

„WIE MISST MAN GUTEN SPORTUNTERRICHT?“ – ERFASSUNG ZENTRALER DIMENSIONEN VON UNTERRICHTS- QUALITÄT IM SCHULFACH SPORT

von Christian Herrmann, Sara Seiler, Uwe Pühse und Erin Gerlach

ZUSAMMENFASSUNG | Die Unterrichtsqualität nimmt für den Lernerfolg von Schülern einen zentralen Stellenwert ein (Hattie, 2008). Während für die Kernfächer bereits validierte Erhebungsinstrumente vorliegen, steht dies für das Fach Sport noch aus. Im vorliegenden Beitrag werden etablierte Fragebogeninventare aus den Kernfächern zur Erfassung der Unterrichtsdimensionen *Klassenführung* und *Schülerorientierung* auf das Fach Sport angepasst und empirisch geprüft. Die Testbatterie umfasst acht Merkmale guten Unterrichts, welche die beiden Unterrichtsdimensionen adäquat abbilden. Die Konstruktvalidität und die psychometrischen Qualitätskriterien konnten über explorative und konfirmatorische Faktorenanalysen bestätigt werden. Mit dieser entwickelten Testbatterie kann die Unterrichtsqualität aus Schülersicht ökonomisch und valide erfasst werden.

Schlüsselwörter: Unterrichtsmerkmale, Klassenführung, Schülerorientierung, Faktorenanalyse.

„HOW TO MEASURE QUALITATIVE PHYSICAL EDUCATION?“ – ASSESSMENT OF CENTRAL DIMENSIONS OF TEACHING QUALITY IN PHYSICAL EDUCATION

ABSTRACT | High quality teaching is a vital component for the successful learning of students (Hattie, 2008). Validated measurement instruments for the core subject areas of teaching already exist, however this is not the case for Physical Education (PE). In this article, the established and validated instruments of the core subjects are adjusted to better suit Physical Education and they are then empirically tested. The validated instruments have been shown to effectively capture two teaching dimensions; *classroom management* and *student orientation*. The test battery for validating a PE-specific measurement instrument consists of eight characteristics of qualitative teaching, which encompass the two teaching dimensions adequately. The construct validity and psychometric quality criteria were confirmed using exploratory and confirmatory factor analyses. The resultant test battery is valid and reliable in measuring teaching quality from the student's perspective.

Key Words: Teaching characteristics, classroom management, student orientation, factor analysis.

„WIE MISST MAN GUTEN SPORTUNTERRICHT?“ – ERFASSUNG ZENTRALER DIMENSIONEN VON UNTERRICHTS- QUALITÄT IM SCHULFACH SPORT

1 | PROBLEMSTELLUNG

Im Rahmen einer „evidenzbasierten Bildungspolitik und Schulentwicklung“ (Altrichter & Maag Merki, 2010, S. 35) etablierte sich in den vergangenen Jahren eine am Output orientierte Qualitätssicherung. Dabei entstand eine kritische Diskussion um Bildungsstandards und Kompetenzen. Diese sollen dazu dienen, die Effekte von schulischem Unterricht empirisch fassbar zu machen (Klieme et al., 2003). Damit diese Reformbemühungen tatsächlich die Schüler¹ erreichen, ist es jedoch nicht ausreichend, ausschließlich den Output zu erfassen, sondern die vorgelagerten Lehr- und Lernprozesse differenziert zu analysieren. Entsprechend folgern Oelkers und Reusser (2008), dass sowohl für das Lehren als auch für das Lernen gilt, dass es keine Qualität ohne entsprechend definierte Prozessmerkmale gibt. Sie räumen damit der Qualität der Lehr-Lern-Prozesse einen zentralen Stellenwert ein. Dies wird durch die Metaanalyse von Hattie (2008) empirisch gestützt, in der nachgewiesen werden konnte, dass die Unterrichts- und Lehrervariablen ein Drittel der Leistungsvarianz von Schülern erklären.

Neben der Formulierung von Bildungsstandards und der fachspezifischen Kompetenzmodellierung liegt die zentrale Herausforderung darin, Prozessmerkmale für die Erreichung bestimmter Lernziele zu identifizieren und zu erfassen (vgl. Gräsel & Göbel, 2011). Während sich die Forschungsbemühungen zunächst auf die Kernfächer konzentrierten, widmet sich mittlerweile auch das Fach Sport zunehmend diesem Thema. Spätestens seit der Jahrestagung der dvs-Sektion Sportpädagogik im Jahr 2004 mit dem Titel „Qualität im Schulsport“ werden die Fragen gestellt: „Was macht guten Sportunterricht aus? Wonach und wie lässt sich seine Qualität beurteilen? Wo kann man ansetzen, um sie wirksam zu verbessern?“ (Kleindienst-Cachay & Kurz, 2005, S. 9). Empirische Bemühungen, die Qualität von Sportunterricht zu erfassen, liegen aktuell nur sehr eingeschränkt vor (vgl. Wolters & Kemna, 2011). So mangelt es (erheblich) an fachspezifischen Instrumenten zur Messung von Unterrichtsqualität im Fach Sport (Wolters & Kemna, 2011) als Bestandteil einer empirischen Bildungsforschung in der Sportwissenschaft.

An dieser Stelle setzt der vorliegende Beitrag an. Es werden im Folgenden die fachübergreifenden Merkmale von Unterrichtsqualität erläutert, um daraus Erhebungsinstrumente für die Erfassung zentraler Dimensionen von Unterrichtsqualität abzuleiten. Die bisher vorhandenen, meist fachübergreifenden Instrumente werden auf das Fach Sport angepasst und auf deren psychometrische Qualitätskriterien empirisch überprüft. Abschließend wird untersucht, ob sich die Dimensionen von Unterrichtsqualität im Sportunterricht von anderen Fächern unterscheiden.

1 Zur einfacheren Lesbarkeit der Arbeit wird im Folgenden nur noch der männliche Term für Personengruppen verwendet. Trotzdem schließen die Aussagen auch Schülerinnen und Lehrerinnen ein.

2 | THEORETISCHER HINTERGRUND: WAS IST GUTER (SPORT-)UNTERRICHT?

In Abhängigkeit von der gewählten Zielperspektive des Unterrichts werden unterschiedliche *Qualitätsbegriffe* verwendet. Beispielsweise liegt der Fokus der kritisch-konstruktiven Didaktik von Klafki (1985) in der Berücksichtigung der Allgemeinbildung im Unterricht, indem epochaltypische Schlüsselprobleme behandelt werden. Während die Schuleffektivitätsforschung von einem engen, an den Sozialwissenschaften orientierten, Bildungsbegriff (vgl. Gräsel, 2011) ausgeht und auf einen Unterricht abzielt, der zu positiven Lern- und Leistungsergebnissen führt.

Dabei wird deutlich, dass die Frage, was als *qualitativ* hochwertig einzuschätzen ist, in hohem Maße davon abhängt, welche Personengruppe mit welchen Wertvorstellungen welche Ziele des Unterrichts verfolgt (Gräsel & Göbel, 2011). Die Begründung und Herleitung von Unterrichtsqualität kann daher sehr unterschiedlich erfolgen.

Die *Schuleffektivitätsforschung* folgte in der Vergangenheit dem *Prozess-Produkt-Paradigma*, das guten Unterricht als einen effektiven bzw. lernwirksamen Unterricht versteht (vgl. Gräsel & Göbel, 2011; Reusser & Pauli, 2010). Im Sinne dieses Paradigmas werden verschiedene Merkmale guten Unterrichts (z. B. Klarheit, Verständlichkeit) identifiziert, die mit erfolgreichen Unterrichtsergebnissen zusammenhängen (Gräsel & Göbel, 2011). Diese vereinfachten Ursache-Wirkungs-Beziehungen zwischen Unterrichtsangebot und Lernerfolg reduzieren jedoch die Unterrichtsrealität unangemessen. Entsprechend wurde das Prozess-Produkt-Paradigma zum *Prozess-Mediations-Produkt-Paradigma* weiterentwickelt (Helmke, 2010). Auf Basis eines sozial-konstruktivistischen Lernverständnisses wurden die Wahrnehmung und die aktive Nutzung des Unterrichts durch die Schüler als bedeutsame Aspekte von Unterrichtsqualität ergänzt. Es folgt der Annahme, dass die Lehrkräfte zwar Lernmöglichkeiten als Angebote an die Schüler schaffen können (= Prozess), deren Nutzung jedoch einzig bei den Schülern liegt (= Mediation). Nur wenn die Schüler mit diesen Angeboten produktiv umgehen, können im nächsten Schritt Lernerfolge (= Produkte) erzielt werden. Hierbei äußert sich ein Verständnis von Bildung, welches vor allem als *Selbstbildung* verstanden und durch guten Unterricht lediglich unterstützt werden kann (vgl. Reusser & Pauli, 2010).

Was also als guter Unterricht bezeichnet wird, hängt in der Schuleffektivitätsforschung grundlegend davon ab, welche *Ziele* im Unterricht angestrebt werden. Unter dem Begriff der „multiplen Bildungsziele von Unterricht“ (z. B. Klieme, 2008; Kunter, 2005) wird in der Unterrichtsforschung diskutiert, dass, neben Leistungsindikatoren, auch weitere Merkmale als Unterrichtsziele einbezogen werden müssen. Beispielsweise werden von Lipowsky (2009) die Vereinbarkeit von Leistungs- und Motivationsförderung und von Reusser und Pauli (2010) die Förderung metakognitiver, kommunikativer und motivationaler Schülerfähigkeiten als Unterrichtsziele genannt. Mittlerweile haben sich in den meisten Studien und Fachdidaktiken die Unterrichtsziele der Leistungs- und Motivationsentwicklung etabliert (z. B. Klieme & Rakoczy, 2008; Kunter et al., 2013a).

Im Fach Sport ist diese Diskussion um (Bildungs- bzw. Unterrichts-)Ziele im Kontext der verschiedenen fachdidaktischen Konzepte (im Überblick Balz, 2009) noch nicht beendet, wenngleich die

übergreifende Konzeption des „Erziehenden Sportunterrichts“ (Prohl, 2010) weitestgehend konsensfähig ist. Diese im Doppelauftrag des Schulsports verankerte pädagogische Leitidee geht davon aus, dass Sportunterricht

- ... zur kritisch-konstruktiven Teilhabe an der Sport- und Bewegungskultur in, neben und nach der Schule befähigen (Erziehung zum Sport) und
- ... zur persönlichen Entwicklung, Bildung und Subjektwerdung der Schüler beitragen soll (Erziehung im und durch Sport).

Im vorliegenden Beitrag wird, im Rückgriff auf ein enges Bildungsverständnis vor allem die Befähigung zur Teilhabe an der Sport- und Bewegungskultur im Fokus stehen. Ein breites Verständnis von Bildung, welches in enger Verbindung zur Persönlichkeitsentwicklung steht (Gräsel, 2011, S. 13), wird zwar mitgedacht, muss jedoch zunächst ausgeklammert bleiben und ist somit nicht Gegenstand vorliegender Untersuchung. Solch ein enges Bildungsverständnis lässt sich durch folgende Annahmen konstituieren: Für die Teilhabe an der Sport- und Bewegungskultur ist die Verfügbarkeit grundlegender Leistungsdispositionen (z. B. motorische Basiskompetenzen; Herrmann & Gerlach, 2014) sowie die Motivation, Sport zu treiben, eine nötige, wenn auch nicht hinreichende Voraussetzung (Balz, 2009; Gogoll, 2012; Scheid & Prohl, 2011). So ist auch für das Fach Sport festzuhalten, dass erkennbare Leistungen angezielt und erbracht werden müssen (vgl. Wolters, Klinge, Klupsch-Sahlmann & Sinning, 2009). Leistung wird hierbei als „voraussetzungsgemäßes Bemühen um die Bewältigung einer Bewegungsaufgabe“ (Funke-Wieneke, 2003, S. 6-9) definiert und beinhaltet auch motivationale Aspekte. Damit gilt die Förderung von Leistung und Motivation als zentrales Ziel des Sportunterrichts, wenngleich sie nicht den gesamten Bildungsauftrag widerspiegelt.

Lässt man sich darauf ein, diese beiden angesprochenen Ziele im Sportunterricht zu fokussieren, so lassen sich deutliche Parallelen zu den anderen (Kern-)Fächern und deren Unterrichtsforschung herstellen.

2.1 | MERKMALE VON UNTERRICHTSQUALITÄT

In der Forschungslinie des *Prozess-Produkt-Paradigmas* wurden zunächst Unterrichtsmerkmale herausgearbeitet, die fachübergreifend den Kompetenzerwerb und die Motivationsentwicklung von Schülern unterstützen. Darauf aufbauend, wurde zunehmend versucht, auch fachspezifische Merkmale zu identifizieren. Es wurden diverse Merkmalslisten entwickelt, die zentrale Aspekte von Unterrichtsqualität zusammenfassen. Im englischsprachigen Raum werden vor allem die zwölf fach- und stufenübergreifend konzipierten Merkmale guten Unterrichts nach Brophy (2000) rezipiert (u. a. Lipowsky, 2006; Klieme, 2008; Kiel, 2010). Im deutschsprachigen Raum sind die Merkmale guten Unterrichts von Helmke (2010) sowie Meyer (2005) verbreitet (vgl. Praetorius, 2014). Helmke (2010, S. 168-267) unterscheidet zehn fach- und stufenübergreifende Merkmale guten Unterrichts. Diese bieten eine gute Orientierungshilfe für die Unterrichtspraxis und stellen zugleich das Gerüst für Unterrichtsbeobachtung und Unterrichtsbeurteilung dar.

Angelehnt an die Merkmalsliste von Meyer (2005), formuliert Gebken (2005, S. 236ff.) acht Merkmale guten Sportunterrichts, welche sehr stark an Meyer (2005) orientiert sind, und daher nur wenig Erkenntnisfortschritt bieten. Balz (2010) bestimmt ebenfalls Merkmale guten Unterrichts, die er in drei Dimensionen aufteilt: (1) Inputmerkmale zur Strukturqualität, (2) Verlaufsmerkmale zur Prozessqualität und (3) Outputmerkmale zur Produktqualität. Diese Ordnung ist fachübergreifend formuliert und bietet einen weiteren Strukturierungsvorschlag, welcher zwar bedingt anschlussfähig an die empirische Bildungsforschung ist, aber eine Präzisierung vermissen lässt. Einen fachspezifischen Entwurf legen Wolters und Kemna (2011) mit ihrem „Rahmenmodell für Qualitätskriterien des Sportunterrichts“ vor. Die Bewegung als Gegenstand des Fachs Sport wird hier unterteilt in die quantitative und qualitative Dimension der *Handlung* sowie in die *Reflexion* der eigenen und gesellschaftlichen Praxis. Eine Formulierung von konkreten Unterrichtsmerkmalen, welche für eine Operationalisierung geeignet wären, bleibt jedoch auch hier aus. Insgesamt ist im Fach Sport die fachspezifische Identifikation von relevanten Unterrichtsmerkmalen noch in den Anfängen und primär normativ geprägt. Empirische Ergebnisse liegen bis auf Einzeluntersuchungen zu Teilaspekten (z. B. zum Klima: Gerlach, 2005; Heemsoth & Miethling, 2012; Heemsoth, 2014) ebenfalls nur eingeschränkt vor. Daher scheint für die Erfassung von Unterrichtsqualität im Fach Sport eine Orientierung an der empirischen Unterrichtsforschung sinnvoll und praktikabel.

2.2 | DIMENSIONEN VON UNTERRICHTSQUALITÄT

In der aktuellen empirischen Unterrichtsforschung wird die Verwendung der genannten Merkmalslisten zunehmend kritisiert, da zwischen den einzelnen Merkmalen inhaltliche Überlappungen (bspw. zwischen den Merkmalen der Schülerorientierung und des lernförderlichen Klimas) existieren und keine Aussagen darüber getroffen werden, wie die Merkmale zusammenhängen. Entsprechend wird eine theoretische Fundierung gefordert (u. a. Hamre & Pianta, 2010; Köller, 2008). „Eine Stoßrichtung der modernen Unterrichtsforschung ist deshalb die Identifikation von übergeordneten Faktoren ..., insbesondere von empirisch unterscheidbaren, voneinander hinreichend unabhängigen und sparsamen Kernprinzipien der Qualität des Fachunterrichts“ (Klieme, 2008, S. 306).

Im Zuge der „TIMSS-Studie“ wurde für den Mathematikunterricht ein dreidimensionales Modell der Unterrichtsqualität (Klieme, Schümer & Knoll, 2001; vgl. auch Klieme, Lipowsky, Rakoczy & Ratzka, 2006) entwickelt. Darin werden Unterrichtsmerkmale theoriegeleitet zu Grunddimensionen zusammengefasst und Annahmen darüber getroffen, welcher Zusammenhang zwischen diesen Dimensionen und den Lernzuwächsen besteht. Dieses fachübergreifende Modell wird im deutschsprachigen Raum häufig rezipiert und liegt den groß angelegten Studien wie „Pythagoras“ (Mathematik; u. a. Klieme et al., 2006), „DESI“ (Deutsch und Englisch; u. a. (Klieme, 2008) oder „COACTIV“ (Mathematik, u. a. Kunter et al., 2013a) zugrunde. Im englischsprachigen Raum hat sich ein nahezu deckungsgleiches Modell von Hamre und Pianta (2010) etabliert. In diesen Studien konnten durchgängig drei Grunddimensionen von Unterrichtsqualität faktorenanalytisch bestätigt werden, welche als eine empirisch verankerte Systematik didaktischer Prinzipien verstanden wer-

den können. Guter Unterricht, durch dessen Bausteine sowohl die Leistung der Schüler als auch ihre Motivation gefördert werden, zeichnet sich aus durch (a) strukturierte, klare und störungspräventive *Unterrichts- und Klassenführung*² [classroom management], (b) unterstützendes, *schülerorientiertes Sozialklima*² [emotional support], und (c) *kognitive Aktivierung* [instructional support] (Stronge, Tucker, Hindman, 2004; Hamre & Pianta, 2010; Klieme & Rakoczy, 2008). „Diese drei Basisdimensionen erscheinen geeignet, die theoretischen Konzepte und Wirkungshypothesen der Unterrichtsforschung zu integrieren“ (Klieme et al., 2006, S. 130).

a. Klassenführung

Unter guter *Klassenführung* wird ein Unterricht verstanden, der den Schülern hinreichend Zeit zum Lernen ermöglicht und in dem möglichst wenig Disziplinprobleme auftreten. Eine optimale Zeitnutzung gibt den Schülern Zeit für das Lernen und schafft viele Gelegenheiten, in denen sich die Schüler in Lernprozesse involvieren können. Entsprechend müssen Lehrkräfte in effektiver Weise präventiv oder intervenierend mit Unterbrechungen und Disziplinproblemen umgehen, um die Unterrichtsstörungen zu minimieren und die Aufmerksamkeit der Schüler zu sichern (Borich, 2007; vgl. den Klassiker von Kounin, 1976).

Merkmale einer guten *Klassenführung* sind ...

- ... klar formulierte, verbindliche Regeln und Ziele,
- ... eine gute Organisation, Klarheit und Strukturierung des Unterrichts sowie
- ... wenige Störungen und eine geringe Zeitverschwendung (Borich, 2007).

b. Schülerorientierung

Bei der *Schülerorientierung* stehen die sozial-emotionalen Unterrichtsmerkmale, welche für den Lernerfolg von Bedeutung sind, im Fokus. Das Klassenklima und die Motivierungsqualität des Unterrichts sind für die Lern- und Interessensentwicklung sowie für die intrinsische Lernmotivation bedeutsam (Gräsel & Göbel, 2011). Bei der Konzeptualisierung dieser Dimension wird auf die Selbstbestimmungstheorie (Deci & Ryan, 1993) und das Bedürfnis nach sozialer Eingebundenheit verwiesen (ausführlich Hamre & Pianta, 2010).

Merkmale der *Schülerorientierung* sind ...

- ... individuelle Lernunterstützung und diagnostische Kompetenz des Lehrers im Sozialbereich,
- ... eine positive Lehrer-Schüler-Beziehung,
- ... konstruktives, positives und individuelles Lehrerfeedback sowie ein positiver Umgang mit Schülerfehlern (Klieme et al., 2006; Klieme et al., 2001).

c. Kognitive Aktivierung

Die *kognitive Aktivierung* spiegelt die Komplexität von Aufgabenstellungen und die Intensität des fachlichen Lernens wider (Klieme et al., 2001). Eine besondere Schwierigkeit dieser Dimension liegt darin, dass sie nicht direkt beobachtbar und erfassbar ist. Entsprechend wird das

2 Zur leichteren Lesbarkeit wird im Folgenden *Unterrichts- und Klassenführung* mit *Klassenführung* und *schülerorientiertes Sozialklima* mit *Schülerorientierung* abgekürzt.

beobachtbare Lehrerverhalten herangezogen, um auf die *kognitive Aktivierung* der Schüler zu schließen (Lipowsky, 2006). Die bisherige Befundlage aus den Fächern Mathematik (u. a. „Pythagoras-Studie“, Klieme et al., 2006; „COACTIV“, Kunter et al., 2013a) und Sprache (Deutsch bzw. Englisch „DESI-Studie“, Klieme, 2008) zeigt, dass die kognitive Aktivierung in jedem Fach unterschiedlich definiert werden muss. Dagegen definieren Hamre und Pianta (2010, Pianta et al. 2008) ihre Dimension „instructional Support“ vollständig fachunspezifisch und machen dies an fachübergreifenden Beobachtungsindikatoren des Unterrichts fest (z. B. am Vorliegen von Warum- und Wie-Fragen der Lehrkraft). Die Befundlage hinsichtlich der Fachspezifik dieser Dimension kann daher als uneinheitlich bezeichnet werden und bedarf weiterer Forschungsbestrebungen (Klieme & Rakoczy, 2008).

2.3 | MESSUNG VON UNTERRICHTSQUALITÄT UND FORSCHUNGSFRAGEN

Unterrichtsqualität wird über die Einschätzung von Unterrichtsprozessen erfasst, da sie nicht objektiv messbar ist, wie beispielsweise die Lernergebnisse von Unterricht. Die Einschätzungen werden von Lehrern, externen Beobachtern oder Schülern vorgenommen (Praetorius, 2014). Mit diesen drei Personengruppen als Datenquelle zur Erfassung von Unterrichtsqualität gehen verschiedene Vor- und Nachteile einher.

So liegt beispielsweise bei den **Lehrereinschätzungen** die Schwierigkeit darin, dass die Lehrer als Akteure im Unterricht ihre eigenen Handlungen nicht neutral beurteilen können. Daraus können Verzerrungen in der Selbstwahrnehmung im Unterrichtsgeschehen und in der Lehrer-Schüler-Interaktion eintreten (Clausen, 2002; Praetorius, 2014, Desimone, Smith & Frisvold, 2010).

Dagegen wird der Erfassung von Unterrichtsqualität über geschulte **externe Beobachter** eine höhere Validität zugesprochen und die Vergleichsmöglichkeiten zwischen den Lehrkräften sind in einem höheren Ausmaß gegeben. Nachteile sind der sehr hohe Zeit- und Kostenaufwand, der punktuelle Beobachtungszeitraum sowie mögliche Reaktivitätseffekte durch das Vorhandensein eines Beobachters bzw. einer Kamera (Clausen, 2002; Praetorius, 2014).

An diesen Nachteilen der Beobachtung setzen die Stärken der **Schülereinschätzungen** an. Sie liegen in der ökonomischen und damit kostengünstigen Erhebungsweise durch Fragebögen, mit der auch große Stichproben erfasst werden können. Die generierten Daten spiegeln dabei – im Vergleich zu punktuellen Beobachtungen – einen relativ langen Beurteilungszeitraum wider. Durch das Aggregieren der Schülerdaten auf Klassenebene gewinnen die Schülerratings an Reliabilität. Der Hauptvorteil liegt wohl darin, dass die Schülerwahrnehmungen vom Unterricht für das Lernen und die Motivation von Schülern relevanter sind als deren beobachtbares Verhalten. *Nachteile* sind, dass Schüler nicht über das methodisch-didaktische Wissen verfügen, um Unterricht adäquat zu bewerten, sie stark in das unterrichtliche Geschehen involviert sind und Tendenzen von sozialer Erwünschtheit gegenüber der Lehrkraft auftreten können (Praetorius, 2014). Schülereinschätzungen sind daher zur Erfassung von Aspekten der Unterrichtsqualität, deren Wahrnehmung subjektiv geprägt ist (z. B. Unterrichtsklima), sinnvoll und praktikabel.

Forschungsfragen

Der vorliegende Beitrag knüpft an die Vorarbeiten der empirischen Unterrichtsforschung an, indem etablierte Fragebogeninventare aus den Kernfächern an das Fach Sport angepasst wurden, um die Unterrichtsqualität über Schülereinschätzungen zu erfassen. Ziel war die Entwicklung eines fachspezifischen sparsamen Fragebogeninventars. Hierbei wurden zunächst die Dimensionen *Klassenführung* sowie *Schülerorientierung* bearbeitet, da diese Dimensionen fachübergreifend formuliert sind und demnach eine Verwendung im Fach Sport sinnvoll erscheint. Durch die Operationalisierung dieser beiden Dimensionen soll auch ein Vergleich von Daten zwischen verschiedenen Fächern möglich werden. Die Dimension *kognitive Aktivierung* musste aufgrund der ungeklärten Frage, ob diese für das Fach Sport fachunspezifisch oder fachspezifisch definiert werden sollte, zunächst außen vor gelassen werden.

Es werden folgende Forschungsfragen untersucht:

- Können Erhebungsinstrumente aus den Kernfächern für die beiden Dimensionen der Unterrichtsqualität *Klassenführung* und *Schülerorientierung* auf das Fach Sport adaptiert werden?
- Inwieweit genügen die angepassten Erhebungsinstrumente psychometrischen Kriterien?
- Kann die Zuordnung der Unterrichtsmerkmale zu den beiden Dimensionen der Unterrichtsqualität für das Fach Sport bestätigt werden?

3 | METHODE

3.1 | DATENGRUNDLAGE UND STICHPROBENBESCHREIBUNG

Zur Beantwortung der Forschungsfragen wurden die Daten der IMPEQT-Studie (**IM**plementation in **Physical Education and the Quality of Teaching**)³ herangezogen, in welcher der Einfluss von Qualitätsmerkmalen im Unterricht auf Lernergebnisse über ein Jahr analysiert wurde. Die Datenerhebungen fanden zu zwei Messzeitpunkten (Januar/Februar 2012 bzw. 2013) in unterschiedlichen Siedlungsgebieten der Kantone Zürich, Baselland und Aargau statt. Es wurden 48 Klassen von 39 Lehrern (62.0% männlich; Alter: M = 35.1 Jahre; SD = 8.9; Range 23-56) der siebten und darauf folgenden achten Jahrgangsstufe erhoben. Für die hier verwendete Fragebogenuntersuchung wurde eine Unterrichtsstunde im Klassenraum genutzt.

Die Datenanalyse umfasst zum ersten Messzeitpunkt 884 Schüler (50.1% männlich; Alter: M = 13.2 Jahre; SD = .65; Range 12-15) der siebten Jahrgangsstufe und zum zweiten Messzeitpunkt 767 Schüler (47.5% männlich; Alter: M = 14.2 Jahre; SD = .61; Range 13-16) der achten Jahrgangsstufe. Durch den Einsatz des *Full Information Maximum Likelihood* (FIML) Schätzverfahrens konnten auch Schüler berücksichtigt werden, die einzelne fehlende Werte aufwiesen (Muthén & Muthén, 2012). Die vollständig geschätzte Stichprobe beider Messzeitpunkte umfasst 1013 Schüler (Tab. 1). Die Erhebungsinstrumente wurden zur Reduzierung der Fragebogenlänge auf zwei Messzeitpunk-

3 Dieses Forschungsprojekt wird durch die Eidgenössische Forschungskommission der Schweiz im Zeitraum von September 2011 bis Dezember 2013 gefördert (Projektleiter: Erin Gerlach, Universität Potsdam).

te aufgeteilt. Es wird davon ausgegangen, dass die Einschätzungen über die Zeit weitestgehend stabil bleiben und damit die unterschiedlichen Erhebungsinstrumente zu einem querschnittlichen Datensatz zusammengefügt werden können.

Tabelle 1: Stichprobenbeschreibung

		t1	t2	t1/t2 ^a
Mädchen	n %	441 (49.9 %)	403 (52.5 %)	504 (49.5 %)
Jungen	n %	443 (50.3 %)	364 (47.5 %)	515 (50.5 %)
Gesamt	n %	884 (100.0 %)	767 (100.0 %)	1019 (100.0 %)

^a Vollständig geschätzte Stichprobe

3.2 | MESSUNG DER UNTERRICHTSQUALITÄT

Ausgehend von den beiden Unterrichtsdimensionen *Klassenführung* und *Schülerorientierung* wurden Erhebungsinstrumente ausgewählt und für das Fach Sport adaptiert. Die Anpassung der Skalen erfolgte über die Umformulierung der Testitems, indem bspw. „Mathematiklehrer“ durch „Sportlehrer“ oder „Mathematikunterricht“ durch „Sportunterricht“ ersetzt wurde. Folgende Tab. 2 listet die beiden Dimensionen sowie die verwendeten Skalen mit ihren Testitems und Itemkennwerten auf.

Tab. 2: Beschreibung der Testitems zur Erfassung der Unterrichtsmerkmale

	Abk.	Item	M	SD	r _{It-i} ^a	r _{It-i} ^b
Klassenführung	1. Zeitnutzung (angelehnt an „QUASSU-Studie“; Ditton & Merz, 2000, „Pythagoras-Studie“; Lipowsky, Rakoczy, Buff & Klieme, 2005)					
	ZN1	Im Sportunterricht stehen wir häufig herum und machen gar nichts. (-)	3.36	.83	.41	.58
	ZN2	Im Sportunterricht wird viel Zeit vertrödelt. (-)	2.93	.92	.07	–
	ZN3	Eine Menge Zeit des Unterrichts wurde für anderes als Sport verwendet (-).	3.11	.94	.38	.68
	2. Disziplin („IGLU-Studie“ Bos & Buddeberg, 2005, S. 206; „DESI-Studie“; Wagner, Helmke & Rösner, 2009, S. 112)					
	DISZ1	Im Sportunterricht können wir ungestört arbeiten.	3.16	.73	-.34	–
	DISZ2	Wir hören nicht auf das, was unser Sportlehrer uns sagt. (-)	3.18	.86	.45	.51
	DISZ3	Unser Sportlehrer muss lange warten, bis Ruhe eintritt. (-)	2.38	.93	.51	.51
	DISZ4	Im Sportunterricht ist es laut und es geht alles durcheinander. (-)	3.10	.90	.60	.67
	DISZ5	Zu Beginn der Stunde vergeht viel Zeit, in der gar nichts passiert. (-)	2.94	.94	.48	.57
	DISZ6	Im Sportunterricht geht es ganz schön drunter und drüber. (-)	2.91	.92	.50	.53
	DISZ7	Im Sportunterricht gibt es viele Störungen durch die Mitschüler. (-)	2.71	.94	.57	.62
	3. Regelklarheit („IGLU-Studie“; Bos & Buddeberg, 2005, S. 182)					
	REGEL1	Im Sportunterricht sind die Verhaltensregeln, die man einhalten muss, allen bekannt.	3.21	.80	.45	.45
	REGEL2	Im Sportunterricht hat unser Lehrer klargemacht, was passiert, wenn man (Verhaltens-)Regeln verletzt.	2.88	.96	.45	.45
	REGEL3	Im Sportunterricht ist klar, was man machen darf und was nicht.	3.16	.80	.53	.53
	4. Zielklarheit („Pythagoras-Studie“; Lipowsky et al., 2005, S. 61)					
	ZIELK1	Unser Sportlehrer gibt uns am Anfang des Unterrichts bekannt, was er mit uns üben will.	3.44	.73	.60	.60

Klassenführung	ZIELK2	Unser Sportlehrer sagt immer, was in nächster Zeit im Unterricht geübt wird.	3.21	.81	.52	.52	
	ZIELK3	Bevor der Sportlehrer mit der Stunde beginnt, macht er uns die Ziele klar.	2.88	.94	.54	.54	
	ZIELK4	Bevor der Sportlehrer mit der Stunde beginnt, sagt er uns, was er durchführen will.	3.21	.85	.66	.66	
	5. Diagnosekompetenz („Pythagoras-Studie“; Lipowsky et al., 2005, S. 44)						
	DKOMP1	Unser Sportlehrer weiß genau, was jeder von uns kann.	2.97	.79	.56	–	
	DKOMP2	Unser Sportlehrer merkt sofort, wenn etwas nicht richtig verstanden wird.	2.83	.81	.61	.60	
	DKOMP3	Unser Sportlehrer merkt sofort, wenn ein Schüler/eine Schülerin im Unterricht nicht mitkommt.	2.92	.84	.62	.60	
DKOMP4	Unser Sportlehrer weiß sofort, was jemand nicht verstanden hat.	2.73	.85	.71	.71		
DKOMP5	Unser Sportlehrer weiß genau, bei welchen Aufgaben wir Schwierigkeiten haben.	2.83	.81	.65	.64		
DKOMP6	Unser Sportlehrer weiß genau, warum jemand etwas im Unterricht nicht kann.	2.58	.89	.62	.61		
Schülerorientierung	6. Fürsorglichkeit der Lehrkraft („LASSO“; Saldern & Litting, 1985)						
	FSLP1	Unser Sportlehrer kümmert sich um die Probleme der Schüler.	2.75	.92	.52	.53	
	FSLP2	Unsere Sportlehrerin/unser Sportlehrer hilft uns wie ein Freund.	2.55	.91	.56	.55	
	FSLP3	Unser Sportlehrer ist bereit, mit uns zu reden, wenn etwas nicht stimmt.	3.18	.80	.57	.55	
	FSLP4	Wenn wir etwas mit unserem Sportlehrer bereden wollen, dann hat er auch Zeit dafür.	3.03	.82	.56	.53	
	FSLP5	Unser Sportlehrer hilft jedem von uns, wenn jemand etwas nicht kann.	3.37	.74	.41	–	
	7. Schülermitbestimmung („DESI-Studie“; Wagner et al., 2009, S. 106)						
	SMITB1	Mein Sportlehrer lässt uns bei der Gestaltung des Sportunterrichts Themen und Aufgaben auswählen.	2.42	.86	.52	.59	
	SMITB2	Mein Sportlehrer geht auf unsere Vorschläge ein.	2.85	.83	.62	.66	
	SMITB3	Mein Sportlehrer ermutigt uns, unsere eigene Meinung auszudrücken.	2.63	.90	.52	–	
	SMITB4	Wenn jemand eine gute Idee hat, dann geht mein Sportlehrer darauf ein.	2.77	.84	.68	.62	
	SMITB5	Mein Sportlehrer gibt mir Gelegenheit, meine Meinung zu sagen.	2.90	.86	.58	–	
	SMITB6	Mein Sportlehrer interessiert sich für das, was ich zu sagen habe.	2.85	.82	.54	–	
	8. Individuelle Bezugsnormorientierung (Schwarzer, Lange & Jerusalem, 1982)						
	INBO1	Wenn Schüler ihre Leistungen gegenüber früher verbessern, werden sie dafür vom Sportlehrer besonders gelobt.	2.87	.85	.50	.50	
	INBO2	Unser Sportlehrer lobt auch die schlechteren Schüler, wenn er merkt, dass sie sich verbessern.	3.29	.81	.55	.55	
	INBO3	Wenn ich eine gute Leistung gezeigt habe, lobt mich der Sportlehrer meistens, auch wenn viele Schüler noch besser sind als ich.	3.02	.84	.59	.59	
	9. Differenzierung („QUASSU-Studie“; Ditton & Merz, 2000, S. 45)						
	DIFF1	Unser Sportlehrer gibt den Schülern unterschiedliche Aufgaben, je nach ihrem Können.	2.02	.95	.53	.53	
	DIFF2	Unser Sportlehrer stellt Arbeitsgruppen nach dem Können der Schüler zusammen.	2.18	.96	.46	.44	
	DIFF3	Unser Sportlehrer stellt den besseren Schülern schwierigere Aufgaben.	2.32	1.02	.56	.60	
	DIFF4	Unser Sportlehrer stellt den schwächeren Schülern leichtere Aufgaben.	2.26	.96	.62	.62	
	DIFF5	Unser Sportlehrer stellt Arbeitsgruppen nach Interesse der Schüler zusammen.	2.32	.94	.33	–	
codierung: 1 = stimmt nicht, 2 = stimmt kaum, 3 = stimmt ziemlich, 4 = stimmt genau (-) Items invers codiert a Itemtrennschärfe aller Items b Itemtrennschärfe des finalen Itemsets							

3.3 | ANALYSEVERFAHREN

Die Datenaufbereitung wurde in SPSS 21 vorgenommen. Alle weiterführenden Berechnungen wurden im Programm Mplus 7.0 (Muthén & Muthén, 2012) durchgeführt. Allen hier referierten Werten liegt die vollständig geschätzte Stichprobe von 1019 Schülern zugrunde. Einzig bei den deskriptiven Itemkennwerten in Tab. 2 werden zur leichteren Interpretierbarkeit die tatsächlichen, nicht geschätzten Werte aus SPSS 21 (deletion listwise) referiert. Die Datenauswertung erfolgte in drei Schritten: (1) Die faktorielle Struktur der Unterrichtsmerkmale wurde mit einer exploratorischen Faktorenanalyse ermittelt (Modell 1a, 1b) und (2) konfirmatorisch geprüft (Modell 2a, 2b). Im Anschluss wird (3) mittels konfirmatorischen Faktorenanalysen zweiter Ordnung geprüft, ob die theoretisch angemessene Aufteilung zu den zwei Dimensionen von Unterrichtsqualität gelingt (Modell 3a, 3b, 3c).

In **Modell 1a** werden mit allen 42 Items der neun Faktoren (Tab. 2) exploratorische Faktorenanalysen⁴ mit obliquer GEOMIN-Rotation für sieben bis neun Faktoren berechnet. Die Faktorladungen und Residualvarianzen wurden dabei zur Schätzung freigegeben und Nebenladungen wurden zugelassen. Auf dieser Basis wurden Items mit hohen Nebenladungen identifiziert und ausgeschlossen, was zu einer erneuten Analyse mit einer gekürzten Testbatterie mit 34 Items führte (**Modell 1b**).

Auf Basis dieser Berechnungen wird die explorierte Struktur der gekürzten Testbatterie mit acht Faktoren in **Modell 2a** unter restriktiveren Bedingungen der konfirmatorischen Faktorenanalyse geprüft. Hierbei werden Nebenladungen nicht zugelassen. Faktorladungen und Residualvarianzen werden für jedes Testitem frei geschätzt. Im **Modell 2b** wird eine weitere Restriktion vorgenommen, indem innerhalb der konfirmatorischen Faktorenanalyse die nichtstandardisierten Faktorladungen der Testaufgaben für jeden Faktor gleich gesetzt werden (essenziell tauäquivalente Ladungen; Bühner, 2011, S. 400). Damit wird geprüft, ob eine Gleichgewichtung der Faktorladungen und damit die Bildung eines Faktorsummenwerts über die Testaufgaben eines Faktors zulässig ist.

Im **Modell 3a** werden, ausgehend vom Modell 2a, zwei latente Faktoren zweiter Ordnung (Dimensionen der Unterrichtsqualität) in die konfirmatorische Faktorenanalyse eingeführt, welche jeweils vier latente Faktoren erster Ordnung (Unterrichtsmerkmale) verknüpfen. Für beide Ordnungen werden Nebenladungen nicht zugelassen sowie Faktorladungen und Residualvarianzen frei geschätzt. Zur Absicherung dieser Ergebnisse wurde im **Modell 3b** eine zweifaktorielle Faktorenanalyse berechnet, innerhalb der die Unterrichtsmerkmale als manifeste Faktormittelwerte eingehen und die Dimensionen der Unterrichtsqualität bei gleicher Zuordnung wie in Modell 3a zwei latente Faktoren bilden.

Da in der vorliegenden Stichprobe keine Zufallsauswahl von Schülern untersucht wurde, sondern ganze Schulklassen mit deren Schülern, wurde im abschließenden **Modell 3c** – aufbauend auf dem Modell 3b – die Mehrebenenstruktur mitberücksichtigt. Eine Möglichkeit, mit den Spezifika von

4 Aufgrund der Programmlogik von Mplus handelt es sich hierbei um exploratorische Strukturgleichungsmodelle (ESEM). Diese bieten u. a. den Vorteil, dass Fit-Indizes zur Bewertung der Modellanpassung zur Verfügung gestellt werden.

Daten in Mehrebenenstrukturen umzugehen, besteht darin, in der konfirmatorischen Faktorenanalyse die Standardfehler zu korrigieren⁵ (Muthén & Satorra, 1995; Rakoczy, 2006). Aufgrund der eingeschränkten Stichprobe von 48 Klassen kann diese Korrektur nur für das Modell 3b mit den manifesten Faktormittelwerten der Unterrichtsmerkmale vorgenommen werden, da die zu schätzenden Parameter die Anzahl der verfügbaren Klassen im Datensatz nicht übersteigen dürfen.

Die Beurteilung der Anpassungsgüte der Modelle erfolgt anhand der in der Literatur vorgeschlagenen Fit-Indizes (Bühner, 2011; Hu & Bentler, 1999; Schermelleh-Engel, Moosbrugger & Müller, 2003; Schreiber, Stage, King, Nora & Barlow, 2006). Folgende Cut-offs deuten auf eine (sehr) gute Modellanpassung hin: Comparative Fit Index (CFI) > .90, Root Mean Square Error of Approximation (RMSEA) < .06, Standardized-Root-Mean-Residual (SRMR) < .11. Für die bessere Interpretierbarkeit der Ergebnisse werden folgend durchgängig die standardisierten Koeffizienten berichtet.

4 | ERGEBNISSE

Tab. 3: Ergebnisse der Analysen

Modell	Analyse		χ^2	df	p	CFI	RMSEA	SRMR
1a	EFA	7 Faktoren (42 Items)	1172.13	588	< .05	.949	.031	.025
	EFA	8 Faktoren (42 Items)	1005.32	553	< .05	.960	.028	.023
	EFA	9 Faktoren (42 Items)	858.49	519	< .05	.970	.025	.021
1b	EFA	7 Faktoren (34 Items)	719.44	344	< .05	.959	.033	.023
	EFA	8 Faktoren (34 Items)	564.45	317	< .05	.973	.028	.019
	EFA	9 Faktoren (34 Items)	454.74	291	< .05	.982	.023	.017
2a	CFA	8 Faktoren (34 Items)	1043.82	499	< .05	.942	.033	.041
2b	CFA	8 Faktoren tauäquivalent	1236.57	525	< .05	.923	.036	.055
3a	CFA	8 Faktoren 1. Ordnung, 2 Faktoren 2. Ordnung	1132.05	518	< .05	.934	.034	.052
3b	CFA	2 Faktoren manifest	93.88	19	< .05	.948	.062	.046
3c	CFA	2 Faktoren manifest, Korrektur des Standardfehlers	55.59	19	< .05	.957	.043	.046

EFA Exploratorische Faktorenanalyse / CFA Konfirmatorische Faktorenanalyse

4.1 | EXPLORATION DER STRUKTUR DER UNTERRICHTSMERKMALE

Die exploratorische Faktorenanalyse im **Modell 1a** wurde mit allen 42 Testitems (Tab. 2) für sieben bis neun Faktoren berechnet. Der Eigenwertverlauf ([1] 8.83, [2] 3.89, [3] 3.37, [4] 1.83, [5] 1.48, [6] 1.31, [7] 1.14, [8] 1.04, [9] 0.99) legte unter Anwendung des Scree-Kriteriums die Extraktion von acht Faktoren nahe. Die achtfaktorielle Lösung zeigte gute Modellanpassung, während die siebenfaktorielle Lösung etwas schlechtere und die neunfaktorielle Lösung etwas bessere Werte aufwies (Tab. 3). Bei der neunfaktoriellen Lösung war der neunte Faktor inhaltlich nicht interpretierbar, da er lediglich die Nebenladungen der anderen Faktoren bündelte. Daher wurde die achtfaktorielle Lösung favorisiert, bei der die Testitems der Skalen „Disziplin“ und „Zeitnutzung“ zu einem

5 Dieser Analyseschritt erfolgt mit der in Mplus implementierten „type = complex“-Funktion für geschachtelte Datensätze.

Faktor zusammenfielen. Es ließen sich acht Items (ZN2; DISZ1; DKOMP1; FSLP5; SMITB3; SMITB5; SMITB6; DIFF5; vgl. Tab. 2) keinem Faktor klar zuordnen, da sie Nebenladungen auf verschiedene Faktoren besaßen. Nach inhaltlicher Kontrolle und unter Beachtung der Itemtrennschärfe (Tab. 2) wurden diese acht Testitems aus den folgenden Analysen ausgeschlossen.

Die verbleibenden 34 Testitems wurden einer erneuten exploratorischen Faktorenanalyse im **Modell 1b** unterzogen. Auf Basis des Eigenwertverlaufs und der inhaltlichen Interpretierbarkeit war die achtfaktorielle Lösung mit guter Modellanpassung (Tab. 3) abermals zu favorisieren. Die Skalen „Disziplin“ und „Zeitnutzung“ fielen wiederum in einem Faktor zusammen. Bei dieser finalen Lösung waren die Nebenladungen der einzelnen Testitems durchweg sehr gering (Tab. 4).

Tab. 4: Faktorstruktur der exploratorischen Faktorenanalyse

Faktor	Disziplin/ Zeitnutzung	Zielklarheit	Regelklarheit	Diagnosekompetenz	Fürsorglichkeit Lehrer	Schülermitbestimmung	Bezugsnormorientierung	Differenzierung
Eigenwert	7.29	3.34	3.16	1.74	1.44	1.20	1.08	0.98
ZN1	.57*	-	-	-	-	-	-	-
ZN3	.71*	-	-	-	-	-	-	-
DISZ2	.54*	-	-	-	-	-	-	-
DISZ3	.57*	-	-	-	-	-	-	-
DISZ4	.75*	-	-	-	-	-	-	-
DISZ5	.57*	-	-	-	-	-	-	-
DISZ6	.63*	-	-	-	-	-	-	-
DISZ7	.71*	-	-	-	-	-	-	-
REGEL1	-	.51*	-	-	-	-	-	-
REGEL2	-	.50*	-	-	-	-	-	-
REGEL3	-	.45*	.26*	-	-	-	-	-
ZIELK1	-	-	-	-	-	-	-	.57*
ZIELK2	-	-	-	-	-	-	-	.54*
ZIELK3	-	-	.27*	-	-	-	-	.43*
ZIELK4	-	-	-	-	-	-	-	.82*
DKOMP2	-	-	.56*	-	-	-	-	-
DKOMP3	-	-	.67*	-	-	-	-	-
DKOMP4	-	-	.76*	-	-	-	-	-
DKOMP5	-	-	.68*	-	-	-	-	-
DKOMP6	-	-	.69*	-	-	-	-	-
FSLP1	-	-	-	-	.60*	-	-	-
FSLP2	-	-	-	.25*	.45*	-	-	-
FSLP3	-	-	-	-	.54*	-	-	-
FSLP4	-	-	-	-	.48*	-	-	-
SMITB1	-	-	-	.66*	-	-	-	-
SMITB2	-	-	-	.81*	-	-	-	-
SMITB4	-	-	-	.69*	-	-	-	-
IBNO1	-	-	-	-	-	.43*	-	-
IBNO2	-	-	-	-	-	.50*	-	-
IBNO3	-	-	-	-	-	.85*	-	-
DIFF1	-	-	-	-	-	-	.61*	-
DIFF2	-	-	-	-	-	-	.44*	-
DIFF3	-	-	-	-	-	-	.79*	-
DIFF4	-	-	-	-	-	-	.71*	-

Ladungen unter .20 werden nicht dargestellt (-).

4.2 | KONFIRMATORISCHE ÜBERPRÜFUNG DER STRUKTUR DER UNTERRICHTSMERKMALE

Eine Replikation dieser achtfaktoriellen Struktur unter restriktiveren Bedingungen (d. h. ohne Nebenladungen) im Rahmen der konfirmatorischen Faktorenanalyse im **Modell 2a** erbrachte eine gute Modellanpassung (Tab. 3). Die Faktorladungen bewegten sich in einem akzeptablen Bereich mit Werten von .52 bis .79 (Tab. 5). Damit konnte die durchgeführte konfirmatorische Faktorenanalyse die Ergebnisse der exploratorischen Faktorenanalyse und damit die achtfaktorielle Struktur bestätigen.

Tab. 5: Faktorstruktur der Unterrichtsmerkmale (Modell 2a)

Disziplin/Zeitnutzung		Zielklarheit		Regelklarheit		Diagnosekompetenz	
ZN1	.65**	ZIELK1	.69**	REGEL1	.55**	DKOMP2	.70**
ZN3	.75**	ZIELK2	.58**	REGEL2	.58**	DKOMP3	.65**
DISZ2	.57**	ZIELK3	.67**	REGEL3	.77**	DKOMP4	.79**
DISZ3	.56**	ZIELK4	.76**			DKOMP5	.71**
DISZ4	.72**					DKOMP6	.66**
DISZ5	.64**						
DISZ6	.59**						
DISZ7	.68**						
Fürsorglichkeit der Lehrperson		Schülermitbestimmung		Individuelle Bezugsnormorientierung		Differenzierung	
FSLP1	0.62**	SMITB1	0.68**	IBNO1	0.62**	DIFF1	0.62**
FSLP2	0.71**	SMITB2	0.78**	IBNO2	0.69**	DIFF2	0.52**
FSLP3	0.64**	SMITB4	0.76**	IBNO3	0.74**	DIFF3	0.70**
FSLP4	0.63**					DIFF4	0.78**

**< .001 * < .05

Die Interkorrelationen zwischen den acht latenten Faktoren fielen heterogen aus und bewegten sich zwischen $r = .01$ und $r = .71$ (Tab. 6). Während beispielsweise Unterrichtsmerkmale wie „Fürsorglichkeit durch die Lehrperson“ und „individuelle Bezugsnormorientierung“ hoch korrelierten, bestand zwischen der „Disziplin/Zeitnutzung“ und der „Differenzierung“ kein Zusammenhang.

Die Mittelwerte der acht Skalen bewegten sich zwischen 2.21 und 3.17. Das Unterrichtsmerkmal „Zielklarheit“ war am höchsten ausgeprägt, die „Differenzierung“ am niedrigsten, Decken- und Bodeneffekte bei den Skalen waren nicht zu erkennen. Die Cronbachs α -Werte wurden mit der vollständig geschätzten Stichprobe in Mplus berechnet und waren durchgängig zufriedenstellend. Einzig der Wert für die „Regelklarheit“ fiel mit .66 niedrig aus, konnte jedoch aufgrund der geringen Itemzahl von drei Items als noch akzeptabel bewertet werden (Tab. 6).

Die Unterschiede zwischen den Schulklassen lassen sich durch die Intraklassenkorrelation (Intra-class Correlation Coefficient, [ICC]) charakterisieren, welche die Varianz zwischen Klassen zu der Gesamtvarianz in Beziehung setzt. Ein hoher ICC-Wert bedeutet, dass es bei dem entsprechenden Merkmal große Unterschiede zwischen Klassen gibt, deren Ursache auf Klassenebene (z. B. Klassenzusammensetzung) zu suchen ist. Die ICC-Werte zwischen .13 und .29 zeigten (Tab. 6), dass es mittlere bis hohe Unterschiede zwischen den Klassen gibt. Raudenbush und Bryk (2002) emp-

fehlen ab einer ICC > .05 für weiterführende Analysen die Verwendung von hierarchisch linearen Modellen (HLM), um die Mehrebenenstruktur der Daten zu berücksichtigen.

Tab. 6: Korrelation der latenten Variablen und psychometrische Kennwerte der Skalen

	(1) Disziplin/ Zeitnutzung	(2) Ziel- klarheit	(3) Regel- klarheit	(4) Diagnose- kompetenz	(5) Fürsorg- lichkeit Lehrer	(6) Schülermit- bestimmung	(7) Bezugs- normorientie- rung	(8) Differenzie- rung
(1)	1							
(2)	.24**	1						
(3)	.29**	.68**	1					
(4)	.31**	.62**	.64**	1				
(5)	.36**	.31**	.27**	.36**	1			
(6)	.26**	.26**	.17**	.29**	.70**	1		
(7)	.36**	.31**	.21**	.29**	.71**	.55**	1	
(8)	.01	.11*	.13*	.21**	.43**	.34**	.42**	1
MZP	t_1	t_2	t_2	t_2	t_1	t_1	t_1	t_1
M^a	2.96	3.17	3.08	2.78	2.87	2.67	3.06	2.21
SE^a	.02	.02	.02	.02	.02	.02	.02	.02
Varianz^a	.40	.41	.43	.41	.43	.49	.45	.54
C's α^a	.85	.77	.66	.83	.75	.78	.72	.75
ICC^b	.27	.17	.17	.19	.21	.29	.15	.13

^a Auf Basis der vollständig geschätzten Stichprobe
^b Auf Basis der manifesten Faktormittelwerte

Aufbauend auf der konfirmatorischen Faktorenanalyse des Modells 2a, wurden im **Modell 2b** die nichtstandardisierten Ladungen innerhalb eines Faktors gleich gesetzt (essentiell tauäquivalente Ladungen). Trotz dieser starken Restriktion war die Modellanpassung noch zufriedenstellend (Tab. 3). Die Berechnung eines Faktormittelwerts war damit statistisch zulässig, ohne die faktorielle Struktur der Testbatterie zu verändern. Demnach muss in der Anwendung der vorliegenden Testbatterie nicht zwingend eine latente Modellierung der Unterrichtsmerkmale vorgenommen werden. Durch die Verwendung der manifesten Faktormittelwerte wird die Handhabung der Testbatterie deutlich vereinfacht.

4.3 | ÜBERPRÜFUNG DER STRUKTUR DER DIMENSIONEN VON UNTERRICHTSQUALITÄT

Aufbauend auf der konfirmatorischen Faktorenanalyse des Modells 2a, wurden im **Modell 3a** die Interkorrelationen zwischen den acht latenten Faktoren (= Unterrichtsmerkmale) durch zwei übergeordnete Faktoren zweiter Ordnung (= Dimensionen von Unterrichtsqualität) ersetzt. Damit wurden die Unterrichtsmerkmale a priori den übergeordneten Dimensionen der Unterrichtsqualität theoriegeleitet zugeordnet und auf ihre Modellanpassung geprüft. Das Messmodell (Beziehung der Testitems zu den latenten Faktoren der Unterrichtsmerkmale) blieb davon unbeeinflusst. Die Struktur dieses Modells konnte aufgrund der guten Modellanpassung bestätigt werden (Tab. 3).

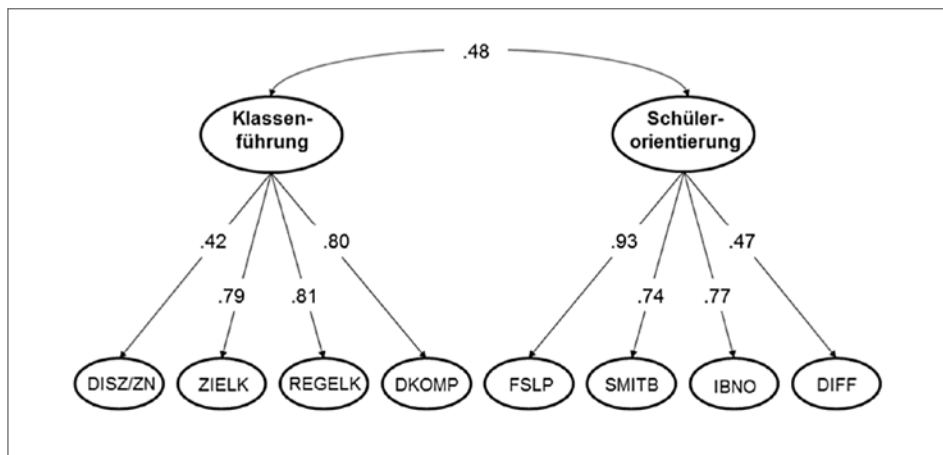


Abb. 1: Konfirmatorische Faktorenanalyse zweiter Ordnung (Modell 3a)

Beim Faktor zweiter Ordnung *Klassenführung* verfügten die Unterrichtsmerkmale „Zielklarheit“, „Regelklarheit“ und „Diagnosekompetenz“ über hohe und homogene Faktorladungen zwischen .79 und .80. Das Unterrichtsmerkmal „Disziplin & Zeitnutzung“ fiel mit .42 etwas ab. Beim Faktor zweiter Ordnung *Schülerorientierung* waren die Faktorladungen der Unterrichtsmerkmale „Fürsorglichkeit durch die Lehrperson“, „Schülermitbestimmung“ und „individuelle Bezugsnormorientierung“ zwischen .74 und .93, während der Faktorwert des Unterrichtsmerkmals „Differenzierung“ mit .47 geringer ausfiel. Die Interkorrelation der beiden latenten Faktoren zweiter Ordnung lag bei .48 und zeigte, dass diese sich in zwei Faktoren trennen ließen⁶, jedoch nicht unverbunden nebeneinander standen (Abb. 1).

Die im Modell 3a gezeigte Struktur wurde im **Modell 3b** auf Basis der manifesten Faktormittelwerte der Unterrichtsmerkmale erster Ordnung berechnet. Hierfür wurden die Faktormittelwerte in eine zweifaktorielle Faktorenanalyse eingebracht. Das Modell 3b verfügte über eine zufriedenstellende Modellanpassung (Tab. 3). Wenngleich diese deutlich unter der Modellanpassung des latenten Modells 3a liegt, wurde deutlich, dass sich diese zweifaktorielle Struktur auch auf manifester Ebene zeigen lässt.

Aufbauend auf Modell 3b, wurde abschließend im **Modell 3c** eine konfirmatorische Faktorenanalyse für hierarchisch geschachtelte Daten berechnet. Durch die resultierende Korrektur der Standardfehler verbesserte sich die Modellanpassung im Vergleich zum Modell 3b. Dies zeigte, dass die zweifaktorielle Struktur auch unter Berücksichtigung der Mehrebenenstruktur erhalten bleibt.

6 Zur Absicherung dieses Ergebnisses wurde das Modell 3a mit nur einem Faktor 2. Ordnung (g-Faktor) berechnet. Die Modellfits ($\chi^2 = 1447.78$, $df = 519$, $p > .05$, $CFI = 899$, $RMSEA = .042$, $SRMR = .072$) fallen deutlich schlechter aus als in Modell 3a. Dies bestätigt zusätzlich die zweifaktorielle Struktur 2. Ordnung.

5 | DISKUSSION⁷

Ausgangspunkt des vorliegenden Beitrags waren die Forderungen der Bildungs- und Schulpolitik nach empirischen Belegen für die Effektivität des Schulsystems. Es besteht dabei Konsens, dass neben dem Output auch die Prozessqualität von Schule und Unterricht bestimmt werden muss. Dabei sind die einzelnen Schulfächer gefordert, Messinstrumentarien zu entwickeln, die einerseits Anschluss an die Erkenntnisse der fachübergreifenden Unterrichtsforschung finden und andererseits Anpassungen an die Spezifika des Fachs Sport vornehmen. Hierfür wurden etablierte Fragebogeninstrumente aus der empirischen Unterrichts- und Bildungsforschung ausgewählt und an die Belange des Fachs Sport angepasst. Dabei erfolgte mit der Erfassung von Merkmalen der *Klassenführung* und der *Schülerorientierung* zunächst eine Reduzierung auf zwei der drei betrachteten Dimensionen von Unterrichtsqualität. Die dritte Dimension *kognitive Aktivierung* musste zunächst außen vor gelassen werden, da eine fachspezifische Definition und theoretische Fundierung dieser Dimension im Fach Sport bislang noch nicht ausreichend vollzogen worden ist.

Die dargestellte empirische Prüfung konnte zeigen, dass die Umformulierung der Testinstrumente und Testitems auf die Bedingungen des Sportunterrichts für die beiden ausgewählten Dimensionen *Klassenführung* und *Schülerorientierung* gelang. Exploratorische und konfirmatorische Faktoranalysen konnten die Konstruktvalidität der Merkmale und der übergeordneten Dimensionen der Unterrichtsqualität bekräftigen. Einzelne Testitems wurden aufgrund psychometrischer Kriterien aus den Skalen ausgeschlossen. Bei inhaltlicher Betrachtung wurde deutlich, dass diese Items die Prozesse und Handlungen des Sportunterrichts nicht widerspiegeln (z. B. „Im Sportunterricht können wir ungestört arbeiten“).

Beachtenswert ist vor allem, dass sich die acht Unterrichtsmerkmale trotz ihrer inhaltlichen Nähe in der Wahrnehmung der Schüler und teilweise identischer Itemstämme (viele Items begannen mit „Unser Sportlehrer ...“) statistisch eindeutig trennen lassen. Dies ist ein Hinweis darauf, dass Schüler der Sekundarschule I in der Lage sind, Unterricht in seinen verschiedensten Facetten hinreichend differenziert zu beurteilen.

Für die Anwendung der vorliegenden Testbatterie ist eine latente Modellierung der Unterrichtsmerkmale nicht zwingend nötig und kann durch die manifeste Berechnung von Faktormittelwerten ersetzt werden. Die Verwendung der manifesten Skalenmittelwerte erleichtert die Anwendung der Testbatterie auch für Zwecke der internen und externen Evaluation. Bei der Erhebung von hierarchischen Daten im Kontext der Schule (Schüler sind Klassen zugeordnet) ist jedoch die Mehrebenenstruktur zu beachten, um die Unterschiede zwischen den Klassen und die Unterschiede zwischen den Schülern aufklären zu können (Hochweber & Hartig, 2012). Auf Basis der statistischen Prüfung lässt sich schlussfolgern, dass die vorgestellte Testbatterie ein geeignetes Mittel ist, um Merkmale guten Sportunterrichts zu erfassen.

⁷ Das Autorenteam möchte sich bei den konstruktiven GutachterInnen bedanken, durch deren inhaltliche und methodische Anmerkungen die Qualität des Beitrags deutlich gesteigert werden konnte.

Abschließend soll auf Limitationen aufmerksam gemacht werden, die die Aussagekraft der Ergebnisse begrenzen und weitere Forschungsbemühungen anregen.

Aufgrund des aus zwei Messzeitpunkten zusammengesetzten Datensatzes und dem Vorliegen einer Mehrebenenstruktur musste auf eine Kreuzvalidierung (randomisierte Teilung der Stichprobe und Durchführung der exploratorischen und konfirmatorischen Faktorenanalysen an unterschiedlichen Stichproben) verzichtet werden. Hier bedarf es weiterführender Studien, welche die vorliegende Testbatterie anwenden und die in dieser Studie gefundene Struktur an einer unabhängigen Stichprobe überprüfen. Weiterhin konnten mit der vorliegenden Studie mit der *Klassenführung* und der *Schülerorientierung* nur zwei der drei betrachteten Dimensionen der Unterrichtsqualität analysiert werden. Ausstehend ist aus unserer Sicht die fachspezifische Anpassung der Dimension *kognitive Aktivierung*. Dies stellt eine besondere Herausforderung dar, da der Kernbestand des Sportunterrichts eben nicht primär im kognitiven, sondern vorwiegend im motorischen Bereich liegt. Diese fachlichen und inhaltsbezogenen Faktoren des Unterrichts spielen jedoch für die Erklärung von Schülerleistungen eine zentrale Rolle (Seidel & Shavelson, 2007). Demnach sollten zukünftige Forschungsbemühungen – neben den allgemeinen Merkmalen der Unterrichtsqualität – fachspezifische Aspekte der Unterrichtsqualität im Fach Sport zunehmend empirisch untersuchen und sich vor allem dieser Dimension mit der notwendigen fachspezifischen Expertise widmen (Gräsel & Göbel, 2011).

Zudem ist darauf hinzuweisen, dass in der empirischen Unterrichtsforschung die Qualitätskriterien des Unterrichts in einem Wechselspiel von theoretischer Fundierung und empirischer Analyse gewonnen wurden. Im vorliegenden Beitrag konnte nur die Konstruktvalidität des Instrumentariums nachgewiesen werden. Die finale Hürde liegt zukünftig in Wirkungsanalysen, in denen die Effekte der Dimensionen von Unterrichtsqualität auf die Lernleistung und Motivation von Schülern empirisch untersucht werden. Die Grunddimension *Schülerorientierung* ist dabei wahrscheinlich primär motivationsfördernd, *kognitive Aktivierung* ist für die Lernleistung besonders relevant, während die *Klassenführung* vermutlich die Voraussetzung für beides darstellt (Klieme & Rakoczy, 2008; vgl. auch Heemsoth, 2014). Eine abschließende Prüfung der prognostischen Validität, wie in anderen Fächern (u. a. Brunner et al., 2006; Rakoczy et al., 2007), mithilfe von geeigneten Outputmerkmalen im motorischen, motivational-volitionalen oder kognitiven Kompetenzbereich steht derzeit noch aus und sollte das zentrale Ziel zukünftiger Forschungsbemühungen sein.

Eine Übereinstimmungsvalidität der vorliegenden Testbatterie mithilfe von Beobachtungsinstrumenten, wie dem Classroom Assessment Scoring System (CLASS, Pianta et al., 2008; in der Sportpädagogik Richartz & Zöller, 2011), bietet eine interessante Forschungsperspektive. Je nach Grad der empirischen Bewährung der einzelnen Erhebungsverfahren sollte dies längerfristig zu einem Methodenmix in der Erfassung von Unterrichtsqualität führen.

Die Dimensionen von Unterrichtsqualität werden im Kontext des Prozess-Mediations-Produkt-Paradigmas als Mediatoren zwischen der professionellen Handlungskompetenz der Lehrer und der Lernleistungen der Schüler begriffen (Kunter et al., 2013a; Seiler, Herrmann & Gerlach, 2014).

Damit sind die vorliegenden Forschungsbemühungen unmittelbar anschlussfähig an die Diskussionen um professionelle Kompetenzen von Sportlehrkräften. Den theoretischen Rahmen bildet hier das „Modell professioneller Handlungskompetenz“ von Baumert und Kunter (2006, S. 482), das Professionswissen, Überzeugungen/Werthaltungen, motivationale Orientierungen und selbst-regulative Fähigkeiten unterscheidet. Im Rahmen der „COACTIV-Studie“ konnten Wirkungen der professionellen Handlungskompetenz auf die Dimensionen von Unterrichtsqualität empirisch bestätigt werden (u. a. Kunter et al., 2013b). Auch in der Sportpädagogik wird seit Kurzem die Erfassung von Professionskompetenz von Sportlehrpersonen von verschiedenen Arbeitsgruppen intensiv bearbeitet (Baumgartner, 2013; Meier, 2013; Oesterheld, Gröschner, Seidel & Sygusch, 2012; Messmer & Brea, 2014). Hierbei könnten mit dem vorgelegten Instrument die mediierenden Wirkungen der erfassten Professionskompetenz überprüft werden – ob also eine „kompetente“ Lehrperson in der Wahrnehmung der Schüler wirklich einen besseren Sportunterricht durchführt (Kunter et al., 2013b).

Abschließend lässt sich feststellen, dass unserer Ansicht nach hiermit eine Studie vorgelegt werden konnte, in der Merkmale der Unterrichtsqualität im Fach Sport empirisch geprüft und hinsichtlich ihrer Beziehung untersucht wurden. Das entwickelte Instrumentarium bietet die Möglichkeit, einen wichtigen Teil der Prozessqualität von Sportunterricht zu erfassen und eine empirische Bildungsforschung in der Sportwissenschaft voranzubringen.

LITERATUR

- Altrichter, H. & Maag Merki, K. (2010). Steuerung der Entwicklung des Schulwesens. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 15-39). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Balz, E. (2009). Fachdidaktische Konzepte update oder: Woran soll sich der Schulsport orientieren? *sportpädagogik*, 33 (1), 25-32.
- Balz, E. (2010). Guter Sportunterricht – Merkmale und Beispiele. *sportpädagogik*, 34 (2), 50-53.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469-520.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47, 133-180.
- Baumgartner, M. (2013). Kompetenzprofile von Sportlehrpersonen der Berufsfachschulen. In F. Oser, T. Bauder, P. Salzmann & S. Heinzer (Hrsg.), *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzprofilen bei Lehrpersonen und Berufsbildungsverantwortlichen* (S. 96-126). Bad Heilbrunn: Klinkhardt.
- Borich, G. D. (2007). *Effective teaching methods* (6. ed). Upper Saddle River, NJ: Pearson Merrill.
- Bos, W. & Buddeberg, I. (2005). *IGLU. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Münster, München [u. a.]: Waxmann.
- Brophy, J. E. (2000). *Teaching*. Brüssel: International Academy of Education.
- Brunner, M., Kunter, M., Krauss, S., Klusmann, U., Baumert, J., Blum, W., ... Tsai, Y.-M. (2006). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht. Eine Zwischenbilanz des COACTIV-Projekts.

- In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 54-82). Münster: Waxmann.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München [u. a.]: Pearson Studium.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39, 223-228.
- Desimone, L. M., Smith, T. M. & Frisvold, D. E. (2010). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy*, 24, 267-329.
- Ditton, H. & Merz, D. (2000). *Qualität von Schule und Unterricht. Kurzbericht über erste Ergebnisse einer Untersuchung an bayerischen Schulen*. Zugriff am 14. November 2011 unter <http://www.quassu.net/Bericht1.pdf>.
- Funke-Wieneke, J. (2003). „Alle Zöglinge sind gemessen, die Aufgaben sind damit in gehöriges Verhältnis gebracht“. Leisten und Leistung im Bewegungsunterricht. *sportpädagogik*, 27 (5), 4-9.
- Gebken, U. (2005). Guter Sportunterricht für alle! In A. Gogoll (Hrsg.), *Qualität im Schulsport* (S. 234-239). Hamburg: Czwalina.
- Gerlach, E. (2005). Prima Klima? Einflussgrößen und Effekte. *Sportunterricht*, 54, 243-247.
- Gogoll, A. (2012). Sport- und bewegungskulturelle Kompetenz - ein Modellentwurf für das Fach Sport. In A. C. Richter, E. Balz, J. Frohn & P. Neumann (Hrsg.), *Kompetenzorientiert Sport unterrichten* (S. 39-52). Aachen: Shaker.
- Gräsel, C. (2011). Was ist Empirische Bildungsforschung? In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Strukturen und Methoden* (Lehrbuch, S. 13-27). Wiesbaden: VS.
- Gräsel, C. & Göbel, K. (2011). Unterrichtsqualität. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung: Gegenstandsbereiche* (S. 87-98). Wiesbaden: VS.
- Hamre, B. K. & Pianta, R. C. (2010). Classroom environments and developmental processes. Conceptualization and measurement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling, and human development* (pp. 25-41). New York: Routledge.
- Hattie, J. (2008). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Heemsoth, T. (2014). Unterrichtsklima als Mediator des Zusammenhangs von Klassenführung und Motivation im Sportunterricht. *Psychologie in Erziehung und Unterricht*, 61 (3), 203-215.
- Heemsoth, T. & Miethling, W.-D. (2012). Schülerwahrnehmungen des Unterrichtsklimas. Entwicklung eines Fragebogens und Befunde zum Sportunterricht. *Sportwissenschaft*, 42 (4), 228-239.
- Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (3. Aufl.). Seelze-Velber: Klett-Kallmeyer.
- Herrmann, C. & Gerlach, E. (2014). Motorische Basiskompetenzen in der Grundschule. Pädagogische Zielentscheidung und Aufgabenentwicklung. *Sportunterricht*, 63 (11) 322-328.
- Herrmann, C., Leyener, S. & Gerlach, E. (2014). *IMPEQT-Studie (Implementation of Physical Education and the Quality of Teaching). Dokumentation der Erhebungsinstrumente*. Zugriff am 12. September 2014 unter <http://edoc.unibas.ch/dok/A6289245>.
- Hochweber, J. & Hartig, J. (2012). Mehrebenenanalyse. In S. Maschke & L. Stecher (Hrsg.), *Enzyklopädie Erziehungswissenschaft Online (EEO), Fachgebiet Methoden der empirischen erziehungswissenschaftlichen Forschung*. Weinheim: Juventa.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6 (1), 1-55.
- Kiel, E. (2010). Unterrichtsforschung. In: R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (3., durchgesehene Aufl.) (S. 773-790). Wiesbaden: Springer.
- Klafki, W. (1985). *Neue Studien zur Bildungstheorie und Didaktik. Beiträge zur kritisch-konstruktiven Didaktik*. Weinheim: Beltz.

- Kleindienst-Cachay, C. & Kurz, D. (2005). Vorwort. In A. Gogoll (Hrsg.), *Qualität im Schulsport* (S. 9-10). Hamburg: Czwalina.
- Klieme, E. (Hrsg.). (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim: Beltz.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., ... Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Expertise*. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 128-146). Münster: Waxmann.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54 (2), 222-237.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur“ und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.), *TIMSS – Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (BMBF publik, S. 43-57). Bonn.
- Köller, O. (2008). Bildungsstandards in Deutschland: Implikationen für die Qualitätssicherung und Unterrichtsqualität. In M. A. Meyer, M. Prenzel & S. Hellekamps (Hrsg.), *Perspektiven der Didaktik* (S. 47-59). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kounin, J. S. (1976). *Techniken der Klassenführung*. Bern: Huber.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9, 231-251.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2013a). *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV Project*. New York, NY: Springer.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013b). Professional competence of teachers: Effects on instructional quality and student Development. *Journal of Educational Psychology*, 105 (3), 805-820.
- 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 & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern. Ausbildung und Beruf* (S. 47-70). Weinheim: Beltz Verlag.
- Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 73-101). Berlin, Heidelberg: Springer.
- Lipowsky, F., Rakoczy, K., Buff, A. & Klieme, E. (2005). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“: Pythagoras*. Frankfurt am Main: DIPF.
- Meier, S. (2013). Kompetenzentwicklung angehender Sportlehrkräfte. In F. Mess, M. Gruber & A. Woll (Hrsg.), *Sportwissenschaft grenzenlos?!* (S. 211). Hamburg: Czwalina.
- Messmer, R. & Brea, N. (2014). Fachdidaktisches Können von Sportlehrpersonen – ein Kompetenzmodell. *Zeitschrift für Sportpädagogische Forschung*, 2, 63-76.
- Meyer, H. (2005). *Was ist guter Unterricht?* (3., Aufl). Frankfurt am Main: Scriptor.
- Muthén, L. K. & Muthén, B. O. (2012). *Mplus user's guide. Statistical analysis with latent variables* (7. Aufl.). Los Angeles: Muthén & Muthén.
- Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 25, 267-316.
- Oelkers, J. & Reusser, K. (2008). *Qualität entwickeln – Standards sichern – mit Differenz umgehen*. Berlin: BMBF.
- Oesterheld, V., Gröschner, A., Seidel, T. & Sygusch, R. (2012). Pädagogische Vorerfahrungen und Kompetenzeinschätzungen im Kontext eines Praxissemesters - Domänenspezifische Betrachtungen am Beispiel der Sportlehrerausbildung. *Lehrerbildung auf den Prüfstand*, 1, 29-46.

- Pianta, R. C., La Paro, Karen M & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS) manual, pre-K*. Baltimore, MD: Paul Brookes Publishing.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Prohl, R. (2010). Fachdidaktische Konzepte des Sportunterrichts. In N. Fessler, A. Hummel & G. Stibbe (Hrsg.), *Handbuch Schulsport* (S. 169-179). Schorn-dorf: Hofmann.
- Rakoczy, K. (2006). Motivationsunterstützung im Mathematikunterricht: Zur Bedeutung von Unterrichtsmerkmalen für die Wahrnehmung von Schülerinnen und Schülern. *Zeitschrift für Pädagogik*, 52 (6), 823-843.
- Rakoczy, K., Klieme, E., Drollinger-Vetter, B., Lipowsky, F., Pauli, C. & Reusser, K. (2007). Structure as a quality feature in mathematics instruction: Cognitive and motivational effects of a structured organization of the learning environment vs. a structured presentation of learning content. In M. Prenzel (Hrsg.), *Studies on the educational quality of schools. The final report on the DFG priority program* (pp. 102-121). Münster: Waxmann.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.
- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität: Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 9-32). Münster: Waxmann.
- Richartz, A. & Zöllner, R. (2011). Die Qualität von Unterrichtsprozessen erfassen. In H. Lange, T. Duttler, A. Leffler, A. Siebe & M. Zimlich (Hrsg.), *Bewegungsbezogene Bildungskonzeptionen. Positionen und Analysen im Spannungsfeld zwischen Konzeption, Implementation und Evaluation* (S. 75-87). Baltmannsweiler: Schneider.
- Saldern, M. & Litting, K. E. (1985). Die Konstruktion der Landauer Skalen zum Sozialklima. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 17 (2), 138-149.
- Scheid, V. & Prohl, R. (2011). *Sportdidaktik. Grundlagen, Lehrplan, Bewegungsfelder*. Wiebelsheim, Hunsrück: Limpert.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23-74.
- Schreiber, J. B., Stage, F. K., King, J., Nora, A. & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99 (6), 323-337.
- Schwarzer, R., Lange, B. & Jerusalem, M. (1982). Die Bezugsnorm des Lehrers aus der Sicht des Schülers. In F. Rheinberg (Hrsg.), *Jahrbuch fuer Empirische Erziehungswissenschaft. Bezugsnormen zur Schulleistungsbewertung. Analyse u. Intervention* (S. 161-172). Düsseldorf: Paedag. Verl. Schwann.
- 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.
- Seiler, S., Herrmann, C. & Gerlach, E. (in Vorbereitung). *Influence of classroom management and student orientation on output indicators of 7th grade students*.
- Stronge, J. H., Tucker, P. D. & Hindman, J. L. (2004). *Handbook for qualities of effective teachers*. Alexandria/VA: ASCD.
- Wagner, W., Helmke, A. & Rösner, E. (2009). *Deutsch Englisch Schülerleistungen international. Dokumentation der Erhebungsinstrumente für Schülerinnen und Schüler, Eltern und Lehrkräfte*. Frankfurt am Main: DIPF.
- Wolters, P. & Kemna, P. (2011). Qualitätskriterien für den Sportunterricht – Entwurf eines Rahmenmodells und einer Skala zur erlebten Sicherheit. In K.-O. Bauer & N. Logemann (Hrsg.), *Unterrichtsqualität und fachdidaktische Forschung. Modelle und Instrumente zur Messung fachspezifischer Lernbedingungen und Kompetenzen* (S. 159-186). Münster, New York, NY, München, Berlin: Waxmann.
- Wolters, P., Klinge, A., Klupsch-Sahlmann, R. & Sinning, S. (2009). Was ist nach unseren Vorstellungen guter Sportunterricht? *Sportunterricht*, 58 (3), 67-72.