

Peter Wulff¹
 David Buschhüter¹
 Anna Nowak¹
 Andreas Borowski¹

¹Universität Potsdam

Computer-basierte Sprachanalyse von Reflexionstexten in der Physik

Die Erfassung und Förderung der Reflexionskompetenz angehender Lehrkräfte stellt ein Kernelement effektiver professioneller Ausbildungsprogramme dar (Darling-Hammond, Hyler & Gardner, 2017; Zeichner, 2010). Dies gilt insbesondere auch in wissenszentrierten Fächern wie der Physik (Hume, 2009; Sorge, Neumann, Neumann, Parchmann & Schwanewedel, 2018). Zur Erfassung und Förderung wurden dabei Prozessmodelle der Reflexionskompetenz vorgeschlagen, die eine strukturierte Reflexion ausmachen und als Vorlage für Studierende dienen (Korthagen, 1999; Lin, Hmelo, Kinzer & Secules, 1999). In narrativ-monologischen Förderansätzen verfassen angehende Lehrkräfte dabei Reflexionsberichte zu ihren Erfahrungen, die dann wiederum als Ausgangspunkt für Diagnostik und Feedback fungieren. Eine systematische und individualisierte Implementation der Erfassung und Förderung der Reflexionskompetenz ist allerdings ressourcenintensiv und in größerem Umfang mit herkömmlichen Mitteln kaum zu bewältigen (Poldner, van der Schaaf, Simons, van Tartwijk & Wijngaards, 2014; Ullmann, 2019).

Computer-basierte Auswertemethoden stellen hierbei eine potentielle Ergänzung zu herkömmlichen Verfahren der Erfassung und Förderung der Reflexionskompetenz dar (Buckingham Shum, Sándor, Goldsmith, Bass & McWilliams, 2017; Zhai, Yin, Pellegrino, Haudek & Shi, 2020). In der vorliegenden Studie werden computer-basierte Auswerteverfahren im Kontext der Erfassung und Förderung der Reflexionskompetenz angehender Physiklehrkräfte vorgestellt, die verschiedene Aspekte der Güte von Reflexionstexten erfassen.

Reflexionstexte in der Lehrkräftebildung

Reflexion im Kontext der Lehrkräftebildung wird als Denkprozess definiert mit dem Ziel eigene Erfahrungen und eigenes Wissen zu restrukturieren und damit seine professionelle Entwicklung zu fördern (Korthagen, 2001; von Aufschnaiter, Fraij & Kost, 2019). Von Aufschnaiter et al. (2019) unterscheiden dabei reflexionsbezogene Dispositionen und reflexionsbezogene Denkprozesse. Dispositionen beziehen sich dabei auf im Individuum angelegte Wissensbestände, Einstellungen oder Werthaltungen zu Reflexion und Denkprozesse auf die situationsspezifische Umsetzung von Aspekten wie der Beschreibung und Deutung von Unterrichtssituationen. Reflexionstexte stellen in diesem Zusammenhang ein Medium dar, in dem reflexionsbezogene Denkprozesse erfasst werden können (Buckingham Shum et al., 2017). Dabei haben Nowak, Kempin, Kulgemeyer und Borowski (2019) ein Modell vorgeschlagen und erste Validierungsstudien durchgeführt (Nowak, Ackermann & Borowski, 2018), um reflexionsbezogene Denkprozesse in Reflexionstexten zu fördern. Sie unterscheiden dazu folgende fünf Elemente einer Reflexion: in den Rahmenbedingungen (1) wird der Kontext spezifiziert (z.B.: Lernziele), in der Beschreibung (2) wird die zu reflektierende Unterrichtssituation detailliert geschildert, in der Bewertung (3) werden Handlungen der Lehrkraft und der Schülerinnen und Schüler evaluiert und gedeutet,

in den Alternativen (4) werden weitere Möglichkeiten der Umsetzung bestimmter Handlungen angegeben und in den Konsequenzen (5) wird die Bedeutung der Erfahrungen dieser Unterrichtssituation für zukünftiges Handeln und die persönliche professionelle Weiterentwicklung reflektiert. Des Weiteren ist im Reflexionsmodell das Professionswissen als wichtige Einflussgröße für gehaltvolle reflexionsbezogene Denkprozesse identifiziert, das in Anlehnung an die Unterteilung nach Shulman (1986) in pädagogisches, fachdidaktisches und fachliches Wissen, unterteilt wird. Aspekte der Güte von Reflexionstexten sind demnach die Vollständigkeit (wurden alle Elemente thematisiert) und die Umsetzung des Professionswissens.

Forschungsfragen: Computer-basierte Auswertung von Reflexionstexten

Zur computer-basierten Erfassung der Vollständigkeit und Umsetzung des Professionswissens können überwachte und unüberwachte Machine-Learning-Algorithmen verwendet werden, die Performanz in einer Aufgabe durch Erfahrung (Daten) optimieren (Mitchell, 1997).

Vollständigkeit der Reflexionstexte (Forschungsfrage 1): Überwachte Machine-Learning-Algorithmen können zur Erfassung der Vollständigkeit (Sind alle Elemente einer Reflexion im Text enthalten?) von Reflexionstexten verwendet werden. Wulff et al. (2020) haben gezeigt, dass Machine-Learning-Algorithmen zur automatisierten Klassifikation der Modellelemente prinzipiell geeignet sind. Die Klassifikationsgenauigkeit lag bei 56%. In der vorliegenden Studie soll in einer Forschungsfrage 1 geprüft werden, inwieweit ein vortrainiertes Sprachmodell, das an die Klassifikationsaufgabe angepasst wird, diese Genauigkeit verbessern kann (Ostendorff et al., 2019).

Umsetzung von Wissen in den Reflexionstexten (Forschungsfrage 2): Die Erfassung der Umsetzung von Professionswissen in Texten erfolgt über unüberwachte Ansätze des Information Retrieval wie word embeddings (Mikolov, Chen, Corrado & Dean, 2013). In Bezug auf Reflexionstexte existieren hier zwar erste lexikon-basierte Ansätze (Ullmann, 2017), allerdings noch nicht auf Basis von word embeddings im Bereich der Physikdidaktik. In einer Forschungsfrage 2 soll deshalb explorativ geprüft werden, inwieweit vortrainierte word embeddings dazu verwendet werden können, um Sätze in Reflexionstexten zu identifizieren, die einen hohen oder geringen Anteil an physikalischem Fachwissen haben.

Methode

Forschungsfrage 1 wird mit einem vortrainierten Deep-Learning-Modell bearbeitet (Devlin, Chang, Lee & Toutanova, 2018). Ziel ist es, Sätze der Reflexionstexte den Elementen des Reflexionsmodells zuzuordnen (Klassifikation). Das vortrainierte Modell berücksichtigt dabei grundlegende Abhängigkeiten und Zusammenhänge in der deutschen Sprache, die nicht mehr neu trainiert werden müssen (Ostendorff et al., 2019). Die Feinabstimmung der Modellparameter auf die Klassifikationsaufgabe erfolgt durch zusätzliche Netzwerk-Ebenen in diesem vortrainierten Modell. Die Regressionsgewichte für diese Ebenen werden durch vorkodierte Daten (etwa 4500 menschlich vorkodierte Segmente zu den Elementen des Reflexionsmodells aus $N=292$ Reflexionstexten) trainiert. Die Überprüfung der Klassifikationsgenauigkeit des Modells erfolgt auf der Basis des F1-Werts, der sowohl falsch positive als auch falsch negative Klassifikationen berücksichtigt. Forschungsfrage 2 wird auf der Basis vortrainierter Wortvektoren der deutschsprachigen Wikipedia bearbeitet (siehe: <https://deepset.ai/german-word-embeddings>). Es werden auf Wortebene

Ähnlichkeitsanalysen zwischen Worten in den Reflexionstexten und einem vordefinierten Lexikon gerechnet. Als Lexikon wurde zur Machbarkeitsanalyse das Stichwortverzeichnis eines Physikdidaktik-Lehrbuches verwendet (Schecker, Wilhelm, Hopf & Duit, 2018). Hier ergaben sich 54 Begriffe, die jeweils physikbezogene Konzepte darstellen (Beispiele: „Stromkreis“, „Druck“).

Ergebnisse

Im Vergleich zu den Machine-Learning-Algorithmen aus Wulff et al. (2020) konnte in der vorliegenden Studie auf Basis vortrainierter Sprachmodelle eine Verbesserung des F1-Wertes auf 78% festgestellt werden (Forschungsfrage 1). Hierbei konnten alle Elemente der Reflexion mit F1-Werten größer 69% identifiziert werden. Dies ist deshalb wichtig, da in der Regel die höherwertigen Elemente wie Alternativen und Konsequenzen weniger Beispieldaten haben und damit schlechter identifiziert werden können (Wulff et al., 2020). In Bezug auf die Identifikation von Sätzen mit einem hohen Fachwissensanteil (Forschungsfrage 2) ergaben sich Unterschiede bei wenig und hoch fachwissenschaftlichen Sätzen in Bezug auf Anteile tatsächlicher fachwissenschaftlicher Begriffe. Der Vergleich eines Satzes mit niedriger Ähnlichkeit zu Fachwissensbegriffen („Meine erste unterrichtete Stunde war die zweite Hälfte eines Physikblocks in einer Klasse.“) und eines Satzes mit hoher Ähnlichkeit („Lernziele: Die SuS können die zeitabhängige Entladespannung eines Kondensators experimentell aufnehmen“) lassen vermuten, dass die Erfassung der Umsetzung des Fachwissens über word embeddings prinzipiell möglich ist.

Diskussion

In der vorliegenden Studie konnte ein vortrainiertes Deep-Learning-Modell zur Klassifikation von Segmenten der Reflexionstexte nach den Elementen einer Reflexion erfolgreich implementiert werden. Die Klassifikationsgenauigkeit konnte bisherige Shallow-Learning-Algorithmen übertreffen. Hier muss einschränkend gesagt werden, dass die Shallow-Learning-Modelle auf kleineren Datensätzen trainiert wurden, sodass die verbesserte Klassifikationsgenauigkeit lediglich als erster Hinweis gedeutet werden darf. Weitere Analysen mit stärkerer Kontrolle der Rahmenbedingungen werden momentan durchgeführt. Alle Elemente einer Reflexion konnten im Deep-Learning-Modell akzeptabel klassifiziert werden. Des Weiteren konnten vortrainierte word embeddings als Methode genutzt werden, um über Ähnlichkeitsanalysen Segmente mit hohem und niedrigem Fachwissensanteil in den Reflexionstexten zu identifizieren. Da es sich hierbei um einen unüberwachten Ansatz handelte, wurden die Ergebnisse zunächst nicht systematisch validiert, sondern lediglich auf Plausibilität überprüft. Eine grundsätzliche Machbarkeit dieser Analysen scheint für die Reflexionstexte gegeben. Weitere Analysen in Bezug auf die systematische Erstellung eines Lexikons an Fachwissensbegriffen sowie der Validierung der Ergebnisse werden momentan durchgeführt. Aufgrund der Bedeutung der Reflexionskompetenz in der universitären Lehrkräftebildung ist eine skalierbare und analytische Auswertung von Reflexionstexten vielversprechend (Poldner et al., 2014; Ullmann, 2019). Wir konnten zeigen, dass überwachte und unüberwachte Machine-Learning-Algorithmen verschiedene Aspekte der Analyse von Reflexionstexten angehender Physiklehrkräfte ausführen können. Diese Algorithmen können nun zur Generierung automatisierten Feedbacks verwendet werden.

Literaturverzeichnis

- Buckingham Shum, S., Sándor, Á., Goldsmith, R., Bass, R. & McWilliams, M. (2017). Towards Reflective Writing Analytics: Rationale, Methodology and Preliminary Results. *Journal of Learning Analytics*, 4(1), 58–84. <https://doi.org/10.18608/jla.2017.41.5>
- Darling-Hammond, L., Hyler, M. E. & Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute.
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*, 1810.04805.
- Hume, A. (2009). Promoting higher levels of reflective writing in student journals. *Higher Education Research & Development*, 28(3), 247–260.
- Korthagen, F. A. (1999). Linking Reflection and Technical Competence: the logbook as an instrument in teacher education. *European Journal of Teacher Education*, 22(2-3), 191–207. <https://doi.org/10.1080/0261976899020191>
- Korthagen, F. A. (2001). *Linking practice and theory. The pedagogy of realistic teacher education*. Mahwah, NJ: Erlbaum. Retrieved from <http://www.loc.gov/catdir/enhancements/fy0634/00057273-d.html>
- Lin, X., Hmelo, C. E., Kinzer, C. & Secules, T. (1999). Designing Technology to Support Reflection. *Educational Technology Research and Development*, 47(3), 43–62.
- Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv*, (1301.3781v3).
- Mitchell, T. (1997). *Machine learning*. New York, NY: McGraw-Hill Education.
- Nowak, A., Ackermann, P. & Borowski, A. (2018). Rahmenthema „Reflexion“ im Praxissemester Physik [Reflection in pre-service physics teacher education]. In A. Borowski, A. Ehlert & H. Precht (Hrsg.), *PSI Potsdam* (S. 217–230). Potsdam: Universitätsverlag Potsdam.
- Nowak, A., Kempin, M., Kulgemeyer, C. & Borowski, A. (2019). Reflexion von Physikunterricht [Reflection of physics lessons]. In C. Maurer (Hrsg.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe* (S. 838). Jahrestagung in Kiel 2018. Regensburg: Gesellschaft für Didaktik der Chemie und Physik.
- Ostendorff, M., Bourgonje, P., Berger, M., Moreno-Schneider, J., Rehm, G. & Gipp, B. (2019). Enriching BERT with Knowledge Graph Embeddings for Document Classification. *arXiv*, (1909.08402v1).
- Poldner, E., van der Schaaf, M., Simons, P. R.-J., van Tartwijk, J. & Wijngaards, G. (2014). Assessing student teachers' reflective writing through quantitative content analysis. *European Journal of Teacher Education*, 37(3), 348–373. <https://doi.org/10.1080/02619768.2014.892479>
- Schecker, H., Wilhelm, T., Hopf, M. & Duit, R. (2018). *Schülervorstellungen und Physikunterricht*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-57270-2>
- Shulman, L. S. (1986). Those Who Understand. Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4–14.
- Sorge, S., Neumann, I., Neumann, K., Parchmann, I. & Schwanewedel, J. (2018). Was ist denn da passiert? *MNU Journal*, 6, 420–426.
- Ullmann, T. D. (2017). Reflective writing analytics. Empirically Determined Keywords of Written Reflection. *ACM International Conference Proceeding Series*, 163–167. LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference. <https://doi.org/10.1145/3027385.3027394>
- Ullmann, T. D. (2019). Automated Analysis of Reflection in Writing: Validating Machine Learning Approaches. *International Journal of Artificial Intelligence in Education*, 29(2), 217–257. <https://doi.org/10.1007/s40593-019-00174-2>
- Von Aufschnaiter, C., Fraij, A. & Kost, D. (2019). Reflexion und Reflexivität in der Lehrerbildung. 144-159 Seiten / Herausforderung Lehrer_innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion, Bd. 2 Nr. 1 (2019): Herausforderung Lehrer_innenbildung - Ausgabe 2. <https://doi.org/10.4119/UNIBI/HLZ-144>
- Wulff, P., Buschhüter, D., Nowak, A., Westphal, A., Becker, L., Robalino, H. et al. (2020). Computer-Based Classification of Preservice Physics Teachers' Written Reflections. *Journal of Science Education and Technology*. <https://doi.org/10.1007/s10956-020-09865-1>
- Zeichner, K. M. (2010). Rethinking the connections between campus courses and field experiences in college- and university-based teacher education. *Journal of Teacher Education*, 61(1-2), 89–99.
- Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C. & Shi, L. (2020). Applying machine learning in science assessment: a systematic review. *Studies in Science Education*, 56(1), 111–151. <https://doi.org/10.1080/03057267.2020.1735757>