine Vorkenntnisse anzunehmen waren. Die Testpersonen wurden durch ein Video auf das Experiment vorbereitet (4 Minuten), in dem der Aufbau und die Methode schriftlich und mit einer Skizze erklärt werden (vgl. Abb. 1). Diese Information war zudem auf dem ausgeteilten Fragenbogen jederzeit zugänglich.

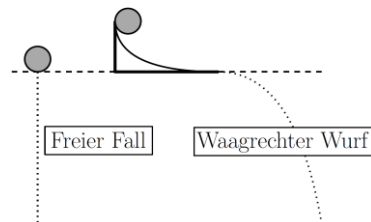


Abb. 1: Skizze des experimentellen Aufbaus.

Im vorgeführten Experiment wird ein Gegenstand im freien Fall fallen gelassen, wobei die Zeit gemessen wird. Anschließend wird der gleiche Gegenstand von der gleichen Höhe mit einer horizontalen Geschwindigkeit – waagerechter Wurf – fallen gelassen, auch hier wird die Zeit gemessen. Die Testpersonen wurden zuvor gefragt, welcher Gegenstand die längste Fallzeit haben wird. Nach dieser Frage wurden die Testpersonen zufällig in eine von drei Gruppen eingeteilt, wobei sich die Gruppen in der Anzahl der Nachkommastellen der jeweiligen Fallzeiten unterschieden. Gruppe A erhielt die Daten mit zwei, Gruppe B mit drei und Gruppe C mit vier Nachkommastellen. Tabelle 1 zeigt die Messwerte der Fallzeiten (in Sekunden) des Freien Falls und des waagerechten Wurfs für Gruppen A, B, und C:

Freier Fall			Waagerechter Wurf		
A	B	C	A	B	C
0,53	0,535	0,5351	0,53	0,535	0,5350
0,53	0,535	0,5353	0,53	0,535	0,5352
0,53	0,535	0,5355	0,53	0,534	0,5347
0,53	0,534	0,5347	0,53	0,535	0,5354
0,54	0,534	0,5349	0,53	0,535	0,5351
0,53	0,534	0,5348	0,53	0,535	0,5352

Tabelle 1: Die Fallzeiten (in Sekunden) des freien Falls und des waagerechten Wurfs für Gruppen A, B, und C.

Anschließend wurden die Testpersonen befragt, wie sie die Daten bewerten würden. Die Multiple-Choice-Optionen hierbei sind: 1 = die Summe der Messreihen vergleichen, 2 = die Durchschnitte vergleichen, 3 = den Median vergleichen, 4 = die Daten Schritt für Schritt paarweise vergleichen, 5 = den Modus vergleichen, 6 = die Messreihen auf Unterschiede prüfen, 7 = die Differenzen von Paaren miteinander vergleichen, oder 8 = keine der genannten Möglichkeiten. Im weiteren Verlauf wurden die Testpersonen gefragt, was für einen Unterschied sie zwischen den Messreihen sehen: einen deutlichen, einen kleinen, oder keinen Unterschied. Abschließend wurden die Probanden nochmals gefragt, welcher Gegenstand die längste Fallzeit hat und wurden um eine Begründung gebeten. Die Begründungen wurden nach quantitativen Aussagen (numerische Äußerungen wie „0,01 s Unterschied“), qualitativen Aussagen (nicht-numerische Aussagen wie „ein kleiner Unterschied“ oder theoretische Begründung), oder Sonstiges klassifiziert. Die Begründungen sind doppelkodiert und erhalten eine sehr gute Übereinstimmung (Cohens Kappa $\kappa = .93$). Vier Testpersonen sind als „Sonstiges“ kodiert, fünf Testpersonen haben keine Begründung angegeben und wurden deshalb aus der Statistik entfernt.

Resultate

Tabelle 2 zeigt die prozentuale Verteilung, wie die Testpersonen ihre Hypothese gewechselt haben. Der Unterschied zwischen den Verteilungen der Gruppen ist signifikant, $\chi^2(8) = 15.55, p < .05$, mit einem mittelstarken Effekt.

Kategorie	Gruppe		
	A (n = 52)	B (n = 52)	C (n = 49)
Korrekt → Korrekt (CC)	35 %	25 %	18 %
Falsch → Korrekt (FC)	40 %	31 %	33 %
Korrekt → Falsch (CF)	0 %	9 %	10 %
Falsch → Andere Falsche (FoF)	13 %	27 %	16 %
Falsch → Gleiche Falsche (FF)	12 %	8 %	23 %

Tabelle 2: Prozentuale Verteilung von Testpersonen über die verschiedenen Hypothesenwechselkategorien.

Die Anzahl der Testpersonen, die bei einer korrekten Hypothese enden (CC + FC) beträgt 75 % bei Gruppe A und 56 % sowie 51 % für die Gruppen B und C. Auffällig ist der Wechsel der Testpersonen in den Gruppen B und C zu je 10 % von einer richtigen zu einer falschen Hypothese (CF). Diese Ursache konnte nicht in der benutzten Auswertungsstrategie gefunden werden. Es gab keinen signifikanten Unterschied zwischen der Strategieverteilung in den Gruppen A, B und C ($p > .1$). Im Durchschnitt geben 76 % der Testpersonen an, dass sie den Mittelwert oder die Summe von Reihen miteinander vergleichen. Weiterhin lässt sich kein signifikanter Unterschied zwischen den Verteilungen, wie die Testpersonen den Unterschied zwischen Datensätzen sehen, feststellen ($p > .1$).

Im Durchschnitt geben 83 % der Testpersonen an, dass sie einen kleinen Unterschied sehen. Bei der qualitativen und quantitativen Begründung der Hypothese gibt es einen signifikanten Unterschied zwischen den Verteilungen der Gruppen A und B+C (zusammen) mit $\chi^2(1) = 4.24$, $p < .05$. In Gruppe A geben 32 % eine quantitative Begründung, in den Gruppen B und C sind es lediglich 17 %. Tabelle 3 zeigt die prozentuale Verteilung qualitativer und quantitativer Begründungen als Funktion verschiedener Hypothesen. Der Unterschied zwischen den Verteilungen ist signifikant mit $\chi^2(2) = 10.13$, $p < .01$.

	Quantitativ (n = 32)	Qualitativ (n = 112)
Gleich	84 %	54 %
Freier Fall	6 %	30 %
Waagrechter Wurf	10 %	16 %

Tabelle 3: Die Prozentualverteilung von qualitativer und quantitativer Begründung und der Hypothese.

Von den Testpersonen, die eine quantitative Begründung geben, wählen 84 % die richtige Hypothese, nachdem sie die Datensätze analysiert haben. Die Testpersonen hingegen, die eine qualitative Begründung angeben, wählen zu 54 % die richtige Hypothese.

Konklusion

Die Anzahl von Nachkommastellen kann die Hypothesenwahl beeinflussen. In Gruppe A – zwei Nachkommastellen – enden 75 % bei der richtigen Hypothese, in den Gruppen B und C – drei bzw. vier Nachkommastellen – sind es nur jeweils 56 % und 51 %. Darüber hinaus wechseln 10 % der Testpersonen in den Gruppen B und C von einer richtigen zu einer falschen Hypothese. Obwohl 76 % der Testpersonen angeben, den Mittelwert oder die Summe von Reihen zu vergleichen, sind sie nicht in der Lage diesen Unterschied im vorliegenden Versuch richtig einzuschätzen. Das Zeigen von Messwerten mit mehr Nachkommastellen führt weiterhin dazu, dass Testpersonen weniger quantitative Begründungen geben. Testpersonen mit einer quantitativen Begründung wählen aber häufiger die richtige Hypothese.

Wie können wir nun SuS bei Aufgaben helfen, in denen Datensätze miteinander verglichen werden müssen? Das Reduzieren von Nachkommastellen erscheint auf Grundlage unserer Ergebnisse als eine Option, dies halten wir im Sinne einer Förderung des Umgangs mit Daten nicht für richtig. Unsere Ergebnisse weisen jedoch ebenso darauf hin, dass es sinnvoll erscheint, Varianz und Messunsicherheiten expliziter im naturwissenschaftlichen Unterricht zu adressieren, sofern man den Umgang mit Daten bei SuS fördern möchte. Resultate beim Experimentieren sollten nicht nur aus einem (Mittel)Wert bestehen, sondern auch aus einer Art Konfidenzintervall. Hiermit bekommen SuS ein quantitatives Hilfsmittel, womit sie den Unterschied zwischen zwei Mittelwerten einschätzen können. Wir vermuten, dass dieses Intervall SuS in die Lage versetzt, unabhängig von der Anzahl der Nachkommastellen quantitative Begründungen anzustellen und so bessere Aussagen in Hinblick auf die Güte von Daten zu treffen.

Literatur

- Allie, S., Buffler, A., Lubben, F., & Campbell, B. (2002). Point and Set Paradigms in Students' Handling of Experimental Measurements. In H. Behrendt, H. Dahncke, R. Duit, W. Gräber, M. Komorek, A. Kross, & P. Reiska (Eds.), *Research in Science Education - Past, Present, and Future* (pp. 331–336). Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47639-8_47
- 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>
- 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. <https://doi.org/10.1002/tea.20020>
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1–48. [https://doi.org/10.1016/0364-0213\(88\)90007-9](https://doi.org/10.1016/0364-0213(88)90007-9)
- 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>
- Masnack, A. M., & Morris, B. J. (2008). Investigating the Development of Data Evaluation: The Role of Data Characteristics. *Child Development*, *79*(4), 1032–1048. <https://doi.org/10.1111/j.1467-8624.2008.01174.x>
- McNeill, K. L., & Krajcik, J. (2007). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. In M. Lovett & P. Shah (Eds.), *Thinking with Data* (pp. 233–267). Psychology Press.
- Priemer, B., & Hellwig, J. (2016). Learning About Measurement Uncertainties in Secondary Education: A Model of the Subject Matter. *International Journal of Science and Mathematics Education*, *16*(1), 45–68. <https://doi.org/10.1007/s10763-016-9768-0>