

Unterrichtsqualität in den Naturwissenschaften – Die Suche nach dem Heiligen Gral

Die Suche nach den Merkmalen guten Unterrichts beschäftigt die empirische naturwissenschaftsdidaktische Forschung seit Jahrzehnten. Dabei hat sie nicht nur mit einigen Mythen aufräumen können, wie z.B. dass mehr Schülerexperimente per se lernförderlicher sind. Sie konnte auch zentrale Dimensionen guten Unterrichts wie z.B. Klarheit und Strukturiertheit oder kognitive Aktivierung für den naturwissenschaftlichen Unterricht bestätigen. Die Ergebnisse zeigen jedoch auch, dass diese Merkmale zwar einen Einfluss auf Schülerleistung haben, die Formel aber z.B. nicht ohne Weiteres lauten kann: Je höher die kognitive Aktivierung desto besser die Schülerleistung. Der vorliegende Beitrag fasst die wesentlichen Erkenntnisse aus der Forschung zur Unterrichtsqualität (in den Naturwissenschaften) zusammen, zeigt auf, warum die Frage nach der Gestaltung guten Unterrichts immer noch weitgehend unbeantwortet ist, und schließt mit einem Plädoyer für die Entwicklung und empirische Erprobung von Konzepten guten Unterrichts in den Naturwissenschaften.

Die prozess-produkt-orientierte Unterrichtsqualitätsforschung

Die frühe Forschung zur Unterrichtsqualität (in den Naturwissenschaften) fokussierte zunächst auf das Lehrerverhalten. Auf der Basis von unter Lehrerausbildern anerkannter Ansichten (d.h. „Expertenmeinungen“) über gutes Lehrerverhalten, wurden Studien des tatsächlichen Verhaltens von Lehrkräften und gezielten Interventionen zur Veränderung des Lehrerverhaltens durchgeführt, die durch wiederholte Unterrichtsbeobachtungen evaluiert wurden. In den 1960er Jahren verschob sich der Fokus der Unterrichtsqualitätsforschung zunehmend auf die Wirkung des Lehrerverhaltens (für eine ausführliche Darstellung siehe Rosenshine & Furst, 1971). Dies spiegelt sich unter anderem auch in dem von Carroll (1963) vorgeschlagenen Modell schulischen Lernens wieder. Dieses Modell beschreibt schulisches Lernen als Funktion des Verhältnisses von aufgewandter zu benötigter Zeit. Die aufgewandte Zeit hängt dabei nach Carroll (1963) von der Quantität des Unterrichts und Ausdauer der Lernenden, die benötigte Zeit von den Fähigkeiten der Lernenden und der Qualität des Unterrichts ab. Die Qualität des Unterrichts wiederum wird bestimmt durch die Klarheit, mit der Lernziele präsentiert werden, die Strukturiertheit, mit der Lernaktivitäten sequenziert werden, und die Aktivierung der Lernenden durch Lernmaterialien (Carroll, 1989). Mit diesem Modell bereitete Carroll (1989) die theoretische Basis für die Interpretation der Befunde zur Wirkung unterrichtlicher Merkmale auf das Lernen der Schülerinnen und Schüler.

In der Folge entstand eine Reihe von Forschungsarbeiten, die die Wirkung unterschiedlicher Merkmale unterrichtlichen Lernens speziell des Verhaltens von Lehrkräften auf das Lernen untersuchten (Rosenshine & Furst, 1971). Neben der reinen Quantität bestimmter Merkmale wurde dabei zunehmend auch deren Qualität in den Blick genommen; das heißt, bei den Unterrichtsbeobachtungen kamen neben niedrig-inferenten mehr und mehr auch hoch-inferente Bewertungssysteme zum Einsatz. Rosenshine und Furst (1971) identifizierten insgesamt elf unterschiedliche Merkmale – darunter die Klarheit z.B. der Lehreraussagen, die Strukturiertheit z.B. mit der die Lehrkraft durch den Unterrichtsgang führt, die Variabilität z.B. instruktionaler Methoden, oder auch der Enthusiasmus der Lehrkraft. Mit Blick auf verschiedene Beschränkungen der den identifizierten Merkmalen zugrundeliegenden Forschung – unter anderem der Problematik, mehrere hoch-inferente Merkmale im Rahmen einer Unterrichtsbeobachtung zu kodieren – plädierten Rosenshine und Furst (1971) für weitere Forschung zur Fundierung der von ihnen identifizierten Merkmale auf der Basis eines weiter entwickelten

methodischen Zugangs. In einer Erweiterung des rudimentären Zugangs der Untersuchung der Wirkung ausgewählter Unterrichtsmerkmale auf das Lernen der Schülerinnen und Schüler, beschreiben Dunkin und Biddle (1974) neben Merkmalen des Unterrichtsprozesses (z.B. dem Verhalten des Lehrers) zwei weitere Klassen von Merkmalen: Merkmale der Lehrkraft (z.B. der Persönlichkeit der Lehrkraft) und des Kontexts (z.B. der Ausstattung des Klassenraums). Zudem erweitern Dunkin und Biddle (1974) den Raum potenzieller Wirkungen dieser Merkmale über Schülerleistung hinaus auf weitere Produkte (z.B. auf die Einstellungen der Schülerinnen und Schüler).

Das Prozess-Produkt-Modell von Dunkin und Biddle (1974) führte zu einer massiven Zunahme der Forschung zu den Wirkungen verschiedener Merkmale des Unterrichtsprozesses oder mit Einfluss auf den Unterrichtsprozess auf entsprechende Produkte – speziell der Schülerleistung. Die Forschung belegte unter anderem die Bedeutung der Quantität des Unterrichts (Anderson, 1981), sowie verschiedener Merkmale der Unterrichtsqualität wie z.B. der Klarheit (Coker, Medley, & Soar, 1980), der Strukturiertheit (Dunkin, 1978), und der Aktivierung (Good & Grouws, 1977), aber auch der Variabilität (Sweitzer & Anderson, 1983) oder des Enthusiasmus (Armento, 1977). Aus einer Zusammenführung der verschiedenen vorliegenden Modellierungen (der Einflussfaktoren) des (unterrichtlichen) Lernens sowie entsprechender empirischer Erkenntnisse entwickelte Walberg (1981) sein „Model of Educational Productivity“. Diese Modell identifiziert neun Merkmale, die Lernen im Sinne der Entwicklung affektiver, kognitiver und Verhaltensmerkmale aufweist: 1) Fähigkeiten, 2) Entwicklung, 3) Motivation, 4) Quantität des Unterrichts, 5) Qualität des Unterrichts, 6) häusliche Umgebung, 7) schulische Umgebung, 8) Freundeskreis und 9) Massenmedien (vgl. Fraser, Walberg, Welch, & Hattie, 1987).

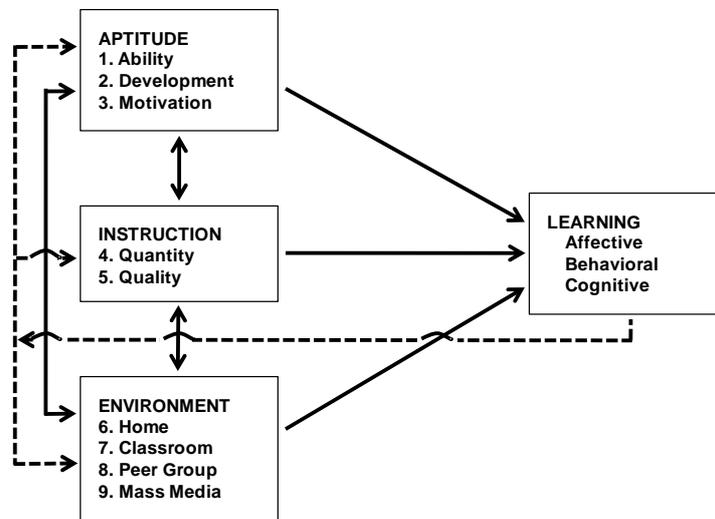


Abbildung 1. Walbergs (1981) „Model of Educational Productivity“.

Das Modell von Walberg (1981) bildete wiederum die Grundlage für eine Reihe von Metaanalysen zur empirischen Fundierung des Modells und damit zur Konsolidierung der Befunde der Prozess-Produkt-Forschung (Brophy & Good, 1986; Fraser et al., 1987; Walberg, Haertel, Pascarella, Junker, & Boulanger, 1981). Walberg et al. (1981) belegen z.B. anhand von Regressionsanalysen auf Basis von Daten aus dem National Assessment of Educational Progress

(NAEP) Programm den Einfluss der Motivation, der Unterrichtsqualität, der häuslichen Umgebung und der schulischen Umgebung. Mit Blick auf Merkmale einer (höheren) Quantität des Unterrichts führen Brophy und Good (1986) anhand einer der umfassendsten Aufarbeitung der Literatur neben dem Umfang der Lerngelegenheiten (im Sinne des Umfangs der unterrichteten Inhalte) die Klassenführung (im Sinne einer aktiven Zeitnutzung zwecks Erhöhung der effektiven Lernzeit) an. Mit Blick auf Merkmale einer (höheren) Qualität von Unterricht identifizieren sie Aspekte der Klarheit (z.B. der Lernziele), der Strukturiertheit (z.B. der angemessenen Sequenzierung von Inhalten) und der Kognitiven Aktivierung (z.B. des kognitiven Niveaus der Lehrerfragen), sowie Aspekte weiterer Dimensionen (z.B. Umgang mit Fehlern als Aspekt eines lernförderlichen Klimas). Fraser et al. (1987) zielten in einem breiteren Zugang auf die Fundierung des „Models of Educational Productivity“. In einer Synthese mehrerer Metaanalysen prüften Sie nicht nur die Einflussfaktoren sondern quantifizierten deren Wirkung. Fraser et al. (1987) konnten dabei im Wesentlichen den Einfluss der drei Merkmalsgruppen *Aptitude*, *Instruction*, und *Environment* auf verschiedenen Lernergebnisse bestätigen. Den größten Effekt auf die Schülerleistung finden sie für die Merkmalsgruppe *Aptitude*, in der die kognitiven Fähigkeiten den größten Faktor ausmachen ($r = .44$). In der Merkmalsgruppe *Instruction* hat die Qualität einen leicht größeren Einfluss ($r = .47$) gegenüber der Quantität ($r = .38$), der aber kaum bedeutsam sein dürfte. Die Arbeit von Fraser et al. (1987) ist insbesondere auch deshalb interessant, weil sich die Autoren der Frage nach der Vorhersage von Lernergebnissen nicht nur im Allgemeinen widmen sondern ihr auch speziell im Kontext des naturwissenschaftlichen Unterrichts nachgehen. Mit Bezug auf die Metanalyse von Wise und Okey (1983) identifizieren sie als Merkmale einer hohen Unterrichtsqualität in den Naturwissenschaften vor allem die Klarheit und Strukturiertheit (als Nutzung von Strategien zur Strukturierung des Lernprozesses, $d = .57$) und die kognitive Aktivierung (als Länge der Wartezeit zwischen der Frage der Lehrkraft und einer Antwort der Schülerinnen und Schüler, $d = .90$, oder als Nutzung von *hands-on*-Strategien, $d = .57$).

Zusammengefasst hat die frühe Unterrichtsforschung unter dem Paradigma „Guter Unterricht ist Unterricht, in dem gelernt wird“ eine Vielzahl von Merkmalen identifiziert, die Einfluss auf die Schülerleistung haben. Diese lassen sich in einige wenige Dimensionen der Unterrichtsqualität zusammenfassen. Da wären zunächst die eigentlich unter Quantität von Unterricht geführte Klassenführung zu nennen, als Merkmal eines Unterrichts, der die effektiv genutzte Zeit – die *time-on-task* – durch aktive Maßnahmen der Lehrkraft (z.B. Störungsprävention) optimiert. Weitere Dimensionen sind die Klarheit und Strukturiertheit, die sich häufig nicht eindeutig voneinander trennen lassen, da Strukturiertheit auch die Klarheit positiv beeinflusst und umgekehrt; sowie außerdem die kognitive Aktivierung. Die frühe Forschung hat damit einerseits zentrale Dimensionen der Unterrichtsqualität identifiziert und diese andererseits gut beschrieben. Unklar blieb aber weitgehend, inwieweit diese doch sehr breiten Beschreibungen über verschiedene Fächer hinweg anwendbar sind (bzw. wie sich eine fach-, hier: naturwissenschaftsspezifische Operationalisierung der Dimensionen darstellt), wie man sich das Zusammenspiel der einzelnen Dimensionen vorstellen muss (bzw. ob die kombinierte Verbesserung der Unterrichtsqualität in den einzelnen Dimensionen auch einen kumulativen Effekt auf die Lernleistung oder andere Lernergebnisse hätte).

Die videobasierte Unterrichtsforschung

Mit der zunehmenden Verfügbarkeit und Erschwinglichkeit von Videotechnik erlebte die Unterrichtsqualitätsforschung Ende der 1990er Jahre ein Revival (Petko, Waldis, Pauli, & Reusser, 2003). Die Möglichkeit, Unterricht aufzuzeichnen und wiederholt mit Blick auf unterschiedliche Merkmale zu untersuchen, erlaubte die differenzierte Analyse der komplexen Unterrichtsabläufe. Insbesondere konnten Merkmale im Zusammenspiel und in ihrer Verän-

derung im zeitlichen Verlauf über den Unterricht hinweg analysiert – und damit Unterrichtsmuster – identifiziert werden. Im Rahmen der so genannten Third International Mathematics and Science (TIMS) Videostudie wurde das Verfahren der Videoanalyse eingesetzt um Mathematik- und Naturwissenschaftsunterricht in ausgewählten Ländern zu untersuchen und Unterrichtsmerkmale zu identifizieren, die als Erklärung für die in der TIMS Studie beobachteten Unterschiede in den Schülerleistungen erklären zu können (Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999; Roth et al., 2006). Dazu wurden je teilnehmendem Land eine stratifizierte, randomisierte Stichprobe von ungefähr 100 Schulen gezogen. In diesen Schulen wurde dann zufällig eine Klasse für die Videoaufzeichnung einer Doppelstunde des Mathematik- oder Naturwissenschaftsunterrichts ausgewählt. Wurde die Teilnahme an der Studie verweigert, wurden eine Ersatzschule und –klasse gezogen. Mit den zur Teilnahme bereiten Schulen bzw. Lehrkräften wurde ein Termin vereinbart, so dass die aufgezeichneten Stunden sich mehr oder weniger gleichmäßig über das Schuljahr verteilten. Bei der Aufzeichnung der Unterrichtsstunden wurde mit zwei Videokameras gearbeitet. Die so genannte Aktionskamera folgte dem jeweils aktuellen Unterrichtsgeschehen, filmte also z.B. die Tafel, wenn die Lehrkraft etwas darauf festhielt. Sie wurde von einer Kameraperson bedient. Die so genannte Totalkamera wurde so aufgebaut, dass mit ihr das Geschehen im gesamten Klassenraum aufgezeichnet werden konnte. Die Vergleichbarkeit der Videoaufnahmen wurde durch schriftliche Anweisungen und intensive Trainings der Kamerapersonen erreicht. Die Kodierung erfolgte anhand aufwändig entwickelter und sorgfältige erprobter schriftlicher Kodiermanuale (Petko et al., 2003; siehe auch Fischer & Neumann, 2012).

Die TIMSS Videostudie zum Naturwissenschaftsunterricht umfasste die Aufzeichnung und Analyse von Naturwissenschaftsunterricht in fünf Ländern: Australien, Japan, den Niederlanden, der Tschechoslowakei und den Vereinigten Staaten von Amerika (Roth et al., 2006). Die Analyse beruhte auf Variablen in drei Klassen, die aus einer gründlichen Aufarbeitung der Literatur zur prozess-produkt-orientierten Unterrichtsqualitätsforschung gewonnen wurden: Naturwissenschaftliche Inhalte, Lehrerverhalten und Schülerverhalten. Die Ergebnisse zeigten zunächst, dass in den videografierten Stunden der Lehrervortrag oder ein lehrergeführtes Klassengespräch überwog. In nahezu allen Stunden wurden in der ein oder anderen Form neue Inhalte eingeführt und in einem Großteil der Stunden wurden in irgendeiner Form praktische Aktivitäten der Schülerinnen und Schüler beobachtet, auch wenn die konkreten Anteile zwischen den einzelnen Stunden zum Teil stark variierten. Darüber hinaus zeigten sich – wie zuvor auch schon in den Videostudien des Mathematikunterrichts (Hiebert, 2003; Stigler et al., 1999) – länderspezifische Unterrichtsmuster: Der Unterricht in der Tschechoslowakei war demnach vor allem durch das Klassengespräch über herausfordernden, oft theoretische Inhalte (naturwissenschaftliche Fakten, Konzepte, Theorien, usw.) gekennzeichnet, der Unterricht in den Niederlanden überwiegend durch unabhängiges, selbst-gesteuertes Lernen der Schülerinnen und Schüler. Sowohl in Japan als auch in Australien war der Unterricht in den Naturwissenschaften durch den Fokus auf die evidenzbasierte Erklärung von Phänomenen geprägt. Japanische Schülerinnen und Schüler untersuchten die Phänomene dabei vielfach in selbst-gesteuerten praktischen Aktivitäten. Australische Schülerinnen und Schüler interagierten im Vergleich seltener direkt mit dem Phänomen und arbeiteten in der Folge weniger mit selbst generierten Daten. Der Naturwissenschaftsunterricht in den Vereinigten Staaten ließ sich am besten als Mischung verschiedener Skripte einschließlich der für die anderen Länder beschriebenen charakterisieren (für eine ausführliche Darstellung siehe Roth et al., 2006). Mit Blick auf ein einheitliches Unterrichtsmuster über die vier hochleistenden Länder in der Studie hinweg, kamen Roth et al. (2006) zu dem Schluss, dass ein einheitliches Unterrichtsmuster nicht zu konstatieren ist; wohl aber zwei Unterrichtsmerkmale in allen diesen Ländern zu beobachten sind: Hohe fachliche Standards und eine Fokussierung auf Inhalte, die allerdings jeweils

unterschiedlich implementiert werden z.B. die Dichte und Schwierigkeit der Inhalte oder die Eigenständigkeit der Schülerinnen und Schüler bei der Erarbeitung der Inhalte betreffend.

Obwohl oder womöglich gerade weil die TIMS Videostudie(n) über die Beschreibung von Unterrichtsmustern für hochleistende Länder hinaus keinen Bezug zwischen einem den beobachteten Mustern entsprechenden Unterricht und den Leistungen der Schülerinnen und Schüler in Folge dieses Unterrichts herstellten, erfuhr die auf die Identifikation von Unterrichtsmustern und die Untersuchung deren Wirkung auf die Schülerleistung gerichtete Unterrichtsforschung weiteren Aufschwung. Dabei blieb die Videoanalyse die zentrale Methode. Allerdings wurde die zugrundeliegenden Modelle, die Forschungsdesigns, die Videoaufzeichnung und -analyse selber, sowie die Identifikation von Unterrichtsmustern bzw. Wirkgefügen auf Basis der Modelle weiterentwickelt.

Die Modelle betreffend wurden ausgehend von den im Rahmen der prozess-produkt-orientierten Unterrichtsforschung entwickelten differenzierte Modelle entwickelt. Diese neue Klasse von Modellen stellte den unterrichtlichen Prozess des Lehrens und Lernens in den Mittelpunkt. Sie werden in der Folge auch als Angebots-Nutzungs-Modelle bezeichnet. Neben weiteren Merkmalen, die den Lehr-Lern-Prozess beeinflussen (können), berücksichtigten diese Modelle vor allem auch die Interaktion der verschiedenen Merkmale. Ein Beispiel für ein solches Modell ist das Angebots-Nutzungs-Modell unterrichtlichen Lernens nach Lipowsky (2006; Abbildung 2).

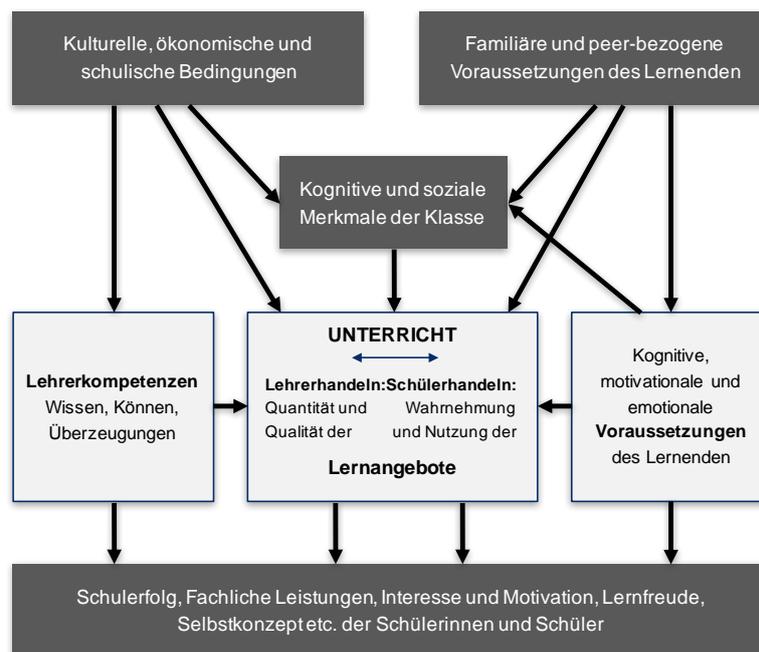


Abbildung 2. Angebots-Nutzungs-Modell unterrichtlichen Lernens nach Lipowsky (2006)

Die Untersuchung von Unterricht auf der Basis dieser weiterentwickelten Modelle erforderte auch eine Weiterentwicklung der Forschungsdesigns. Um über eine Beschreibung von Mustern hinaus die Wirkung der unterrichtlichen Lehr-Lern-Prozesse zu untersuchen, wurden nun vor allem Messwiederholungs-Designs eingesetzt, in denen die Schülerleistung vor und nach

einer Unterrichtseinheit erfasst und ausgewählte Stunden der Einheit videografiert werden (z.B. Seidel et al., 2007). Zusätzlich werden im Rahmen dieser Designs üblicherweise weitere Merkmale der Schülerinnen und Schüler, der Lehrkraft, sowie der Schule erhoben (z.B. Neumann, Fischer, & Sumfleth, 2008). Seidel et al. (2007) führten zudem zwei Neuerungen ein, um entsprechende Limitationen der TIMS Videostudien zu adressieren: 1) Da die Aufzeichnungen in den TIMS Videostudien über das ganze Schuljahr verteilt wurden, variierten die in den videografierten Unterrichtsstunden adressierten Inhaltsbereiche substanziell. Seidel et al. (2007) entschieden sich vor diesem Hintergrund dafür, pro Lehrkraft zwei Unterrichtsstunden zu verschiedenen Inhaltsbereichen zu videografieren und in beiden Fällen eine Unterrichtsstunde zu videografieren, in der neue Inhalte eingeführt werden. 2) In den TIMS Videostudien wurden Spezifika der Schulsysteme, die die Generalisierbarkeit der Befunde einschränken könnten, nicht berücksichtigt. In Deutschland hätte z.B. die Dreigliedrigkeit des Schulsystems berücksichtigt werden müssen, entweder durch eine Beschränkung auf eine Schulform oder durch Ziehen einer Stichprobe, die jeweils Teilstichproben umfasst, die repräsentativ für die einzelnen Schulformen gewesen wären. Da die Berücksichtigung der verschiedenen Schulformen neue Probleme mit sich bringen kann – so werden nicht immer die gleichen Inhalte im gleichen Jahrgang unterrichtet –, beschränkten sich Seidel et al. (2007) auf zwei Schulformen (für eine ausführliche Darstellung siehe Seidel, Prenzel, & Kobarg, 2005).

Neben den zugrundeliegenden Modellen und den Forschungsdesigns wurde auch das Verfahren der Videoaufzeichnung und -analyse selber weiterentwickelt. Dies betrifft einerseits die Aufnahmetechnik und andererseits die Kodiersysteme (Fischer & Neumann, 2012). Neben stärker standardisierten Videoaufzeichnungen konnte dadurch vor allem die Objektivität (d.h. die Interraterreliabilität) der inzwischen überwiegend hoch-inferenten Kodiersysteme verbessert werden. So wurden qualitativ hochwertige Kodiersysteme unter anderem für die Strukturiertheit des Unterrichts (z.B. Brückmann, 2009; Wackermann, 2008) und die kognitive Aktivierung (z.B. Lau, 2011; Neumann et al., 2008) entwickelt. Diese Kodiersysteme erlauben eine ausreichend objektive Messung für einen Vergleich verschiedener Unterrichtsstunden. Mit diesen Weiterentwicklungen entwickelte sich die Videoanalyse zu der (wenn auch aufwändigen) Standardmethode für die Analyse von Unterricht und löste direkte Unterrichtsbeobachtungen weitgehend ab – eine Entwicklung, die sich mit der immer günstigeren Videotechnik bis heute weiter fortsetzt, selbst bzw. gerade bei groß angelegten Vergleichsstudien (vgl. Fischer & Neumann, 2012).

Nicht zuletzt wurden auch die Verfahren zur Identifikation von Unterrichtsmustern (z.B. Hugener et al., 2009) bzw. zur Untersuchung von Wirkgefügen (z.B. Rakoczy, Harks, Klieme, Blum, & Hochweber, 2013) weiterentwickelt. Dies betrifft insbesondere die zunehmende Nutzung neuerer statistischer Verfahren wie der Pfadanalyse oder der Strukturgleichungsmodellierung. Wie diese Weiterentwicklungen im Zusammenspiel mit den Weiterentwicklungen im Bereich der zugrundeliegenden Modelle, der Forschungsdesigns, sowie die Videoaufzeichnung und -analyse zur Gewinnung neuer Erkenntnisse über die Unterrichtsqualität (in den Naturwissenschaften) beitragen, und welche Limitationen immer noch bestehen, soll im Folgenden am Beispiel des Projekts „Quality of Instruction in Physics“ (QuIP) dargestellt werden.

Das Projekt QuIP war als internationale Vergleichsstudie des Physikunterrichts in drei Ländern (Deutschland, Finnland und der Schweiz) angelegt. Die theoretische Basis markierte dabei ein Modell der Unterrichtsqualität in Physik in Anlehnung an das Angebots-Nutzungsmodell unterrichtlichen Lernens nach Lipowsky (2006). Dieses Modell beschreibt neben verschiedenen Merkmalen der Schülerinnen und Schüler sowie der Lehrkraft Merkmale der Unterrichtsqualität auf der Ebene der Oberflächen- und der Tiefenstruktur des Unterrichts sowie drei für den Physikunterricht zentrale Lernergebnisse: Fachwissen, Experimentierfähigkeiten,

sowie Interesse und Motivation (Fischer, Neumann, Labudde, & Viiri, 2014; Abbildung 3). Auf Grundlage dieses Modells sollten im Projekt QuIP Unterrichtsmuster identifiziert und unter Berücksichtigung von Rahmenbedingungen zur Entwicklung des Fachwissens, der Experimentierfähigkeiten, sowie des Interesses und der Motivation der unterrichteten Schülerinnen und Schüler in Beziehung gesetzt werden sollen.

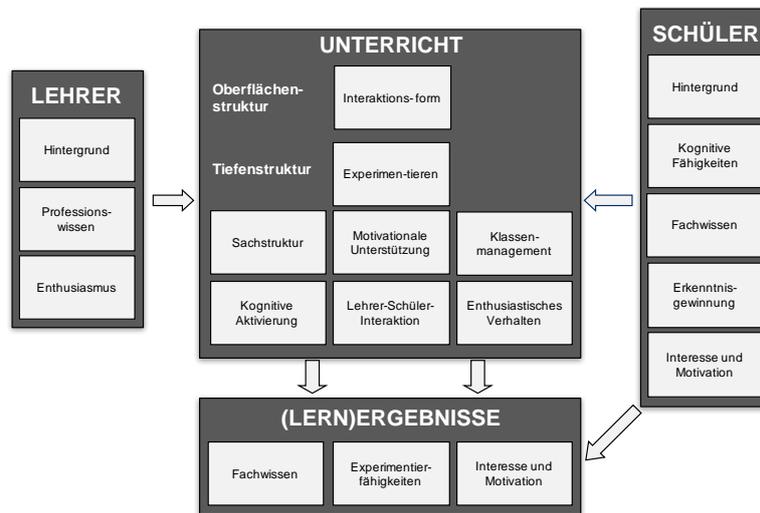


Abbildung 3. Unterrichtsqualität in Physik (Fischer, Neumann, et al., 2014)

Im Zentrum der im Prä-Post-Design angelegten Untersuchung stand die Videoaufzeichnung und -analyse von zwei aufeinanderfolgenden Unterrichtsstunden. Um die Unterrichtsvideos vergleichbar zu halten, wurde das Thema des zu videografierten Unterrichts auf den „Zusammenhang von elektrischer Energie und Leistung“ festgelegt. Aufgrund der Themenwahl wurde die Erhebung in Finnland und der Schweiz in der 9. Jahrgangsstufe, in Deutschland in der 10. Jahrgangsstufe durchgeführt. In jeder teilnehmenden Klasse wurde zu Beginn des Halbjahres, in dem die Videoaufzeichnung stattfand, eine Prä-Testung, am Ende des Halbjahres eine Post-Testung durchgeführt. Insgesamt wurden 47 deutsche, 25 finnische und 31 schweizerische Klassen untersucht. Für Deutschland wurde dabei darauf geachtet, die verschiedenen Schulformen im Verhältnis angemessen abzubilden. Im Rahmen der Prä-Testung wurden die kognitiven Fähigkeiten, die Motivation und das Vorwissen erfasst, sowie die Fähigkeit der Schülerinnen und Schüler, Experimente zu planen, und der familiäre Hintergrund der Schülerinnen und Schüler. Auf Seite der Lehrperson wurden deren Enthusiasmus, Aspekte ihres pädagogischen Wissens sowie der Ausbildungshintergrund erhoben. Im Rahmen der Post-Testung wurde das Wissen der Schülerinnen und Schüler sowie die experimentellen Fähigkeiten als wesentliche Aspekte von Kompetenz sowie außerdem die Motivation erhoben; von der Lehrperson das Fach- und das fachdidaktische Wissen zum Thema Elektrizitätslehre. Zur Analyse wurden in Abhängigkeit von der konkreten Fragestellung unterschiedliche Verfahren eingesetzt von Regressions-, über Pfad-, bis hin zu Strukturgleichungsanalysen (z.B. Geller, Neumann, Boone, & Fischer, 2014; Keller, Neumann, & Fischer, 2017; für einen Überblick siehe Fischer, Labudde, Neumann, & Viiri, 2014)

Die Ergebnisse bestätigten zunächst einmal auch für den Lernzuwachs in der untersuchten Unterrichtseinheit die Überlegenheit der finnischen Schülerinnen und Schüler (Geller et al 2014). Die beobachteten Unterschiede entsprechen dabei ziemlich genau den Unterschieden,

die man aufgrund der Befunde der internationalen Vergleichsstudien wie TIMSS oder PISA erwarten durfte. Diese Ergebnisse wurden im Projekt umfassend zu Unterrichtsmerkmalen auf verschiedenen Ebenen in Beziehung gesetzt. Hier sollen nur beispielhaft ausgewählte Erkenntnisse zu den zentralen Dimensionen der Unterrichtsqualität Klassenführung, Klarheit und Strukturiertheit sowie kognitive Aktivierung dargestellt werden.

Mit Blick auf die Klassenführung wurden in Anlehnung an einschlägige Konzeptionen der Klassenführung die Merkmale Zeitnutzung, Regelklarheit, Instruktionklarheit, Monitoring, Disziplinprobleme, Störungsprävention und time-on-task untersucht. Das Merkmal der Zeitnutzung unterschied sich dabei insofern vom Merkmal time-on-task als die Zeitnutzung beschreibt, in welchem Maße der Unterricht konsistent ist, ohne Verzögerung abläuft und ob die Zeit optimal für das Lehren und Lernen genutzt wird. Das Merkmal Time-on-task auf der anderen Seite bezieht sich auf die tatsächliche Zeit, die für den Lernprozess zur Verfügung stand. Es ist damit ein quantitatives Maß, während die Zeitnutzung eher ein qualitatives Maß darstellt. Die Ergebnisse zur Wirkung der Klassenführung auf die Schülerleistung sind in Tabelle 1 dargestellt (eine detaillierte Beschreibung wie die Ergebnisse zustande gekommen sind findet sich bei Junge, von Arx, & Labudde, 2014).

Tabelle 1. Einfluss der Klassenführung auf den Lernzuwachs (Junge et al., 2014).

| Kategorie | F | p | η^2 |
|---------------------|----------|----------|----------------------------|
| Zeitnutzung | 4.462 | .014 | .097 |
| Regelklarheit | 3.894 | .024 | .087 |
| Instruktionklarheit | 1.792 | .173 | .040 |
| Monitoring | 1.694 | .190 | .039 |
| Disziplinprobleme | 0.682 | .509 | .017 |
| Störungsprävention | 3.520 | .034 | .078 |
| Time-on-Task | 2.921 | .059 | .066 |

Die Ergebnisse zeigen statistisch bedeutsame Effekte für die die Regelklarheit, die Störungsprävention und die Zeitnutzung. Interessant ist hier insbesondere das Ergebnis, dass die Zeitnutzung als qualitatives Maß einen bedeutsameren (und größeren) Einfluss auf die Schülerleistung hat als time-on-task als quantitatives Merkmal. Dies belegt, dass die Klassenführung nicht ausschließlich ein quantitatives Merkmal ist, sondern dass vielmehr qualitative Merkmale der Klassenführung, die letztendlich zu einer erhöhten Lernzeit und einer besseren Nutzung dieser Lernzeit führen, berücksichtigt werden müssen.

Die Klarheit und Strukturiertheit des Unterrichts wurde von Helaakoski und Viiri (2014) in Anlehnung an das von Brückmann (2009) verwendete Verfahren der Sachstrukturanalyse untersucht. Schrittweise wurden die aufgezeichneten Lehrer-Schüler-Dialoge in konzeptuelle Netzwerke (so genannte Concept Maps, Novak, 1995) übersetzt. Anhand dieser Concept Maps wurden anschließend Parameter einer stärkeren Konnektivität, d.h. einer stärkeren Verknüpfung von Begriffen im Unterricht bestimmt. Dabei zeigte sich einerseits, dass im finnischen Physikunterricht gegenüber dem deutschen Physikunterricht Inhalte stärker zueinander in Beziehung gesetzt werden, und andererseits, dass sowohl eine größere Zahl von Konzepten, aber auch die stärkere Verknüpfung dieser Konzepte zu einem höheren Lernzuwachs der Schülerinnen und Schüler führt (Abbildung 4; für weitere Details siehe Helaakoski & Viiri, 2014).

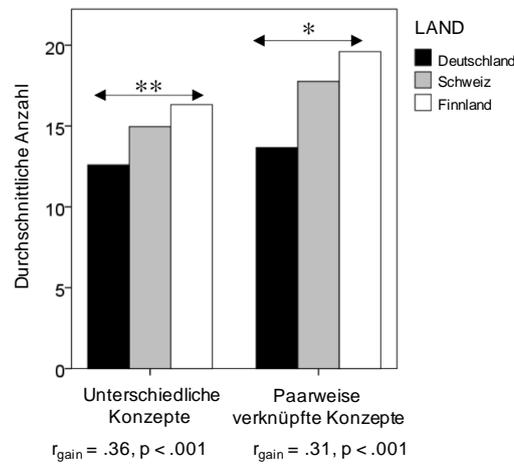


Abbildung 4. Einfluss der Strukturiertheit auf den Lernzuwachs (Helaakoski & Viiri, 2014).

Die kognitive Aktivierung als drittes Merkmal der Unterrichtsqualität, das hier besprochen werden soll, wurde im QuIP-Projekt als der prozentuale Anteil der Aufgaben gewertet, bei denen die Schülerinnen und Schüler in ihrer Reaktion auf die Aufgabenstellung durch die Lehrkraft das von der Lehrkraft geforderte Niveau auch erreichen (vgl. Ergöneç, Neumann, & Fischer, 2014). Die Auswertung zeigt statistisch bedeutsame Unterschiede zwischen den drei Ländern Finnland, Deutschland und der Schweiz (Abbildung 5). Der deutsche Physikunterricht besitzt dabei die niedrigsten Anteile bei denen eine Passung zwischen kognitiver Anforderung der Aufgabenstellung und dem kognitiven Niveau der Bearbeitung durch die Schülerinnen und Schüler (kurz: Passung) gegeben ist, der finnische Unterricht die höchsten Anteile. Darüber hinaus zeigt sich ein statistisch bedeutsamer Zusammenhang zwischen der in dieser Form operationalisierten kognitiven Aktivierung und der Leistung der Schülerinnen und Schüler. Auffällig sind allerdings die vergleichsweise hohen absoluten Anteile der Passung, die dafür sprechen, dass die Lehrkräfte das kognitive Niveau der Aufgaben möglicherweise niedriger ansetzen, als nötig.

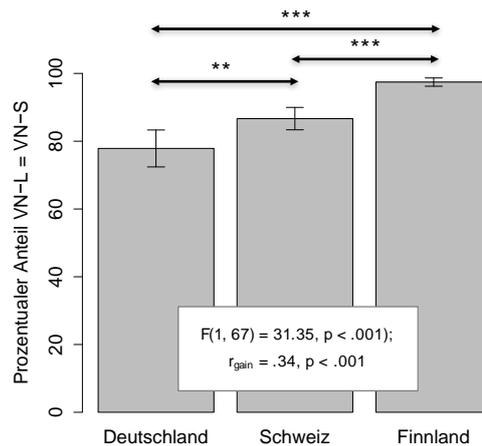


Abbildung 5. Einfluss der kognitiven Aktivierung auf den Lernzuwachs

In weiteren Analysen konnten im Projekt QuIP differenzierte Erkenntnisse über die Unterrichtsqualität in Physik gewonnen werden (Fischer, Labudde, et al., 2014). Dabei zeigte sich z.B. die besondere Bedeutung des Enthusiasmus der Lehrkraft für die Motivation der Schülerinnen und Schüler (Keller et al., 2017). Die Erkenntnisse stimmen weitgehend mit den Erkenntnissen anderer Studien zum deutschen Physikunterricht überein; die zeigen zunächst, dass der Physikunterricht durch ein durch die Lehrkraft eng geführtes, fragend-entwickelndes Unterrichtsgespräch gekennzeichnet ist (Seidel et al., 2007). Sie belegen auch einen positiven Effekt z.B. eines anspruchsvollen Klassengesprächs (d.h. einer hohen kognitiven Aktivierung) auf affektive Merkmale der Schülerinnen und Schüler (Seidel, Rimmel, & Prenzel, 2003).

Das QuIP Projekt ist damit in gewisser Weise prototypisch für die videobasierte Unterrichtsforschung der letzten Jahre. Diese Forschung bestätigt zunächst die klassischen Dimensionen der Unterrichtsqualität: Klarheit und Strukturiertheit, kognitive Aktivierung und Klassenführung. Dabei wurden die einzelnen Dimensionen jeweils fachspezifisch operationalisiert, so dass nicht nur die empirische Basis die Bedeutung dieser Dimensionen für Lernen erweitert wurde, sondern auch fachspezifische Erkenntnisse gewonnen wurden, durch welche Merkmale z.B. kognitiv aktivierender Unterricht in der Physik gekennzeichnet ist. Die Arbeiten zur Unterrichtsqualität der letzten Jahre sind zudem stärker multi-kriterial angelegt. Das heißt, es werden über das Fachwissen hinaus weitere Zielkriterien des (naturwissenschaftlichen) Unterrichts in den Blick genommen, wie z.B. Interesse und Motivation der Schülerinnen und Schüler. Obwohl diese Arbeiten damit ohne Frage sowohl theoretisch als auch methodisch zur Weiterentwicklung der Unterrichtsqualitätsforschung beigetragen haben, wurde ein zentrales Versprechen nicht oder nicht wirklich eingelöst: das Versprechen, über die verschiedenen Merkmale hinweg Muster guten Unterrichts zu identifizieren. Im Gegenteil zeigen sich teilweise sogar methodische Schwierigkeiten. So führt die Berücksichtigung verschiedener Unterrichtsmerkmale bei der Erforschung der Wirkung des fachdidaktischen Wissens auf die Schülerleistung bei Ergönenç et al. (2014) dazu, dass die Bedeutsamkeit einzelner Pfade im Wirkmodell zurückgeht (S. 155). Hier zeigt sich, dass die Auflösung komplexer Wirkmodelle mit den herkömmlichen Forschungsdesigns und Analysemethoden nicht ohne Weiteres zu bewerkstelligen ist. Zudem stellt sich die Frage inwieweit möglicherweise jenseits der klassischen Dimensionen der Unterrichtsqualität (in ihrer naturwissenschaftsspezifischen Operationalisierung) stärker auch naturwissenschaftsspezifische Dimensionen der Unterrichtsqualität berücksichtigt werden sollten. Eine systematische Forschung hierzu fehlt bisher weitgehend.

Die Unterrichtsqualitätsforschung (in den Naturwissenschaften) heute

Ein Plädoyer für eine stärkere Berücksichtigung fachspezifischer Dimensionen der Unterrichtsqualität findet sich in gewisser Weise bereits bei Seidel und Shavelson (2007). Sie können in ihrer Meta-Analyse im Vergleich die größten Effekte für fachspezifische Lerngelegenheiten wie z.B. *scientific inquiry* nachweisen; das heißt für Lerngelegenheiten, die möglichst dicht an angestrebten Lernergebnissen sind. Dies wirft die Frage auf, ob nicht die Qualität der Umsetzung von Unterrichtskonzeptionen wie z.B. der des *inquiry-based learning* als Dimension der Unterrichtsqualität in den Naturwissenschaften in den Blick genommen werden sollten – zumindest, wenn Kompetenzen im Bereich Erkenntnisgewinnung angestrebt werden. Dabei stellt sich die Frage, ob die klassischen Dimensionen der Unterrichtsqualität weiterhin einfach anwendbar sind oder möglicherweise in Abhängigkeit von der jeweiligen Unterrichtskonzeption anders interpretiert werden müssen. Für eine individuelle Interpretation der klassischen Dimensionen spricht die Arbeit von Praetorius et al. (2014). In dieser Arbeit weisen die Autoren mittels *generalizability theory* nach, dass die klassischen Dimensionen der Unterrichtsqualität unterschiedlich stabil über verschiedene Unterrichtsstunden hinweg sind. Insbesondere zeigt sich, dass kognitive Aktivierung als mit Blick auf die angestrebten Lernergebnisse proximalere Dimension der Unterrichtsqualität gegenüber der Klassenführung als

distalere Dimension über mehrere Unterrichtsstunden hinweg stärker variiert. Während die Klassenführung bereits nach einer Unterrichtsstunde ausreichend reliabel eingeschätzt werden kann, erfordert die reliable Erfassung der kognitiven Aktivierung nach Praetorius et al. (2014) mindestens neun Unterrichtsstunden. Praetorius et al. (2014) diskutieren drei mögliche Erklärungsansätze: 1) Kognitive Aktivierung ist abhängig vom behandelten Inhalt. Das heißt, die Einführung des Kraftbegriffs erlaubt möglicherweise eine höhere oder zumindest andere kognitive Aktivierung als die Erarbeitung des Reflexionsgesetzes. 2) Kognitive Aktivierung hängt vom Lernziel ab. Die Einführung neuen Wissens dürfte z.B. eine höhere oder zumindest andere kognitive Aktivierung erlauben, als das Einüben von Routinen. 3) Kognitive Aktivierung hängt vom Unterrichtskonzept ab. Die Einführung neuer Konzepte durch *inquiry-based learning* könnte eine höhere oder zumindest andere Form von kognitiver Aktivierung erlauben als die Einführung eines neuen Konzepts durch Lehrervortrag. Zusammen mit den Erkenntnissen von Seidel und Shavelson (2007) wirft dies die Frage auf, ob man Unterrichtsqualität stärker vor dem Hintergrund des jeweiligen Unterrichtskonzepts bzw. Lernziels oder Inhalts betrachten und dabei neben den klassischen auch fachspezifische Merkmale berücksichtigen muss.

Diese Idee ist nicht neu: Bereits Fischer, Reyer, Wirz, Bos und Höllrich (2002) erfassen die Lehrziele, die Lehrkräfte verfolgen, und vermuten, dass die Erarbeitung der einzelnen Lehrziele spezifischen Mustern folgt (siehe auch Reyer, 2003). Und auch die Vorgehensweise von Brückmann (2009) und Wüsten, Schmelzing, Sandmann und Neuhaus (2010) legt nahe, dass die Autoren jeweils vermuten, die Qualität von Unterrichtsstrukturen sei abhängig vom konkreten Inhalt oder müsse zumindest abhängig vom konkreten Inhalt bewertet werden. Bisher fehlt es jedoch prinzipiell an Maßen, die es erlauben würden, die Qualität von Unterrichtsstrukturen zu untersuchen; einfach, weil generische Maße wie das der Zentralität oder der Zahl der Verknüpfungen sich als nicht geeignet erwiesen haben, die Qualität von Unterrichtsstrukturen über verschiedene (z.T. nicht einmal innerhalb einzelner) Inhalte hinweg zu bewerten oder weil es an spezifischen Maßen für die Unterrichtsstrukturen zu einzelnen Inhalten fehlt. Der Versuch, aus der Deskription von erfolgreichem Unterricht – d.h. Unterricht, der zu einem hohen Lernzuwachs der Schülerinnen und Schüler führt – Kriterien herauszudestillieren, dürfte hier ob der schierigen Menge unterschiedlicher Inhalte kaum zielführend sein. Erfolgversprechender *dürfte der Ansatz sein, für unterschiedliche Unterrichtskonzeptionen wie z.B. inquiry-based learning* (z.B. National Science Foundation [NSF], 1996), *project-based learning* (z.B. Krajcik & Czerniak, 2014) oder *context-based learning* (Demuth, Parchmann, & Ralle, 2007) Merkmale einer qualitativ hochwertigen Umsetzung zu beschreiben.

Eine Schwierigkeit dürfte die Tatsache sein, dass naturwissenschaftsdidaktische Unterrichtskonzeptionen in der Realität, zumindest in Reinform, kaum zu beobachten sind. Hier bieten sich verschiedene Optionen an – z.B. können die Unterrichtskonzepte anhand der Merkmale einer qualitativ hochwertigen Umsetzung implementiert und im Rahmen von systematischen Interventionen evaluiert werden (z.B. Schulz, 2011). Alternativ können Lehrkräfte im Hinblick auf eine qualitativ hochwertige Umsetzung der Unterrichtskonzeptionen fortgebildet werden (z.B. Trendel, Wackermann, & Fischer, 2008). Ein weiteres Problem ist die Identifikation von Mustern. Auch die systematische Intervention oder die Fortbildung von Lehrkräften können nicht garantieren, dass eine qualitativ hochwertige Umsetzung hinsichtlich aller Merkmale erreicht wird. Welches sind also kritische Merkmale? Welche Muster von Merkmalen charakterisieren eine qualitätsvolle Umsetzung? Welchen Einfluss haben Rahmenbedingungen? Die Beantwortung dieser Fragen erfordert weitere empirische Forschung. Dabei stellt sich das Problem, dass sich auch der Unterrichtsforschung der vergangenen Jahre stellte – die Frage, wie die vielen verschiedenen erhobenen Merkmale sinnvoll in Unterrichtsmuster kombiniert werden können. Hier erscheint es lohnenswert, das Potenzial des Methodenreper-

toire der Big Data auszulösen; das heißt, Methoden zur Reduktion der Komplexität und Erkennung von Mustern in großen Datenmengen zu nutzen. Dieser Zugang könnte es auch ermöglichen, Verläufe von bestimmten Unterrichtsmerkmalen (z.B. des kognitiven Niveaus der von der Lehrkraft angebotenen Aufgaben) zu analysieren. Dieser Ansatz wurde bisher nicht wirklich erfolgreich verfolgt. Nicht zuletzt sollten auch bei der Zusammenführung von Mustern und Schülerleistung neue Wege exploriert werden. Sicher haben Pfadanalysen und Strukturgleichungsmodellierungen neue Möglichkeiten eröffnet (oder im Hinblick auf mehrerebenenanalytische Zugänge auch ganz konkrete Limitationen klassischer Analyseverfahren behoben). Sie sind sicher auch noch nicht ausgereizt. Dennoch stoßen auch diese Zugänge an Grenzen, wenn die Zahl der berücksichtigten Merkmale zu groß und die modellierten Wirkgefüge zu komplex werden. Auch hier sollten entsprechend neue Verfahren wie z.B. *Cognitive Diagnostic Modeling* oder innovative Anwendungen bestehender Verfahren wie z.B. *Latent Class Analysen* – natürlich immer orientiert an der Fragestellung – exploriert werden.

Zusammenfassung und Ausblick

Zusammengefasst hat die Unterrichtsqualitätsforschung zunächst in unzähligen Studien eine Vielzahl verschiedener Unterrichtsmerkmale identifiziert. Durch Metaanalysen (z.B. Sweitzer & Anderson, 1983) und Synthesen von Metaanalysen (z.B. Fraser et al., 1987) konnten aus diesen Merkmalen einige wenige Dimensionen der Unterrichtsqualität herausgearbeitet werden; darunter Klarheit und Strukturiertheit, kognitive Aktivierung und Klassenführung. Die sich rasant entwickelnde Videotechnik eröffnete schließlich die Möglichkeit, Unterrichtsstunden hinsichtlich aller dieser Dimensionen zu untersuchen und Unterrichtsmuster zu beschreiben, um diese anschließend – unter Berücksichtigung von Rahmenbedingungen des Unterrichts – zu Lernergebnissen wie naturwissenschaftlichem Wissen, experimentellen Fähigkeiten, sowie dem Interesse an und der Motivation zur Auseinandersetzung mit den Naturwissenschaften in Beziehung zu setzen. In verschiedenen Studien ist es gelungen, die genannten Dimensionen der Unterrichtsqualität für die Naturwissenschaften in einer fachspezifischen Operationalisierung nachzuweisen und damit Erkenntnisse darüber zu gewinnen, wie kognitiv aktivierender Physikunterricht gestaltet sein könnte. Das Versprechen, Unterrichtsmuster zu beschreiben und damit Muster guten Physikunterrichts wie z.B. das eines klar strukturierten, kognitiv aktivierenden, gleichzeitig aber wenig Klassenführung erfordernden Physikunterrichts in ihrer positiven Wirkung auf Lernen zu bestätigen, wurde nicht eingelöst. Dies liegt darin begründet, dass es an Methoden mangelt, entsprechende Muster in geeigneter Weise zu identifizieren, aber auch an Theorien darüber, welche Muster erfolgversprechend sein könnten und warum – nicht zuletzt, weil es der naturwissenschaftsdidaktischen Forschung an fachspezifischen Theorien der Unterrichtsqualität fehlt. Die Entwicklung solcher Theorien kann mit großer Wahrscheinlichkeit nicht *bottom-up* aus der Deskription lernförderlichen oder etwas motivierenden Unterrichts gelingen. Sie muss vielmehr mit der Identifikation typischer naturwissenschaftlicher oder sogar chemie- oder physikspezifischer Unterrichtskonzeptionen einhergehen, gefolgt von der Formulierung spezifischer Merkmale einer qualitätsvollen Umsetzung dieser Unterrichtskonzeptionen. Diese könnten durch sorgfältig geplante experimentelle Studien sowie groß angelegte Feldstudien validiert werden. Dafür ist auch eine methodische Weiterentwicklung notwendig. Dabei beginnt die naturwissenschaftsdidaktische Forschung in keiner Weise bei Null; sie kann vielmehr auf eine lange Tradition entsprechender Arbeiten zurückgreifen, die eine gute theoretische und methodische Basis bilden, die aber eben auch Raum für die dringend notwendige Weiterentwicklung unter der Prämisse einer stärkeren Fachspezifität lässt. Oder, um es mit Leuders (2007) zu formulieren: „Ein Konzept der Unterrichtsqualität wird erst dann zu einem fachdidaktischen, indem es Bezug auf die Spezifika des Unterrichtsfaches aufnimmt“ (S. 205).

Literatur

- Anderson, L. (1981). Instruction and Time-on-Task: a Review. *Journal of Curriculum Studies*, 13(4), 289–303.
- Armento, B. J. (1977). Teacher behaviors related to student achievement on a social science concept test. *Journal of Teacher Education*, 28(2), 46–52.
- Brophy, J. E., & Good, T. L. (1986). Teacher behaviour and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York: Macmillan.
- Brückmann, M. (2009). *Sachstrukturen im Physikunterricht: Ergebnisse einer Videostudie*. Berlin: Logos-Verl.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*.
- Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher*, 18(1), 26–31.
- Coker, H., Medley, D. M., & Soar, R. S. (1980). How valid are expert opinions about effective teaching? *The Phi Delta Kappan*, 62(2), 131–149.
- Demuth, R., Parchmann, I., & Ralle, B. (2007). *Chemie im Kontext: Handreichungen für den Unterricht*. Berlin: Cornelsen.
- Dunkin, M. J. (1978). Student characteristics, classroom processes, and student achievement. *Journal of Educational Psychology*, 70(6), 998.
- Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. Holt, Rinehart & Winston.
- Ergöneç, J., Neumann, K., & Fischer, H. E. (2014). The impact of pedagogical content knowledge on cognitive activation and student learning. *Quality of Instruction in Physics*, 145–160.
- Fischer, H. E., Labudde, P., Neumann, K., & Viiri, J. (2014). *Quality of Instruction in Physics: Comparing Finland, Switzerland and Germany*. Waxmann Verlag.
- Fischer, H. E., & Neumann, K. (2012). Video analysis as a tool for understanding science instruction. *The World of Science Education*, 115–140.
- Fischer, H. E., Neumann, K., Labudde, P., & Viiri, J. (2014). Theoretical framework. *Quality of Instruction in Physics—comparing Finland, Germany and Switzerland*, 13–30.
- Fischer, H. E., Reyer, T., Wirz, C., Bos, W., & Höllrich, N. (2002). Unterrichtsgestaltung und Lernerfolg im Physikunterricht. *Zeitschrift Für Pädagogik*, 45, 124–138.
- Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of Educational Productivity Research. *International Journal of Educational Research*, 11, 145–252.
- Geller, C., Neumann, K., Boone, W. J., & Fischer, H. E. (2014). What makes the Finnish different in science? Assessing and comparing students' science learning in three countries. *International Journal of Science Education*, 36(18), 3042–3066.
- Good, T. L., & Grouws, D. A. (1977). Teaching effects: A process-product study in fourth-grade mathematics classrooms. *Journal of Teacher Education*, 28(3), 49–54.
- Helaakoski, J., & Viiri, J. (2014). Content and content structure of physics lessons and students' learning gains: Comparing Finland, Germany and Switzerland. *Quality of Instruction in Physics. Comparing Finland, Germany and Switzerland*, 93–110.
- Hiebert, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. Washington, DC: U.S. Gov. Print. Off. Retrieved from <http://purl.access.gpo.gov/GPO/LPS31989>
- Hugener, I., Pauli, C., Reusser, K., Lipowsky, F., Rakoczy, K., & Klieme, E. (2009). Teaching patterns and learning quality in Swiss and German mathematics lessons. *Learning and Instruction*, 19(1), 66–78.
- Junge, C., von Arx, M., & Labudde, P. (2014). 10. Classroom Management. *Quality of Instruction in Physics: Comparing Finland, Switzerland and Germany*, 161.
- Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest: PHYSICS TEACHERS' KNOWLEDGE AND MOTIVATION. *Journal of Research in Science Teaching*, 54(5), 586–614. <https://doi.org/10.1002/tea.21378>
- Krajcik, J. S., & Czerniak, C. M. (2014). *Teaching Science in Elementary and Middle School: A Project-Based Approach* (4th ed.). Routledge.
- Lau, A. (2011). *Passung und vertikale Vernetzung im Chemie- und Physikunterricht*. Berlin: Logos.
- Leuders, T. (2007). Fachdidaktik und Unterrichtsqualität im Bereich Mathematik. *Arnold, K.-H. (Hg.): Unterrichtsqualität Und Fachdidaktik. Bad Heilbrunn: Klinkhardt*, 205–234.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an: Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda (Ed.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (pp. 47–70). Weinheim; Basel: Beltz.
- National Science Foundation [NSF]. (1996). *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology*. Arlington, VA: Author.
- Neumann, K., Fischer, H. E., & Sumfleth, E. (2008). Vertikale Vernetzung und kumulatives Lernen im Chemie- und Physikunterricht. *Pädagogische Professionalität Als Gegenstand Empirischer Forschung*, 141–151.

- Novak, J. D. (1995). Concept mapping: A strategy for organizing knowledge. In S. M. Glynn & R. Duit (Eds.), *Learning science in the schools: Research reforming practice* (pp. 229–245). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Petko, D., Waldis, M., Pauli, C., & Reusser, K. (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik. *ZDM*, 35(6), 265–280.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, 27, 63–73. <https://doi.org/10.1016/j.learninstruc.2013.03.002>
- Reyer, T. (2003). *Oberflächenmerkmale und Tiefenstrukturen im Unterricht: Exemplarische Analysen im Physikunterricht der gymnasialen Sekundarstufe*. Berlin: Logos-Verl. Retrieved from <http://www.worldcat.org/oclc/723466988>
- Rosenshine, B., & Furst, N. (1971). Research on Teacher Performance Criteria. In B. O. Smith (Ed.), *Research in teacher education - a symposium*. Englewood Cliffs; Nj: Prentice-Hall.
- Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T., ... Gallimore, R. (2006). *Teaching science in five countries: Results from the TIMSS 1999 video study statistical analysis report*. (NCES 2006-011). U.S. Department of Education; National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Retrieved from <http://nces.ed.gov>
- Schulz, A. (2011). *Experimentierspezifische Qualitätsmerkmale im Chemieunterricht: Eine Videostudie* (Vol. 113). Logos Verlag Berlin GmbH.
- Seidel, T., Prenzel, M., & Kobarg, M. (2005). *How to run a video study*. Münster: Waxmann.
- Seidel, T., Prenzel, M., Rimmel, R., Herweg, C., Kobarg, M., & Schwindt, K. (2007). Science teaching an learning in German physics classrooms. In M. Prenzel (Ed.), *Studies in the educational quality of schools. The final report on the DFG Priority Programme* (pp. 79–99). Münster: Waxmann.
- Seidel, T., Rimmel, R., & Prenzel, M. (2003). Gelegenheitsstrukturen beim Klassengespräch und ihre Bedeutung für die Lernmotivation. *Unterrichtswissenschaft*, 31(2), 142–165.
- 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), 454–499.
- Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study. Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan and the United States*. Washington D.C.: U.S. Department of Education.
- Sweitzer, G. L., & Anderson, R. D. (1983). A meta-analysis of research on science teacher education practices associated with inquiry strategy. *Journal of Research in Science Teaching*, 20(5), 453–466.
- Trendel, G., Wackermann, R., & Fischer, H. E. (2008). Lernprozessorientierte Fortbildung von Physiklehrern. *Zeitschrift Für Pädagogik*, 54(3), 322–340.
- Wackermann, R. (2008). *Überprüfung der Wirksamkeit eines Basismodell-Trainings für Physiklehrer*. Berlin: Logos.
- Walberg, H. J. (1981). A psychological theory of educational productivity. In F. H. Farley & N. Gordon (Eds.), *Psychology and education: The state of the union* (pp. 81–108). Berkeley: McCutchan.
- Walberg, H. J., Haertel, G. D., Pascarella, E., Junker, L. K., & Boulanger, F. D. (1981). Probing a model of educational productivity in science with national assessment samples of early adolescents. *American Educational Research Journal*, 18(2), 233–249.
- Wise, K. C., & Okey, J. R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. *Journal of Research in Science Teaching*, 20(5), 419–435.
- Wüsten, S., Schmelzing, S., Sandmann, A., & Neuhaus, B. J. (2010). Sachstrukturdiagramme – eine Methode zur Erfassung inhaltspezifischer Merkmale der Unterrichtsqualität. *Zeitschrift Für Didaktik Der Naturwissenschaften*, 16, 7–23.