

Carolin Führer, Marco Magirus, Thorsten Bohl, Bernd-Stefan Grewe,
Wolfgang Polleichtner, Fahimah Ulfat (Hrsg.)

Relativität und Normativität von Beurteilungen

Fachübergreifende und fachspezifische Analysen

SCHRIFTENREIHE NR. 2 DER TÜBINGEN SCHOOL OF EDUCATION BAND 03

LEHRERBILDUNG. KREATIV. INNOVATIV. UNGEWÖHNLICH.
HERAUSGEGEBEN VON THORSTEN BOHL, NINA BECK, SIBYLLE MEISSNER

Relativität und Normativität von Beurteilungen

Carolin Führer, Marco Magirius, Thorsten Bohl, Bernd-Stefan Grewe,
Wolfgang Polleichtner, Fahimah Ulfat (Hrsg.)

Relativität und Normativität von Beurteilungen



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie, detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung – Nicht kommerziell – Keine Bearbeitungen 4.0 International Lizenz. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode> oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.

Die Online-Version dieser Publikation ist auf dem Repositorium der Universität Tübingen frei verfügbar (Open Access):

<http://hdl.handle.net/10900/144580>
<http://nbn-resolving.de/urn:nbn:de:bsz:21-dspace-1445809>
<http://dx.doi.org/10.15496/publikation-85924>

Tübingen Library Publishing 2023
Universitätsbibliothek Tübingen
Wilhelmstraße 32
72074 Tübingen
druckdienste@ub.uni-tuebingen.de
<https://tlp.uni-tuebingen.de>

ISBN (PDF): 978-3-946552-91-8

Layout und Umschlaggestaltung: Gabriele von Briel, Tübingen School of Education (TüSE)
Satz: Cornelia True, Universitätsbibliothek Tübingen
Druck und Bindung: Druckhaus Sportflieger in der Medialis Offset GmbH
Printed in Germany

Inhaltsverzeichnis

Carolin Führer	
Einleitung: Relativität und Normativität von Beurteilungen.....	9
Samuel Merk und Sarah Bez	
Relativität und Normativität in der unterrichtspraktischen und psychometrischen Leistungsbewertung	13
Georg Breidenstein	
Zensuren: Risiken und Nebenwirkungen	23
Benjamin Fauth	
Bewertungen von Lehrenden und Lernenden im Unterricht: Evaluation und Feedback... ..	31
Silvia-Iris Beutel	
Alternative Leistungsbeurteilung als Entwicklungsaufgabe von Schule	37
Philippe Wampfler und Christian Albrecht	
Zugänge zu einer zeitgemäßen Prüfungskultur	45
Jan-Philipp Burde und Walther Paravicini	
Mehr als Fakten und Formeln – Normativität im Kontext der Fächer Mathematik und Physik.....	53
Wolfgang Polleichtner	
Graecum und Latinum. Zwischen normativem Anspruch und relativer Ungleichbehandlung	61
Stefan Hofstetter	
Zur Bewertung mündlicher Leistungen (nicht nur) im Fremdsprachenunterricht.....	69

Marco Magirus und Fabian Wiez	
Zum Bewerten von Interpretationen im Deutschabitur	77
Helga Gese	
Form oder Funktion? Zur Begründung von Korrektornormen für grammatische Leistungen.....	85
Alexander Schneider	
Lernen und Bewerten im Kunstunterricht der Sekundarstufe I.....	93
Friedrich Schweitzer und Fahimah Ulfat	
Relativität und Normativität von Leistung und ihrer Beurteilung im Religionsunterricht.....	101
Christian Fischer und Johanna Grad	
Wege zu neuen Bewertungssystemen in der Hochschullehre: Gameful Learning und Pädagogische Psychologie.....	107
Jessica Heesen	
Informationen aus dem Internet beurteilen	113
Abbildungsverzeichnis	118
Tabellenverzeichnis	119

Carolin Führer

Einleitung: Relativität und Normativität von Beurteilungen

„Am Ende steht eine Note ...! – wie relativ diese Note ist, bzw. dass Zensuren angesichts der Vielfalt von Faktoren in der Leistungsmessung niemals angemessen der individuellen Anstrengung, dem Lernfortschritt oder dem Persönlichkeits- und Begabungsspektrum von Lernenden gerecht werden, ist spätestens seit Beginn der modernen bürgerlichen Gesellschaft selbstverständlich geworden.

„Leistung‘ [...] fungiert [dennoch, CF] als Bindeglied zwischen freiheitlicher Aufklärung und kapitalistisch orientierter Wirtschaft und ‚adelt‘ die ‚Arbeitsgesellschaft‘; nichts scheint daher naheliegender, als auch in ihr einen Grundbegriff der Moderne zu sehen.“ (Ricken & Reh, 2018, S. 3)

Ein Symbol für die Funktionalität dieser Leistungsorientierung ist die Einführung von Beurteilung und Bewertung in das Schulwesen (Reh et al., 2021) und deren langanhaltende Bedeutung und Wirkung bis in die Gegenwart hinein. Aber so wie Leistung sozialer Konstruktion bedarf und nicht nur historisch betrachtet ein systematisch unscharfer Begriff ist (Verheyen, 2018), gilt dies eben auch für Beurteilungs- und Bewertungsprozesse. Der politische Aufwind standardisierter Leistungsmessung und -evaluation im Bildungswesen kann darüber nicht hinwegtäuschen, ist vielmehr Ausdruck am Festhalten an diesen ideellen Grundlagen unserer Gesellschaft.

„Die überwiegende Mehrheit der empirisch-quantitativen Arbeiten nimmt an, mit der ‚Leistung‘ eine latente Variable – also eine Variable, die nicht direkt, sondern nur mittelbar über Indikatoren gemessen werden kann (Borsboom, 2008) – zu messen (Edelsbrunner & Dablander, 2019).“ (Merk & Bez in diesem Band, S. 17)

Internationale wie nationale Studien belegen die starke Relativität von Leistung, beispielhaft und wirkmächtig steht dafür die Reproduktion der soziale Ordnungsmacht im deutschen Bildungssystem (Dietrich & Fricke, 2013).

Dennoch werden Kinder in Deutschland (zu) früh mit den Normativitäten (und Normierungen) schulischer Beurteilung vertraut gemacht (Beutel, 2005). Die ‚Tradition‘ der Kritik daran findet bildungspolitisch immer wieder Niederschlag in Schulversuchen, in denen verstärkt alternative und formative Beurteilungen oder Bewertungen mit individueller Bezugsnorm erprobt werden. Die forcierte Konzentration der Beurteilung auf individualisierte Leistungen kann dabei nicht nur für das Individuum (und seine Psyche) sowie das Schulwesen und sein Versprechen der

Anerkennung von Diversität, Inklusion und sozialer Gerechtigkeit zur (Er-)Lösung und gleichzeitig zum Problem werden. Dabei geht es nicht nur um Fragen der Umsetzbarkeit und das Denken und Handeln von Lehrkräften (Greiten, 2017) im Zusammenspiel mit der Standardisierung im Bildungswesen. Vielmehr hat ein Aufweichen der Leistungsnorm Folgen für die Demokratie (Sandel, 2020) und die Generierung und Konsolidierung von Wissen in der Öffentlichkeit. Denn Beurteilen ist auch ein Modus des Umgangs mit Wissen, aus welchem sowohl Schulfächer wie universitäre Disziplinen historisch ihre Legitimität ziehen (vgl. Reh & Caruso, 2020, S. 616). Werden diese Wissenspraktiken verändert oder gar aufgelöst, erzeugt das auch eine neue epistemische Relativität oder gar Transformation in den „Modi der Weltbegegnung“ (Baumert, 2002).

Mit dem Festhalten an einer meritokratischen Gesellschaft („Meritokratie als Selbstbetrug“, Schimank, 2018, S. 24) bedarf es angesichts der Vielfalt der Relativitäten im Bewerten und Beurteilen konsensfähiger Normen. Normen der schulischen Beurteilung werden derzeit hergestellt durch eine Orientierung an relativ objektivierbaren, normativ festgelegten (Qualitäts-) Maßstäben. Die Diskussion der Bezugsnormen ist hierbei maßgeblich geworden (Heckhausen, 1974; Rheinberg, 1980), jedoch sowohl empirisch als auch unterrichtspraktisch zu relativieren. Im schulischen Rahmen orientieren sich Beurteilungen zunächst an institutionellen Vorgaben (Bildungsstandards, Bildungspläne etc.), deren Stellenwert im Kontext der Zentralisierung von Abschlussprüfungen in den letzten Jahren noch an Bedeutung gewonnen haben dürfte (zum Abitur vgl. Hoffmann et al., 2022). Zugleich werden Leistungsbeurteilungen durch ein langanhaltendes pädagogisches und didaktisches Brauchtum bestimmt. Der Abituraufsatz im Fach Deutsch ist ein bedeutsames Beispiel dafür, wie Bewertung und Beurteilungen wirkmächtigen Traditionsbildungen unterliegen, die mit den gesetzlichen Rahmen eine eigene Interaktion eingehen (Eiben-Zach, 2022).

In der vorliegenden Herausgeberschrift war es uns ein Anliegen, das Spannungsfeld von Normativität und Relativität von Beurteilungen nicht nur auf der Ebene dieser bestehenden Praxis(-normen) transparent zu machen. Vielmehr ging es darum, Normen der Beurteilung in den fachlichen und fachdidaktischen Dimensionen der einzelnen Unterrichtsfächer als auch bildungs- und erziehungswissenschaftlichen Perspektiven zu diskutieren. In erziehungs- und bildungswissenschaftlicher Perspektive kommen bezüglich der Normsetzungen des (schulischen) ‚Leistungsparadigmas‘ (Ricken & Reh, 2018) deren Funktionen, Maßstäbe und Formen ebenso in den Blick wie die unterschiedlichen Akteure des Beurteilens und Bewertens. In fachdidaktischer Perspektive wird die Relativität von Beurteilung in je unterschiedlicher fachlicher Ausformung thematisch, das kann Wissenskonstruktionen wie die der Sprachnorm im Fach Deutsch ebenso betreffen wie Fragen

des Bildungsziels, z. B. bei Interpretation, aber auch der Entwicklung von Wert- oder gar Geschmacksurteilen wie in Philosophie und Kunst. Scheinbar alle Fächer kennzeichnet dabei in der Tendenz eine Spannung in den Bezugsnormen. Diese werden verstärkt durch Antinomien in der pädagogischen, fachlichen und fachdidaktischen Normenbildung zum Beurteilen. Im Verbund mit den unterrichtlichen Ungewissheiten und der Kontingenz pädagogischen Handelns stehen diese systemischen Anforderungen der Normierung und Standardisierung (und ihrer Umsetzung in formativen Assessments etc.) dabei potenziell ebenso entgegen wie die Notwendigkeiten individueller Orientierungsfunktion und individualisierter Bewertung. Dies alles wirft komplexe aufgaben- und prüfungstheoretische wie -praktische Fragen der Objektivierbarkeit, Vergleichbarkeit und Umsetzbarkeit von Bewertungen auf, wie sie dann insbesondere im Zusammenhang der Selektion und Vergabe von Berechtigungen an Brisanz gewinnen. Schon allein deshalb täte es der wissenschaftlichen Auseinandersetzung gut, den Diskurs über Bewerten und Beurteilen fortlaufend wach zu halten. Dass die wissenschaftliche Auseinandersetzung mit der Frage „Wie sich Leistung bewerten lässt“ weit hinter anderen, scheinbar relevanteren Themen zurücksteht, hat die unter diesem Titel im Wintersemester 2021/22 durchgeführte Ringvorlesung an der Universität Tübingen offengelegt. In diesem Sinne sind die in diesem Umfeld entstandenen und hier versammelten Beiträge auch als interdisziplinärer Denkanstoß und Anregung zu verstehen. Denn jenseits öffentlichkeitswirksamer Diskurse über Noten-inflation bleibt zu fragen, wie Bewerten und Beurteilen fachlich, pädagogisch und fachdidaktisch den Anforderungen einer zukunftsfähigen Bildung gerecht werden kann, die die sich rasant verändernden außerschulischen Kontexte berücksichtigt. Aus diesem Grunde haben wir exemplarisch zwei Beiträge aufgenommen, die (zukünftige) Hochschulbewertung und die Beurteilung von Sachinformationen in einer (post-) digitalen Welt als Grundlage jeder Bewertung in den Blick nehmen. So ist die Frage der fachlichen Gegenstandskonstitution und -erkenntnis (und ihrer Beurteilung), die im Zuge der Kompetenzorientierung in den Hintergrund trat, neu zu stellen: Universale Verfügbarkeit von Fachgegenständen, -problemen und Lösungswegen – bei gleichzeitiger Pluralität der Antworten im Netz – fordert Formen schulischer Normierung und Orientierung insbesondere in fachlichen Beurteilungsprozessen heraus, die dieser (post-)digitalen Wissensordnung und -aneignung auch entsprechen.

Im Namen der Mitherausgeber*innen danke ich den Autor*innen für ihre Beiträge, die ihnen durch das Format und die Adressatengruppe besondere Offenheit und Mut zur Pointierung abverlangten. Ein Dank geht auch an Dr. Britta Eiben-Zach, die die Ringvorlesung wesentlich mitorganisiert und begleitet hat. Den Mitgliedern der Tübinger Special Interest Group „Relativität, Normativität und Orientierung in der fachlichen Bildung“ der Tübingen School of Education sei für die gemeinsamen Vorarbeiten (Führer et al., 2022) und den Austausch gedankt, ohne die dieser Band nicht zustande gekommen wäre.

Literatur

- Baumert, J. (2002). Deutschland im internationalen Bildungsvergleich. In: N. Killius, J. Kluge & L. Reisch (Hrsg.), *Die Zukunft der Bildung*. Frankfurt/M.: Suhrkamp.
- Beutel, S.-I. (2005). Zeugnisse aus Kindersicht. Beiträge zur Professionalisierung der Leistungsbeurteilung, *Schriftenreihe der Max-Traeger-Stiftung, Bd. 42*. Weinheim & München: Juventa.
- Heckhausen, H. (1974). *Leistung und Chancengleichheit*. Göttingen: Hogrefe.
- Eiben-Zach, B. (2022). *Literatur als Gegenstand Westberliner Abituraufsätze der 1950er und 1960er Jahre. Literaturkonzepte und Interpretationen, Aufgabentypen und Aufsatzarten*. Bern: Peter Lang.
- Dietrich, F. & Fricke, U. (2013). Reproduktion sozialer Ungleichheit im Prozess schulischer Leistungsbewertung. In F. Dietrich, M. Heinrich & N. Thieme (Hrsg.), *Bildungsgerechtigkeit jenseits von Chancengleichheit: Theoretische und empirische Ergänzungen und Alternativen zu "PISA"* (S. 259-292). Wiesbaden: Springer VS.
- Führer, C., Schweitzer, F., Tesch, B., Eiben-Zach, B., Ulfat, F., Thomas, P., Pollechner, W., Grewe, B. & U. Küchler (Hrsg.) (2022). *Relativität und Bildung. Fachübergreifende Herausforderungen und fachspezifische Grenzen*. Münster: Waxmann.
- Schäfer, A. & Thompson, C. (Hrsg.) (2015). *Leistung. Reihe: Pädagogik – Perspektiven*. Paderborn: Schöningh.
- Greiten, S. (2017). "Leistung" wird relativ. Konzeptveränderungen von Sonder- und Regelschullehrkräften durch Unterricht mit integrativen Lerngruppen. In: A. Textor, S. Grüter, I. Schiermeyer-Reichl & Streese, B. (Hrsg.), *Leistung inklusive? Inklusion in der Leistungsgesellschaft. 2. Unterricht, Leistungsbewertung und Schulentwicklung* (S. 157-164). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Reh, S. & Ricken, N. (Hrsg.) (2018). *Leistung. Zur Entstehung und Transformation eines pädagogischen Konzepts*. Berlin: Springer.
- Reh, S., Bühler, P., Hofmann, M. & Moser, V. (Hrsg.) (2021). *Schülerauslese, schulische Beurteilung und Schülertests 1880-1980*. Bad Heilbrunn: Klinkhardt. (Bildungsgeschichte. Forschung - Akzente - Perspektiven).
- Schimank, U. (2018). Leistung und Meritokratie in der Moderne. In: S. Reh & N. Ricken (Hrsg.), *Leistung als Paradigma* (S. 19-42). Wiesbaden: Springer VS. Wiesbaden.
- Hoffmann, L., Schröter, P., Groß, A., Schmid-Kühn, S. M. & P. Stanat (Hrsg.) (2022). *Das unvergleichliche Abitur. Entwicklungen – Herausforderungen – Empirische Analysen*. Bielefeld: wbv.
- Sandel, M. J. (2020). *Vom Ende des Gemeinwohls. Wie die Leistungsgesellschaft unsere Demokratien zerreit*. Frankfurt a.M.: S. Fischer.
- Verheyen, N. (2018). *Die Erfindung der Leistung*. Berlin: Hanser.
- Weinert, F. E. (Hrsg.) (2014). *Leistungsmessungen in Schulen*. Weinheim: Beltz.

Angaben zur Autorin

- Prof. Dr. Carolin Führer, Professorin für Deutsche Philologie/ Didaktik der deutschen Literatur, Universität Tübingen, Deutsches Seminar, carolin.fuehrer@uni-tuebingen.de

Samuel Merk und Sarah Bez

Relativität und Normativität in der unterrichtspraktischen und psychometrischen Leistungsbewertung

1. Einleitung

Der vorliegende Beitrag führt in die häufig zitierte Unterscheidung von Bezugsnormen ein, um anhand dieser zu diskutieren, welche Rolle Relativität und Normativität bei der Leistungsbewertung in Bildungspraxis und -forschung spielen. Die zentrale These lautet dabei, dass sowohl die (empirisch quantitative) Bildungsforschung als auch die Bildungspraxis Bewertungen primär durch Relativierung einzelner Messwerte an einer Verteilung dieser Messwerte konstruieren, wobei Normativität bei der Wahl der betrachteten Größe (z. B. intraindividuelle Leistungszuwachs; Abstand von einem Referenzwert) sowie der Transformation von Prozenträngen in Noten oder andere Bewertungsskalen (z. B. Notenpunkte) erzeugt wird.

2. Bezugsnormen

2.1. Einführendes Beispiel

Angenommen, eine Grundschullehrkraft führt in ihrer Klasse alle drei Wochen einen Test zum sinnkonstruierenden Satzlesen durch, in dem Schüler*innen unvollständige Sätze (z. B. ‚Mama holt meinen ab‘) durch die Auswahl eines Wortes aus einer Liste (z. B. ‚Bruder – Kopf – Sinn – Schatten‘) sinnvoll ergänzen müssen. Die Schüler*innen haben drei Minuten Zeit, um möglichst viele (der gleich schwierigen) Sätze zu bearbeiten (Speedtest; Rost, 2004). Abschließend wird gezählt, wie viele der Aufgaben korrekt bearbeitet wurden. Angenommen, der Lehrkraft liegen die folgenden Ergebnisse ihrer Klasse vor (siehe Tab. 1). Wie sind die Ergebnisse des letzten Tests zu bewerten? Was macht eine ‚gute‘ Leistung in diesem Fall aus?

Legt man Lehrkräften eine solche Aufgabe vor, sind vor allem drei Argumentationsstränge zu beobachten (Rheinberg, 1980): Zum einen wird argumentiert, dass Schüler*in C eine sehr gute Leistung aufweist, da Schüler*in C in jedem Test jeweils mehr Punkte erzielt hat als alle Mitschüler*innen. Zum anderen wird Schüler*in B eine sehr gute Leistung attestiert, da Schüler*in B den größten Leistungszuwachs zeigt. Schließlich ist auch noch zu beobachten, dass das

Erzielen von z. B. mindestens zehn korrekten Sätzen als eine gute Leistung bewertet wird, da dies z. B. einem im Bildungsplan verankerten Ziel entspräche.

Schüler*in	Woche				
	1	4	7	10	13
A	5	7	7	9	10
B	2	4	4	7	11
C	10	9	11	12	13
D	4	5	6	7	6
E	1	2	1	3	3
F	4	3	4	5	8

Tab. 1. Fiktive Ergebnisse eines Leseverständnistests. W1, W4, etc. entsprechen Woche 1, Woche 4, etc.

2.2. Kriteriale, individuelle und soziale Bezugsnorm

Diese drei Argumentationsstränge werden in den Bildungswissenschaften Bezugsnormen (engl. *reference norms*) genannt. Darunter versteht man den Standard, der zur Bewertung einer Leistung herangezogen wird (Heckhausen, 1974). Die ‚soziale Bezugsnorm‘ liegt vor, wenn das zu bewertende Ergebnis mit den Ergebnissen einer Vergleichsgruppe, die denselben Test durchlaufen hat, abgeglichen wird. Die ‚kriteriale Bezugsnorm‘ (auch ‚sachliche Bezugsnorm‘; Rheinberg & Fries, 2010) legt a priori fest, was eine ‚gute‘ Leistung auszeichnet. Ein typisches Beispiel aus dem Schulalltag sind etwa Wertungstabellen des Sportunterrichts. Dort ist bspw. festgelegt, dass eine Oberstufenschülerin 14 Notenpunkte für 200 Meter Rückenschwimmen bekommt, wenn sie weniger als 223,5 Sekunden aber mehr als 217,7 Sekunden dafür benötigt (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2023). Schließlich wird in der Literatur noch die ‚individuelle Bezugsnorm‘ (auch ‚temporale‘, ‚intraindividuelle‘ oder ‚ipsative Bezugsnorm‘) genannt. Sie vergleicht (wie die soziale Bezugsnorm) zur Bewertung einer Leistung diese ebenfalls mit anderen realen (empirisch gemessenen) Leistungen. Allerdings stammen jene Leistungen vom selben Merkmalsträger (siehe Abb. 1).

2.3. Die Relativität und Normativität der Bezugsnormen

Die drei Bezugsnormen unterscheiden sich – wie oben angeführt – im Standard, auf den sie sich zur Bewertung einer einzelnen Leistung beziehen. Damit scheint der Wahl der Bezugsnorm ein stark normatives Moment innezuwohnen,

definiert sie doch einen jeweils unterschiedlichen Idealzustand. Jedoch erfolgt in allen Bezugsnormen die Bewertung anhand von Abständen auf der Leistungsvariablen (siehe Abb. 1).

Im Folgenden wird daher dafür argumentiert, dass alle drei Bezugsnormen Relativierungen darstellen, die sich von der sozialen Bezugsnorm lediglich um eine Translation (kriteriale Bezugsnorm) oder bezüglich der zugrunde gelegten Variablen unterscheiden (individuelle Bezugsnorm). Dies soll am Beispiel aus Tab. 1 für alle drei Bezugsnormen illustriert werden.

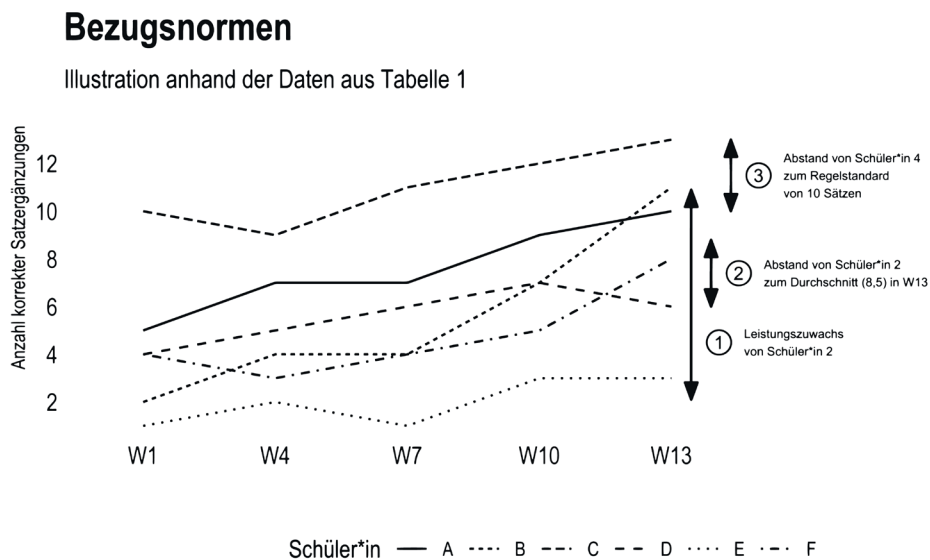


Abb. 1. Veranschaulichung der Bezugsnormen. Die Datengrundlage bildet Tab. 1.

Legt sich die Grundschullehrkraft aus dem einführenden Beispiel normativ auf die individuelle Bezugsnorm fest, ergibt die Relativierung des letzten Messwertes von Schüler*in C am ersten Messwert eine Steigerung um 3 Sätze, wohingegen sich Schüler*in F um 4 Sätze gesteigert hat. Die Lehrkraft muss nun nicht nur normativ entscheiden, welche dieser Leistungssteigerungen einer ‚wie guten‘ Leistung entspricht. Es stellt sich darüber hinaus die Frage, wie differenziert diese Transformation sein soll. Macht das Ausgangsniveau der Leistung einen Unterschied? Oder ob Deutsch die Zweit- oder Erstsprache der jeweiligen Schüler*innen ist? Es ist nicht plausibel anzunehmen (wenngleich nach unserem Wissen nicht empirisch belegt), dass Lehrende diese Fragen via a priori abgeleiteten Prinzipien beantworten. Vielmehr ist zu erwarten, dass bei dieser Bewertung die Leistungssteigerung

eines Individuums am Erfahrungshorizont von bisher wahrgenommenen Leistungssteigerungen relativiert wird. Danach hätte also ein*e Schüler*in unter Annahme der individuellen Bezugsnorm eine ‚gute Leistung‘ erbracht, wenn die Leistungssteigerung im Vergleich zu Leistungssteigerungen implizit assoziierter anderer überdurchschnittlich ist. Damit erfolgt auch unter Nutzung der individuellen Bezugsnorm die Leistungsbewertung durch soziale Relativierung.

Die Nutzung einer kriterialen Bezugsnorm scheint zunächst frei von einer solchen interindividuellen Relativierung, da es sich ja um eine Idealnorm handelt (Klauer et al., 1972) und bei dieser die Transformation des Messwerts zur Bewertung (Abstand Messwert vs. Kriterium; siehe Pfeil 3 in Abb. 1) „in der Sache selbst liegt“ (Rheinberg & Fries, 2010, S. 61). Dass dies jedoch praktisch kaum der Fall sein kann, wird sehr schnell plausibel: Angenommen, eine Lehrkraft weiß qualitativ alles über den Lerngegenstand – also in unserem Beispiel über den Schriftspracherwerb (z. B. diverse Stufenmodelle, effektive Unterrichtspraktiken, typische Lernschwierigkeiten, etc.) – aber nichts über typische quantitative Resultate: Sie würde kaum ableiten und begründen können, warum welche Anzahl an richtigen Sätzen als ‚gute Leistung‘ bewertet werden sollten. Vielmehr ist plausibel, dass bei der Festlegung der Schwellenwerte Wissen über die Resultate (wer macht mit welchen Ergebnissen typischerweise guten Fortschritt in der nächsten Klassenstufe etc.) eine zentrale Rolle spielen. Damit definiert wiederum eine soziale Relativierung, was eine ‚gute Leistung‘ ausmacht. Im Unterschied zur sozialen Bezugsnorm stammen die Resultate, anhand derer die Relativierung stattfindet, jedoch auch von Merkmalsträgern außerhalb der zu bewertenden Schüler*innen. Einschränkend muss an dieser Stelle allerdings ergänzt werden, dass nach unserem Wissen kaum tragfähige empirische Untersuchungen bei Lehrpersonen vorliegen, die der Frage der sozialen Relativierung bei der Setzung von Referenzwerten für kriteriale Bezugsnormen nachgehen und die hier aufgestellten Hypothesen verifizieren oder falsifizieren könnten.

3. Relativität und Normativität in der unterrichtspraktischen Leistungsbewertung

In der unterrichtlichen Praxis der (summativen) Leistungsbewertung findet man kaum die ausschließliche Anwendung genau einer Bezugsnorm. Vielmehr herrscht Einigkeit in der Literatur, dass in der unterrichtlichen Praxis die Anwendung mehrerer Bezugsnormen bei der Messung und Bewertung von Leistungen weit verbreitet ist (Bohl, 2004; Lintorf & Buch, 2021; Pant, 2020), was sich erstaunlicherweise mit vielen Abschnitten aus Verordnungen und Gesetzen der Bundesländer zur Notenvergabe (Pant, 2020) deckt.

Wie schnell eine Vermischung von Bezugsnormen in der unterrichtlichen Praxis erfolgen kann, soll folgendes Beispiel illustrieren: Eine Lehrkraft unterrichtet eine Einheit und schließt diese mit einer summativen Leistungserfassung (z. B. einer Klassenarbeit) ab. Die Unterrichtseinheit deckt sich im Großen und Ganzen mit dem Bildungsplan, lediglich ein einzelner im Bildungsplan geforderter Teilaspekt wird im Unterrichtsverlauf nur gestreift. Der Definition der Noten gemäß orientiert sich die Lehrkraft bei der Erstellung der Klassenarbeit am Bildungsplan und wählt je Teilaspekt im Bildungsplan zwei Aufgaben für die Klassenarbeit aus (kriteriale Bezugsnorm). Da die so konstruierte Klassenarbeit aus Sicht der Lehrkraft zu viele Aufgaben enthält, streicht sie diejenigen beiden Aufgaben, deren Inhalte im Unterricht zuvor nur sehr kurz behandelt wurden (Verschiebung des Kriteriums). Danach wählt die Lehrkraft die maximal zu erreichenden Punkte für jede Aufgabe, wobei einfache Aufgaben (im Sinne hoher erwarteter Lösungshäufigkeiten) weniger Punkte erhalten (soziale Bezugsnorm). Abschließend ergänzt die Lehrkraft die Klassenarbeit um eine Zusatzaufgabe (Abweichung von der kriterialen Norm). Diese antizipiert die Lehrkraft als eher schwierig, versieht sie aber dennoch mit wenigen Punkten (Abweichung von der sozialen Norm). Nach der Korrektur legt die Lehrkraft zunächst ihren üblichen Notenschlüssel an und berechnet den Durchschnitt. Dieser erscheint ihr in Anbetracht der wahrgenommenen Lernfortschritte während der Unterrichtseinheit zu schlecht, weshalb sie den Schlüssel abändert (individuelle Bezugsnorm).

4. Relativität und Normativität in der Bildungsforschung

Auch in Verfahren der Leistungsmessung, die in der empirischen Bildungsforschung zum Einsatz kommen, spielen Normativität und Relativität eine erhebliche Rolle. Da an dieser Stelle keine Einführung in die Psychometrie gegeben werden kann, werden stattdessen einige konzeptuelle Aspekte anhand eines Beispiels illustriert und auf vertiefende Literatur verwiesen.

Ein erstes normatives Moment der Leistungsmessung in der Bildungsforschung wird bereits bei der Wahl der metatheoretischen Grundannahmen sichtbar: Die überwiegende Mehrheit der empirisch-quantitativen Arbeiten nimmt an, mit der ‚Leistung‘ eine latente Variable zu messen (Edelsbrunner & Dablander, 2019) – also eine Variable, die nicht direkt, sondern nur mittelbar über Indikatoren gemessen werden kann (Borsboom, 2008). Dazu wird ein mathematisches Modell spezifiziert, das den Zusammenhang zwischen der empirischen Datenstruktur und der latenten Variablen (‚Leistung‘) beschreibt. Diese Modellierung erfolgt derzeit für fachliche Kompetenzen meist anhand unidimensionaler Item-Response-Modelle (IRT; Kelava & Moosbrugger, 2020). Diese postulieren unter anderem, dass die

Wahrscheinlichkeit, dass ein*e Schüler*in p eine Testfrage i richtig beantwortet $P(X_{pi} = 1)$, lediglich von der Schwierigkeit der Frage β_i und der Ausprägung der latenten Variable (Leistung) von Schüler*in θ_p abhängt, wobei diese Größen wie folgt in Beziehung stehen.

$$P(X_{pi} = 1) = \frac{e^{\theta_p - \beta_i}}{1 + e^{\theta_p - \beta_i}}$$

Andere Variablen (z. B. weitere Merkmale der Schüler*innen) oder Parameter, die die $P(X_{pi} = 1)$ beeinflussen, werden als nicht-existent angenommen. Diese Annahme des Modells, welche auch Rasch-Homogenität genannt wird, trägt insofern Normativität in den Prozess der Leistungsmessung ein, als dass sie mitbestimmt, welche Aufgaben der Leistungstest final enthält: Denn Aufgaben, deren empirische Datenmuster der Annahme der Rasch-Homogenität widersprechen, werden im Prozess der Testkonstruktion entfernt bzw. ersetzt. Die Verwendung von IRT gilt im Kontext von *Large-Scale-Assessments* als wissenschaftlicher Konsens (Borsboom & Wijsen, 2017). Nichtsdestotrotz stellt diese Wahl selbst bereits eine normative Entscheidung dar, da sie – wie beschrieben – mit bestimmten Annahmen einhergeht.

Genauso würde die Entscheidung für eine alternative Modellierung (z. B. Knowledge Space Theory [KST]; Falmagne et al., 1990) eine normative Entscheidung darstellen, da diese ebenfalls (wenngleich auch andere) Annahmen trifft.

Relativität und Normativität

in der Kompetenzstufenmodellierung

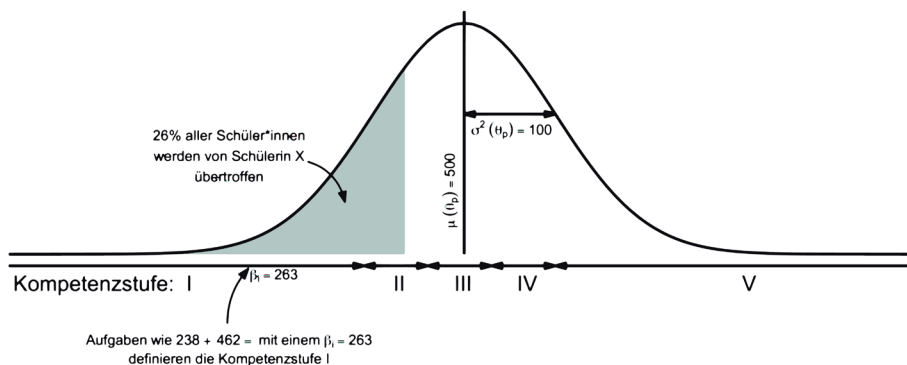


Abb. 2. Simultan kriteriale und soziale Normierung in der IRT.

Relativität wird bei der Messung von Leistung anhand IRT dadurch erzeugt, dass die Fähigkeitsparameter θ_p einer Normalverteilung folgen (müssen). Damit kann aus Kenntnis des Erwartungswertes der θ_p sowie deren Streuung jedes θ_p in Perzentile transformiert werden (siehe Abb. 2). In Abb. 2 wird zudem deutlich, dass ein*e Schüler*in X, welche*r bei diesem Leistungstest 425 Punkte erzielt hat, 26% aller Schüler*innen übertrifft. Obwohl die IRT damit (im Gegensatz zur KST) die Anwendung einer sozialen Bezugsnorm impliziert, erlaubt ein solcher Konstruktionsprozess der Kompetenzmodellierung auch die Anwendung der individuellen und kriterialen Bezugsnorm: Die individuelle Bezugsnorm etwa wird besonders einfach anwendbar, da die Leistung einer Person mit nur einem Parameter θ_p beschrieben werden kann. Durch die sog. Verankerung von Items wird es sogar möglich, θ_p anhand überschneidungsfreier Aufgabensets zu bestimmen und damit Lernfortschritte in der Einheit von θ_p zu bestimmen. Die Anwendung der kriterialen Bezugsnorm wird durch die Bestimmung sog. Ankerpunkte (Beaton & Allen, 1992) für die β_i ermöglicht. Die Bestimmung dieser Ankerpunkte stellt eine Post-Hoc-Analyse (Analyse nach Erhebung der Daten) von Iteminhalten dar, in der fachdidaktische Expert*innen das Leistungskontinuum in diskrete Intervalle unterteilen (Kompetenzstufen, siehe Abb. 2). Dabei werden Aufgabeninhalte und Aufgabenanforderungen derjenigen Aufgaben untersucht, die empirisch gesehen besonders spezifisch für einen Ankerpunkt auf der Kompetenzskala bzgl. notwendiger kognitiver Operationen, inhaltlicher Kriterien (z. B. Wortschatz eines Lesetextes), Aufgabenformate etc. sind (Hartig, 2007). Die Schnittmenge dieser Aufgabencharakteristika wird dann zur Definition der Kompetenzstufe herangezogen. Da die Schwierigkeitsparameter β_i und die Fähigkeitsparameter θ_p dieselbe Metrik aufweisen (‚auf derselben Achse liegen‘), kann die Leistung θ_p kriterial im Sinne von ‚ist fähig, Aufgaben der Art/Kompetenzstufe X zu lösen‘ interpretiert werden.

5. Fazit

Eine Leistungsbewertung erfordert notwendigerweise die Relativierung eines Messwertes. Im Beitrag wurde argumentiert, dass Lehrkräfte diese Relativierung (implizit) entlang normativ gewählter Bezugsnormen vornehmen, wobei diese im Prozess oftmals wechselt, während in psychometrischen Modellierungen von schulischer Leistung bereits die Wahl des mathematischen Modells eine normative (wenngleich auch gut begründete) Entscheidung darstellt. Relativierungen von Leistungen sind in den häufig vorkommenden Modellen der IRT sowohl sozial als auch kriterial und intraindividuell leicht möglich und inhaltlich interpretierbar.

6. Literatur

- Beaton, A. E. & Allen, N. L. (1992). Interpreting Scales Through Scale Anchoring. *Journal of Educational Statistics*, 17(2), 191. <https://doi.org/10.2307/1165169> [25.01.2023]
- Bohl, T. (2004). *Prüfen und Bewerten im Offenen Unterricht*. Weinheim: Beltz.
- Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research & Perspective*, 6(1-2), 25-53. <https://doi.org/10.1080/15366360802035497> [25.01.2023]
- Borsboom, D. & Wijsen, L. D. (2017). Psychology's atomic bomb. *Assessment in Education: Principles, Policy & Practice*, 24(3), 440-446. <https://doi.org/10.1080/0969594X.2017.1333084> [25.01.2023]
- Edelsbrunner, P. A. & Dablander, F. (2019). The Psychometric Modeling of Scientific Reasoning: A Review and Recommendations for Future Avenues. *Educational Psychology Review*, 31(1), 1-34. <https://doi.org/10.1007/s10648-018-9455-5> [25.01.2023]
- Falmagne, J.-C., Doignon, J.-P., Koppen, M., Villano, M. & Johannesen, L. (1990). Introduction to knowledge spaces: How to build, test, and search them. *Psychological Review*, 97(2), 201-224. <https://doi.org/10.1037/0033-295X.97.2.201> [25.01.2023]
- Hartig, J. (2007). Skalierung und Definition von Sprachniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen: Konzepte und Messung* (S. 83-99). Weinheim: Beltz.
- Heckhausen, H. (1974). *Leistung und Chancengleichheit*. Göttingen: Hogrefe.
- Holmeier, M. (2013). *Leistungsbeurteilung im Zentralabitur*. Wiesbaden: Springer VS.
- Kelava, A. & Moosbrugger, H. (2020). Einführung in die Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 369-409). Springer: Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4_16 [25.01.2023]
- Klauer, K. J., Fricke, R., Herbig, M., Rupperecht, H. & Schott, F. (Hrsg.). (1972). *Lehrzielorientierte Tests*. Düsseldorf: Schwann.
- Lintorf, K. & Buch, S. R. (2021). Stabile Präferenz oder flexibel am Diagnoseziel orientiert? – Die Bezugsnormwahl angehender Lehrkräfte. *Zeitschrift für Pädagogische Psychologie*, 35(2-3), 107-118. <https://doi.org/10.1024/1010-0652/a000271> [25.01.2023]
- Ministerium für Kultus, Jugend und Sport Baden-Württemberg (2023). *Bewertungskriterien/Wertungstabellen Schwimmen und Leichtathletik für das Fach Sport in den vier Halbjahren der Qualifikationsphase und in der Abiturprüfung 2023*. <https://km-bw.de/Lde/startseite/schule/dokumente-sportabitur> [30.01.2023]
- Pant, H. A. (2020). Notengebung, Leistungsprinzip und Bildungsgerechtigkeit. In S.-I. Beutel & H. A. Pant (Hrsg.), *Lernen ohne Noten. Alternative Konzepte der Leistungsbeurteilung* (S. 22-58). Stuttgart: Kohlhammer.
- Rheinberg, F. (1980). *Leistungsbewertung und Lernmotivation*. Göttingen: Hogrefe.
- Rheinberg, F. & Fries, S. (2010). Bezugsnormorientierung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 61-67). Weinheim: Beltz.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- Wilbert, J. & Gerdes, H. (2009). Die Bezugsnormwahl bei der Bewertung schulischer Leistungen durch angehende Lehrkräfte des Förderschwerpunktes Lernen. *Heilpädagogische Forschung*, 35, 122-135.

Angaben zum Autor und zur Autorin:

- Prof. Dr. Samuel Merk, Juniorprofessor für Empirische Schul- und Unterrichtsforschung, Pädagogische Hochschule Karlsruhe, samuel.merk@ph-karlsruhe.de
- Sarah Bez, wissenschaftliche Mitarbeiterin, Institut für Erziehungswissenschaft an der Universität Tübingen und Institut für Schul- und Unterrichtsentwicklung an der Pädagogischen Hochschule Karlsruhe, sarah.bez@uni-tuebingen.de

Georg Breidenstein

Zensuren: Risiken und Nebenwirkungen

1. Das Projekt „Leistungsbewertung in der Schulklasse“

Im Folgenden wird zusammenfassend von Ergebnissen eines Forschungsprojektes berichtet, das über mehrere Jahre hinweg den Alltag schulischer Leistungsbewertung beobachtet hat. Das Forschungsinteresse richtete sich darauf, die Bedeutung zu ergründen, die den Zensuren im Unterrichtsalltag deutscher Schulen zukommt. Um zu verstehen, was hier vor sich geht, so die Ausgangsüberlegung des Projektes, muss man die Praktiken selbst in den Blick nehmen, die sich um die Leistungsbewertung ranken. Welche Eigendynamik und Selbstläufigkeit entwickeln diese Praktiken? Welche Inszenierungen, Routinen und Rituale knüpfen sich an die alltägliche Handhabung der Zensuren? Die zentrale Untersuchungsfrage lautete: Was leisten die Praktiken der Bewertung für den Unterrichtsalltag?

Die genannten Fragestellungen erforderten die Beobachtung, Aufzeichnung und Analyse des unterrichtlichen Geschehens selbst. Bei den Praktiken und Ritualen, die die Noten im Schulalltag in Szene setzen, handelt es sich um Routinen des alltäglichen Handelns, die der Reflexion nur zum Teil zugänglich sind und über die die Teilnehmer*innen auch nicht ohne Weiteres im Interview berichten könnten. Insofern bestand die zentrale Methode in der teilnehmenden Beobachtung. Gegenstand der Beobachtung waren die alltäglichen Situationen schulischen Bewertens: Tests, Klassenarbeiten, mündliche Prüfungen – und vor allem die Rückgabe von Arbeiten, die Besprechung von Zensuren und schließlich Zeugnisausgaben. Das Feld der teilnehmenden Beobachtung bildeten zwei Schulklassen in zwei kontrastierenden Schulen, einem Gymnasium und einer Sekundarschule (eine Schulform, die Haupt- und Realschulzweig umfasst). Unsere Erhebungen haben zu Beginn der 5. Klasse eingesetzt und führten bis ins 7. Schuljahr hinein. (Zaborowski, Meier und Breidenstein 2011)

2. Befunde: Die Legitimierung von Noten in der schulischen Praxis

Was lässt sich auf der Basis unserer Analysen über das grundlegende Funktionieren, über die Eigenlogik der Praxis der Notengebung sagen? Fast alle der beobachteten Praktiken beziehen sich auf Probleme der Legitimierung der zu vergebenden Noten. Dabei geht es weniger um die Legitimität von Noten an sich, die scheint kaum zur Debatte zu stehen, sondern darum, dass die konkret von der

Lehrerin oder dem Lehrer vergebenen Noten von den Betroffenen in der Situation als ‚gerecht‘ anerkannt werden. Der Aufwand, den Lehrkräfte in der Begründung von Noten betreiben, verweist auf erhebliche Legitimationsprobleme. Dass ‚Objektivität‘ im Kontext schulischer Leistungsbewertung nicht erreichbar ist, weil diese vielfältigen Verzerrungen und Erwartungseffekten unterliegt, ist allen Beteiligten (mehr oder weniger) bewusst – Schüler*innen der fünften Klasse haben bereits die unterschiedlichen Bewertungsmaßstäbe unterschiedlicher Lehrpersonen kennen gelernt; die Lehrkraft hat möglicherweise nicht die entsprechenden Befunde der empirischen Forschung präsent, aber sie weiß zumindest intuitiv, wie prekär die Aufgabe der Zensurengebung ist. Für die Legitimität der Notengebung scheint die Beschwörung der Fiktion einer neutralen und objektiven Bewertung dennoch oder gerade deswegen unverzichtbar zu sein.

2.1. Mathematisierung

Eine überaus weit verbreitete Praxis der Legitimierung der Noten stellt die Mathematisierung der Notenvergabe dar. Die Note wird gefunden, indem sie ‚errechnet‘ wird: Sie wird als Durchschnitt aus vielen Einzelnoten gebildet. Dieses Verfahren ist zwar aus Sicht der statistischen Theorie äußerst fragwürdig, denn die Zensuren bilden eine Ordinalskala (Werte werden sortiert), und verrechnen dürfte man nur Werte aus einer metrischen Skala (die tatsächlich zu messen in der Lage ist), aber dieses Problem wird bis in die Erlasslagen hinein ignoriert, zu wichtig scheint die Suggestion der Objektivierung qua Mathematisierung zu sein, als dass die Praxis darauf verzichten wollte. Die (scheinbare) Mathematisierung der Zensuren stellt eine (im Effekt) überzeugende Rhetorik der ‚Objektivität‘ zur Verfügung, die im Gefüge der Praktiken der Legitimierung von Noten eine zentrale Rolle spielt.

2.2. Quantität der Produktion an Einzelnoten

Der Versuch, die Objektivität der Notengebung vorwiegend mathematisch abzusichern, hat noch einen weiteren Effekt, der den schulischen Alltag deutlich prägt. Die Rede ist von der Quantität der Produktion an Einzelnoten. Eine Logik des Je-mehr-desto-Besser scheint die Gesamtnote zu legitimieren. Zusätzlich gewonnene Noten verringern das Gewicht der einzelnen Note: Schüler*innen haben mehr ‚Chancen‘ erhalten, die Leistungen sind kontinuierlicher und umfassender überprüft worden, der Spielraum für Fehleinschätzungen erscheint kleiner.

Tatsächlich haben wir ein verblüffendes Ausmaß der Notenproduktion beobachtet. Es verging kaum ein Tag, an dem nicht Zensuren vergeben oder verkündet wurden. Mit der Permanenz der Bewertungen wird auch der Bewährungsdruck auf

Dauer gestellt: Schüler*innen können sich nie sicher sein, ob nicht schon wieder die nächste Leistungsüberprüfung naht und mit ihr neue Noten vergeben werden. Andererseits entwickeln Schüler*innen eine große Routine im Umgang mit Leistungsbewertungen, wobei schließlich eine Inflationierung des Wertes einzelner Zensuren unvermeidlich scheint: Wie soll bei einer Zahl von Hundert oder mehr Benotungen im Halbjahr der einzelnen Zensur noch Bedeutung verliehen werden? Die Prüfung wird so sehr zum Bestandteil des Unterrichtsalltages, dass der ‚Leistungsdruck‘ zwar auf Dauer gestellt wird – sich aber möglicherweise gerade darin auch wieder bricht.

2.3. Einbeziehung der Schüler*innen

Eine andere Praxis der Legitimierung besteht in der ‚Einbeziehung‘ von Schüler*innen in die Notenfindung: Schüler*innen sollen sagen, was für eine Note sie sich ‚selbst geben würden‘. Diese Einbeziehung stellte letztlich in allen beobachteten Fällen nur eine Pseudo-Einbeziehung dar, denn das Lehrerurteil wurde immer auch gegen abweichende Selbsteinschätzungen aufrechterhalten – in vielen Fällen wurde aber eben doch erreicht, dass die Schüler*innen der Bewertung öffentlich zustimmen und sie dadurch legitimierten.

2.4. Alleinige Verantwortung der Schüler für die erbrachten Leistungen

Das vermutlich folgenreichste Problem im Rahmen der Legitimierung von Noten besteht darin, dem Schüler oder der Schülerin die bewertete Leistung vollständig zuzurechnen. Dies ist keineswegs trivial, wenn man sich klarmacht, dass jede schulische Leistungsbewertung den Unterrichtserfolg überprüft, der sich aus der Qualität der Vermittlung *und* den Aneignungsbemühungen bzw. -kapazitäten zusammensetzt. Gerade bei schlechten Noten muss aber gewährleistet werden, dass nicht etwa der Unterricht daran (mit-)schuldig ist. Es muss klargestellt werden, dass die gemessene Leistung nicht etwa darin begründet liegt, dass die Erklärung des in Frage stehenden Sachverhaltes im Unterricht mangelhaft war oder die unterrichtlichen Lerngelegenheiten ungünstig waren. Die Note muss voll und ganz dem betreffenden Schüler oder der Schülerin zugerechnet werden können, denn die Note soll den individuellen Leistungsstand ausdrücken. Noten erscheinen als Urteile über Schülerleistungen nur dann legitim, wenn diese ohne Abstriche für ihr Zustandekommen verantwortlich gemacht werden.

Wenn die Noten sich wenig verändern, reduziert das den Begründungsbedarf, denn dann ist die Verbindung zwischen Note und Person etabliert und unproblematisch. Verbesserungen von Noten können genutzt werden, um Erfolgsgeschichten zu inszenieren. Das eigentliche Erklärungsproblem werfen offensichtlich

Verschlechterungen von Noten auf, denn in diesen Fällen ist dokumentiert, dass die jeweilige Schülerin oder der Schüler ‚eigentlich‘ mehr kann und ihr oder sein ‚Leistungsvermögen‘ diesmal nicht abgerufen hat. Diese Fälle scheinen besonders prekär für die Lehrkraft, denn es muss der Verdacht ausgeräumt werden, dass ihr Unterricht diesen (eigentlich) guten Schüler*innen die Sache nicht vermitteln konnte. Hier finden sich jedenfalls die ausführlichsten Kommentierungen und Begründungen von Noten, die sich auf das (Fehl-)Verhalten des betreffenden Schülers oder der Schülerin beziehen, wahlweise aber auch auf deren Persönlichkeit oder ihr soziales Umfeld (Breidenstein, 2012). Das Dementi des eigenen Anteils bzw. das Insistieren auf der alleinigen Verantwortung der Schüler*innen für die erbrachten Leistungen zieht sich in frappierender Konsequenz durch unsere sämtlichen Beobachtungen von Szenen der Zensurenbekanntgabe. Wir haben keine einzige Situation beobachtet, in der eine Lehrperson klassenöffentlich in Erwägung gezogen hätte, dass der eigene Unterricht für enttäuschende Leistungen von Schüler*innen mitverantwortlich sein könnte.

3. Zensuren: Objekte der Distanzierung und des Kampfes um Anerkennung

In der Analyse des Alltags der Leistungsbewertung lassen sich also mindestens drei aufeinander bezogene Praktiken der Legitimierung von Noten herausarbeiten: Die Noten werden als Ergebnis mathematischer Berechnungen präsentiert; es wird ein Verfahren etabliert, das die öffentliche Zustimmung der Betroffenen zu ihren Noten hervorbringt; und der Unterricht der Lehrperson wird von seinem Anteil an der bewerteten Leistung vollständig entlastet. In der Kombination dieser drei Praktiken wird die Note als objektivierte Bewertung einer individuellen Schülerleistung etabliert. Die ‚Objektivität‘ dieser Noten ist also genau in diesen sozialen Praktiken erzeugt und keineswegs unabhängig davon gegeben. Kurz: Schulische Zensuren sind ‚objektiv‘, insoweit sie in komplexen sozialen Praktiken objektiviert worden sind. – Dieser Zusammenhang darf den Beteiligten allerdings nicht bewusst sein, sonst könnte die Objektivität der Zensuren kaum beansprucht werden. Die Beobachtung des Unterrichtsalltages zeigt eine erstaunliche Stabilität der Legitimität von Zensuren sowohl bei Lehrpersonen als auch bei Schüler*innen. Die sozialen Praktiken, die sich auf die Objektivierung der Zensuren richten, schaffen es offenbar in bemerkenswerter Weise, nicht nur die Objektivität von Zensuren zu erzeugen, sondern auch ihren eigenen Anteil an diesem Prozess unsichtbar zu machen.

Es ist nun insgesamt ein Kennzeichen der beobachteten Unterrichtspraxis, dass Zensuren in das Zentrum der Auseinandersetzung um die Anerkennung der

Bedeutung des Schulischen durch die Schüler*innen rücken. Die Notengebung wird mobilisiert, um wenigstens ein Mindestmaß an Engagement von Schüler*innen zu erzwingen, aber auch weiterreichendes Engagement und bedeutendere Leistungen von Schüler*innen werden in erster Linie durch gute Noten honoriert. Das Engagement von Schüler*innen im Rahmen des Unterrichts wird wesentlich durch die Zensurengebung aufrechterhalten – allerdings zugleich auch begrenzt. Denn die Etablierung der Noten im Kern des Unterrichtsgeschehens als zentrales Mittel der Sanktionierung und Motivierung funktioniert zwar fast unhinterfragt und erscheint doch ausgesprochen prekär. Unsere Beobachtungen zeigen, wie empfindlich und letztlich hilflos Lehrpersonen agieren, wenn die Anerkennung der Relevanz der Zensur gefährdet erscheint.

Die Zensurengebung ist darauf angewiesen, dass die Schüler*innen die Bedeutsamkeit der Noten anerkennen. Diese Anerkennung muss durch passende Reaktionen demonstriert werden: Freude über gute Noten und Zerknirschung bei schlechten Noten. Dabei steht viel auf dem Spiel: Da mit der Note über die Sinnhaftigkeit und Bedeutsamkeit des Schulischen schlechthin verhandelt wird, geht es ums Ganze. Ihre Brisanz erhält die Auseinandersetzung um das Ernstnehmen der Zensuren durch ihren öffentlichen Charakter: Die Auseinandersetzung spielt sich zwar zwischen einzelnen Schüler*innen und der Lehrperson ab – aber immer vor dem Publikum der Mitschüler*innen. Das macht die Sache zusätzlich heikel, denn jeder Einzelfall ist potenziell verallgemeinerbar, die Verweigerung einer einzelnen Schülerin oder eines einzelnen Schülers kann jederzeit für die Verweigerung einer größeren Gruppe oder gar der ganzen Klasse stehen.

Wir haben vor allem an der Sekundarschule gesehen, wie Lehrpersonen sich komplexe Rituale ausdenken, um Kindern die Bedeutsamkeit von Zensuren zu vermitteln; wie dünnhäutig Lehrpersonen reagieren, wenn die angemessen zerknirschte Reaktion auf eine schlechte Note ausbleibt und insgesamt wie umkämpft die Anerkennung der Relevanz der Zensuren ist. Dieser Kampf um die Anerkennung des Schulischen mittels Zensurengebung stellt sich als sich selbst verstärkender Prozess dar: Das Streben nach guten Noten wird mit guten Noten belohnt und die Geringschätzung schlechter Noten wird mit noch härteren Noten bestraft. Die Praxis der Zensurengebung entwickelt in diesem Prozess eine ungeheure Dynamik: Zensuren lösen sich von der Funktion der Leistungsrückmeldung und erscheinen als reines Mittel zum Zweck, der in der Aufrechterhaltung (eines Minimums) von Unterrichtsbereitschaft liegt. Dieser Zweck allerdings wird selbst wiederum gefährdet, wenn er sich nur noch mit dem Mittel der Zensurengebung erreichen lässt. Denn letztlich ist zu vermuten, dass sich mit der Fokussierung schulischen Unterrichts auf die Zensurengebung auch auf Seiten der Schüler*innen das Interesse am Unterricht auf dessen ‚Ergebnisse‘ in Form von Zensuren reduziert.

Zensuren werden in unmittelbarer und sehr gegenständlicher Art und Weise zum Objekt der Distanzierung, der Auseinandersetzung und des Kampfes um Anerkennung. Im Zuge der schulischen Sozialisation entwickeln sich notwendigerweise Praktiken der Distanzierung von Noten. Je länger Schüler*innen schulischer Leistungsbewertung in Form von Zensuren ausgesetzt sind, desto unwahrscheinlicher wird, dass sie diese im Sinne einer ‚Leistungsrückmeldung‘ ernst nehmen. Das heißt nicht, dass Zensuren bedeutungslos würden. Wenn beispielsweise die Versetzung gefährdet ist, wird von Schüler*innen sehr genau gerechnet und wenn es auf das Abitur zugeht, sind die Punktestände durchaus im Fokus der Aufmerksamkeit. Aber der Umgang mit den Noten insgesamt ist als zunehmend pragmatisch und instrumentell zu beschreiben. So legen Schüler*innen mit zunehmender Erfahrung ein ausgesprochen ökonomisches Kalkül in der Dosierung ihres Einsatzes an den Tag und investieren nur dort, wo sich noch eine Verbesserung der Zensur erreichen lässt (Breidenstein, 2006).

Wenn man versucht, Sinnprobleme schulischen Unterrichts mithilfe der Zensurengebung zu bearbeiten, handelt man sich weitreichende Folgeprobleme ein. In unseren Beobachtungen hat sich gezeigt, wie der Kampf um die Anerkennung des Sinns von Schule und Unterricht groteske Formen annehmen kann. Zu befürchten ist, dass die Schule, indem sie immer unvermittelter und verzweifelter auf Zensuren als letztes Mittel in diesem Kampf setzt, sich in eine Auseinandersetzung verwickelt, die sie nicht gewinnen kann.

4. Konsequenzen: Reflexion und Distanz schaffen

Der hier skizzierten Effekte der Notengebung ist man sich im Handeln nicht bewusst und erst recht hat sie keiner gewollt. Es handelt sich um *nicht-intendierte* Effekte, die sich in der *Eigenlogik* des praktischen Vollzugs der Zensurengebung herstellen und die sich nur der detaillierten Beobachtung und Analyse dieses Vollzugs erschließen. Abschließend ist zu fragen, welche Hinweise sich aus der Analyse für die praktische Handhabung der Zensurengebung ergeben – wenn man davon ausgeht, dass die Notengebung in weiten Teilen zumindest des staatlichen Schulwesens auf absehbare Zeit noch Bestand haben wird. Ich will drei Überlegungen thesenartig zur Diskussion stellen.

Es scheint erstens angeraten, die Bedeutung der Noten in der alltäglichen unterrichtlichen Kommunikation bewusst zu begrenzen. Wenn man sieht, wie die Praxis der Notengebung dazu neigt, sich zu verselbständigen und die Kommunikation zwischen Lehrpersonen und Schüler*innen zu dominieren, muss man sich dieser Dynamik wohl bewusst widersetzen. Das System der Ziffernoten übt eine eigene Anziehungskraft und Faszination auf alle Beteiligten, Lehrer*innen, Schüler*innen

und Eltern aus, die als solche reflektiert und thematisiert werden sollte. Die Bedeutung der Noten zu begrenzen hieße z. B. explizit ‚notenfreie‘ Räume im Unterricht zu etablieren und sowohl in der Kommunikation mit den Schüler*innen als auch mit den Eltern andere Formen der (Leistungs-) Rückmeldung zu entwickeln.

Zweitens wird es darauf ankommen, reflexiv mit der Fiktion der ‚Objektivität‘ der Noten umzugehen. Zwar ist diese Fiktion für die Legitimierung der Zensuren konstitutiv und unvermeidbar, andererseits zeitigt der Versuch der Objektivierung der Noten die beschriebenen problematischen Effekte: eine auf Dauer gestellte Benotung, den Pseudo-Einbezug von Schüler*innen in die Notengebung und die vollständige Schuldzuweisung an die Schüler*innen bei schlechten Noten. Diese Effekte lassen sich wohl nur begrenzen, wenn man auch zu der Idee der Objektivität der Noten so viel Distanz entwickelt, dass man einzuräumen bereit ist, dass die gebotene Objektivität der Noten auch bei bestem Bemühen immer wieder verfehlt wird. Dieses Eingeständnis, das Schüler*innen wie Eltern gegenüber zu artikulieren wäre, würde Spielräume für einen reflexiven und distanzierten Umgang mit der Praxis der Notengebung eröffnen.

Schließlich scheint es geboten, insbesondere *pädagogische* Ambitionen, die sich an die Notengebung knüpfen, kritisch zu reflektieren. Wie die Analyse von Szenen der Begründung von Noten zeigt, ist es gerade der Versuch, den Schüler*innen anhand der Noten individuelle, erzieherisch wirksame Rückmeldungen zu geben, der zu weitreichenden und entgrenzenden Zuschreibungen führt. Es kommt darauf an, sich der Risiken des diagnostizierenden Blicks bewusst zu bleiben, der auch bei Wohlwollen Schüler*innen immer zugleich auf bestimmte Charaktereigenschaften, Positionen in der Klasse und Leistungsvermögen festschreibt.

Vermutlich besteht der beste Schutz vor den problematischen Effekten einer sich verselbstständigenden Praxis der Zensurenggebung darin, sich die eigene Praxis gelegentlich vor Augen zu führen und auf ihre Effekte zu reflektieren. Vielleicht sollte man ab und zu das Geschehen aufzeichnen, wenn man eine Klassenarbeit zurückgibt oder mit Schüler*innen über Zensuren spricht. Die subtilen Praktiken der Zuschreibung, Identifizierung und auch Stigmatisierung jedenfalls zeigen sich erst der nachträglichen und detailgenauen Analyse. Bei einer solchen Analyse der eigenen Praxis käme es auch darauf an, sich die Situation der Bewerteten vor Augen zu führen: Welche Handlungsoptionen haben die Schüler*innen? Bleibt ihnen etwas anders übrig, als sich von dieser Art des Bewertet-Werdens zu distanzieren?

5. Literatur

Breidenstein, G. (2006). *Teilnahme am Unterricht. Ethnographische Studien zum Schülerjob*. Wiesbaden: Springer VS.

Breidenstein, G. (2012). *Zeugnisnotenbesprechung. Zur Analyse der Praxis schulischer Leistungsbewertung*. Opladen: Budrich.

Breidenstein, G. (2015). Zensuren: Anwendungsgebiete, Risiken und Nebenwirkungen. In Huber, St. G. (Hrsg.), *Jahrbuch Schulleitung. Befunde und Impulse zu den Handlungsfeldern des Schulmanagements* (S. 60-68). Köln: Wolters Kluver Deutschland.

Zaborowski, K., Meier, M., Breidenstein, G.: *Leistungsbewertung und Unterricht. Ethnographische Studien zur Bewertungspraxis in Gymnasium und Sekundarschule*, Wiesbaden: VS.

Angaben zum Autor:

- Prof. Dr. Georg Breidenstein, Professur für Erziehungswissenschaft mit dem Schwerpunkt Grundschulpädagogik, Martin-Luther-Universität Halle-Wittenberg, Institut für Schulpädagogik und Grundschuldidaktik, georg.breidenstein@paedagogik.uni-halle.de.

Benjamin Fauth

Bewertungen von Lehrenden und Lernenden im Unterricht: Evaluation und Feedback

1. Einleitung

„Wie sich Leistung bewerten lässt“ – das war die Frage der Vorlesungsreihe des Studium Generale, die auch Anlass für diesen Beitrag ist. Im Folgenden soll nicht nur gefragt werden wie sich die Leistungen von Schüler*innen bewerten lassen, sondern auch wie unterschiedlich die Bewertung der Leistung von Lehrkräften aussehen kann und wie pädagogisch sinnvolle Formen der Leistungsrückmeldung, die im Rahmen von Untersuchungen an Schüler*innen erforscht wurden, auch für Rückmeldungen an Lehrkräfte zu ihrem Unterricht lehrreich sein können. Gehen wir davon aus, dass Lehrkräfte nicht bloß Lehrende, sondern immer zugleich auch Lernende sind, so liegt es nahe, auch bei der Bewertung von Lehrkräften nach Ansätzen zu suchen, die vor allem für ihre professionelle Weiterentwicklung hilfreich sind. Es liegt auf der Hand, dass der Unterricht als ‚Kerngeschäft‘ des Lehrkräfteberufs hier besonders im Fokus steht, denn der Unterricht ist der Ort, an dem sich das Lernen der Schüler*innen und das Lernen der Lehrkräfte am Ende treffen muss.

2. Bewertungen von Lehrkräften

Flächendeckende Bewertungen von Lehrkräften im Sinne einer Beurteilung finden in Deutschland im Grunde nur in der Lehrkräfteausbildung statt. Im Rahmen des Referendariats werden angehende Lehrkräfte unter anderem auf der Grundlage von Unterrichtsbeobachtungen beurteilt. Die Ergebnisse empirischer Studien zeigen allerdings, dass es bei Unterrichtsbeobachtungen – auch wenn sie von geschulten Beobachter*innen vorgenommen werden – zu deutlichen Unterschieden in der Unterrichtsqualität kommen kann, wenn dieselbe Lehrkraft in unterschiedlichen Klassen unterrichtet. Empirisch sind diese Unterschiede unter anderem durch die Motivation (z. B. Interesse und Anstrengungsbereitschaft) der unterrichteten Schüler*innen zu erklären (Fauth et al., 2020).

Die bei Unterrichtsbesuchen gemachten Beobachtungen der ‚Leistungen‘ einer Lehrkraft müssen also immer *relativ* zu der Klasse bewertet werden, in der gerade unterrichtet wird (ebd.; Voss et al., 2022). Für die Qualität des Unterrichts einer Lehrkraft kann es einen sehr großen Unterschied machen, ob sie in dieser oder in jener Klasse beobachtet und bewertet wird. Beobachter*innen müssen sich dessen bewusst sein.

Nach Abschluss der Lehramtsausbildung finden Bewertungen des Unterrichts in Deutschland vornehmlich im Rahmen der externen Evaluation an Schulen statt (z. B. Gärtner, 2013). Üblicherweise statten mehrere Evaluatord*innen oder Inspektord*innen den Schulen Besuche ab, machen Unterrichtsbeobachtungen und verfassen einen Evaluationsbericht. Die Rückmeldung erfolgt in den Bundesländern, die externe Evaluation betreiben, auf der Schulebene – nicht auf der Ebene einzelner Lehrkräfte. Auch sind diese Evaluationssysteme keine High-Stakes-Evaluationen, d. h. aus der Evaluation ergeben sich keine direkten Konsequenzen, was zum Beispiel die finanzielle Ausstattung von Schulen oder die Bezahlung, Beförderung oder Entlassung von Lehrkräften angeht.

Solche Konsequenzen externer Evaluation klingen für deutsche Ohren sehr ungewohnt. Die Vorstellung, aufgrund eines schlechten Evaluationsergebnisses weniger Gehalt zu bekommen oder gar entlassen zu werden, ist hierzulande sehr fremd. Anders ist dies in Ländern, in denen standardmäßig solche *high-stakes teacher evaluation systems* implementiert sind, beispielsweise in den Vereinigten Staaten (Lavigne, 2014). Auch dort kommen Unterrichtsbeobachtungen zum Einsatz, für die eine Reihe von standardisierten und wissenschaftlich geprüften Instrumenten genutzt werden (National Council of Teacher Quality, 2018).

Daneben wird die Leistung von Lehrkräften in vielen Evaluationssystemen auch anhand von sogenannten Value-Added-Scores bewertet. Diese Werte basieren auf standardisierten Leistungstests der Schülerinnen und Schüler einer Lehrkraft. Unter Berücksichtigung von individuellen Hintergrundmerkmalen der Schüler*innen wird ein Erwartungswert berechnet, der hinsichtlich der fachlichen Leistungen in einer Klasse oder einem Kurs erzielt werden müsste (Opper, 2019). Abweichungen von diesem Erwartungswert nach unten oder oben werden dann auf die Leistung der Lehrkraft zurückgeführt. Der Bildungsökonom Eric Hanushek bringt die dahinter stehende Sichtweise folgendermaßen auf den Punkt:

„I use a simple definition of teacher quality: good teachers are ones who get large gains in student achievement for their classes; bad teachers are just the opposite.“ (Hanushek, 2002, S. 3)

Im Jahr 2010 veröffentlichte die Los Angeles Times – eine seriöse Tageszeitung für den Großraum Los Angeles – die Value-Added-Scores von allen ca. 11.500 Grundschullehrkräften, die damals im Los Angeles Unified School District tätig waren (<https://projects.latimes.com/value-added/>). Diese Veröffentlichung wurde auch in den Vereinigten Staaten kontrovers diskutiert. Aber allein die Tatsache, dass die Los Angeles Times es für eine gute Idee hielt, die Value-Added-Scores der Lehrkräfte zu veröffentlichen, zeigt wie unterschiedlich auf die Bewertung von Lehrkräften geblickt werden kann.

Relativiert wird die Leistung von Lehrkräften im LA Teacher Rating Project wohl-gemerkt ausschließlich an den Leistungen der Kolleg*innen: Es wurden Rangrei-hen gebildet und jede Lehrkraft wurde einem Perzentil zugeordnet. Die angelegte Bezugsnorm ist damit eine soziale.

3. Bewertungen von Lernenden: Formative Assessments

Bewertungen von Schülerinnen und Schülern im Unterricht sollen unter-schiedliche Funktionen erfüllen (z. B. Selektion, Sozialisation, Information und Rückmeldung). Von Seiten der Bildungswissenschaften und der Fachdidaktiken wurde verschiedentlich darauf hingewiesen, dass Leistungsrückmeldungen in Form von Schulnoten die didaktisch-pädagogische Funktion nicht sehr gut erfüllen. Folgt man Hattie und Timperley (2007), so ist Feedback im Unterricht vor allem dann hilfreich, wenn es Lernenden Informationen zu drei Fragen liefert:

- Was ist der aktuelle Lernstand? (*How am I going?*)
- Was ist das angestrebte Lernziel? (*Where am I going?*)
- Was sind die nächsten Schritte zum Lernziel? (*Where to next?*)

In einem Review-Artikel kommen Hattie und Timperley (2007) zu dem Schluss, dass Feedback vor allem dann lernwirksam wird, wenn es den Lernenden Informationen zu diesen drei Fragen zur Verfügung stellt (vgl. auch Käfer, Herbein & Fauth, 2021). Es scheint, dass so genannte formative Assessments diesen Ansprüchen in besonderer Weise gerecht werden (Black & William, 2009). Die Grundidee hinter formativen Assessments ist, dass sich Leistungsmessungen (Assessments) nicht nur für die Beurteilung von Schüler*innenleistungen verwenden lassen, wie dies beispielsweise bei einer benoteten Klassenarbeit am Ende einer Lerneinheit der Fall ist (summatives Feedback). Stattdessen werden im formativen Ansatz Leistungsmessungen in den laufenden Unterricht integriert und schon während des Lernprozesses wiederholt eingesetzt, um Lehrkräften wie Schüler*innen Informationen zur Lernentwicklung innerhalb einer Einheit geben zu können – und nicht erst an deren Ende.

Der entscheidende Punkt ist, dass die so gewonnenen Informationen nicht zur No-tengebung verwendet werden. Sie können einerseits von den Schüler*innen als Feedback im obigen Sinne genutzt werden und andererseits von den Lehrkräften verwendet werden, um ihren Unterricht anzupassen und einzelne Schüler*innen in-dividuell zu fördern (Dumont, 2019). Formativen Assessments, in denen Leistungs-messungen dann kein ‚Add-on‘ zum Unterricht mehr sind, sondern integraler Bestandteil, wird empirisch eine hohe Wirksamkeit bescheinigt (Hattie, 2009). In den

Meta-Studien von Hattie (ebd.) sind die Faktoren *providing formative evaluation* und *feedback* regelmäßig auf den vordersten Plätzen aller Indikatoren zu finden.

Diese hohe Wirksamkeit mag auch mit den hier angelegten Bezugsnormen zusammenhängen. Durch die wiederholte Erfassung von Leistungen kann besonders gut eine individuelle, zeitliche Bezugsnorm angelegt werden.

4. Unterricht beobachten und gemeinsam reflektieren

Auch ohne die kritische Diskussion um die oben vorgestellten Ansätze von high-stakes Evaluationen im Detail auszuführen, kann vermutet werden, dass Ansätze des formativen Feedbacks, wie sie hier für das Feedback an Schülerinnen und Schüler dargestellt wurden, auch für die professionelle Weiterentwicklung von Lehrkräften genutzt werden können. Begreift man Lehrkräfte konsequent auch als Lernende, so liegt ein solcher Ansatz durchaus nahe.

Allerdings stellt sich auch hier die Frage nach dem richtigen Instrument, das für ein formatives Feedback zum Unterricht genutzt werden kann. Dies gilt umso mehr, wenn bei der Beobachtung die sogenannten unterrichtlichen Tiefenstrukturen in den Blick genommen werden sollen, die in der empirischen Unterrichtsforschung einerseits als besonders wirksam für die Leistungsentwicklung von Schüler*innen gelten (Fauth & Trautwein, 2018), von denen man aber zum anderen auch weiß, dass sie im Rahmen von Unterrichtsbeobachtungen besonders schwer zu erfassen sind (Fauth & Leuders, 2021). Zur Erfassung der Tiefenstrukturen braucht es besonders gute und besonders gut validierte Instrumente, um sie der Erfahrung zugänglich zu machen und sie für Feedback nutzen zu können.

In den Bundesländern werden eine ganze Reihe unterschiedlicher Verfahren und Hilfsmittel zur Unterrichtsbeobachtung verwendet. Diesen liegen unterschiedliche theoretische Modelle und Begrifflichkeiten zu Grunde und auch in Bezug auf Umfang und Auswahl der berücksichtigten Unterrichtsmerkmale gibt es große Unterschiede. Eine explizite Überprüfung der psychometrischen Güte der eingesetzten Instrumente findet allerdings meistens nicht statt.

Vor diesem Hintergrund wurde am Institut für Bildungsanalysen Baden-Württemberg (IBBW) der „Unterrichtsfeedbackbogen Tiefenstrukturen“ (UFB; Fauth, Herbein & Maier, 2022) entwickelt, der den Anspruch hat, besonders relevante Aspekte der unterrichtlichen Tiefenstrukturen beobachtbar und damit für die gemeinsame Reflexion und Weiterentwicklung des Unterrichts nutzbar zu machen. Im Rahmen des Gesamtprojekts „Qualitätsentwicklung durch Unterrichtsbeobachtung und Feedback“ (QUBE-F), das vom Zentrum für Schulqualität und Lehrerbildung (ZSL) initiiert wurde, wird der Bogen in verschiedenen Studien validiert und

intensiv wissenschaftlich begleitet. Dabei wird Expertise aus empirischer Unterrichtsforschung und der Praxis der Unterrichtsentwicklung zusammengeführt. Erste Ergebnisse aus der wissenschaftlichen Begleitung belegen eine hinreichend hohe Beobachtungsübereinstimmung und eine gute konvergente Validität in Bezug auf Instrumente, die bislang in wissenschaftlichen Studien zur Erfassung der Tiefenstrukturen eingesetzt wurden (Maier et al., 2021).

Der Bogen umfasst insgesamt 11 Items, die in einem Manual ausführlich erläutert werden. Dazu gehört zum einen ein konzeptioneller Hintergrund zu jedem Item und zum anderen konkret beobachtbare Positiv- und Negativindikatoren, die eine kriteriengeleitete Einschätzung der Unterrichtsqualität ermöglichen. Dabei ist die eigentliche Unterrichtsbeobachtung nur ein Baustein in einem größeren Prozess der (gemeinsamen) Reflexion von Unterricht und der Unterrichtsentwicklung. Das kann im Rahmen der kollegialen Hospitation geschehen, aber auch in allen Beratungskontexten in der Aus- und Fortbildung von Lehrkräften.

Der UFB eignet sich somit in besonderer Weise für ein formatives Feedback zum Unterricht im oben skizzierten Sinn. Schaut man auf die Bezugsnormen, die dann angelegt werden, so steht hier nicht eine soziale Bezugsnorm im Vordergrund (wie beim LA Teacher Rating Project), sondern eine kriteriale, wobei die Kriterien definiert sind durch die Items und Indikatoren des Beobachtungsbogens (Fauth et al., 2022). Für die Unterrichtsentwicklung wird es besonders wertvoll sein, daneben auch eine individuelle, zeitliche Bezugsnorm anzulegen, indem gemeinsam reflektiert wird, welche Änderungen im Unterricht seit dem letzten Feedback zu beobachten sind.

5. Fazit

Als Fazit kann festgehalten werden: Es lohnt sich in Schule und Unterricht über lernwirksame Formen des Feedbacks nachzudenken und dabei kann die Unterscheidung von eher summativen und eher formativen Ansätzen hilfreich sein. Jene Ansätze, die sich in Bezug auf das Lernen von Schüler*innen als hilfreich und wirksam erwiesen haben, können Vorbild sein, wenn man über Ansätze der Evaluation von Lehrkräften nachdenkt, die für die professionelle Weiterentwicklung von Lehrkräften nützlicher sind als die oben beschriebenen high-stakes Evaluationen. Es ist deutlich geworden, dass die unterschiedlichen Relativierungen der erfassten Leistungen in Form von unterschiedlichen Bezugsnormen dabei eine zentrale Rolle spielen.

6. Literatur

Black, P. & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.

Dumont, H. (2019). Neuer Schlauch für alten Wein? Eine konzeptionelle Betrachtung von individueller Förderung im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 22, 249-277.

Fauth, B., Göllner, R., Lenske, L., Praetorius, A. & Wagner, W. (2020). Who sees what? Theoretical considerations on the measurement of teaching quality from different perspectives. *Zeitschrift für Pädagogik*, 66, 138-155.

Fauth, B., Herbein, E. & Maier, J. L. (2022). *Beobachtungsmニュアル zum Unterrichtsfeedbackbogen Tiefenstrukturen*. Stuttgart: Institut für Bildungsanalysen Baden-Württemberg (IBBW). Verfügbar unter: <https://ibbw.kultus-bw.de/Lde/Startseite/Empirische-Bildungsforschung/unterrichtsfeedbackbogen>

Fauth, B. & Leuders, T. (2021). Kognitiv aktivierenden Unterricht beobachten. *Pädagogik*, (11), 32-37.

Fauth, B. & Trautwein, U. (2018). Unterrichtsqualität: Konzeptualisierung, Erfassung und Wirkung. *Infobrief Schulpsychologie BW*, 18(1), 1-10.

Fauth, B., Wagner, W., Bertram, C., Göllner, R., Roloff, J., Lüdtke, O., Polikoff, M. S., Klusmann, U. & Trautwein, U. (2020). Don't blame the teacher? The need to account for classroom characteristics in evaluations of teaching quality. *Journal of Educational Psychology*, 112(6), 1284-1302.

Gärtner, H. (2013). Praxis und Verhältnis interner und externer Evaluation im Schulsystem im internationalen Vergleich. *Zeitschrift für Erziehungswissenschaft*, 16(4), 693-712.

Hanushek, E. (2002). Teacher Quality. In W. F. Evers & L.T. Izumi (2013), *Teacher quality* (S. 1-12). Stanford: Hoover Institution Press.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112.

Käfer, J., Herbein, E. & Fauth, B. (2021). *Formatives Feedback im Unterricht. Reihe Wirksamer Unterricht Band 4*. Verfügbar unter: <https://ibbw.kultus-bw.de/Lde/Startseite/Empirische-Bildungsforschung/Publikationsreihe-Wirksamer-Unterricht> [25.01.2023].

Lavigne, A. (2014). Exploring the Intended and Unintended Consequences of High-Stakes Teacher Evaluation on Schools, Teachers, and Students. *Teachers College Record*, 116(1), 1-29.

Maier, J. L., Panetta, T., Philipp, J., Ruth-Herbein, E. & Fauth, B. (2021). Unterrichtsfeedbackbogen Tiefenstrukturen. Präpilotierung und Validierung: Erste Ergebnisse der wissenschaftlichen Begleitung. Stuttgart: Institut für Bildungsanalysen Baden-Württemberg. https://ibbw-bw.de/site/pbs-bw-km-root/get/documents_E-2093645246/KULTUS.Dachmandant/KULTUS/Dienststellen/ibbw/Empirische%20Bildungsforschung/Programme-und-Projekte/Unterrichtsfeedbackbogen/UFB_Ergebnisbericht_Pr%C3%A4pilotierung%20Ergebnisse.pdf [19.07.2022].

National Council of Teacher Quality (2018). *2017 state teacher policy yearbook. National summary*. Washington, DC: NCTQ. <https://www.nctq.org/publications/2017-State-Teacher-Policy-Yearbook> [25.01.2023].

Opper, I., (2019). *Value-Added Modeling 101: Using Student Test Scores to Help Measure Teaching Effectiveness*. Santa Monica, CA: RAND Corporation.

Voss, T., Zachrich, L. Fauth, B. & Wittwer, J. (2022). The same yet different? Teaching quality differs across a teacher's classes, but teachers with higher knowledge make teaching quality more similar. *Learning & Instruction*, 80, 101614.

Angaben zum Autor

- apl. Prof. Dr. Benjamin Fauth, Leiter der Abteilung 4 – Empirische Bildungsforschung am Institut für Bildungsanalysen Baden-Württemberg (IBBW) und außerplanmäßiger Professor am Hector-Institut für Empirische Bildungsforschung, benjamin.fauth@ibbw.kv.bwl.de

Silvia-Iris Beutel

Alternative Leistungsbeurteilung als Entwicklungsaufgabe von Schule

1. Die Fragestellung

Dieser Beitrag fragt danach, wie die anhaltende Notendebatte in den Schulen zugunsten von Alternativen überwunden werden kann. Und wie können dabei wissenschaftliche Begründungen, bereits vorhandene schulpraktische Erfahrungen mit bildungspolitischen Anregungen und Forderung einhergehen?

In Blick auf die Allgemeine Didaktik und Unterrichtsforschung hat die Suche nach modernen, angemessenen und die Effizienz des Lernens sowie die Förderung der Lernenden in den Mittelpunkt stellenden Formen der Leistungsbeurteilung eine zentrale Bedeutung. Dies gilt insbesondere für die Aufbereitung qualitativ hochwertiger didaktischer Arrangements des Unterrichts und weitergehender Lerngelegenheiten sowie die damit verbundenen Aufgaben der dialogisch-partizipativen Begleitung von Kindern und Jugendlichen. Es gilt auch in Hinblick auf die mit der Leistungsbeurteilung stets korrespondierenden Gefahren, Fehler zu vermeiden wie etwa Unangemessenheit von Beurteilungen und Noten, deren fehlende Transparenz oder für die Lernenden nicht nachvollziehbare Begründungen. Damit verbunden sind letztlich fundamentale Ungerechtigkeiten gegenüber den individuell Lernenden und die Gefahr der verminderten Chancen sowie falscher Zuteilungen im Zusammenspiel von Leistungsbeurteilung und Berechtigungswesen.

2. Zur Relativität und Normativität von schulischen Beurteilungen und Bewertungen

Es ist inzwischen anerkannt, dass die Frage der demokratischen Schulentwicklung grundlegend mit der Herausforderung verbunden ist, moderne, diversitätssensible und der Gerechtigkeit und Lernförderung verbundene Formen der Leistungsbeurteilung zu entwickeln, zu implementieren und anzuwenden (Ruberg, Beutel & Espermüller, 2022). Zugleich ist hinreichend erforscht, dass es nicht um die polarisierende Dualität zwischen Ziffernzensur und Formen verbaler Beurteilung allein geht. Auch verbale Beurteilungen sind fehleranfällig, sie erfordern zudem eine besondere Form professioneller Handhabung, kollegialer Standardsetzung und regelhafter Evaluation.

Vielmehr drehen sich die aktuellen Diskurse darum, kontextsensible Lösungen zu erarbeiten, die auf der Professionalität und anwendungsklaren Überzeugung sowie begleitenden Reflexion der beurteilenden Lehrkräfte einerseits sowie der kommunikativen Einbeziehung der Lernenden andererseits basieren.

Dabei sprechen Konzepte und Erfahrungen aus der schulischen Entwicklungsarbeit für sich. Denn Alternativen in Form und Praxis der Leistungsbeurteilung finden an profilierten und qualitätsausweisenden Schulen zunehmend Resonanz. In dieser Form der schulbezogenen Professionalisierung spiegeln sich gesellschaftliche und empirisch fundierte Erwartungen daran, mehr Bildungsgerechtigkeit, Differenzsensibilität und Demokratieerleben in den didaktischen Kern der inklusiven Schulentwicklung zu stellen. Dabei stehen die begleitende Auswertung, Rückmeldung und lernförderliche Beurteilung fachlicher und lernbezogener Leistungen der Schüler*innen im Zentrum solcher Formen der Leistungsbeurteilung.

Entscheidend sind die Einsichten darin, an die Stelle summativ konstruierender Leistungsnachweise Anwendungsmuster formativer Leistungsrückmeldung sowie -beurteilung zu stellen: weg von der curricularen Bezugsnorm als innerem Zentrum schulischer Leistungsmessung hin zu der Beschreibung, Rückmeldung und Auswertung von Lernprozessen gemeinsam und im Widerspiel von Lehrenden und Lernenden. Dabei entstehen differente Instrumente und Anwendungsformen, die sich gleichwohl schwerpunktmäßig auf die individuelle Förderung der Lernenden beziehen – was keinesfalls mit der Dominanz allein einer individuellen Bezugsnormorientierung gleichzusetzen ist.

Das zeigen eindrücklich die Praxen, Instrumente und Anwendungsformen von Bewerberschulen für den Deutschen Schulpreis – insbesondere jener, die im Qualifikationsverfahren dieser Ausschreibung unter die Spitzengruppe kommen. In nahezu allen diesen sehr differenzierten pädagogischen Konzepten gilt der Leistungsrückmeldung und einer differenten Normorientierung eine große Aufmerksamkeit. Das lässt sich sowohl auf die schriftliche Darstellung der pädagogischen Profile solcher Schulen beziehen, als auch auf die kritische Beobachtung und Auseinandersetzung mit deren Praxis bei den qualitätsdokumentierenden Schulbesuchen.

Kollegien, die an dieser Herausforderung über alle Jahrgangsstufen hinweg arbeiten und sich dazu professionell in ihrem Arbeitsalltag organisieren, stellen dabei ein pädagogisches Lern- und Leistungsverständnis heraus, in dessen Kern Individualisierung und Differenzierung stehen. Damit wird zwangsläufig die individuelle Bezugsnormorientierung gestärkt, ohne diese zu verabsolutieren. Vielmehr geht es dabei darum, Lernpotenziale diagnostisch zu erfassen und zu fördern sowie diese Förderung im Team optimieren.

Die biographischen Lern- und Lebenslagen der Schüler*innen werden zum Ausgangspunkt einer integrativ wirkenden Leistungsbeurteilung, die sich unterschiedlicher, zugleich aufeinander abgestimmter Formate der Planung, Selbst- und Fremdrelexion bedient, so können z. B. Logbücher als Planungshilfe und Dokumentationsmappe mit kompetenzausweisenden Reflexionsbögen zusammen eingesetzt und in sog. Bilanz- und Zielgesprächen gemeinsam ausgewertet werden (Beutel & Pant, 2020; Beutel & Xylander, 2021). Die Erprobung von Leistungsvermögen und die Erfahrung der Bedeutung des eigenen Tuns sowie ihrer Lernfortschritte bei den Schüler*innen selbst spielen dabei eine wesentliche Rolle. Das schließt die Diagnose von Lernlücken und die Erarbeitung von Strategien der Lernförderung gemeinsam mit den Lernenden ein. Und es korrespondiert mit einer pädagogischen Praxis und schulischen Lernumgebung, in der vielfältige Lernanlässe geboten werden. In der Folge treten die Prozessqualitäten des Lernens besonders hervor.

Elementarer Bestandteil dabei ist die Dokumentations- und Verständigungssicherheit über Lernwege und diese begründenden Streckenabschnitte, förderlich sind zudem die Einrichtung selbst gewählter Prüfungszeiten. Regelmäßige interne und externe Evaluationen geben Anlass, die Leistung der Schule gerade mit der Perspektive ihrer Wirkung für die Lernenden einzuschätzen. Entsprechende Formen sind beispielsweise ‚Navigationskarten‘, die Lernstände, Lernaufgaben und Lernwege formulieren und dokumentieren. Oftmals gehören auch Hospitationen und Feedbackrunden durch die Schüler*innen für die Lehrenden zur Qualität eines solchen Konzeptes von Unterricht dazu. Diese werden im Idealfall zum Anlass schulinterner Fortbildungen. Sie spielen für die kollegiale Schulentwicklung eine bedeutsame Rolle. So wird eine geteilte Entwicklungsperspektive – eine Art Ko-Intention – zwischen Lehrer*innen und Schüler*innen angelegt. Vielfältige Beispiele zu einer solchen Praxis sind mittlerweile in den Medien verbreitet, anschaulich kommentiert und teilweise auch bereits für die Fortbildungsarbeit, durchaus auch in digitalen Formaten, nutzbar gemacht (Beutel & Ruberg, 2021).

Grundsätzlich kann man aus Sicht der wissenschaftlichen Schulentwicklungsbegeleitung feststellen, dass die Frage einer qualitativ hochstehenden und reflektierten Form der Leistungsbeurteilung heute eine größere Rolle in der Zielbestimmung sowie der Auswahl von Fortbildungsthemen und zugehörigen Methoden spielt, als dies bislang der Fall war. Es wird zudem sichtbar, dass die konfrontativen Zuspitzungen der Diskurse in diesem Feld, wie sie die Reformdiskussionen der 1980er-Jahre und danach bestimmt haben, erheblich reduziert werden konnten. Es ist deutlich geworden, dass eine Positionsnahme zwischen numerischen Beurteilungsformen und verbalen Formen der Leistungsrückmeldung allein keines der normativen und relationalen Probleme dieser pädagogischen Kernaufgabe

mindert. Festgestellt werden kann auch, dass im Rahmen der Stärkung diagnostischer Kompetenzen innerhalb des Professionsverständnisses der Lehrerbildung und des Lehrberufes generell die Notwendigkeit entsteht, dem Thema einer lernförderlichen Leistungsbeurteilung in allen Phasen der Lehrerbildung größere fachliche und wissenschaftliche Aufmerksamkeit zu widmen – ohne dass deshalb das Kernproblem einer fehlenden grundständigen Professionalisierung in diesem Feld schon umfassend behoben wäre.

Schließlich ist seit den 2000er-Jahren auch im Zuge der verstärkten Diskurse und praktischer Modelle zur Demokratisierung der Schule, zu ihrer sozialisatorischen Funktion in Blick auf belastbare demokratische Werthaltungen und eine entsprechende Kompetenzförderung in der Schule – auch über den fachdidaktischen Kern politischer Bildung im engeren Sinne hinaus – die Frage einer gerechten, transparenten, für die Schüler*innen und deren Eltern nachvollziehbaren Beurteilungspraxis verstärkt in den Blick geraten. Dass Leistungsbeurteilung und Notengebung auch immer eine Verführung zur einseitigen Machtausübung sowie eine exkludierende Form der Hierarchisierung innerhalb der Peer-Gruppe von Lernenden mit sich bringt, wird innerhalb der aktuellen Demokratiebildung gesehen und diskutiert (Beutel et al., 2021). Demokratische Schulentwicklung ohne Weiterführung und Differenzierung der angewandten Formen von Leistungsrückmeldung ist nicht darstellbar und praktisch nicht umzusetzen.

3. Ansätze zum Umgang mit aktuellen Herausforderungen: Neue Leistungsbeurteilungssysteme in Schule implementieren

Veränderungen und Neuanlagen von alternativen Leistungsbeurteilungssystemen sowie deren Implementation an Einzelschulen – und auch als Zusammenschluss in entwicklungsstarken Netzwerken – lassen sich seit langem in allen Bundesländern beobachten. In den länderspezifischen Orientierungsrahmen, Erlassen, Verordnungen und Gesetzen werden dazu pädagogische Spielräume mit einer formativen Leistungsbeurteilung und entsprechend kompetenzausweisenden Instrumenten und Verfahren eröffnet. Sie werden dort auch pädagogisch begründet und zur Einführung in Grund- und weiterführenden Schulen empfohlen (Beutel & Xylander, 2021). Eine begleitende Feedbackkultur, die vor allem die Schüler*innen, aber auch Eltern anspricht und alle in verlässliche Mitverantwortung bringt, gehört ebenfalls dazu.

Entgegen naheliegenden, auch wissenschaftlich begründeten Erwartungen hebt diese Erneuerungslinie jedoch nicht die bestehende Praxis der Notengebung auf. Vielmehr zeigt sie sich als Investition in eine Doppelstrategie, die bisher keine dauerhafte und grundsätzliche Revision des gesamten Prüfungs- und

Zertifikatswesens eingeleitet hat. Reformen sind zwar angestoßen, die grundsätzliche Frage nach der Gültigkeit der gängigen Anwendung des bestehenden Leistungsprinzips in der Schule wird jedoch nicht gestellt. Noten werden weiter als unterkomplexe Form der Leistungsbeurteilung akzeptiert, als leicht verstehbare Währung mit geschichtlicher Tradition fortgeschrieben und auch in der notwendigen Aufgabe, mehr Bildungsgerechtigkeit im System differenzierender Schularten und -formen zu erreichen, nicht grundlegend in Zweifel gezogen:

„In der schulischen Praxis wie auch den gesetzlichen Regelungen zur Leistungsbeurteilung, auf die sich die Praxis beziehen soll, ist diese ‚Kontaminierung‘ fast durchgängig angelegt“ (Pant, 2020, S. 52).

Verschiedene Gründe für die Beharrungskraft der Notengebung und ihrer Mängel sind in der Fachliteratur vielfach herausgearbeitet und auch für praxiswirksame Formen der Qualifizierung beschrieben und benannt worden (ebd., S. 22ff.) Aus heutiger Perspektive werden sie neu und kritisch unter Bezugnahme auf das meritokratische Modell gesehen, welches das Versprechen beinhaltet, dass ein jeder in der Gesellschaft durch Leistung zu Ansehen, Geld, Macht und Glück kommen kann. Hinzu kommt die anhaltende Diskussion um dessen Anwendungsform, Wirkkraft und letztlich auch Beitrag zu einer ausgeglichenen Spannung zwischen individuellen Anrechten auf Selbstverwirklichung, Anerkennung und Verdienst einerseits sowie sozialer und bildungsbeeinflusster Chancen- und Lebengerechtigkeit hin andererseits. Zweifel an der Dominanz meritokratischer Leistungsentfaltung auf Basis und in der Folge schulischer Zertifikate sowie damit einhergehender Berechtigungen auf staatliche Studien-, Berufs- und Qualifizierungsgänge gelten verstärkt in Blick auf den Zusammenhang zwischen ungerechter Wirkung schulischer Beurteilungspraxis in Hinblick auf eine auch politisch zwischenzeitlich erwünschte Anerkennungskultur von Leistungsvielfalt in der inklusiven Schule. Pant (ebd.) spricht hier von diffusen Vorstellungen zu Gerechtigkeitsmodellen.

Es ist anerkannt, dass Einzelprüfungen und auf statisches Wissen zielende Überprüfungen von Lernleistungen kaum geeignet erscheinen, um valide Aussagen an den Bildungsübergängen im Schulsystem, aber auch zu den Anrechten auf Hochschulbildung zu treffen oder gar auf eine moderne Arbeitswelt in einer pluralen Gesellschaft, auf Team- und Diskursfähigkeit, auf Kontroversenmoderation und Kompromissuche hin vorzubereiten: Mehr denn je ist deutlich, dass flexibles Anwendungssetting und kreative Lösungsfähigkeiten bei komplexen Problemlagen andere Kompetenzmodellierungen ansprechen als Formen statischen Wissens, auch „[...] die sozialen und emotionalen Kompetenzen müssen sich in den Beurteilungen widerspiegeln“ (Schleicher, 2019, S. 332).

Die Zukunftsagenda 2030 (BMWZE, 2017) mit den „17 globalen Zielen für eine bessere Zukunft“ der Vereinten Nationen gilt als Leitlinie für Bildung, Lern- und Leistungskonzepte in der verbundenen, globalisierten und digitalisierten Welt. Sie lässt Mehrdeutigkeit im Lernen zu, verlangt reflektierte Erkenntnis und Nachhaltigkeit und die Förderung von Resilienz angesichts stetiger und stets wiederkehrender Ungewissheit. Dies ist auch eine gesellschaftspolitische, die Demokratie bestandsfähig machende Aufgabe neuer Leistungsbeurteilungssysteme. Resilienzförderung und reflektierter Umgang mit Ungewissheiten verändern eben auch die curricularen Erwartungen an schulisches Lernen – und setzen damit auch auf Ebene dieser Bezugsnormen andere Orientierungspunkte und Maßstäbe.

Nicht zuletzt hat die weltweite pandemische Lage in der Covid-Pandemie erkennen lassen, dass Lernen nicht länger in ausschließlicher Gebundenheit an Schule als Ort und in Jahrganggruppen zu organisieren ist. Vielmehr können vielfältige Lernumgebungen genutzt werden. Neue individuelle Lern- und Denkräume sowie virtuelle Gruppenräume sind entstanden, die über die Beziehungs- und Kooperationspflege in der eigenen Schule weit hinausgehen können und damit auch bestehende Curricula im Sinne der Überwindung alleinigen Nachvollzugs modernisieren helfen. Orientierung in den analogen und digitalen Lernstraßen geben aufeinander bezogene Planungs- und Reflexionsinstrumente, Kompetenzraster mit Lerninhalten und Wahlniveaus – damit wächst zugleich aus den neuen Foren von Örtlichkeit des Lernens sowie Kommunikation und Mobilität ein höheres Maß an Mitbestimmung bei der Leistungsbeschreibung durch die Lernenden selbst.

Das äußert sich in einer variantenreichen Fülle kreativer Formen, die es anzuwenden gilt: Checklisten für Online-Klausuren, Portfolios für Präsentationsprüfungen, ‚Frei-Days‘ mit Projektarbeit und agilem digitalem Management, Bögen für formative Evaluationen und ausblickhaftes Lernfeedback können eingesetzt werden. Eine vertikale und horizontale Individualisierungs- und Differenzierungsorganisation mit tutorieller Begleitung sichert den notwendigen stillen Lernrückzug. Ein Time-Out *ebenso* wie Versionen des Peer-Learnings als Helfer-, Markt- und Expertensystem (Wehe, 2021) im open classroom stützen diesen Formenreichtum. Für die neue Art der Leistungsbeurteilung prägend ist dann eine Überwindung der Bindung von Lernen, vom Nachvollzug und der Gleichzeitigkeit von Prüfungssituationen. Flexibilisierung und Entwicklungsorientierung rücken dann in den Mittelpunkt von Bildungserfahrungen in einem starken Sozial- und Förderraum.

4. Und was folgt daraus?

Von der Bildungspolitik ist deshalb in naher Zukunft erwartbar, dass sie Schulen, die in eine formativ-lernförderliche und partizipative notenfreie Leistungsbeurteilung durch nachgewiesene Konzeptstärke nicht nur für einzelne Jahrgänge oder Lerngruppen investieren und die zugleich gute Implementationsstrategien haben, eine Erweiterung ihrer Selbstverwaltung ermöglichen und an diese eine verbindliche Rechenschaftslegung anknüpfen. Auch aus der Perspektive der Leistungsbeurteilung ist die Stärkung der Autonomie und pädagogischen Freiheit der Einzelschule das Mittel der Wahl. An beidem – Selbstverwaltung und Rechenschaftslegung – kann den an einer gerechteren Lernzukunft für alle Schüler*innen orientierten Schulen nur selbst gelegen sein. Sie sollten dabei von einer beratenden und forschungsseitigen Unterstützung profitieren können, von der Schulaufsicht interessiert, aber mit dem Zutrauen zu dieser pädagogischen Freiheit begleitet und von den Kommunen unterstützt werden.

5. Literatur

Beutel, S.-I. & Pant, H.A. (2020). *Lernen ohne Noten. Alternative Konzepte der Leistungsbeurteilung*. Stuttgart: Kohlhammer.

Beutel, S.-I. & Ruberg, C. (2021). Chancen einer anderen Leistungsbeurteilung im Distanzlernen. *Bildung+Schule digital*, 1, 6-9.

Beutel, S.-I. & Xylander, B. (2021). Gerechte Leistungsbeurteilung. Impulse für den Wandel (Reihe Bildung und Unterricht). Ditzingen: Reclam.

Beutel, S.-I., Beutel, W. & Gloe, M. (2021): Demokratische Schulentwicklung. *Wochenschau Politik und Wirtschaft unterrichten: Demokratiebildung*, 72(21s), 14-20.

Pant, H. A. (2020). Notengebung, Leistungsprinzip und Bildungsgerechtigkeit. In S.-I. Beutel & H. A. Pant, *Lernen ohne Noten. Alternative Konzepte der Leistungsbeurteilung* (S. 22-58). Stuttgart: Kohlhammer.

Ruberg, C., Beutel, S.-I. & Espermüller-Jug, E. (2022). Partizipative Leistungsbeurteilung. In W. Beutel, M. Gloe, M., G. Himmelmann, D. Lange, V. Reinhardt & A. Seifert (Hrsg.), *Handbuch Demokratiepädagogik* (S. 758-770). Frankfurt a.M.: Wochenschau Verlag.

Schleicher, A. (2019). *WELTKLASSE. Schule für das 21. Jahrhundert gestalten*. Bielefeld: wbv-Publikationen.

Wehe, I. (2021). Das Helfersystem in den Lernzeiten. Förderung sozialer Kompetenzen an der Ganztagschule. In A.-M. Davis, N. Fischer, R. Mounajed, K.-J. Tillmann & W. Vogelsaenger (Hrsg.), *Friedrich Jahresschrift 39: Chance Ganztage* (S. 110-111). Hannover: Friedrich.

Angaben zur Autorin:

- Prof. Dr. phil. habil. Silvia-Iris Beutel, Professorin für Schulpädagogik und Allgemeine Didaktik, TU Dortmund, Institut für Allgemeine Didaktik und Schulpädagogik, silvia-iris.beutel@tu-dortmund.de

Philippe Wampfler und Christian Albrecht

Zugänge zu einer zeitgemäßen Prüfungskultur

1. Einleitung

Im Rahmen der digitalen Transformation der Gesellschaft haben sich in den letzten 20 Jahren neue Rollen- und Berufsbilder entwickelt, die sich stark auf Vorstellungen der Zusammenarbeit und des Lernens auswirken. Bildung sollte im 21. Jahrhundert junge Menschen dazu ermächtigen, „Gesellschaft aktiv mitzugestalten“ (Wagner, 2021, S. 11). Wie Wagner (ebd.) ausführt, kann eine an Zertifikaten und Wissensvermittlung orientierte Bildung diese Funktion nicht wahrnehmen. Die etablierte Prüfungs- und Bewertungskultur ist statisch und isoliert Lernende – primär deshalb, weil sie gesellschaftlich der Allokation dient.

Lernen ist aber ein dynamischer, kollektiver Prozess. Da Prüfungen Schulen stark prägen – Unterricht soll Lernende befähigen, Prüfungen erfolgreich zu absolvieren – ist ein Wandel der Prüfungskultur Bedingung dafür, eine Lernkultur zu etablieren, die jungen Menschen erlaubt, aktuelle und zukünftige gesellschaftliche, wirtschaftliche und ökologische Herausforderungen zu bewältigen.

Das Institut für zeitgemäße Prüfungskultur hat Merkmale einer Prüfungs- und Bewertungskultur festgehalten, welche geeignet sind, Bildungsprozesse zu initiieren und zu stützen (Institut für zeitgemäße Prüfungskultur, 2021): Verdichtet geht es darum, Schüler*innen aktiv in die Frage der Leistungsbeurteilung einzubeziehen. Indem sie über Kriterien und Lernprozesse nachdenken, ihre eigene Arbeit selbst beurteilen und Feedback von Lehrpersonen und Peers aktiv einfordern, können sie entsprechenden Beurteilungsprozessen einen Sinn für ihr eigenes Lernen abgewinnen. Sie verstehen und erleben, wie Rückmeldungen ihnen dabei helfen, Lernprozesse zu verbessern – und arbeiten grundsätzlich in kooperativen Settings.

2. Herausforderungen bei der Etablierung einer zeitgemäßen Prüfungskultur

2.1. Trennung von Lernen und Leisten

Konzeptionell und rhetorisch kann die Tatsache, dass summative Prüfungen viele Bedingungen und Eigenschaften realer Lernprozesse ausblenden, dadurch verdeckt werden, dass Lernen von Leisten bzw. Lernaufgaben von Leistungsaufgaben getrennt werden (Büchter & Leuders, 2006). Produktorientierte Aufgaben,

die unter Zeitdruck und externen Vorgaben isoliert bearbeitet werden müssen, werden Leistungsaufgaben genannt und von Lernaufgaben unterschieden, die sich in einen Lernprozess integrieren lassen. Leistungsaufgaben werden summativem, Lernaufgaben formativem Feedback zugeordnet.

Das hinter summativen Prüfungsformaten stehende Konzept, Fähigkeiten und Fertigkeiten einer Person in situativer Verdichtung zur sicht- und bewertbaren Entfaltung bringen zu können, erweist sich insbesondere im Kontext der Kompetenzorientierung als problematisch. Kompetenzen lassen sich nämlich nicht an isolierten Schülerleistungen ablesen (vgl. Winter, 2014, S. 58), sondern als entsprechender Outcome nur in einem mehrdimensionalen und zeitextensiven Beobachtungsprozess ermitteln, der weitere Lerndokumente, Gespräche mit den Lernenden sowie deren Reflexionen und Selbsteinschätzungen einschließt (ebd.). Ebenso erweist sich diese enge Definition eines sogenannten an den Lernprozess angedockten Leistungsraumes als zu hermetisch. Eine mit Zeitdruck und Fehlervermeidung assoziierte Prüfungssituation unterminiert ein förderliches Lernklima (vgl. Berner, 2021, S. 64) und reduziert darüber hinaus die Bedeutung der vielen Varianten formativer Leistungsbeurteilung (Winter, 2018). Diese Probleme können durch konzeptionelle oder begriffliche Sonderstellungen von Leistungsaufgaben nicht gelöst werden – Leisten lässt sich von Lernen nicht trennen (Zumbansen, 2021). Gemeint sind hier eigentlich Prüfungsaufgaben, die den Anforderungen einer an sozialer Selektion orientierten Prüfungskultur genügen, nicht aber die Leistung von Lernen erfassen.

2.2. Eigenständigkeit der Leistung

Die Vorstellung, eine Leistung müsste einem Individuum zurechenbar sein, hat die aktuelle Prüfungskultur geprägt: Wer geprüft wird, arbeitet allein und ohne Möglichkeiten, über ein vordefiniertes Minimum hinaus mediale Inhalte hinzuziehen zu können. Auch jede Form von Kommunikation ist ausgeschlossen.

In einer vernetzten und digital geprägten Gesellschaft allerdings spiegelt eine solche Prüfungspraxis die Lernwirklichkeit einer von Gemeinschaftlichkeit und Referentialität geprägten Kultur der Digitalität (Stalder, 2017) nicht länger angemessen wider: Leistung wird heute in einem sozialen, interaktiven Prozess erbracht, in dem die auf Expertentum fokussierte Priorität des Individuellen einer „Priorität des Sozialen“ (Reusser, 2006, S. 155) und der Arbeit in kollektiven Netzwerken weicht, sodass insbesondere die Leistung des interpersonalen sozialen Systems (vgl. Giesecke, 2005, S. 24) Gegenstand von Prüfungssituationen sein muss. In diesem Zusammenhang ist oft die Rede von den 4K: Sie bezeichnen die vier ‚21st Century Skills‘ Kommunikation, Kollaboration, Kritisches Denken und Kreativität und die mit

ihnen verbundenen, mitunter überfachlichen Wissensbestände, Fertigkeiten, Einstellungen, Werte und Haltungen (Binkley et al., 2012; Fadel et al., 2015, S. 72ff.; Thornhill-Miller et al., 2023). Diese nicht-automatisierbaren Kompetenzen sind im Alltag und in der Berufswelt des 21. Jahrhunderts von herausragender Bedeutung. Geht es also darum, Leistungen zu beurteilen, dann müssten diese Dimensionen berücksichtigt werden (Albrecht, 2021; 2022).

Der Mythos, Leistungen würden eigenständig erbracht, verhindert das. Er blendet alles aus, was Kollaboration oder Kommunikation explizit erfordern würden. Tatsächlich gibt es keine isoliert erbrachten Lernerfolge: Zum einen grundiert Leistung jedes Menschen auf sozialen Parametern, die sich nicht unmittelbar auf das Individuum zurückführen lassen (vgl. Verheyen, 2018, S. 199ff.), zum anderen wird Leistung stets in sozialen Kontexten erbracht. Denn wer lernt, lernt von und mit anderen Menschen. Wenn übliche Prüfungsformate nun implizieren, dass Leistungen isoliert erbracht werden könnten, dann führt das zu einer massiven Verzerrung und zu einer Einseitigkeit, die insbesondere einen zeitgemäßen Einsatz digitaler Medien und Kulturtechniken verunmöglicht, da deren Affordanz 4K-Arbeitsformen voraussetzen.

2.3. Normen und Vergleiche

Viele Prüfungen orientieren sich am Ideal einer standardisierten Bewertung: Die Beurteilung sollte Lernenden grundsätzlich ermöglichen, sich im Vergleich mit anderen ein Bild darüber zu verschaffen, wie gut sie vorgegebene Erwartungen erfüllen. Die wenigen vorliegenden empirischen Befunde zeigen, dass dieses Ideal deutschlandweit kaum eine Realität ist: Aufgabenformate, Bewertungen und Erwartungshorizonte differieren; Referenzgruppeneffekte verzerren Beurteilungen; leistungsfremde Einflüsse wirken sich auf die Bewertung aus – um nur eine Auswahl der Probleme zu benennen (ein Überblick über die Forschung findet sich bei Kötter-Mathes & van Ackeren, 2022, S. 25).

Geht es nun bei der Entwicklung einer zeitgemäßen Prüfungskultur unter anderem darum, kollaborative Lernprozesse in die Leistungsbeurteilung einzubeziehen und möglichst sinnvolle Rückmeldungen für Lernende zu generieren, so ist die Orientierung an der Vorstellung, Bewertungen müssten primär Vergleichbarkeit herstellen, problematisch. Diese Vergleichbarkeit wird in Bezug auf die Allokationsfunktion von Bewertungen als Rangfolge von einzelnen Schüler*innen verstanden. Eine Beurteilungspraxis, die primär dazu dient, individuelles Feedback zu ermöglichen, kann keine Rangfolge herstellen. Gleichwohl besteht diese Erwartung, weil die etablierten Formen des Prüfens mit scheinengenauen Prüfungsnoten (vgl. Nölte & Wampfler, 2021, S. 39) dazu mathematisch in der Lage sind.

2.4. Allokation

Alle bereits diskutierten Probleme haben damit zu tun, dass Prüfungsergebnisse für Allokationsprozesse genutzt werden. Die meritokratische Vorstellung, wonach die erbrachte Leistung der Lernenden Ausgangspunkt für soziale Selektion darstellen soll, führt zu Prozessen, welche diese Leistung möglichst objektiv mit Ziffernnoten als statistische Basis messen sollen.

Diese Prozesse geraten schnell mit einer anderen Funktion der Schule in Konflikt, nämlich Kinder und Jugendliche beim Lernen zu unterstützen. „Lernen muss nicht bewertet werden“ (ebd., S. 9) – diese Grundeinsicht stimmt im aktuellen Bildungssystem nur deshalb nicht, weil der Schule die Allokationsfunktion zugewiesen wird. Sie schädigt Kinder direkt, wie Stojanov (2011, S. 106) festgestellt hat:

„Die Folge dieser planwirtschaftlich ausgerichteten Forderung für die Betroffenen ist, dass sie schon vor ihrem eigentlichen Eintritt in das Arbeitsleben als Menschen mit bestimmten Kompetenzen, Fähigkeitspotenzialen oder Begabungen festgelegt werden. [D]iese Festlegung von außen [unterdrückt] in vielen Fällen das Selbstwertgefühl und letztlich die Autonomiefähigkeit des Einzelnen.“

Allokation schadet – sie ist auch der Grund für alle Missverständnisse und Probleme rund um die aktuell vorherrschende Prüfungskultur (Stojanov, 2022). Vergleiche von Individuen mit scheinengenauen, summativen Werten braucht es deshalb, um Menschen klassifizieren zu können. Summative Prüfungsverfahren sind nicht dazu da, um Lernenden zu helfen.

2.5. Objektivität und Messung

Bereits das im Prüfungskontext gebräuchliche Wort der ‚Leistungsmessung‘ verweist darauf, an welchem Modell sich die traditionelle Prüfungskultur orientiert: am Modell der experimentellen Wissenschaften. Der Schlüsselbegriff des experimentellen Modells ist die Relation des Dokumentierens: In einem Experiment werden Instrumente in einen Ausschnitt der Natur eingebracht. Sie ermöglichen durch ihre Konstruktion, dass sie nicht beobachtbare Entitäten dokumentieren und für den Menschen erkennbar und – im Falle von Messungen – ablesbar machen. Instrumente eines naturwissenschaftlichen Experiments reagieren kausal auf das nicht Beobachtbare und dokumentieren in dieser Reaktion für die Beobachtenden dasjenige, was ihnen ansonsten nicht zugänglich wäre.

Die hinter diesem Modell stehenden Annahmen sind in Bezug auf Lernprozesse falsch oder problematisch: Erstens ist der Lernprozess, der vermessen werden

soll, durch die Messung selbst beeinflusst. Prüfungen sind Orientierungspunkt für Lernprozesse, sie stören diese gleichsam, wenn *teaching to the test* oder *learning for the test* stattfindet. Zweitens messen Prüfungen nicht objektiv, denn eine Reihe von Verzerrungen und Fehlern beeinflussen die Vergabe von Noten (Nölte & Wampfler, 2021, S. 42ff.). Drittens beziehen sich Prüfungen meist auf (hand-)schriftliche Lernprodukte, in denen sich nur Ausschnitte von Lernprozessen zeigen können. Und viertens ist es möglich, Prüfungsaufgaben auch dann schematisch zu lösen, wenn entsprechende Lernprozesse gar nicht stattgefunden haben.

Konventionelle Prüfungen messen aus diesen Gründen im besten Falle, wie gut Schüler*innen im Moment der Prüfung vorgegebene Aufgaben im Vergleich mit einem Erwartungshorizont lösen können. Sie sind kein taugliches Mittel, um Rückschlüsse auf Lernprozesse zu ziehen.

3. Lösungsansätze

Die in Abschnitt 2 kritisierte Prüfungskultur ist über rechtliche Vorgaben tief ins Bildungssystem eingeschrieben. Um diese Vorgaben herum haben sich Praktiken und professionelles Handlungswissen etabliert, an denen sich Schulen und Lehrpersonen bei der Bewertung von Schüler*innen orientieren. Die Vision einer zeitgemäßen Prüfungskultur, die Lernenden primär dabei hilft, über vielfältige formative Rückmeldungen über ihren Lernprozess nachzudenken, Wertschätzung für ihre Leistungen zu erhalten und ihre Perspektiven durch kooperative Arbeitsformen und dialogische Lernreflexion zu erweitern, reicht nicht aus, um Vorgaben auszuhebeln und traditionsreiche Prüfungen zu verbessern. Hinzu kommen psychologisch-systemische Faktoren, die Winfried Kronig bei der Beantwortung der Frage, weshalb Jahrzehnte von wissenschaftlichen Belegen, dass Schulnoten Leistungen nicht messen, keine Veränderung bewirkt haben, wie folgt formuliert:

„Die gegenwärtige Praxis der Leistungsbeurteilung ist für das System funktional und organisatorisch sinnvoll. Sie erfüllt das Bedürfnis der Verwaltung und Lenkung von immer wieder neuen Schülerströmen in regional unterschiedlich ausgestaltete Strukturen. Dies zwingt die Schule dazu, die leistungsgerechte Notenvergabe eher zu inszenieren, als sich nach ihr zu richten. Die Schule kann gar nicht anders, als den von ihr gefällten Notentscheidungen eine rituelle Glaubwürdigkeit zu verleihen, wenn sie nicht selbst unglaubwürdig werden will.“ (Kronig, 2022, S. 31)

Diese Perspektive führt zur grundsätzlichen Frage, ob ein Wandel des Systems überhaupt denkbar ist, ob es Alternativen gibt, die in Bezug auf die heute noch

nötige Allokation und Verwaltung von Schüler*innen ähnliche Funktionen übernehmen können wie die Bewertung von Schüler*innen über Prüfungen und Noten. Der Ansatz dürfte über die Glaubwürdigkeit und die Akzeptanz von Noten führen. Wenn Formen der Beurteilung für Lernende einen Sinn erhalten, indem sie sich auf Lernprozesse beziehen und nachvollziehbar dazu beitragen können, dass diese Prozesse zu besseren Ergebnissen führen, dann können damit verbundene Bewertungen die Glaubwürdigkeit von Schulen und Lehrpersonen gegenüber dem aktuellen System steigern.

Dadurch ergibt sich ein weiterer Lösungsansatz: Individuelle Formen der Bewertung. Lernende können beispielsweise mitentscheiden, welche Kriterien für ihre Arbeitsformen überhaupt relevant sind, für welche Aspekte ihnen Rückmeldungen weiterhelfen könnten. Dadurch wird der Eindruck zerstreut, Bewertungen könnten eine objektive Vergleichbarkeit herstellen. Mit Bezug auf die (grundsätzlich fragwürdige) Allokationsfunktion würden Lernende so in diesen Prozess miteinbezogen: Durch die individuelle Wahl von Projekten und Kriterien zeigen sie, welche Ziele sie anstreben und welche Ausbildungswege oder Berufschancen sie sich erschließen möchten. Lernstandsgespräche müssten hier Transparenz schaffen und Schüler*innen darüber informieren, wie die Bewertung ihrer Arbeit sich auf ihre Zukunft auswirkt. Besonders für bildungsbenachteiligte Schüler*innen ist es dabei von Bedeutung, individuelle Stärken, Potenziale und Möglichkeiten auch jenseits innerer und äußerer Schranken aufzuzeigen, um primären und sekundären Herkunftseffekten (vgl. Pfeifer, 2011, S. 59) entgegenwirken zu können.

Ein letzter Lösungsansatz, der hier kurz vorgestellt werden kann, betrifft die Eigenaktivität von Lernenden. Aktive Lernende setzen sich eigene Ziele, deren Erreichung sie mitbeurteilen können. Projektergebnisse oder Portfolios wären der zeitgemäße Ersatz für Prüfungen an sich. Sie stellen Stärken und das Engagement von Schüler*innen in den Mittelpunkt und erlauben kooperative Arbeitsformen in einer Kultur der Digitalität.

Gleichwohl werden sie unter den aktuellen politischen Rahmenbedingungen wieder zu einer Bewertung führen müssen, die mit Vergleichen und Selektion verbunden ist. Solange sich das nicht ändern lässt, werden Lehrpersonen auch immer wieder auf traditionelle Prüfungsformen zurückgreifen, weil sie effizienter sind: Sie lassen sich mit Disziplinierung von Schüler*innen verbinden und entlasten Lehrende davon, Beurteilungen begründen und individuell sinnvoll gestalten zu müssen. Die Vorstellung, mit Lernenden individuelle Zielvereinbarungen zu treffen und sie bei Lernprozessen mit formativem Feedback zu begleiten, ist für viele Lehrkräfte mit einem Mehraufwand verbunden. Lösungsansätze müssen deshalb einerseits eine Entlastung bieten – z. B. von monotonen Korrekturarbeiten unter Zeitdruck –, andererseits auch eine magnetische Wirkung entfalten. Die Akzeptanz

durch Lernende und die damit verbundene Glaubwürdigkeit Lehrender, die sich nicht hinter Prüfungen und Erwartungshorizonten verstecken müssen, könnte in der Vision einer zeitgemäßen Prüfungskultur eine solche Wirkung haben.

Dieser Text stützt sich auf Vorarbeiten von Lars Zumbansen und Lars Mecklenburg (Institut für zeitgemäße Prüfungskultur).

4. Literatur

- Albrecht, C. (2021). Prüfungsformate im digitalen Wandel. In W. Klee, Ph. Wampfler & A. Krommer (Hrsg.), *Hybrides Lernen. Zur Theorie und Praxis von Präsenz- und Distanzlernen* (S. 130-146). Weinheim: Beltz.
- Albrecht, C. (2022). *Auf dem Prüfstand. Plädoyer für eine zeitgemäße Prüfungskultur*. <https://blog.pruefungskultur.de/2022/02/01/auf-dem-pruefstand-plaedoyer-fuer-ein-zeitgemaesse-pruefungskultur/> [25.01.2023].
- Berner, N. (2021). Lernen begleiten. Lernprozesse im Kunstunterricht: beraten und fördern. *Kunst + Unterricht*, 457-458, 58-64.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M. & Rumble, M. (2012). Defining Twenty-First Century Skills. In P. Griffin, B. McGaw & E. Care (Hrsg.), *Assessment and Teaching of 21st Century Skills* (S. 17-66). Dordrecht: Springer.
- Büchter, A. & Leuders, T. (2006). Was ist eine gute Aufgabe? Das kommt darauf an! *Praxis der Naturwissenschaften – Chemie in der Schule*, 55(8), 9-15.
- Fadel, C., Bialik, M. & Trilling, B. (2015). *Four-Dimensional Education: The Competencies Learners Need to Succeed*. Boston: Center for Curriculum Redesign.
- Giesecke, M. (2005). Auf der Suche nach posttypographischen Bildungsidealen. *Zeitschrift für Pädagogik*, 51(1), 14-29.
- Institut für zeitgemäße Prüfungskultur (2021). *Vorteile einer zeitgemäßen Prüfungskultur*. <https://pruefungskultur.de/vorteile.html>
- Kötter-Mathes, S. & van Ackeren, I. (2022). Vergleichbare Prüfungsanforderungen und Noten? Empirische Befunde zur bundesweiten Vergleichbarkeit zentraler Prüfungsstrukturen. *Friedrich Jahresheft*, 40, 24-25.
- Kronig, W. (2022). Wenn die Schulklasse die Note mitbestimmt. Über die Bedeutung des Referenzgruppenfehlers bei Karlheinz Ingenkamp. *Friedrich Jahresheft*, 40, 26-31.
- Nölte, B. & Wampfler, Ph. (2021). *Eine Schule ohne Noten*. Bern: HEP.
- Pfeifer, M. (2011). *Bildungsbenachteiligung und das Potenzial von Schule und Unterricht. Lesekompetenz bei sozioökonomisch benachteiligten Schülern*. Wiesbaden: VS Verlag.
- Reusser, K. (2006). Konstruktivismus – vom epistemologischen Leitbegriff zur Erneuerung der didaktischen Kultur. In M. Baer, M. Fuchs, P. Füglistner, K. Reusser & H. Wyss (Hrsg.), *Didaktik auf psychologischer Grundlage. Von Hans Aebli's kognitionspsychologischer Didaktik zur modernen Lehr- und Lernforschung* (S. 151-68). Bern: HEP.
- Stalder, F. (2017). *Kultur der Digitalität* (2. Aufl.). Berlin: Suhrkamp.
- Stojanov, K. (2011). *Bildungsgerechtigkeit. Rekonstruktion eines umkämpften Begriffs*. Wiesbaden: VS Verlag.
- Stojanov, K. (2022). Warum "Leistung" kein tragendes Prinzip von Bildungsgerechtigkeit sein kann. *Friedrich Jahresheft*, 40, 20-21.

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., et al. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence*, 54(11), 1-32. <https://doi.org/10.3390/jintelligence11030054>.

Verheyen, N. (2018). *Die Erfindung der Leistung*. Bonn: Bundeszentrale für politische Bildung.

Wagner, A.C. (2021). *Berufen statt zertifiziert. Neues Lernen, neue Chancen*. Bern: HEP.

Winter, F. (2014). *Leistungsbewertung. Eine neue Lernkultur braucht einen anderen Umgang mit den Schülerleistungen*. Baltmannsweiler: Schneider Verlag Hohengehren.

Winter, F. (2018). *Lerndialog statt Noten. Neue Formen der Leistungsbeurteilung*. Weinheim: Beltz.

Zumbansen, L. (2021). Lernen und Bewerten. Leistungsbeurteilung im Kunstunterricht. *Kunst+Unterricht*, 457-458, 65-70.

Angaben zu den Autoren:

- Dr. Christian Albrecht, Wissenschaftlicher Mitarbeiter am Lehrstuhl für Didaktik der deutschen Sprache und Literatur, FAU Erlangen-Nürnberg, Mitglied des Instituts für zeitgemäße Prüfungskultur, christian.albrecht@fau.de
- Philippe Wampfler, Dozent für Fachdidaktik Deutsch an der Universität Zürich, Mitglied des Instituts für zeitgemäße Prüfungskultur, philippe.wampfler@uzh.ch

Jan-Philipp Burde und Walther Paravicini

Mehr als Fakten und Formeln – Normativität im Kontext der Fächer Mathematik und Physik

1. Einleitung

Den ‚harten‘ Fächern Mathematik und Physik haftet in der Schule oftmals das Stigma an, es ginge nur um ‚Fakten und Formeln‘. Vor diesem Hintergrund soll im vorliegenden Beitrag die Bedeutung der Normativität u. a. in Hinblick auf das Wechselspiel von Unterrichtsgestaltung und Bewertungssituationen in den beiden Fächern näher beleuchtet werden.

2. Normen im Fach Physik: Die vier Kompetenzbereiche

Für das Fach Physik soll im Folgenden der Zusammenhang zwischen normativer Festlegung der Unterrichtsinhalte und deren schulischen Bewertung z. B. im Rahmen von Klassenarbeiten diskutiert werden. Dies geschieht u. a. vor dem Hintergrund, dass „im [bisherigen] Physikunterricht eher der Erwerb von Fakten, Gesetzen und Formalismen betont [wurde]“ (Wiesner et al., 2018, S. 20), der Anspruch des Physikunterrichts sich nicht zuletzt seit der mit den Bildungsstandards eingeführten Kompetenzorientierung jedoch nicht mehr auf die alleinige Vermittlung von Fach- und Formelwissen beschränkt. Vielmehr wird sein normatives Ziel darin gesehen, im Sinne von Scientific Literacy zu einer naturwissenschaftlichen Grundbildung beizutragen (Bybee, 2002), die ein Verständnis der Denk- und Arbeitsweisen der Physik, der Gültigkeit und Grenzen physikalischer Theorien und Modelle sowie der gesellschaftlichen Bedeutung der Physik einschließt (vgl. Wiesner et al., 2018, S. 16). Neben dem inhaltsbezogenen Kompetenzbereich „Fachwissen“ werden daher entsprechend der KMK-Beschlüsse auch die prozessbezogenen Kompetenzbereiche „Erkenntnisgewinnung“, „Kommunikation“ und „Bewertung“ als normativ zentral für die Gestaltung des Physikunterrichts erachtet. Nimmt man für jeden dieser vier Kompetenzbereiche an, dass diese sich in drei Anforderungsbereiche unterteilen lassen, so lässt sich diese normative Setzung in eine Kompetenzmatrix entsprechend Tab. 1 überführen (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2004).

3. Kompetenzorientierung in Bewertungssituationen

Aus unterrichtspraktischer Sicht besteht die Herausforderung nun darin, die verschiedenen durch die KMK für die Physik normativ festgelegten Kompetenzbereiche in schulischen Bewertungssituationen auch in angemessener Breite und Tiefe abzubilden. Als „Spiegelbild des Unterrichts“ (Wiesner et al., 2018, S. 6) gilt dies in besonderem Maße für Klassenarbeiten, die eine sehr weit verbreitete Form der Leistungsfeststellung an Schulen darstellen. Um sicherzustellen, dass das gesamte unterrichtliche Kompetenzspektrum in angemessener Weise im Rahmen der Leistungsdiagnose abgedeckt wird, bietet sich grundsätzlich ein Rückgriff auf die in Tab. 1 dargestellte Kompetenzmatrix an. Mit ihrer Hilfe können Defizite wie z. B. eine einseitige Fokussierung auf das traditionell zentrale und verhältnismäßig leicht überprüfbare „Fachwissen“ in Klassenarbeiten leicht erkannt werden (vgl. ebd., S. 66). Im Folgenden sollen nun exemplarisch einige spezifische Herausforderungen beleuchtet werden, die sich bei der schulischen Überprüfung der normativ festgelegten Kompetenzbereiche „Fachwissen“, „Erkenntnisgewinnung“ und „Bewertung“ ergeben.

		Kompetenzbereich			
		Fachwissen	Erkenntnisgewinnung	Kommunikation	Bewertung
Anforderungsbereich	I	Wissen wiedergeben	Fachmethoden beschreiben	Mit vorgegebenen Darstellungsformen arbeiten	Vorgegebene Bewertungen nachvollziehen
	II	Wissen Anwenden	Fachmethoden nutzen	Geeignete Darstellungsformen nutzen	Vorgegebene Bewertungen beurteilen und kommentieren
	III	Wissen transferieren und verknüpfen	Fachmethoden problembezogen auswählen und anwenden	Darstellungsformen selbständig auswählen und nutzen	Eigene Bewertungen vornehmen

Tab. 1. Kompetenzmatrix bestehend aus Kompetenz- und Anforderungsbereichen in Anlehnung an die KMK-Beschlüsse (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2004, S. 13f.)

Im Kompetenzbereich „Fachwissen“ sollte u. a. darauf geachtet werden, dass die reine Reproduktion und Anwendung von Faktenwissen sowie physikalische Rechenaufgaben, die primär auf formale Lösungsroutinen abzielen, keinen zu großen Stellenwert einnehmen. Auch wenn sich Aufgaben, die vor allem auf die Berechnung von Zahlenwerten durch Termumformungen abzielen, leicht korrigieren lassen, besteht hier die Gefahr, dass korrekte Zahlenwerte „[...] bei den Lernenden ebenso wie bei den Lehrkräften ein inhaltliches Verständnis vor[täuscht], das möglicherweise gar nicht gegeben ist“ (Schecker & Duit, 2018,

S. 5). Die Herausforderung in diesem Kompetenzbereich besteht also darin, Aufgaben zu finden, die zum Nachdenken über physikalische Zusammenhänge anregen und auf das qualitativ-konzeptionelle Verständnis der Lernenden abzielen – idealerweise mit Lebensweltbezügen (Schecker & Klieme, 2001). Bei der Konzeption entsprechender Aufgaben bietet es sich an, die Ergebnisse der Schülervorstellungs- sowie Interessensforschung zu berücksichtigen (Hoffmann et al., 1998; Schreiner & Sjøberg, 2004; Schecker et al., 2018).

Als „wesentliches Merkmal naturwissenschaftlichen Arbeitens“ (Tesch & Duit, 2004, S. 51) wird der Physikunterricht entscheidend durch das Experiment geprägt. Konkret konnten Studien zeigen, dass ca. ein Drittel der Unterrichtszeit mit der Durchführung von Experimenten verbracht wird und sich inklusive Vor- und Nachbereitung ganze zwei Drittel der Unterrichtszeit um das Experiment drehen (Tesch & Duit, 2004). Entsprechend der herausragenden Bedeutung des Experiments nimmt dieses auch im Kompetenzbereich „Erkenntnisgewinnung“ neben der Nutzung von Modellen eine zentrale Stellung ein. Vor diesem Hintergrund stellt sich die Frage, welche Rolle das Experiment als Teil des Kompetenzbereichs „Erkenntnisgewinnung“ in schulischen Bewertungssituationen wie z. B. Klassenarbeiten spielt bzw. spielen kann. Eine praktische Herausforderung liegt dabei z. B. in der Organisation und Bereitstellung entsprechender Materialien in Klassenstärke, um die Durchführung von Versuchen z. B. im Rahmen von fachpraktischen Aufgabenstellungen überhaupt erst zu ermöglichen (vgl. Wiesner et al., 2018, S. 129). Ist in der jeweiligen Beurteilungssituation eine praktische Durchführung von Experimenten keine viable Option, können aber auch experimentelle Teilkompetenzen wie z. B. das Planen von Experimenten, das Aufstellen von Hypothesen oder das Interpretieren und Aufbereiten von Daten innerhalb des Kompetenzbereichs „Erkenntnisgewinnung“ fokussiert werden (Nawrath et al., 2011).

Der Kompetenzbereich „Bewertung“ bezieht sich neben dem kritischen Umgang mit (medialen) Informationen, also z. B. die Unterscheidung zwischen nicht-wissenschaftlichen und wissenschaftlich belegten Aussagen, insbesondere auf die Nutzung physikalischen Wissens zur Bewertung von Chancen und Risiken u. a. von modernen Technologien. Darin spiegelt sich auch die Forderung der KMK nach einer stärkeren Kontextorientierung im Physikunterricht wider, um den Lernenden durch sinnhafte Lebensweltbezüge die Bedeutung der Physik für ihr Leben aufzuzeigen und sie zu befähigen, ihr Wissen in alltagsnahen Anforderungssituationen nutzen zu können (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2004). Mit Blick auf die Gestaltung von Klassenarbeiten besteht eine Herausforderung im Kompetenzbereich „Bewertung“ folglich u. a. darin, Aufgaben zu konstruieren, die aktuelle gesellschaftliche Diskussionen zu naturwissenschaftlichen Themen (z. B.

Energieversorgung, Elektromobilität, Mobilfunk etc.) aufgreifen und trotz ihrer Komplexität dazu geeignet sind, dass Lernende zu diesen Themen auf Basis ihres physikalischen Wissens sowie ggf. unter Berücksichtigung ökonomischer, ökologischer und sozialer Aspekte eine sachlich begründete Position beziehen können. Als Grundlage für entsprechende Aufgaben können sowohl Internet-Quellen als auch authentische Zeitungsartikel dienen (Kuhn et al., 2011).

4. Normativität in der Mathematik: Erfahrungen und Kompetenzen

Offenkundig herrscht allgemeiner Konsens darüber, dass Mathematik in der Schulbildung einen großen Raum einnehmen sollte. Weit weniger einig ist man sich jedoch, was die Ziele des Mathematikunterrichts, über grundlegende Rechenfertigkeiten hinaus, sein sollen. Um Vorschläge für Zielsetzungen besser einordnen zu können, kann man versuchen, sie zwischen den Polen ‚Bildung‘ – im Sinne einer individuellen, biographischen Entwicklung (Heymann, 2013) – und ‚Literacy‘ – im Sinne einer funktionalen Grundbildung, welche anwendbar für die jetzige Teilhabe an einer Kultur und anschlussfähig für nachfolgendes Lernen ist (OECD, 2019) – anzuordnen.

Eine der bekanntesten Konzeptualisierungen möglicher Ziele des Mathematikunterrichts stammt von Winter (1996, S. 35), welcher den Bildungswert der Mathematik in drei „Grunderfahrungen“ sieht, die wohl eher dem Pol ‚Bildung‘ zuzuordnen sind:

1. Erscheinungen der Welt um uns, die uns alle angehen oder angehen sollten, aus Natur, Gesellschaft und Kultur, in einer spezifischen Art wahrzunehmen und zu verstehen,
2. mathematische Gegenstände und Sachverhalte, repräsentiert in Sprache, Symbolen, Bildern und Formeln, als geistige Schöpfungen, als eine deduktiv geordnete Welt eigener Art kennen zu lernen und zu begreifen,
3. in der Auseinandersetzung mit Aufgaben Problemlösefähigkeiten, die über die Mathematik hinaus gehen, (heuristische Fähigkeiten) zu erwerben.

Im Rahmen der Konzeption der PISA-Studien (Reiss et al., 2019) wurde jedoch verstärkt auf die Anwendbarkeit der Mathematik abgehoben. So werden in den PISA-Studien etwa die mathematischen Testaufgaben grundsätzlich in einen Kontext gesetzt. Kontextunabhängige, innermathematische Aufgaben, etwa aus der ebenen Geometrie, kommen in PISA nicht vor – die zweite Wintersche Grunderfahrung wird hier zu Gunsten der ersten offenkundig weniger stark gewichtet.

Der Einfluss des Literacy-Begriffs auf die deutschen Bildungsstandards nach dem Jahr 2000 ist unübersehbar (KMK, 2003). Im Sinne einer Output-Orientierung werden Kompetenzen benannt, welche die Schüler*innen im Mathematikunterricht erwerben sollen. Neben inhaltlichen Kompetenzen sind dies auch prozessbezogene Kompetenzen, von denen die prominentesten die winterschen Grunderfahrungen in die Sprache der Kompetenzorientierung übersetzen: 1. Modellieren, 2. Argumentieren und Beweisen, 3. Mathematische Probleme lösen.

Damit sind drei Kompetenzen benannt, die über das bloße Anwendenkönnen von Formeln und Standardverfahren hinausgehen – womit aber gleich auch die Frage verbunden ist, wie man solche Kompetenzen lehren und prüfen kann. Dazu gibt es kreative und ausgefeilte Vorschläge (z. B. Blum et al., 2008; Rott, 2018).

5. Echte Probleme in Klausuren?

Betrachten wir das Lösen mathematischer Probleme als Beispiel. Im Kontext Schule versteht man unter einem mathematischen Problem eine Aufgabe, welche nicht mit Standardverfahren lösbar ist. Hier ist ein elementares Beispiel für eine Aufgabe, welche für die meisten Menschen ein Problem in diesem Sinne darstellen sollte (Grieser, 2013, S. 13):

Wie viele Nullen finden sich am Ende der Zahl $1 \cdot 2 \cdot 3 \cdot \dots \cdot 99 \cdot 100$?

Die allermeisten werden keine fertige Formel für solche Fragen parat haben. Viele haben aber vermutlich trotzdem das notwendige Fachwissen aus der Sekundarstufe I, um diese Aufgabe zumindest im Prinzip lösen zu können. Neben dem Fachwissen spielen aber auch andere Faktoren eine Rolle, etwa die Kenntnis von mehr oder weniger allgemeinen Strategien („Heuristiken“), mit denen man solche Aufgaben angehen kann – aber auch die Fähigkeit zur Selbstregulation und zum planvollen Einsatz der eigenen Ressourcen und schließlich auch grundsätzliche Einstellungen und die eigene Motivation (Schoenfeld, 1985; Rott, 2018).

All dies muss von Lehrkräften beachtet werden, wenn Sie ihren Schüler*innen helfen wollen, bessere Problemlöser*innen zu werden. Das ist eine im Schulalltag herausfordernde Aufgabe, und es gibt inzwischen eine reichhaltige Literatur zur Frage, wie man das Problemlösen in den Schulunterricht einbinden kann; manches ist dabei gut vermittelbar: Bei der obigen Aufgabe etwa läge es für eine Expertin oder einen Experten nahe, nicht gleich das Produkt über alle Zahlen von 1 bis 100 zu betrachten, sondern erst einmal das Produkt der Zahlen nur bis 5, 6 oder 10. Vielleicht erkennt man ja ein Muster? Versuchen Sie es selbst! Weiter unten gibt es dann noch Hinweise zur Lösung.

Die Kompetenz, mathematische Probleme zu lösen, ist in den letzten beiden Jahrzehnten in den deutschen Bildungsplänen als eine der zentralen prozessbezogenen Kompetenzen verankert worden. Eine Einbindung des Problemlösens in den Mathematikunterricht – so viel war von Anfang an klar – kann nur in der Breite funktionieren, wenn die Kompetenz im Lösen mathematischer Probleme auch in Prüfungen relevant wird. Das ist der Fall, wenn sie etwa im Abitur getestet wird.

Hier nun stoßen wir auf eine Inkongruenz zwischen den institutionellen und formalen Rahmenbedingungen von Klassenarbeiten, Klausuren und insbesondere dem schriftlichen Abitur einerseits und den Anforderungen an einen validen Test für prozessbezogene Kompetenzen wie das Problemlösen andererseits. Wie beschrieben kennzeichnet das Problemlösen die Unverfügbarkeit einer standardisierten Lösungsmethode. Probleme sind von der Tendenz her offen, was Lösungswege angeht, und das Finden eines Lösungswegs ist ein kreativer, gelegentlich langwieriger und streckenweise auch einmal frustrierender Prozess. Wenn Sie sich gerade an obigem Problem probieren, wollen Sie vielleicht einmal eine Pause machen und später zum Problem zurückkehren. Vielleicht wollen Sie Ihre Ideen auch mit anderen diskutieren. All dies ist in den typischen Klausursituationen natürlich kaum möglich. Hier wäre langfristig eine Öffnung hin zu anderen Prüfungsformen notwendig, welche die Lehrkraft als Beobachter*in von Prozessen mehr Spielraum gibt.

Wie wenig ein Kulturwandel auch auf Seiten der Schüler*innen vollzogen zu sein scheint, belegt die folgende anekdotische Beobachtung. In die heutigen Abiturprüfungen werden schon länger Aufgaben einbezogen, die im Rahmen der Möglichkeiten auch Problemlöseanteile haben, auch wenn man dabei noch nicht von echten Problemen sprechen kann. Im Frühjahr 2019 gab es, so Presseberichte, Proteste von zehntausenden Schüler*innen aus neun Bundesländern wegen der Abituraufgaben in Mathematik (Spiegel Online, 2019, o. S.). „Die Schüler begründen ihren Protest mit Aufgabenstellungen, die vorher kaum im Unterricht vorgekommen seien.“ Auch wenn diese Regung verständlich ist, so wäre es ja genau das Lösenkönnen solcher Probleme, was wir in Zukunft unseren Schüler*innen mitgeben wollen.

[Hinweise zur Lösung des obigen Problems: Die Anzahl der Nullen am Ende einer Zahl ist gleich der Anzahl der Faktoren $10 = 2 \cdot 5$, welche in der Zahl ‚stecken‘. Wenn man das Produkt nach und nach durchgeht, so erscheint die erste 5 – und damit die erste Null – bei $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$. Als Nächstes steckt die 5 als Faktor in der 10, dann in der 15 und so weiter. Am Ende sind es 24 (nicht 20) Nullen.]

6. Literatur

- Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (Hrsg.) (2008). *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichts Anregungen und Fortbildungsideen*. Berlin: Cornelsen Verlag Scriptor.
- Bybee, R. W. (2002). Scientific Literacy — Mythos oder Realität? In W. Gräber, P. Nentwig, T. Koballa & R. Evans (Hrsg.), *Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (S. 21-43). Opladen: Leske und Budrich.
- Grieser, D. (2013). *Mathematisches Problemlösen und Beweisen*. Wiesbaden: Springer Vieweg.
- Heymann, H. W. (2013). *Allgemeinbildung und Mathematik*. Weinheim: Beltz.
- Hoffmann, L., Häußler, P. & Lehrke, M. (1998). *Die IPN-Interessenstudie Physik*. Kiel: IPN.
- KMK (2003) = Kultusministerkonferenz (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss*. München: Luchterhand.
- Kuhn, J., Müller, A., Müller, W. & Vogt, P. (2011) "Zeitungsaufgaben" und andere authentische Problemstellungen. Impulse für die Aufgabenkultur aus der physikdidaktischen Forschung. *Naturwissenschaften im Unterricht – Physik*, 22(121), 4-10.
- Nawrath, D., Maiseyenko, V., Schecker, H. (2011). Experimentelle Kompetenz - Ein Modell für die Unterrichtspraxis. *Praxis der Naturwissenschaften – Physik in der Schule*, 60(6), 42-48.
- OECD. (2019). PISA 2018 assessment and analytical framework. Paris: OECD Publishing. <https://doi.org/10.1787/b25efab8-en> [25.01.2023].
- Reiss, K., Weis, M., Klieme, E. & Köller, O. (Hrsg.) (2019). *PISA 2018. Grundbildung im internationalen Vergleich*. Münster & New York: Waxmann.
- Rott, B. (2018). Empirische Zugänge zu Heuristiken und geistiger Beweglichkeit in den Problemlöseprozessen von Fünft- und Sechstklässlern. *mathematica didactica*, 41(1), 47-75.
- Schecker, H. & Duit, R. (2018). Schülervorstellungen und Physiklernen. In H. Schecker, T. Wilhelm, M. Hopf & R. Duit (Hrsg.), *Schülervorstellungen und Physikunterricht. Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis* (S. 1-21). Berlin: Springer-Spektrum.
- Schecker, H. & Klieme, E. (2001). Mehr Denken, weniger Rechnen: Konsequenzen aus der internationalen Vergleichsstudie TIMSS für den Physikunterricht. *Physikalische Blätter*, 57(7-8), 113-117.
- Schecker, H., Wilhelm, T., Hopf, M. & Duit, R. (Hrsg.) (2018). *Schülervorstellungen und Physikunterricht. Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis*. Berlin: Springer-Spektrum.
- Schoenfeld, A. (1985). *Mathematical Problem Solving*. London: Academic Press.
- Schreiner, C. & Sjøberg, S. (2004). *The relevance of science education. Sowing the Seed of ROSE*. Oslo: Acta Didactica.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2004). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss (Jahrgangsstufe 10)*. München, Neuwied: Wolters Kluwer Deutschland.
- Spiegel Online. (2019). *Schüler wehren sich gegen Matheaufgaben*. <https://www.spiegel.de/lebenundlernen/schule/abitur-schueler-verzweifeln-an-mathepruefungen-a-1265918.html> [25.01.2023].
- Tesch, M. & Duit, R. (2004). Experimentieren im Physikunterricht – Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 51-69.
- Wiesner, H., Schecker, H. & Hopf, M. (2018). *Physikdidaktik kompakt*. Seelze: Aulis Verlag.
- Winter, H. (1996). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 61, 37-46.

Angaben zu den Autoren:

- Jun.-Prof. Dr. Jan-Philipp Burde, Juniorprofessor für Didaktik der Physik, AG Didaktik der Physik, Universität Tübingen, jan-philipp.burde@uni-tuebingen.de
- Prof. Dr. Walther Paravicini, Professor für Mathematik und ihre Didaktik, Mathematisches Institut, Universität Tübingen, w.paravicini@uni-tuebingen.de

Wolfgang Polleichtner

Graecum und Latinum. Zwischen normativem Anspruch und relativer Ungleichbehandlung

1. Was ist das Graecum? Was ist das Latinum?

Graecum, Latinum und Hebraicum – das sind Sprachprüfungen in Griechisch Latein und Hebräisch. Der eigentümliche Name geht eigentlich auf die Bezeichnung *examen Latinum*, *examen Graecum* sowie *examen Hebraicum* zurück (zur Geschichte dieser Prüfungen: Brüssel, 2018). Diese Sprachprüfungen muss man in Österreich, der Schweiz und Deutschland bestanden haben, wenn man bestimmte Fächer studieren möchte. Allerdings haben alle drei genannten Länder auch zum Teil intern jeweils unterschiedliche Richtlinien entwickelt – sowohl für die Prüfungsinhalte, Prüfungszeitpunkte, die genauen Prüfungsbezeichnungen und auch für die Fächer, für die man Sprachkenntnisse, wie sie durch diese Prüfungen nachgewiesen werden, benötigt. Im Folgenden beschränke ich mich auf die derzeitige Situation in der Bundesrepublik Deutschland und auf Latinum und Graecum.

Der gesamte Themenkomplex ist bildungspolitisch eigentlich seit dem Fall des Abiturmonopols des Humanistischen Gymnasiums im Jahr 1900 nachgerade vermint. Sämtliche mit den genannten drei altsprachlichen Examina verbundenen Fragen sind durch die Kontroversen aber gleichzeitig fest mit der Geschichte der Bildungspolitik verwachsen und spielen auch in aktuellen Diskussionen – etwa um die Gemeinschaftsschule in Baden-Württemberg – immer wieder eine Rolle.

In Deutschland regelt derzeit die „Vereinbarung über das Latinum und das Graecum“ der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) die Verhältnisse (aktualisiert zuletzt am 22.09.2005). Am 12.3.2008 wurde diese Vereinbarung von derselben Ständigen Konferenz um „Richtlinien für die Zuerkennung des Latinums an deutschen Schulen im Ausland“ ergänzt. Ein Latinum und ein Graecum, das nach diesen Vereinbarungen abgelegt wird, ist ‚das‘ Latinum bzw. Graecum der Kultusministerkonferenz. Es gibt neben diesen Sprachprüfungen noch regional, das heißt auf Landes- oder auch Universitätsebene unterschiedlich geregelte Prüfungen, die gelegentlich ‚Kleines Latinum‘, ‚Großes Latinum‘, ‚Theologen-Graecum‘ oder ähnlich genannt werden. Diese Prüfungen gehen über die Standards der Kultusministerkonferenz hinaus oder bleiben dahinter zurück. Im Folgenden sprechen wir also von Latinum und Graecum im Sinne der zitierten KMK-Vereinbarung aus dem Jahr 2005.

Weshalb gibt es diese Vereinbarung der Kultusministerkonferenz? Am 21.5.1970 wurden von der Kultusministerkonferenz „Rahmenbestimmungen zu den Anforderungen der wissenschaftlichen Prüfung für das Lehramt an Gymnasien“ erlassen. Sie legten für verschiedene Fächer als Prüfungsvoraussetzung „Lateinkenntnisse“ oder „Griechischkenntnisse“ fest. Ein Prüfungsstandard wurde aufgrund unterrichtsgeschichtlicher Tradition im „Latinum“ und „Graecum“ gefunden. Die Anforderungen und die Erwerbsmöglichkeiten dieser „Kenntnisse“ deutschlandweit einheitlich zu regeln, war das Ansinnen einer Festlegung dieser Prüfungsvoraussetzungen am 26.10.1979, die 2005 fortgeschrieben wurden, „um die Anforderungen an das Latinum und das Graecum und die Möglichkeiten ihres Erwerbs auch künftig einheitlich zu gestalten und der didaktischen Entwicklung in den Ländern der Bundesrepublik Deutschlands anzupassen.“ Graecum und Latinum stellen so einen wichtigen Beitrag zu einem bundesweit einheitlichen Abitur an Schulen im In- und Ausland dar. Worin allerdings die Anpassungen an die didaktische Entwicklung in den Ländern zwischen 1979 und 2005 lagen, wird nicht beschrieben. Das Papier beinhaltet keine didaktischen Hinweise.

Latinum und Graecum gelten wie auch das Hebraicum als Ergänzungsprüfungen zum Abitur. Sie werden daher entweder von der Abiturprüfungskommission einer Schule zuerkannt oder von Kommissionen, in denen staatlich geprüfte Lehrerinnen und Lehrer sitzen, die mit dieser Aufgabe von den Schulämtern oder Regierungspräsidien betraut sind.

Da das Fach Hebräisch nicht mehr in allen Bundesländern ordentliches Schulfach ist, ergeben sich für dieses Fach natürlich besondere Schwierigkeiten, da es mancherorts eben etwa keine Lehrkräfte mit entsprechender Prüfung mehr gibt, aber auch keine bundeseinheitliche KMK-Vereinbarung zum Hebraicum. Im Musterentwurf der KMK zum Formular des Zeugnisses der Allgemeinen Hochschulreife vom 8.1.1974 in der Fassung vom 20.9.2007 steht lediglich in Fußnote 15, dass die Länder über das Hebraicum in eigener Zuständigkeit entscheiden.

Das ist für Latein insofern wichtig, als dieselbe Fußnote die gleiche Regelung für das „Kleine Latinum“ trifft. Die KMK kannte 2007 also einerseits ‚ihr‘ Latinum und erkennt andererseits an, dass auch ein „Kleines Latinum“ in Länderhoheit vergeben werden kann. Von der umgekehrten Möglichkeit, dass es Gründe geben könnte, über das KMK-Latinum hinausgehende Standards festzulegen, scheint die KMK nicht auszugehen. In der neuesten Fassung von 2018 formuliert die erwähnte Fußnote 15, die jetzt zur Fußnote 16 wurde, allerdings nichts mehr zu einem „Kleinen“ Latinum, überlässt nur sämtliche Entscheidungen zu Latinum, Graecum und Hebraicum in Abiturzeugnissen den Ländern.

Ein Graecum, das ‚kleiner‘ oder ‚größer‘ wäre als das durch die KMK-Vereinbarung festgelegte, scheint es überhaupt nicht zu geben, obwohl das leider auch nicht mehr der Realität entspricht. Gerade theologische Fakultäten geben sich heutzutage aus unterschiedlichen Gründen für manche ihrer Studiengänge auch mit Griechischkenntnissen zufrieden, die nicht dem KMK-Graecum genügen, dafür aber inhaltlich auf theologische Texte abgestimmt sind.

Aber die „Ländergemeinsamen inhaltlichen Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung“ der KMK vom 16.10.2008 in der Fassung vom 16.5.2019 kennen lediglich ein Latinum oder Graecum. Diese „Anforderungen“ haben, die oben erwähnten „Rahmenbestimmungen“ für das gymnasiale Lehramt von 1970 mittlerweile ersetzt. Für zukünftige Lehrkräfte in Latein und Griechisch werden Latinum und Graecum als Studienvoraussetzungen festgeschrieben. Studienordnungen für diese Lehramtsfächer beruhen in ganz Deutschland auf dieser Voraussetzung.

Nun gerät aber gerade das Latinum durch die unterschiedlichen Schwierigkeitsgrade tatsächlich abgehaltener Prüfungen, die teilweise schon über zentralen Abiturprüfungen liegen, in eine massive Glaubwürdigkeits- und Legitimationskrise (Niehl, 2021). Denn es dürfte nach dem oben Gesagten an sich überhaupt nicht passieren, dass eine Latinumsprüfung schwieriger zu bewältigen wäre als eine Abiturprüfung. Die Prüfungsanforderungen für Latinum und Graecum sind eigentlich streng geregelt:

2. Prüfungsanforderungen

Nach Ziffer 2 der KMK-Vereinbarung müssen die Prüflinge im Latinum

„lateinische Originaltexte im sprachlichen Schwierigkeitsgrad inhaltlich anspruchsvollerer Stellen (bezogen auf Bereiche der politischen Rede, der Philosophie und der Historiographie) in Inhalt, Aufbau und Aussage [...] erfassen.“ „Dieses Verständnis ist durch eine sachlich richtige Übersetzung in angemessenem Deutsch, ggf. zusätzlich durch eine vertiefende Interpretation nachzuweisen. Hierzu werden Sicherheit in der für die Texterschließung notwendigen Formenlehre und Syntax, ein ausreichender Wortschatz und die erforderlichen Kenntnisse aus den Bereichen römische Politik, Geschichte, Philosophie und Literatur vorausgesetzt.“

Es steht den Ländern frei, die Benutzung eines Wörterbuchs bei den Übersetzungsleistungen zuzulassen.

Für Griechisch wird die Beschränkung der lateinischen Prüfung auf politische Rede, Philosophie und Geschichtsschreibung – worunter unter Umständen nicht nur Prosa fallen könnte – explizit auf „inhaltlich anspruchsvollere Platontexte“ konkretisiert. Allerdings schreibt der Text eben nicht vor, dass nur Prosatexte in den Prüfungen zur Übersetzung gestellt werden könnten, sondern dass sich ihr inhaltlicher Schwierigkeitsgrad an diesen „inhaltlich anspruchsvolleren Stellen“ orientieren soll.

Wann sollen diese Kenntnisse nachgewiesen werden? Wollen Prüflinge das Latein oder Graecum durch Schulunterricht erwerben, so gilt ein Mindestunterrichtsvolumen von lediglich vier Schuljahren, wobei am Ende eben solche Textstellen behandelt werden müssen, die dem erwähnten Anforderungsniveau entsprechen. Wenn weniger Schulunterricht erteilt wird, muss eine Prüfung abgelegt werden, die aus einer schriftlichen und mündlichen Teilprüfung besteht (Zu Problemen dieser doppelten Prüfung bei der Anrechnung für Bachelor-Studiengänge siehe Polleichtner, 2020.). Die Abschlussnote bzw. Prüfungsnote muss zur Zuerkennung des Graecums oder Latinums mindestens 4,0 lauten.

Diese letztere Regelung einer Doppelprüfung besteht dann auch für Prüflinge, die nach ihrem Schulbesuch – als Nicht-Schüler*innen – ihre Abiturprüfungen entsprechend nachträglich erweitern wollen. Hierfür werden dann vor eigens bestimmten Prüfungsausschüssen der Schulbehörden zunächst eine schriftliche und dann eine mündliche Prüfung abgelegt. Die schriftliche Prüfung dauert 180 Minuten für 180 lateinische bzw. 195 griechische Wörter. Dieser Prüfung folgt eine mündliche Prüfung von 20-minütiger Dauer, vor der eine halbe Stunde Vorbereitungszeit angesetzt ist. Für die mündliche Prüfung wird ein neuer Text im Lateinischen von 50 Wörtern und im Griechischen von 60 Wörtern gestellt.

Was mit „inhaltlich anspruchsvolleren Stellen“ für die Klausur gemeint ist, wird durch Beispiele im Anhang der KMK-Vereinbarung verdeutlicht. Für Latein kommen diese Beispieltex te aus Ciceros 2. Philippischer Rede, aus Ciceros *de natura deorum*, aus seinen *Tusculanischen Gesprächen* und aus der Chronik Ottos von Freising. Für Griechisch stammen die Beispieltellen alle von Platon; sie sind aus dem *Lysis*, dem *Kriton* und der *Apologie* genommen. Was aber das Wort „anspruchsvoller“ genau meint, wird über diese Beispiele hinaus nicht definiert. Die Textstellen müssten sich aber wohl über dem durchschnittlichen Anspruchsniveau der Werke Platons usw. bewegen. Wir stellen also fest, dass der primäre Fokus der Latinums- und Graecumskurse auf den Inhalt, nicht unbedingt auf die Sprache der Textstellen gerichtet ist.

Diese Feststellung stimmt überein mit einer Möglichkeit, die die KMK-Vereinbarung für Graecum und Latinum eröffnet: Statt einer Klausur, die in Gänze aus einer

Übersetzung besteht, gibt es die Möglichkeit, eine Klausur zu stellen, die zu zwei Dritteln eine Übersetzungsaufgabe und zu einem Drittel Interpretationsaufgaben beinhaltet. Die Endnote würde dann konsequenterweise im Verhältnis 2:1 errechnet werden. Persönlich kenne ich allerdings keinen Standort für außerschulische Latinums- und Graecumskurse bzw. deren Prüfungen, an dem diese Prüfungsform tatsächlich Verwendung finden würde. Es würde nämlich zusätzliche Vorbereitungszeit erfordern, diese Aufgabenformate einzuführen. Diese Vorbereitungszeit steht aber nicht zur Verfügung. Die meisten Kursleiter*innen, die ich kenne, sagen, dass die Kursteilnehmer*innen erst in den buchstäblich letzten beiden Wochen der typischerweise zweisemestrigen Vorbereitungskurse ein Niveau von Übersetzungskompetenz erreichen, das nötig ist, um die Abschlussprüfung auch mit Erfolgsaussicht angehen zu können.

Hieraus ergibt sich ein Problem: Die Aufgaben, die gestellt werden, orientieren sich an den sogenannten EPA, den Einheitlichen Prüfungsanforderungen in der Abiturprüfung. Eine Aufgabe mit ‚Übersetzen‘ als Operator stellt eine Aufgabe in den Anforderungsbereich III. Eine reine Übersetzungsaufgabe ist also offensichtlich schwerer als eine Klausur, die aus einer Übersetzung und Zusatzaufgaben besteht, deren Teilaufgaben auch aus den beiden ersten Anforderungsbereichen stammen. Und jetzt müssen wir die Leistungen schulfremder Prüflinge bzw. von Prüflingen von spätbeginnenden Sprachkursen an Schulen, die ja, wie wir gesehen haben, beide diese ausführlich dargelegten Prüfungen ablegen müssen, mit dem vergleichen, was die Schüler*innen der anderen Ausbildungsgänge zu leisten haben: Hier gibt es, soweit ich sehe, in keiner Schule noch Klassenarbeiten oder Schulaufgaben, die aus einer reinen Übersetzungsaufgabe bestehen würden.

Obendrein ergibt sich aus der Beispielklausur, die dieses Aufgabenformat im Rahmen der KMK-Vereinbarung vorstellt, dass die KMK davon ausgeht, dass diese Interpretationsaufgabe an einem zusätzlich zum Übersetzungstext neu eingeführten Text, der im Original und in deutscher Übersetzung gegeben wird, gelöst werden soll. Es gibt Fachdidaktiker*innen bzw. Kultusbeamt*innen, die im Gegensatz zu der Beispielklausur der KMK darauf bestehen, dass bei Klassenarbeiten in der Schule der zu interpretierende Text auch der zu übersetzende Text sein muss. Auch diese Frage ist nicht trivial zu beantworten, sondern fachintern bisweilen eine Glaubensfrage. Denn wenn man den Übersetzungstext falsch übersetzt, müssen und werden die Interpretationsfragen auf jeden Fall falsch beantwortet werden. Umgekehrt könnten Interpretationsfragen Prüflinge auch auf die richtige Spur bei der Übersetzung zurückbringen. Andererseits: Durch den neuen Text mit deutscher Begleitübersetzung kann (oder muss) eben in einem zweiten Ansatz ein neuer Text behandelt werden. Diese Identität oder Nicht-Identität der Texte von Übersetzungs- und Interpretationsaufgabe ist durch den Text der KMK-

Vereinbarung allerdings explizit weder gefordert noch ausgeschlossen, nur durch das Beispiel gesetzt.

Dieses Problem ergibt sich für den mündlichen Prüfungsteil nicht: Hier stehen nach der mündlichen Vorübersetzung inhaltliche und sprachliche Erklärungen rund um den Übersetzungstext auf dem Programm, wobei sich sprachliche Fragen auf eine Basisgrammatik beschränken sollen – wie auch immer diese definiert ist –, ohne dass aber ein genauer Bewertungsschlüssel von Übersetzungsleistung und weiteren Fragen durch die KMK-Vereinbarung festgelegt wäre.

3. Fazit

Wir müssen an dieser Stelle also konstatieren: Das Graecum und das Latinum stehen mit einem enormen normativen Anspruch da, den sie in der Praxis nicht mehr erfüllen. Sie behandeln Schüler*innen in den L1- und L2-Lehrgängen auf der einen Seite und diejenigen in L3- oder L4-Lehrgängen oder Schulfremde auf der anderen Seite (siehe wie oben Niehl, 2021) nicht gleich. Zusätzlich eröffnen die KMK-Vereinbarungen noch Möglichkeiten, alle Gruppen untereinander ungleich zu behandeln. Dabei, und das ist ebenfalls wichtig festzustellen, wird die Heterogenität der Ausgangspositionen der Prüflinge noch gar nicht berücksichtigt.

Schauen wir uns nur zum Beispiel den Fall sogenannter Bildungsausländer*innen an. Die Hochschulen in Deutschland eröffnen den Hochschulzugang mit Sprachzertifikaten. Manchmal sind sie mit Deutschkenntnissen auf dem Niveau B2 zufrieden, manchmal wird C1 verlangt. Wir müssen auch in letzter Zeit konstatieren, dass selbst Bildungsinländer*innen nicht immer die besten Deutschkenntnisse mitbringen. Dennoch erscheint es merkwürdig, das Latinum und Graecum in gleicher Weise von Prüflingen zu erwarten, die sich auf allen Kompetenzstufen des Deutschen von B2 bis C2 bewegen dürfen.

An der äußeren Norm und Form von Latinum und Graecum wurde und wird festgehalten. Am inneren Gehalt wird aber trotz der Darstellung in Gesetznormen nach außen die Anforderungsschraube nach unten gedreht. Die Anschlussfähigkeit insbesondere des Lehramtsstudiums in Latein und Griechisch an den Schulunterricht ist dadurch in hohem Maße gefährdet.

4. Literatur

Brüssel, M. (2018). *Altsprachliche Erwachsenendidaktik in Deutschland. Von den Anfängen bis zum Jahr 1945*. Propyläen: Heidelberg.

Niehl, R. (2021). Zur Entwicklung des Latein-Abiturs. *Latein und Griechisch in Baden-Württemberg*, 49(2), 5-30.

Polleichtner, W. (2020). Zu alten und neuen Fragen rund um Graecum und Latinum. Die Vereinbarungen der Kultusministerkonferenz, die Studiengangplanung der Bachelor- und Masterabschlüsse und vom Leben zwischen Baum und Borke. *Latein und Griechisch in Baden-Württemberg*, 48(2), 21-36.

Angaben zum Autor:

- Dr. Wolfgang Polleichtner, Arbeitsbereich Fachdidaktik der Alten Sprachen am Philologischen Seminar der Universität Tübingen, wolfgang.polleichtner@philologie.uni-tuebingen.de

Stefan Hofstetter

Zur Bewertung mündlicher Leistungen (nicht nur) im Fremdsprachenunterricht

Wo geringe Normierung auf große pädagogische Freiheit trifft – Herausforderungen, Chancen und praktisches Handeln

1. Eine erste Annäherung

In einer vielzitierten Studie ließ Manfred Arendt die mündliche Englischprüfung des Muttersprachlers Dan per Video aufzeichnen und anschließend von insgesamt 26 verschiedenen Lehrkräften an 22 Schulen bewerten (Arendt, 2006). Dabei hatte Arendt penibel darauf geachtet, dass die zugrundeliegende Prüfung inhaltlich alle relevanten Aspekte korrekt umfasste und auch sichergestellt, dass in sprachlicher Hinsicht, wie von einem Muttersprachler nicht anders zu erwarten, nichts zu beanstanden war. Ernüchterndes Ergebnis: Nur zwei Lehrkräfte entschieden sich für eine ‚sehr gute‘ Bewertung, die große Mehrheit (23) vergab Noten im ‚guten‘ bis ‚befriedigenden‘ Notenbereich und einmal wurde diese Prüfung sogar allen Ernstes als ‚nicht bestanden‘ eingestuft. Viel eindrucksvoller kann man die Relativität der Bewertung mündlicher Schülerleistungen wohl kaum demonstrieren.

Auch in meiner eigenen Erfahrung führen gerade mündliche Noten im Schulalltag oftmals zu Konflikten, mit Schüler*innen selbst, aber auch mit deren Eltern oder innerhalb des Kollegenkreises, und das deutlich häufiger als bei schriftlichen Leistungsbewertungen. Nachstehend soll daher dargelegt werden, warum gerade mündliche Bewertungen im Fremdsprachenunterricht aber auch in anderen Fächern für Lehrkräfte eine besondere Herausforderung sind. Gleichzeitig möchte ich jedoch auch darstellen, welche Chancen mündliche Noten Lehrkräften eröffnen und Handlungsvorschläge für die Praxis unterbreiten, wie Lehrkräfte mit dem für mündliche Bewertungen typischen Spannungsfeld aus kaum vorhandener Normativität einerseits und einem hohen Maß an persönlicher Entscheidungsfreiheit andererseits umgehen können.

2. Herausforderungen und Chancen mündlicher Leistungsbewertung

2.1. Problemfelder

Zunächst ist festzustellen, dass Lehrkräfte mündliche Noten häufig ‚nebenbei‘ vergeben, während der Unterricht weiterläuft und sie etwa auch auf Inhaltliches, Disziplin im Klassenzimmer u. Ä. achten müssen, was die Einnahme einer unabhängigen Beobachterposition unmöglich macht. Außerdem entstehen im Bereich der Mündlichkeit anders als bei schriftlichen Leistungen keine dauerhaften Ergebnisse, die hinterher in einem ruhigen Umfeld objektiv verglichen werden könnten und zudem besteht hier selbst bei kritischen Entscheidungen, etwa wenn es um die (Nicht-)Versetzung von Schüler*innen geht, nicht die Möglichkeit, Noten von Kolleg*innen gegenchecken zu lassen. Ebenfalls im Gegensatz zur schriftlichen Leistungsmessung werden mündliche Noten in der Regel auch nicht auf Basis von Punkten oder Bewertungsskalen vergeben, was diese weniger objektiv erscheinen lässt und nicht selten Beschwerden seitens der Schüler*innen oder deren Eltern hervorruft (wenngleich solche Punkte/Skalen *per se* natürlich auch nicht zwangsläufig zu einem hohen Grad an Objektivität führen müssen; hier geht es aber nur um die subjektive Wahrnehmung der beteiligten Akteur*innen). Insbesondere bei der Vergabe von Eindrucksnoten, der vermutlich am weitesten verbreiteten Praxis (siehe 3), besteht weiterhin die Gefahr der Verfälschung von Ergebnissen über längere Zeiträume hinweg, wenn einzelne, besonders gute Beiträge am Ende der Bewertungsperiode überschätzt werden oder zu Beginn Gesagtes in Vergessenheit gerät.

Auch der Einfluss subjektiver Faktoren ist bei mündlichen Bewertungen sicherlich größer; an dieser Stelle sei lediglich exemplarisch auf den *halo effect* und den logischen Fehler, persönliche Sympathien bzw. Antipathien (beispielsweise ruhige und angenehme versus undisziplinierte Schüler*innen) oder individuelle Tendenzen von Lehrkräften wie die Vermeidung schlechter Noten (‚Kein*e Schüler*in ist so schlecht, dass er/sie eine 5 oder 6 verdient.‘) oder die Tendenz zur Mitte (‚Niemand kann meinen hohen Anforderungen an eine 1 wirklich gerecht werden.‘) verwiesen (zu subjektiven Faktoren vgl. [u. v. a.] Jürgens & Sacher, 2008, S. 54-71 oder Grimm et al., 2015, S. 272ff.). Ebenso ist häufig eine Beeinflussung mündlicher Noten durch schriftliche Leistungen zu beobachten mit der Tendenz, dass erstere die letzteren eher bestätigen als erheblich davon abzuweichen: So erfordert es Mut von Lehrkräften, etwa einem Schüler, der schwache Essays schreibt, eine richtig gute mündliche Note zu geben oder umgekehrt jemandem, der hervorragende schriftliche Leistungen aufweist, die Gesamtnote durch eine schlechte mündliche Teilnote zu ‚ruinieren‘. (Diese Auflistung wurde erstellt in Anlehnung an

Wengert, 2011, S. 334ff. und 343f. sowie Jäger, 2000, S. 173-210; da diese Aspekte meines Erachtens als unumstritten angesehen werden können, habe ich hier aus Platzgründen auf eine genaue Quellenangabe bei jedem Einzelpunkt verzichtet.)

Speziell im Fremdsprachenunterricht besteht zudem innerhalb des heute üblichen *communicative approach* die Gefahr einer kompletten Überforderung leistungsschwächerer Schüler*innen, wenn ausschließlich freies Sprechen beurteilt wird, wo verschiedenste Anforderungsbereiche (Wortschatz, Grammatik, Aussprache, aber auch Inhaltliches oder rhetorische Fähigkeiten) zusammenkommen. Dies wird nach meiner Erfahrung häufig erst in höheren Klassenstufen überhaupt möglich, weshalb hier unbedingt auch auf den Einsatz einfacherer, weniger integrativer mündlicher Aufgabenstellungen zu achten ist.

Zusätzlich erschwert wird die mündliche Leistungsbewertung außerdem von einer ganzen Reihe juristischer Fallstricke wie etwa der sogenannten ‚Bringschuld‘ der Schüler*innen, welcher die ‚Holschuld‘ der Lehrkräfte gegenübersteht oder der Tatsache, dass bei mündlichen Noten lediglich die Qualität, nicht aber die Quantität der Beiträge berücksichtigt werden soll (letztere darf sich nur in der Kopfnote der Schüler*innen für ‚Mitarbeit‘, nicht aber in einzelnen Fachnoten niederschlagen), was in der Praxis jedoch oftmals kaum zu trennen ist. (Ich beschränke mich hier auf die Erwähnung dieser beiden juristischen Probleme; interessierte Leser*innen seien auf die ausführlichere Darstellung in Reip & Gayer, 2012, S. 108-134 verwiesen.)

2.2. Chancen

Trotz dieser allgegenwärtigen Schwierigkeiten und Herausforderungen sollten Lehrkräfte die Pflicht zur Bewertung mündlicher Leistungen aber nicht nur als lästige Bürde ansehen, sondern sich auch die Chancen bewusst machen, die damit unmittelbar einhergehen: So führt die Vergabe mündlicher Noten zu einem zusätzlichen Anreiz für Schüler*innen, sich aktiv am Unterricht zu beteiligen (Wengert, 2011, S. 343) und auch wenn Lehrkräfte sich selbstverständlich ohnehin um eine ansprechende und motivierende Unterrichtsgestaltung bemühen sollten, schadet dieser zusätzliche Beteiligungsanreiz sicherlich nicht. Darüber hinaus stellen mündliche Noten auch eine ganzheitlichere Art der Bewertung als die meisten Formen schriftlicher Leistungsmessung dar und sind deshalb als aussagekräftiger einzustufen, denn während letztere immer nur Auskunft über den Leistungsstand zu einem ganz bestimmten Zeitpunkt bezüglich einer spezifischen Aufgabenstellung geben, sagen erstere etwas über die Leistungsfähigkeit von Schüler*innen über einen längeren Zeitraum hinweg und hinsichtlich unterschiedlicher

Leistungskomponenten aus (vor allem bei klassischen Eindrucksnoten, siehe 3). Weiterhin zeichnen sich mündliche Noten auch durch einen geringeren Wettbewerbscharakter aus, da die einzelnen mündlichen Beiträge meist eigenständig bewertet werden und nicht mit dem Klassenverbund verglichen wird. Im Sprachunterricht bieten mündliche Bewertungen auch Schüler*innen, die sich im Schriftlichen etwa aufgrund mangelnder Rechtschreib- oder Zeichensetzungskenntnisse schwertun und deshalb dort nie gute oder gar sehr gute Resultate erzielen, die Möglichkeit, ihre Note aufzubessern und so ihre Motivation für ein Unterrichtsfach aufrechtzuerhalten. Und schließlich leben speziell Fremdsprachen davon, ‚gesprochen‘ zu werden: Es ist mindestens genauso wichtig, Schüler*innen auf fremdsprachliche Begegnungen in gesprochener Form vorzubereiten (bspw. im Rahmen schulischer Austauschprogramme) wie in geschriebener und daher ist es nur folgerichtig, wenn sich dieser hohe Stellenwert der Mündlichkeit auch in der Notengebung niederschlägt.

3. Häufige Praktiken mündlicher Notenvergabe

Im schulischen Kontext können zwei Grundrichtungen der Vergabe mündlicher Noten unterschieden werden: Zum einen die zeitraumorientierte Notenvergabe, wie sie etwa bei klassischen Eindrucksnoten Anwendung findet, wobei alle Beiträge von Schüler*innen über einen längeren Zeitraum hinweg Berücksichtigung finden und zum anderen zeitpunktbasierte Methoden, z. B. in Form des Abfragens einzelner Schüler*innen zu Stundenbeginn über wesentliche Inhalte der Vorstunde, der Präsentation von Hausaufgaben oder Übungen, der mündlichen Vorstellung von Stundenprotokollen oder von Referaten der Schüler*innen zu einem bestimmten Thema, was häufig im Fremdsprachenunterricht geschieht. Wenig überraschend gehen beide Grundformen dabei gleichermaßen mit Vor- wie Nachteilen einher: Zeitraumorientierte Verfahren sind als sehr ganzheitliche Methode anzusehen und können auch bei komplexeren Aufgabenstellungen wie dem freien Sprechen im Fremdsprachenunterricht angewandt werden. Gleichzeitig kann dieser längere Zeitraum der Notenerhebung Eindrücke aber auch massiv verfälschen, diese Noten entstehen quasi ‚nebenbei‘ (siehe 2.1) und vielen Lehrkräften fällt es schwer, im normalen Unterrichtsgespräch ihrer ‚Holschuld‘ den Schüler*innen gegenüber gerecht zu werden. Um zumindest ersterem Problem entgegenzuwirken, wurde vorgeschlagen, im unmittelbaren Anschluss an jede Stunde anhand einer Strichliste o. Ä. Eindrücke direkt zu dokumentieren (Wengert, 2011, S. 345), was meines Erachtens bei Lehrkräften mit hohem Deputat, die von einem Klassenzimmer zum nächsten eilen, jedoch kaum praktikabel ist. Bei den zeitpunktbasierten Vorgehensweisen liegt der Fokus hingegen auf der Notenfindung selbst und die Lehrkraft kann ihre ‚Holschuld‘ problemlos erfüllen, doch hier

ergeben sich andere Schwierigkeiten: Schüler*innen laufen Gefahr, vor der Klasse bloßgestellt zu werden (dies betrifft übrigens nicht nur schwache, sondern auch sehr leistungsstarke Schüler*innen, die nicht als ‚Streber‘ abgestempelt werden wollen), wertvolle Unterrichtszeit geht in jeder Stunde verloren, die Aussagekraft ist teils eingeschränkt (Schüler*innen können Zeitpunkte erwischen, an denen sie fit und ausgeschlafen oder aber übernächtigt, erkältet, etc. sind) und mit fortschreitendem Unterrichtsjahr wird für die verbleibenden Schüler*innen ihre anstehende Abfrage zusehends vorhersagbarer, so dass sich hier letzten Endes auch die Gerechtigkeitsfrage stellt.

Im Gegensatz zur schriftlichen Leistungsmessung wird bei mündlichen Noten häufig auch die Frage der Beteiligung der Schüler*innen an der Notenvergabe, die mittels einer Schüler selbstbewertung (*self assessment*) und/oder Fremdbeurteilung durch andere Schüler*innen (*peer assessment*) erfolgen kann, gestellt und durchaus kontrovers diskutiert: Während beispielsweise Anne Lewis die Lehrkraft uneingeschränkt ins Zentrum aller Bewertungen rückt (Heflebower et al., 2014, S. 44), wird eine solche Schülerbeteiligung in Buschmann (2009, S. 22) oder Walvoord und Johnson Anderson (2011, S. 67f.) vehement eingefordert; anderswo, etwa in Bach und Timm (2013, S. 227) oder Nieweler (2006, S. 261ff.) finden sich diesbezüglich etwas zurückhaltender formulierte Forderungen, weshalb ich hier meine eigenen Erfahrungen anfügen möchte: Insbesondere zu Beginn meiner Lehrtätigkeit habe ich mit verschiedenen Formen von *self* und *peer assessment* experimentiert, allerdings mit wenig Erfolg. Bei ersterer Methode hat sich herausgestellt, dass Schüler*innen nicht notwendigerweise besonders gut darin sind, ihr eigenes Leistungsvermögen angemessen einzuschätzen, sowohl bei leistungsschwächeren Schüler*innen (keiner gibt sich gern selbst eine 5 oder Schlechteres, selbst wenn dies angebracht wäre) als auch bei sehr leistungsstarken Schüler*innen, die sich – vielleicht aufgrund falscher Bescheidenheit – zu keiner Note im sehr guten Bereich trauten. Aber auch letzteres Vorgehen erwies sich als problematisch: Vor allem jüngere Schüler*innen waren häufig nicht in der Lage, bei der Bewertung ihrer Mitschüler*innen über persönliche Belange wie Freundschaften oder Animositäten hinwegzusehen und so mutete das Endergebnis solcher Noteneinholungen oftmals eher wie ein Popularitätsranking innerhalb einer Klasse an, wo am einen Ende des Rankings ablesbar war, wer besonders beliebt ist, und das andere Ende Aufschluss über die Außenseiter einer Klasse gab, das Ganze aber wenig mit einer objektiven Leistungsbewertung zu tun hatte. Solche Schülerbeteiligungen sind zudem immer mit einem erheblichen Mehraufwand für Lehrkräfte verbunden (Erhebung; Auswertung) und am Ende stellt sich auch eine grundsätzliche Frage, nämlich nach dem Umgang mit Abweichungen zwischen den Einschätzungen der Lehrkraft und denen der Schüler*innen: Fließen die Bewertungen der Schüler*innen selbst in die Endnote ein (zu welchem Prozentsatz auch immer),

besteht sofort die Gefahr des Missbrauchs, indem die Schüler*innen sich oder anderen bewusst bessere Noten geben; setzt sich im Zweifelsfall hingegen die Meinung der Lehrkraft durch, entsteht bei Schüler*innen sehr schnell der Eindruck, dass sie ohnehin nicht ernstgenommen werden und ihre Beteiligung nur *pro forma* erfolgt – ein echtes Dilemma. Alles in allem sehe ich daher *self/peer assessment* im Kontext mündlicher Noten eher kritisch und es verwundert nicht, dass Schülerbeteiligungen derzeit „noch weit von gängiger Unterrichtspraxis entfernt“ sind (Haß, 2006, S. 289).

4. Fazit und Praxisvorschläge

Insgesamt zeigt sich also, dass Lehrkräfte bei der mündlichen Notengebung über ein hohes Maß an pädagogischer Freiheit verfügen und dieser Bereich zugleich kaum normiert ist. Dies steht im Gegensatz zu vielen schriftlichen Formen der Leistungsmessung (als Extrembeispiel seien die rigiden Korrekturvorgaben der schriftlichen Abiturprüfung erwähnt; vgl. „Korrekturrichtlinien“, S. 39-60), was Lehrkräften Chancen bietet, aber auch zu Komplikationen führen kann. Abschließend möchte ich deshalb noch drei konkrete Handlungsempfehlungen aussprechen: Erstens rate ich bei der praktischen Notenvergabe zu einer Kombination aus einem zeitraum- und einem zeitpunktbasierten Verfahren, um die jeweiligen Nachteile zumindest teilweise gegenseitig auszugleichen: So vergebe ich selbst z. B. vier mündliche Eindrucksnoten pro Jahr, die ich anhand einer speziellen *Cluster*-Methode festlege (vgl. Abb. 1), um Reihenfolgeeffekten einer Klassenliste vorzubeugen und gleichzeitig verlange ich von allen Schüler*innen eine Präsentation zu einem selbstgewählten Thema (Buch, Film, Musikgruppe, Reiseziel, Sportler*in, etc.), wobei sich die freie Themenwahl für die Motivation der Schüler*innen als äußerst zuträglich erwiesen hat. Zweitens ist es hilfreich, sich der ausgeführten Schwierigkeiten überhaupt bewusst zu sein, da ein entsprechendes Problembewusstsein direkt dazu beitragen kann, manche dieser Fallstricke zu vermeiden. Und drittens empfiehlt sich eine möglichst frühzeitige Beschäftigung angehender Lehrkräfte mit Fragen der Leistungsmessung bereits während ihres Studiums, um zu verhindern, dass Referendar*innen in ihrem zweiten Ausbildungsabschnitt ins kalte Wasser geworfen werden, wie dies derzeit leider häufig der Fall ist.



Abb. 1. Beispiel einer praktischen Notenvergabe

5. Literatur

- Arendt, M. (2006). Beurteilung mündlicher Leistungen. Eine Untersuchung. *Praxis Fremdsprachenunterricht*, 3(3), 3-10 und 3(4), 3-8.
- Bach, G. & Timm, H.-P. (Hrsg.). (2013). *Englischunterricht. Grundlagen und Methoden einer handlungsorientierten Unterrichtspraxis* (5. Aufl.). Tübingen: A. Francke.
- Buschmann, R. (2009). Beteiligung von Schülern an der Bewertung von Leistung. *Pädagogik*, 61(6), 22-25.
- Grimm, N., Meyer, M. & Volkmann, L. (2015). *Teaching English*. Tübingen: Narr Francke Attempto.
- Haß, F. (Hrsg.). (2006). *Fachdidaktik Englisch. Tradition, Innovation, Praxis*. Stuttgart: Ernst Klett Sprachen.
- Heflebower, T., Hoegh, J. K. & Warrick, Ph. (2014). *A School Leader's Guide to Standards-Based Grading*. Bloomington IND: Marzano Research.
- Jäger, R. (2000). *Von der Beobachtung zur Notengebung. Ein Lehrbuch. Diagnostik und Bewertung in der Aus-, Fort- und Weiterbildung*. Landau: Verlag Empirische Pädagogik.
- Jürgens, E. & Sacher, W. (2008). *Leistungserziehung und pädagogische Diagnostik in der Schule. Grundlagen und Anregungen für die Praxis*. Stuttgart: Kohlhammer.
- Korrekturrichtlinien = Ministerium für Kultus, Jugend und Sport Baden-Württemberg (2022, April). *Beurteilungs- und Korrekturrichtlinien für die Abiturprüfung an den allgemein bildenden Gymnasien, gültig für die Abiturprüfung 2022*. Abgerufen am 21. Januar 2023 von <https://www.km-bw.de/site/pbs-bw-new/get/documents/KULTUS.Dachmandant/KULTUS/KM-Homepage/Artikelseiten%20KP-KM/Schularten/Gymnasium/Abitur%202022/Beurteilungs-%20und%20Korrekturrichtlinienerlass%202022.pdf>.
- Nieweler, A. (Hrsg.). (2006). *Fachdidaktik Französisch. Tradition, Innovation, Praxis*. Stuttgart: Ernst Klett Sprachen.
- Reip, S. & Gayer, B. (2012). *Schul- und Beamtenrecht in Baden-Württemberg* (10. Aufl.). Haan-Gruiten: Europa Lehrmittel.
- Walvoord, B. E. & Johnson Anderson, V. (2011). *Effective Grading. A Tool for Learning and Assessment in College*. San Francisco: John Wiley and Sons.
- Wengert, H. G. (2011). Leistungsbeurteilung in der Schule. In G. Bovet & V. Huwendiek (Hrsg.), *Leitfaden Schulpraxis. Pädagogik und Psychologie für den Lehrberuf* (6. Aufl., S. 324-349). Berlin: Cornelsen.

Angaben zum Autor:

Dr. Stefan Hofstetter, Universität Tübingen, Englisch Seminar, Didaktik und Linguistik des Englischen, stefan.hofstetter@uni-tuebingen.de

Marco Magirius und Fabian Wiez

Zum Bewerten von Interpretationen im Deutschabitur

1. Einleitung

Im vorliegenden Beitrag fragen wir nach den besonderen Herausforderungen, die das Bewerten von Interpretationsaufsätzen insbesondere im Rahmen von Abiturprüfungen mit sich bringt. Hierzu beschreiben wir literarische Texte als konstitutiv mehrdeutige und ästhetische Gegenstände. Diese Eigenschaften und die sich daraus ergebende freie Textrezeption werden mit den Anforderungen einer schulischen Prüfungssituation kontrastiert. Zum Schluss stellen wir Ansätze zum Umgang mit diesen Herausforderungen vor, wie sie insbesondere in der literaturdidaktischen Aufgabenforschung diskutiert werden.

2. Herausforderungen bei der Bewertung von Interpretationsaufsätzen

2.1. Literarische Texte als polyvalente und ästhetische Lerngegenstände

Mit Blick auf das Bewerten von Schülerleistungen beim Interpretieren literarischer Texte ergibt sich folgendes grundlegende Problem: Wie geht man mit Deutungsspielräumen um? Dieses Problem ist unhintergebar mit literarischen Gegenständen verbunden, da die Rezeption ebendieser von der Individualität der Rezipient*innen gesteuert wird – ihren Werthaltungen, ihren Erfahrungen etc. (Berger, 2010). Zwar findet bei allen Leseprozessen neben einem ‚Herauslesen‘ auch ein ‚Hineinlesen‘ statt (Köster, 2006a), gerade bei literarischen Texten spielt das individuelle „Wertungsgeschehen“ (Berger, 2010, S. 100, Magirius & Führer, 2023) aber eine entscheidende Rolle.

Literarische Texte legen eine ästhetische Wahrnehmung nahe. Diese beschreibt Seel (2003/2019, S. 223) mit Kant gesprochen als „interesselose Betrachtung“. Damit ist nicht gemeint, dass das Interesse am Gegenstand unterbleiben sollte – im Gegenteil. Man interessiert sich bei der ästhetischen Wahrnehmung für die Wirkung der „individuellen Konstellation“ (ebd., S. 176) seiner Elemente und ihres „Zusammenspiel[s]“ (ebd., S. 94). Diese Wirkung wird nicht bloß theoretisch angenommen, sondern erlebt. Hierfür ist eine Rezeptionshaltung erforderlich, die sich auf die Wirkung des Gegenstandes einlässt. Diese Haltung begnügt sich nicht mit einem „gedanklichen Festhalten“, sondern „verweilt in einem sinnlichen Verfolgen“ (ebd., S. 90). Das Interesse am Gegenstand richtet sich nicht bloß auf

Bestimmung und Bestimmbarkeit. Stattdessen geraten Aspekte seiner „Unbestimmbarkeit“ (ebd., S. 95) in den Blick. Doch auch ästhetische Wahrnehmung zielt auf Erkenntnis. Bei dieser soll der ästhetische Gegenstand aber nicht nur klassifiziert und kategorisiert (vgl. ebd., S. 70), sondern sinnlich erfahren und so in seiner Individualität verstanden werden. Angewandt auf literarisches Verstehen ist hierzu ein „forschende[r] Blick“ (Szondi, 1978, S. 267) vonnöten, der nicht mit fixen, sondern veränderlichen Begriffen – beispielsweise von Textsorten und Epochen – operiert, welche in Wechselwirkung mit dem literarischen Text Veränderung erfahren können. Unter „Forschung“ versteht Szondi insbesondere das Stellen von „Fragen“ (ebd.). Fragen an den Text und damit nach dem Textsinn entstehen laut aktuellen literaturdidaktischen Beiträgen durch „Staunen“, „Irritation“ (Freudenberg & Lessing-Sattari, 2020) oder auch „Stolpern“ (Freudenberg, 2014). Besonders prominent ist der Begriff der „Verstrickung“, den bereits Kreft (1977, S. 379) von Iser übernommen hat (vgl. auch Zabka, 2015, S. 137) und der die affizierte Rezeption beschreibt, welche am Anfang der Textauseinandersetzung stehen sollte. Die Bedeutung der Verstrickung für das Textverstehen ist aus der Perspektive der meisten Literaturdidaktiker*innen konsensfähig (kritisch hierzu: Carl, 2018) und kognitionswissenschaftlich begründbar (Magirius & Führer, 2023).

Diese Überlegungen sind besonders mit Lernsituationen vereinbar, in denen der interpretative Umgang mit Texten gelernt wird (vgl. Köster, 2003, S. 19f). Prüfungssituationen dagegen stellen die Schüler*innen vor besondere Herausforderungen. Nachdem die ästhetische Wahrnehmung inklusive Wertungsprozessen idealerweise dazu verholfen hat, einen Zugang zum Text zu finden (vgl. Berger, 2010, S. 107) und eine sogenannte Entdeckungsprozedur anzustoßen und zu begleiten, ist im Rahmen der Prüfung auch eine Rechtfertigungsprozedur zu bewältigen. In dieser wird der Versuch unternommen, „einen mehr als möglichen, den wahrscheinlichen, den – bis auf weiteres – gültigen Sinn dieses Textes [argumentativ] herauszuarbeiten“ (Kurz, 2018, S. 258).

2.2. Literarische Texte als Prüfungsgegenstände

In Prüfungssituationen reicht der von der Individualität der Schüler*innen – z. B. ihrem Emotionsreservoir, ihren Lektüreerfahrungen und Werthaltungen – geprägte Vollzug beider Prozeduren allerdings nicht aus. Der Fokus muss darauf liegen, einen literarischen Text nicht nur zu durchdringen, sondern die Durchdringung für Beurteilende nachvollziehbar werden zu lassen. Der erhöhte Druck der Prüfungssituation kann die Schüler*innen dazu verleiten, bekannte, konventionelle Zusammenhänge zu präsentieren, statt sie anhand des konkreten Gegenstandes mutig infrage zu stellen. Auch kann dieser zusätzliche Druck schon auf die

Entdeckungsprozedur einwirken, sodass von einer im obigen Sinne „interesselosen Betrachtung“ kaum mehr die Rede sein kann.

Prüfungssituationen, die zur Bewertung literarischen Verstehens führen sollen, stellen auch die bewertenden Lehrkräfte vor besondere Herausforderungen. Die Bewertung der bloßen Deskription des literarischen Textes, der klassifizierenden Analyse seines Soseins, die als Teil eines Interpretationsaufsatzes üblicherweise gefordert wird, stellt das kleinere Problem für die bewertenden Lehrkräfte dar. Es gehört jedoch zum Konsens der Literaturdidaktik, dass aus der Textanalyse allein kein Textsinn entsteht. Die eigentliche Deutung als Beantwortung der Frage, „[w]arum [...] der Text] so [ist], wie er ist“ (Kindt, 2015, S. 96), stellt den Kern der Interpretation und damit des zu bewertenden Schreibprodukts dar. Besonders dann, wenn die Schüler*innen ihr Wissen nicht nur schematisch anwenden, sondern in Wechselwirkung mit dem literarischen Text eine Aktualisierung des Wissens zulassen – wenn sie also mit einem forschenden Blick auf den Text schauen –, ist die Lehrkraft gefordert, dies nicht als deviant, sondern als wünschenswert zu markieren. Eine positive Bestimmung fällt Lehrkräften im Allgemeinen aber schwer (vgl. Kurz, 2018, S. 12). Wenn eine Interpretation nicht „plausibel“ (Zabka, 1999, S. 9f.) ist, d. h. wenn sie in sich nicht stimmig ist oder am Text falsifiziert werden kann, oder wenn sie zwar plausibel, aber nicht mit dem Entstehungskontext kompatibel – nicht „angemessen“ (ebd.) – ist, dann mag es zwar leichtfallen, sie zu bemängeln. Im Allgemeinen entziehen sich Interpretationsleistungen aber einer simplen Richtig-falsch-Dichotomie.

Fraglich bleibt, inwiefern die eben genannten Herausforderungen es erlauben, für Prüfungen Lösungserwartungen zu formulieren. Dies ist hochgradig abhängig von der Aufgabenstellung.

3. Ansätze zum Umgang mit den genannten Herausforderungen

3.1. Aufgabenstellungen

Die besonderen Herausforderungen, welche Interpretationsaufgaben stellen, lassen sich mit folgenden zwei Kategorien der literaturdidaktischen Aufgabenforschung beschreiben (vgl. Winkler, 2010, S. 108). Erstens bezeichnet der Begriff „Entscheidungsspielraum“ den Grad der Determiniertheit von Ausgangssituation, Prozess und Ergebnis der Aufgabenlösung. Zweitens spricht man vom sogenannten „Integrationsgrad“. Dieser ist abhängig von der Anzahl der zur Aufgabenlösung relevanten Variablen (Textstellen, Vorwissenselemente) und dem Grad ihrer Vernetztheit (vgl. ebd.). Traditionell ist angenommen worden, dass mehrdeutige Gegenstände Aufgabenstellungen mit großem Entscheidungsspielraum benötigen.

Vor Aufgaben mit zu starker Lenkung ist dagegen verschiedentlich gewarnt worden (vgl. dazu Steinmetz, 2020, S. 6). Wenn in einer Aufgabe jedoch verlangt wird, einen Text zu interpretieren, ohne dass dabei eine Lenkung stattfindet, ist sowohl der Entscheidungsspielraum besonders groß als auch der Integrationsgrad besonders hoch. Dann besteht die Gefahr der Überforderung und einer eigenmächtig von den Schüler*innen vorgenommenen „wilde[n] Komplexitätsreduktion“ der Aufgabe (vgl. Köster & Lindauer, 2008, S. 153). Dies kann zu Interpretationen führen, die sehr schematisch auf jenes Vorwissen zurückgreifen, welches besonders zu den Erwartungen passt, die die Schüler*innen auf die prüfenden Lehrkräfte projizieren (vgl. Stark & Zabka, 2012, S. 21). Die Frage, ob diese Interpretationen tatsächlich mit dem literarischen Text kompatibel sind, tritt bei der Aufgabenbearbeitung in den Hintergrund – ein forschender Blick bleibt aus.

Um diese wilden Komplexitätsreduktionen zu vermeiden, schlägt die Aufgabenforschung Support als Teil der Aufgabenstellung vor (Steinmetz, 2020). Mit Blick auf Abituraufsätze differenzieren Stark und Zabka (2010) verschiedene Supportmöglichkeiten zur Steuerung der Wissensanwendung wie z. B. den Hinweis auf Kontexte, die Vorgabe eines Vergleichstexts und/oder eines Themas oder eines Analysekriteriums oder auch das Mitliefern relevanten Kontextwissens. Ferner besteht die Möglichkeit zur Zugabe weiterer Primärliteratur, aber auch von verschiedenster Sekundärliteratur – etwa von Rezeptionsdokumenten, Urteilen von Kritiker*innen, Auszügen aus wissenschaftlichen Texten oder auch poetologischen Maximen der Autor*innen. Indem beispielsweise die Passung zwischen den genannten Zusatzinformationen und dem Text zu untersuchen ist, richtet sich das Augenmerk der Schüler*innen automatisch auf ebendiesen (vgl. ebd., S. 24), wodurch ein forschender Blick wahrscheinlicher wird.

3.2. Bewertungswerkzeuge

Indem Aufgabenstellungen bestimmte Arbeitsschritte oder Analyseschwerpunkte mittels Support vorgeben, wird die Menge der möglichen Bearbeitungsergebnisse eingegrenzt und das Erstellen von Erwartungshorizonten überhaupt erst sinnvoll möglich.

„Denn je offener und komplexer die Aufgabenstellung, desto schwieriger wird die Generierung guter Korrekturanleitungen. Je mehr Entscheidungsspielraum die Prüflinge haben, desto abstrakter werden die Kriterien der Korrekturanleitung, die dann in vielfältiger Verzweigung beispielhaft ausdifferenziert werden müssen.“ (Köster 2010, S. 32)

Diese Kriterien können sowohl von Lehrkräften für ihre individuelle Beurteilungspraxis erstellt als auch zentral vorgegeben werden. Dabei lassen sich holistische von analytischen Erwartungshorizonten unterscheiden.

Holistische Erwartungshorizonte sollen eine ganzheitliche Beurteilung von Schülertexten unterstützen und wurden somit lange in Form längerer Fließtexte formuliert, die mögliche zu erwartende Schülerleistungen darstellten, sodass beurteilende Lehrkräfte tatsächliche Schülerleistungen an dieser ‚Musterlösung‘ messen konnten. Analytische Erwartungshorizonte zergliedern dagegen zu erwartende – meist gute und ausreichende – Leistungen in einzelne Teilleistungen, sodass Schülertexte bei der Beurteilung an einer Reihe von Items gemessen werden können. So können analytische Erwartungshorizonte ein breiteres Spektrum möglicher Lösungen angeben, da sie nicht die Form eines zusammenhängenden Fließtextes haben, der seinerseits um eine zusammenhängende Argumentation bemüht sein muss. Inzwischen werden vor allem Mischformen der beiden Varianten verwendet, die erwartete Leistungen in Form isolierter Items angeben. Analytischere Formen zeichnen sich hier dadurch aus, dass sie durch die Verbindung einzelner Items mit einer bestimmten Punktzahl die Gewichtung der Teilleistungen mit vorgeben, während holistischere Varianten die Gewichtung den beurteilenden Lehrkräften überlassen.

Aufgrund des oben beschriebenen Entscheidungsspielraums und der Komplexität der zu lösenden Aufgaben gilt für alle Arten von Erwartungshorizonten, dass sie offen für individuelle Deutungsmöglichkeiten sein müssen. Sie können und dürfen also nicht den Eindruck erwecken, ‚die eine‘ richtige Lösung einer Aufgabe darzustellen, und sind daher mit entsprechenden Metakommentaren zu versehen (vgl. Zabka & Stark, 2010, S. 26). Zugleich sollten sie aber so formuliert sein, dass die Geltungskraft der vorgeschlagenen Lösungen klar wird und nicht von den beurteilenden Lehrkräften vermutet werden muss (vgl. ebd.). Bei der Formulierung von Erwartungshorizonten sollte deshalb möglichst genau angegeben werden, welche Teilleistungen von den Schüler*innen erbracht werden müssen und welche nicht obligatorisch, sondern lediglich möglich sind und daher Alternativen zulassen (vgl. ebd., S. 28). Diese Erwartungen an bestimmte Teilleistungen ergeben sich aus der jeweiligen Aufgabenstellung und dem verwendeten Operator, wobei nicht jede Teilleistung explizit in der Aufgabenstellung gefordert werden muss. Es wird vorausgesetzt, dass die Schüler*innen mit den Operatoren soweit vertraut sind, dass sie wissen, welche Teilleistungen der jeweilige Operator erfordert.

Um möglichst offen für individuelle Lösungen zu sein, geben Erwartungshorizonte erwartete Teilleistungen meist zunächst abstrakt an. Sie geben so lediglich eine grobe Struktur des zu verfassenden Interpretationstextes vor. Auf einer zweiten Ebene werden die abstrakten Teilleistungen dann durch konkrete, am Gegenstand

orientierte Verstehensziele ergänzt. So heißt es etwa in einem Erwartungshorizont zu einer Interpretationsaufgabe zu Hilde Domins Gedicht *Vorsichtige Hoffnung* des öffentlich zugänglichen Pools von Abituraufgaben des Instituts zur Qualitätsentwicklung im Bildungswesen zunächst: „Die Schülerinnen und Schüler [...] formulieren das Thema bzw. eine Deutungsthese des Gedichts, etwa [...]“ (IQB, 2021, S. 4). Unter dieser abstrakten Erwartung werden dann zwei konkrete Vorschläge gemacht, wie das Thema des Gedichts lauten könnte. Diese haben dann meist den Status von möglichen und nicht zwingenden Lösungen. Neben dieser Inhaltsleistung formulieren Erwartungshorizonte Kriterien für die Darstellungsleistung, z. B. bezüglich bestimmter Merkmale des Schülertextes, der korrekten Verwendung der Standardsprache und relevanter Fachbegriffe. Da so nicht alle möglichen Deutungsergebnisse angegeben werden können, müssen Lehrkräfte in tatsächlichen Beurteilungssituationen Schülerleistungen als mehr oder weniger passend zu den angegebenen Kriterien bewerten (vgl. Kötter-Mathes, 2020, S. 65). Dabei müssen sie die Erwartungshorizonte gegebenenfalls individuell ergänzen, um eine Passung zwischen abstrakten Kriterien und konkreten Schülertexten herzustellen. Diese Auslegungen der Erwartungshorizonte sind stark von den individuellen Merkmalen der jeweiligen Lehrkräfte beeinflusst. So sind Lehrkräfte in unterschiedlichem Maße dazu in der Lage, die Erwartungshorizonte zu verstehen und auf einen konkreten Schülertext anzuwenden (vgl. ebd., S. 75). Die bewertenden Lehrkräfte neigen darüber hinaus ihrerseits zu individuellen Lesarten der Texte und bestimmter Kontexte, aus denen sie dem literarischen Texten Bedeutung zuweisen. Sie haben ferner eigene Werthaltungen und nicht zuletzt Zielvorstellungen bezüglich ihres Literaturunterrichts, deren Erreichen sie überprüfen wollen.

Literarisches Verstehen ist als zu prüfende Fähigkeit aufgrund der für das Interpretieren zentralen Subjektivität der Prüflinge nicht „restlos objektivierbar“ (Berger, 2010, S. 101). Mehrdeutige und ästhetische Gegenstände erfordern Interpretationsfreiheit. Deshalb ist das „Postulat der Freiheit und Konstruktivität des Schülersubjekts“ unbedingt gegenüber dem „Postulat normierter Leistungsstandards“ (Köster, 2006b, S. 80) aufrechtzuhalten. Doch auch auf Seiten der Lehrkräfte muss jede Bewertung durch das „Nadelöhr der Subjektivität“ (Begriff von Radtke 1996, S. 39) gehen. Deshalb kann man davon ausgehen, dass unterschiedliche Lehrkräfte auch beim Einsatz von Erwartungshorizonten ein und dieselbe Schülerleistung unterschiedlich beurteilen. Die genannten Lösungsansätze können deshalb lediglich dahingehend unterstützend wirken, die unhintergehbare Subjektivität der Bewertung mit Prüfungsgerechtigkeit auszubalancieren.

4. Literatur

Berger, T. (2010). Nachweis und Bewertung literarischen Verstehens in schulischen Leistungssituationen und Testaufgaben. In I. Winkler, N. Masanek & U. Abraham (Hrsg.), *Poetisches Verstehen: Literaturdidaktische Positionen – empirische Forschung – Projekte aus dem Deutschunterricht* (S. 98-113). Baltmannsweiler: Schneider Verlag Hohengehren.

Carl, M.-O. (2018). Stolpern oder jonglieren, staunen oder spielen? Eine kognitionspsychologische wie hermeneutische Kritik der ästhetischen Theorie analogischen Verstehens. In D. Scherf & A. Bertschi-Kaufmann (Hrsg.), *Ästhetische Rezeptionsprozesse aus didaktischer Perspektive* (S. 41-55). Weinheim: Juventa,

Freudenberg, R. (2014). „Arme, arme Schlange!“ Ein Aufruf zur Ausbildung von Irritationsbereitschaft. *kj&m*, 66(1), 60-68.

Freudenberg, R., Lessing-Sattari, M. (2021). Zur Einführung. In R. Freudenberg & M. Lessing-Sattari (Hrsg.), *Zur Rolle von Irritation und Staunen im Rahmen literärästhetischer Erfahrung. Theoretische Perspektiven, empiriebasierte Beobachtungen und praktische Implikationen* (S. 7-16). Frankfurt a.M.: Peter Lang.

IQB (2021). *Gemeinsame Aufgabenpools der Länder. Pool für das Jahr 2021. Aufgabe für das Fach Deutsch*. <https://www.iqb.hu-berlin.de/abitur/pools2021/deutsch/> [25.01.2023].

Jakobson, R. O. (1979/2016). *Poetik: ausgewählte Aufsätze, 1921-1971*. Frankfurt a.M.: Suhrkamp.

Kindt, T. (2015). Deskription und Interpretation. Handlungstheoretische und praxeologische Reflexionen zu einer grundlegenden Unterscheidung. In M. Lessing-Sattari, M. Löhden & D. Wieser, (Hrsg.), *Interpretationskulturen. Literaturdidaktik und Literaturwissenschaft im Dialog über Theorie und Praxis des Interpretierens* (S. 93-112). Frankfurt a.M.: Peter Lang.

Köster, J. (2006a). Von der Lebenswelt zur Literatur. Zu Erich Kästner „Fauler Zauber“ (4. Schuljahr). In C. Kammler (Hrsg.), *Literarische Kompetenzen – Standards im Literaturunterricht. Modelle für die Primar- und Sekundarstufe* (S. 50-65). Seelze: Kallmeyer (in Verbindung mit Klett).

Köster, J. (2006b). Das Deutschabitur in Zeiten von Bildungsstandards – Vergleichbarkeit von Prüfungsleistungen und ihrer Bewertung. *Didaktik Deutsch*, 11(21), 78-90.

Köster, J. (2010). Leistung und Qualität von Korrekturanleitungen im Zentralabitur Deutsch. *Der Deutschunterricht*, 62(1), 31-39.

Köster, J. & Lindauer, T. (2008). Zum Stand wissenschaftlicher Aufgabenreflexion aus deutschdidaktischer Perspektive. *Didaktik Deutsch*, 13(25), 148-161.

Kötter-Mathes, S. (2020). *Leistungsbeurteilung in zentralen Prüfungen*. Wiesbaden: Springer Fachmedien.

Kreft, J. (1977). *Grundprobleme der Literaturdidaktik: Eine Fachdidaktik im Konzept sozialer und individueller Entwicklung und Geschichte*. Heidelberg: Quelle und Meyer.

Kurz, G. (2018). *Hermeneutische Künste*. Stuttgart: J. B. Metzler.

Magirius, M. (2022). Zum Umgang mit der Relativität literarischen Verstehens – Positionen von Deutschlehrkräften und -studierenden. In C. Führer, F. Schweitzer, B. Tesch, B. Eiben-Zach, F. Ulfat, P. Thomas, W. Polleichtner, B. Grewe & U. Küchler, (Hrsg.), *Relativität und Bildung* (S. 157-168). Münster: Waxmann.

Magirius, M. & Führer, C. (2023). Evaluative ästhetische Rezeption und Bildung – Einleitung in den Band. In M. Magirius, C. Meier, S. Kubik & C. Führer, (Hrsg.), *Ästhetische evaluative Rezeption* (S. 7-24). München: Kopaed.

Radtke, F.-O. (1996). *Wissen und Können. Die Rolle der Erziehungswissenschaft in der Erziehung*. Opladen: Leske und Budrich.

Schmidt, S. J. (1991). *Grundriß der Empirischen Literaturwissenschaft*. Frankfurt a.M.: Suhrkamp.

Seel, M. (2003/2019). *Ästhetik des Erscheinens* (6. Aufl.). Frankfurt a.M.: Suhrkamp.

Steinmetz, M. (2020). *Verstehensupport im Literaturunterricht. Theoretische und empirische Fundierung einer literaturdidaktischen Aufgabenorientierung*. Wiesbaden: VS Springer.

- Szondi, P. (1978). Über philologische Erkenntnis. In J. Bollack, H. Beese, W. Fietkau, H.-H. Hildebrandt, G. Mattenklott, S. Metz & H. Stierlin (Hrsg.), *Schriften I* (S. 263-344). Frankfurt a.M.: Suhrkamp.
- Zabka, T. (1999). Subjektive und objektive Bedeutung. Vorschläge zur Vermeidung eines konstruktivistischen Irrtums in der Literaturdidaktik. *Didaktik Deutsch*, 4(7), 4-23.
- Zabka, T. (2003). Interpretationskompetenz als Ziel der ästhetischen Bildung, *Didaktik Deutsch*, 8(15), 18-32.
- Zabka, T. (2015). Was ist Hochschulreife im Umgang mit Literatur?, *Didaktik Deutsch*, 20(38), 136-150.
- Zabka, T., Stark, T. (2010). Aufgabenstellungen und Erwartungshorizonte als Steuerungsinstrumente. Zum Umgang mit Problemen der Literaturinterpretation im Zentralabitur. *Der Deutschunterricht*, 62(1), 19-29.

Angaben zu den Autoren:

- Dr. Marco Magirius, Vertretung des Fachbereichs Literaturdidaktik/Neuere deutsche Literatur am Institut für Deutsche und Niederländische Philologie der FU Berlin, mail@marcomagirius.de
- Fabian Wiez, Lehramtsreferendar in Hamburg und Doktorand an der Universität Hamburg, Didaktik der sprachlichen und ästhetischen Fächer (EW 4), fabian.wiez@uni-hamburg.de

Helga Gese

Form oder Funktion? Zur Begründung von Korrektornormen für grammatische Leistungen

1. Einleitung

Im Vergleich zur Bewertung inhaltlicher Leistungen erscheint die Bewertung sprachlicher Leistungen im Deutschunterricht zunächst als ein Bereich mit besonders stabilen und klaren Bezugsnormen. So mag es verwundern, dass die Diskussion um die Bewertungspraxis in der Sprachdidaktik – und hier insbesondere zu grammatischen Leistungen – nicht abreißt. Dreh- und Angelpunkt dieser Diskussion ist die Frage, inwieweit Lehrkräfte bei der Korrektur grammatischer Leistungen als „zu normative“ (Kilian, 2021, S. 501) „Sprachnormautoritäten“ (Ammon, 2005, S. 36) auftreten. Wann ist eine Korrektur ‚zu normativ‘ und auf welche Normen können sich Lehrkräfte bei der Bewertung sprachlicher Leistungen überhaupt stützen? Der vorliegende Beitrag plädiert für ein gestuftes Vorgehen: Zunächst ist es didaktisch legitim, Korrekturen an formbezogenen Normen auszurichten, auch wenn diese linguistisch betrachtet als zu streng erscheinen. Diese nur übergangsweise geltende Formorientierung stellt eine Transitnorm und nicht die eigentlichen Zielnorm dar. Sie sollte im Lernverlauf schrittweise durch eine Orientierung an linguistisch begründeten Normen der funktionalen Angemessenheit abgelöst werden.

2. Herausforderungen beim Bewerten sprachlicher Leistungen

2.1. Deutschlehrkräfte als Sprachnormautoritäten

Beginnen wir mit dem *locus communis* jeder Auseinandersetzung mit der Bewertung sprachlicher Leistungen im Deutschunterricht: Deutschlehrkräfte sind „Sprachnormautoritäten“ (ebd.). Erst durch Sprachnormautoritäten, „existieren Sprachnormen als solche“, da nur sie „über ausreichende Macht verfügen oder dies glaubhaft machen können, um das Sprachhandeln anderer Personen (der Normsubjekte) zu korrigieren“ (ebd.). Der Status als Sprachnormautorität ist dabei weniger Privileg als Verpflichtung.

Der Begriff der Sprachnorm ist vielschichtig. Grundlegend für ein Verständnis der Kontroversen um die Bewertungspraxis von Deutschlehrkräften sind zwei Verwendungen: Zurückgehend auf Coserius Unterscheidung von System (den in einer Sprache möglichen grammatischen Strukturen) und Norm (den in einer Sprache

tatsächlich realisierten Sprachstrukturen) dominiert in der Linguistik ein deskriptiver Normbegriff, der Normen als das in einer Sprache Gebräuchliche versteht (Coseriu, 1975). In didaktischen Kontexten ist dagegen ein präskriptiver Normbegriff anzusetzen, der Sprachnormen als sozial konstituierte Erwartungen an die Sprachverwendung versteht (Gloy, 2004). Solche präskriptiv verstandenen Sprachnormen können sich in ihrer Begründung wiederum auf unterschiedliche sprachlich-kommunikative Domänen beziehen (vgl. Beuge, 2019, S. 56ff.): Geht es um die Erwartung, dass sprachliche Produkte „so gestaltet sein müssen, dass eine Rezeption und Interpretation möglich und problemlos in einer Sprachgemeinschaft gegeben ist“ (ebd.), so haben wir es mit grammatisch-semantischen Normen zu tun. Diese sind für die Bewertung sprachlicher Formen grundlegend. Angemessenheitsnormen wiederum bewerten „den Sprachgebrauch eines Individuums in Rückbindung an die jeweilige Sprachgebrauchssituation hinsichtlich funktionaler Kriterien“ (ebd. S. 58). Zusätzlich können noch „Normen der Äußerungsqualität“ (Hartung, 1977, S. 37), z. B. „ästhetische Normen“ (Fix, 1995, S. 37), herangezogen werden. Es stellt sich die Frage, wann welches Normkonzept angelegt werden sollte.

In Deutschland bilden die nationalen Bildungsstandards die rechtliche Basis für die Bewertungspraxis. Nach diesen ist die „sprachliche Richtigkeit“ (KMK, 2012, S. 23) wesentliches Kriterium für die Korrektur schriftlicher Arbeiten im Fach Deutsch. Für die Bewertung entscheidend ist zudem die Frage der Erfüllung standardsprachlicher Normen: So ist eine Abiturarbeit z. B. dann als sprachlich gut zu bewerten, wenn darin „die sichere Beherrschung standardsprachlicher Normen nachgewiesen wird“ (KMK, 2012, S. 42). „Abweichungen von der Standardsprache [werden nur dann] akzeptiert, wenn sie deutlich einem Gestaltungswillen untergeordnet sind“ (KMK, 2004, S. 30). Ein solches legitimes Abweichen von der Standardsprache zu erkennen, ist freilich alles andere als leicht.

2.2. Das Spannungsfeld von Standardsprache und Sprachvariation als Herausforderung für die Bewertung

Wenn Lehrkräfte qua Amtes in ihren Korrekturen verpflichtet sind, die Erfüllung standardsprachlicher Normen zu beurteilen, so „stellt sich besonders dringlich die Frage, welche Sprachformen denn standardsprachlich sind und welche nicht“ (Ammon, 2005, S. 28). Untersuchungen zur Korrekturpraxis deuten auf „eine hohe Variabilität in der Fehlermarkierung“ (Hennig, 2012, S. 145) und einen fehlenden „Korrekturstandard“ (Häcker, 2009) im Bereich der Grammatik (ähnliche Befunde gibt es für Zeichensetzung, Langlotz 2017, und Rechtschreibung, Ransmayr 2021). Dabei ließen sich die Anstreichungen „nicht immer anhand des Kodexes erklären“ (Davies, 2006, S. 486). Hierzu ist anzumerken, dass es für die

Grammatik (anders als für Rechtschreibung und Zeichensetzung) in Deutschland keinen amtlichen Kodex gibt, so ist z. B. die Grammatik des privaten Dudenverlags nicht rechtlich bindend. Der Bereich der Grammatik ist daher stärker als Rechtschreibung und Zeichensetzung durch Sprachvariation geprägt. Sprachvariation wiederum generiert Zweifelsfälle und Ermessensspielräume, die es im Unterrichtsgeschehen zu nutzen gilt (siehe Praxis Deutsch Nr. 264/2017 sowie Gese 2022).

"[I]n der täglichen Korrekturarbeit muss man sich als Lehrer aber entscheiden, ob eine Formulierung noch akzeptabel ist oder man sie mit einem Korrekturzeichen versieht" (Dürscheid, 2011, S. 163f.).

In dieser „Situation [der] Bedrängnis“ (Müller, 2009, S. 64) urteilen Lehrkräfte oftmals strenger als die einschlägigen Kodizes (u. a. Davies, 2006, 2017; Dürscheid, 2011; Häcker, 2009; Hennig, 2012; Müller, 2009; Osterroth, 2015; zu einem anderen Ergebnis kommt z. T. Kilian, 2021). Dies mag auf eine Orientierung an subsistenten Normen zum schulischen Sprachgebrauch und zur jeweils selbst erlebten Korrekturpraxis zurückzuführen zu sein. Diese gelebte Korrekturpraxis nämlich erscheint „in vielen Fällen als ein traditionelles Brauchtum [...], das auf eine strenge Durchsetzung sprachlicher Normvorstellungen ausgerichtet ist“ (Müller, 2009, S. 64). Exemplarisch sei auf Beispiele aus Studien von Davies (2001) verwiesen:

„Der temporale Gebrauch von wo ist [...] sehr unbeliebt: Keine einzige Lehrkraft meint, dass an dem Tag, wo akzeptabel sei, weder in der gesprochenen noch in der geschriebenen Standardsprache. Nur 31.1% akzeptieren welcher als Relativpronomen und noch weniger (15.9%) betrachten bräuchte und brauchen ohne zu als akzeptabel in der Schriftsprache und in der gesprochenen Standardsprache“. (Davies, 2006, S. 487)

Weitere prominente Beispiele für standardsprachlich legitime, aber im Schulkontext tabuisierte Gebrauchsweisen betreffen die Kategorien Tempus und Modus: der Gebrauch des Perfekts in Berichten und Erzählungen und die Verwendung des Indikativs in der indirekten Rede. In all diesen Fällen treten Lehrkräfte, so die Ergebnisse der oben genannten Studien, als „zu normative Normautoritäten auf“ (Kilian, 2021), indem sie Konstruktionen sanktionieren, die in einschlägigen Grammatiken als standardsprachlich geführt werden.

3. Ansätze zu einer didaktisch begründeten und linguistisch fundierten Bewertung grammatischer Leistungen

Normen gewinnen Legitimität durch „ihre Kompatibilität mit anerkannten Werten“ (Ammon, 2005, S. 40). Für Normen der Bewertung grammatischer Leistungen ist ein solcher Wert ohne Zweifel die „systemlinguistische Begründbarkeit“,

d. h. die Grammatikalität: Was ungrammatisch ist, ist bei der Korrektur anzustreichen. Wie oben gezeigt, werden in schulischen Kontexten aber oft weitaus strengere Kriterien angelegt, m. a. W. auch grammatische Konstruktionen werden sanktioniert. Im Folgenden werden zwei Begründungslinien skizziert, auf die sich die Bewertung grammatischer Leistungen jenseits der Grammatikalität stützen kann: 1) der didaktische Nutzen für sprachliche Lehr- und Lernprozesse und 2) die kommunikative Effizienz, d. h. die funktionale Angemessenheit der sprachlichen Mittel.

3.1. Formal ausgerichtete Faustregeln als Transitnormen

Einen Bericht schreibt man im Präteritum. In der indirekten Rede verwendest du den Konjunktiv. Solche und ähnliche Faustregeln kennen Lernende zur Genüge. Sie erscheinen in Bewertungsrastern und prägen die Korrekturpraxis im Deutschunterricht. Das Problem ist, dass solchermaßen streng ausgelegte, rein formal ausgerichtete Sprachnormen den Erwerb einer Textsortenkompetenz eher hemmen als fördern (Kilian, 2021). Andererseits, so betont Kilian (ebd.) mit Verweis auf Fix (2006, S. 91ff.),

„kann eine solche Orientierung an Textsortennormen sprachdidaktisch geboten sein, kann die didaktische Sprachkritik begründen, dass und inwiefern sie das sprachliche Lernen fördern und zum Aufbau sprachlichen Wissens beitragen können“.

Inwiefern ein Agieren als „zu normative Normautorität“, z. B. ein linguistisch nicht gerechtfertigtes Festhalten an rein formal ausgerichteten Faustregeln der Standardsprachlichkeit didaktisch begründet sein könnte, versteht man, wenn man sich Folgendes klar macht: „[I]n einem positiven Sinne wirken Normen komplexitätsreduzierend, stabilisierend, typisierend“ (Gloy, 1980, S. 367). Indem die prinzipiell unendlichen sprachlichen Möglichkeiten auf ein für die Lernenden bewältigbares Maß reduziert werden, liefern Sprachnormen für den Lernenden ein Gerüst, das den Spracherwerb stützt. In diesem Sinne können Sprachnormen im Deutschunterricht als „transitorische Normen“ (Feilke 2012, S. 155) fungieren und als solche legitimiert sein. Es stellt sich allerdings die Frage, bis wann solche schulisch tradierten Transitnormen im Sinne eines Scaffolding-Ansatzes nötig sind. Auch wenn es hierfür keinen fixen Zeitpunkt gibt, so erscheint doch eines sinnvoll: Lernende sollten beim Erstkontakt die Einhaltung dieser Faustregeln streng befolgen: indirekte Rede im Konjunktiv, Beschreiben im Präsens, Berichten und Erzählen im Präteritum; weil als Verbletztsatz-einleitende Subjunktion, *brauchen* mit zu-Infinitiv, *wegen* mit Genitiv. Wichtig erscheint es aber, solche Transitnormen nicht mit

Zielnormen zu verwechseln, sondern als Gerüst, das zu gegebener Zeit wieder abgebaut wird:

„Schulische Sprachnormen und schulische Sprachgebrauchsnormen sind - als schulische - präskriptiv. Allerdings sind sie im Regelfall keine definiten Zielnormen [...] ihr Erwerb kennzeichnet [...] einen didaktisch bestimmten Zustand des Wissens und Könnens, der im Regelfall selbst wieder zu überwinden ist.“ (ebd., S. 155)

3.2. Funktional ausgerichtete Angemessenheitsnormen als Schlüssel zur Bewertung

Es stellt sich die Frage, nach welchem Maßstab die Bewertung erfolgen sollte, nachdem Transitnormen in fortgeschrittenen Lernergruppen wieder ‚abgebaut‘ sind? Hierauf geben die Bildungsstandards Hinweise (siehe der in Kap. 2.1 erwähnte „Gestaltungswille“), die Sprachdidaktik präzisiert: Entscheidend sind funktionale Angemessenheit und kommunikative Effizienz, denn Grammatikbetrachtung, die einen positiven Effekt auf sprachliche Kompetenzen entfalten soll, darf sich nicht in der isolierten Betrachtung von Formen erschöpfen (siehe Funke 2018 zur Auswirkung von Grammatikunterricht auf sprachpraktische Kompetenzen). Vielmehr sind formal-strukturelle Aspekte stets in Verzahnung mit funktionalen Aspekten zu betrachten. Dies gilt für das Unterrichtsgeschehen genauso wie für die Korrekturpraxis, denn auch bei der Korrektur geht es letztlich darum, „bei den Schülerinnen und Schülern ein Bewusstsein dafür zu schaffen, was als situationsangemessenes Sprechen und Schreiben gilt“ (Dürscheid, 2013, S. 273). Lehrkräfte sollten somit zumindest in höheren Klassenstufen die rein formal ausgerichteten sprachlichen Transitnormen zurücknehmen und im sprachreflexiven Unterricht wie bei der Korrektur einen funktionalen Blickwinkel einnehmen. Sehen wir uns die folgenden Beispiele an:

- (1) Er bekam eine Mahnung, obwohl er die Rechnung bezahlt gehabt hat/hatte.
- (2) Wenn er eine Mahnung bekommen hätte, obwohl er die Rechnung bezahlt gehabt hätte, wäre er sicher verärgert.
- (3) Wenn er eine Mahnung bekommen hätte, obwohl er die Rechnung bezahlt hätte, wäre er sicher verärgert.

Doppelperfekte und Doppelpräteritumperfekte wie in (1) und (2) gelten in der geschriebenen Standardsprache als nicht korrekt. In bestimmten Verwendungskontexten ist eine Verwendung dieser Formen aber funktional legitimiert. Dies ist z. B. dann der Fall, wenn im Konjunktiv auf die Vorvergangenheit Bezug genommen werden soll, vgl. (2). Das deutsche Tempussystem ist hier auf die doppelten

Vergangenheitstempora angewiesen, denn das standardsprachliche Präteritumperfekt drückt im Konjunktiv keine Vorvergangenheit aus, vgl. (3). Auch im Indikativ kann das Doppelpräteritumperfekt funktional legitimiert sein: Als Ausdruck der Vor-Vorvergangenheit „kommt [ihm] eine einzigartige – wenn auch selten ausgenutzte – Position im Rahmen des deutschen Tempussystems zu.“ (Breuer & Dorow 1996, zitiert nach Duden 2016, S. 526) Ohne diese kontextuelle Legitimation sind die Doppel-Vergangenheitstempora nur in der gesprochenen Sprache angemessen – im schulischen Kontext z. B. in einem verschrifteten Dialog, in dem Umgangssprache zur Figurenkonturierung verwendet wird.

Bei der Beurteilung der Standardsprachlichkeit dürfen Lehrkräfte solche funktionalen Aspekte nicht ausblenden. Hinweise darauf, dass eine funktionale Orientierung der Bewertung sprachlicher Leistungen zumindest in Ansätzen erfolgt, liefert eine Korrekturstudie (n=64). Angehende Lehrkräfte sollten für Sätze wie (1) bis (3) angeben, ob sie die Verwendung für (i) korrekt, (ii) für eine nicht als Fehler zu sanktionierende/für sie unklare Abweichung oder (iii) für einen Fehler halten. Zudem wurden Korrekturbegründungen erhoben. Auch wenn die Doppelformen schlechter bewertet wurden als die standardsprachlich etablierten Tempora (und dies auch in Kontexten, in denen die Standardtempora funktional unterlegen sind, vgl. (2)/(3) oben), so zeigen die Ergebnisse doch einen Unterschied zwischen kontextuell legitimierten, funktional angemessenen Verwendungen, Bsp. (2), und solchen Verwendungen, in denen die Doppelformen funktional keinen Mehrwert liefern, vgl. (1) oben: Erstere wurden besser bewertet als letztere. Dieser deskriptiv zwar nur geringe Bewertungsunterschied erweist sich in der statistischen Analyse mittels ANOVA als signifikant ($F(1,63) = 7.077$, $p = .010$). Interessant sind in diesem Kontext die Begründungen, die die angehenden Lehrkräfte für ihre Einschätzungen gaben: Zwar sind diese mehrheitlich als formal zu klassifizieren (Bsp. „die Form ist falsch gebildet“), allerdings ist die zweithäufigste Begründungskategorie Funktion (Bsp. „die Form dient hier dem Ausdruck von Vorvergangenheit“). Erst danach folgen Begründungen, die sich auf Register/Stil (Bsp. „die Form ist umgangssprachlich“), Sprachgefühl o. a. („hört sich falsch an“) beziehen. Die Bewertungsstudie liefert erste Hinweise, dass (angehende) Lehrkräfte die Funktion als Bewertungskriterium in ihre Korrekturen einbeziehen, dass dieser Einbezug aber noch gestärkt werden sollte.

Lehrkräfte darin zu bestärken, die Einschätzung der funktionalen Angemessenheit nichtstandardsprachlicher grammatischer Erscheinungen fachwissenschaftlich fundiert und fachdidaktisch begründet zu treffen, war ein Anliegen des vorliegenden Beitrags. Entscheidend für die Bewertung grammatischer Leistungen sollte letztlich die Antwort auf folgende zwei Fragen sein: (i) Ist das

Ansetzen eines rein formbezogenen Maßstabs unter Ausblendung funktionaler Aspekte in einer Lerngruppe im Sinne einer Transitnorm noch zielführend? Wenn nein: (ii) Liefert die Verwendung einer nichtstandardsprachlichen bzw. grammatisch-peripheren Konstruktionen einen funktionalen Mehrwert?

4. Literatur

Ammon, U. (2005). Standard und Variation: Norm, Autorität, Legitimation. In L. M. Eichinger & W. Kallmeyer (Hrsg.), *Standardvariation. Wie viel Variation verträgt die deutsche Sprache?* (S. 469-520). Berlin: De Gruyter.

Beuge, P. (2019). *Was ist gutes Deutsch? Eine qualitative Analyse laienlinguistischen Sprachnormwissens*. Berlin: De Gruyter.

Coseriu, E. (1975). System, Norm und Rede [1952]. In E. Coseriu (Hrsg.), *Sprachtheorie und allgemeine Sprachwissenschaft. Fünf Studien* (S. 11-101). München: Fink.

Davies, W. (2001). Standardisation and the school: norm tolerance in the educational domain. *Linguistische Berichte*, 188, 393-414.

Davies, W. (2006). Normbewusstsein, Normkenntnis und Normtoleranz von Deutschlehrkräften. In E. Neuland (Hrsg.), *Variation im heutigen Deutsch: Perspektiven für den Deutschunterricht* (S. 483-491). Frankfurt a.M.: Peter Lang.

Davies, W. & Langer, N. (2014). Die Sprachnormfrage im Deutschunterricht: das Dilemma der Lehrenden. In A. Plewia & A. Witt (Hrsg.), *Sprachverfall? Dynamik – Wandel – Variation* (S. 299-321). Berlin: De Gruyter.

Davies, W. (2017). Gymnasiallehrkräfte in Nordrheinwestfalen als SprachnormvermittlerInnen und Sprachnormautoritäten. In W. Davies, A. Häcki Buhofer, R. Schmidlin, M. Wagner, M. & E. L. Wyss (Hrsg.), *Standardsprache zwischen Norm und Praxis. Theoretische Betrachtungen, empirische Studien und sprachdidaktische Ausblicke* (S. 123-146). Tübingen: Narr.

Dürscheid, C. (2011). Zweifeln als Chance? Zweifeln als Problem? Sprachliche Zweifelsfälle im Deutschunterricht. In K.-M. Köpcke & A. Ziegler (Hrsg.), *Grammatik – Lehren, Lernen, Verstehen. Zugänge zur Grammatik des Gegenwartsdeutschen* (S. 155-173). Berlin: De Gruyter.

Feilke, H. (2012). Schulsprache - Wie Schule Sprache macht. In S. Günthner, W. Imo, D. Meer & J. G. Schneider (Hrsg.), *Kommunikation und Öffentlichkeit. Sprachwissenschaftliche Potentiale zwischen Empirie und Norm* (S. 149-175). Berlin: De Gruyter.

Fix, U. (1995). Textmusterwissen und Kenntnis von Kommunikationsmaximen. Voraussetzung, Gegenstand und Ziel einer kommunikationsbezogenen Sprachberatung. In B. Biere, B. & R. Hoberg (Hrsg.), *Bewertungskriterien in der Sprachberatung* (S. 62-73). Tübingen: Narr.

Funke, R. (2018). Working on grammar at school: Empirical research from German-speaking regions. *L1 Educational Studies in Language and Literature*, 18, 1-39.

Gese, H. (2022). Norm und Variation beim Tempusgebrauch. *Praxis Deutschunterricht*, 4, 43-49.

Gloy, K. (1980). Sprachnorm. In H. P. Althaus, H. Henne & H. E. Wiegand (Hrsg.), *Lexikon der Germanistischen Linguistik* (S. 363-368). Berlin: De Gruyter.

Gloy, K. (2004). Sprachnormierung und Sprachkritik in ihrer gesellschaftlichen Verflechtung. In W. Besch, A. Betten, O. Reichmann, O. & S. Sonderegger (Hrsg.), *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung* (S. 396-406). Berlin: De Gruyter.

Häcker, R. (2009). Wie viel? Wozu? Warum? Grammatik in der Schule. In M. Konopka & B. Strecker (Hrsg.), *Deutsche Grammatik – Regeln, Normen, Sprachgebrauch* (S. 309-332). Berlin: De Gruyter.

Hartung, W. (1977). Zum Inhalt des Normbegriffs in der Linguistik. In W. Hartung (Hrsg.), *Normen in der sprachlichen Kommunikation* (S. 9-69). Berlin: Akademie-Verlag.

Hennig, M. (2012). Was ist ein Grammatikfehler? In S. Günthner, W. Imo, D. Meer & J. G. Schneider (Hrsg.), *Kommunikation und Öffentlichkeit. Sprachwissenschaftliche Potenziale zwischen Empirie und Norm* (S. 125-148). Berlin: De Gruyter.

KMK (2004). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss*.
https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-BS-Deutsch-MS.pdf
[15.6.2022].

KMK (2012). *Bildungsstandards im Fach Deutsch für die Allgemeine Hochschulreife*.
https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Deutsch-Abi.pdf [15.6.2022].

Kilian, J. (2018). Didaktische Sprachkritik als Öffentlichkeitsarbeit. Zur Bewertung sprachlicher Leistungen in der Praxis des Deutschunterrichts. *Muttersprache*, 128(1), 52-65.

Kilian, J. (2020). Didaktische Sprachkritik und Deutschunterricht. In: Th. Niehr, J. Kilian & J. Schiewe (Hrsg.), *Handbuch Sprachkritik* (S. 413-421). Berlin: J. B. Metzler.

Kilian, J. (2021). Didaktische Sprachkritik im Spannungsfeld zwischen Norm und Variation. In T. Hoffmeister, M. Hundt & S. Naths (Hrsg.), *Laien, Wissen, Sprache* (S. 491-516.). Berlin: De Gruyter.

Kilian, J., Niehr, Th. & Schiewe, J. (2016). *Sprachkritik. Ansätze und Methoden der kritischen Sprachbetrachtung* (2. Aufl.). Berlin: De Gruyter.

Langlotz, M. (2018). Der zweite Nebensatz lässt sich besser als einzelner Hauptsatz lesen - Interpunktionskorrekturen in Schülertexten. In K. George, M. Langlotz, U. Milevski, K. Siedschlag (Hrsg.), *Interpunktion im Spannungsfeld zwischen Norm und stilistischer Freiheit* (S. 125-156). Frankfurt a. M.: Peter Lang.

Müller, C. (2009). Wat ihr normal nennt... Ein didaktisches Dilemma und die Chance, aus der Not eine Tugend zu machen. In M. Hennig & C. Müller (Hrsg.), *Wie normal ist die Norm? Sprachliche Normen im Spannungsfeld von Sprachwissenschaft, Sprachöffentlichkeit und Sprachdidaktik* (S. 60-83). Kassel: Kassel University Press.

Osterroth, A. (2015). *Linguistisch begründete Sprachkritik in der Schule*. Baltmannsweiler: Schneider Verlag Hohengehren.

Ransmayr, J. (2021). Wo der Rotstift Pause hätte: Zur Überkorrektur in orthographischen Spielräumen. *Mitteilungen des Deutschen Germanistenverbandes*, 68(1), 94-108.

Wöllstein, A. & Dudenredaktion (2016): Duden. Die Grammatik. Unentbehrlich für richtiges Deutsch (9. Aufl.). Berlin: Dudenverlag.

Angaben zur Autorin:

- Dr. Helga Gese, Deutsches Seminar der Universität Tübingen, Linguistische Abteilung - Sprachdidaktik, helga.gese@uni-tuebingen.de

Alexander Schneider

Lernen und Bewerten im Kunstunterricht der Sekundarstufe I

Epirrhema

Müset im Naturbetrachten
Immer eins wie alles achten;
Nichts ist drinnen, nichts ist draußen:
Denn was innen das ist außen.
So ergreift ohne Säumnis
Heilig öffentlich Geheimnis.

Freuet euch des wahren Scheins,
Euch des ernstesten Spieles:
Kein Lebendiges ist ein Eins,
Immer ist's ein Vieles.

(Johann Wolfgang von Goethe, ca. 1819/20)

1. Einleitung

Als ein primär praktisches Fach wird der Kunstunterricht häufig als ein Ausgleich zur Kopfarbeit sog. ‚harter Fächer‘ gesehen. Aber nicht nur diese vermeintlich kompensatorische Funktion, sondern auch Reformen wie das achtjährige Gymnasium mit seiner verkürzten Stundentafel und ein steigender Mangel an Fachlehrer*innen, haben zur Marginalisierung des Unterrichtsfaches in der Sekundarstufe I beigetragen (BDK, 2015). Doch ungeachtet dieser strukturellen und personellen Herausforderungen soll im Kunstunterricht nach wie vor gelernt und bewertet werden. Damit stellt sich die Frage, wie sich unter den skizzierten Bedingungen sowohl Stärken als auch Entwicklungsnotwendigkeiten und Fehler in Bewertungssituationen sinnvoll berücksichtigen lassen. Dieser Beitrag entwickelt eine Annäherung an einen gelingenden Umgang damit, indem er zunächst die Lehr-Lernsituation des Faches beleuchtet und davon ausgehend den Akt des Bewertens als ein Oszillieren zwischen mehreren Vergleichshorizonten beschreibt. Die diesem Band zugrundeliegende Frage von Normativität und Relativität wird hierbei im Verhältnis von ‚Objektivität‘, ‚Subjektivität‘ und ‚Intersubjektivität‘ reflektiert.

2. Lernen und Bewerten zwischen Freiraum und Anleitung

2.1. Grundsätzliches zur Lehr-Lern-Situation

Lehren und Lernen konstituiert sich im Kunstunterricht auf der Basis von Produktion und Rezeption. Beide Bereiche sind eng miteinander verwoben und aufeinander bezogen. Gleichwohl fokussiert die Leistungsbewertung in der Sekundarstufe I hauptsächlich das bildnerisch-praktische Arbeiten. Entsprechende Aufgabenstellungen können dabei unterschiedliche inhaltliche und formale Schwerpunkte aufweisen: Ein zur Bewertung stehendes Gestaltungsergebnis kann beispielsweise darauf ausgerichtet sein, Wirklichkeit zeichnerisch abzubilden, zu deuten oder auch zu bewältigen; das Gestaltungsergebnis kann aber auch plastiziert oder skulpiert worden sein und eine Vision sichtbar machen oder einen praktischen Nutzen erfüllen usw. Gestaltungsergebnisse entstehen und stehen somit im Rahmen der angedeuteten breiten Differenzierung.

In der Unterrichtspraxis zeigt sich allerdings zunehmend, dass „u. a. wegen fehlender Zeitbudgets und aus Disziplinierungsgründen“ kaum zwischen Lern- und Leistungsaufgaben unterschieden wird (Wagner, 2010, S. 8). Erschwerend kommt hinzu, so hat es Andreas Schoppe treffend herausgearbeitet, dass ein an konstruktivistischen Lerntheorien orientierter Kunstunterricht – und man könnte reformpädagogische Strömungen, die dem schöpferischen Genius des Kindes frönen, ergänzend hinzunehmen – ein vermeintlich freies, individuelles und kreatives Lernumfeld vorstellt. In diesem Umfeld seien dann Schüler*innen in der Lage, ihr Wissen und Können selbst zu generieren (vgl. Schoppe, 2016, S. 7). Elementare didaktische Kategorien, wie Lernen und Üben, erscheinen unter diesen lerntheoretischen und erziehungswissenschaftlichen Vorzeichen als restriktiv und damit vernachlässigbar. Tatsächlich läuft jedoch ein vor diesem Hintergrund eingeforderter pädagogischer Freiraum Gefahr, in Beliebigkeit und Willkür zu münden. Entsprechend wird solch ein vermeintlich freiheitlich-schöpferisch aufgefasster und praktizierter Kunstunterricht landläufig mit dem Makel einer rein subjektiven Notengebung in Verbindung gebracht.

Erst in jüngerer Vergangenheit haben sich u. a. Timo Bautz sowie Jochen Krautz und Hubert Sowa für die didaktische Rehabilitierung von ‚Lernen‘ und ‚Üben‘ im Kunstunterricht stark gemacht (Bautz, 2015; Krautz, 2020; Krautz & Sowa, 2012). Dieser Umstand ist im Kontext dieses Beitrags insofern relevant, als Leistungen erst dann (wieder) intersubjektiv nachvollzogen und bewertet werden können, wenn ein curricular strukturiertes Unterrichtskonzept vorliegt, das gestalterisches Wissen und Können systematisch durch Lernen und Üben aufbaut.

2.2. Lernstands-, Übungs- und Leistungsaufgaben im Kunstunterricht

Systematisches Lernen und Üben erfordert eine spiralcurriculare Struktur. Eine solche Struktur ordnet den Lehr- und Lernstoff nicht linear an, sondern in Form einer aufsteigenden Spirale. Entsprechend kehren einzelne Themenfelder und Aufgabenformate im Laufe der Schulzeit auf jeweils höheren Niveaus und in differenzierterer Form wieder. Sowa, der diese aufsteigende Zirkelbewegung für den Kunstunterricht konzeptualisiert hat, unterscheidet drei curriculare Stränge, die einerseits interdependent sind und andererseits punktuelle Fokussierungen zulassen. Es handelt sich hierbei erstens um „[h]andwerkliches Können und Wissen“, zweitens um „[g]estalterisches Wissen und Können“ und drittens um „Können und Wissen im Bereich inhaltlicher Sinndeutung und Sinnggebung“ (Sowa, 2014, S. 493).

Solch ein dreisträngiges Spiralcurriculum liefert nicht nur ein Organisationsschema für unterrichtliches Lehren und Lernen. Es ermöglicht zusätzlich auch intersubjektive Nachvollziehbarkeit von Leistungsfeststellungen, weil es die Kontinuität von Lernprozessen sicherstellt. Eine solche Kontinuität ist auch für den Kunstunterricht essentiell, um überhaupt Stärken und Schwächen von Lernenden diagnostizieren zu können und individuelle Entwicklungsspielräume oder eben auch Entwicklungsnotwendigkeiten aufzuzeigen. Denn die aufsteigende Spiralbewegung bleibt nicht nur auf die Makroebene, also das Curriculum, beschränkt, sondern bezieht auch Meso- und Mikroebenen, wie die Reihen- und Stundenplanung, mit ein. So hat Schoppe (2016) u. a. auf der Basis von Sowas Ausführungen ein entsprechend gestuftes Phasenmodell zur Unterrichts- und Reihenplanung entworfen, in dem sich Lernprogression auch auf der Ebene der Aufgabentypen widerspiegelt. Dieses verläuft – analog zum Spiralcurriculum – aufsteigend: Unten hat er die Diagnoseaufgaben angeordnet, die den Lernstand und die Lernausgangslage prüfen (vgl. ebd. S. 13). Hiermit verknüpft und/oder darauf aufbauend folgen Lernaufgaben, die sowohl handwerkliche Fertigkeiten als auch inhaltliches Wissen vermitteln, bevor es zu einer auf das Gelernte und Geübte bezogenen Leistungsbewertung kommt (vgl. ebd.).

3. Bewerten als komparative Mehrdimensionalität

3.1 Vergleichendes Sehen und Perspektivwechsel

Die diesem Beitrag leitmotivisch vorangestellten Zeilen Goethes liefern nun noch ein erkenntnistheoretisches Fundament, das im Kontext eines spiralcurricularen Verständnisses von Lehren, Lernen und Bewerten relevant sein kann. Im

dritten und vierten Vers von Goethes Gedicht werden wir mit den Worten „Nichts ist drinnen, nichts ist draußen / Denn was innen das ist außen.“ daran erinnert, dass Intro- und Extrospektion nicht separiert voneinander existieren. Im Gegenteil: Die Eigenschaften der Objektwelt manifestieren bzw. spiegeln sich in der Befindlichkeit des Subjekts wider und umgekehrt. Die hierin anklingende Relationalität mündet bei Goethe dann in einer Conclusio: „Kein Lebendiges ist ein Eins, / Immer ist's ein Vieles.“ (V. 9f.) Transzendiert man dieses von Goethe literarisch formulierte erkenntnistheoretische Verständnis auf die Bewertungssituation im Kunstunterricht, kommt man nicht umhin festzustellen, dass es „ein Vieles“ an Faktoren gibt, das den Gestaltungsprozess beeinflusst, begleitet und moderiert. Diese Vielheit an Faktoren bewegt sich hierbei im Verhältnis von subjektiven Sichtweisen, Gefühlen sowie Empfindungen und kulturell geronnenen Objektivationen. In Anbetracht dieser Teil-Ganzes-Relation bedarf es für die Bewertung von bildnerischen Gestaltungsleistungen einer erhöhten Sensibilität. Das bedeutet, dass man das Gestaltungsergebnis auf keine Checkliste mit isolierten Gestaltungskriterien reduzieren kann oder generell in ein Punktesystem mit voneinander getrennten Teilleistungen übertragen sollte (vgl. hierzu auch grundsätzlich Peez, 2020).

Infolgedessen plädiert dieser Beitrag dafür, die Praxis der Leistungsbewertung anschauungsgeleitet im Sinne eines vergleichenden Sehens zu kultivieren und auf ein formalistisches Punktesystem zu verzichten. Durch einen wechselnden Blick zwischen mehreren Gestaltungsergebnissen kann ein Prozess des Gruppierens, Sortierens und Hierarchisierens angestoßen werden. Doch hierbei ist zu bedenken, dass es währenddessen zu Überblendungen kommen kann oder dass sich die Aufmerksamkeit der bzw. des Bewertenden im Laufe des Vergleichens verschiebt. Um im Ergebnis also das vorstehende, in Anlehnung an Goethe ausgebreitete Erkenntnisverständnis nicht doch wieder rein subjektivistisch zu verkürzen, ist es sinnvoll den Prozess des Vergleichens reflexiv zu begleiten. Hierfür können beispielsweise Diagnose- bzw. Lernstandsaufgaben herangezogen werden. Zusätzlich oder alternativ könnte man individuelle wie auch kollektive Reflexionsgespräche oder/und Reflexionstexte der Lernenden in einer Bewertung berücksichtigen. Denn erst eine Perspektivvielfalt ermöglicht es der Lehrperson, auch blinde Flecken beim eigenen vergleichenden Sehen zu lokalisieren. Der nächste Punkt entfaltet und konkretisiert mögliche, für die intersubjektive Verständigung relevante Vergleichshorizonte.

3.2. Plurale Vergleichshorizonte

Dass Leistungsbewertungen komparativ angelegt sein können, ist innerhalb der Kunstpädagogik nicht neu und wurde in der Vergangenheit immer wieder herausgestellt. So lautet es bereits 1969 bei Gunter Otto:

„Eine [...] häufige Praxis geht von einer verhältnismäßig groben, aber in der Regel leicht zu treffenden Qualitätsunterscheidung aus; alles vorliegende Material wird in *drei Gruppen* [Herv. im Org.] – gut, mittel, schlecht – eingeteilt und hernach, wenn erwünscht, innerhalb der Gruppen weiter differenziert. Dieses Verfahren [...] geht oft von Evidenzurteilen aus, führt aber häufig während der Anwendung durch Vergleich und Gruppierungen zu einer allmählichen Herausbildung von anwendbaren Kriterien.“ (Otto, 1969, S. 225f.)

Diese komparative Verfahrensweise ist nicht nur später von Reinhard Pfennig (1974) erneut aufgegriffen worden, ihre Sedimente finden sich auch noch im heutigen Schulalltag, wenngleich sie auch in neueren Einführungen in die Kunstpädagogik durchaus kritisch besprochen werden (vgl. Eid et al., 2002, S. 269). Was aus Sicht dieses Beitrags in dem bei Otto angesprochenen und später bei Pfennig erneut verfolgten komparativen Gruppieren und Ordnen von Schülerarbeiten vernachlässigt wird, ist die oben entfaltete Relationalität.

Relationalität, wie sie ausgehend von Goethes *Epirrhema* vorgestellt wurde, verlangt eine Pluralität von Vergleichen, die bereits in der Unterrichtsplanung grundgelegt sein sollte. Bewertungskriterien sollten sich also nicht – wie es bei Otto anklingt – lediglich am Ende ‚allmählich‘ herausbilden.

Indem auch bereits zu Beginn einer spiralcurricular gedachten Unterrichtsreihe eine kurze Diagnoseaufgabe gestellt wird, um den Lernstand abzufragen, ergibt sich bereits mit Blick auf das später zur Bewertung stehende Gestaltungsergebnis ein lernbiographisch zu berücksichtigendes Vergleichspaar: Lernstand ↔ Lernfortschritt. Unmittelbar hiermit korreliert zusätzlich die Frage, inwieweit im Gestaltungsergebnis die in der Aufgabenstellung formulierten Anforderungen umgesetzt und erfüllt worden sind. Bevor diese Vergleiche jedoch gezogen werden können, empfiehlt es sich, noch während des Arbeitsprozesses Rückmeldebögen einzusetzen – beispielsweise als Peer-Feedback. Die Schüler*innen können sich anhand eines solchen Bogens gegenseitig Rückmeldungen geben, indem sie beispielsweise den Einsatz von Form und Farbe beim Erstellen von Piktogrammen reflektieren oder die Lesbarkeit eines Comics oder einer Karikatur kontrollieren. Zwar haben Feedbackbögen primär eine beratende Funktion und sollen Überarbeitungsmöglichkeiten aufzeigen und anregen (vgl. hierzu auch Seydel, 2006). Nichtsdestoweniger kann auf sie auch im Rahmen der Leistungsbewertung zurückgegriffen werden, weil sie ein Reflexionsfolie für individuell zu berücksichtigende Lernfortschritte bildet. Zusammenfassend ergeben sich die folgenden möglichen Vergleichshorizonte:

- Vergleich aller vorliegenden Gestaltungsergebnisse durch die Lehrkraft

- Vergleich des einzelnen Gestaltungsergebnisses mit den Anforderungen der Aufgabenstellung
- Vergleich von individuellem Lernstand und individuellem Lernfortschritt
- Vergleich von Gestaltungsergebnissen unter den Lernenden (Peer-Feedback)

Die vorstehend konturierten Vergleichspunkte sind selbstverständlich exemplarisch; sie dienen lediglich dazu, die Mehrdimensionalität des Bewertungsvorgangs zu skizzieren. Die Vergleichspunkte können je nach Lernkontext und Aufgabenstellung kleinschrittiger differenziert, modifiziert und/oder ergänzt werden. Dass die Vergleichshorizonte darüber hinaus nicht linear, sondern relational verlaufen, illustriert der hier abgebildete Knoten (vgl. Abb. 1). Da eine Bewertung immer konstatierend angelegt ist, wurde mit dem Knoten eine geschlossene, sich nicht entwickelnde Figur als visuelle Vorstellungshilfe gewählt. Der Knoten weist einen gemeinsamen Mittelpunkt auf, welcher sich auf der Basis von Verschlingungen konstituiert. Die Schlingen können je nach Unterrichtsgegenstand umfangreich oder weniger umfangreich sein; hier wurde analog zu den oben angeführten vier Vergleichshorizonten ein Knoten mit vier Schlingen gewählt. Jede Schlinge repräsentiert einen Vergleichshorizont, wobei jeder Horizont in Wechselbeziehung mit einem je anderen Horizont bzw. mit einer je anderen Schlinge steht. Je nach unterrichtlicher Situation kann natürlich eine Schlinge mal weiter und mal enger gefasst oder mal dicker und mal dünner beschaffen sein. Auch kann je nach Kontext eine Schlinge hinzukommen oder wegfallen.

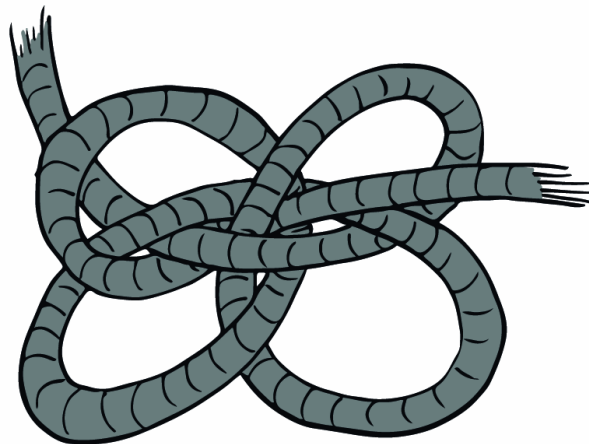


Abb. 1. Relationalität und Pluralität als Knoten (Vorstellungshilfe)

Letztlich ist es nur wichtig, die Teil-Ganzes-Relation nicht aufzulösen. Indem man die jeweils für eine Bewertung herangezogenen Vergleichshorizonte darüber hinaus für die Schüler*innen transparent macht, öffnet man einen Raum für intersubjektive Nachvollziehbarkeit.

4. Literatur

Bund Deutscher Kunstlehrer (BDK) (4/2015). *Gegen die Kürzung und den Abbau des Faches Kunst an Schulen! Stellungnahme des geschäftsführenden Bundesvorstands des BDK, Fachverband für Kunstpädagogik e.V.* https://static.bdk-online.info/bis2020/2015/04/BDK_Gegen-die-K%C3%BCrzung-und-den-Abbau-des-Faches-Kunst-an-Schulen_2015.pdf [25.01.2023].

Bautz, T. (2015). Üben – blinder Fleck oder verlorenes Terrain. *BÖKWE*, 4, 104-107.

Eid, K., Langer, M. & Rupprecht, H. (2002). *Grundlagen des Kunstunterrichts. Eine Einführung in die kunstdidaktische Theorie und Praxis*. (6. durchgesehene Auflage). Stuttgart: UTB.

Krautz, J. & Sowa, H. (Hrsg.). (2012). Lernen – Üben – Können. *Kunst+Unterricht*, 369/370.

Krautz, J. (2020). *Kunstpädagogik. Eine systematische Einführung*. Paderborn: Wilhelm Fink.

Otto, G. (1969). *Kunst als Prozeß im Unterricht* (2. erw. Auflage) Braunschweig: Westermann.

Peez, G. (2020). Einführung. Beurteilen und Bewerten im Kunstunterricht. In G. Peez (Hrsg.), *Beurteilen und Bewerten im Kunstunterricht. Modelle und Unterrichtsbeispiele zur Leistungsmessung und Selbstbewertung* (6. Auflage, S. 10-21). Hannover: Klett/Kallmeyer.

Pfennig, R. (1974). Beurteilung und Bewertung. In R. Pfennig (Hrsg.), *Gegenwart der bildenden Kunst. Erziehung zum bildnerischen Denken* (S. 180-182). Oldenburg: Isensee.

Schoppe, A. (2016). Aufgaben im Kunstunterricht. Motoren für Lernprozesse – Werkzeuge der Diagnose – Schlüssel zum Kompetenzerwerb. *Kunst+Unterricht*, 399/400, 6-14.

Seydel, F. (2006). Zwischendurch Rückmeldung geben – aber wie? Beratung im Kunstunterricht. *KUNST 5 bis 10*, 2, 46-47.

Sowa, H. (2015). Grundlagen der Kunstpädagogik – anthropologisch und hermeneutisch. In A. Glas, U. Heinen, J. Krautz, M. Miller, H. Sowa & B. Uhlig (Hrsg.), *Kunstunterricht verstehen. Schritte zu einer systematischen Theorie und Didaktik der Kunstpädagogik* (S. 481-517). München: Kopaed.

Wagner, E. (2010). Aufgaben, Bildungsstandards, Kompetenzen. Versuch einer Klärung der Begriffsvielfalt. *Kunst+Unterricht*, 341, 4-11.

Angaben zum Autor:

- Dr. Alexander Schneider, Studienrat an einem Gymnasium und Mitglied des IMAGO-Forschungsverbundes für Kunstpädagogik, alexander.schneider@gmx.de

Friedrich Schweitzer und Fahimah Ulfat

Relativität und Normativität von Leistung und ihrer Beurteilung im Religionsunterricht

1. Einleitung

Der Religionsunterricht unterscheidet sich von den meisten, wenn nicht allen anderen Schulfächern dadurch, dass die Leistungsthematik mit zu seinen Kernthemen gehört, insbesondere im Blick auf die Würde des Menschen sowie die Bedeutung von Gnade als Gegenpol aller Leistungsanforderungen.

2. Leistung und Leistungsbewertung als Thema im Religionsunterricht

Leistung und Leistungsbewertung sind moderne Begriffe, die in den religiösen Traditionen als solche nicht vorkommen. Vor allem im Neuen Testament findet sich jedoch der Begriff der „Werke“, als Übersetzung des griechischen *erga* (ἔργα), was in bestimmten Fällen durchaus mit Leistung übersetzt werden kann, zumal es hier darum geht, nach welchen Maßstäben der Mensch beurteilt werden kann. Die für den Religionsunterricht zentralen Begriffe sind jedoch Gnade und Barmherzigkeit, die beide auf die Relativität von Leistungsmaßstäben und Leistungsbewertungen verweisen. Das gilt für den christlichen ebenso wie für den jüdischen und islamischen Religionsunterricht. Die stärkste Zuspitzung hat das jedes Leistungsprinzip relativierende Prinzip der Gnade allerdings in der evangelischen Tradition gefunden. Dass der Mensch bei Gott „allein aus Gnade“ Anerkennung finden kann, gehört zu den Spitzensätzen der reformatorischen Theologie. Auch in koranischer Perspektive bedarf Gnade keiner Leistung, sie ist unverdient. Meist wird die Gnade Gottes mit seiner Barmherzigkeit zusammengedacht. Beide Begriffe gehören zur göttlichen Güte. Gnade wird auch mit Dankbarkeit in Verbindung gebracht. Sie wird zwar ohne Leistung erlangt, soll jedoch mit Dankbarkeit beantwortet werden.

Im Religionsunterricht wird die Leistungsthematik in eine transzendenzbezogene Perspektive gerückt und auf diese Weise mit der menschlichen Suche nach Sinn und Anerkennung für das eigene Leben verbunden. Dabei wird im Religionsunterricht deutlich gemacht, dass die Würde des Menschen von Gott verliehen und dem Menschen also ganz unabhängig von allen seinen Verdiensten geschenkt wird, weshalb sie nicht leistungsabhängig sein kann und auch nicht nach anderen Kriterien, wie Religion, Geschlecht oder Hautfarbe skalierbar. Auch Menschen, die jedenfalls im üblichen Sinne keine Leistungen erbringen können, die als

gesellschaftlich nützlich anerkannt sind, verlieren in dieser Sicht keinesfalls ihre Würde. Das Leistungsprinzip kann folglich nur eine begrenzte und insofern relative Geltung beanspruchen.

Damit verbunden ist auch ein besonderes Verständnis von Gerechtigkeit, das über die in der philosophischen und pädagogischen Tradition seit Aristoteles gebräuchliche Unterscheidung zwischen Verteilungsgerechtigkeit (gemäß Leistung), ausgleichender Gerechtigkeit (orientiert an individueller menschlicher Bedürftigkeit) sowie Strafgerechtigkeit (in der Entsprechung zwischen Vergehen und Strafe) in einem grundsätzlichen Sinne noch einmal hinausführt. In theologischer Sicht zielt Gerechtigkeit auf einen neuen Zustand eines erhofften umfassenden Friedens sowie insgesamt gelingender Beziehungen zwischen den Menschen sowie zwischen den Menschen und Gott. An dieser umgreifenden Vision sind alle menschlichen Auffassungen von Gerechtigkeit und damit auch von Leistung zu messen, wodurch sie erneut in einem grundsätzlichen Sinne relativiert werden. Dies kommt auch beispielsweise in zu fast geflügelten Worten gewordenen Formulierungen gut zum Ausdruck - beispielsweise: Der Mensch sieht, was vor Augen ist, Gott aber sieht das Herz an (1. Sam 16,7).

Allerdings soll nicht außer Acht gelassen werden, dass in religiösen Traditionen auch der Aspekt der Anerkennung, Würdigung, Entlohnung, Sanktionierung, Vergeltung oder Entschädigung von menschlichen Handlungen durch die göttliche Gerechtigkeit sowohl im Diesseits als auch im Jenseits eine Rolle spielt. Ein zentrales Thema im Koran ist die Anerkennung, Würdigung und Entlohnung für rechtschaffenes Verhalten und Taten der menschlichen Barmherzigkeit. Es wird aber auch die Bedeutung der Vergeltung von begangenen Unrecht deutlich zum Ausdruck gebracht. Als Maßstab für die Rechtleitung des Menschen dient dabei sowohl die Gerechtigkeit Gottes als auch seine Barmherzigkeit, Güte und Gnade. Besonders hervorzuheben ist, dass in Sure 6 Vers 12 Gott sich selbst die spezifische Eigenschaft der Barmherzigkeit (رحمة *rahma*) ‚auferlegt‘ hat bzw. sich dazu verpflichtet. Barmherzigkeit ist dementsprechend grundlegend für sein Handeln. Das kann aber in einem Spannungsverhältnis zu seinem Handlungsprinzip der Gerechtigkeit stehen, was zu kontroversen Positionen innerhalb der muslimischen Theologiegeschichte geführt hat (dazu bspw. Hildebrandt, 2007; Ghandour, 2014). Im vorliegenden Zusammenhang ist besonders wichtig, dass die menschliche Beurteilung menschlicher Leistung nie abschließend oder absolut sein kann. Menschliches Urteil ist stets relativ, nur das Urteil Gottes ist absolut.

Auch im christlichen Bereich finden sich Vorstellungen von Gericht, Beurteilung und Bestrafung oder Belohnung, wofür besonders das Matthäusevangelium mit seinen Vorstellungen von einem endzeitlichen Weltgericht (Mt 25, 31ff.) prägend geworden ist. Schon in der Bibel und vor allem in der Theologiegeschichte finden

sich dazu jedoch sehr unterschiedliche Auslegungen, die zum Teil stärker auf Ver-söhnung oder sogar eine Allversöhnung zielen. Aus heutiger Sicht lässt sich ein verabsolutiertes gesellschaftliches oder schulisches Leistungsdenken, das sich seiner Grenzen und damit seiner Relativität nicht mehr bewusst ist, theologisch jedenfalls nicht legitimieren. Auf die Grenzen menschlicher Urteilsfähigkeit in biblischer Sicht wurde oben bereits verwiesen.

Wie auch immer man die Beziehung zwischen menschlichem und göttlichem Urteil theologisch bestimmt, es kann in der gegenwärtigen Religionspädagogik als Konsens angesehen werden, dass der persönliche Glaube der Schülerinnen und Schüler sich der Bewertung im Unterricht entzieht. Diese Grenze für alle Bewertungsversuche gilt ebenso aus theologischen wie auch pädagogischen Gründen. Vielfach wird auf diese Grenze entsprechend in den Einleitungen zu Bildungsplänen ausdrücklich hingewiesen. Zugleich gibt es aber auch im Religionsunterricht Bereiche wie den Erwerb von Wissen, von Fähigkeiten des Deutens und Verstehens sowie der Urteilsbildung, die in ähnlicher Weise wie in anderen Fächern bewertet werden können und, der Logik der Schule folgend, dann auch bewertet werden sollten (Schweitzer, 2016).

Die Begriffe von Relativität und Normativität spielen in der Religionsdidaktik auch dort, wo es der Sache nach um die damit bezeichneten Zusammenhänge geht, noch keine terminologische Rolle. Die Relativität von Leistungsbewertungen wird zwar durchaus auch mit diesem Begriff bezeichnet, aber die Spannung zwischen Relativität und Normativität wird beispielsweise noch nicht dazu eingesetzt, einen kategorialen religionsdidaktischen Zugang zur Leistungsthematik zu gewinnen. Es könnte lohnend sein, die damit verbundenen Fragen in Zukunft weiter zu vertiefen.

Die Berufung auf das Prinzip der Gnade und einer leistungsunabhängigen Anerkennung der Menschenwürde bedeutet theologisch nicht, dass jeder normative Anspruch von Leistungsanforderungen abgelehnt würde. Sofern die grundsätzliche Relativität von Leistungsbewertungen und deren Grenzen bewusst bleiben, können sie auch theologisch in ihrer – möglichen – gesellschaftlichen und auch pädagogischen Nützlichkeit bejaht werden. Auch bei dieser Bejahung muss allerdings die grundlegende Relativität und Begrenztheit menschlicher Leistungsbeurteilungen und -ansprüche bewusst bleiben. Dies wird gerade bei der schulischen Leistungsbewertung sichtbar.

Trotz aller Anerkennung des Leistungsprinzips in seiner gesellschaftlichen Funktion wurde und wird in Theologie und Religionspädagogik immer wieder eine deutliche Spannung zwischen der Thematik des Religionsunterrichts auf der einen und besonders der Vergabe versetzungserheblicher Noten auf der anderen Seite empfunden. In der Folge wurde lang darüber gestritten, ob sich der Religionsunterricht

solchen Formen der Leistungsbewertung nicht grundsätzlich verschließen sollte (Nipkow, 1979). Durchgesetzt hat sich am Ende jedoch die Auffassung, dass der Religionsunterricht sich nicht einfach aus der Reihe und Gemeinschaft der Schulfächer, die allesamt von der Problematik der schulischen Leistungsbewertung und der Vergabe von Ziffernnoten betroffen sind, verabschieden könne. Deshalb wurde und wird bejaht, dass der Religionsunterricht beispielsweise auch bei Prüfungen im Abitur beteiligt ist.

Führende Vertreter in der entsprechenden Diskussion, die vor allem in den 1970er Jahren geführt wurde (neben Nipkow, 1979 auch etwa Jendorff, 1979), verbanden eine entsprechende Empfehlung für die Leistungsbewertung im Religionsunterricht allerdings mit der Forderung, dass sich Religionsunterricht und Religionsdidaktik fortan verpflichtet sehen müssten, sich um pädagogisch sinnvolle sowie für Kinder und Jugendliche förderliche Formen der Leistungsbewertung zu bemühen.

3. Ansätze zum Umgang mit Leistungsbewertung und Notengebung im Religionsunterricht

Trotz der soeben genannten Forderung kann nur auf relativ wenige größere Veröffentlichungen zur Leistungsthematik im Religionsunterricht verwiesen werden (Überblick: Zimmermann, 2020). Die Hoffnung, dass sich die Religionspädagogik gerade in dieser Hinsicht besonders engagieren und profilieren werde, hat sich offenbar nicht erfüllt. Aus der Praxis ist allenfalls mitunter zu hören, dass die Religionsnoten insofern eher ‚gnädig‘ vergeben würden, als sie in aller Regel recht gut ausfallen. Für Abschlussprüfungen wie das Abitur dürfte das allerdings kaum zutreffen (zu Abituraufgaben im Bereich des Religionsunterrichts Muth, 2021).

Ein Teil der Religionslehrkräfte scheint, wie allerdings nur informell aus Erfahrungsberichten bekannt ist, das Spektrum der vergebenen Noten bewusst einzuschränken. Die Kritik an der Vergabe von Ziffernnoten nimmt dann wohl gelegentlich sogar die Gestalt einer Karikatur des Notensystems an, wenn alle dieselbe Note erhalten – ganz so wie im berühmten neutestamentlichen Gleichnis von den Arbeitern im Weinberg (Mt 20), die trotz ungleichen Arbeitseinsatzes und entsprechender unterschiedlicher Leistung am Ende denselben Lohn empfangen.

Die bislang sehr wenigen empirischen Erkenntnisse in der religionspädagogischen Forschung zum Thema Leistungsbeurteilung zeigen, dass Lehrkräfte Leistungsbeurteilung im Religionsunterricht als Dilemma erleben, da der Religionsunterricht zum einen ein Fach wie jedes andere sein, zum anderen aber seine besondere Charakteristik beibehalten soll, die sich im Beziehungsgeschehen zur und in der Klasse ausdrückt (Pirker & Juen, 2018). Mit Blick auf Relativität und Normativität

von Leistungsbeurteilungen verstehen sie Leistungsbeurteilung nicht im Sinne von Selektion und Allokation, sondern als „kritischen Gegenpol zu einem auf ‚Effizienzsteigerung‘ ausgerichtetem Bildungsverständnis“ (ebd., S. 177). Es kann davon ausgegangen werden, dass auch Schüler*innen dieses Dilemma wahrnehmen, auch wenn es dazu noch keine Erhebungen gibt.

Eine weitere Variante der Debatte besteht in der Forderung nach alternativen Formen der Leistungsbewertung, die dann von der Religionspädagogik her nicht anders als in anderen Fachdidaktiken diskutiert wird. Im Zusammenhang beispielsweise der Notenhebung im Religionsunterricht kann das Thema der schulischen Leistungsbewertung aber auch ausdrücklich aufgenommen und zur Diskussion gestellt werden, grundsätzlich aufgrund der im Religionsunterricht vertretenen Sicht von Menschenwürde, aber auch aktuell, wenn beispielsweise Schülerinnen und Schüler auf die Spannung zwischen dieser Sicht und der Notengebung speziell im Religionsunterricht verweisen. Ist ‚Gnade‘ im Religionsunterricht lediglich ein abstraktes Thema oder hat dieses Thema auch Folgen für die Praxis von Religionsunterricht, eben bis hin zur Vergabe von Noten?

Auch wenn die Rückmeldefunktion der Leistungsbewertung auf der pädagogischen Ebene als sinnvolles Instrument selbst im Religionsunterricht angesehen werden kann, wird die Anreiz- und Selektionsfunktion auf individueller Ebene problematisiert. Daher wird für eine pädagogische Diagnostik plädiert, die den individuellen Entwicklungsverlauf im Fach in den Vordergrund stellt (Zimmermann, 2008; 2020).

Große Übereinstimmung besteht in der Religionspädagogik darüber, dass in diesem Unterricht alle Schülerinnen und Schüler gefördert werden sollen. Die dafür erforderliche, speziell fachbezogene Diagnostik, die eine Identifikation von Lernproblemen erlaubt und damit einen Beitrag zu deren Überwindung leisten kann, wird in der Religionsdidaktik aber erst in neuerer Zeit stärker zum Thema. Entsprechend wenig entwickelt sind daher auch die Instrumente, die bei einer solchen Diagnostik eingesetzt werden können (Reis & Schwarzkopf, 2015).

4. Ausblick

Leistung und ihre Beurteilung tauchen also, wenn man so will, im Religionsunterricht in vierfacher Gestalt auf. Zum ersten ist sie ein theologisches Problem und ein inhaltliches Thema des Religionsunterrichts. Zum zweiten aber impliziert dieses theologische Problem eine Spannung, die für den Religionsunterricht eigentümlich ist: Inwieweit kollidiert die theologische Sicht von Gottes Gnade und Gerechtigkeit mit der schulischen Notwendigkeit der Leistungsmessung und

Notenvergabe? Zum dritten ist die Frage der Notenvergabe und ihrer pädagogischen Angemessenheit aber auch eine theoretische und in jüngerer Zeit auch empirische Herausforderung der Religionspädagogik. Schließlich, zum vierten, geht es um die Praxis Leistungsbewertung und deren Ausgestaltung im Religionsunterricht.

5. Literatur

Ghandour, A. (2014). Die Freiheit Gottes und des Menschen. Zur Theorie des kasb in der aš'arītischen Theologie. In M. Karimi & M. Khorchide (Hrsg.), *Jahrbuch für Islamische Theologie und Religionspädagogik. Gott und Schöpfung - Freiheit und Determination* (S. 141-167). Breisgau: Kalam Verlag.

Hildebrandt, T. (2007). *Neo-Mu'tazilismus? Intention und Kontext im modernen arabischen Umgang mit dem rationalistischen Erbe des Islam*. Leiden: Brill.

Jendorff, B. (1979). *Leistungsmessung im Religionsunterricht. Methoden und Beispiele*. München: Kösel.

Muth, K. (2021). *Bewertungskriterien ethischer und religiöser Urteilskompetenz. Eine qualitative Studie über Prüfungsaufgaben und Bewertungsvorgaben im schriftlichen Abitur des Faches Evangelische Religionslehre*. Leipzig: EVA.

Nipkow, K. E. (1979). *Religionsunterricht in der Leistungsschule. Gutachten - Dokumente*. Gütersloh: Gütersloher Verlagshaus Mohn.

Pirker, V. & Juen, M. (2018). *Religion - (k)ein Fach wie jedes andere: Spannungsfelder und Perspektiven in der kompetenzorientierten Leistungsbeurteilung*. Stuttgart: Kohlhammer.

Reis, O. & Schwarzkopf, T. (Hrsg.). (2015). *Diagnose im Religionsunterricht. Konzeptionelle Grundlagen und Praxiserprobungen*. Berlin: LIT.

Schweitzer, F. (2016). Leistungsmessung und Leistungsbewertung, Kompetenzen und Standards: Was ist im Religionsunterricht messbar?, *entwurf*, 1, 4-8.

Zimmermann, M. (2008). Leistungsmessung und Leistungsbewertung des Religionsunterrichts im Schulalltag: Probleme – Analysen – Perspektiven. *Zeitschrift für Pädagogik und Theologie*, 60, 296-309.

Zimmermann, M. (2020). *Leistungsmessung, Leistungsbewertung*. https://doi.org/10.23768/wirelex.Leistungsmessung_Leistungsbewertung.100028 [20.03.2022]

Angaben zum Autor und zur Autorin:

- Prof. Dr. Dr. h.c. Friedrich Schweitzer, Seniorprofessor für Praktische Theologie mit Schwerpunkt Religionspädagogik, Universität Tübingen, Evangelisch-Theologische Fakultät, friedrich.schweitzer@uni-tuebingen.de
- Prof. Dr. Fahimah Ulfat, Professur für Islamische Religionspädagogik, Universität Tübingen, Zentrum für Islamische Theologie, fahimah.ulfat@uni-tuebingen.de

Christian Fischer und Johanna Grad

Wege zu neuen Bewertungssystemen in der Hochschullehre: Gameful Learning und Pädagogische Psychologie

Ein wichtiger Bestandteil im Alltag von Wissenschaftlerinnen und Wissenschaftlern an deutschen Hochschulen ist die Ausbildung von Studierenden in universitären Lehrveranstaltungen. Während traditionelle Lehrformate wie frontal gehaltene Vorlesungen und vortragsbasierte Seminare die Lehre an deutschen Universitäten oftmals dominieren, können neue und innovative Kurs-Designs neue Impulse zur Verbesserung von Lehr-Lern-Prozessen anstoßen. Ein Beispiel hierfür ist Gameful Learning, eine Theorie, welche die Prinzipien des Game-Designs und der Pädagogischen Psychologie vereint (Hayward & Fishman, 2020). Dabei ist die Philosophie des Gameful Learnings als Lehrstrategie in der Selbstbestimmungstheorie und dem Konstruktivismus verankert, was zur aktiven Verbesserung des Engagements auf Seiten der Lernenden, deren intrinsischer Motivation und resultierend ihren Lernerfolgen beitragen kann. Das Ziel von Gameful Learning ist die Erschaffung und Bereitstellung vielseitiger Lernumgebungen, welche neben der Förderung der Autonomie, des sozialen Zugehörigkeitsgefühls und der Kompetenz des Lernenden auch sichtbare Fortschritte innerhalb seiner individuellen, unmittelbaren Entwicklung bereiten soll (Fishman et al., 2022).

Dabei nutzen Kurse im Gameful Learning Design zwar Elemente von spielerischen Anleitungen und spielerischem Lernen. Allerdings erfordert Gameful Learning weitaus mehr als nur eine simple Gamification mit Abzeichen- und Punktevergabe ohne Anpassung des generellen Lehrkonzeptes (Fishman et al., 2022; van Roy & Zaman, 2017). Schulische Bewertungen können aus der Perspektive der Lernenden oft relativ erscheinen, vor allem da es häufig zu schwerwiegenden Unterschieden in den Bewertungen zwischen Lehrkräften kommt (Ingenkamp, 1997). Ebenso kann eine Note selbst bei vollständiger Normierung von traditionellen Bewertungskriterien nicht zu einer umfassenden Darstellung der Kompetenz von Lernenden herangezogen werden (Trautwein et al., 2008). Im Gegensatz dazu orientieren sich Lernumwelten, die nach der Theorie des Gameful Learnings kreiert werden, an den Grundprinzipien der Selbstbestimmungstheorie nach Ryan und Deci (2000). Dieser theoretische Rahmen betont anstelle von traditionellen Bewertungskriterien die Relevanz und Transparenz von Wahlmöglichkeiten zur Erreichung von klar formulierten Lernzielen (Autonomie), sowie die Unterstützung der Beziehungen zwischen Lernenden untereinander sowie mit den Lehrenden (soziale Zugehörigkeit). Ferner sollen Aufgaben sowohl das Kompetenzerleben der

Lernenden fördern als auch Möglichkeiten zur Übung und Überprüfungen von Wissen bieten (Kompetenz).

Das innovative Lernmanagement-System *GradeCraft* (www.gradecraft.com) unterstützt eine solche spielerische Lehr- und Lernmethode. Eine Veranschaulichung eines Kursdesign im Rahmen einer Einführungsveranstaltung für Studierende der Fakultät für Wirtschafts- und Sozialwissenschaften, die den Prinzipien des Gameful Learnings folgen, ist unter www.chrisfi.com/teaching zu finden. Im Folgenden wird das Kursdesign beschrieben und erläutert, wie Prinzipien des Gameful Learnings erfüllt werden.

Autonomie: In ‚traditionellen‘ Kursformaten gibt es häufig nur einen Weg, der die Lernenden zum Erfolg führt. Es ist hierbei gängig, Noten durch summative Evaluationen wie beispielsweise Klausuren und Tests zu bestimmen, wobei die finale Note die Leistung eines Lernenden über verschiedene Verfahren hinweg repräsentiert. Dadurch sind die Wahlmöglichkeiten für Lernende oftmals eingeschränkt und jede Prüfung wird hochrelevant, was leistungsorientiertes Lernen oftmals nicht nur erforderlich, sondern auch erstrebenswert macht. Dagegen abgrenzbar sind Gameful Learning Kurse, bei welchen die Lernenden beispielsweise 10.000 Punkte sammeln müssen, um eine 1,0 als Note zu erreichen, wobei es aber innerhalb des Kurses Möglichkeiten gibt, deutlich mehr Punkte zu sammeln (bspw. 15.000 Punkte). Um Punkte zu sammeln, können Studierende bspw. Reading Reactions zu den wöchentlichen Artikeln in einem Diskussionsforum erstellen und kommentieren, kursinterne Blogbeiträge verfassen, oder Podcasts und Erklärvideos erstellen. Dies sind alles Beispiele für formative Evaluationen. Diese Struktur soll Lernende befähigen, ihre Lernerfahrung in einem Maße zu personalisieren, das mit den individuellen Lernzielen der Person übereinstimmt. Als Alternative zu klassischen summativen Evaluationen können u. a. das Schreiben eines Research Proposals, das Erstellen von Postern oder das Verfassen eigener Manuskripte zählen. Diese Struktur erlaubt es den Lernenden, eine Vielzahl an Wegen zum Lernziel zu nutzen. Jede Aufgabe ermöglicht das Sammeln von Punkten, wobei alle aufgabenspezifischen Punktzahlen über die Dauer des Kurses aufsummiert werden, um die Note des Lernenden zu determinieren. Lernende haben somit wesentlich mehr Transparenz und Möglichkeiten über den eigenen Lernerfolg mitzubestimmen. Dabei sind die Punktezahlen der einzelnen Aufgaben nicht willkürlich festgelegt, sondern basieren auf den pädagogischen Entscheidungen der Lehrperson bezogen auf den Wert einer Aufgabe zur Vertiefung des Verständnisses der Kursthematik. Hierbei können digitale Tools, wie ein Grade Predictor in *Grade-Craft*, Lernende dabei unterstützen, ihren personalisierten Lernweg zu erstellen und zu überwachen sowie für mehr Transparenz im Hinblick auf die Zusammenstellung der finalen Note zu sorgen. Um Punkte für die finale Note zu sammeln,

können Lernende zusätzlich zur Erfüllung von Aufgaben digitale Abzeichen gewinnen, welche wünschenswerte Arbeitshaltungen fördern und belohnen. Der Erhalt eines Abzeichens kann Lernenden erlauben, von einmaligen Möglichkeiten wie bspw. einer Fristverlängerung von Kursabgaben Gebrauch zu machen.

Soziale Zugehörigkeit: Ein weiterer Aspekt, der durch das Gameful Learning gefördert werden kann, ist die soziale Zugehörigkeit innerhalb der Lerngruppe. Insbesondere die Pandemie hat verdeutlicht, dass das soziale Zugehörigkeitsgefühl von Lernenden zu einem Kurs und zu ihrer Universität von fundamentaler Wichtigkeit ist, um gute Leistungen in einem wissenschaftlichen Umfeld zu erzielen (Jaeckel, Scheiter & Göllner, 2021). Gameful Learning Kurse können von sogenannten Open-Door-Policies gekennzeichnet sein. Dies bedeutet, dass private Meetings mit einzelnen Gruppen von Studierenden durch das Semester hinweg gehalten werden, aber auch, dass aktives Lernen innerhalb der Vorlesungen betont wird, indem beispielsweise Break-Out-Räume für mehr Engagement und Mitarbeit erstellt werden. Außerdem sind Gameful Learning Kurse so angelegt, dass sie Gruppenarbeit und Kollaboration fördern können, ohne diese verpflichtend zu machen. So können Lernende Freundschaften innerhalb des Kurses schließen und kooperativ an Projekten über das Semester hinweg arbeiten. Beispielhaft für kollaborative Zusammenarbeit unter Lernenden können die kreativen 20%-Projekte als Veranschaulichung angeführt werden (Xu, 2020). Das gemeinsame Arbeiten in der Gruppe an einem selbstgewählten Thema fördert kreatives Problemlösen, kritisches Denken und interdisziplinäres Arbeiten im Team und wird nicht am Gruppenprozess, sondern an dem Endprodukt, welches die Gruppe präsentiert, bewertet (Vogel & Fischer, 2020).

Kompetenz: Weiterhin sollen Gameful Learning Kurse das Kompetenzerleben der Studierenden fördern. Kurse, die auf den Charakteristiken von Gameful Learning basieren, weisen oftmals Merkmale von einem „low floor“, von „wide walls“ und einer „high ceiling“ auf (Resnick et al., 2009, S. 63). „Low floors“ bezeichnet den einfachen Einstieg in den Kurs für Neuzugänge, mit einer Vielfalt an Aufgabentypen zum Ausprobieren, wobei die Aufgaben an Schwierigkeitsgrad zunehmen, wenn Lernende kontinuierlich an einem Aufgabentyp arbeiten. Da Lernende bei den verschiedenen Aufgabentypen im Anfängerniveau nicht ihre Kursnote aufs Spiel setzen, werden sie dazu ermutigt sich auszuprobieren, Risiken einzugehen und Fehlschläge als Lernerfahrung in Kauf zu nehmen. Insbesondere leistungsorientierte Lernende haben dadurch die Chance, ihre Kompetenzen in verschiedenen Aufgabenbereichen zu testen, ohne negative Konsequenzen im Hinblick auf ihre Note befürchten zu müssen. „[W]ide walls“ (ebd.) bezeichnen die verschiedenen möglichen Formen von aktivem Engagement im Kurs (siehe auch „Autonomie“). Lernende können ihr Wissen und ihre Fähigkeiten erweitern, und dafür ihre

persönlichen Präferenzen in Bezug auf ihren Lernweg nutzen. „[H]igh ceilings“ (ebd.) beschreiben, dass Lernende von komplexen und herausfordernden Lernmöglichkeiten Gebrauch machen können, welche die Erwartungen für das Abschließen des Kurses mit einer 1,0 weitaus überschreiten können. Beispielsweise könnten sich Lernende entscheiden, Kursabgaben iterativ so weit zu verbessern, dass diese für akademische Konferenzen oder Publikationen akzeptiert werden. Alternativ könnten Lernende ihren kontinuierlichen Lernprozess in einem eigenen Blog posten, um mit einer Art Lernportfolio (z. B. Pfeifer & Kriebel, 2007) ihr Wissen auch für mögliche Arbeitgeber*innen zu präsentieren.

Studien belegen die positiven Effekte, die mit Kursen im Gameful Learning Design einhergehen. Fishman und Hayward (2022) zeigten, dass Lernende durchgehend bessere Noten in Gameful Learning Kursen erzielen als Lernende in vergleichbaren traditionellen Kursen. Außerdem sind Lernende in ihren Kursen aktiver, wenn sie ihr Handeln als autonom wahrnehmen (Fishman & Hayward, 2022). Hayward et al. (2021) zeigten, dass bei Berücksichtigung des Vorwissens der Lernenden keine signifikanten Unterschiede in der Lernleistung für traditionell-unterrepräsentierte Studierendengruppen (z. B. nach sozio-ökonomischen Hintergrund, Geschlecht etc.) existieren. Technische Erweiterungen – wie der bereits angesprochene Grade Predictor in *GradeCraft* – begünstigen und unterstützen das Gefühl von Autonomie bei den Lernenden, was wiederum zu mehr aktivem Einsatz auf Seiten der Lernenden führt (Holman et al. 2013). Aguilar et al. (2018) untersuchten, wie Lernende das Lernen und Arbeiten in einem Gameful Learning Kurs wahrnehmen. Dabei geben Lernende an, dass Bewertungssysteme aus ihrer Sicht fairer als die herkömmliche Notengebung sind und sie ein stärkeres Gefühl der Kontrolle über ihre Bewertung haben. Zusätzlich macht das genutzte Bewertungssystem den Erhalt der Note, die man erzielen möchte, einfacher. Dabei ist dies kein Indikator dafür, dass Lernende weniger oder weniger intensiv arbeiten. Im Gegenteil – Lernende arbeiten oftmals mehr in Gameful Learning Kursen und vervollständigen ihre Aufgaben öfter. Kurzum, Gameful Learning Kurse haben Potenzial, die Lehr- und Lernprozesse im Alltag der deutschen Hochschullehre nachhaltig zu unterstützen.

Literatur

Aguilar, S.J., Holman, C. & Fishman, B. J. (2018). Game-Inspired Design: Empirical Evidence in Support of Gameful Learning Environments. *Games and Culture*, 13(1), 44-70.
<https://doi.org/10.1177/1555412015600305> [03.04.2023]

Fishman, B. & Hayward, C. (2022). Gameful Learning: Leveraging the Learning Sciences to Improve the “Game of Learning”- How can gameful learning foster greater student engagement? *Rapid Community Reports*, 2-10.

Fishman, B., Hayward, C. & Niemer, R. (2022). Improve Student Engagement with Gameful Learning. Teaching in the Game-Based Classroom: Practical Strategies for Grades 6-12. In D. Seelow (Hrsg.), *Teaching in the game-based classroom: Practical strategies for grades 6-12*. Oxfordshire: Routledge.

Hayward, C. & Fishman, B. (2020). *Gameful Learning: Designing with Motivation in Mind*. In M. Gresalfi & I. S. Horn (Hrsg.), *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020* (2. Auflage, S. 1007-1014). Nashville, Tennessee: International Society of the Learning Sciences.

Hayward, C., Schulz, K. & Fishman, B. (2021). Who wins, who learns? Exploring gameful pedagogy as a technique to support student differences. In LAK21: *11th International Learning Analytics and Knowledge Conference (LAK21)*, 12.-16. April, 2021, Irvine, CA, USA. ACM, New York, NY, USA. <https://doi.org/10.1145/3448139.3448198> [25.01.2023].

Holman, C., Aguilar, S. J. & Fishman, B. (2013). GradeCraft: What can we learn from a gameinspired learning management system? In D. Suthers, K. Verbert, E. Duval & X. Ochoa (Hrsg.), *LAK '13: proceedings of the Third International Conference on Learning Analytics and Knowledge* (S. 260-264). ACM.

Ingenkamp, K. (1997). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz

Jaekel, A.-K., Scheiter, K. & Göllner, R. (2021). Distance Teaching During the COVID-19 Crisis: Social Connectedness Matters Most for Teaching Quality and Students' Learning. *AERA Open*, 7,1-14. <https://doi.org/10.1177/23328584211052050> [25.01.2023].

Pfeifer, S. & Kriebel, J. (2007). Lernen mit Portfolios: neue Wege des selbstgesteuerten Arbeitens in der Schule. Göttingen: Vandenhoeck & Ruprecht.

Ryan, R. M. & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54-67.

Resnick, M., Maloney, J., Monroy-Hernandez, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B. & Kafai, Y. (2009). Scratch: Programming for all. *Communications of the ACM*, 52(11), 60-67.

Trautwein, U., Lüdtke, O., Becker, M., Neumann, M. & Nagy, G. (2008). Die Sekundarstufe I im Spiegel der empirischen Bildungsforschung: Schulleistungsentwicklung, Kompetenzniveaus und die Aussagekraft von Schulnoten. In E. Schlemmer & H. Gerstberger (Hrsg.), *Ausbildungsfähigkeit im Spannungsfeld zwischen Wissenschaft, Politik und Praxis*. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90839-7_5 [25.01.2023].

van Roy, R. & Zaman, B. (2017). Why Gamification Fails in Education and How to Make It Successful: Introducing Nine Gamification Heuristics Based on Self-Determination Theory. In M. Ma & A. Oikonomou (Hrsg.) *Serious Games and Edutainment Applications*. Cham: Springer. https://doi.org/10.1007/978-3-319-51645-5_22 [04.03.2023]

Vogel, F. & Fischer, F. (2020). Computerunterstütztes kollaboratives Lernen. In H. Niegemann & A. Weinberger (Hrsg.), *Handbuch Bildungstechnologie*. Berlin & Heidelberg: Springer. https://doi.org/10.1007/978-3-662-54368-9_3 [25.01.2023].

Xu, T. (2020). *Side Project Programs Can Have Major Benefits for Employers*. Retrieved from: <https://builtin.com/software-engineering-perspectives/20-percent-time> [25.01.2023].

Angaben zum Autor und zur Autorin

- Prof. Dr. Christian Fischer, Tenure-Track-Professor für Educational Effectiveness, Universität Tübingen, Hector-Institut für Empirische Bildungsforschung, christian.fischer@uni-tuebingen.de
- Johanna Grad, Studentische Hilfskraft, Universität Tübingen, Hector-Institut für Empirische Bildungsforschung, johanna.grad@student.uni-tuebingen.de

Jessica Heesen

Informationen aus dem Internet beurteilen

Wahrheit und Quellenbeurteilung in Sozialen Medien und digitalen Diensten aus Perspektive der Medienethik

1. Einleitung

Der vorliegende Beitrag fragt nach der Bedeutung der Beurteilung von Informationen aus dem Internet für die schulische Bildung. Dabei werden mit der Frage nach der Wahrheit von Information und der Glaubwürdigkeit von Informationsquellen wichtige Bezugspunkte für die Beurteilung genannt. Die Urteilsfähigkeit über Informationen im Internet lässt sich über diese Bezugspunkte entfalten, kommt aber nie ohne eine kritische Reflexion der Kontexte von Information aus.

Im Folgenden werden einerseits Analysen zur Bewertung von Information aus dem Normverständnis der Medienethik entwickelt und andererseits konkrete Wege zur Vermittlung der kritischen Reflexion von Online-Information für die schulische Praxis vorgestellt.

Aus Perspektive der Medienethik steht im Zusammenhang von schulischer Bildung und schulischem Alltag insbesondere die Befähigung zur Medienmündigkeit im Zentrum. Vor diesem Hintergrund geht es um die Schaffung von gesicherten und vertrauenswürdigen Lernvoraussetzungen mit digitalen Medien trotz der mit ihnen verbundenen Unsicherheiten. Die Mediennutzung von Kindern und Jugendlichen wird in ihrem Alltag ganz maßgeblich durch Soziale Medien und das Internet bestimmt (KIM-Studie, 2020; JIM-Studie, 2021). Gleichzeitig verweben sich zunehmend Lehr- und Lernmethoden mit Informationsangeboten unterschiedlichster Online-Quellen und -Dienste. Insofern erfordert ein aufgeklärter Umgang mit digitalen Informationsangeboten eine schulische Beurteilung, für die nicht die Betrachtung der Leistungen der einzelnen Schüler*innen im Vordergrund steht, sondern die Tauglichkeit der pluralen Informationsangebote einer digitalen Gesellschaft für den schulischen Bildungsauftrag. Schule als Institution und schulische Bildung bleiben von Des- und Falschinformationen nicht unberührt und müssen Fragen von vertrauenswürdiger Kommunikation für ein gutes und mündiges Aufwachsen in einer digitalen Gesellschaft in die jeweiligen Bildungspläne integrieren.

Bei der Beurteilung von Informationen geht es zumeist um ihre Relevanz und um ihre Richtigkeit bzw. Wahrheit. Die Wahrheit von Informationen ist eng verbunden mit der Beurteilung ihrer Entsprechung zur Realität. Eine wahre Information ist in

der Regel eine solche, die einer geteilten Realitätsvorstellung entspricht (zur Korrespondenztheorie der Wahrheit vgl. Grundmann, 2018). Diese Einschätzungen von Realität können in Abhängigkeit von Kulturen, Weltregionen, Geschlecht oder individueller Lebenssituation unterschiedlich sein. Eine entscheidende Rolle für die Wahrnehmung von Realität und dem, ‚was gilt‘, spielt die Medienkommunikation. Medien legen sich gleichsam wie ein Filter über die Sichtweisen individueller und kollektiver Lebenswelten. Positionen aus dem Konstruktivismus (Pörksen, 2014) würden sogar sagen, dass Medien die Realitätswahrnehmung nicht nur selektieren, sondern dass die Weisen, wie wir durch Medien kommunizieren, dafür bestimmend sind, wie unsere gesellschaftliche Realität zustande kommt. Medien wirken insofern daran mit, wie gesellschaftliche Realitäten erfahren, aber auch, wie sie hergestellt werden.

Neben diesem grundlegenden erkenntnistheoretischen Zugang muss für eine kritische Spiegelung des Wahrheitsbegriff die Medienkommunikation als solche betrachtet werden. Bei diesem Zugang scheidet sich die Frage nach richtiger und wahrer Information maßgeblich an der Autorität der Quelle. Die Autorität der Quelle begründet sich aus einer Vielzahl von Annahmen, die von Plausibilitätsannahmen für die Argumentation in Medienbeiträgen über stilistische Fragen bis hin zu Reputationserwartungen in Hinsicht auf Medienhäuser oder einzelne Journalist*innen reichen. Die Autorität der Quelle ist an das Vertrauen daran gekoppelt, dass Aussagen über die Realität richtig wiedergegeben werden und die entsprechenden Personen oder Institutionen zuverlässig sind.

Für beide dieser zentralen Methoden zur Beurteilung von Information, also die Abstimmung von Realitätsvorstellungen und die Prüfung der Quellen, liegen in der digitalen Kommunikation viele Hindernisse vor. Probleme in der Bewertung von Information liegen allgemein häufig in dem intransparenten Einfluss von Technik auf die Auswahl und Herstellung von Information, dem Meinungsstreit in Sozialen Medien, der schieren Masse von Informationen und einer mangelnden Vertrauenswürdigkeit von Quellen bis hin zu Manipulation und Falschinformation begründet.

Ein einflussreicher und umstrittener Akteur wie der ehemalige US-Präsident Trump z. B. kann für die Online-Kommunikation sein eigenes Soziales Netzwerk gründen. So geschehen mit dem Dienst *Truth Social* Anfang 2022. Auf diese Weise kann Trump, der vorher wegen Falschinformationen und Aufrufen zu Gewalt vom Kurznachrichtendienst Twitter gesperrt wurde, seine eigenen Ansichten über das äußern, was als wahr zu gelten hat. Über den neuen Kommunikationsdienst kann er eine hinreichend große Menge von Menschen erreichen, die dazu bereit ist, seine Interpretationen als wahr anzuerkennen. Diese Menge kann so zahlreich werden, dass sie als kritische Masse eine Schwelle überwindet, um die öffentliche Meinungsbildung zu beeinflussen.

Ein weiteres Beispiel betrifft die sogenannte Ein-Prozent-Regel: Danach ist nur ein sehr kleiner Anteil der Nutzer*innen des Internets aktiv an der Herstellung von Inhalten beteiligt (in der Forschung liegen die Angaben dazu zwischen 1 und 10%). Trotzdem dominiert diese kleine Gruppe die Sozialen Medien, die als Stimme der ‚Allgemeinheit‘ gelten. Die daraus resultierenden Verzerrungseffekte sind auch bekannt für die herkömmlichen Massenmedien. In Hörfunk und Fernsehen ist es aber transparent und selbstverständlich, dass relativ wenige (professionell geschulte) Medienschaffende aktiv sind. Die Sozialen Medien sind vorgeblich partizipativer und haben trotzdem ganz eigene, häufig technisch gesteuerte Regeln für die Kuratierung der Inhalte („Content Moderation“). Dies führt dazu, dass häufig ungeprüfte Informationen einzelner eine hohe Reichweite erlangen und sich unter Gleichgesinnten verstärken (vgl. ‚Echokammern‘, Vaccari, 2013).

2. Kritische Reflexion von Online-Information in der schulischen Praxis

Für den schulischen Zusammenhang gibt es insbesondere zwei Perspektiven, aus denen ein mündiger und reflektierter Umgang mit Informationen im Netz erfolgen sollte: Die allgemeine demokratische Perspektive (1) und die Perspektive der Nutzung von Informationen aus dem Netz unmittelbar für die schulische Bildung (2).

(1) Schulische Bildung verfolgt das Ziel, die Schüler*innen dazu zu befähigen, als mündige Bürger*innen am demokratischen Gemeinwesen und seinen Partizipationsprozessen teilzuhaben. Öffentliche Information, Austausch, Meinungsstreit und Verständigung sind die Herzkammer der liberalen Demokratien. Umso wichtiger ist ein grundlegendes Verständnis von den Funktionsweisen der öffentlichen Kommunikation in Abhängigkeit von den jeweiligen Medientechniken und ihren spezifischen Logiken. Wie funktioniert die Kuratierung der Inhalte auf Social-Media-Plattformen? Was bedeutet Plattformökonomie? Warum gibt es einen öffentlich-rechtlichen Rundfunk? Die Beantwortung solcher grundlegender Fragen sollte Gegenstand der schulischen Bildung in ihren jeweiligen Kontexten sein, damit deutlich wird, dass Medien nie einfach ‚da‘ sind oder ‚neutral‘ vermitteln, sondern durch verschiedene Akteur*innen mit unterschiedlichen Interessen eingesetzt und dementsprechend auch für bestimmte Zwecke gestaltet werden.

(2) Für die Bewertung von Online-Informationen für die Herstellung von fachbezogenem Wissen in der Schule gilt es sich grundsätzlich kritisch mit den Produktionsbedingungen von Informationen im Internet auseinanderzusetzen. Hier ist es wichtig, ein kritisches Bewusstsein für die gesamte ‚Lieferkette‘ für Informationen in den Blick zu nehmen. Dazu gehören begriffliche Unterscheidungen zwischen

Daten, Fakten, Meinungen, Behauptungen, Informationen, Wissen usw. und die Reflexion der unterschiedlichen Rollen und Aufgaben, die Akteur*innen im Internet einnehmen. Journalist*innen sind etwa an eine Berufsethik gebunden, die mit bestimmten Standards bei Recherche und Berichterstattung einhergeht. Für Schüler*innen ist es wichtig, nachvollziehen zu können, dass es für die Herstellung von Informationen Ressourcen, Techniken und Prozesse gibt, die in der Regel durch die im Journalistikstudium erlernten Berufsstandards und durch Redaktionen abgesichert werden und deren Zweck (zumindest idealerweise) der demokratischen Meinungsbildung dient.

Grundlegend anders verläuft die Informationsverbreitung in den Sozialen Medien. Hier kommunizieren Individuen, Organisationen und Unternehmen privat, politisch oder kommerziell mit unterschiedlichsten Zielsetzungen. Diese Absichten – z. B. Public Relations, individuelle Selbstdarstellung, politische Stellungnahmen – gilt es zu identifizieren und Informationen vor diesem Hintergrund kritisch zu bewerten. Erst wenn auf diesem Weg Misstrauen kultiviert wird, kann das Vertrauen in Informationen ermöglicht werden, das für die Herstellung von anerkanntem Orientierungswissen notwendig ist.

Zur Vermittlung dieses kritischen Denkens in der schulischen Bildung kann auf ganz unterschiedliche Methoden und Hilfsmittel zurückgegriffen werden. Etablierte Ansätze im Bereich der Medienkompetenz (Baacke, 1996) oder neuere Konzepte unter den Begriffen Data oder Computer Literacy (Schüller, K., Koch, H. & Rampelt F. (2021) sind hier hilfreich, aber nicht ausreichend, um auf die vielfältigen und immer wieder neuen Herausforderungen eines Alltags mit digitalen Medien reagieren zu können.

Sie bedürfen der Erweiterung durch kontextbezogene Lehrangebote für die unterschiedlichen Fächer und gewinnen an zusätzlicher Überzeugungskraft, wenn Expert*innen von außen zu Unterrichtsbesuchen eingeladen werden. Zu denken ist hier z. B. an Verantwortliche („Faktenchecker“) aus Contentcentern, also Einrichtungen u. a. zur Verifizierung von Bild- und Videomaterial, aus dem Journalismus (z. B. das Recherchezentrum *Correctiv*) oder von NGOs wie dem *Chaos Computer Club* („Chaos macht Schule“).

Für die Unterstützung der Unterrichtsgestaltung in Bezug auf Kompetenzen zur Informationsbewertung kann auf einen Fundus von Anregungen im Landesmedienzentrum Baden-Württemberg zurückgegriffen werden. Im Angebot sind hier u. a. Peer-to-Peer-Ansätze in einem Mentorenprogramm, Konzepte für die pädagogische Arbeit von Lehrenden und Informationen zu aktuellen Trends in der Mediennutzung von Kindern und Jugendlichen. Konkretes Unterrichtsmaterial liefert ein Angebot des Bayerischen Rundfunks oder die Website Philovernetz.de mit

Bausteinen für den Philosophie- und Ethikunterricht aus Perspektive der Philosophiedidaktik.

Weiterhin können sich Kinder und Jugendliche über verschiedene Apps und Websites selbst mit dem Thema „Online-Informationen richtig bewerten“ beschäftigen. Hilfreich ist hier z. B. eine App (*Streitkultur 3.0*) der *Berghof Foundation* mit ausführlicher Beschreibung von Lehreinheiten, die zur Vertiefung der Nutzung einer eigens erstellten App im Unterricht genutzt werden kann. Eigene Versuche zur Überprüfung der Authentizität von Bildern und Videos können Schüler*innen auf der Website *deepware* unternehmen.

3. Literatur

Baacke, D. (1996). *Medienkompetenz – Begrifflichkeit und sozialer Wandel*. In A. von Rein (Hrsg.), *Medienkompetenz als Schlüsselbegriff* (S. 112-124). Bad Heilbrunn: Klinkhardt-Verlag.

Bayerischer Rundfunk (n.d.). <https://www.br.de/sogehmedien/stimmt-das/fake-news-durchschauen/index.html> [25.01.2023].

Berghof Foundation: N. Rieber, C. Sokele & S. Brunkhorst (2019). *Leitfaden zur Nutzung der Lern-App Streitkultur 3.0 für Lehrkräfte und Pädagog*innen*. https://www.vielfalt-mediathek.de/wp-content/uploads/2021/03/Berghof-Foundation_-Leitfaden_Streitkultur3.0_vielfalt-mediathek.pdf [25.01.2023].

Chaos Computer Club (CCC) (n.d.). „*Chaos macht Schule*“. <https://www.ccc.de/schule> [25.01.2023].

Correctiv. Recherchen für die Gesellschaft (n.d.). <https://correctiv.org/faktencheck/> [25.01.2023].

Deepware: Scan & Detect Deepfake Videos (n.d.). <https://deepware.ai/> [25.01.2023].

Grundmann, T. (2018). *Philosophische Wahrheitstheorien*. Stuttgart: Reclam.

Landesmedienzentrum Baden-Württemberg (n.d.). <https://www.lmz-bw.de/landesmedienzentrum/> [25.01.2023].

Philovernetz. Bausteine für den Philosophie- und Ethikunterricht (n.d.). <http://www.philovernetz.de/> [25.01.2023].

Pörksen, B. (2014). *Konstruktivismus: Medienethische Konsequenzen einer Theorie-Perspektive*. Wiesbaden: Springer VS.

Schüller, K., Koch, H. & Rampelt, F. (2021). *Data-Literacy-Charta. Version 1.2*. Berlin: Stifterverband. https://www.stifterverband.org/sites/default/files/data-literacy-charta_v1_2.pdf [25.01.2023].

Vaccari, C. (2013). From echo chamber to persuasive device? Rethinking the role of the Internet in campaigns. *New Media & Society*, 15(1), 109-127.

Angaben zur Autorin:

- PD Dr. Jessica Heesen, Leiterin des Forschungsschwerpunkts Medienethik, Technikphilosophie & KI am Internationalen Zentrum für Ethik in den Wissenschaften der Universität Tübingen, jessica.heesen@uni-tuebingen.de

Abbildungsverzeichnis

S. 15	Merk u. Bez/Abb. 1	Veranschaulichung der Bezugsnormen.
S. 18	Merk u. Bez/Abb. 2	Simultan kriteriale und soziale Normierung in der IRT.
S. 75	Hofstetter/Abb. 1	Beispiel einer praktischen Notenvergabe
S. 98	Schneider/Abb. 1	Relationalität und Pluralität als Knoten

Tabellenverzeichnis

S. 14	Merk/Bez	Tab. 1	Fiktive Ergebnisse eines Leseverständnistests.
S. 54	Burde/Paravicini	Tab. 1.	Kompetenzmatrix

Relativität und Normativität von Beurteilungen

Bewertungen und Beurteilungen sind in einer Leistungsgesellschaft omnipräsent. Aber wie sicher können – und wollen – wir uns der Verfahren und der (Qualitäts-) Maßstäbe sein, die dabei zum Einsatz kommen? An welchen fachspezifischen oder überfachlichen Werten und Normen orientiert sich schulische und hochschulische Leistungsbewertung, um ihre unterschiedlichen Funktionen zu erfüllen? Und wie „relativ“ darf Bewertung ausfallen, um den individuellen Voraussetzungen von Schülerinnen und Schülern gerecht zu werden? Während Fragen der Bewertung und Beurteilung vielfältig diskutiert werden, beispielsweise mit Blick auf diagnostische Gütekriterien, bleiben die dahinterliegenden Normen und Werte häufig wenig thematisiert. Gleichzeitig prägen sie die Beurteilungsprozesse in hohem Maße.

Der Herausgeberband widmet sich solchen Fragen und Herausforderungen des Schul- sowie Hochschulalltags unterschiedlichster Fächer und verbindet auf diese Weise erziehungswissenschaftliche, bildungswissenschaftliche und fachdidaktische Perspektiven, um das Problemfeld der Relativität und Normativität von Bewertung und Beurteilung zu erhellen.

[Prof. Dr. Carolin Führer](#) ist Lehrstuhlinhaberin für Deutsche Philologie/Didaktik der deutschen Literatur an der Eberhard Karls Universität Tübingen.

[Dr. Marco Magirius](#) ist als Literatur- und Mediendidaktiker Gastdozent an der Freien Universität Berlin.

[Prof. Dr. Thorsten Bohl](#) ist Erziehungswissenschaftler und seit 2015 Direktor der Tübingen School of Education.

[Prof. Dr. Bernd-Stefan Grewe](#) ist Historiker und leitet seit 2017 das Institut für Geschichtsdidaktik und Public History.

[Dr. Wolfgang Polleichtner](#) ist Klassischer Philologe und seit 2016 an der Universität Tübingen insbesondere mit der Fachdidaktik des Lateinischen und Griechischen betraut.

[Prof. Dr. Fahimah Ulfat](#) ist Religionspädagogin und seit 2020 Leiterin des Instituts für islamisch-religionspädagogische Forschung.