

Unterrichtsqualität in der Geographie – Eine fachspezifische Sekundäranalyse der sächsischen Schulevaluation

Instructional Quality in Geography – A Subject-Specific Secondary Analysis of the Saxon School Evaluation

Calidad de la instrucción en Geografía: un análisis secundario de la evaluación de las escuelas sajonas

Natalie Bienert  , Nina Scholten , Rainer Mehren

Abstract Qualitativer (Fach-)Unterricht gilt als zentraler Prädiktor für den Lernerfolg von Schülerinnen und Schülern, wobei derzeit keine empirisch belastbaren Aussagen zur geographischen Unterrichtsqualität vorliegen. Ziel dieser quantitativen Sekundäranalyse ist die Erfassung des Status quo des Geographieunterrichts in seinen Qualitätsdimensionen sowie Einflussfaktoren zur erstmaligen Generierung empirischer Einblicke. Datengrundlage bilden 679 durch externe Raterinnen und Rater des Landesamtes für Schule und Bildung Sachsen (LASUB) mithilfe eines Beobachtungsbogens geratete Geographieunterrichtsstunden. Insgesamt fällt die geographische Unterrichtsqualität tendenziell positiv aus ($M=3,53$). Ferner wurden faktorenanalytisch sechs Subskalen geographischer Unterrichtsqualität gebildet: Klassenmanagement, Klarheit & Struktur, Konsolidierung, Schülerorientierung, inhaltliche Auseinandersetzung und Individualisierung des Lernens. Signifikante Unterschiede zeigen sich teils zwischen den Subskalen, Jahrgangsstufen und Schulformen.

Schlüsselwörter Unterrichtsqualität, Fachspezifität, Geographieunterricht, Status Quo-Analyse, Unterrichtsforschung

Abstract Instructional quality is supposed to predict students' learning success. Due to the lack of empirical evidence concerning instructional quality in geography, this study aims to capture the status quo of geography lessons in its dimensions as well as its influencing factors to generate first empirical insights. A data set containing 679 geography lessons comprise the basis of this secondary analysis. External raters evaluated the lessons using an observation sheet. Results indicate that, overall, the instructional quality of geography lessons is rather positive ($M=3.53$). A factor analysis shows six sub-scales of geographic instructional quality: classroom management, structure & clarity, consolidation, student-orientation, content-based involvement, and individualization of the learning process. Significant differences emerge between sub-scales, grades, and types of school.

Keywords instructional quality, subject specificity, geography lessons, status quo-analysis, instructional research

Resumen Se supone que la calidad de la instrucción predice el éxito del aprendizaje de los estudiantes. Debido a la falta de evidencia empírica sobre la calidad de la enseñanza en Geografía, este estudio tiene como objetivo capturar el status quo de las lecciones de Geografía en sus dimensiones, así como sus factores que influyen para generar unos primeros conocimientos empíricos. Un conjunto de datos que contiene 679 lecciones de Geografía constituye la base del presente análisis secundario. Los revisores externos evaluaron las lecciones utilizando una hoja de observación. Los resultados indican que, en general, la calidad de la enseñanza de las lecciones de Geografía es bastante positiva ($M=3,53$). Un análisis factorial muestra seis subescalas de calidad de la instrucción Geográfica: gestión del aula, estructura y claridad, consolidación, orientación de los estudiantes, participación basada en el contenido e individualización del proceso de aprendizaje. Surgen diferencias significativas entre subescalas, grados y tipos de escuela.

Palabras clave calidad de la instrucción, especificidad de la asignatura, lecciones de Geografía, análisis del status quo, investigación educativa

1. Anlass und Problemstellung

„Gelungene Lernprozesse in der Schule sind [...] hoch bedeutsam für die Lebensläufe der Schülerinnen und Schüler und die Gesellschaft insgesamt“ (DRECHSEL & SCHINDLER, 2019, S. 354). Qualitativ hochwertiger Unterricht stellt hierfür eine zentrale Gelingensbedingung dar, denn empirische Erkenntnisse weisen darauf hin, dass qualitativ vollere Unterricht zu größerem Lernerfolg führt (vgl. KUNTER & VOSS, 2011). Gleichzeitig macht Unterrichtsqualität einen der wichtigsten beeinflussbaren Faktoren auf Schülerleistungen aus (vgl. HATTIE, 2013). Die empirische Erfassung von Unterrichtsqualität ermöglicht nicht nur Aussagen zu deren Status Quo, sondern erlaubt auch eine differenzierte Identifikation zentraler Stellschrauben zur Qualitätssteigerung (vgl. KLIEME, 1999, in WÜSTEN, 2010). Im aktuellen Forschungsdiskurs wird vermehrt die Notwendigkeit einer fachspezifischen Erfassung von Unterrichtsqualität als Ergänzung zu deren bisher überwiegend generischen Erfassung von Unterrichtsqualität diskutiert (vgl. überblicksartig PRAETORIUS & GRÄSEL, 2021), welche sich in einer zunehmend fachspezifischen Adaption von Modellen wie dem der Basisdimensionen von Unterricht (vgl. KLIEME ET AL., 2006) manifestiert.

Auch in der Geographiedidaktik werden derzeit zwar erste theoretische Erweiterungen diskutiert (vgl. OHL & STREITBERGER, 2017; MEHREN & MEHREN, 2020; 2022; SCHOLTEN ET AL., 2023, in diesem Heft), allerdings steht die empirische Beforschung geographischer Unterrichtsqualität bislang noch weitgehend aus. Dabei entfaltet die Evidenzgenerierung zur Unterrichtsqualität für die Geographie eine besondere Relevanz aus einander ergänzenden Perspektiven:

(1) *Forschungsperspektive*: Zahlreiche andere Fachdidaktiken sind bereits in ILSAS (*International Large Scale Assessments*) wie PISA, TIMSS (*Trends in International Mathematics and Science Study*) oder im Ländervergleich des IQB (Institut zur Qualitätsentwicklung im Bildungswesen) repräsentiert (vgl. STANAT ET AL., 2022) und nehmen umfangreiche Studien wie die Mathematik mit COACTIV (*Cognitive Activation in the Classroom*, vgl. BAUMERT ET AL., 2010) oder die IPN Videostudie der Physik (vgl. SEIDEL ET AL., 2006) in Angriff, sodass eine grundlegende Datenlage zur jeweiligen Unterrichtsqualität existiert. Diese wird in den o.g. Studien orientiert am *outcome* als Leistungen der Schülerinnen und Schüler aufgefasst. Hingegen ist die Geographie in solch großskaligen Bildungsstudien bislang, trotz Bemühungen wie

der – gescheiterten – internationalen TIGAS-Initiative (*Trends in International Geography Assessment Study*, vgl. SOLEM ET AL., 2018), nicht vertreten. Eine empirische Beforschung geographischer Unterrichtsqualität bietet eine fachdidaktische Perspektive auf das Konstrukt und ist im Diskurs um Fachspezifik und Generik von Unterrichtsqualität zu verorten.

(2) *Unterrichtspraktische Perspektive*: Anhand von Unterrichtsqualitätsanalysen können gegenwärtige Stärken und Schwächen des Geographieunterrichts offengelegt werden. Nicht nur wird damit dem durch die Schulgeographinnen und Schulgeographen hervorgebrachten Einwand einer Defizitorientierung geographiedidaktischer Forschung begegnet (vgl. HEMMER ET AL., 2020), sondern es lassen sich vor allem Ansatzpunkte zur weiteren Verbesserung der Unterrichtsqualität in der Praxis ableiten. So können die durch die Schulgeographie eingeforderten Hilfestellungen (vgl. HEMMER ET AL., 2020) gegeben werden, wofür die Kenntnis des Status Quo unabdingbar ist.

(3) *Transferbezogene Perspektive*: „Aus Sicht der Praktiker [...] mutet geographiedidaktische Forschung häufig wenig bedeutsam für ihre Praxis an“ (HEMMER ET AL., 2020, S. 20). Der auch in der Roadmap-Studie von der Schulgeographie – und ebenfalls in den Naturwissenschaftsdidaktiken – erhobene Vorwurf der Praxisferne (FÖGELE ET AL., 2022) weist auf einen weiteren neuralgischen Punkt hin: Das fachliche Unterrichtsgeschehen wird *in vivo* bislang kaum beforscht. Weder liegen in ausreichendem Maße Erkenntnisse darüber vor, wie sich Geographieunterricht artikuliert, noch welche Qualität dieser aufweist. Entsprechend besteht eine Diskrepanz zwischen geographiedidaktischen Forschungsfeldern und der situierten Unterrichtspraxis. Die Ausrichtung des Forschungsinteresses auf den *tatsächlichen* Geographieunterricht in seiner Prozessualität kann daher zu der geforderten Relationierung der verschiedenen Phasen geographischer Lehrkräftebildung beitragen und auf die Praxis abgestimmte Forschungsfelder identifizieren.

2. Theoretischer und empirischer Forschungsstand

Die Beforschung von Unterricht steht in einer langen Tradition (u.a. LONG, 1980), wobei insbesondere der PISA-Schock eine Akzeleration der empirischen Zuwendung induzierte. Dabei entfaltet das Konstrukt „Unterrichtsqualität“ eine zentrale Rolle für bildungspolitische Qualitätssicherung, empirische Unterrichtsforschung und Fachdidaktiken gleichermaßen (vgl. LINDMEIER & HEINZE, 2020, S. 257).

2.1 Evidenzorientierung im Bildungswesen

Die Diskussion um die Erhebung von Unterrichtsqualität lässt sich im Kontext einer zunehmenden Evidenzorientierung im Bildungssektor verorten (LAMBRECHT & RÜRUP, 2012). Die Etablierung evidenzbasierter Governancestrukturen zur Gewährleistung schulischer Qualität involviert eine Vielzahl bildungspolitischer Steuerungsinstrumente wie externe Unterrichtsevaluationen, Schulinspektionen und Qualitätsanalysen (vgl. KOTTHOFF & BÖTTCHER, 2010; LAMBRECHT & RÜRUP, 2012). Zentrum des Forschungsinteresses bildet das Unterrichtsgeschehen, wofür als methodischer Zugriff häufig Videoanalysen oder Unterrichtsbeobachtungen mittels Beobachterratings genutzt werden. Als Unterform des *classroom assessments* fokussieren Unterrichtsbeobachtungen Fragestellungen nach dem Status Quo des Unterrichts in Bezug auf Stärken und Schwächen (vgl. CROSS, 1998). Zugrundeliegend ist die Annahme, dass die Kenntnis über den unterrichtlichen IST-Zustand entscheidend dafür ist, diesen über passgenaue und zielführende Maßnahmen weiter zu optimieren.

2.2 Konzeptionen von Unterrichtsqualität

Mitunter sind die Auffassungen davon, was *guten* Unterricht auszeichnet, divergent; so kann Unterrichtsqualität einerseits, im Sinne des *good teaching*, vor dem Hintergrund normativer Vorstellungen, andererseits im Sinne des *effective teaching* auf der Ebene von Effekt- bzw. Wirkungsmessungen aufgefasst werden (vgl. BERLINER, 2005). Im letzteren Verständnis gilt Unterricht dann als gut, wenn er lernwirksam ist (vgl. HELMKE, 2015). Die Erfassung von Unterrichtsqualität anhand empirisch fundierter Dimensionen im Sinne des *effective teaching* hat gegenüber dem *good teaching* den Vorteil, normative Vorstellungen gering zu halten (vgl. LINDMEIER & HEINZE, 2020).

Dem Ansatz des *effective teaching* folgend wird Unterrichtsqualität verstanden als „[...] Bündel von Merkmalen, die sich als Prozessqualität positiv auf

die Produktqualität auswirken“ (EINSIEDLER, 2002, S. 195). Ergänzend definieren WEINERT ET AL. (1989, S. 899) Unterrichtsqualität, auf das *Educational Effectiveness*-Paradigma rekurrierend (vgl. SEIDEL & SHAVELSON, 2007), als „[...] jedes stabile Muster von Instruktionsverhalten, das als Ganzes oder durch einzelne Komponenten die substantielle Vorhersage und/oder Erklärung von Schulleistung erlaubt“. Unterrichtsqualität wird demnach funktional im Sinne der Effektivität verstanden, lässt sich bezüglich der prozess- bzw. produktbezogenen Ausrichtung differenzieren und wird anhand von Kriterien gefasst.

2.3 Sicht- und Tiefenstrukturen

Im Diskurs ist eine Unterscheidung zwischen Sicht- und Tiefenstrukturen des Unterrichts etabliert. Für die Lernwirksamkeit von Unterricht sind insbesondere die Tiefenstrukturen wesentlich (vgl. OSER & BAERISWYL, 2001; KUNTER & TRAUTWEIN, 2013). Während Sichtstrukturen sich auf organisatorische Elemente des Unterrichts beziehen und direkt beobachtet werden können (z.B. Sozialform und Methode), umfassen Tiefenstrukturen die Qualität der inhaltlichen Auseinandersetzung (z.B. kognitive Aktivierung). Letztere sind oft nur durch Interpretation zugänglich. Dabei sind Oberflächen- und Tiefenstrukturen größtenteils unabhängig voneinander. So bestimmt beispielsweise die Arbeit allein oder in Gruppen (Sichtstruktur) nicht die kognitive Aktivierung der Schülerinnen und Schüler (Tiefenstruktur) (vgl. KUNTER & TRAUTWEIN, 2013).

2.4 Basisdimensionen der Unterrichtsqualität

Strukturell haben sich in der deutschsprachigen Bildungsforschung die drei Basisdimensionen als grundlegende Unterrichtsqualitätsmerkmale erwiesen (vgl. KLIEME & RAKOCZY 2008), welche anhand empirischer Zugriffe (u.a. Videoanalysen, Schüler- und Lehrkräftebefragungen) als relevante Einflussfaktoren identifiziert wurden (vgl. SCHIEPE-TISKA ET AL., 2016). Es wird zwischen den Dimensionen *effektive Klassenführung*, *konstruktive Unterstützung* und *kognitive Aktivierung* unterschieden. Dabei umfasst *effektive Klassenführung* eine störungspräventive Unterrichtsführung, *konstruktive Unterstützung* ein unterstützendes und schülerorientiertes Sozialklima und *kognitive Aktivierung* das kognitive Anregungspotential der Schülerinnen und Schüler (vgl. KLIEME & RAKOCZY, 2008). Das Modell postuliert Gültigkeit über alle Klassen, Schulformen und Fächer hinweg.

2.5 Generik und Spezifik von Unterrichtsqualität

Wenngleich das Modell der Basisdimensionen dementsprechend als generisches Modell gilt, wird zunehmend die Fachspezifik einzelner (Sub-)Dimensionen diskutiert. Beispielsweise führen KLIEME und RAKOCZY (2008) an, dass kognitive Aktivierung eine fachspezifische Komponente aufweist. In einigen Fächern wird demnach die Notwendigkeit einer fachspezifischen Ausdifferenzierung des Modells der Basisdimensionen bereits verfolgt (vgl. PRAETORIUS & GRÄSEL, 2021). Eine fachspezifische Erfassung von Unterrichtsqualität ist notwendig, da erstens die Befundlage zur prädiktiven Validität des Modells derzeit inkonsistent ist (vgl. PRAETORIUS ET AL., 2020). Zweitens zeigt eine Metaanalyse von SEIDEL und SHAVELSON (2007), dass sich insbesondere die fachspezifischen Merkmale als die erklärungsstärksten Komponenten für den Nachweis der Effektivität von Lehr- und Lernprozessen darstellten. Drittens scheint es nicht plausibel, dass die fachliche Richtigkeit der Ausführungen der Lehrkraft (=Vermeidung fachlicher Fehler) im Modell unberücksichtigt bleibt, obwohl dieser Faktor aus fachdidaktischer Sicht eine entscheidende Rolle für effektive Lernprozesse einnimmt (BRUNNER, 2018). Eine ausführliche Auseinandersetzung bezüglich Fachspezifik und Generik von Unterrichtsqualität in der Geographiedidaktik nehmen SCHOLTEN ET AL. (2023, in diesem Heft) vor.

2.6 Unterrichtsqualität in der Geographiedidaktik

Insgesamt ist ein weitgehend durch normative Vorstellungen und konzeptioneller Entwicklungen geprägter fachlicher Diskurs bezüglich der Frage nach der Qualität von Geographieunterricht (*good teaching*) zu konstatieren:

(1) *Konzeptionelle Überlegungen*: Bislang lassen sich vor allem Bemühungen der theoretisch-konzeptionellen Fassung von Unterrichtsqualität ausmachen. Erste Impulse liefern die Beiträge MEHREN und MEHRENS (2020; 2022) sowie SCHOLTEN ET AL. (2023, in diesem Heft), in welchen eine geographiespezifische Erweiterung des Modells der Basisdimensionen von Unterricht um die Dimensionen *Fachliche Strukturiertheit*, *Gegenstandsbezogene Qualität* und *Adaptiv-inhaltliche Unterstützung* theoretisch hergeleitet wird. Auch OHL und STREITBERGER (2017) bieten mit dem Augsburger Analyse- und Evaluationsraster für Bildungsmedien eine geographiespezifische Modellierung an, die im Vergleich zum abstrakteren Modell von MEHREN und MEHREN (2020)

spezifischer ausfällt (vgl. SCHOLTEN ET AL. 2023, in diesem Heft).

(2) *Normativer Zugriff*: Auch im Sinne des *good teaching* existieren vielfältige Vorstellungen, wie sie sich implizit u.a. in geographiedidaktischen Einführungswerken äußern (u.a. KANWISCHER, 2013; REINFRIED & HAUBRICH, 2015; RINSCHDE & SIEGMUND, 2019). Im Interviewband REMPLERS (2018) wurden zudem Geographiedidaktikerinnen und Geographiedidaktiker bezüglich ihrer Auffassung von lernwirksamem Geographieunterricht befragt. Dabei zeigt sich eine große Vielfalt unterschiedlicher Sichtweisen (vgl. hierzu auch SCHOLTEN ET AL., 2023). Einen empirischen Zugriff auf normative Geographieunterrichtsvorstellungen offeriert THUME (2023, 314), indem sie anhand vier empirisch identifizierter Perspektiven auf Geographieunterricht (konzeptionell-abstrahierend, verknüpfend-mündig, alltäglich-handelnd und interessant-bewältigend) aufzeigt, wie heterogen die Auffassungen bezüglich *gutem* Geographieunterricht ausfallen.

(3) *Empirische Erhebungen*: Während die explizite Beforschung fachspezifischer Unterrichtsqualität insbesondere in den Naturwissenschaften didaktiken (vgl. auch Kap. 1) bereits etabliert ist, ist dies in der Geographiedidaktik bislang kaum der Fall. Auch wenn insgesamt nur wenige geographiedidaktische Feldstudien existieren, die die Qualität des Unterrichtsgeschehens selbst beforschen, lassen sich doch einzelne Studien identifizieren, die sich unterschiedlichen Facetten geographischer Lehr-Lern-Prozessen widmen und thematisch fokussierte Einblicke in den Status Quo des gegenwärtigen Geographieunterrichts erlauben. Diese Ausnahmen, ausgewählt nach deren methodischer Anlage als Unterrichtsbeobachtungen, werden im Folgenden skizziert:

(a) *Argumentieren*: In einer Studie BUDKES (2021) wurden anhand strukturierter Unterrichtsbeobachtungen, die während Schulpraktika von hospitierenden Studentinnen und Studenten erhoben wurden, mittels Beobachtungsbögen die Schreibprozesse in 178 Geographiestunden in Real- und Hauptschulen untersucht. Die Befunde zeigen, dass in 78% der beobachteten Stunden Schreibprozesse stattfanden, das Anforderungsniveau aber niedrig ist (BUDKE, 2021).

(b) *Postkolonialismus*: In ihrer qualitativen Dissertationsstudie untersuchte SCHWARZE (2021) u.a. anhand von in-vivo-Unterrichtsbeobachtungen, welche Bilder von Subsahara-Afrika konstruiert werden. Mittels eines inhaltsanalytischen Vorgehens identifiziert sie vier Narrative: Subsahara-Afrika als

Problemraum, unterstützungsbedürftiger Raum, Faszinationsraum und Raum einer Agrargesellschaft. Demnach besteht ein Problemfokus im Unterricht, dem latente postkoloniale Denkstrukturen zugrunde liegen. Entsprechend erfolgt der tatsächliche unterrichtliche Zugriff auf Subsahara-Afrika größtenteils unreflektiert und undifferenziert.

(c) *Forschend-entdeckendes Lernen*: Ferner untersucht ein derzeit laufendes Dissertationsprojekt die Unterrichtsqualität im österreichischen Sachunterricht in Hinblick auf die Lernwirksamkeit forschend-entdeckender Lerngelegenheiten (NEUBAUER, 2022). In einer auf dem TIMSS 2019-Datensatz basierenden Sekundäranalyse bezüglich der Einsatzhäufigkeit spezifischer Unterrichtsmerkmale im naturwissenschaftlichen Sachunterricht wurden in einer explorativen Faktorenanalyse vier Skalen herausgestellt: involvierend aktiv-forschender Unterricht, kognitiv aktivierender forschender Unterricht, initiiert lehrkraftzentrierter moderierender Unterricht und materialbasierter, lebensweltbezogener Unterricht. (d) *Unterrichtsmethoden*: Im internationalen Kontext beschäftigt sich eine Studie von DOMAČINOVIĆ und VUK (2022) mit der Häufigkeit und Frequenz von Unterrichtsmethoden im kroatischen Geographieunterricht in der Primarstufe. Mittels eines Selbstbeobachtungsbogens sowie anschließenden strukturierten Interviews wurden 32 Unterrichts-

stunden in Hinblick auf den Einsatz von Unterrichtsmethoden durch Lehrkräfte selbst beobachtet. Die Ergebnisse zeigen eine Dominanz der Unterrichtsdiskussion, gefolgt vom indirekten Einsatz von Grafiken (=lehrkraftseitiges Zeigen von Fotos, Karten, Diagrammen) und Textarbeit. Selten kamen Schülervorträge und der direkte Einsatz von Grafiken (=schülerseitige Produktion von Karten, Diagrammen, Zeichnungen) vor.

Den aufgeführten Studien ist gemein, dass sich in der Tendenz größere Defizite bezüglich der (inhaltlichen) unterrichtlichen Auseinandersetzung zeigen, sodass eine prägnante Theorie-Praxis-Diskrepanz besteht. Weiterhin lässt sich aufgrund der fehlenden thematischen Fokussierung kein kohärentes Gesamtbild darüber, wie sich Qualität im Geographieunterricht artikuliert, ausmachen. Ferner ist festzustellen, dass es sich bei allen Studien um Querschnitterhebungen zu einzelnen Zeitpunkten handelt, während längsschnittliche (Prozess-)Untersuchungen, die Aussagen über die Entwicklung zuließen, bislang ausstehen. Auch finden bislang kaum Quantifizierungen im Sinne der Messung der Unterrichtsqualität nach large-scale-Erfassungen statt. Entsprechend ist ein umfassendes Forschungsdesideratum zum Status Quo des Geographieunterrichts und dessen Entwicklung festzustellen.

3. Forschungsfragen und Zielsetzung

Der nur in Teilen existente und wenig kohärente geographische Forschungsstand bedingt den explorativen Charakter dieser quantitativen Studie und begründet ein forschungsfragengeleitetes anstelle eines hypothesenüberprüfenden Vorgehens. Wie herausgestellt wurde, liegen trotz der großen Relevanz der Kenntnis des IST-Zustandes der Unterrichtsqualität bislang kaum geographiedidaktische Evidenzen vor (vgl. Kap. 2). Daraus ergibt sich die Hauptforschungsfrage:

HF 1: Wie ist die Qualität des Geographieunterrichts ausgeprägt?

Da es sich bei Unterrichtsqualität um ein multidimensionales Konstrukt handelt, wird dieses im Forschungsdiskurs (vgl. Kap. 2) in der Regel strukturell untergliedert. Zwecks Erlangung differenzierterer Einblicke wird nach empiriebasierten Teilaspekten geographischer Unterrichtsqualität gefragt:

TF 1.1: Wie artikulieren sich die unterschiedlichen Facetten der Qualität des Geographieunterrichts?

In der empirischen Bildungsforschung werden zudem weitere Einflussfaktoren auf Unterrichtsqualität ausgewiesen, deren Wirkungen auf den Geographieunterricht bisher noch kaum bekannt sind:

TF 1.2: Wie wirken die Einflussfaktoren Schulform, Jahrgangsstufe und Erhebungszeitpunkt auf die geographische Unterrichtsqualität?

Zielsetzung dieser Studie bildet demnach erstens die Generierung empirischer Erkenntnisse zum Status Quo der Qualität des Geographieunterrichts in ihren verschiedenen Facetten mittels einer large-scale-Erfassung anhand eines Datensatzes aus Sachsen, der durch in-vivo-Unterrichtsbeobachtungen durch externe Raterinnen und Rater erhoben wurde. Auf Basis dieser Sekundäranalyse werden zweitens Stärken und Schwächen zur Optimierung geographischer Bildungsprozesse identifiziert sowie drittens weitere Forschungsdesiderata abgeleitet.

4. Untersuchungsdesign

Die vorliegende Studie stellt eine fachspezifische Sekundäranalyse eines Datensatzes des sächsischen Landesamts für Schule und Bildung (LASUB) vor. Das ursprüngliche Interesse des LASUB war, anhand einer externen, fächerübergreifenden Schulevaluation „Schulen im Hinblick auf die Qualität der Arbeit und der erreichten Ergebnisse“ zu bewerten, um „Ergebnisse der schulischen Qualitätsanstrengungen zu ermitteln und [...] die weiteren schulischen Qualitätsmaßnahmen zu planen“ (LASUB, 2020, o. S.). Der Datensatz, der den Geographieunterricht betrifft, wurden den Autorinnen und dem Autor durch das LASUB in standardisierten Excel-Formularen zur fachspezifischen Auswertung zur Verfügung gestellt. Zur Erhöhung der Nachvollziehbarkeit ist das Kapitel in die Kontextualisierung der Primärerhebung durch das LASUB (Kap. 4.1) und die durch die Autorinnen und den Autor vorgenommene, fachspezifische Sekundäranalyse (Kap. 4.2) unterteilt.

4.1. Kontextualisierung der Primärerhebung

Sample und Samplingverfahren

Der im Rahmen der externen Schulevaluation erhobene Datensatz basiert auf in-vivo-Beobachtungen von $n=679$ Geographieunterrichtsstunden in Sachsen. Das sächsische Schulwesen zeichnet sich durch ein hohes Bildungsniveau aus: das Bundesland belegt im IQB-Ländervergleich in den Bereichen Lesen, Zuhören und Mathematik Spitzenpositionen (STANAT ET AL., 2022). Das Fach Geographie wird in Sachsen am Gymnasium in den Klassenstufen 5-12, an Mittel-/Oberschulen in den Klassenstufen 5-10 als eigenständiges Fach konsekutiv unterrichtet. Mittel-/Oberschulen entsprechen teilentegrativen Gesamtschulen ohne gymnasialen Zweig, an denen Haupt- und Realschulabschlüsse erworben werden können. Insgesamt ist das Fach in Sachsen im Bundeslandvergleich bezüglich Durchgängigkeit und Eigenständigkeit verhältnismäßig gut aufgestellt (vgl. FÖGELE ET AL., 2022).

Die Primärerhebung fand in den Schuljahren 2006/2007 bis 2015/2016 an Gymnasien ($n=195$), Mittel-/Oberschulen ($n=395$), Förderschulen ($n=88$) und einer berufsbildenden Schule ($n=1$) statt. Die Teilnahme der Schulen war im Schuljahr 2006/2007 freiwillig, in darauffolgenden Schuljahren (ab 2007/2008) obligatorisch. Die Auswahl der Schulen erfolgte ab 2007 per Losverfahren als randomisiertes Samplingverfahren und kann als repräsentative Auswahl für Sachsen gelten. Während die Schulbesuche vorher angekündigt wurden

und dem Lehrpersonal entsprechend bekannt waren, erfolgte die konkrete Auswahl der beobachteten Unterrichtsstunden per Zufallsprinzip erst vor Ort. Demnach bestand vorab Unkenntnis darüber, welche Stunde evaluiert wird. Entsprechend sind auch die Klassen zufällig ausgewählt (=keine längsschnittliche Begleitung derselben Klasse über den Untersuchungszeitraum hinweg).

Externe Raterinnen und Rater

Bei den externen Raterinnen und Ratern handelte es sich um abgeordnete Lehrkräfte und Schulleitungen aller Schulformen und Fächer, die systematisch in Bezug zu theoretischen Grundlagen sowie praktischer Handhabung des Erhebungsinstruments sowohl vor Erhebungsbeginn als auch erhebungsbegleitend intensiv geschult wurden. In der Schulung wurden u.a. Videoanalysen von Unterrichtsstunden durchgeführt, die die Raterinnen und Rater zunächst unabhängig voneinander mittels des Beobachtungsbogens beurteilten und anschließend kommunikativ validierten. In der begleitenden systematischen Evaluation der Videobeobachtungsergebnisse der Raterinnen und Rater zeigten sich hohe Korrelationen, sodass insgesamt gute bis akzeptable Reliabilitätswerte erreicht wurden. Zwar sind Ratereffekte nicht auszuschließen (vgl. HOYT, 2000), allerdings konnten durch dieses Vorgehen die in der Empirie häufig angeführten Probleme in der Reliabilität bei Unterrichtsbeobachtungen weitgehend vermieden werden (vgl. PRAETORIUS, 2013). Zur Gewährleistung der Objektivität wurden die externen Raterinnen und Rater nicht an der eigenen Schule zur Evaluation eingesetzt. Da die Zielsetzung der externen Schulevaluation nicht die Bewertung fachinhaltlicher, sondern generischer Aspekte der Unterrichtsqualität umfasste, wurde die Fakultas der Raterinnen und Rater nicht systematisch in den Primärdaten erfasst. Entsprechend waren diese, wie in Unterrichtsbeobachtungen im Rahmen von Schulevaluationen üblich, wenn dann nur zufällig Fachlehrkräfte der Geographie.

Erhebungsinstrument

Mithilfe eines standardisierten, grundsätzlich zur generischen Erfassung von Unterrichtsqualität ausgelegten Unterrichtsbeobachtungsbogens des LASUB bewerteten externe Raterinnen und Rater (s.u.) den Geographieunterricht auf einer fünfstufigen Likert-Skala (1=trifft nicht zu, 5=trifft vollkommen zu). Unterrichtsqualität wurde in diesem durch das LASUB auf Grundlage extensiver Litera-

turreviews mittels 18 Items operationalisiert (vgl. Fig. 1) und anhand von vier literaturbasierten Dimensionen (Aufrechterhalten von Aufmerksamkeit, Förderung von Verstehen, Förderung von Anwendungsbezug, Förderung intrinsischer Motivation) gruppiert. Dem Erkenntnisinteresse der Primärerhebung folgend ist der Beobachtungsbogen auf die überfachliche Erfassung von Unterrichtsqualität ausgerichtet. Entsprechend umfasst das Instrument neben Items, die die pädagogisch-organisatorische Qualität von Unterricht berücksichtigen (z.B. Klassenführung) auch Items, die sich auf Lehr- und Lernprozesse beziehen, damit eine gegenstandsbezogene Qualität berücksichtigen und entsprechend stärker fachspezifisch aufgeladen sind (z.B. Vernetzung). Um den interpretativen Charakter des Beobachtungsbogens zu reduzieren, wurde jedes Item anhand mehrerer Indikatoren als Hilfestellung für die externen Raterinnen und Rater konkretisiert. Da die Terminologie der aufgeführten Items teils anders aufgefasst wird als im geographiedidaktischen Fachdiskurs (z.B. Stimulierung: „SuS werden zur kritischen Auseinandersetzung mit den Unterrichtsinhalten angeregt“, LASUB, 2020), bedarf es einer Verständnisklärung. Aus Platzgründen wird die Itemoperationalisierung exemplarisch in Fig. 1 im Ergebnisteil präsentiert. Das gesamte Erhebungsinstrument kann bei den Autorinnen und dem Autor angefragt werden. Zur Gewährleistung der Vergleichbarkeit wurde derselbe Beobachtungsbogen bei allen Schulformen, Jahrgangsstufen und Fächern eingesetzt. Die Validierung des Erhebungsinstruments erfolgte anhand von Expertendiskussionen (u.a. Raterinnen und Rater, Forscherinnen und Forscher), in welchen die Validität diskursiv geprüft wurde.

Externe Einflussfaktoren

Weiterhin wurden die Variablen Schulform, Jahrgangsstufe und Erhebungszeitpunkt erfasst, die auch in der Forschungsliteratur als äußere Einflussfaktoren auf Unterrichtsqualität empirisch ausgewiesen werden.

(a) *Schulform*: Ergebnisse der COACTIV-Studie weisen auf Differenzen zwischen gymnasialen und nicht-gymnasialen Schulformen bezüglich der Schülerleistung in Mathematik, Lesen und allgemeiner kognitiver Grundfertigkeiten in der 9./10. Klasse hin (vgl. BAUMERT ET AL., 2010). Dazu anschlussfähig sind die Befunde einer Studie von KUGER ET AL. (2017), nach welcher die Schulform (gymnasial vs. nicht-gymnasial) einen starken Prädiktor für den Stand der Schülerleistungen im Mathematikunterricht der 9. Klasse bildet. Mit Blick auf die Dimensionen von Unterricht zeigen KUNTER ET AL. (2005) anhand einer

Schüler- und Lehrkräftebefragung, dass die kognitive Aktivierung im Mathematikunterricht in Gymnasien signifikant höher ist als in Hauptschulen. (b) *Jahrgangsstufe*: In Bezug auf die Jahrgangsstufe konnte u.a. in der amerikanischen MET-Studie (*Measures of Effective Teaching Project*, KANE ET AL., 2015) für die Fächer Mathematik und Englisch ein Rückgang der Unterrichtsqualität zwischen den Klassenstufen 4-8 gezeigt werden (MIHALY & MCCAFFREY, 2015). (c) *Erhebungszeitpunkt*: In Ermangelung des Vorliegens längsschnittlicher Daten zur zeitlichen Veränderung der Unterrichtsqualität lassen sich Befunde der PISA-Studien zu den Leistungstrends deutschen Schülerinnen und Schüler heranziehen (OECD, 2019): Alle drei Dimensionen weisen Zugewinne bis 2012 vor, denen seither stetige Rückgänge folgen, welche für Mathematik und Naturwissenschaften gar als „umfassende negative Trends“ betitelt werden (OECD, 2019, S. 4). Zurückzuführen sind diese Trends teils auf ein verändertes demographisches Profil der Schülerinnen und Schüler.

Erhebungsdurchführung

Die beobachteten Unterrichtsstunden umfassten 45 Minuten, sodass der gesamte Unterrichtsprozess Gegenstand einer Urteilsbildung zu jedem der Indikatoren war. Wurde bei den ersten Schulbesuchen zunächst die Unterrichtsstunde durch zwei-drei Raterinnen und Rater gemeinsam beobachtet und deren Urteile anschließend verglichen, kam bei den anschließenden Unterrichtsbeobachtungen grundsätzlich jeweils nur eine Raterin bzw. ein Rater zum Einsatz.

4.2. Geographiespezifische Sekundäranalyse

Die sekundäranalytische Nutzung von Daten ist ein etabliertes Standardvorgehen in der Bildungsforschung und umfasst die erneute Auswertung eines Datensatzes in Hinblick auf eine neue Fragestellung und/oder verbesserte Methodik (vgl. DÖRING & BORTZ, 2016). Besondere Potentiale ergeben sich dabei in der sekundäranalytischen Auswertung von Schulinspektionsdaten. Schulinspektionen sind institutionell verankert, erfolgen kriteriengestützt und standardisiert (vgl. KOTTHOFF & BÖTTCHER, 2010). Im Fokus steht die wissenschaftsbasierte, „[...] kriteriengeleitete Feststellung und Rückmeldung von Schulqualität auf der Basis verlässlicher Befunde und Forschungsverfahren“ (LAMBRECHT & RÜRUP, 2012, S. 59). Resultat ist die Akkumulation von mit hohem Aufwand erhobenen empirischer Daten über Schul- und Unterrichtsqualität, die nur selten fachsystematisch ausgewertet werden

(LAMBRECHT & RÜRUP, 2012, S. 57). Auch wenn die Daten der sächsischen Schulevaluation mit einer anderen primären Funktion erhoben wurden (s.o.) und u.a. bezüglich des Umfangs nicht vergleichbar mit ILSAS sind, bieten sie große – bislang für das Fach Geographie weitgehend ungenutzte – Potentiale zur Status Quo-Analyse von Unterricht im Sinne einer „Innenperspektive“ und erlauben eine empirisch untermauerte Ableitung erster Tendenzen geographischer Unterrichtsqualität.

Dabei wird das für Sekundäranalysen zentrale Kriterium der Passung zwischen existierendem Datensatz und neuem Forschungsinteresse (JOHNSTON, 2017) insofern erfüllt, da das sekundäranalytische Forschungsinteresse (Analyse geographischer Unterrichtsqualität) eine Konkretisierung des ursprünglichen Forschungsinteresses (Analyse allgemeiner Unterrichtsqualität) darstellt. Kritisch zu betrachten ist jedoch, dass eine genuin fach- bzw. geographiespezifische Herangehensweise ggf. eine andere Operationalisierung des Konstrukts Unterrichtsqualität im Erhebungsinstrument sowie eine andere, fachspezifische Qualifizierung der Raterinnen und Rater erfordern würde, insbesondere in Anbetracht der aktuellen fachdidaktischen Diskussion um Generik und Fachspezifik von Unterrichtsqualität (vertiefend hierzu s. SCHOLTEN ET AL., 2023, in diesem Heft). Somit kann die vorliegende Sekundäranalyse zwar nicht das Zusammenspiel generischer und fachspezifischer Merkmale der Unterrichtsqualität darstellen, wohl aber als Annäherung an die Qualität des Geographieunterrichts in ihrer generischen Betrachtung dienen. Entsprechend ist das sekundäranalytische Forschungsinteresse auch als genuin neu zu bewerten, da der Datensatz bislang nicht geographiespezifisch ausgewertet wurde.

Die Beschaffung des Datensatzes erfolgte auf Anfrage der Autorinnen und des Autors, welchen das sächsische Landesamt für Schule und Bildung für das Vorhaben dieser Sekundäranalyse als einziges aller angefragten 16 Bundesländer zur Verfügung stellte.

5. Ergebnisdarstellung

5.1. Ausgewählte Ergebnisse

Auf der Gesamtskala beziffert sich die Qualität des Geographieunterrichts über alle Schulformen, Erhebungszeitpunkte sowie Jahrgangsstufen hinweg ($n=679$) im Mittelwert auf 3,53 ($SD=0,63$) und liegt damit über dem arithmetischen Mittel von 3. Einen differenzierteren Einblick gewährt das Ranking auf Einzelitemebene (vgl. Fig. 1).

Die Top 3-Items (Wertschätzendes Lehrkraftverhalten, Klassenführung und Klarheit) weisen Mittelwerte >4 auf. Dabei werden diese Items stark über das Steuerungsverhalten der Lehrkraft operationalisiert.

Sekundäranalytisches Auswertungsverfahren

Im Fokus der Sekundäranalyse steht, den vorliegenden generischen Datensatz für das Fach Geographie auszuwerten. In der quantitativen Sekundärauswertung kamen sowohl deskriptive als auch inferenzstatistische Verfahren zum Einsatz. Zunächst erfolgte eine Berechnung der Mittelwerte und Standardabweichungen auf Gesamtskalen- sowie auf Einzelitemebene. Letztere wurden dann in eine Rangreihenfolge gebracht. Eine anschließende explorative Faktorenanalyse ergab vier Subskalen, die jedoch nicht die theoretisch angenommenen Dimensionen des Erhebungsinstrumentes (Aufrechterhalten von Aufmerksamkeit, Förderung von Verstehen, Förderung von Anwendungsbezug, Förderung intrinsischer Motivation) replizieren und sich einer alternativen inhaltlichen Interpretation entzogen. Deshalb wurde das Verfahren gedreht, indem die Autorinnen und der Autor eigene Subskalen bildeten. In einer anschließenden konfirmatorischen Faktorenanalyse bestätigten sich diese sechs didaktisch sinnvollen Subskalen (vgl. Fig. 1). Aufgrund der sekundäranalytischen Anlage, die keine Beeinflussung des Erhebungsinstrumentes zulässt, treten für vereinzelte Items (z.B. Interesse wecken) nicht zu vermeidende Doppelladungen auf. Daher erfolgte die Faktorenbildung nicht ausschließlich anhand statistischer Kennwerte, da als weiteres Kriterium der didaktische Erklärungswert im Sinne einer inhaltlichen Stichhaltigkeit angelegt wurde. Die Subskalen weisen trotz des teils kleinen Skalenumfanges akzeptable bis gute Reliabilitätswerte (Cronbachs α) auf (vgl. Fig. 1). Weiterhin wurden die Einflussfaktoren Erhebungszeitpunkt, Jahrgangsstufe und Schulform hinsichtlich ihres Einflusses auf geographische Unterrichtsqualität ausgewertet.

Die letztplatzierten Items hingegen wurden mit Werten zwischen 2,39 und 3,13 beurteilt und beziehen sich in ihrer Operationalisierung stärker auf die Unterstützung des schülerseitigen Lernprozesses. Einzig die Items *Stimulierung* und *Differenzierung* liegen unterhalb des skalentheoretischen Mittelwertes.

5.2. Subskalen geographischer Unterrichtsqualität

Zur weiteren Differenzierung wurden faktorenanalytisch sechs Subskalen gebildet (vgl. Fig. 2):

- (a) *Klassenmanagement*: SK1 umfasst Items (Wertschätzendes Lehrkraftverhalten, Klassen-

Platz	Item	M	SD	SK
1	Wertschätzendes Lehrkraftverhalten (z.B. Anerkennung)	4,25	0,86	SK1
2	Klassenführung (z.B. feste Verhaltensregeln der Schülerinnen und Schüler)	4,21	0,88	SK1
3	Klarheit (z.B. präzise Formulierungen)	4,18	0,85	SK2
4	Strukturiertheit (z.B. verschiedene Darstellungsformen)	4,13	0,88	SK2
5	Schülerbeteiligung (z.B. aktive Mitwirkung im Unterricht)	4,07	0,86	SK4
6	Engagement der Lehrkraft (z.B. Begeisterungsfähigkeit)	4,03	0,94	SK1
7	Vernetzung (z.B. Erkennen von Zusammenhängen)	3,79	0,92	SK5
8	Festigung (z.B. Wiederholung von Inhalten & Themen)	3,77	1,00	SK3
9	Variabilität (z.B. Variation von Sozialformen & Methoden)	3,35	0,97	SK4
10	Überprüfung (z.B. Anwendung & Überprüfung von Wissen)	3,49	0,98	SK3
11	Interesse wecken (z.B. Wecken der Neugier auf Unbekanntes)	3,21	1,01	SK4
12	Selbstkonzeptstärkung (z.B. entwicklungsfördernde Kritik)	3,20	0,96	SK6
13	Kritisches Prüfen (z.B. kritische inhaltliche Auseinandersetzung)	3,14	1,12	SK5
14	Anwendungsbezug (z.B. multiperspektivische Problembetrachtung)	3,14	1,14	SK5
15	Flexibilität (z.B. Integration von Interessen der S*S)	3,13	1,08	SK4
16	Autonomieunterstützung (z.B. eigenverantw. Aufgabenlösung)	3,13	1,00	SK6
17	Stimulierung (z.B. Anregung zur kritischen Auseinandersetzung)	2,92	1,00	SK5
18	Differenzierung (z.B. versch. Schwierigkeitsgrade von Aufgaben)	2,39	0,84	SK6
	Gesamtskala	3,53	0,63	

Fig. 1. Ranking der Einzelitems unter Angabe der exemplarischen Indikatoren (kursiv), Mittelwerte (M), Standardabweichungen (SD), und Subskalenzugehörigkeit (SK) ($n=679$). Die Beispiele sind aus dem Erhebungsinstrument entnommen und wurden zur Darstellung z.T. gekürzt. Das komplette Messinstrument mit allen Indikatoren für jedes Item kann beim Autorenteam angefragt werden (Quelle: Autorinnen und Autor)

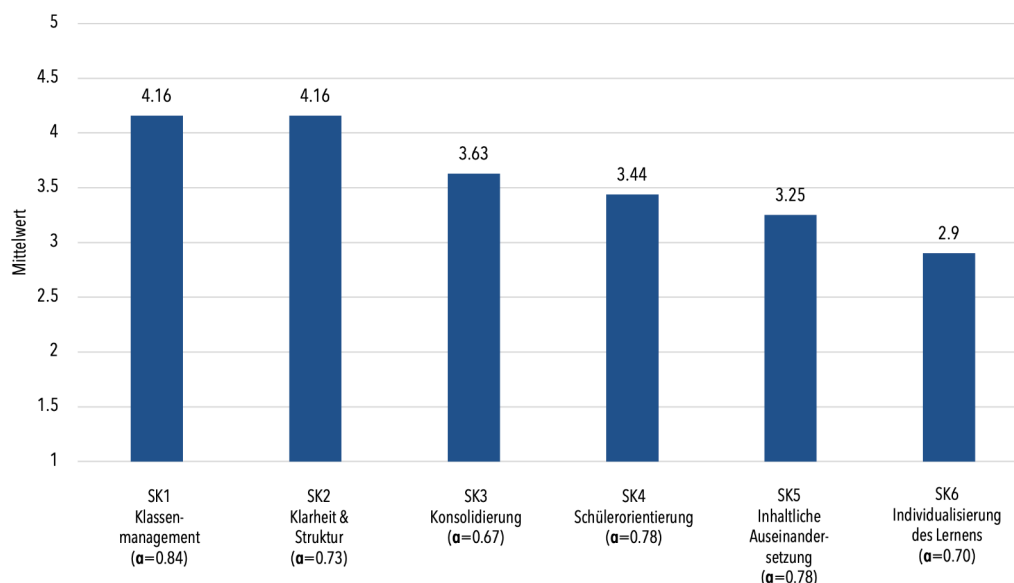


Fig. 2. Sechs Subskalen geographischer Unterrichtsqualität in der Praxis auf Basis $n=679$ Geographieunterrichtsstunden, mit Reliabilitätswerten (Cronbachs α) (Quelle: Autorinnen und Autor)

führung, Engagement der Lehrkraft), die auf lehrkräftsbezogene Steuerung der Unterrichtsorganisation rekurrieren.

(b) *Klarheit & Struktur*: Auch SK2 (Klarheit, Strukturiertheit) bezieht sich auf die Lehrkraft, wobei hier die durch Sprache und Darstellungsformen erzeugte Verständlichkeit des Unterrichtsgeschehens im Fokus steht.

(c) *Konsolidierung*: Hingegen bezieht sich SK3 (Festigung, Überprüfung) stärker auf die durch Transfer und Wiederholung erzeugte Stabilisierung von Wissensbeständen im Lernprozess.

(d) *Schülerorientierung*: Eine klare Fokussierung der Schülerinnen und Schüler findet sich in SK4 (Schülerbeteiligung, Variabilität, Interesseweckung, Flexibilität), welche die Berücksichtigung kognitiver und affektiver Aspekte im Unterricht umfasst und stärker auf die Lerngruppe als Ganzes bezogen ist.

(e) *Inhaltliche Auseinandersetzung*: SK5 (Vernetzung, Kritische Prüfung, Anwendungsbezug, Stimulierung) bezieht sich dagegen stärker auf den fachlich-inhaltlichen Wissenserwerb inklusive kritisch-reflexiver Elemente. Aufgrund ihrer stärker fachlichen Ausrichtung ist SK5 von besonderem Interesse. Auffällig ist die im Skalenvergleich größte Heterogenität der Mittelwerte der Einzelitems ($\Delta=0,87$): Das innerhalb der Subskala am höchsten geratete Item *Vernetzung* ($M=3,79$) liegt mit seinem Mittelwert deutlich höher als die anderen drei Items *Sti-*

mulierung ($M=2,92$), *Kritische Prüfung* ($M=3,14$) und *Anwendungsbezug* ($M=3,14$).

(f) *Individualisierung des Lernens*: Im Mittelpunkt von SK6 (Selbstkonzeptstärkung, Autonomieunterstützung, Differenzierung) steht die Unterstützung des Lernprozesses mit Blick auf die Voraussetzungen der einzelnen Schülerinnen und Schüler. Im Vergleich zur Schülerorientierung (SK4) fokussiert diese Subskala auf die individualisierten Lernprozesse.

Insgesamt liegen fünf der sechs Subskalen über dem arithmetischen Mittel und deuten auf eine grundsätzlich hohe Beurteilung geographischer Unterrichtsqualität hin. *Klassenmanagement* und *Klarheit & Struktur* bilden die beiden am höchsten bewerteten Subskalen (jeweils $M=4,16$). *Konsolidierung* ($M=3,63$), *Schülerorientierung* ($M=3,44$) und *Inhaltliche Auseinandersetzung* ($M=3,25$) sind ebenfalls hoch bewertet, jedoch niedriger als die vorausgegangenen. Die *Individualisierung des Lernens* liegt als einzige Subskala unterhalb des skalentheoretischen Mittelwertes ($M=2,90$) und ist damit am niedrigsten bewertet.

5.3 Einflussfaktoren

Erhebungsjahr

Alle Subskalen zeigen 2007/08 ihre niedrigsten Werte (vgl. Fig. 3). Dieser Rückgang zu 2006 ist mutmaßlich auf das veränderte Samplingverfahren von freiwilliger zu obligatorischer Teilnahme zurückzuführen (vgl. Kap. 4). Zur Gewährleistung der

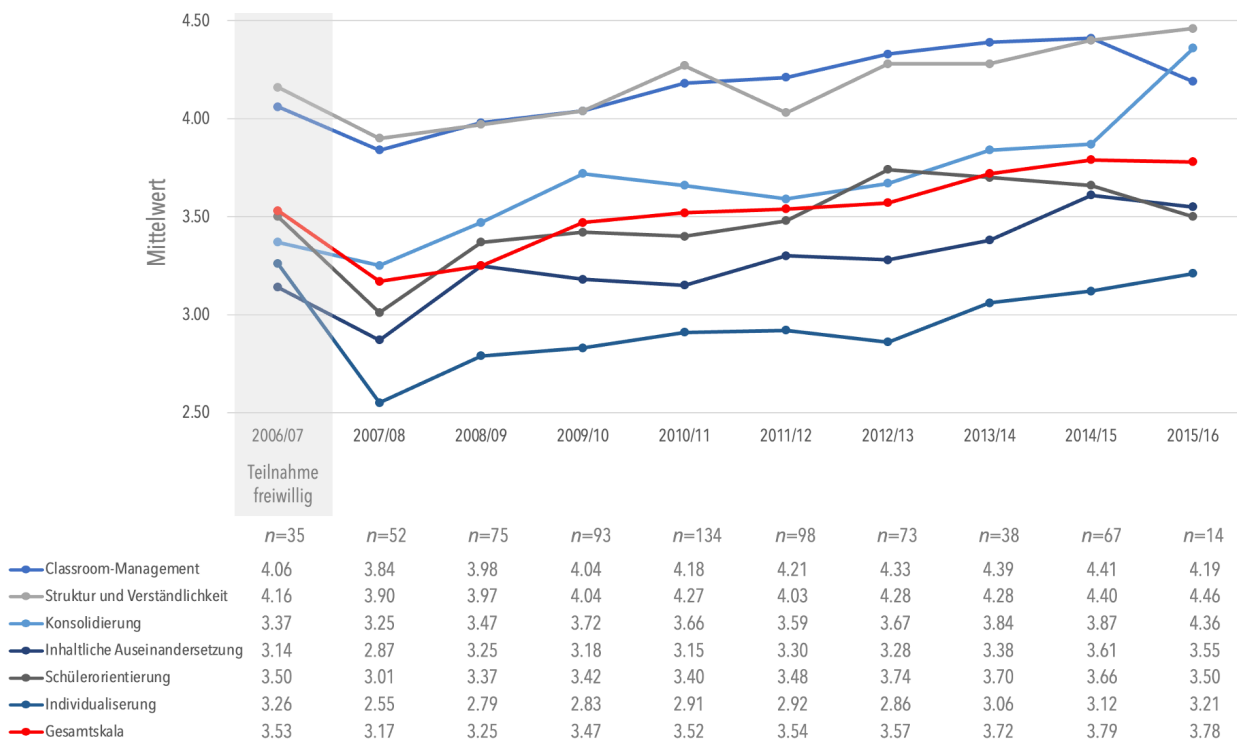


Fig. 3. Geographische Unterrichtsqualität in Abhängigkeit vom Erhebungsjahr, differenziert nach Subskalen & Gesamtskala ($n=679$) (Quelle: Autorinnen und Autor)

Vergleichbarkeit beziehen sich die folgenden Ausführungen auf die Schuljahre zwischen 2007/08 und 2015/16. Während dieses Erhebungszeitraums zeigt sich überwiegend eine durchgängige leichte Steigerung der Unterrichtsqualität.

Die im Skalenvergleich über alle Erhebungsjahre hinweg am höchsten geratete Skala *Klarheit und Struktur* zeigt eine Zunahme bis zum Schuljahr 2010/11, gefolgt von einem Rückgang (-0,2) zum darauffolgenden Schuljahr. Zwischen 2011/12 bis zum Erhebungsende ist hingegen eine erneute Zunahme zu verzeichnen, sodass diese Subskala zum Erhebungsende ihren Höchstwert von 4,46 erreicht. Für die Subskala *Klassenmanagement* zeigt sich eine nahezu lineare Steigerung (+0,6) zwischen 2007/08 und 2014/15. Zum Schuljahr 2015/16 fällt die Qualität auf 4,2 zurück. In diesem Zeitraum ist auch ein leichter Rückgang der Subskala *Inhaltliche Auseinandersetzung* von 3,66 auf 3,5 auszumachen. Konträr hierzu steht die Subskala *Konsolidierung*, die zwischen 2014/15 und 2015/16 steigt (+0,5) und so ihren Höhepunkt erreicht. Insgesamt bildet dies die Skala mit dem größten Zuwachs (+1,11) und weist zum Erhebungsende mit 4,36 den zweithöchsten Wert im Skalenvergleich auf. Einen deutlichen Anstieg verzeichnet ebenfalls die Subskala *Individualisierung des Lernens*: liegt diese 2007/08 noch bei 2,55 und damit unter dem skalentheoretischen Mittelwert, ist ein nahezu durchgängiger Anstieg auf 3,21 im letzten Erhebungsdurchgang zu erkennen. Anders verhält es sich mit der Skala *Schülerorientierung*, welche bereits 2012/13 ihren Hochpunkt erreicht und seitdem bis zum Erhebungsende kontinuierlich abnimmt.

Schulform

Neben dem Erhebungszeitpunkt lassen sich Differenzen bezüglich der schulformbezogenen Geographieunterrichtsqualität zwischen Mittel-/Ober-

schule ($n=395$) und Gymnasium ($n=169$) identifizieren (Fig. 4), wobei zur Vergleichbarkeit ausschließlich Daten der Sekundarstufe I herangezogen wurden. Diese dichotome Auswahl ist aufgrund deren Äquivalenz zu Schulformen in anderen Bundesländern zu legitimieren, während der Ausschluss der anderen Schulformen der Erhebung auf deren hohen Spezialisierungsgrad zurückzuführen ist.

Zwar liegt auf der Gesamtskala die Unterrichtsqualität in beiden Schulformen deutlich über dem skalentheoretischen Mittelwert und damit im hohen Bereich, allerdings zeigen sich signifikante Unterschiede ($p<.05$) zwischen Mittel-/Oberschule ($M=3,48$) und Gymnasium ($M=3,59$). Damit stellt sich die Schulform als signifikanter Einflussfaktor auf geographische Unterrichtsqualität im Gesamten heraus. Auch auf Subskalenebene fällt die Unterrichtsqualität am Gymnasium durchgängig höher aus. Statistisch signifikant ($p<.05$) ist dies allerdings nur bei der Subskala *Inhaltliche Auseinandersetzung*, deren Ausprägung bei der Mittel-/Oberschule bei $M=3,19$, beim Gymnasium hingegen bei $M=3,35$ liegt.

Jahrgangsstufe

Ebenfalls bestehen Unterschiede der Unterrichtsqualität in Abhängigkeit von der Jahrgangsstufe (vgl. Fig. 5). Die Ergebnisse werden zunächst differenziert nach den Schulformen dargelegt, um so die Entwicklung der Unterrichtsqualität entlang der jeweils typischen Beschulung zu skizzieren, und anschließend zu vergleichen.

Die Unterrichtsqualität am Gymnasium weist zum Zeitpunkt des Abiturs einen nahezu identischen Wert auf ($M=3,78$) wie zum Zeitpunkt der Einschulung in der 5. Klasse ($M=3,77$). Dabei handelt es sich jedoch nur scheinbar um eine Stagnation, da die Entwicklung zunächst durch einen di-

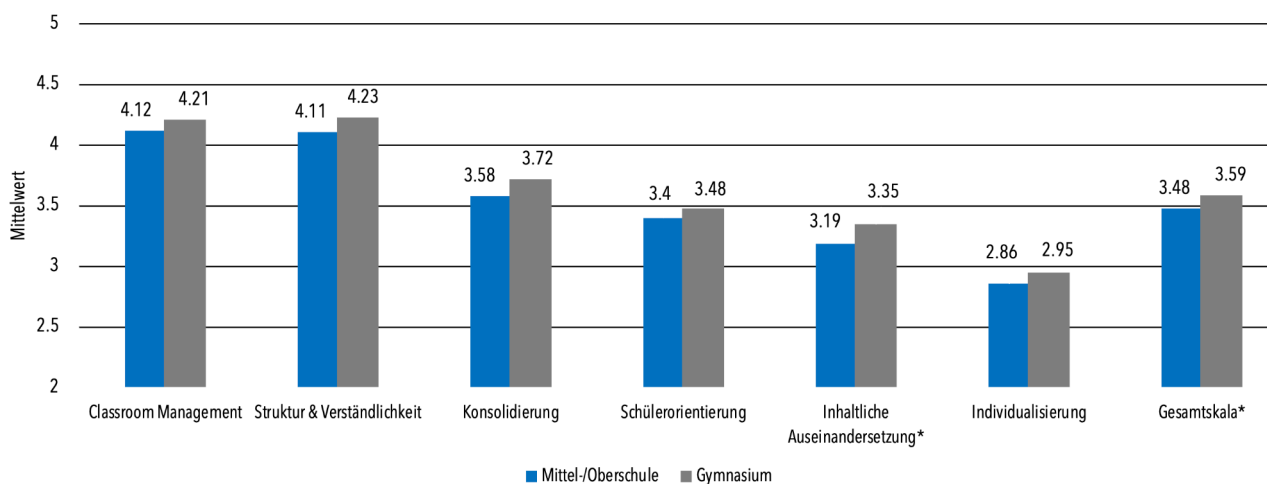


Fig. 4. Unterrichtsqualität in der Sekundarstufe I in Abhängigkeit der Schulart, differenziert nach Subskalen & Gesamtskala ($n=564$) (Quelle: Autorinnen und Autor)

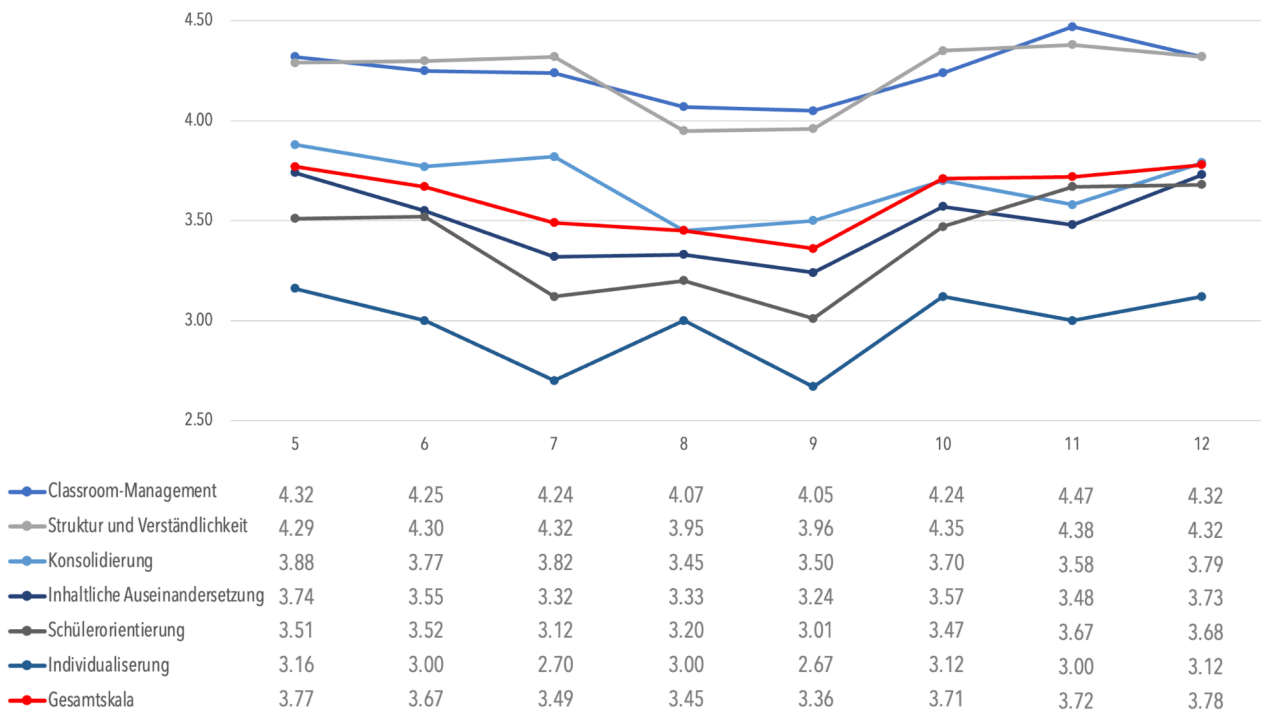


Fig. 5. Unterrichtsqualität an Gymnasien in Abhängigkeit der Jahrgangsstufe, differenziert nach Subskalen & Gesamtskala ($n=195$) (Quelle: Autorinnen und Autor)

stinktiven Rückgang der Unterrichtsqualität in der Mittelstufe gekennzeichnet ist: Die Tiefstwerte aller Subskalen lassen sich in der 8./9. Klasse verorten. Dem folgt ein sprunghafter Anstieg (+0,4) zur 10. Klasse, wobei der Gesamtskalenwert ($M=3,71$) zu diesem Zeitpunkt nicht das Ausgangsniveau der 5. Klasse erreicht: Dieses wird erst im Laufe der Oberstufe minimal übertroffen. Die beiden höchstbewerteten Subskalen *Klassenmanagement* und *Klarheit und Struktur* zeigen einen ähnlichen Verlauf: Beide Subskalen weisen zu den Anfangs- und Endzeitpunkten der Beschulung in etwa gleiche Werte um 4,3 auf, wobei in der Mittelstufe zwischenzeitlich ein Rückgang (ca. -0,3) zu verzeichnen ist. Eine ähnliche Tendenz zeigt sich bei den Subskalen *Konsolidierung* und *Schülerorientierung*, die aber insgesamt niedriger geratet sind. Beachtenswert ist der Verlauf der Subskala *Inhaltliche Auseinandersetzung*: diese stellt die einzige Subskala dar, die zwischen der 5. und 12. Klasse insgesamt an Unterrichtsqualität gewinnt, wobei sich auch hier zwischenzeitlich ein Rückgang auf den Tiefstwert (3,01) in der 9. Klasse zeigt. Ab diesem Zeitpunkt nimmt die Qualität dann kontinuierlich zu, erreicht in der 12. Klasse ihren Höchstpunkt ($M=3,68$) und übertrifft damit den Ausgangswert der 5. Klasse um 0,17. Die insgesamt am niedrigsten geratete Subskala *Individualisierung des Lernens* nimmt zwischen Beschulungsbeginn und -ende minimal (-0,04) ab. Der Entwicklungsverlauf ist dabei charakterisiert durch Schwankungen um den skalentheoretischen Mittelwert, wobei der

Tiefstwert in der 9. Klasse erreicht wird ($M=2,67$), bevor ihr größter Anstieg (+0,35) zur 10. Klasse erfolgt. Zum Zeitpunkt des Abiturs liegt diese mit 3,12 knapp unter dem Ausgangswert der 5. Klasse.

Zwischen der 5. und 10. Klasse der Mittel-/Oberstufe nimmt die Qualität des Geographieunterrichts um 0,12 leicht zu (vgl. Fig. 6). Beachtenswert ist dabei der Verlauf, da die Unterrichtsqualität von der 5. bis zur 9. Klasse kontinuierlich von 3,57 auf 3,36 sinkt, anschließend einen verhältnismäßig steilen Anstieg zur 10. Klasse nimmt und somit zum Beschulungsende ihren Höchstwert (3,65) erreicht. Auf Subskalenebene fällt auf, dass die insgesamt am höchsten geratete Subskala *Klassenmanagement* zwischen der 5. und 10. Klasse um 0,11 sinkt, wobei ihr Tiefpunkt in der 7. Klasse ($M=3,94$) liegt. Trotz einer anschließenden kontinuierlichen Steigerung wird das Ausgangsniveau (4,27) allerdings nicht mehr erreicht. Ein minimaler Qualitätsrückgang ist auch bei der am zweithöchsten gerateten Subskala *Klarheit und Struktur* zu verzeichnen, die in der 5. Klasse noch bei ihrem Höchstwert von 4,18 liegt, dann in der 7. Klasse auf ihren Tiefstwert (3,94) fällt und zum Beschulungsende knapp ihr Ausgangsniveau verfehlt (4,1).

Im Kontrast hierzu weisen die Subskalen *Schülerorientierung*, *Inhaltliche Auseinandersetzung* und *Individualisierung des Lernens* insofern ähnliche Tendenzen auf, als dass diese Nettozugewinne zum Beschulungsende verzeichnen: Ab der 5./6. Klasse sinkt die jeweilige Facette der Unterrichtsqualität zunächst kontinuierlich ab und erreicht in

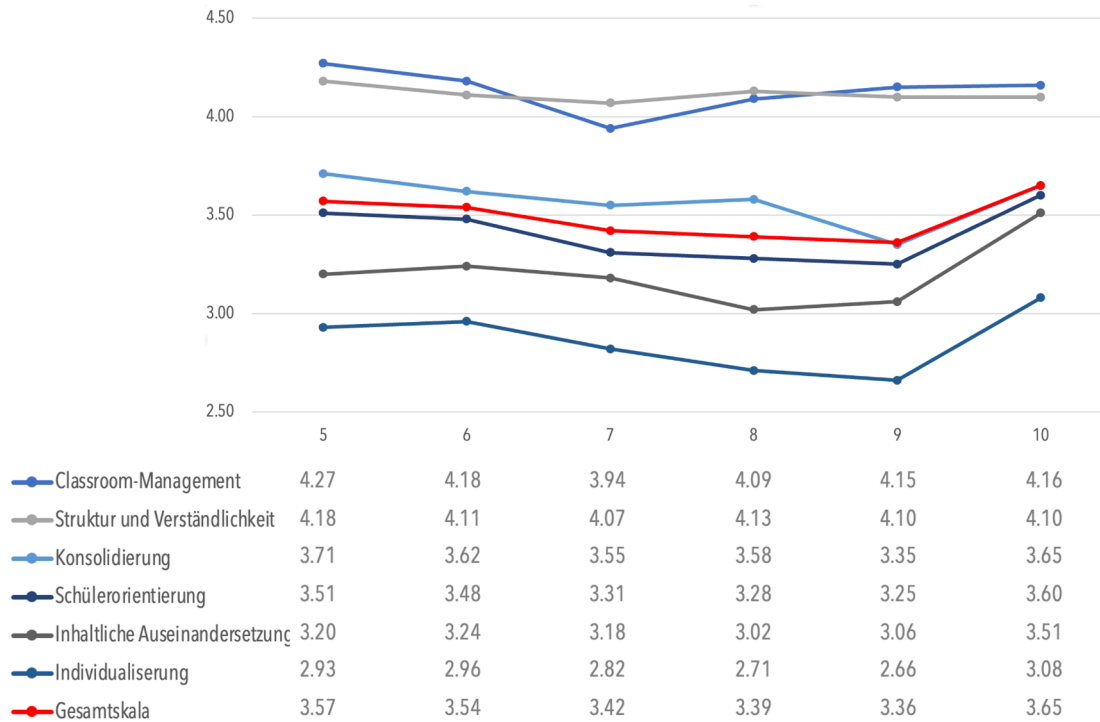


Fig. 6. Unterrichtsqualität an Mittel-/Oberschulen in Abhängigkeit der Jahrgangsstufe, differenziert nach Subskalen & Gesamtskala ($n=395$) (Quelle: Autorinnen und Autor)

der 8./9. Klasse ihren Tiefpunkt, bevor ein sprunghafter Anstieg auf die jeweiligen Skalenhöchstwerte in der 10. Klasse zu verzeichnen ist. Am ausgeprägtesten ist dieser Zugewinn in der Subskala *Inhaltliche Auseinandersetzung*, die innerhalb dieses Schuljahres um knapp 0,5 ansteigt. Eine Parallele hierzu zeigt die Subskala *Individualisierung des Lernens*, die in dieser Zeit um 0,4 ansteigt. Lag diese zuvor konstant in einem leicht negativen Bereich, erreicht diese in der 10. Klasse erstmalig einen über dem skalentheoretischen Mittel liegenden Wert ($M=3,08$). Diese skizzierten Unterschiede in Abhängigkeit der Jahrgangsstufe sind bezüglich der Individualisierung statistisch signifikant ($p<.05$).

Während die Globalskalenwerte der Mittel-/Oberschulen bei Schulabgang in Klasse 10 die der Einschulung leicht übertreffen ($+0,08$), sind diese an Gymnasien zum Zeitpunkt des Abiturs nahezu identisch ($+0,01$); zum Zeitpunkt der 10. Klasse ist an Gymnasien sogar ein leichter Nettorückgang zu verzeichnen ($-0,06$). Trotz dessen ist die Unterrichtsqualität im Gesamten in der 10. Klasse an Gymnasien ($M=3,71$) noch höher als an Mittel-/Oberschulen

($M=3,65$). Diese Differenz vergrößert sich bis zum Zeitpunkt des Abiturs weiter ($M=3,78$). Damit wird die Geographieunterrichtsqualität zum Zeitpunkt des Regelabschlusses an Gymnasien höher als an Mittel-/Oberschulen evaluiert. Tiefstwerte der Unterrichtsqualität werden bei beiden Schulformen in der Mittelstufe erreicht, wobei der größte Qualitätsanstieg bei beiden Schulformen zwischen der 9. und 10. Klasse zu verzeichnen ist. Dieser fällt beim Gymnasium größer aus ($+0,36$) als an Mittel-/Oberschulen ($+0,29$). Auf Subskalenebene ist in der Sekundarstufe I bezüglich der Subskala *Inhaltliche Auseinandersetzung* an Mittel-/Oberschulen ein Nettozugewinn ($+0,31$) zu verzeichnen, während an Gymnasien ein leichter Nettoverlust ($-0,04$) zu konstatieren ist. Auch die Gesamtwerte der 10. Klasse fallen bei Gymnasien ($M=3,47$) leicht geringer aus als an Mittel-/Oberschulen ($M=3,51$). Während an Gymnasien die Jahrgangsstufe weder auf Gesamt- als auf Subskalenebene einen signifikanten Einflussfaktor auf Unterrichtsqualität bildet, erweist sich an Mittel-/Oberschulen die Jahrgangsstufe als signifikanter Einflussfaktor auf die Subskala *Individualisierung des Lernens*.

6. Diskussion der Forschungsfragen

Die *Ausprägung geographischer Unterrichtsqualität* (HF1) fällt auf Gesamtskalenebene positiv ($M=3,53$) aus und ist damit anschlussfähig zum hervorragenden Abschneiden sächsischer Schülerinnen und Schüler in Bildungsvergleichsstudien

(STANAT ET AL., 2022). Auch wenn das Fach Geographie per se kein Teil dieser Vergleichsstudien ist, so deuten deren Ergebnisse auf eine im Bundeslandvergleich insgesamt höhere Unterrichtsqualität hin.

Die *Facetten geographischer Unterrichtsqualität* (TF1.1) sind anhand von sechs Subskalen (s.o.) differenzierbar, die einen ersten Zugriff zur Erfassung von Unterrichtsqualität bieten. Es zeigt sich, dass die Unterrichtsqualität im Fach Geographie in lehrkraftsseitigen, stärker pädagogisch-organisatorischen Subskalen höher ausfällt als in stärker schülerseitigen, den fachlichen Lernprozess betreffenden Subskalen. So stellen die Subskalen *Klassenmanagement* und *Klarheit und Struktur* die am höchsten gerateten Facetten (je $M=4,16$) dar. Zwar können diese in ihrer Ausprägung auch fachliche Aspekte miteinschließen (z.B. terminologische Klärung von Fachbegriffen), jedoch beziehen sie sich grundsätzlich und überwiegend auf fachunspezifische Aspekte des Unterrichts (z.B. präzise Formulierungen, Sitzordnung, Verhaltensregeln, Phasierung des Unterrichts). Auch in anderen Fachdidaktiken werden Klarheit und Strukturiertheit als generische Merkmale aufgefasst (u.a. BRUNNER, 2018). Hingegen werden die stärker fachlich ausgerichteten Subskalen, die die vertiefte Auseinandersetzung der Schülerinnen und Schüler mit dem Lerngegenstand umfassen, niedriger bewertet: So liegt die *Inhaltliche Auseinandersetzung* als vorletztplatzierte Subskala ($M=3,25$) nur in einem mittleren Bereich und die letztplatzierte *Individualisierung des Lernens* ist gering ausgeprägt ($M=2,92$). Diesbezüglich ist auf Einzelitemebene interessant, dass das Item *Vernetzung* ($M=3,79$) deutlich höher geratet ist als z.B. *Kritisches Prüfen* ($M=3,14$, z.B. kritische Auseinandersetzung mit Lerninhalten). Fachfremde Raterinnen und Rater erkennen also das der Geographie als Systemwissenschaft inhärente Erkennen von Zusammenhängen als eine deutliche Stärke, während die kritische Auseinandersetzung mit Inhalten und die Prüfung der Relevanz von Informationen schwächer ausgeprägt ist. Dies ist kongruent zu Befunden der empirischen Lehr-Lern-Forschung, in denen gezeigt wurde, dass qualitativvolles Klassenmanagement leichter zu erreichen ist als beispielsweise ein hohes kognitives Aktivierungspotenzial: So zeigt die SCHOLASTIK-Studie im Primarstufenbereich, dass bezüglich Klarheit/Strukturiertheit und Klassenführung alle Klassen überdurchschnittliche Werte aufweisen (WEINERT & HELMKE, 1997). Dabei fallen die Effektstärken einer gelungenen Klassenführung auf den Lernerfolg mit $d=0,35$ mittelgroß aus (HATTIE, 2013). Anhand dessen, dass eine gelingende Klassenführung als unterrichtliche Tiefenstruktur Voraussetzung für lernwirksamen Unterricht bildet (HELMKE & SCHRADER, 2008, S. 37), ist dieser Befund zunächst erfreulich.

Dass hingegen die Fachlichkeit im Geographieunterricht (*Inhaltliche Auseinandersetzung*) vergleichsweise niedrig bewertet wird, ist kritisch zu

sehen. Hierfür sind mehrere Gründe denkbar: Erstens ist dieser Befund kongruent zu Ergebnissen von SCHOLTEN ET AL. (2022), nach denen Geographielehramtsstudentinnen und Geographielehramtsstudenten in unterrichtsnahen Anforderungssituationen kein fachliches Wissen (CK), sondern vor allem pädagogisches Wissen (PK) nutzen, so dass auch hier eine Präferenz allgemeindidaktischer gegenüber fachlich-fachdidaktischer Inhalte besteht. Auch wenn dieser Befund durch die Natur der Erhebungssituation (fachfremde Raterinnen und Rater) in Teilen kritisch zu reflektieren ist, deutet er trotzdem darauf hin, dass sich das Erreichen fachlicher Qualität anspruchsvoller als das Herstellen organisatorisch-pädagogischer Strukturen gestaltet. Demnach bestehen bezüglich der unterrichtlichen fachlichen Tiefe Optimierungspotentiale.

Dies gilt insbesondere auch für die letztplatzierte Subskala *Individualisierung des Lernens*, welche als einzige unterhalb des skalentheoretischen Mittelwertes liegt ($M=2,90$). Dies ist anschlussfähig zum Befund von VOCK ET AL. (2022), nach welchem die innere Differenzierung des Mathematik- und Deutschunterrichts aus Schülereinschätzung unabhängig der Jahrgangsstufe oder Klassentyp gering ausgeprägt ist. Ergänzend zeigen DECRISTAN und JUDE (2017), dass sich Lehrkräfte im Umgang mit Heterogenität häufig überfordert fühlen und Unterstützungsbedarf äußern. Auch in der Geographie zeigen Ergebnisse der roadmap-Studie, dass Geographielehrkräfte Differenzierung als Herausforderung sehen, sodass ein selbstreflexives Bewusstsein vorliegt (FÖGELE ET AL., 2022). Angesichts der hohen Effektstärken, die beispielsweise Feedback ($d=0,70$) oder Scaffolding ($d=0,83$) auf den Lernerfolg haben (vgl. HATTIE, 2013), illustriert der Befund eindrücklich die fortbestehende Relevanz von Diagnostik und Förderung in der Lehrkräftebildung in der Geographie, um die evidente Diskrepanz zwischen der Bedeutung individualisiertem Geographielernens und der tatsächlichen unterrichtlichen Situation zu schließen.

Zudem wurden Erkenntnisse zu *Einflussfaktoren auf geographische Unterrichtsqualität* (TF1.2) generiert:

- Bezüglich des *Erhebungszeitpunkts* zeigt sich eine leichte Steigerung geographischer Unterrichtsqualität bis zum Ende der Erhebungen (2015/16) und weicht damit von den PISA-Trends ab; allerdings weisen einige Subskalen (*Schülerorientierung*, *Inhaltliche Auseinandersetzung*, *Klassenmanagement*) Rückgänge seit 2012 auf, die parallel dazu liegen. Entsprechend weichen die retrospektiven Entwicklungstrends geographischer Unterrichtsqualität teils von den PISA-Ergebnissen zur Schülerleistung ab.

◦ Weiterhin zeigen sich Differenzen der Unterrichtsqualität in Abhängigkeit der *Jahrgangsstufe*. Während am Gymnasium die Unterrichtsqualität zum Zeitpunkt des Abiturs nur minimal die der 5. Klasse übertrifft, ist die Unterrichtsqualität an Mittel-/Oberschulen nie höher als im Abschlussjahr. Gemeinsamkeiten zwischen den Schulformen sind bezüglich der Entwicklung zu erkennen: sowohl an Gymnasien wie an Mittel-/Oberschulen liegen die Tiefstwerte der Unterrichtsqualität in der 8./9. Klasse im Sinne eines *Pubertätsknicks*, dem ein Qualitätssprung von der 9. auf die 10. Klasse folgt. Dieser Befund ist anschlussfähig zum MET-Projekt, in welchem ein Rückgang der Unterrichtsqualität zwischen den Klassen 4-8 für Mathematik und Englisch konstatiert wird (vgl. MIHALY & McCAFFREY, 2015). Der sich anschließende Qualitätssprung zur 10. Klasse könnte auf den Eintritt in die Oberstufe bzw. Erreichen der mittleren Reife zurückzuführen sein, da durch solche externen Anreize möglicherweise eine erhöhte Leistungsbereitschaft auf Seite der Schülerinnen und Schüler und der Lehrkräfte besteht. Entsprechend könnte Unterrichtsqualität auch von den Bedingungen des Bildungssystems sowie der Situation der Schülerinnen und Schüler abhängen.

◦ Bezüglich der *Schulform* ist zu konstatieren, dass Geographieunterrichtsqualität in der Sekundarstufe I an Gymnasien signifikant höher ausfällt als an Mittel-/Oberschulen. Dabei zeigen sich auf Subskalenebene bedeutsame Unterschiede in der *Inhaltlichen Auseinandersetzung*, die an Gymnasien (Sek. I) signifikant höher ausfällt als an Mittel-/Oberschulen. Dabei scheinen die schulformspezifischen Befunde zur Subskala *Konsolidierung* widersprüchlich zu weiteren Evidenzen zu sein. Während das Einüben von Prozeduren und Techniken durch KUNTER ET AL. (2005) als Erklärungsansatz für eine im Vergleich zum Gymnasium deutlich geringere kognitive Aktivierung in der Hauptschule angeführt wird, wird Konsolidierung im Geographieunterricht des Gymnasiums ($M=3,72$) höher geratet als in der Mittel-/Oberschule ($M=3,58$). Auch insgesamt fallen die Schulformunterschiede in der Sekundarstufe I signifikant aus, insbesondere mit Blick auf die bereits diskutierte *Individualisierung des Lernens*. Dass gerade diese Subskala an Mittel-/Oberschulen – welche sich durch eine grundsätzlich höhere Heterogenität der Lernenden auszeichnen und damit umso höhere Differenzierungsmaßnahmen einfordern – signifikant geringer ausfällt als an Gymnasien, ist problematisch.

7. Kritische Reflexion

Bedingt durch die sekundäranalytische Anlage konnte kein Einfluss auf den Beobachtungsbogen genommen werden, welcher Unterrichtsqualität primär an lehrkrafts betreffenden Variablen operationalisiert. Dies stellt einen typischen Kritikpunkt an unterrichtsqualitätsbezogener Forschung dar (vgl. SCHOLTEN ET AL., 2023). Ferner spiegeln die Ergebnisse das Antwortverhalten der Raterinnen und Rater wider. Im Gegensatz zu neueren Forschungsansätzen wird die Perspektive der Schülerinnen und Schüler in begleitenden Fragebögen nicht berücksichtigt. Zudem sind im Beobachtungsbogen nicht nur Items der Tiefenstruktur, sondern vereinzelt auch weniger lernwirksame Indikatoren der Sichtstruktur integriert. Zweitens konnte die Subskalierung nicht theoretisch und in Pilotstudien faktorenanalytisch a priori hergeleitet werden, was in vergleichsweise kleinen, die Validität einschränkenden Subskalen resultiert. Ferner lag der Erhebungsfokus auf der Messung von Unterrichtsqualität in ihrem *Prozessablauf*, allerdings erfolgte damit keine einhergehende *Outcome-Messung*, um beides aufeinander zu beziehen. Weiterhin ist der Beobachtungsbogen, dem ur-

sprünglichen Erkenntnisinteresse folgend, in seiner grundsätzlichen Anlage generisch und wenig fachspezifisch. Zwar ist zudem herauszustellen, dass die Fakultas der Raterinnen und Rater nicht erfasst wurde. Vor dem Hintergrund der umfangreichen Raterschulungen (vgl. Kap. 4.1) ist jedoch davon auszugehen, dass die Vergleichbarkeit der Ratings grundsätzlich gegeben ist. Die Frage, ob Raterinnen und Rater mit Geographiehintergrund zu anderen Ergebnissen als ihre fachfremden Kolleginnen und Kollegen kämen, bietet einen Ansatzpunkt für zukünftige Forschung.

Die räumliche Fokussierung auf Sachsen bildet eine weitere Limitation der Studie, da die Erkenntnisse aufgrund bundeslandabhängig unterschiedlicher Rahmenvorgaben und Bildungssysteme nur eingeschränkt generalisiert werden können, zumal Vergleichsstudien zeigen, dass das Bundesland eine wichtige Einflussvariable auf den Lernerfolg darstellt (STANAT ET AL., 2022). Neben der räumlichen besteht zudem eine zeitliche Limitation, da der Erhebungszeitraum – bedingt durch datenschutzrechtliche Fragen bei der Sekundäranalyse – zeitlich weiter zurückliegt. Dies ist vor

dem Hintergrund der Methodik jedoch keineswegs untypisch, sondern stellt ein etabliertes Vorgehen dar, da beispielsweise auch TIMSS-Datensätze weiterhin für Sekundäranalysen herangezogen werden (u.a. NEUBAUER, 2022). Zu bedenken ist, dass die dargestellten Trends und Entwicklungslinien nicht unreflektiert extrapoliert werden können, da Ergebnisse von weiteren Vergleichsstudien wie beispielsweise der IQB-Bildungstrends seit Erhebungsende, auch vor dem Hintergrund gesellschaftlicher Ereignisse (z.B. Pandemie ab 2020), rückgängig sind (vgl. STANAT ET AL., 2022). Auch die kurzzeitige Evaluationsspanne von 45 Minuten ist vor dem Hintergrund möglicherweise eingeschränkter Reliabilität kritisch zu sehen (vgl. PRAETORIUS, 2013).

Dem stehen besondere Stärken gegenüber. Es handelt sich um objektivierte, extern geratete Daten, die im Gegensatz zu anderen Erhebungsmethoden, wie z.B. Selbstbeschreibungen von Lehr-

kräften, weniger stark Verzerrungseffekten unterliegen und damit eine höhere Güte aufweisen (vgl. NILSEN ET AL., 2016).

Weiterhin ist der hohe Samplingumfang mit 679 beobachteten Geographieunterrichtsstunden hervorzuheben. Damit fokussiert diese Studie den einzigen derzeit vorliegenden längsschnittlichen large-scale-Datensatz zur Erfassung der Unterrichtsqualität im Fach Geographie. Ein immenser Mehrwert liegt zudem im doppelten Prozesscharakter der Studie: Erstens handelt es sich um eine in-vivo-Feldstudie, welche den Unterrichtsprozess untersuchte. Dies stellt beinahe ein Alleinstellungsmerkmal in der Geographiedidaktik dar, in welcher der Großteil der Untersuchungen meist als Labor- oder Interventionsstudien angelegt sind. Zweitens ist die mehrjährige, längsschnittliche Anlage hervorzuheben, die retrospektive Aussagen über Entwicklungstrends zulässt.

8. Ausblick

Forschungsperspektive 1: Replikationsstudie zur Verringerung der Limitationen

Weiterer Forschungsbedarf besteht bezüglich der fachspezifischen Erfassung geographischer Unterrichtsqualität. Nicht nur vor dem Hintergrund des Diskurses um die fachspezifische Ausdifferenzierung von Unterrichtsqualität (vgl. PRAETORIUS & GRÄSEL, 2021), sondern auch anhand des Erkenntnisses dieser Studie, dass die Unterrichtsqualität auf den stärker fachspezifischen Subskalen *Inhaltliche Auseinandersetzung* und *Individualisierung des Lernens* geringer ausfällt als auf den eher generischen, ist die Notwendigkeit einer geographiedidaktischen Ausdifferenzierung zu begründen. Hierfür sind zunächst eine fachspezifische Modellierung und Operationalisierung geographischer Unterrichtsqualität (vgl. SCHOLTEN ET AL., 2023, in diesem Heft) unablässig. Neben der ausstehenden Entwicklung eines geographiespezifischen Beobachtungsbogens sollte in zukünftigen Vorhaben, auch angesichts der Kritik LINDMEIER und HEINZES (2020), welche die Relevanz des Fachhintergrundes herausstellen, geographiedidaktisch gebildete Raterinnen und Rater eingesetzt werden. Um der häufig lehrkräftsbezogenen Operationalisierung entgegenzuwirken, sind weitere Begleiterhebungen mitsamt zugehöriger Instrumententwicklung (z.B. ergänzendes Lernendenfeedback mittels Fragebögen, vgl. WISNIEWSKI & ZIERER, 2020) zur Datentriangulation vorzunehmen. Ebenfalls sollte, dem Prozess-Produkt-Paradigma folgend neben

der Beforschung des Geographieunterrichtsprozesses analog zu Studien der Naturwissenschaftsdidaktiken künftig auch dessen Outcome (z.B. Testung des kognitiven Lernzuwachses) gemessen sowie weitere Einflussfaktoren (z.B. Hintergrund der Lehrkraft) identifiziert und erhoben werden (vgl. MÖLLER ET AL., 2002), um noch fundiertere Aussagen zur Wirksamkeit von Geographieunterricht treffen zu können. Hierzu ist auch eine Ausdehnung um weitere Bundesländer, die im Bildungsvergleich eine größere Varianz aufweisen, zur Vervollständigung der Datenlage anzustreben, um Hinweise auf mögliche regionale Unterschiede geographischer Unterrichtsqualität zu generieren.

Forschungsperspektive 2: Von allgemeiner Geographieunterrichtsqualität hin zur Untersuchung gegenstandsbezogener Geographieunterrichtsqualität

Ungeklärt ist bisher, wie sich geographische Unterrichtsqualität bezüglich inhaltlicher (z.B. Unterrichtsqualität beim Thema Regenwald), konzeptioneller (z.B. Anbahnung von Basiskonzepten) oder methodischer (z.B. Exkursionen) Foki artikuliert. Aufgrund des forschungsökonomischen Aufwandes bietet sich Arbeitsteilung in Verbundanträgen an: beispielsweise könnte eine große Anzahl von Geographieunterrichtsstunden videographiert werden, welche standortgebunden fokusspezifisch ausgewertet werden. Auch können die entstehenden Videoartefakte sowohl von fachfremden als auch

geographisch gebildeten Raterinnen und Ratern ausgewertet werden, um so die Diskussion um Generik und Fachspezifität von Unterricht (vgl. PRAETORIUS & GRÄSEL, 2021) weiter auszuscharfen. Eine lohnenswerte Ergänzung zu den o.g. Stuserhebungen der Qualität des Geographieunterrichts ist eine längsschnittliche Verlaufserfassung, bei welcher z.B. der Geographieunterricht weniger Lehrkräfte über einen längeren Zeitraum erhoben wird. Weiterführend wäre die Fokussierung auf eine einzelne Basisdimension der fachlichen Unterrichtsqualität (z.B. strukturierte Gesprächsführung) ertragreich, da dies eine tiefgehende Analyse ermöglicht.

Literatur

- BAUMERT, J., KUNTER, M., BLUM, W., BRUNNER, M., VOSS, T., JORDAN, A., KLUSMANN, U., KRAUSS, S., NEUBRAND, M., & TSAI, Y.-M. (2010). [Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress](#). *American Educational Research Journal*, 47(1), 133-80.
- BERLINER, D. C. (2005). [The Near Impossibility of Testing for Teacher Quality](#). *Journal of Teacher Education*, 56(3), 205-13.
- BRUNNER, E. (2018). [Qualität von Mathematikunterricht: Eine Frage der Perspektive](#). *Journal für Mathematik-Didaktik*, 39, 257-84.
- BUDKE, A. (2021). Bedeutung des argumentativen Schreibens im Geographieunterricht im Kontext aller Schreibtätigkeiten – eine empirische Studie auf der Grundlage von Unterrichtsbeobachtungen. In A. BUDKE & F. SCHÄBITZ (Hg.), *Argumentieren und Vergleichen: Beiträge aus der Perspektive verschiedener Fachdidaktiken* (S. 57-76). LIT.
- CROSS, K. P. (1998). [Classroom Research: Implementing the Scholarship of Teaching](#). *New Directions for Teaching and Learning*, 75, 5-12.
- DECRISTAN, J., & JUDE, N. (2017). Heterogenitätskategorie Schulleistung/Leistung von Schülerinnen und Schülern. In T. BOHL, J. BUDDE & M. RIEGER-LADICH (Hg.), *Studienbuch Umgang mit Heterogenität in Schule und Unterricht* (S. 109-22). UTB.
- DOMAĆINOVIĆ, F., & VUK, R. (2022). [Research on the Frequency and Effectiveness of Teaching Methods in Geography Classes: An Example of Self-Observation of Teaching Practices](#). *Journal of the Geographical Institute Jovan Cvijic, SASA*, 72(3), 307-22.
- DÖRING, N., & BORTZ, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer.
- DRECHSEL, B., & SCHINDLER, A.-K. (2019). Unterrichtsqualität. In D. URHAHNE, M. DRESEL & F. FISCHER (Hg.), *Psychologie für den Lehrberuf* (S. 353-72). Springer.
- EINSIEDLER, W. (2002). Das Konzept Unterrichtsqualität. *Unterrichtswissenschaft*, 30(3), 194-96.
- FÖGELE, J., MEHREN, R., & THUME, S. (2022). [Die roadmap-Studie: Zur Situation des Schulfachs Geographie aus der Sicht von Lehrkräften](#). HGD.
- HATTIE, J. (2013). *Lernen sichtbar machen: Überarbeitete deutschsprachige Ausgabe von "Visible Learning"*. Schneider Verlag Hohengehren.
- HELMKE, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Klett/Kallmeyer.
- HELMKE, A., & SCHRADER, F. W. (2008). Merkmale der Unterrichtsqualität: Potenzial, Reichweite und Grenzen. In B. SCHAAL & F. HUBER (Hg.), *Qualitätssicherung im Bildungswesen. Auftrag und Anspruch der bayerischen Qualitätsagentur* (S. 69-108). Waxmann.
- HEMMER, M., HOFFMANN, K. W., & MEHREN, M. (2020). Lehrerprofessionalität und Lehrerprofessionalisierung im Fach Geographie – Annäherungen aus geographiedidaktischer Perspektive. In M. HEMMER, A.-K. LINDAU, C. PETER, M. RAWOHL & G. SCHRÜFER (Hg.), *Lehrerprofessionalität und Lehrerbildung im Fach Geographie im Fokus von Theorie, Empirie und Praxis: Ausgewählte Tagungsbeiträge zum HGD-Symposium 2018 in Münster. Geographiedidaktische Forschungen (Band 72)* (S. 1-34). HGD.
- HOYT, W. T. (2000). [Rater Bias in Psychological Research: When Is It a Problem and What Can We Do About It?](#) *Psychological Methods*, 5(1), 64-86.

- JOHNSTON, M. P. (2017). Secondary Data Analysis: A Method of Which the Time Has Come. *Qualitative and Quantitative Methods in Libraries*, 3(3), 619-26.
- KANE, T. J., KERR, K. A., & PIANTA, R. C. (2015). *Designing Teacher Evaluation Systems: New Guidance From the Measures of Effective Teaching Project*. Jossey-Bass.
- KANWISCHER, D. (Hg.) (2013). *Geographiedidaktik: Ein Arbeitsbuch zur Gestaltung des Geographieunterrichts*. Borntraeger.
- 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 (Hg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127-146). Waxmann.
- KLIEME, E., & RAKOCZY, K. (2008). [Empirische Unterrichtsforschung und Fachdidaktik: Outcome-orientierte Messung und Prozessqualität des Unterrichts](#). *Zeitschrift für Pädagogik*, 54(2), 222-37.
- KOTTHOFF, H.-G., & BÖTTCHER, W. (2010). Neue Formen der „Schulinspektion“: Wirkungshoffnungen und Wirksamkeit im Spiegel empirischer Bildungsforschung. In H. ALTRICHTER & K. MAAG MERKI (Hg.), *Handbuch Neue Steuerung im Schulsystem* (S. 295-325). VS Verlag für Sozialwissenschaften.
- KUGER, S., KLIEME, E., LÜDTKE, O., SCHIEPE-TISKA, A., & REISS, K. (2017). [Mathematikunterricht und Schülerleistung in der Sekundarstufe: Zur Validität von Schülerbefragungen in Schulleistungsstudien](#). *Zeitschrift für Erziehungswissenschaft*, 20(Suppl. 2), 61-98.
- KUNTER, M., BRUNNER, M., BAUMERT, J., KLUSMANN, U., KRAUSS, S., BLUM, W., JORDAN, A., & NEUBRAND, M. (2005). [Der Mathematikunterricht der PISA-Schülerinnen und -Schüler](#). *Zeitschrift für Erziehungswissenschaft*, 8(4), 502-520.
- KUNTER, M., & TRAUTWEIN, U. (2013). *Psychologie des Unterrichts*. Schöningh.
- KUNTER, M., & VOSS, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. KUNTER, J. BAUMERT, W. BLUM, U. KLUSMANN, S. KRAUSS & M. NEUBRAND (Hg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85-113). Waxmann.
- LAMBRECHT, M., & RÜRUP, M. (2012). Bildungsforschung im Rahmen einer *evidence based policy*: Das Beispiel „Schulinspektion“. In A. WACKER, U. MAIER & J. WISSINGER (Hg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung: Empirische Befunde und forschungsmethodische Implikationen* (S. 57-78). VS Verlag für Sozialwissenschaften.
- LASUB (SÄCHSISCHES LANDESAMT FÜR SCHULE UND BILDUNG) (2020). *Evaluation*. sachsen.de. Aufgerufen am 16. März 2023 [hier](#).
- LINDMEIER, A., & HEINZE, A. (2020). Die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung: (bisher) ignoriert, implizit enthalten oder nicht relevant? In A.-K. PRAETORIUS, J. GRÜNKORN & E. KLIEME (Hg.), *Empirische Forschung zu Unterrichtsqualität: Theoretische Grundfragen und quantitative Modellierungen* (S. 255-268). Beltz Juventa.
- LONG, M. H. (1980). [Inside the 'Black Box': Methodological Issues in Classroom Research on Language Learning](#). *Language Learning*, 30(1), 1-42.
- MEHREN, M., & MEHREN, R. (2020). Über die Tiefenstrukturen des (Geographie-)Unterrichts. *Praxis Geographie*, 50(4), 4-9.
- MEHREN, M., & MEHREN, R. (2022). Fachbezogene Tiefenstrukturen: Woran erkenne ich die Qualität eines Geographieunterrichts? *Praxis Geographie*, 52(7/8), 4-11.
- MIHALY, K., & MCCAFFREY, D. F. (2015). Grade-Level Variation in Observational Measures of Teacher Effectiveness. In T. J. KANE, K. A. KERR & R. C. PIANTA (Eds.), *Designing Teacher Evaluation Systems: New Guidance From the Measures of Effective Teaching Project* (pp. 9-49). Jossey-Bass.
- MÖLLER, K., JONEN, A., HARDY, I., & STERN, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung. In M. PRENZEL & J. DOLL (Hg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 176-91). Beltz.
- NEUBAUER, D. (2022). Unterrichtsqualität im Sachunterricht: Zur Bedeutung und Lernwirksamkeit forschend-entdeckender Lerngelegenheiten in Österreichs Primarstufe. In E. BOXHOFER, M. KRAMER, D. NEUBAUER & A. WEINBERGER (Hg.), *PH forscht: Bericht 2019-2021* (S. 120-21). Private Pädagogische Hochschule der Diözese Linz.
- NILSEN, T., GUSTAFSSON, J.-E., & BLÖMEKE, S. (2016). Conceptual Framework and Methodology of This Report. In T. NILSEN & J.-E. GUSTAFSSON (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (pp. 1-20). Springer Open.

- OECD (2019). *Deutschland - Ländernotiz - Ergebnisse PISA 2018*. Aufgerufen am 19. Januar 2024 [hier](#).
- OHL, U., & STREITBERGER, S. (2017). Einsatzmöglichkeiten des Augsburger Analyse- und Evaluationsrasters für Bildungsmedien in der Geographiedidaktik: Eine domänenspezifische Analyse am Beispiel eines kostenlosen Online-Unterrichtsmaterials zur globalen Produktionskette von Smartphones. In C.-C. FEY & E. MATTHES (Hg.), *Das Augsburger Analyse- und Evaluationsraster für analoge und digitale Bildungsmedien (AAER): Grundlegung und Anwendungsbeispiele in interdisziplinärer Perspektive* (S. 141–66). Verlag Julius Klinkhardt.
- OSER, F. K., & BAERISWYL, F. J. (2001). Choreographies of Teaching: Bridging Instruction to Learning. In V. RICHARDSON (Ed.), *Handbook of Research on Teaching* (pp. 1031–65). American Educational Research Association.
- PRAETORIUS, A.-K. (2013). *Messung von Unterrichtsqualität durch Ratings*. Waxmann.
- PRAETORIUS, A.-K., & GRÄSEL, C. (2021). [Noch immer auf der Suche nach dem heiligen Gral: Wie generisch oder fachspezifisch sind Dimensionen der Unterrichtsqualität?](#) *Unterrichtswissenschaft*, 49, 167–188.
- PRAETORIUS, A.-K., HERRMANN, C., GERLACH, E., ZÜLSDORF-KERSTING, M., HEINITZ, B., & NEHRING, A. (2020). [Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum - zwischen Generik und Fachspezifik](#). *Unterrichtswissenschaft*, 48, 409–46.
- REINFRIED, S., & HAUBRICH, H. (2015). *Geographie unterrichten lernen. Die Didaktik der Geographie*. Cornelsen.
- REMPFLER, A. (Hg.) (2018). *Wirksamer Geographieunterricht*. Schneider Verlag Hohengehren.
- RINSCHADE, G., & SIEGMUND, A. (2019). *Geographiedidaktik*. UTB.
- SCHIEPE-TISKA, A., HEINE, J. H., LÜDTKE, O., SEIDEL, T., & PRENZEL, M. (2016). Mehrdimensionale Bildungsziele im Mathematikunterricht und ihr Zusammenhang mit den Basisdimensionen der Unterrichtsqualität. *Unterrichtswissenschaft*, 44(3), 211–25.
- SCHOLTEN, N., BIENERT, N., MEHREN, M. & MEHREN, R. (2023). Unterrichtsqualität im Fach Geographie – Herleitung und kritische Diskussion eines Modells. *Zeitschrift für Geographiedidaktik | Journal of Geography Education*, 51(3), 106–22.
- SCHOLTEN, N., DOLL, J., & MASANEK, N. (2022). [How Preservice Teachers Refer to Different Knowledge Domains When Evaluating a Lesson Plan on the Tropical Rainforest](#). *Journal of Geography*, 121(3), 91–9.
- SCHWARZE, S. (2021). *Die Konstruktion des subsaharischen Afrikas im Geographieunterricht der Sekundarstufe I*. ULB Münster.
- SEIDEL, T., PRENZEL, M., RIMMELE, R., DALEHEFTE, I. M., HERWEG, C., KOBARG, M., & SCHWINDT, K. (2006). [Blicke auf den Physikunterricht: Ergebnisse der IPN Videostudie](#). *Zeitschrift für Pädagogik*, 52(6), 799–821.
- SEIDEL, T., & SHAVELSON, R. J. (2007). [Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results](#). *Review of Educational Research*, 77(4), 454–99.
- SOLEM, M., STOLTMAN, J., LANE, R., BOURKE, T., CHANG, C. H., & VIEHRIG, K. (2018). An Assessment Framework and Methodology for a Trends in International Geography Assessment Study (TIGAS). *Geographical Education*, 31, 7–15.
- STANAT, P., SCHIPOLOWSKI, S., SCHNEIDER, R., SACHSE, K.A., WEIRICH, S., & HENSCHEL, S. (Hg.) (2022). *IQB-Bildungstrend 2021: Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Waxmann.
- THUME, S. (2023). *Die Relevanz ausgewählter Unterrichtsprinzipien für das geographische Lernen und Lehren aus Sicht unterschiedlicher Akteure der geographischen Bildung mittels Q-Methode* (Dissertation).
- VOCK, M., GRONOSTAJ, A., KRETSCHMANN, J., & WESTPHAL, A. (2022). [Wie bewerten begabte und leistungsstarke Jugendliche in separaten Spezialklassen ihren Unterricht?: Unterrichtsqualität in Deutsch und Mathematik in der Sekundarstufe](#). *Zeitschrift für Erziehungswissenschaft*, 25, 1151–73.
- WEINERT, F. E., & HELMKE, A. (1997). *Entwicklung im Grundschulalter*. Beltz Psychologie Verlags Union.
- WEINERT, F. E., SCHRADER, F. W., & HELMKE, A. (1989). [Quality of Instruction and Achievement Outcomes](#). *International Journal of Educational Research*, 13(8), 895–914.
- WISNIEWSKI, B., & ZIERER, K. (2020). [Empirische Arbeit: Entwicklung eines Online-Fragebogens zur Erhebung von Unterrichtsqualität durch Lernendenfeedback und erste Validierungsschritte](#). *Psychologie in Erziehung und Unterricht*, 67(2), 138–55.
- WÜSTEN, S. (2010). *Allgemeine und fachspezifische Merkmale der Unterrichtsqualität im Fach Biologie: Eine Video- und Interventionsstudie*. Logos Verlag.

Autorinnen und Autor

✉ **Dr. Natalie Bienert**

Universität Münster
Institut für Didaktik der Geographie
Heisenbergstraße 2
48149 Münster
E-Mail: natalie.bienert@uni-muenster.de

Dr. Nina Scholten

Universität Münster
Institut für Didaktik der Geographie
Heisenbergstraße 2
48149 Münster

Prof. Dr. Rainer Mehren

Universität Münster
Institut für Didaktik der Geographie
Heisenbergstraße 2
48149 Münster