

<b>Zeitschrift:</b>	Schweizerische Zeitschrift für Bildungswissenschaften = Revue suisse des sciences de l'éducation = Rivista svizzera di scienze dell'educazione
<b>Herausgeber:</b>	Schweizerische Gesellschaft für Bildungsforschung
<b>Band:</b>	38 (2016)
<b>Heft:</b>	3
<b>Artikel:</b>	Anmerkungen zum Qualitätsbegriff in der Bildungsforschung
<b>Autor:</b>	Sauerwein, Markus / Klieme, Eckhard
<b>DOI:</b>	<a href="https://doi.org/10.5169/seals-786819">https://doi.org/10.5169/seals-786819</a>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

**Download PDF:** 12.01.2026

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

# Anmerkungen zum Qualitätsbegriff in der Bildungsforschung

**Markus Sauerwein und Eckhard Klieme**

«Die Frage: „Was ist Bildung? Was ist Qualität?“ wird immer ähnlich komplexe Antworten erfordern wie die Frage: „Was ist Wahrheit?“» (Becker 1962, S. 367)

*In dem Beitrag werden die Bedeutung und die Verwendung des Qualitätsbegriffs in der Bildungsforschung exemplarisch anhand von PISA und StEG betrachtet. Hierbei werden zwei Modi des Sprechens über Qualität in der Bildungsforschung unterschieden: ein empirisch analytischer und ein deskriptiver. Unabhängig vom Modus des Argumentierens hat Qualität in der Bildungsforschung eine wertende Konnotation. Grundlage dieser Wertungen sind immer normative Annahmen, die sich – je nach Modus in unterschiedlicher Form – in der Qualitätsbeurteilung niederschlagen. Bildungsforschung kann den normativen Gehalt ihrer Qualitätsaussagen explizieren und multiple Kriterien berücksichtigen und ist dadurch gerade nicht – wie einige kritisieren – «politikhörig».*

Empirische Bildungsforschung versteht sich – wie empirische Forschung allgemein – zunächst als «Tatsachenforschung». Ihre Grundelemente – wissenschaftstheoretisch: ihre Beobachtungssätze, ihr empirisches Relativ – bilden Feststellungen über Eigenschaften von «Objekten»<sup>1</sup>. Die entscheidende Frage für den hier zu führenden Meta-Diskurs lautet: Wie werden aus diesen «Tatsachenfeststellungen» Qualitätsfeststellungen, also wertende Urteile über bestimmte «Objekte» wie z.B. eine Schule, ein Bildungssystem, eine Unterrichtssequenz, wenn es darum geht, diese «Objekte» nicht nur zu beschreiben?

Dieser Diskurs ist auch vor dem Kontext einer wiederkehrenden Kritik an empirischer Bildungsforschung zu führen. So werden z.B. Vorwürfe erhoben, empirische Bildungsforschung sei politikhörig und vermische Normbeschreibung und Normdurchsetzung (Forster, 2015; Bellmann, 2015). Eine weniger pauschalisierende Version der Kritik weist auf die «weitgehend unkoordinierte, ja divergierende» (Hupka-Brunner, Imdorf, Grunder & Bergman 2015, S. 8) Verwendung des Qualitätsbegriffs hin (vgl. auch Heid, 2013a).

Im Folgenden soll auf einer Meta-Ebene aufgezeigt werden, dass die Kritik in ihrer provokativ vorgetragenen Form nicht zutreffend ist. Hierfür wird zunächst dargestellt, wie Qualitätsurteile in der empirischen Bildungsforschung gefällt werden und dass hierbei zwischen politischer Auftragsforschung und dem Kerngeschäft empirischer Bildungsforschung – wissenschaftlichen Studien – zu differenzieren ist. Indem zwei Modi aufgezeigt werden, wie aus Tatsachenbeschreibungen Qualitätsurteile werden und an welcher Stelle normative Vorannahmen in den Urteilsprozess einfließen, soll der Qualitätsbegriff systematisiert und zugleich begründet werden, dass die Grundsatzkritik, empirische Bildungsforschung verwendet einen unreflektierten, rein technologischen oder gar politik-hörigen Qualitätsbegriff, nicht zutreffend ist. Am Schluss soll zudem diskutiert werden, wie der wissenschaftliche Diskurs mit der Perspektive der eigenverantwortlich handelnden Subjekte, vor allem der professionellen Akteurinnen und Akteure im Bildungssystem, vermittelt werden kann.

## Qualität in der Bildungsforschung – zwei Beispiele

Seit etwa 30 Jahren hat es sich eingebürgert, weite Teile der Bildungsforschung - ob es um akademische Grundlagenforschung geht oder um wissenschaftliche Dienstleistungen wie etwa das sogenannte Bildungsmonitoring – als Untersuchungen zu «Schulqualität» (Fend, 1986, 1988; OECD, 1989), «Bildungsqualität von Schule» (Prenzel, Mertens & Noack, 1999), allgemeiner auch «Qualität im Bildungswesen» (Fend 2001) «Quality in education» oder «Quality education» (Colby, 2000) darzustellen. In diesem durchaus unscharfen Sinne untersucht Bildungsforschung die Qualität von Schulsystemen, Schulen, Unterricht, extracurricularen Bildungsangeboten (z.B. Ganztagsangeboten) etc. Für diese Untersuchungen wurden Modelle entwickelt, die zwischen verschiedenen strukturellen Bedingungen (Inputs), Aktivitäten und Maßnahmen (Prozesse) und den (Lern-)Ergebnissen der Schüler/-Innen (Outcomes) unterscheiden. Zusätzlich wird dabei zwischen verschiedenen Handlungsebenen differenziert, die von einer Systemebene (des Bildungswesens) über eine Schul- und Klassenebene bis zu den Schülerinnen und Schüler reichen (Klieme & Vieluf, 2013; OECD, 2013). Dieses Modell wird in der Literatur als «CIPO-Modell» (Context-Input-Process-Outcome-Modell) bezeichnet (Creemers & Kyriakidēs, 2008; vgl. weitere Arbeiten hier zu Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000), die auch auf Ganztagschulen adaptiert wurden (Spillebeen & Willems, 2014; Willems & Becker, 2015). Im Fokus des vorliegenden Artikels stehen jedoch nicht diese Modelle, die letztlich zu einer Strukturierung von Aussagen über Qualität dienen, sondern es wird reflektiert, wie (implizite) normative Setzungen die Qualitätsaussagen allgemein bestimmen, unabhängig davon, ob dies auf System-, Schul- oder Klassenebene geschieht.

Zur Veranschaulichung werden hierfür zwei Studien herangezogen: Die erste Studie, das Programme for International Student Assessment (PISA) der OECD, gilt vielfach als Prototyp von Studien, die Aussagen über die Qualität von Bildungssystemen treffen. Hier werden alle drei Jahre repräsentative Daten erhoben, um Indikatoren zu bilden und «*Wissen* zu schaffen über Faktoren, die die Qualität von Bildungssystemen beeinflussen» (Klieme & Vieluf, 2013, S. 229). In mittlerweile mehr als 70 Staaten werden dazu ca. eine halbe Million Jugendliche befragt und getestet, dazu Schulleitungen, Lehrkräfte und Eltern. Erfasst werden bekanntlich die Kompetenzen von Schülerinnen und Schülern in den Domänen Lesen, Mathematik und Naturwissenschaften. Daneben werden jedoch auch kognitive Schlüsselkompetenzen (Problemlösefähigkeit), soziale Kompetenzen (Perspektivenübernahme, Kooperationsfähigkeit, Zusammenarbeit in interkulturellen Kontexten), Meta-Kompetenzen (Metakognition, Selbstregulation, Selbstwirksamkeit) emotionale Kompetenzen (Empathie, Verantwortungsübernahme etc.), Verhaltenstendenzen (Lernstrategien, prosoziales Verhalten etc.) und affektive Merkmale (Lernmotivation, Lesefreude, Fachinteresse für Mathematik und Naturwissenschaften etc.) mit Testaufgaben geprüft oder als Selbstbeschreibungen erfasst. Neben diesen «Outcomes», die Bildungs- und Erziehungszielen entsprechen und deren Erreichen messbar machen sollen, werden bei PISA auch «Input»-Faktoren (z.B. Ausbildungsstand der Lehrkräfte) und Prozessmerkmale von Schule und Unterricht erfasst, welche ebenfalls als Aspekte von «Bildungsqualität» gelten können.

Eine weitere Studie, die hier exemplarisch vorgestellt werden soll, ist die Studie zur Entwicklung von Ganztagschulen (StEG). StEG beinhaltet eine Schulleitungsbefragung, bei der im Abstand von drei Jahren Schulleitungen über bestimmte ganztagschulspezifische Kriterien befragt werden wie z.B. den Umfang und die Art von Ganztagsangeboten an ihrer Schule oder die Verbindung von Ganztagsangeboten mit dem Unterricht. Daneben werden in StEG verschiedene Längsschnittstudien durchgeführt. Eine dieser Längsschnittstudien, an der die Autoren des vorliegenden Artikels in den vergangenen Jahren beteiligt waren, betrachtet u.a. Effekte des Besuchs sowie der Qualität von Ganztagsangeboten auf das Leseverstehen, die soziale Selbstwirksamkeit oder den Selbstwert der Schüler und Schülerinnen nach deren Übergang in die Sekundarstufe (Teilprojekt StEG-S). Das Projekt StEG ist somit ein Hybrid aus Bildungsmonitoring (Schulleitungsbefragung) einerseits und theoriebezogener Wirkungsstudie andererseits. Wie PISA ist auch StEG in der multidisziplinären Bildungsforschung verankert, nutzt deren Konstrukte und Messinstrumente, um Ziele zu überprüfen und Prozesse zu beschreiben.

Trotz dieser Unterschiedlichkeit geht es in allen Studien (PISA, StEG Schulleitungsstudie, StEG-Längsschnittstudie und den ihnen zugrunde liegenden Arbeiten aus der Bildungsforschung) um Qualität: In PISA werden Bildungssysteme verglichen, in der StEG Schulleitungsbefragung werden einzelne zuvor festgelegte Merkmale von Ganztagschule erhoben, u.a. um Aussagen über die

Qualität von Ganztagschulen in Deutschland zu machen, und in der Längsschnittstudie StEG-S werden u.a. Effekte bestimmter Qualitätsmerkmale in Ganztagsangeboten betrachtet, um Hinweise darüber zu erhalten, welche dieser Merkmale beispielsweise dazu beitragen können, den Selbstwert der Schülerinnen und Schüler zu stärken.

Diese exemplarische Darstellung soll zeigen, dass «Qualität» in der Bildungsforschung ein geläufiger Begriff zu sein scheint. Die Fragen nach der Qualität von Bildungssystemen, (Ganztags-)Schulen, Unterricht und auch außerunterrichtlichen Settings sind Gegenstand von Bildungsforschung. Welche Bedeutung der Qualitätsbegriff in der Bildungsforschung erhält und welche impliziten Annahmen hinter Qualitätsurteilen stecken, soll in diesem Beitrag betrachtet werden. Es handelt sich also um einen theoretischen Beitrag, dessen Argumentation über Beispiele aus den PISA und StEG Studien gestützt und veranschaulicht werden soll.

## Der Qualitätsbegriff in der Bildungsforschung

Qualität ist zunächst kein genuin wissenschaftlicher oder gar erziehungswissenschaftlicher Begriff, sondern Bestandteil unserer Alltagssprache. Diese alltagssprachliche Bedeutung kann beispielsweise über das digitale Wörterbuch der Deutschen Sprache ([www.dwds.de](http://www.dwds.de)) erschlossen werden, das auf den Arbeiten der Brüder Grimm aufbaut. Demnach hat Qualität (lat.: *qualitas* = Beschaffenheit, Merkmal, Eigenschaft, Zustand) zwei Bedeutungen:

- «Gesamtheit der Eigenschaften eines Produktes [...]»
- besonders gute Beschaffenheit» ([www.dwds.de](http://www.dwds.de) Stand 06.2016)

Einerseits kann Qualität folglich als die Beschreibung der Eigenschaften eines Produktes oder Objektes verstanden werden (*neutral*), andererseits als die Bewertung, im Sinne einer besonders guten Beschaffenheit, eines Objektes (*bewertend*) (Heid, 2000, 2013a; Klieme, 2013; OECD, 1989; Zollondz, 2010). Während die beschreibende Komponente intersubjektiv überprüfbar ist, trifft dies auf die präskriptive Komponente nicht zu, weil die dahinterliegenden normativen Bewertungskriterien nicht prinzipiell als wahr oder falsch gelten (Heid, 2015). In diesem Sinne ist Qualität selbst kein Objekt der Beurteilung, sondern das Resultat der Beurteilung eines Objektes (Heid, 2000). In jedem Fall kann und muss zwischen einer beschreibenden Aussage und einer bewertenden Aussage unterschieden werden.

An dieser Stelle kann im pädagogischen Kontext eine weitere Differenzierung vorgenommen werden. Berliner (2005) unterscheidet zwischen «good teaching» und «successful teaching». Unter «good teaching» ist das Einhalten bestimmter professioneller Standards zu verstehen, während «successful teaching» die Effektivität, also das Erreichen der Lernziele des Unterrichtens beschreibt. Wir schlagen an dieser Stelle vor, den Begriff des «good teaching» zu

erweitern und nicht nur professionelle Standards hierunter zu subsummieren, sondern auch ethische und moralische Vorstellungen wie etwa das Einhalten bestimmter (Kinder-)Rechte, und deshalb von *moralischen Qualitätsurteilen* zu sprechen. Beispielsweise begründet Prengel (2013) die Qualität pädagogischer Beziehungen nicht nur zweckbezogen, sondern berücksichtigt auch moralphilosophische oder tugendethische Konzepte<sup>2</sup>. Nach Oser (1998) sind die drei Werte Gerechtigkeit, Vertrauen und Wahrhaftigkeit konstitutiv für das Professionethos von Lehrkräften und Fend (1986; 1988) bezieht in seinem sozial-ökologischen Modell u.a. Angaben der Lehrkräfte zur Arbeitszufriedenheit sowie zur pädagogischen Zuwendung zur Schülerschaft mit ein, um «gute» und «schlechte» Schulen per Definition (!) zu unterscheiden. Qualität und Effektivität sind demnach nicht untrennbar miteinander verbunden, wie vielfach angenommen (u.a. Willems & Becker, 2015).

Demgegenüber steckt hinter «successful teaching» ein zweck- oder zielgebundenes Qualitätsverständnis. Sicherlich würde niemand der Tatsache widersprechen, «dass ein Unterricht ‚gut‘ ist, wenn er seinen Zweck erfüllt [...]. Aber tatsächlich gibt es eine Vielzahl unterschiedlicher Konkretisierungen dessen, was ‚der‘ Zweck, ‚der‘ Inhalt [...] eines als gut beurteilten Unterrichts ‚ist‘» (Heid, 2013a, S. 406). D.h., bevor beispielsweise nach der Qualität von Unterricht gefragt wird, muss zunächst der Zweck des Unterrichts festgelegt werden.

Unabhängig davon, ob moralisch/ethisch oder zweckgebunden über Qualität gesprochen wird - beides führt zwangsläufig dazu, dass Qualitätsurteile auf normativen Grundlagen beruhen, die von Qualitätsbeurteilenden – explizit oder implizit – im Kontext der Qualitätsbeurteilung gesetzt werden. Was Einer als «gut» beurteilt, kann eine Zweite als «schlecht» bewerten (Harvey & Green, 2000; Heid, 2000; Terhart, 2000).

Zusammenfassend ist festzuhalten, dass die Verwendung des Begriffs der Qualität in der Schulpraxis (z.B. bei Schulinspektionen) wie in der Bildungsforschung immer eine wertende Konnotation erhält. Ferner kann differenziert werden zwischen einer moralischen und einer zweck- bzw. zielgebundenen Qualitätsbestimmung. In beiden Fällen beruhen die Qualitätsurteile aber auf mehr oder weniger explizierten normativen Setzungen (Klieme, 2013; Terhart, 2000). An welcher Stelle sich diese normativen Setzungen in der konkreten empirischen Betrachtung von Qualität niederschlagen, soll nachfolgend aufgezeigt werden.

**Zwei Modi, wie normative Setzungen sich in Qualitätsaussagen der empirische Bildungsforschung wiederfinden**  
 In der empirischen Bildungsforschung können grob zwei Arten des Sprechens über Qualität unterschieden werden. Sie stellen zwei Argumentationsmodi dar, wie ausgehend von empirischen «Beobachtungs- oder Tatsachenfeststellungen» zu Qualitätsaussagen gelangt werden kann. Diese zwei Modi sind unabhängig von den oben angesprochenen Qualitätsmodellen anzutreffen.

Im ersten Fall wird Qualität über Wirkungs- bzw. Zusammenhangsanalysen festgestellt (1), im zweiten Fall werden bestimmte Merkmale oder Indikatoren und deren Ausprägung erfasst und unmittelbar als Qualitätsmerkmale interpretiert (2). Während ein moralisch motiviertes Qualitätsverständnis ausschließlich im zweiten Fall empirisch gestützt oder überprüft werden kann, ist es möglich, ein zweck- bzw. zielgebundenes Qualitätsverständnis in beiden Modi zu berücksichtigen, wie nachfolgend aufgezeigt werden soll.

(1) In Zusammenhangsanalysen wird der Einfluss bestimmter (Prozess-) Merkmale auf zuvor ausgewählte abhängige («Outcome»-) Variablen betrachtet. Dabei wird normativ vorausgesetzt, dass die «Outcome»-Variable als tatsächlich wünschenswertes Ergebnis anzusehen ist – auf System-, Schul- oder Klassenebene. Dies bedarf einer vorgängigen Legitimation (siehe unten). Beispielsweise kann bei fast allen «Outcome»-Variablen der PISA-Studien trefflich gestritten werden, ob diese wirklich für im deutschen Bildungssystem erwünschte und relevante Ziele stehen. Lässt sich nun im Verlauf eines Forschungsprozesses – von einzelnen Studien bis zu Reviews und Metaanalysen – für ein bestimmtes Prozess-Merkmal ein bedeutsamer Einfluss auf die abhängige Variable nachweisen, wird es als Qualitätsmerkmal bezeichnet. Die normative Setzung überträgt sich gewissermaßen von dem Zielkriterium (z.B. der mittels Test gemessenen Lesefähigkeit) auf die schulischen Prozesse. Auf den Unterricht bezogen, ergibt sich für Klieme (2013), dass erst von Unterrichtsqualität oder gutem Unterricht gesprochen werden kann, wenn die Ziele des Unterrichts durch bestimmte Unterrichtsmerkmale befördert werden. Unterrichtsqualität kann demnach als «Bündel von Unterrichtsmerkmalen [...] die sich positiv auf Unterrichts- oder Erziehungsziele [...] auswirken» (Einsiedler, 2002, S. 195) verstanden werden. Natürlich werden hierbei nicht irgendwelche Merkmale eingesetzt, sondern es wird basierend auf theoretischen Befunden und vorliegenden empirischen Arbeiten ausgewählt.

Gruschka (2011, S. 37) kritisiert ein solches Verständnis guten Unterrichts: «Guter Unterricht ist demnach der, der wirkt. Betrachtet werden allein die Mittel, aber nicht die Zwecke, denn diese gelten ja als vorgegeben». Richtig ist von dieser Aussage der erste Teil, aber dies bedeutet keineswegs, dass die Zwecke der Betrachtung entzogen werden sollen. Im Gegenteil: Die gesetzten Zwecke müssen konzeptuell so klar formuliert werden, dass sie in Messgrößen umgesetzt werden und ihr Zusammenhang mit Prozessmerkmalen expliziert werden kann. Die gleiche Analyse kann mit alternativen Zielvariablen wiederholt werden. So gehen Klieme und Rakoczy (2008) davon aus, dass neben fachlichem Verständnis auch Motivationszuwachs als Bewertungskriterium von Unterricht fungieren sollte. Mindestens diese zwei «Zwecke» werden in ihren Modellen durchgespielt. Dazu leiten sie Merkmale von Unterricht aus Unterrichts- und Lerntheorien ab, erfassen diese empirisch

und prüfen deren Effekt sowohl auf fachliches Verstehen als auch auf die Motivationsentwicklung, die somit als Bewertungskriterien für die Qualität des Unterrichts fungieren. Ein anderes Ziel von Unterricht würde demnach zu anderen Qualitätsdimensionen führen. Der dieser Logik zugrundeliegende Modus wird deshalb als *empirisch analytische Qualitätsbestimmung* (Abb.1) bezeichnet, weil versucht wird, die Zusammenhänge zwischen Merkmalen und bestimmten Outcomes nicht (nur) theoretisch, sondern empirisch zu bestimmen. Diesem Vorgehen folgt auch die oben dargestellte Längsschnittstudie StEG-S. Theoriegeleitet werden Merkmale von Ganztagsangeboten empirisch erfasst und deren Zusammenhänge mit bestimmten Outcomes betrachtet. Beispielsweise «wirken» sich Partizipationsmöglichkeiten in Ganztagsangeboten aus dem Bereich «Medien» positiv auf den Selbstwert der Schülerinnen und Schüler aus (Sauerwein, 2016) und sind nach der Logik des empirisch analytischen Modells deshalb als Qualitätsmerkmal anzusehen. Das normativ gesetzte Zielkriterium «Förderung des Selbstwertes» rechtfertigt zusammen mit der Wirkungsanalyse, «Partizipationsmöglichkeiten» als Aspekte der Qualität von Ganztagsangeboten zu bezeichnen.

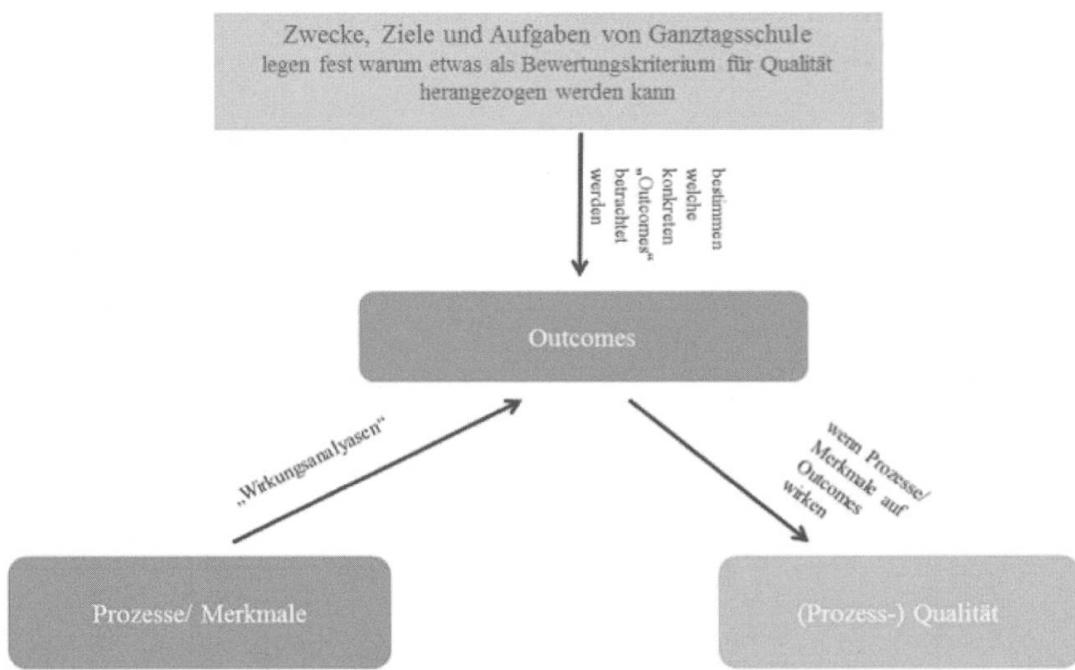


Abbildung 1 Empirische Betrachtung von Qualität: «Empirisch analytisch erklärende Qualitätsbestimmung « am Beispiel von Ganztagschulen

- (2) Eine zweite Variante besteht darin, das Vorhandensein bestimmter zuvor festgelegter Qualitätsmerkmale festzustellen (Abb.2). Die Bewertungskriterien für Qualität werden in diesem Modell nicht über bestimmte Outcomes repräsentiert, deren Erreichen über Zusammenhangsanalysen geprüft wird.

Dabei müssen die Qualitätsmerkmale jedoch theoretisch abgeleitet und argumentativ gestützt werden. D.h. aber auch, dass sich die theoretische Argumentation zugunsten dieser Qualitätsmerkmale entweder auf bestimmte Ziele und Zwecke stützen muss oder auf berufliche, ethische oder moralische Standards, wie etwa bei Prengel (2013), Oser (1998) oder Fend (1986). Heutzutage findet dieser Argumentationsmodus vor allem in sogenannten System- und Bildungsmonitorings und in der Schulevaluation Verwendung. Beispielsweise werden in der StEG Schulleitungsbefragung Schulleitungen zur Verbindung von Unterricht und Ganztagsangeboten befragt, weil angenommen wird, dass durch eine solche Verbindung Schülerinnen und Schüler besser individuell gefördert werden können. Auch diese Art der Qualitätsbestimmung ist wertend, weil es nicht bloß darum geht, Merkmale oder Eigenschaften – in diesem Fall Merkmale von Ganztagschulen – zu beschreiben, sondern um eine Beurteilung dieser Merkmale.

Die Festlegung der Qualitätskriterien kann sich dabei auf Setzungen etwa im Sinne professionsethischer Standards (Beispiel: Verzicht auf körperliche Gewalt in pädagogischen Handlungen) oder theoretischer Annahmen (Beispiel: die Bedeutung der Lebensweltorientierung in der Reformpädagogik) stützen, ohne dass diese empirisch abgesichert wären, aber es können auch empirische Ergebnisse herangezogen werden. Beispielsweise hat sich das Unterrichtsmerkmal «kognitive Aktivierung» in empirischen Studien als lernwirksam erwiesen (Lipowsky et al., 2009; Klieme, Schümer & Knoll, 2001), könnte nun für die Schulevaluation als Qualitätskriterium herangezogen und über fragebogengestützte Expertenratings eingeschätzt werden, ohne dass die Wirksamkeit kognitiver Aktivierung nochmals empirisch überprüft werden muss. Es ist demnach über einen solchen Modus durchaus möglich, die Qualität des Unterrichts zu beurteilen (Pietsch, 2010). Diese Art der Qualitätsbestimmung kann als *systematisch deskriptive* bezeichnet werden (Abb. 2).

Die *systematisch deskriptive* Art der Qualitätsbestimmung kann ebenfalls eingesetzt werden, wenn die Zwecke, die über ein bestimmtes Qualitätsmerkmal erreicht werden sollen, nur schwer empirisch abzubilden sind. Beispielsweise sollen Partizipationsgelegenheiten auf das Leben in der Demokratie vorbereiten. Der Umgang mit der Demokratie ist jedoch etwas Abstraktes, welches sich nicht umstandslos empirisch operationalisieren lässt, gerade weil Demokratien tolerant gegenüber verschiedenen Lebensweisen sind und sich damit beliebig viele Outcomes ableiten lassen. Im Interesse (!) der Demokratieförderung könnte man daher auf Operationalisierungen verzichten und Partizipation als ein Qualitätsmerkmal sui generis normativ setzen; tatsächlich finden solche Setzungen u.a. in der Sozialpädagogik häufig statt. Wer mit derartigen Setzungen arbeitet, unterschätzt jedoch mitunter die Möglichkeiten empirischer Forschung – etwa wenn es gilt, auch Wirkungen von Partizipationsgelegenheiten empirisch zu betrachten. In StEG-S wird

beispielsweise analysiert, ob Partizipation in Ganztagsangeboten die soziale Selbstwirksamkeit der Schülerinnen und Schüler positiv beeinflussen kann, weil dies als ein relevantes Merkmal angesehen wird, um sich später in die Gesellschaft aktiv einbringen zu können (Sauerwein, 2016).

Allgemein gesprochen: Empirische Befunde mögen Anlass geben, normative Setzungen zu hinterfragen, aber sie können solche Setzungen nicht aufheben. Ein historisches Beispiel hierfür ist die Tatsache, dass sozialdemokratische Bildungsreformer aufgrund gesellschaftspolitischer Überzeugungen an der Gesamtschule festhielten, obwohl die Studien von Fend deren Wirksamkeit in Frage stellten (von Friedeburg, 1989).

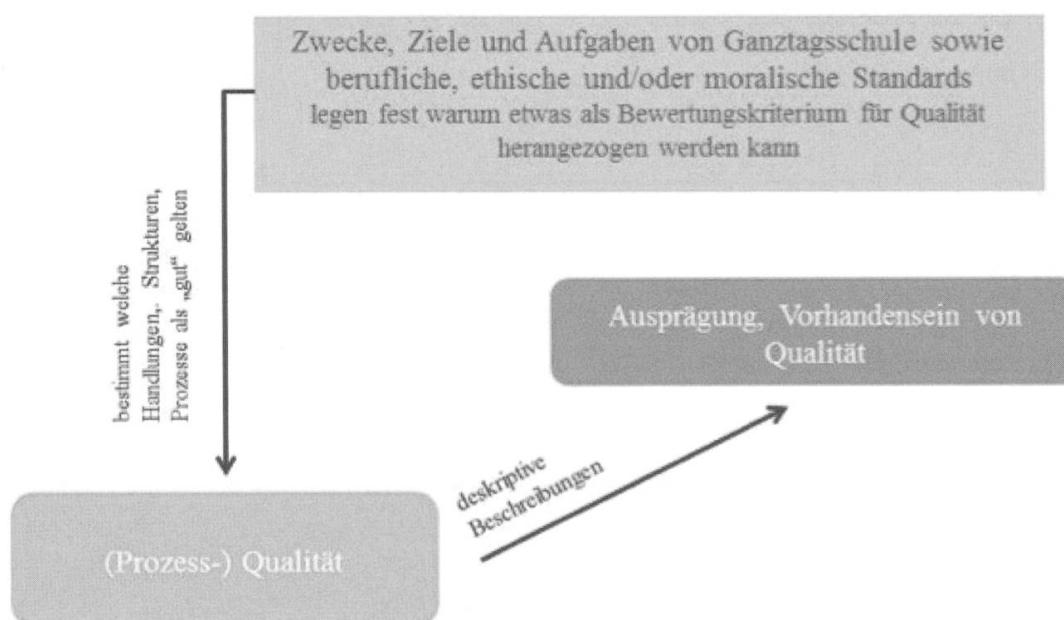


Abbildung 2: Empirische Betrachtung von Qualität: deskriptive Qualitätsbestimmung am Beispiel von Ganztagschulen

Beiden Modi der Qualitätsbestimmung – dem *empirisch analytisch erklärenden* (Abb.1) und dem *deskriptiven* (Abb.2) – ist gemein, dass Qualität sich nur über eine normative Setzung erschließt, weil die Bestimmung der Ziele bzw. Zwecke oder der moralischen bzw. ethischen Standards des zu betrachten Objekts normativ ist. Beide Modelle unterscheiden sich hinsichtlich der Art und Weise, wie sich die normativen Setzungen in den jeweiligen Modellen wiederfinden lassen. Während in der *empirisch-analytisch erklärenden Qualitätsbestimmung* letztlich der Zusammenhang mit Outcomes, die bestimmten Bildungs- und Erziehungszielen entsprechen, entscheidend dafür ist, ob ein Merkmal zu Qualität wird, ist beim *deskriptiven Modus* die Frage des Zusammenhangs zwischen Qualitätsmerkmalen und Zielen in den theoretischen und normativ-professionellen Diskurs verlagert. Auf welcher Ebene (System-, Schul-, Klassenebene) die Qualitätsfeststellung letztlich stattfindet, ist für die

Systematisierung des Forschungsprozesses relevant, für die Logik der Qualitätsbestimmung jedoch unerheblich.

Der *theoretische Diskurs* der Begründung von Qualitätskriterien ist freilich Gegenstand von Bildungsforschung und Erziehungswissenschaft. Wie bereits angesprochen geht es damit sehr wohl um die Ziele und Zwecke von Unterricht. Beim Beispiel der Partizipation etwa lässt sich auf Dewey (1949) rekurrieren, demzufolge Partizipation durch den Prozess des Partizipierens erlernt werden soll (*learning by doing*). Anders formuliert, Partizipation ist theoretisch als Qualitätsmerkmal anzusehen, weil Jugendliche durch den Prozess des Partizipierens Fähigkeiten zur Partizipation erlernen und diese Fähigkeiten als notwendig zum Erhalt der Demokratie angesehen werden. Gegenüber diesem wissenschaftlichen Diskurs zeichnet sich der *professionelle Diskurs*, etwa unter Lehrkräften oder Expertinnen und Experten der Schulaufsicht und Schulevaluation, durch die Frage aus, welche Kriterien in Kenntnis von wissenschaftlichen Befunden, gesellschaftlichen Erwartungen, politischen Vorgaben und tradierten Praktiken pragmatisch ausgewählt werden sollen; solche Diskurse werden dann etwa in einem «Qualitätsmodell der Schulinspektion» (Hessisches Kultusministerium, 2011) gebündelt. Der wissenschaftliche Diskurs kann – aufgrund seiner Orientierung an allgemeinen Wahrheitszielen und methodischen Standards der Erkenntnisgewinnung – nie mit normativen Setzungen abgebrochen werden. Der professionelle Diskurs hingegen kann, ja muss im Interesse administrativer und pädagogischer Handlungsfähigkeit immer wieder pragmatisch beendet werden. In diesem Sinne – und nicht im Sinne einer Entpolitisierung oder Unterwerfung der Wissenschaft, wie Forster (2015) meint – ist die Aussage von Klieme (2013) zu verstehen, dass eine gezielte Setzung von Kriterien keine Wissenschaft sei, sondern typisch für professionelle, administrative und politische Handlungszusammenhänge.

### Implizite Festlegung der Normen der Qualitätssetzung

Qualitätsurteile beruhen, wie es in den vorangestellten Abschnitten bereits angeklungen ist, stets auf normativen Grundlagen (Heid, 2000, 2013a; Klieme, 2013; Klieme & Tippelt, 2008; Terhart, 2000), unabhängig davon, ob sie moralisch motiviert erfolgen oder ziel- bzw. zweckgebunden. Bezugnehmend auf Heid (2000) betonen Klieme und Tippelt (2008, S. 9) deshalb, «dass Qualitätsfeststellungen einer normativen Grundlage bedürfen, d. h. einer zumindest impliziten Setzung von Gütekriterien, die nicht wahr oder falsch, sondern nur unterschiedlich gut legitimierbar oder nützlich sein kann». Der normative Gehalt der Qualitätsurteile und damit verbundene Herausforderungen für die Bildungsforschung sollen nachfolgend genauer betrachtet werden – auch mit dem Ziel der Klarstellung von Missverständnissen, die manchen Kommen-

taren im Anschluss an die Diskussion in der Zeitschrift für Erziehungswissenschaft 2013 zwischen Heid (2013a; 2013b) und Klieme (2013) innewohnen (Bellmann, 2015; Forster, 2015). Es soll hier nochmals deutlich gemacht werden, dass empirische Bildungsforschung ihre normativen Grundlagen und damit ihre Verantwortung für das, was sie tut und nicht tut, keineswegs leugnet.

Zweifelsohne sind Bildungsforschende in einem Dilemma: Einerseits sollte die Wissenschaftlerin bzw. der Wissenschaftler bei der Bestimmung der normativen Bewertungskriterien neutral sein, klar ist aber auch, dass Wissenschaft nicht frei von Wertvorstellungen und Werturteilen ist (Becker, 1962, S. 371). In Arbeiten, die die Wirkung von Unterrichtsqualität auf Leistung betrachten (Helmke & Klieme, 2008; Klieme, Lipowsky, Rakoczy & Ratzka, 2006; Klieme, Pauli & Reusser, 2009; Klieme & Rakoczy, 2008; Klieme et al., 2001), ist beispielsweise der Zweck, an dem die Qualität des Unterrichts bestimmt wird, die Leistung bzw. die Verbesserung der Leistung von Schülerinnen und Schülern und gleichzeitig ein positiver Einfluss auf affektive Merkmale wie z.B. die Lernmotivation. Dabei bleiben notwendigerweise andere mögliche Ziele unberücksichtigt (so etwa «Mündigkeit», um eine Alternative zu benennen). Auch wenn Fragen herkunftsbedingter Disparitäten im Bildungssystem betrachtet werden (Baumert, 2001, 2002; 2003), erfolgt dies auf Basis eines bestimmten Verständnissen, zumeist in der Hinsicht, dass herkunftsbedingte Disparitäten als ein Problem für das meritokratische Verständnis unserer Gesellschaft aufgefasst werden. Die Setzung, gute Bildungssysteme gewährleisteten Chancengerechtigkeit, ist demnach die normative Setzung eines Bewertungskriteriums für die Qualität von Bildungssystemen (Heinrich, 2010). Forschungsfragen, die sich mit Qualität in Bildungssettings oder von Bildungssystemen beschäftigen, sind also werturteilsbehaftet, weil entweder (implizite) normative Zuschreibungen der Ziele des Objekts die Qualität bedingen – die Förderung der Leistung der Schülerinnen und Schüler als Zweck des Unterrichts, die Herstellung von Chancengerechtigkeit als Zweck des Schulsystems – oder professionelle, moralische oder ethische Vorstellungen gesetzt werden. Zu Recht darf hierbei kritisch betrachtet werden, dass oftmals nur Leistung als Bewertungskriterium fungiert (Scharenberg, 2015) oder andere Bewertungskriterien in den Hintergrund rücken. Wie eingangs erwähnt, werden immerhin in PISA und StEG vielfältige Bildungsziele operationalisiert, in der öffentlichen Wahrnehmung scheint jedoch Leistung zu überwiegen.

Forster (2015) wirft der Bildungsforschung an dieser Stelle eine Verzahnung von Normdurchsetzung und Normbeschreibung vor und geht davon aus, dass Wertungen Bestandteile des gesamten Forschungsprozesses sind. Diese «Erkenntnis» ist jedoch keineswegs neu und mag bei genauerer Betrachtung trivial erscheinen.<sup>3</sup> Natürlich sind Bildungsforschende nicht frei von normativen Implikationen und politischen Impulsen, die in die eigene Forschungspraxis Eingang finden, weil letztlich kein Mensch und keine Wissenschaft frei von Werturteilen ist, wie bereits dargelegt wurde (Becker, 1962; Heinrich, 2015;

Horkheimer, 1937), jedoch ist zu unterscheiden zwischen Wissenschaft und dem gezielten Setzen von Qualitätskriterien durch die Bildungspolitik (Klieme, 2013). In letztem Fall haben wir es demnach auch zwangsläufig mit einer politischen Frage zu tun (Hupka-Brunner, 2015). Bildungsforschung weiß um diese Unterscheidung.

Viele Kritikerinnen und Kritiker empirischer Bildungsforschung differenzieren hier nicht hinreichend zwischen den Kontexten der Qualitätsurteile (als Evaluierung im Rahmen administrativer Qualitätssicherung, als Bericht an Politik und Öffentlichkeit im Rahmen einer anwendungsbezogenen Auftragsforschung, in wissenschaftlichen Publikationen etc.), noch wird sich die Mühe gemacht, die unterschiedlichen Modi zu berücksichtigen auf deren Basis Qualitätsurteile gefällt werden (siehe oben), sondern stattdessen pauschal die «PISA-Forschung» im Konkreten sowie die empirische Bildungsforschung im Allgemeinen als «politikhörig» abgestempelt. So unterstellt Forster (2015, S. 65f.) Klieme «die Vorstellung, dass es keine grundlegenden gesellschaftlichen Interessenkonflikte und damit keine strategische Option oder politische Alternativen, sondern nur sachbezogene Entscheidungen gebe». Bei genauer Betrachtung ist das Gegenteil der Fall: Wenn empirisch arbeitende Wissenschaftlerinnen und Wissenschaftler explizit machen, welche Bewertungskriterien und Argumentationsmodelle sie jeweils nutzen, und wenn sie darüber hinaus die theoretischen und methodischen Setzungen ihrer Studien offen legen (etwa dass sich die Basisdimensionen guten Unterrichts auf die Ziele von Unterricht nach Motivations- und Leistungszuwachs beziehen), so wird der normative Diskurs erst ermöglicht: als Diskurs über Ziele von Bildung und Erziehung, und als Diskurs über die angemessene Art des wissenschaftlichen Umgangs mit diesen Zielen. Kritikerinnen und Kritiker der empirischen Bildungsforschung begehen einen Fehlschluss, wenn sie politikbezogene Berichte gleichsetzen mit dem Kerngeschäft empirischer Bildungsforschung – wissenschaftlichen Studien, Publikationen und Diskussionen – und den Differenzierungsgrad der Forschung, die sie verurteilen, nicht zur Kenntnis nehmen.

Diese Unterscheidung zwischen wissenschaftlichen Publikationen und Auftragsforschung bzw. Evaluation ist wichtig, denn im ersten Fall kann und muss Wissenschaft Bewertungskriterien, wie etwa das Erreichen bestimmter Bildungsstandards selbst setzen, – wobei gute Forschung in der Regel multiple Ziele berücksichtigt und ggf. Zielkonflikte aufzeigt – während im zweiten Fall die Bewertungskriterien durch die Auftraggebenden (also z.B. die Bildungspolitik) vorgegeben werden – freilich informiert durch Ergebnisse empirischer Studien oder theoretischer Reflexion. Es gehört beispielsweise zu den Standards professioneller Evaluation, dass Evaluatoren und Evaluatoren im Prozess der Zielklärung beratend tätig sind, letztlich aber jene Kriterien akzeptieren müssen, die sich die Akteurinnen und Akteure selbst gesetzt haben und nicht stattdessen eigene setzen können. Dies soll an einem konkreten Beispiel verdeutlicht werden: Wenn wir im Kontext der StEG-Studie Politik und Öffentlichkeit

über Qualitätsmerkmale von Ganztagschulen Bericht erstatten, müssen sich die normativen Setzungen weitestgehend an den politischen Erwartungen und Zielen von Ganztagschule (Züchner & Fischer, 2011; Wiere, 2011) orientieren. «StEG-Forscher und Forscherinnen» sind keine demokratisch legitimierten Vertreterinnen und Vertreter, die eigene Ziele von Ganztagschule untereinander aushandeln und statt der politischen setzen können. Jenseits des politischen Kontextes sind Forscherinnen und Forscher frei, eigene Kriterien zu formulieren und zu begründen, die politischen Zielsetzungen konträr gegenüberstehen dürfen; sie müssen diese ggf. vor der wissenschaftlichen Öffentlichkeit vertreten und verteidigen. Wer Auftragsforschung wie PISA oder StEG verantwortungsvoll macht, wird das Design so anlegen, dass unterschiedliche Bewertungskriterien angewandt werden können.

Resümierend ist deshalb festzuhalten, dass Qualitätsurteile der Bildungsforschung immer auf normativen Setzungen beruhen. Entweder werden bestimmte Zwecke oder Ziele des zu untersuchenden Gegenstands thematisiert, oder es werden berufliche, ethische, moralische etc. Standards gesetzt. Quantitative Analysen mögen den Anschein von Werturteilsfreiheit und Neutralität geben; in der Bildungsforschung kann und sollte es jedoch nicht darum gehen, normative Implikationen zu verdecken, sondern diese offen darzulegen. Sicherlich darf von Bildungsforschung eingefordert werden, die normativen Grundlagen ihrer Qualitätsurteile offenzulegen. Vorwürfe einer «Politikhörigkeit», wie u.a. von Forster (2015) erhoben, sind zurückzuweisen, insbesondere mit Verweis auf das Kerngeschäft empirischer Bildungsforschung jenseits einer Berichtslegung für Politik und Öffentlichkeit.

## Fazit

In dem Beitrag wurde das Qualitätsverständnis der Bildungsforschung vor dem Hintergrund aktueller Diskurse dargestellt. Qualität wird in der Bildungsforschung nicht deskriptiv, sondern wertend verstanden und ist deshalb zwangsläufig normativ. In der praktischen Arbeit können in der Bildungsforschung zwei Modi ausfindig gemacht werden, in denen über Qualität gesprochen wird. Ein empirisch analytischer Modus versucht mittels (längsschnittlicher) Analysen Effekte bestimmter Merkmale auf zuvor festgelegte Outcomes zu erfassen<sup>4</sup>. Ein deskriptiver Modus betrachtet vorher definierte Qualitätsmerkmale, weil diese entweder aus ethisch/moralischen Gründen als bedeutsam angesehen werden oder aber theoretisch mit bestimmten Zielen in Verbindung stehen.

Erkenntnisse der Bildungsforschung – wie beispielsweise Qualitätsurteile – sollen auch auf Politik und Praxis zurückwirken, denn nicht zuletzt darüber erhält Bildungsforschung ihre gesellschaftliche Legitimation. Bei diesem Transfer, sei es in die Bildungspraxis oder in die Bildungspolitik, können Wissenschaftler

und Wissenschaftlerinnen die Rezeption ihrer Ergebnisse letztlich nicht beeinflussen und kontrollieren, weil auf der «anderen Seite» ebenfalls selbständige denkende und handelnde Subjekte stehen (Heid, 2015). Deshalb ist dieser «Transfer» nicht als «sozialtechnologische» Umsetzung gedacht, wie erziehungswissenschaftlich reflektierte Bildungsforschung – entgegen den Unterstellungen etwa bei Forster (2015) – durchaus weiß. Zudem stehen Praktikerinnen und Praktiker oftmals vor einer Vielzahl von wissenschaftlichen Arbeiten mit einer hohen Divergenz an Forschungsergebnissen, die zudem kontrovers diskutiert (Heid 2015) und in der praktischen Umsetzung neu kontextualisiert werden müssen (Hartmann, Decristan & Klieme, im Druck), so dass eine nahtlose Implementierung der Qualitätsurteile von Bildungsforschung unmöglich ist. Selbstkritisch einzuräumen ist dennoch, dass empirische Bildungsforschung vermehrt auf Zielkriterien jenseits von Leistung verweisen sollte, denn anscheinend wird empirische Bildungsforschung selbst in Teilen der erziehungswissenschaftlichen Forschungsgemeinschaft ausschließlich als Leistungsmessung verstanden.

In der Bildungspolitik zeichnet sich ab, dass Bildungsforschung zur Legitimation politischen Handelns herangezogen wird (Tröhler, 2010, 2015). Der Ausbau der Ganztagschulen wurde beispielsweise auch über das schlechte Abschneiden Deutschlands in der PISA Studie begründet (Zickgraf & Zickgraf, 2009), obwohl kein empirischer Nachweis eines Zusammenhangs vorlag; die politische Entscheidung war vielmehr unabhängig erfolgt (Tillmann, Dederling, Kneuper, Kuhlmann & Nesssel, 2008). Bildungsforschung steht hier möglicherweise vor dem Problem, unfreiwillig «zur ‚Legitimationswissenschaft‘ gemacht zu werden, die nur dann zur Legitimation politischen Handelns herangezogen wird, wenn die Empfehlungen oder Befunde dem politischen Programm entsprechen» (Heinrich, 2015, S. 293). Allerdings entstehen derartige Probleme systematisch an der Grenze zwischen Wissenschafts- und Politiksystem, auch in anderen gesellschaftlichen Handlungsfeldern. Beim Transfer sowohl in die Bildungspraxis als auch in die Bildungspolitik haben Wissenschaftlerinnen und Wissenschaftler weder das Mandat noch die Macht, ihre eigenen Bewertungskriterien und damit auch ihre eigenen Qualitätsurteile durchzusetzen (Heid, 2015). Dies ist plausibel, für die einzelnen jedoch nicht unkritisch, wenn Ergebnisse fehlerhaft aufgegriffen, uminterpretiert oder in ihrer Komplexität reduziert werden<sup>5</sup>.

In der Präsentation wissenschaftlicher Qualitätsurteile scheint demnach ein weiterer neuralgischer Aspekt zu liegen, insbesondere wenn die normativen Grundlagen, auf denen die Qualitätsurteile beruhen, selbst nicht mehr Gegenstand von professionellen, politischen und wissenschaftlichen Diskursen sind, sondern unhinterfragt übernommen und die Ergebnisse in vermeintlicher Wertneutralität dargestellt werden. Wie in diesem Beitrag dargelegt, gehört es zu den Qualitätskriterien (!) guter wissenschaftlicher Arbeit, dass Bildungsforscherinnen und Bildungsforscher die normativen Grundlagen ihrer Arbeiten transparent darlegen. Selbstkritisch muss angemerkt werden, dass die Reflektion

der normativen Grundlagen weiter verbessert werden könnte, sowohl in anwendungsbezogenen als auch in wissenschaftlichen Publikationen, um sich einerseits vor einer Instrumentalisierung durch die Bildungspolitik zu schützen und gleichzeitig die eigenen Ergebnisse einschließlich möglicher Inkonsistenzen – wie sie beispielsweise entstehen, wenn unterschiedliche Bewertungskriterien und/oder Operationalisierungen angewandt werden – für Praktiker und Praktikerinnen verständlich darzustellen. Akteurinnen und Akteure des Wissenschaftssystems sind dem Prinzip der «Wahrheitssuche» verpflichtet. Sie haben kein Mandat, eigene normative Vorstellungen in anderen gesellschaftlichen Systemen durchzusetzen. Gerade in Feldern, die gesellschaftliche und professionelle Praxis stark beeinflussen können wie z.B. der Gesundheitsforschung, wird aber zugleich erwartet, dass Wissenschaft zu einer «grundlegenden Verbesserung der menschlichen Existenz» (Horkheimer, 1937, S. 284) beiträgt. Dies bedeutet im Falle von Bildungsforschung, Bildungspolitik und Bildungspraxis zu hinterfragen und soweit möglich durch handlungsrelevantes Wissen zu einer Verbesserung von Bildungssystemen, Schule, Unterricht etc. beizutragen – vor dem Kontext einer Explikation der eigenen normativen Grundlagen.

### Anmerkungen

- 1 Selbstverständlich sind auch «Beobachtungssätze» nicht «neutral»: Die Auswahl der Prädikate, die Abgrenzung der zu beobachtenden Einheiten, der Vorgang der Beobachtung selbst beruhen auf theoretischen Vorannahmen, normativen und untersuchungspraktischen Setzungen.
- 2 Zum Beispiel: «Nach Nussbaum (2002, S. 12) gehören Gefühle der Liebe, der Sympathie, der Fürsorge zum ‚innersten Kern des ethischen Lebens‘» (Prengel & Winklhofer, 2014, S. 33).
- 3 Siehe oben Fußnote 2. Für die Auswahl von Fragestellungen und Messbereichen in Bildungsstudien gilt ebenfalls, dass sie von externen, auch politischen Überlegungen abhängen. Beispielsweise greift das Design der PISA-Fragebögen Kenntnisse und Fragestellungen der Schuleffektivitätsforschung auf (Klieme & Vieluf, 2013); die Priorisierung und damit die Auswahl der Instrumente erfolgt letztlich in politischen Gremien. Auch die Auswahl der drei Kernkompetenzen der internationalen Schulleistungsstudien entsprang in den 1960er Jahren politischen Setzungen – wie Tröhler (2010) meint, letztlich als Reaktion auf den Kalten Krieg und den «Sputnikschock», da diese drei Kompetenzen als relevante Kriterien angesehen wurden, um den wissenschaftlichen und ökonomischen Wettkampf gegen das sowjetische System zu gewinnen.
- 4 Eine gesonderte, in den Sozialwissenschaften hoch differenziert und methodisch aufwändig geführte Debatte befasst sich damit, unter welchen Voraussetzungen Effekte kausal interpretiert werden können. Wir können sie an dieser Stelle aus Platzgründen nicht einbeziehen.
- 5 Beispielsweise wird in wissenschaftlichen Arbeiten die Bedeutung einer *kognitiven* Aktivierung der Schülerinnen und Schüler im Vergleich zu einer Handlungs-Aktivierung für Lernprozesse betont (Klieme & Rakoczy, 2008). Praktikerinnen und Praktiker antworten auf die Frage, was für Sie guter Unterricht sei, jedoch nur mit «Aktivierung, die ohne den Zusatz des Kognitiven auskommt» (Hartmann, Decristan & Klieme, im Druck).

## Literaturverzeichnis

- Baumert, J. (Hrsg.) (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J. (Hrsg.) (2002). *Pisa 2000 - die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.
- Baumert, J. (Hrsg.) (2003). *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- Becker, H. (1962). *Quantität und Qualität. Grundfragen der Bildungspolitik*. Freiburg im Breisgau: Rombach & Co.
- Bellmann, J. (2015). Symptome der gleichzeitigen Politisierung und Entpolitisierung der Erziehungswissenschaft im Kontext datengetriebener Steuerung. *Erziehungswissenschaft*, 50, 45–54.
- Berliner, D. C. (2005). The Near Impossibility of Testing for Teacher Quality. *Journal of Teacher Education* 56(3), 205–213.
- Creemers, B.P. M. & Kyriakidēs, L. (2008). *The dynamics of educational effectiveness. A contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- Colby, J. (2000). *Defining Quality in Education. UNICEF working paper*. New York: United Nations Children's Fund.
- Dewey, J. (1949). *Demokratie und Erziehung. Eine Einleitung in die philosophische Pädagogik*. Braunschweig: Georg Westermann Verlag.
- Einsiedler, W. (2002). Das Konzept «Unterrichtsqualität». *Unterrichtswissenschaft* 30(3), 194–196.
- Fend, H. (1986). «Gute Schulen - schlechte Schulen». Die einzelne Schule als pädagogische Handlungseinheit. *Die Deutsche Schule* 78(3), 275–294.
- Fend, H. (1988). Schulqualität. Die Wiederentdeckung der Schule als pädagogische Gestaltungsebene. *Neue Sammlung* 28(4), 537–547.
- Fend, H. (2001). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. 2., bereinigte Aufl. Weinheim: Juventa.
- Forster, E. (2015). Zur Kritik partizipativer Wissenspolitik. *Erziehungswissenschaft*, 50, 65–73.
- Gruschka, A. (2011). *Verstehen lehren. Ein Plädoyer für guten Unterricht*. Ditzingen: Reclam.
- Hartmann, U., Decristan, J. & Klieme, E. (im Druck). Unterricht als Feld evidenzbasierter Bildungspraxis? Herausforderungen und Potenziale für einen wechselseitigen Austausch von Wissenschaft und Schulpraxis. *Zeitschrift für Erziehungswissenschaft*.
- Harvey, L. & Green, D. (2000). Qualität definieren. Fünf unterschiedliche Ansätze. *Zeitschrift für Pädagogik Beiheft*, 41, 17–40.
- Heid, H. (2000). Qualität: Überlegungen zur Begründung einer pädagogischen Beurteilungskategorie. *Zeitschrift für Pädagogik, Beiheft*, 41, 41–54.
- Heid, H. (2013a). Logik, Struktur und Prozess der Qualitätsbeurteilung von Schule und Unterricht. *Zeitschrift für Erziehungswissenschaft* 16(2), 405–431.
- Heid, H. (2013b). Qualitätsurteile als Untersuchungsgegenstände – eine Erwiderung auf Eckhard Klieme. *Zeitschrift für Erziehungswissenschaft* 16(2), 443–446.
- Heid, H. (2015). Bildungsforschung im Kontext gesellschaftlicher Praxis. Über (soziale) Bedingungen der Möglichkeit, Bildungspraxis durch Bildungsforschung zu beeinflussen. *Zeitschrift für Pädagogik* 61(3), 390–409.
- Heinrich, M. (2010). Bildungsgerechtigkeit durch Evidence-based-Policy? Governanceanalysen zu einem bildungspolitischen Programm. In W. Böttcher, J. N. Dicke & N. Hogrefe (Hrsg.), *Evaluation, Bildung und Gesellschaft. Steuerungsinstrumente zwischen Anspruch und Wirklichkeit* (S. 47–68). Münster: Waxmann.
- Heinrich, M. (2015). Neue «Vergessene Zusammenhänge?». Pädagogisches Unbehagen anlässlich Heinz-Elmar Tenorths Verhältnisbestimmung von Bildungspolitik und Bildungsforschung. *Die Deutsche Schule* 107(3), 285–298.

- Helmke, A. & Klieme, E. (2008). Unterricht und Entwicklung sprachlicher Kompetenzen. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 301–312). Weinheim: Beltz .
- Hessisches Kultusministerium (2011). *Hessischer Referenzrahmen Schulqualität, überarbeitete Fassung*. Wiesbaden: Institut für Qualitätsentwicklung.
- Horkheimer, M. (1937). Traditionelle und kritische Theorie. *Zeitschrift für Sozialforschung* 6(2), 243–294.
- Hupka-Brunner, S., Grunder, H.-S., Bergmann, M. & Imdorf, C. (2015). *Qualität in der Bildung*. Bad Heilbrunn: Klinkhardt.
- Hupka-Brunner, S., Imdorf, C., Grunder, H.-U. & Bergmann, M. (2015). Zur Einführung. Der Qualitätsdiskurs im Bildungsbereich - eine kritische Bestandsaufnahme. In S. Hupka-Brunner, H.-U. Grunder, M. Bergmann & C. Imdorf (Hrsg.), *Qualität in der Bildung* (S. 7–12). Bad Heilbrunn: Klinkhardt..Klieme, E. (2013). Qualitätsbeurteilung von Schule und Unterricht: Möglichkeiten und Grenzen einer begriffsanalytischen Reflexion – ein Kommentar zu Helmut Heid. *Zeitschrift für Erziehungswissenschaft* 16(2), 433–441.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. theoretische Grundlagen und ausgewählte Ergebnisse des Projekts Pythagoras. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms [BIQUA]* () Münster: Waxmann,
- Klieme, E., Pauli, C. & Reusser K. (2009). The Pythagoras study: investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Hrsg.), *The power of video studies in investigating and learning in the classroom* (S. 137–160). Münster: Waxmann..
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik* 54(2), 222–237.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: Aufgabenstruktur und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.), *TIMMS - Impulse für Schule und Unterricht* (S. 43–58). Bundesministerium für Bildung und Forschung.
- Klieme, E. & Tippelt, R. (2008). Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz. *Zeitschrift für Pädagogik Beiheft*, 53, 7–15.
- Klieme, E. & Vieluf, S. (2013). Schulische Bildung im internationalen Vergleich. Ein Rahmenmodell für Kontextanalysen in PISA. *Zeitschrift für Pädagogik Beiheft* 59, 229–246.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E. & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction* 19(6), 527–537.
- Nussbaum, M. C. (2012). *Konstruktion der Liebe, des Begehrens und der Fürsorge: Drei philosophische Aufsätze* ([Veränd. Nachdr.]). Reclams Universal-Bibliothek: Nr. 18189. Stuttgart: Reclam.
- OECD (1989). *Schulen und Qualität. Ein internationaler OECD-Bericht*. Unter Mitarbeit von John Lowe. Frankfurt am Main.
- OECD (2013). *PISA 2012 Assessment and Analytic Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Organisation for Economic Cooperation and Development; ebrary, Inc. Paris.
- Oser, F. (1998). *Ethos - die Vermenschlichung des Erfolgs. Zur Psychologie der Berufsmoral von Lehrpersonen*. Opladen: Leske + BudrichPietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft* 13(1), 121–148.
- Prengel, A. (2013). *Pädagogische Beziehungen zwischen Anerkennung, Verletzung und Ambivalenz*. Leverkusen: Barbara Budrich.

- Prengel, A.; & Winklhofer, U. (Hrsg.) (2014): *Kinderrechte in pädagogischen Beziehungen: Forschungszugänge*. Leverkusen: Barbara Budrich.
- Prenzel, M., Merkens, H. & Noack, P. (1999). Die Bildungsqualität von Schule: Fachliches und fächerübergreifendes Lernen im mathematisch-naturwissenschaftlichen Unterricht in Abhängigkeit von schulischen und außerschulischen Kontexten. Antrag an den Senat der DFG zur Einrichtung eines Schwerpunktprogramms. Kiel: IPN.
- Sauerwein, M. (2016). *Qualität von Ganztagsangeboten*. Phil. Dissertation, Frankfurt a.M.
- Scharenberg, K. (2015). Schulische Kontexteffekte im erziehungswissenschaftlichen Qualitäts- und Forschungsdiskurs. In: S. Hupka-Brunner; H.-U. Grunder; M. Bergmann & C. Imdorf (Hrsg.), *Qualität in der Bildung* (S. 51–74). Bad Heilbrunn: Klinkhardt.
- Scheerens, J. & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Spillebeen, L. & Willems, A.S (2014). Wie verändert sich die Schulqualität aus Sicht der Lehrkräfte? Zur Anwendung von latenten Mover-Stayer-Modellen. In K. Drossel, R. Strietholt & W. Bos (Hrsg.), *Empirische Bildungsforschung und evidenzbasierte Reformen im Bildungswesen* () Münster: Waxmann.
- Teddlie, C. & Reynolds, D. (2000). *The international handbook on school effectiveness research*. London: Falmer Press.
- Terhart, E. (2000). Qualität und Qualitätssicherung im Schulsystem. *Zeitschrift für Pädagogik* 46(6), 809.
- Tillmann, K.-J., Dedering, K., Kneuper, D., Kuhlmann, C. & Nessel, I. (2008). *PISA als bildungspolitisches Ereignis. Empirische Fallstudien in vier Bundesländern*. Wiesbaden: VS Verlag für Sozialwissenschaften
- Tröhler, D. (2010). Harmonizing the Educational Globe. World Polity, Cultural Features, and the Challenges to Educational Research. *Studies in Philosophy and Education* 29(1), 5–17.
- Tröhler, D. (2015), The medicalization of current educational research and its effects on education policy and school reforms. *Discourse: Studies in the Cultural Politics of Education* 36(5), 749–764.
- Von Friedeburg, L. (1989). *Bildungsreform in Deutschland. Geschichte und gesellschaftlicher Widerspruch*. Frankfurt: Suhrkamp.
- Wiere, A. (2011). Warum Ganztagschule? Rekonstruktion einer bildungspolitischen Kampagne. In H. Gängler & T. Markert (Hrsg.), *Visoion und Alltag der Ganztagschule. Die Ganztagschulbewegung als bildungspolitische Kampagne und regionale Praxis* (S. 13–32). Weinheim: Beltz Juventa.
- Zickgraf, A. & Zickgraf, P. (2009). Vier Milliarden Euro für einen gerechten Start ins Leben. In F. Prüß, S. Kortas & M. Schöpa (Hrsg.), *Die Ganztagschule: von der Theorie zur Praxis* (). Weinheim: Juventa.
- Willems, A.S & Becker, D. (2015). Ganztagschulen - Qualitätsmodelle, Potenziale und Herausforderungen für die Schulpraxis und die empirische Schul- und Unterrichtsforschung. In H. Wendt & W. Bos (Hrsg.), *Auf dem Weg zum Ganztagsgymnasium. Erste Ergebnisse der wissenschaftlichen Begleitforschung zum Projekt Ganz In* (S. 32–66). Münster: Waxmann.
- Zollondz, H.-D. (2010). *Grundlagen Qualitätsmanagement. Einführung in Geschichte, Begriffe, Systeme und Konzepte*. München: Oldenbourg.
- Züchner, I. & Fischer, N. (2011). Ganztagschulentwicklung und Ganztagschulforschung. In N. Fischer, H.G Holtappels, E. Klieme, T. Rauschenbach, L. Stecher & I. Züchner (Hrsg.), *Ganztagschule: Entwicklung, Qualität, Wirkungen. Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagschulen (StEG)* (S. 9–17). Weinheim: Beltz Juventa.

**Schlagworte:** Qualität, Normativität, Evaluation, Politikberatung

## Considérations à propos du concept de «qualité» dans le champ de la recherche en éducation

### Résumé

Cette contribution met en discussion le sens et l'usage du concept de «Qualité» dans la recherche en éducation, sur la base des recherches PISA et StEG (Study on the Development of German All-Day Schools). Deux modes de discours sur la qualité dans la recherche en éducation sont mis en évidence: le premier est empirique et analytique; le second descriptif. Indépendamment des modes d'argumentation choisis, la «Qualité» demeure connotée dans le domaine de la recherche en éducation. En effet, des présupposés normatifs sont toujours à la base des jugements qui se traduisent – sous différentes formes en fonction des modes de discours – dans l'évaluation de la qualité. Pourtant, la recherche en éducation est en mesure d'expliciter les présupposés normatifs des jugements de qualité et de prendre en considération de multiples critères et pour cette raison, elle n'est pas – comme souvent critiquée – soumise au politique.

**Mots-clés:** Qualité, recherche en éducation, politique de l'éducation

## Considerazioni sul concetto di qualità nella ricerca educativa

### Riassunto

Il contributo discute il significato e l'utilizzo del concetto di qualità nella ricerca educativa basandosi su due studi, PISA e StEG (Study on the Development of German All-Day Schools). Sono stati identificate due diverse modalità di discorso sulla qualità nella ricerca educativa: una modalità empirico-analitica ed una modalità descrittiva. Indipendentemente dalla modalità del discorso, la qualità nella ricerca educativa presenta una connotazione valutativa. Tale connotazione è sempre basata su presupposti normativi che influenzano – in modi diversi a seconda della modalità – la valutazione della qualità. La ricerca educativa è in grado di rendere esplicativi tali presupposti normativi e di considerare criteri multipli della valutazione della qualità e, per questo motivo, essa non è, come alcune critiche vogliono, alla mercé della politica.

**Parole chiave:** Qualità, ricerca in educazione, politica dell'educazione

## Some comments on the term «quality» in educational research

### Summary

The contribution discusses the meaning and use of quality as a concept in educational research based on two studies, PISA and StEG (Study on the Development of German All-Day Schools). Two different modes of discourses on quality in educational research are identified: an empiric-analytic mode, and a descriptive mode. Independent of the mode of discourse, quality in educational research is connoted with value judgements. These value judgements always are based on normative presuppositions; they affect – in different ways depending on the mode – the evaluation of quality. Educational research is in the position to make the normative presuppositions of quality evaluation explicit and to take into account multiple criteria. In doing so, it is not under politics' spell as is sometimes criticized.

**Keywords:** Quality, Normativity, Evaluation, Political Advice