

Panel 3: Bewertung von Abschlussarbeiten im Fach Deutsch – zwischen dem Anspruch auf Objektivität und Überzeugungen von Lehrkräften

Moderator:innen: Jörg Jost (Universität Köln), Dorothee Wieser (Techn. Univ. Dresden)

Die schulische Noten- bzw. Zertifikatsvergabe ist nach wie vor von hoher Relevanz für den weiteren Lebensweg von Schüler:innen (z. B. für d. Hochschulzugang). Vor diesem Hintergrund wurden bislang mit Ausnahme von Rheinland-Pfalz in allen Bundesländern (teil-)zentrale Prüfungen zum Erwerb des Mittleren Schulabschlusses und der Allgemeinen Hochschulreife implementiert. Dabei sollen landesweit zentral gestellte schriftliche Prüfungsaufgaben einheitliche Anforderungen an die Prüflinge gewährleisten und die zugehörigen Erwartungshorizonte identische Kriterien für die Leistungsbeurteilung durch die Lehrkräfte vorgeben (Kötter-Mathes 2020). Für die Abschlussprüfungen der Sekundarstufe II nimmt diesbezüglich der Gemeinsame Abituraufgabenpool der Länder eine zentrale Stellung ein.

Dem bildungspolitischen Ziel möglichst objektiver Bewertungen von Abschlussarbeiten steht aber gerade im Fach Deutsch manches entgegen. Zum einen ist die Bewertung von schriftlichen Arbeiten (z. B. Interpretations- oder Erörterungsaufsätzen) insbesondere aufgrund der Verwobenheit von Verstehens- und Darstellungsleistung, vielfältiger Lösungsmöglichkeiten und des meist großen Bewertungsspielraums komplex. Dies stellt sowohl die Entwicklung von Erwartungshorizonten, die objektive Bewertungen unterstützen, als auch die Lehrkräfte in der Anwendung dieser Bewertungsinstrumente vor nicht unerhebliche Herausforderungen.

Zum anderen ist anzunehmen, dass auch individuelle wie kollektiv geteilte Überzeugungen von Lehrkräften darüber, was einen gelungenen Aufsatz im Fach Deutsch ausmacht, die Nutzung der Erwartungshorizonte rahmen. Insbesondere der Abituraufsatz gilt zudem als Ausdruck einer individuellen Leistung, weshalb nicht nur im Kontext der Interpretation literarischer Texte, sondern generell hinsichtlich der Engführung der Bewertung entlang differenzierter und bindender Bewertungskriterien Vorbehalte bestehen können.

Vor diesem Hintergrund möchte das Panel drei empirische Studien diskutieren, die sich der Wirkung unterschiedlicher Varianten von Erwartungshorizonten (Vortrag 1: Pauline Schröter, Lars Hoffmann & Hannelore Söldner), der lehrer:innenseitigen Einschätzung der sprachlichen Komplexität von Abituraufsätzen (Vortrag 2: Anja Riemenschneider) sowie dem Zusammenhang zwischen kontroversen Benotungen von Interpretationsaufsätzen und professionellen Überzeugungen (Vortrag 3: Fabian Wiez) widmen. Die Studien zeichnen sich nicht nur durch unterschiedliche Perspektiven auf das Forschungsfeld aus, sondern auch durch divergierende methodische Zugänge. Ziel der anschließenden Diskussion im Panel soll es sein, die Ergebnisse der Studien aufeinander zu beziehen und weiterführende Forschungsperspektiven für die Konzeption von Erwartungshorizonten und ihre Implementation, aber auch für die Professionalisierung von Lehrkräften zu entwickeln.

Literatur:

Kötter-Mathes, S. (2020). Leistungsbeurteilung in zentralen Prüfungen. Lehrkräftewahrnehmungen der Beurteilungspraxis unter besonderer Berücksichtigung der landesweit vorgegebenen Erwartungshorizonte. Wiesbaden [u. a.]: Springer VS.

Kontakt: joerg.jost@uni-koeln.de, dorothee.wieser@tu-dresden.de

Mit dem Beschluss der Kultusministerkonferenz im Jahr 2012, für die Fächer Deutsch, Mathematik, Englisch und Französisch auf Basis der Bildungsstandards für die Allgemeine Hochschulreife ländergemeinsame Abituraufgabenpools zu entwickeln, hat ein Annäherungsprozess begonnen, der unter anderem zum Ziel hat, die mit den Abiturprüfungen der Länder verbundenen Anforderungen sukzessive anzugleichen. Um die Vergleichbarkeit von Prüfungsergebnissen sicherzustellen, soll neben den Rahmenbedingungen auch die Gestaltung der Prüfungsaufgaben und die Bewertung der Prüfungsarbeiten vereinheitlicht werden. Für das Fach Deutsch gibt es bislang keine ländergemeinsame Regelung zur Bewertung. In einigen Ländern erfolgt die Bewertung von Abiturarbeiten im Sinne eines holistischen Bewertungsverfahrens, in anderen hingegen wird eine analytische Bewertung von Prüfungsarbeiten vorgenommen, bei der verschiedene Teilleistungen jeweils separat beurteilt und gewichtet werden. Zwischen diesen beiden Polen finden sich Mischformen. Alle Bewertungsmodelle unterscheiden sich darin, wie differenziert die Kriterien zur Bewertung ausgewiesen werden sowie ob und in welcher Form Teilleistungen zu gewichten sind.

Bisherige Forschungsergebnisse zum Vergleich von analytischen und holistischen Urteilen lassen sich nur bedingt auf den deutschen Abiturprüfungskontext übertragen. Spezifische Befunde zur Übereinstimmung der Bewertungen von Abiturarbeiten im Fach Deutsch liegen bisher nicht vor. Um die Entwicklung einer gemeinsamen Regelung zur Bewertung von Abiturarbeiten im Fach Deutsch zu unterstützen, wurden vom Institut zur Qualitätsentwicklung im Bildungswesen (IQB) zwei experimentelle Studien durchgeführt. Die erste Studie zielte darauf ab, die Vor- und Nachteile des holistischen sowie des analytischen Bewertungsmodells zu untersuchen und die Frage zu beantworten, welches der beiden Modelle objektivere Bewertungen ermöglicht. Hierfür wurden 33 Deutschlehrkräfte aus zwei Ländern gebeten, Abiturarbeiten aus dem Prüfungsjahr 2017 entweder auf der Grundlage von holistisch oder von analytisch ausgerichteten Erwartungshorizonten (EWH) zu beurteilen. Aufbauend auf den Ergebnissen ging eine zweite Studie der Frage nach, welche Auswirkungen die Vorgabe unterschiedlich differenzierter Kriterien und deren Gewichtung auf die Bewertung von Abiturarbeiten haben. Neben eindeutig holistischen und analytischen EWH wurden hier auch Mischformen untersucht. Dazu wurden zu einer Abituraufgabe fünf verschiedene EWH-Varianten entwickelt, auf deren Grundlage 91 erfahrene Deutschlehrkräfte Abiturarbeiten aus dem Prüfungsjahr 2017 bewerteten. Die Ergebnisse zur Übereinstimmung der Bewertungen sowie zur Einschätzung der EWH-Varianten seitens der Lehrkräfte aus beiden Studien werden vorgestellt und im Hinblick auf Implikationen für die Bildungspolitik, -administration, aber auch für die Lehrkräftebildung diskutiert.

Die Fähigkeit, grammatikalisch und lexikalisch komplexe Sprache zu verstehen und zu produzieren, eröffnet Schüler:innen die Möglichkeit, komplexe Ideen, Argumente und Perspektiven nachzuvollziehen und zu formulieren (McNamara, Crossley, & McCarthy 2010). In nationalen Bildungsstandards nimmt das Konzept der sprachlichen Komplexität deshalb eine zentrale Rolle ein. Auch im Abitur wird von Lehrkräften erwartet, dass sie Aspekte sprachlicher Komplexität in die Bewertung von Schülerleistungen einbeziehen. Es wird jedoch als schwierig eingeschätzt, quantitative Maße sprachlicher Komplexität, z. B. Frequenz- und Längenmaße, effizient einzuschätzen (NGACBP & CCSSO 2010). Zudem ist eine Beeinflussung der Bewertung durch andere Faktoren wie z. B. der Sprachrichtigkeit nicht auszuschließen. In der vorliegenden Studie wurde untersucht, ob und wie Lehrkräfte sprachliche Komplexität in ihre Bewertung von Schülertexten einfließen lassen. Da sprachliche Komplexität ein multidimensionales Konzept ist und keine allgemeingültige Operationalisierung vorliegt, wurde in einem ersten Schritt exploratorisch vorgegangen und die sprachliche Komplexität von 344 originalen Abituarbeiten im Fach Deutsch mithilfe computerlinguistischer Maße analysiert. Auf dieser empirischen Grundlage wurden für den gegebenen Kontext relevante Maße sprachlicher Komplexität identifiziert und eine Teilstichprobe von Schülertexten ausgewählt, die sich in ihrer sprachlichen Komplexität unterscheiden. Diese Teilstichprobe (n = 16) wurde von 33 erfahrenen Lehrkräften neu bewertet (inhaltliche Leistung, sprachliche Leistung und Gesamtleistung). Zusätzlich nahmen die Lehrkräfte eine Einschätzung der sprachlichen Komplexität der Schülertexte anhand von ausgewählten Aspekten vor. Die Zusammenhänge der computerlinguistischen Messung sprachlicher Komplexität, der Bewertung durch die Lehrkräfte sowie ihrer Einschätzung der sprachlichen Komplexität wurden mit Mehrebenenmodellen analysiert. Die Ergebnisse sprechen dafür, dass sowohl die Bewertung der sprachlichen Leistung als auch die Einschätzung der Komplexität der Schülertexte durch die Lehrkräfte mit der computerlinguistischen Messung sprachlicher Komplexität zusammenhängen, vor allem mit Maßen syntaktischer Komplexität und des Sprachgebrauchs (Wortfrequenzmaße). Die Untersuchung deutet aber auch darauf hin, dass Lehrkräfte bei der Bewertung Schwierigkeiten haben, Aspekte von sprachlicher Komplexität, sprachlicher Richtigkeit und Inhalt zu trennen.

Literatur:

NGACBP & CCSSO (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. Appendix A. Washington, DC: National Governors Association Center for Best Practices (NGACBP), Council of Chief State School Officers (CCSSO).

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86.

Kontakt: anja.riemenschneider@iqb.hu-berlin.de

Der Vortrag soll erste Ergebnisse des Promotionsprojekts *Überzeugungen von Deutschlehrkräften und ihr Einfluss auf Beurteilungen von Interpretationen literarischer Texte* präsentieren. Ziel des Projekts ist es, im Rahmen einer qualitativen Untersuchung unterschiedliche Überzeugungsstrukturen von Deutschlehrkräften festzustellen, um dann zu prüfen, ob sich ein Einfluss dieser Unterschiede bei der Beurteilung von Interpretationsaufsätzen zu einem literarischen Text feststellen lässt. Aufgrund ihrer Filter-, Rahmungs- und Steuerungsfunktion liegt die Vermutung nahe, dass es die Überzeugungen der beurteilenden Lehrkräfte sind, die zu diesen divergierenden Benotungen beitragen (vgl. Fives/Buehl 2012).

Das Projekt arbeitet dabei mit einer Stichprobe von 14 Lehrerinnen und Lehrern, die alle Deutsch in der Oberstufe unterrichten, sich aber sonst durch unterschiedliche Hintergründe auszeichnen. In einer ersten Erhebungsphase, in der die Teilnehmer:innen zwei ausgewählte Schülertexte beurteilen sollten, wurden Beurteilungsdaten (Randkommentare, Beurteilungskommentar, Benotung in Notenpunkten) sowie von den Teilnehmenden erstellte Erwartungshorizonte zur bearbeiteten Aufgabenstellung erhoben. In einer zweiten Erhebungsphase wurden mit allen Beteiligten Interviews geführt, die Aufschluss über den breiteren Überzeugungskontext bieten sollten, vor dem die Beurteilungen der ersten Phase erfolgt waren.

Bei der Auswertung der Daten wurde im Rahmen einer qualitativen Inhaltsanalyse nach Kuckartz (2018) und Mayring (2015) insbesondere der Frage nachgegangen, ob sich Zusammenhänge zwischen unterschiedlichen Überzeugungen und auseinandergehenden Benotungen der Schülertexte ausmachen lassen. Dabei konnte wie in anderen Studien eine breite Streuung der vergebenen Notenpunkte festgestellt werden.

Das qualitative Forschungsprojekt hat einen explorativen Charakter und soll einen Einblick in die verschiedenen Überzeugungen von Deutschlehrerinnen und Deutschlehrern ermöglichen, die diese bei der Beurteilung von Interpretationen literarischer Texte lenken. Exemplarisch deutet sich etwa an, dass verschiedene Lehrkräfte von ihren Schülerinnen und Schülern unterschiedliche Schreibmodi bzw. -strategien erwarten, wenn ein literarischer Text schriftlich interpretiert werden soll. Manche sehen den Interpretationstext als sekundären Beleg einer im Voraus im Kopf des Lernenden vollzogenen Deutung, andere erwarten dagegen, dass die Auseinandersetzung mit der Textgrundlage erst beim Schreiben der Interpretation und die Gesamtdeutung erst am Ende derselben erfolgt. Inwiefern Unterschiede wie dieser in unterschiedlichen Erwartungen an Schülertexte resultieren, soll das Projekt und soll der Vortrag zeigen.

Perspektivisch soll das Projekt Hinweise für die Lehrerbildung, die Konzeption und Auswertung zentraler Leistungsüberprüfungen und das Nachdenken über Interpretationen und ihre Beurteilung durch einzelne Lehrerinnen und Lehrer bieten.

Literatur:

Fives, Helenrose; Buehl, Michelle M. (2012). Spring cleaning for the "messy" construct of teachers' beliefs. What are they? Which have been examined? What can they tell us? In: APA educational psychology handbook, Vol. 2: Individual differences and cultural and contextual factors, 471–499.

Kuckartz, Udo (2018). Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung (4. Aufl.). Weinheim: Beltz-Juventa.

Mayring, Philipp (2015). Qualitative Inhaltsanalyse. Grundlagen und Techniken (12. Aufl.). Weinheim: Beltz.

Kontakt: fabian.wiez@uni-hamburg.de