

Schüler_innen schreiben Texte – Physiklehrkräfte beurteilen diese

Eine Vielzahl von Autor_innen weist darauf hin, dass das Verfügen über bildungs- und fachsprachliche Fähigkeiten entscheidend zum Lernerfolg von Schüler_innen im Physikunterricht beiträgt (z. B. Rincke & Markic, 2018; Schleppegrell, 2004; Wellington & Osborne, 2001). Die Sprachkenntnisse von Schüler_innen beeinflussen nicht nur deren Aneignung von naturwissenschaftlichen Denk und Arbeitsweisen, sondern auch die Art und Weise, in der eine_e Schüler_in demonstrieren kann, welches Fachwissen und welche physikspezifischen Fähigkeiten er_sie bereits erworben hat (Lyon, 2013; Höttecke et al., 2018). Gleichzeitig findet sich in der fachdidaktischen Forschung Evidenz dafür, dass viele Physiklehrkräfte sich nicht für die Förderung sprachspezifischer Fähigkeiten und Kompetenzen zuständig fühlen (Airey, 2012) und dass ihnen nicht bewusst ist, dass die Sprachkenntnisse von Schüler_innen ihr Lernen und ihre Leistung im Physikunterricht beeinflussen (Luykx et al., 2007). Darüber hinaus setzen es viele Physiklehrkräfte als selbstverständlich voraus, dass sich ihre Schüler_innen im Fachunterricht in einer sprachlich versierten Art und Weise ausdrücken können (Feilke, 2012; Tajmel, 2017). Insgesamt scheinen viele Physiklehrkräfte hohe sprachliche Anforderungen in ihrem Fachunterricht zu stellen.

Unsere Vermutung ist, dass sich diese hohen Anforderungen auch in der Art und Weise niederschlagen, wie Physiklehrkräfte die Qualität von Schülertexten im Rahmen einer Klassenarbeit beurteilen. Genauer gesagt: Wir vermuten, dass Physiklehrkräfte bei der Beurteilung von Klassenarbeiten neben fachlich-konzeptuellen stets auch sprachliche Merkmale von Schülertexten berücksichtigen und dass Physiklehrkräfte hierbei hohe Erwartungen an die sprachliche Realisierung eines Schülertextes stellen. Nach unserem besten Wissen sind allerdings empirische Befunde darüber, wie Physiklehrkräfte in ihrem Berufsalltag tatsächlich vorgehen, wenn sie den Sprachgebrauch von Schüler_innen im Rahmen einer Klassenarbeit beurteilen, in der fachdidaktische Literatur bislang nur unzureichend vorhanden. In einer explorativen Studie haben wir daher die spezifischen Alltagspraktiken von Physiklehrkräften beim Beurteilen von Schülertexten in einer Klassenarbeit untersucht. Eine zentrale Forschungsfrage, der wir uns in dieser Studie gewidmet haben, lautet:

Wie stark und in welcher Art und Weise berücksichtigen Physiklehrkräfte bei der Genese von Leistungsurteilen die sprachliche Realisierung von Schülertexten?

Methodisches Vorgehen

In einer Vorstudie haben wir zunächst ein gegenstandsgemessenes Erhebungsinstrument für im Schuldienst aktive Physiklehrkräfte entwickelt (Feser et al., 2016). Dieses Erhebungsinstrument besteht aus einer Laut-Denk-Aufgabe, die die Teilnehmer_innen dazu anregt, entsprechend ihrer alltäglichen Berufspraxis beim Korrigieren von Klassenarbeiten zu agieren. Zunächst werden Teilnehmer_innen gebeten, einen Erwartungshorizont zu einer Klassenarbeitsaufgabe so zu erstellen, wie sie dies unter normalen Umständen auch tun würden, und dabei laut zu denken. Mit Hilfe ihres Erwartungshorizonts korrigieren die Lehrkräfte anschließend (ebenfalls laut-denkend) vier unterschiedliche Schülerlösungstexte. Bei diesen 4 Schülertexten (als Text A-D bezeichnet) handelt es sich um Kontrastfälle, die mit Hilfe eines mehrstufigen Expertenratings aus insgesamt 116 authentischen Schülerlösungen so ausgewählt wurden, dass sich diese hinsichtlich ihrer fachlich-konzeptuellen Qualität

und/oder der Qualität ihrer sprachlichen Realisierung deutlich voneinander unterscheiden (für Details siehe Feser & Höttecke, 2017):

- **Text A** besitzt eine **hohe** fachlich-konzeptuelle Qualität und eine **geringe** Qualität hinsichtlich seiner sprachlichen Realisierung.
- **Text B** besitzt eine **geringe** fachlich-konzeptuelle Qualität und eine **hohe** Qualität hinsichtlich seiner sprachlichen Realisierung.
- **Text C** besitzt sowohl eine **hohe** fachlich-konzeptuelle Qualität, als auch eine **hohe** Qualität hinsichtlich seiner sprachlichen Realisierung.
- **Text D** besitzt sowohl eine **geringe** fachlich-konzeptuelle Qualität, als auch eine **geringe** Qualität hinsichtlich seiner sprachlichen Realisierung.

Die Erhebung der Hauptstudie, in der das eben beschriebene Erhebungsinstrument eingesetzt wurde, fand von April bis September 2016 statt. Dabei wurde eine Gelegenheitsstichprobe von $N = 21$ Hamburger Physiklehrkräften gewonnen, die sich deutlich hinsichtlich ihrer Lehrfahrung (2.5 – 37.0 Jahre) unterscheiden und zum Erhebungszeitpunkt an verschiedensten Schultypen tätig waren (Gymnasium = 13, Stadtteilschule = 7, Gym + StS. = 1; staatliche Schule = 19; Privatschule = 2).

Während des lauten Denkens wurden alle Teilnehmer_innen audiographiert. Diese Audiographien wurden regelgeleitet transkribiert, sowie in einzelne mitvokalisierte Gedanken segmentiert (Green & Gilhooly, 2002). Anschließend wurden alle Transkriptsegmente mit einem induktiv am Material entwickelten Kategoriensystems ($\kappa_{\text{Inter}}=.70$; $\kappa_{\text{Intra}}=.87$) codiert (Schreier, 2014), die prozentuellen Codehäufigkeiten je Lehrkraft bestimmt und schließlich mit Hilfe des Wilcoxon-Vorzeichen-Rang-Tests Medianvergleiche dieser Codehäufigkeiten vorgenommen.

Zentrale Befunde

Durch das beschriebene Analyseverfahren konnten in den erhobenen Daten zwei bemerkenswerte Muster identifiziert werden, die auf die Art und Weise schließen lassen, wie die Teilnehmer_innen unserer Studie sprachliche Merkmale der Schülertexte während ihrer laut-denkenden Korrekturarbeit beurteilten.

Muster 1 zeigt sich in den medianen prozentuellen Häufigkeiten, in denen die Teilnehmer_innen sprachliche Merkmale verschiedener Schülertexte beurteilten. Wie in der linken Hälfte von Abbildung 1 dargestellt, zeigt sich ein moderat ausgeprägter Medianunterschied in der Häufigkeit, in der die Teilnehmer_innen sprachliche Merkmale von Schülertext D und A beurteilten und ein stark ausgeprägter Medianunterschied zwischen sprachbezogenen Beurteilungen, die sich auf Schülertext B und C beziehen. Das Besondere an diesem Muster ist, dass sich Schülertext D und A bzw. B und C nicht hinsichtlich der Qualität ihrer sprachlichen Realisierung voneinander unterscheiden. Sie unterscheiden sich hinsichtlich ihrer fachlich-konzeptuellen Qualität. Dementsprechend weisen die eben beschriebenen Medianunterschiede darauf hin, dass die Teilnehmer_innen unserer Studie sprachliche Merkmale eines Schülertextes insbesondere dann beurteilten, wenn ein Text auch ihren fachlich-konzeptuellen Erwartungen entsprach.

Muster 2 hingegen zeigt sich in den medianen prozentuellen Häufigkeiten, in der sich die Teilnehmer_innen zu sprachlichen bzw. fachlich-konzeptuellen Merkmalen der Schülertexte positiv, negativ oder neutral äußerten. Wie aus der rechten Hälfte von Abbildung 1 hervorgeht, äußerten sich die Teilnehmer_innen zu fachlich-konzeptuellen Merkmalen der Schülertexte überwiegend positiv, wohingegen sie sich hinsichtlich sprachlicher Merkmale vor allem negativ äußerten. Hier zeigen sich also eindeutige Hinweise darauf, dass die Teilnehmer_innen ihrer Beurteilung fachlich-konzeptueller Textmerkmale an (mutmaßlich) bereits vorhandenen fachlichen Fähigkeiten und Wissensbeständen der Schüler_innen

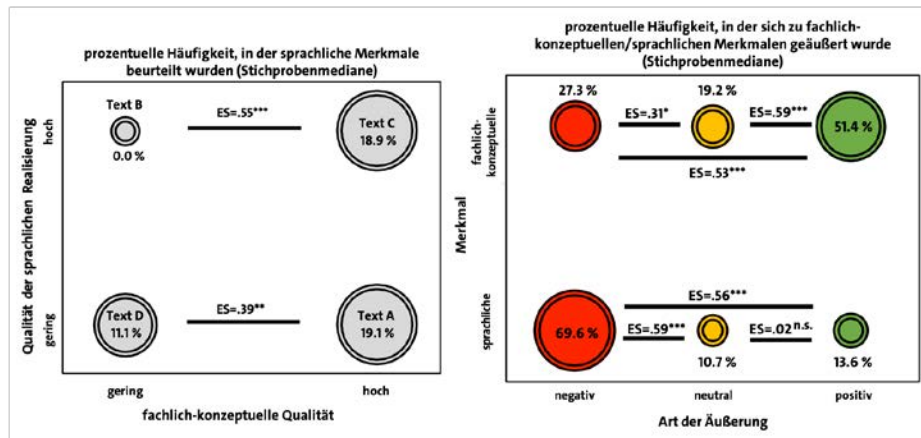


Abb. 1: Zentrale Befunde aus der Analyse der Laut-Denk-Daten (^{n.s.}: $p > .05$; **: $p \leq .01$; ***: $p \leq .001$; Wilcoxon Vorzeichen-Rang-Test; $ES \geq .10$: schwacher Effekt; $ES \geq .30$: moderater Effekt; $ES \geq .50$: starker Effekt).

orientierten, wohingehen sie sprachliche Merkmale der Schülertexte in einer defizitorientierten Art und Weise beurteilten.

Zusammenfassung und Diskussion

Die von uns durchgeführte Studie führte zu dem Befund, dass die teilnehmenden Physiklehrkräfte den Sprachgebrauch von Schüler_innen im Rahmen einer Klassenarbeit in erster Linie dann beurteilten, wenn ein Schülertext auch ihren fachlich-konzeptuellen Erwartungen entsprach. Darüber hinaus zeigte sich, dass die Beurteilung sprachlicher Schülertextmerkmale durch die Teilnehmer_innen unserer Studie defizitorientiert erfolgte. Insgesamt deuten diese Befunde darauf hin, dass die Teilnehmer_innen unserer Studie möglicherweise nicht ausreichend über professionelles Wissen und Können verfügen, um in einer Klassenarbeit den Sprachgebrauch von Schüler_innen ausbalanciert und ressourcenorientiert beurteilen zu können. Fachdidaktische Forschung, die sich mit der Konzeption von evidenzbasierten Aus- und Weiterbildungsmaßnahmen für (angehende) Physiklehrkräfte zum Thema „Leistungsbeurteilung in der Fachunterricht“ befasst, sollten u. E. diese Befunde berücksichtigen.

Unsere explorative Studie hat allerdings eine Vielzahl an Limitationen. Zu nennen sind insbesondere die kleine Stichprobengröße und dass die Teilnehmer_innen unserer Studie lediglich gebeten wurden Schülerlösungen zu einer bestimmten Klassenarbeitsaufgabe zu korrigieren. Aus diesem Grund ist in jedem Fall weitere fachdidaktische Forschung erforderlich, um die Verallgemeinerbarkeit und Übertragbarkeit unserer Befunde zu überprüfen.

Literatur

- Airey, J. (2012). "I don't teach language". The linguistic attitudes of physics lectures in Sweden. *AILA Review*, 25 (1), 64–79.
- Feilke, H. (2012). Schulsprache – Wie Schule Sprache macht. In S. Günther, W. Imo, & D. Meer, J.G. Schneider (Hrsg.), *Kommunikation und Öffentlichkeit. Sprachwissenschaftliche Potenziale zwischen Empirie und Norm* (pp. 151-175). Berlin: De Gruyter.
- Feser, M. S., Höttecke, D., & Ehmke, T. (2016). Testitems zur qualitativen Untersuchung der Ressourcen von Physiklehrkräften beim Bewerten schriftlicher Schülerleistungen. *PhyDid B – Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung*, <http://www.phydid.de/index.php/phydid-b/article/download/677/819> (09.10.2019).
- Feser, M. S. & Höttecke, D. (2017). Klassenarbeiten kriteriengeleitet korrigieren – Wie beurteile ich eine Schülererklärung? *Unterricht Physik*, 158, 15–18.
- Green, C. & Gilhooly, K. (2002). Protocol analysis: practical implementation. In J. T. E. Richardson (Ed.), *Handbook of Qualitative Research Methods for Psychology and the Social Sciences* (reprint, pp. 55–74). Oxford: Blackwell Publishing.
- Höttecke, D., Feser, M. S., Heine, L., & Ehmke, T. (2018). Do Linguistic Features Influence Item Difficulty in Physics Assessments? *Science Education Review Letters*, 2018, 1–6.
- Luykx, A., Lee, O., Hart, J., & Deaktor, R. (2007). Cultural and Home Language Influences on Children's Responses to Science Assessments. *Teachers College Record*, 109 (4), 897–926.
- Lyon, E. G. (2013). What about language while equitably assessing science?: Case studies of preservice teachers' evolving expertise. *Teaching and Teacher Education*, 32, 1–11.
- Rincke, K. & Markic, S. (2018). Sprache und das Lernen von Naturwissenschaften. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 31–48). Berlin: Springer Spektrum.
- Schreier, M. (2014). Quantitative Content Analysis. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Data Analysis* (pp. 170–181). London: Sage Publications Ltd.
- Schleppegrell, M. J. (2004). *The Language of Schooling. A Functional Linguistics Perspective*. New York: Routledge.
- Tajmel, T. (2017). *Naturwissenschaftliche Bildung in der Migrationsgesellschaft. Grundzüge einer Reflexiven Physikdidaktik und kritisch-sprachbewussten Praxis*. Wiesbaden: Springer VS.
- Wellington, J., & Osborne, J. (2001). *Language and Literacy in Science Education*. Buckingham: Open University Press.