

Sophia Chroszczinsky¹
Burkhard Priemer¹
Amy Masnick²

¹Humboldt-Universität zu Berlin
²Hofstra University New York

Der Einfluss des inhaltlichen Kontextes auf die Einschätzung der Qualität von Messdaten

Einleitung

Lernende zeigen vielfältige Schwierigkeiten beim fachgerechten Umgang mit Daten, die z. B. aus Experimenten stammen können und u. U. den eigenen Erwartungen widersprechen (Kanari & Millar, 2004; Kok et al., 2019; Lubben & Millar, 1996; Masnick et al., 2017). Vor diesem Hintergrund liegt unser Erkenntnisinteresse darin, den Einfluss der Eigenschaften von Daten - wie z. B. der Varianz eines Datensatzes - und von inhaltlichen Kontexten - wie z. B. einem physikalischen Themenbereich - auf den Umgang mit Daten von Lernenden zu untersuchen. Umgang bedeutet hier, aus einem Vergleich von Datensätzen Schlüsse zu ziehen.

Theoretischer Hintergrund

Schulz et al. (2018) haben gezeigt, dass Studierende nichtnaturwissenschaftlicher Fächer bestimmte Eigenschaften von Daten oder Datensätzen, z. B. kleinere Varianzen, als „besser“ empfinden und entsprechende Datensätze deshalb bevorzugen. Allerdings können diese eher intuitiven Entscheidungen auch zu fachlich falschen Schlüssen führen (Kok et al., 2019). Ferner wurde auch beobachtet, dass Lernende beim Umgang mit Daten zum Teil nicht-rationale Begründungen heranziehen (Chinn & Brewer, 1998; Ludwig et al., 2019; Masnick & Morris, 2008). Das lässt darauf schließen, dass Lernende über statistische Vor-Vorstellungen verfügen und Entscheidungen auf Basis dieser treffen. Aber Lernende haben nicht nur statistische Vor-Vorstellungen zu Daten, sondern auch zu inhaltlichen Kontexten (z. B. Schecker et al., 2018), was sich z. B. beim Aufstellen einer Hypothese zu einem Experiment zeigt. Es konnte beobachtet werden, dass eher fachlich korrekte Schlussfolgerungen aus Daten gezogen werden, wenn Daten und Kontexte rational ausgewertet werden (Ludwig et al., 2019) oder Daten vorher aufgestellte Hypothesen stützen (Kanari & Millar, 2004). Ergänzend wurde in der Studie von Masnick et al. (2017) beobachtet, dass wenn Daten eine Hypothese stützten, eher datenbasierte Argumentationen auftreten. Wenn die Daten eine Hypothese von der fachlichen Perspektive her allerdings nicht stützen, dann ist bei Lernenden häufiger eine kontextualisierte Argumentation zu beobachten (Masnick et al., 2017) und die Daten werden eher unter dem Einfluss eines *Confirmation Biases* entsprechend der aufgestellten Eingangshypothese interpretiert (Kanari & Millar, 2004). Um zu diesem Themenfeld am Beispiel des Vergleichs von zwei Datensätzen - ein Datensatz steht für eine bestimmte Messmethode - beizutragen, wurden folgende Forschungsfragen untersucht.

Forschungsfragen

- FF1: Welche Unterschiede zeigen sich bei der Wahl eines „bevorzugten“ Messverfahrens, wenn zwei Datensätze ohne Kontext bzw. zwei Datensätze mit Kontext präsentiert werden?
- FF2: Welche Unterschiede zeigen sich bei der Wahl eines „bevorzugten“ Messverfahrens, wenn zwei Kontexte ohne Datensätze bzw. zwei Kontexte mit Datensätzen präsentiert werden?

Method

Die Forschungsfragen wurden im Rahmen einer größeren, fragebogenbasierten empirischen Studie (Pre-Registrierung bei AsPredicted im Mai 2020) beantwortet. Die Daten wurden zwischen September und Dezember 2020 erhoben. Bei den Teilnehmenden handelte es sich um Studierende nicht-physikalischer und nicht-statistischer Fachrichtungen aus den USA und Deutschland. Insgesamt wurden 125 vollständige Datensätze generiert und ausgewertet (USA: 101, Deutschland: 24). Das Durchschnittsalter der Teilnehmenden betrug 21,5 Jahre (USA: 19,2 Jahre, Deutschland: 31,1 Jahre). Die Studierenden sollten zunächst ohne inhaltlich beschriebene Kontexte zwei Datensätze, bestehend aus je sechs Messwerten inklusive angegebene Mittelwert und Standardabweichung, die zu zwei fiktiven Messmethoden gehören, miteinander vergleichen und sich für den für sie „besseren“ Datensatz entscheiden. Anschließend wurden den Teilnehmenden zwei unterschiedliche Methoden zur Bestimmung der Durchschnittsgeschwindigkeit eines Fahrrads (ohne Daten) vorgestellt. Die Teilnehmenden sollten nun ohne Daten bewerten, welches Messverfahren sie für „besser“ halten. Abschließend haben die Studierenden fiktive Messdaten (dieselben wie im ersten Teil der Studie) zu den Messverfahren erhalten und sollten sich nun auf Grundlage der Daten erneut für eines der beiden Messverfahren entscheiden.

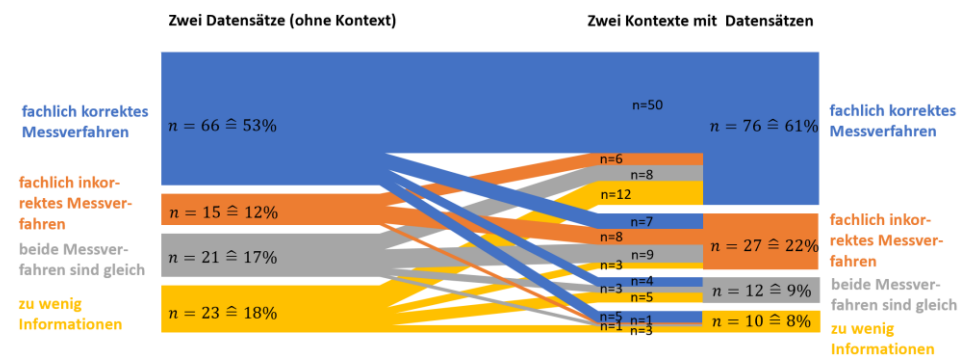


Abb. 1: Entwicklung der Entscheidungen der Teilnehmenden für ein Messverfahren ohne Kontext bzw. mit Daten und Kontext

Ergebnisse

Auf der linken Seite der Abb. 1 wird zunächst dargestellt, wie viele der Teilnehmenden welche datenbasierte (kontextlose) Entscheidung für ein Messverfahren getroffen haben. Die Korrektheit wurde auf Basis der kleineren Standardabweichung beurteilt. Weiterhin sind in Abb. 1 die Entscheidungswechsel der Teilnehmenden zu erkennen, wenn sich später mit Hilfe von Daten und inhaltlichem Kontext entschieden werden musste. Diese spätere datenbasierte Entscheidung im Kontext wird auf der rechten Seite von Abb. 1 dargestellt. Es ist zu erkennen, dass knapp über die Hälfte der Teilnehmenden (53%) bei der rein datenbasierten Wahl eines Messverfahrens den aus fachlicher Perspektive korrekten Datensatz ausgewählt hat. Wenn sich mit Daten und Kontext entschieden werden musste, ist zu erkennen, dass dies 61% der Teilnehmenden korrekt tun. Wenn man diese beiden Verteilungen statistisch mit Hilfe des Wilcoxon-Vorzeichen-Rang-Tests vergleicht, dann ist festzustellen, dass es keine signifikanten Unterschiede bei der Wahl des Messverfahrens mit Kontext ($Mdn=2$, Skala von 0 bis 2) und ohne Kontext ($Mdn=2$) gibt ($T=622$, $p=.53$, $r=-.04$). Wenn nur die Teilnehmenden einbezogen werden, die sich nicht für die fachlich korrekte Messreihe entschieden haben, fällt die Wahl

eines Messverfahrens mit Kontext ($Mdn=1$, Skala von 0 bis 2) signifikant besser aus als ohne Kontext ($Mdn=0$) ($T=92$, $p<.01$, $r=-.36$). Die Effektstärke ist als mittelgroß einzustufen. In Abb. 1 fällt auch auf, dass die Teilnehmenden, die ohne den Kontext angegeben haben, entweder zu wenig Informationen für eine Entscheidung zur Verfügung zu haben oder angeben haben, dass beide Messverfahren gleich gut geeignet sind, mehrheitlich mit Kontext eine davon abweichende Entscheidung treffen.

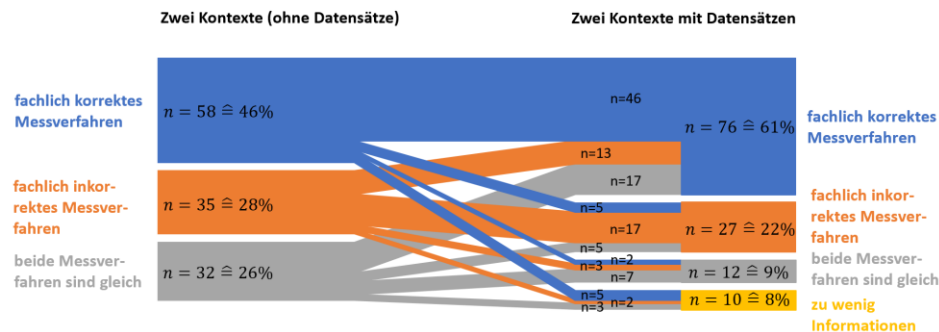


Abb. 2: Entwicklung der Entscheidungen der Teilnehmenden für ein Messverfahren im Kontext ohne Daten bzw. mit Daten und Kontext

Abb. 2 ist analog zu Abb. 1 aufgebaut. Es wurde lediglich die linke Seite der Abbildung ausgetauscht, damit nun erkennbar wird, welche Erwartungen die Teilnehmenden zu den Kontexten hatten, ohne dass Daten zur Beurteilung zur Verfügung standen. Es ist zu erkennen, dass 46% der Teilnehmenden auch ohne Daten die korrekte Erwartung bzgl. des „besseren“ Messverfahrens hatten. Wenn diese Verteilungen statistisch mit Hilfe des Wilcoxon-Vorzeichen-Rang-Tests verglichen werden, ist festzustellen, dass sich die Wahl eines Messverfahrens mit Daten ($Mdn=2$, Skala von 0 bis 2) signifikant verbessert im Vergleich zur datenlosen Vermutung ($Mdn=0$) ($T=230,5$, $p<.01$, $r=-.24$). Die Effektstärke ist allerdings als schwach einzustufen. Wenn auch hier die Teilnehmenden aus dem Test herausgelassen werden, die zuvor schon die fachlich korrekte Erwartung an das Ergebnis hatten, dann ist ebenfalls festzustellen, dass sich die datenbasierte Wahl im Kontext ($Mdn=1$, Skala von 0 bis 2) signifikant verbessert im Vergleich zur datenlosen Vermutung ($Mdn=0$) ($T=0$, $p<.01$, $r=-.48$). Hier handelt es sich um einen starken Effekt. Auch in Abb. 2 ist auf die deskriptiv beobachtbare Tendenz hinzuweisen, dass sich die Teilnehmenden, die zuvor angaben, dass sich beide Messverfahren gleich gut eignen, mehrheitlich mit den neuen Informationen in Form von Daten zu einer Neubewertung der Situation kommen, da nur 7 der 32 Teilnehmenden bei der Entscheidung bleiben.

Diskussion

Das Hinzufügen von Informationen in Form von Daten bzw. Kontext verändert die Auswahl für ein Messverfahren insgesamt nicht stark. Aber Studierende, die zuvor „falsche“ oder eher „ausweichende“ Antworten gegeben haben, beurteilen die Datensätze mit mehr Informationen etwas besser als zuvor. Das heißt, dass die hinzugefügten neuen Informationen für diese Studierenden hilfreich bei der Beurteilung von Datensätzen zu sein scheinen. Diese Effekte müssen allerdings noch genauer untersucht werden. Denn die Ergebnisse der Studie werden dadurch limitiert, dass lediglich ein Inhaltsbereich als Kontext vorgegeben worden ist und die Gelegenheitsstichprobe einen recht kleinen Umfang hatte. Die Studie befindet sich aktuell in weiterer Auswertung.

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
- Kanari, Z. & Millar, R. (2004). Reasoning from Data: How Students Collect Data and Interpret Data in Science Investigations. *Journal of Research in Science Teaching*, 41 (3), 748-769
- Kok, K., Priemer, B., Musold, W. & Masnick, A. (2019). Students' conclusions from measurement data: The more decimal places, the better? *Physics Review Physics Education Research* 15, 010103. <https://doi.org/10.1103/PhysRevPhysEducRes.15.010103>
- Lubben, F. & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education*, 18 (8), 955-968
- Ludwig, T., Priemer, B. & Lewalter, D. (2019). Assessing Secondary School Students' Justifications for Supporting or Rejecting a Scientific Hypothesis in the Physics Lab. *Research in Science Education*. <https://doi.org/10.1007/s11165-019-09862-4>
- Masnick, A. M., Klahr, D. & Knowles, E. R. (2017). Data-Driven Belief Revision in Children and Adults. *Journal of Cognition and Development*, 18 (1), 87-109
- Masnick, A. M. & Morris, B. J. (2008). Investigating the Development of Data Evaluation: The Role of Data Characteristics. *Child Development*, 79 (4), 1032-1048
- Schecker, H., Wilhelm, T. Hopf, M & Duit, R. (2018). *Schülervorstellungen und Physikunterricht*. Heidelberg: Springer Spektrum.
- Schulz, J., Priemer, B. & Masnick, A. (2018). Student's preferences when choosing data sets with different characteristics. Paper presented at 10th International Conference on Teaching Statistics