

Forum Kind Jugend Sport 2023 · 4:88–101  
<https://doi.org/10.1007/s43594-023-00110-2>  
 Eingegangen: 10. Mai 2023  
 Angenommen: 18. September 2023  
 Online publiziert: 12. Oktober 2023  
 © The Author(s) 2023



C. Herrmann<sup>1</sup> · S. Seiler<sup>2</sup> · A. Siffert<sup>3</sup> · L. C. Dapp<sup>4</sup> · V. Gashaj<sup>5</sup> · F. Studer<sup>3</sup>

<sup>1</sup> Forschungsgruppe Didaktik Bewegung und Sport, Pädagogische Hochschule Zürich, Zürich, Schweiz

<sup>2</sup> Bundesamt für Sport BASPO, Jugend- und Erwachsenensport, Magglingen, Schweiz

<sup>3</sup> Eidgenössische Hochschule für Sport Magglingen EHSM, Bundesamt für Sport BASPO, Magglingen, Schweiz

<sup>4</sup> Institut für Psychologie, Universität Bern, Bern, Schweiz

<sup>5</sup> School of Science, Loughborough University, Loughborough, Großbritannien

# Entwicklung und Validierung eines Evaluationsinstruments zur Erfassung der Qualität des Lehrens und Lernens im Schweizer Kinder- und Jugendsport

## Zusatzmaterial online

Die Online-Version dieses Artikels (<https://doi.org/10.1007/s43594-023-00110-2>) enthält den Artikel in englischer Sprache.

## Jugend und Sport

Jugend und Sport (J+S) ist das größte Schweizer Sportförderungssystem und europaweit einzigartig. Es vereint über 90 Sportarten in einem staatlich subventionierten Ausbildungssystem, das strategisch und inhaltlich vom Bundesamt für Sport (BASPO) geregelt wird. Das Ziel des J+S-Sportförderungsmodells ist es, das Sport- und Bewegungsverhalten junger Menschen langfristig positiv zu beeinflussen. Dabei sollen sich Kinder und Jugendliche ganzheitlich entwickeln und sich in eine Sportgemeinschaft integrieren können. J+S-Leiter\*innen werden in ihrer Sportart ausgebildet und bieten Aktivitäten für 5- bis 20-jährige Kinder und Jugendliche in organisierten, außerunterrichtlichen und freiwilligen schulischen Sportangeboten an. Die

Dieser Forschungsbeitrag hat vor der Annahme zur Veröffentlichung ein double-blind Peer-Review-Verfahren durchlaufen.

Grundausbildung zu J+S-Leiter\*in dauert sechs Tage und wird in der jeweiligen Sportart durchgeführt. Die Teilnehmenden werden in den drei Kompetenzbereichen „Sich als J+S-Leiter\*in engagieren“, „Die Sportart verstehen“ und „Die Sportart unterrichten“ ausgebildet. Nach dieser Grundausbildung können sich die J+S-Leiter\*innen in der modularen Weiterbildung in diesen drei Kompetenzbereichen sportartspezifisch oder sportartübergreifend vertiefen (Bundesamt für Sport 2023).

Für die Qualitätssicherung von J+S-Aktivitäten wurde ein J+S-Ausbildungsverständnis entwickelt, welches sich am Magglinger Ausbildungsmodell (Bundesamt für Sport 2019) orientiert und die dortigen Handlungsbereiche integriert. Das J+S-Ausbildungsverständnis richtet sich an wissenschaftlichen Ansprüchen der Unterrichtsforschung und dem etablierten Angebot-Nutzungsmodell (Helmke 2017) schulischer Wirkungsforschung aus. Entsprechend werden qualitätsvolle J+S-Aktivitäten als effektive Lernangebote begriffen, die fachspezifische (sportartbezogene) und fachübergreifende (persönlichkeitsbildende) Kompetenzen sowie erzieherische Effekte anstreben. Bisher fehlt ein

validiertes Evaluationsinstrument, um die Erreichung dieser Ziele empirisch zu überprüfen. In diesem Beitrag soll mit der Konzeption, Entwicklung und Validierung eines Evaluationsinstruments zur Erfassung der Qualität des Lehrens und Lernens bei J+S ein Beitrag dazu geleistet werden, diese normativ-theoretischen Wirkungsannahmen empirisch zu evaluieren.

## Qualität des schulischen und außerschulischen Sports

### Qualitätsverständnis der Unterrichtsforschung

Die Unterrichtsqualitätsforschung entwickelte verschiedene Konzeptualisierungen zur Beschreibung von Unterrichtsqualität, welche weitgehend stufen- und fachübergreifend verstanden wird (im Überblick Charalambous und Praetorius 2018). In verschiedenen Studien (u. a. TIMSS, Klieme und Baumert 2001; DESI, Klieme 2008; COACTIV, Kunter et al. 2013) wurden drei Basisdimensionen von Unterrichtsqualität identifiziert und zu einem theoretisch und empirisch fundierten Modell verknüpft (im Überblick Praetorius et al.

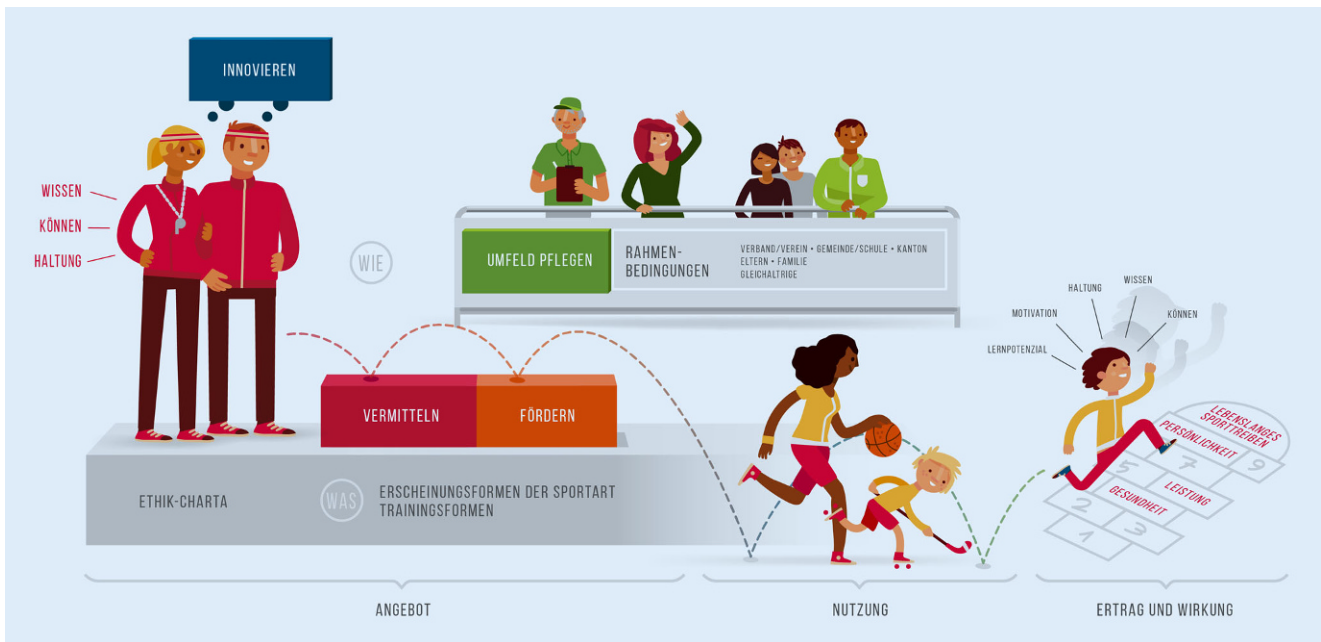


Abb. 1 ▲ J+S-Ausbildungsverständnis Jugendausbildung. (Abbildung: Bundesamt für Sport BASPO)

2018). Qualitätsvoller Unterricht zeichnet sich durch (a) strukturierte, klare und störungspräventive *Unterrichts- und Klassenführung*, (b) unterstützendes, *schülerorientiertes Sozialklima* und (c) das *Potenzial zur kognitiven Aktivierung* aus. Diese Dimensionen können als eine empirisch verankerte Systematik didaktischer Prinzipien verstanden werden, die als geeignet erscheint, um sowohl theoretische Konzepte als auch Wirkungsannahmen der Unterrichtsforschung zu vereinen (Klieme 2008). Diese drei international etablierten Basisdimensionen wurden mittlerweile sowohl theoretisch wie auch empirisch in verschiedenen Fachdidaktiken ausdifferenziert (im Überblick Praetorius et al. 2018).

Auch in der Unterrichtsqualitätsforschung im schulischen und außerschulischen Sport werden die Basisdimensionen als theoretisches Rahmenmodell genutzt (im Überblick Herrmann und Gerlach 2020). Dabei stellt sich die Frage, inwiefern dieses Modell und entsprechende Evaluationsinstrumente aus der Sportunterrichtsforschung auf außerunterrichtliche Sportangebote übertragen werden können. Hier argumentieren Richartz, Maier und Kohake (2021b), dass das außerschulische, sportliche Training ähnliche Ziele wie der schulische (Sport-)Un-

terricht verfolgt. Dazu gehören nicht nur motorische Kompetenzen und sportpezifische Fertigkeiten, sondern auch motivationale Aspekte und sozial-kognitive Lernprozesse.

### Qualitätsverständnis bei Jugend und Sport

Das J+S-Ausbildungsverständnis (Abb. 1; ausführliche Informationen unter [www.jugendundsport.ch](http://www.jugendundsport.ch)), basiert auf den vier *Handlungsbereichen* Innovieren, Vermitteln, Fördern und Umfeld pflegen (farbige Rechtecke). Diese Handlungsbereiche sind der pädagogisch-didaktische Kern der J+S-Ausbildung und bilden den Kompetenzbereich „Sportart unterrichten“ ab. Jeder der oben genannten vier Handlungsbereiche wird durch konkrete *Handlungsfelder*, in denen die J+S-Leitenden aktiv werden, weiter untergliedert. Danach haben J+S-Leitende vor allem die Sportart zu **vermitteln** und die Teilnehmenden zu **fördern**. Durch Reflektion und **Innovation** tragen sie zur persönlichen Entwicklung, zur Weiterentwicklung der Sportart oder des Vereinslebens bei. Weiter sollen sie den Austausch mit dem naheliegenden **Umfeld pflegen** (Vereinszugehörige, Eltern, Verband etc.).

Die drei Basisdimensionen der Unterrichtsqualität (Klassenführung, schüler\*innenorientiertes Klima und kognitive Aktivierung) sind im J+S-Ausbildungsmodell im Handlungsbereich *Vermitteln* verankert und werden in den folgenden vier Handlungsfeldern auf den Kontext von J+S adaptiert (J+S-Ausbildungsteam 2019; Tab. 7):

- „Lernförderliches Klima ermöglichen“,
- „Lernziele priorisieren und Struktur schaffen“,
- „Gruppe sicher und effizient führen“,
- „Attraktive Aufgaben stellen und Rückmeldungen geben“.

Damit das konkrete Handeln der J+S-Leitenden greifbar wird, wurden für alle vier Handlungsfelder praxisrelevante **Handlungsempfehlungen** formuliert, welche in der Unterrichtsqualitätsforschung den Subdimensionen oder Unterrichtsmerkmalen entsprechen. Beispielsweise wird das Handlungsfeld „Lernförderliches Klima ermöglichen“ in die Handlungsempfehlungen (a) Teilnehmende wertschätzen, (b) Teilnehmende einbeziehen, (c) Fehler nutzen und (d) Freude wecken ausdifferenziert (Tab. 7).

C. Herrmann · S. Seiler · A. Siffert · L. C. Dapp · V. Gashaj · F. Studer

## Entwicklung und Validierung eines Evaluationsinstruments zur Erfassung der Qualität des Lehrens und Lernens im Schweizer Kinder- und Jugendsport

### Zusammenfassung

Jugend und Sport (J + S) ist das größte Schweizer Sportförderungssystem und vereint in einem staatlich subventionierten und inhaltlich regulierten Ausbildungssystem über 90 Sportarten. Um die Qualität der Sportaktivitäten zu gewährleisten, bedarf es einer ansprechenden und qualitativ hochwertigen Ausbildung von J + S-Leiter\*innen. Im J + S-Ausbildungsmodell werden im Handlungsbereich „Vermitteln“ Ausbildungsinhalte zu qualitativ vollen J + S-Aktivitäten an die Leitenden weitergegeben, welche sich eng am aktuellen Forschungsstand der Unterrichtsqualitätsforschung orientieren. Zur systematischen Evaluation dieses Handlungsbereichs in der Praxis wurde das J + S-Evaluationsinstrument entwickelt.

Damit sollen die an J + S-Aktivitäten teilnehmenden Kinder und Jugendlichen mittels Fragebogenitems zu vier Handlungsbereichen digital befragt werden. Aufbauend auf drei qualitativen Pilotstudien wurde eine schweizweite quantitative Validierungsstudie durchgeführt. Es wurden zwei Umfragen mit  $N = 1230$  (53,7 % weiblich;  $M = 13,99$  Jahren,  $SD = 2,36$ ) und  $N = 851$  (53,7 % weiblich;  $M = 13,96$  Jahren,  $SD = 2,18$ ) Kinder und Jugendlichen durchgeführt. Zur Prüfung der faktoriellen Validität und Reliabilität des entwickelten J + S-Evaluationsinstruments wurden konfirmatorische Faktoranalysen (CFA) berechnet. Die CFA ergaben nach systematischer Itemreduktion durchgehend eine gute Modellanpassung.

Alle vier Handlungsbereiche mit ihren 17 Handlungsempfehlungen konnten mittels 51 Items valide und reliabel abgebildet werden, sodass das Evaluationsinstrument als valide und reliabel bewertet werden kann. Damit steht ein validiertes Evaluationsinstrument in deutscher, französischer und italienischer Sprache zur Verfügung, das zukünftig jährlich in der Schweiz in J + S-Aktivitäten eingesetzt werden kann.

### Schlüsselwörter

Organisierter Kinder- und Jugendsport · Fragebogenentwicklung · Unterrichtsqualität · Klassenführung · Unterrichtsklima · Aktivierung · Trainingsqualität

## Development and validation of an evaluation instrument for assessing the quality of teaching and learning in Swiss children's and youth sports

### Abstract

Youth and Sport (Y + S) constitutes Switzerland's most extensive sports support system, encompassing over 90 sports within a state-subsidized and content-regulated training system. Upholding the quality of sports activities necessitates effective (appealing and high-quality) education for Y + S instructors. The Y + S education model transmits teaching content for high-quality Y + S activities through the "Teaching" domain of action, which closely aligns with contemporary instructional quality research. We developed the Y + S evaluation instrument to systematically evaluate this field of action in practical scenarios. This questionnaire digitally

surveys children and youth participating in Y + S activities, employing items to assess four fields of action. We realized a nationwide quantitative validation study based on three qualitative pilot studies. Two surveys were administered to  $N = 1230$  (53.7% female;  $M = 13.99$  years,  $SD = 2.36$ ) and  $N = 851$  (53.7% female;  $M = 13.96$  years,  $SD = 2.18$ ) children and youths. We conducted confirmatory factor analyses (CFA) to assess the factorial validity and reliability of the developed Y + S evaluation instrument. Following systematic item reduction, CFAs consistently demonstrated good model fits. All four fields of action

with their 17 recommendations for action, were represented by 51 items, attesting to the validity and reliability of the evaluation instrument. In conclusion, a validated evaluation instrument is now available in German, French, and Italian and will be implemented annually to assess Y + S activities in Switzerland moving forward.

### Keywords

Organized children's and youth sport · Questionnaire development · Teaching quality · Classroom management · Classroom climate · Activation · Training quality

## Maßnahmen zur Qualitätssicherung bei Jugend und Sport

Ausgehend von den theoretischen Überlegungen der Unterrichtsforschung entsteht der Bedarf nach Instrumenten zur empirischen Qualitätssicherung von Sportunterrichts- und -angeboten. In den vergangenen Jahren wurden aufeinander aufbauende Studien durchgeführt, um ein Evaluationsinstrument zur Bewertung der Unterrichtsqualität im Fach

Sport zu entwickeln (u. a. Herrmann und Gerlach 2017; Herrmann et al. 2018, 2015). Auf Basis dieser Vorarbeiten wurde das QUALLIS-Instrument zur Evaluation der *Qualität* des *Lehrens* und *Lernens* im Sportunterricht entwickelt (Herrmann 2019; Herrmann et al. 2023a). Es beinhaltet zentrale Dimensionen und Merkmale von Unterrichtsqualität, die fachspezifisch ausdifferenziert wurden (Herrmann und Gerlach 2020; Herrmann et al. 2020) und welche für die Erreichung der Erziehungs- und Lern-

ziele des Sportunterrichts von Bedeutung sind.

Das QUALLIS-Instrument erfasst die Unterrichtsqualität über die Einschätzung von Unterrichtsmerkmalen anhand konkreter beobachtbarer Fragebogenitems. Die Einschätzungen werden dabei von (1) Lehrpersonen, (2) externen Beobachter\*innen und von (3) Schüler\*innen vorgenommen. Der Vergleich dieser drei Perspektiven und die Identifikation von Unterschieden zwischen den Einschätzungen bietet die Möglich-

keit, den Sportunterricht systematisch und kriteriengeleitet weiterzuentwickeln (Helmke und Lenske 2013; Herrmann 2019).

Inwiefern bestehende Evaluationsinstrumente zur Erfassung der Unterrichtsqualität im Fach Sport auf den organisierten, außerunterrichtlichen und freiwilligen schulischen Sport übertragen werden können, ist noch wenig diskutiert. Das Ziel dieses Beitrags war daher die Entwicklung eines Evaluationsinstrumentes zur Erfassung der Qualität des Lehrens und Lernens bei Jugend und Sport. Dazu wurde das bereits etablierte QUALLIS-Evaluationsinstrument für den Sportunterricht (Herrmann 2019; Herrmann et al. 2023a) an den Kontext von J+S-Aktivitäten angepasst.

Bei der ersten inhaltlichen Prüfung der Dimensionen, Subdimensionen und Testitems des QUALLIS-Evaluationsinstrumentes schien die Erfassung der relevanten J+S-Handlungsfelder und Handlungsempfehlungen zunächst hinreichend präzise zu sein, sodass zunächst von einer inhaltlichen Validität ausgegangen werden konnte. Jedoch waren Anpassungen der Testitems erforderlich (bspw. von Lehrer\*innen auf Leiter\*innen), um das QUALLIS-Evaluationsinstrument im Rahmen der J+S-Aktivitäten anwenden zu können. Zudem wurde das J+S-Evaluationsinstrument nach den sprachlichen Anpassungen – aufgrund der Mehrsprachigkeit in der Schweiz – auf Französisch und Italienisch übersetzt. Die Qualität und Eindeutigkeit der Übersetzungen wurde von Mitarbeitenden des BASPO mit Muttersprache Italienisch bzw. Französisch geprüft. Die vollständige Auflistung der Items findet sich in der online verfügbaren Itemdokumentation (Herrmann et al. 2023b). Für eine bessere Lesbarkeit werden im Folgenden die deutschsprachigen Items benannt, wengleich sich die Analysen auf alle drei Sprachversionen beziehen.

## Validierung des J+S-Evaluationsinstruments

Zur Validierung des J+S-Evaluationsinstrumentes wurden drei qualitative und eine quantitative Studie durchgeführt. Die **qualitativen Studien** dienten der

Überprüfung der Verständlichkeit und Umsetzbarkeit des auf J+S-Aktivitäten übertragenen QUALLIS-Instruments und zielten auf die inhaltliche Validität ab, indem untersucht wurde, ob die ausgewählten QUALLIS-Items die J+S-Handlungsempfehlungen inhaltlich korrekt erfassen. Nach jeder qualitativen Studie erfolgten Anpassungen am Fragebogen. Die **quantitative Studie** prüfte die faktorielle Validität und Reliabilität des entwickelten J+S-Evaluationsinstrumentes. Nachfolgend werden die Validierungsstudien und deren Resultate erläutert.

## Qualitative Pilotierungsstudien

Für die **erste qualitative Validierungsstudie** wurden drei Fokusgruppeninterviews geführt. Ziel war es zu prüfen, ob die Operationalisierung der Handlungsempfehlungen (z.B. Teilnehmende wertschätzen) in Items inhaltlich passend und vollständig vorgenommen wurde (Inhaltsvalidität). Dafür wurden insgesamt 16 Personen ausgewählt: acht Ausbildungsverantwortliche verschiedener Sportverbände, fünf J+S-Mitarbeiter\*innen in der Funktion „Leiter\*innen-Ausbildung Sportarten“ sowie drei Fachwissenschaftler\*innen aus den Bereichen Erziehungswissenschaft, Sportpädagogik und Educational Measurements. In jedem der drei Interviews nahmen fünf bis sechs Personen mit diesen unterschiedlichen Hintergründen teil, um eine maximale Kontrastierung der Rückmeldungen zu erreichen. Die Fokusgruppeninterviews wurden mittels eines halbstrukturierten Leitfadens (u. a. „Denken Sie, dass die Items die Qualität einer J+S-Aktivität in Ihrer Sportart gut repräsentieren?“; „Sehen Sie Probleme beim Einsatz des Instruments?“) durchgeführt und kategorienbasiert inhaltsanalytisch ausgewertet (Mayring 2015).

Die Expert\*innen bestätigten die inhaltliche Passung des Evaluationsinstrumentes in Bezug auf den Handlungsbereich „Vermitteln“ des J+S-Ausbildungsverständnisses (Bundesamt für Sport 2023). Die Übertragung des bereits empirisch validierten QUALLIS-

Instruments aus dem Schulsportkontext auf J+S wurde als sinnvoll erachtet.

Die Sportartenvertreter\*innen bezeichneten die Qualitätskriterien überwiegend als passend für ihre Sportart. Abweichende Meinungen (z. B. in Bezug auf den Aspekt Sicherheit) wurden für die Überarbeitungsphase aufgenommen. Weiter wiesen die Expert\*innen auch auf mögliche selbstdienliche Verzerrungen hin, die bei Selbstbeurteilungen durch die J+S-Leitenden auftreten könnten. Die Beurteilung aus den drei unterschiedlichen Perspektiven (a) *Leitende*, (b) *Beobachter\*innen* und (c) *Teilnehmende* wurde als sinnvoll erachtet. Nach den Fokusgruppeninterviews erfolgte eine erste Überarbeitung, um die inhaltliche Validität des Fragebogens zu verbessern. Es wurden 15 Items angepasst und sechs neue Items hinzugefügt, wobei auf die Erhaltung der Struktur der Handlungsfelder und Handlungsempfehlungen geachtet wurde.

Die **zweite qualitative Validierungsstudie** prüfte das sprachliche Niveau, die Itemschwierigkeit sowie die Verständlichkeit der Items und Antwortmöglichkeiten aus Sicht der Kinder und Jugendlichen. Teilnehmerinnen dieser Pilotierungsstudie waren sieben Mädchen einer Geräteturn-Trainingsgruppe im Alter von zehn bis 13 Jahren. In zwei Gruppen wurde jeweils die Hälfte des Fragebogens (35–36 Items) wie folgt bearbeitet: Die Versuchsleiter\*innen baten die Teilnehmerinnen, je ein Item laut vorzulesen. Dabei wurden bereits erste schwierige Begriffe oder Verständnisprobleme identifiziert. Im Anschluss wurden die Verständlichkeit der Fragen sowie Redundanzen innerhalb des Fragebogens diskutiert und schriftlich von den Versuchsleiter\*innen protokolliert. In den Kleingruppen wurde zweimal 20 Minuten an den Fragebogenitems gearbeitet. Dazwischen organisierte das Versuchsleiter\*innenteam eine zehnmütige Bewegungspause.

Die differenzierten Rückmeldungen der Teilnehmerinnen erlaubten die Identifizierung und anschließende Optimierung schwieriger oder unpräziser Begriffe und Satzkonstruktionen. Beispielsweise wurden Begriffe wie „(Lern-)Aufgabe“ im Rahmen der sportlichen Aktivitäten

missverstanden und in den Items durch den Begriff „Übung“ ersetzt.

Für die inhaltliche Prüfung wurde in einer **dritten qualitativen Studie** das Evaluationsinstrument in neun J+S-Aktivitäten (drei deutsch-, drei französisch- und drei italienischsprachige J+S-Aktivitäten) aus allen drei Perspektiven (Teilnehmende, Leitende, Beobachter\*innen) geprüft. Der Fokus lag auf der Umsetzbarkeit und der Verständlichkeit des J+S-Evaluationsinstruments.

Während der J+S-Aktivität hatten die Beobachter\*innen den Auftrag, das Evaluationsinstrument auszufüllen. Um zu prüfen, ob die Items auch wirklich die intendierten Inhalte erfragen und diese Inhalte auch im Trainingsalltag beobachtbar sind, wurden die Beobachter\*innen gebeten, alle ihre Gedanken während der Kursevaluation laut auszusprechen. Diese Methode des *lauten Denkens* (Think-Aloud) dient dazu, die mentalen Prozesse sichtbar zu machen, welche die Befragten bei der Beantwortung des Fragebogens anwenden. Um den Prozess des *lauten Denkens* zu erleichtern, wird der Einsatz des *gezielten Nachfragens* (Probing) empfohlen. In der vorliegenden Studie verwendeten die Versuchsleitenden entsprechend Probes wie „Was denkst du gerade?“ oder „Auf was schaust du jetzt?“ (Collins 2003).

Am Ende der J+S-Aktivität wurden die Teilnehmenden und die Leitenden gebeten, den J+S-Fragebogen auszufüllen. Während des Ausfüllens wurde ein *Behavior Coding* (Oksenberg et al. 1991) vorgenommen, indem alle Rückfragen, Störungen, Unsicherheiten und das Verhalten der Teilnehmenden schriftlich festgehalten wurden. Nach dem Ausfüllen des Fragebogens wurden den Teilnehmenden im Plenum gezielte Fragen zur Länge und Verständlichkeit des Evaluationsinstruments sowie itemspezifische Fragen gestellt. Nach der Verabschiedung der Teilnehmenden wurden die Leitenden zu ihrer allgemeinen Einschätzung zum Evaluationsinstrument sowie zu einzelnen Items befragt.

Die Resultate des Think-Aloud-Prozesses der Beobachter\*innen wie auch die Rückmeldungen der Leitenden und Teilnehmenden wurden in einer Protokollvorlage handschriftlich festgehalten. Die

se Vorlage war in vorab definierten Kategorien zur Verständlichkeit und Umsetzbarkeit gegliedert. Die Protokolle wurden im Anschluss digitalisiert und kategorienbasiert inhaltsanalytisch ausgewertet (Mayring 2015).

Die Beobachter\*innen schätzten das J+S-Evaluationsinstrument für die Qualitätsbeurteilung von J+S-Aktivitäten als sinnvoll ein. Hingegen empfanden sie den vollständigen Fragebogen, welcher alle vier Handlungsfelder abdeckt, als zu lang. Auch wünschten sich die Beobachter\*innen einen Aufbau des Fragebogens entlang des üblichen Ablaufs der Trainingsaktivität (Aufwärmen, Hauptteil, Abschluss).

Zudem wurden Inhalte identifiziert, die in einer einzelnen Einheit möglicherweise schwer zu beobachten sind, wie beispielsweise die Regelklarheit, die nur bei einem Regelverstoß beobachtbar wäre. Bei den Teilnehmenden ergab das *Behavior Coding* wichtige Erkenntnisse für die Durchführung der Befragung (Klarheit der Instruktion, genügend Abstand zwischen den Teilnehmenden beim Ausfüllen). Die Leitenden beurteilten die Instruktion und die Items als selbsterklärend, wiesen aber auch auf die Redundanzen und Wiederholungen hin. Als Beispiel wurden Fragen zur Freude, wie „Den Teilnehmenden macht das Training Spaß“ und „Die Teilnehmenden sind vom Training begeistert“, als gleichbedeutend betrachtet. Gleiches galt für die Fragen zum Umgang mit Störungen: „Im Training erhalte ich schnell Aufmerksamkeit, wenn ich etwas sage“ und „im Training hören die Teilnehmenden auf mich“. In der Überarbeitung wurde entsprechend nur je eines der beiden Items im Fragebogen behalten. Die Leitenden gaben auch wichtige Hinweise zur Umsetzung, insbesondere wie und wann die Evaluationen in die J+S-Aktivitäten integriert werden können.

In der Summe ergaben die qualitativen Studien, dass der Fragebogen in allen getesteten Aktivitäten eingesetzt werden kann. Die Methode des *lauten Denkens* zeigte, dass die gewählten Qualitätskriterien für die Beobachter\*innen in verschiedenen Sportarten und Settings beobachtbar sind. Für die Leitenden und die Teilnehmenden konnte eine

geplante Bearbeitungszeit des Fragebogens von 15–20 min eingehalten werden, wobei die J+S-Teilnehmenden in dieser Zeit rund 25 Items gewissenhaft beantworteten. Sämtliche sprachlichen und inhaltlichen Rückmeldungen sowie Erkenntnisse bezüglich Durchführbarkeit und Ökonomie der Befragung wurden in die erneute Überarbeitung einbezogen.

## Quantitative Validierungsstudie

Die quantitative Studie prüfte die psychometrischen Qualitätskriterien des J+S-Evaluationsinstruments aus der Perspektive der Teilnehmenden. Dabei wurde die Fragestellung bearbeitet, ob die J+S-Handlungsfelder (z. B. Lernförderliches Klima ermöglichen) und dazugehörigen Handlungsempfehlungen (z. B. Teilnehmende wertschätzen) durch die erfassten Items (z. B. „Unser/e Leiter/in beachtet mich“, [Tab. 7](#)) faktoriell valide und reliabel abgebildet werden. Eines der wichtigsten Kriterien für ein zukünftiges Evaluationssystem zur Messung der Qualität in J+S-Aktivitäten ist die Testökonomie. Unter Einhaltung der psychometrischen Qualitätskriterien sollen daher möglichst wenig Items zur Erfassung eingesetzt werden (Raubenheimer 2004). Entsprechend wurde eine systematische Itemreduktion auf drei Items pro Handlungsempfehlung vorgenommen.

## Methodisches Design und Datenerhebung

Aus der J+S-Datenbank wurden für vier Sportarten zufällig 40 Trainingsgruppen ausgewählt. Damit bestand die Grundgesamtheit aus 160 Trainingsgruppen. Darin vertreten waren Aktivitäten aus der deutsch-, französisch- und italienischsprachigen Schweiz.

Mit Fußball und Volleyball wurden zwei „klassische“ Sportarten für die Validierung gewählt, welche bei gleichbleibender Gruppenzusammensetzung wöchentlich durchgeführt werden und an einem festen Ort stattfinden. Zusätzlich wurden die zwei Sportarten Gerätturnen und Skifahren ausgewählt, die in gewissen Punkten andere Charakteristika aufweisen. Beim Gerätturnen trainieren meist sehr viele Teilnehmende gleichzeitig und mit mehreren Leitenden. Ski-

**Tab. 1** Stichprobenbeschreibung

Stichprobe	N	N-Gruppen	Gruppengröße	Geschlecht	M	SD	95 %-KI
Klima t1	594	71	8,37	47,0% weiblich	13,97 Jahre	2,45	[13,77; 14,18]
Klima t2	410	59	6,95	48,7% weiblich	14,12 Jahre	2,37	[13,88; 14,34]
Gruppe t1	604	69	8,75	59,5% weiblich	14,14 Jahre	2,37	[13,95; 14,33]
Gruppe t2	457	63	7,25	63,1% weiblich	13,83 Jahre	2,16	[13,62; 14,01]
Ziel t1	581	68	8,54	51,9% weiblich	14,21 Jahre	2,35	[14,02; 14,40]
Ziel t2	387	59	6,56	53,6% weiblich	13,97 Jahre	2,06	[13,77; 14,18]
Aufgabe t1	682	75	9,09	55,7% weiblich	13,68 Jahre	2,23	[13,52; 13,85]
Aufgabe t2	513	69	7,44	58,2% weiblich	13,86 Jahre	2,10	[13,68; 14,04]
Gesamt t1	1230	141	8,72	53,7% weiblich	13,99 Jahre	2,36	[13,86; 14,12]
Gesamt t2	851	120	7,08	55,9% weiblich	13,96 Jahre	2,18	[13,80; 14,10]

N Stichprobenumfang, N-Gruppen Anzahl Trainingsgruppen, Gruppengröße durchschnittliche Größe der Trainingsgruppe, M Mittelwert, SD Standardabweichung, KI Konfidenzintervall

fahren unterscheidet sich vor allem im wechselnden Trainingskontext (Indoor, Outdoor, auf dem Schnee, Alternativtraining ohne Schnee) und in der Art der Raumnutzung.

Derselbe J+S-Fragebogen wurde in jeder Trainingsgruppe zu zwei Erhebungszeitpunkten im Abstand von vier bis sechs Wochen eingesetzt. Aufgrund rechtlicher Bestimmungen, dass die Datenerhebung vollständig anonymisiert erfolgen musste, konnten keine individuellen Personendaten erhoben werden, welche es ermöglichen würden, die Daten der beiden Erhebungen auf Individualebene zu verknüpfen. Entsprechend liegen zwei querschnittliche Stichproben vor, die sich in ihrer Zusammensetzung unterscheiden können, da nicht kontrolliert werden konnte, ob zu beiden Erhebungen dieselben Teilnehmenden anwesend waren. Die Daten der ersten Erhebung wurden für explorative Analysen, die Daten der zweiten Erhebung für konfirmatorische Analysen eingesetzt.

Aus testökonomischen Gründen und aufgrund der Ergebnisse der qualitativen Studien wurde den Teilnehmenden nicht der vollständige Fragebogen zur Beantwortung vorgelegt. Pro Trainingsgruppe wurden zu beiden Erhebungen jeweils zwei der vier Handlungsfelder befragt. Die Zuordnung erfolgte randomisiert. Durch dieses Vorgehen entstanden insgesamt sechs verschiedene Kombinationen der abgefragten Handlungsfelder (Klima & Gruppe, Klima & Ziel, Klima & Aufgabe, Gruppe & Ziel, Gruppe & Aufgabe, Ziel & Aufgabe).

Die einleitende Instruktion im Fragebogen lautete: „Wie hast du das Training in den letzten 4 Wochen wahrgenommen?“ bzw. „Wie hast du das Training seit den letzten Ferien wahrgenommen?“. Alle Items wurden auf einer fünfstufigen Likert-Skala beantwortet (0 = trifft nicht zu, 1 = trifft eher nicht zu, 2 = trifft teils zu, 3 = trifft eher zu, 4 = trifft zu), wobei ergänzend ein Antwortfeld „weiß ich nicht“ integriert wurde.

Die Durchführung und Begleitung der Evaluation vor Ort wurde von geschulten Beobachter\*innen vorgenommen, welche im Rahmen der J+S-Ausbildung als J+S-Expert\*innen fungieren und bei der Entwicklung des Evaluationsinstruments beteiligt waren. Die Datenerhebung erfolgte per Tablet oder Handy online am Ende des Trainings. Beim zweiten Erhebungszeitpunkt sollten alle Trainingsgruppen der ersten Erhebung erneut befragt werden, wobei die Datenerhebung identisch ablief.

### Stichprobe

Von den geplanten 160 Aktivitäten konnten zur **ersten Erhebung** Daten von insgesamt 141 J+S-Gruppen mit  $N=1230$  Kindern und Jugendlichen (53,7% weiblich) in einer Altersspanne von zehn bis 20 Jahren ( $M=13,99$ ,  $SD=2,36$ ) erfasst werden. Die Rücklaufquote auf Gruppenebene betrug 88%. Die Datenerhebung erfolgte in den folgenden Anteilen: 29,9% im Fußball, 30,5% im Volleyball, 23,3% im Gerätturnen und 16,3% im Skifahren. Bei der **zweiten Erhebung** wurden von den 141 J+S-Gruppen der ersten Erhebung 120 erneut befragt (Rücklaufquote

85%). Dabei wurden  $N=851$  Kinder und Jugendliche (53,7% weiblich) im Alter von  $M=13,96$  Jahren ( $SD=2,18$ ) erfasst (26,9% Fußball, 29,5% Volleyball, 25,5% Gerätturnen, 18,1% Skifahren).

Die acht Teilstichproben (vier Stichproben der ersten und vier Stichproben der zweiten Erhebung), welche sich aus Handlungsfelder-Kombination ergaben, können **Tab. 1** entnommen werden. Jede Stichprobe bezieht sich somit auf eine der vier Handlungsempfehlungen. Auf Basis der 95%-Konfidenzintervalle zeigte sich, dass sich die Altersverteilungen zwischen den Stichproben nicht bedeutsam unterscheiden.

### Datenauswertung

Die Häufigkeits- und Korrelationsanalysen wurden mit SPSS 28 (IBM Corp. 2021) durchgeführt, die latenten Strukturgleichungsmodelle mit Mplus 8,3 (Muthén und Muthén 2017). Bei allen Strukturgleichungsmodellen wurden potenzielle Abhängigkeiten innerhalb der vorhandenen Mehrebenenstruktur (Teilnehmer\*innen genestet in Trainingsgruppen) berücksichtigt, indem die Standardfehler mit der Mplus-Funktion *type=complex* für geschachtelte Datensätze angepasst wurden.

Der Anteil an fehlenden Werten pro Item lag zwischen 0,5 und 7,0% ( $M=2,60\%$ ,  $SD=1,73\%$ ), wobei auch Werte der Antwortkategorie „Weiß ich nicht“ mitzählten. Fehlende Werte wurden anhand des *Full-Information-Maximum-Likelihood* (FIML)-Algorithmus geschätzt. Die Beurteilung der Anpassungsgüte der verschiedenen Modelle

**Tab. 2** Modellanpassungen der konfirmatorischen Faktoranalysen (CFA)

	Modell	Items	Stichprobe	N	$\chi^2$	df	CFI	TLI	RMSEA
<b>Lernförderliches Klima ermöglichen (KLIMA)</b>									
1	CFA-1	15	Klima-1	594	199,7	84	0,949	0,937	0,048
2	CFA-2	12	Klima-2	410	76,0	48	0,983	0,977	0,038
3	CFA-2 2. Ord.	12	Klima-2	410	77,4	50	0,984	0,978	0,031
<b>Gruppe sicher und effizient führen (GRUPPE)</b>									
1	CFA-1	23	Führung-1	604	488,9	215	0,894	0,876	0,046
2	CFA-2	18	Führung-2	457	218,3	120	0,938	0,921	0,042
3	CFA-2 2. Ord.	18	Führung-2	457	233,4	129	0,934	0,922	0,042
<b>Lernziele priorisieren und Struktur schaffen (ZIELE)</b>									
1	CFA-1	12	Ziel-1	581	201,4	51	0,917	0,893	0,071
2	CFA-2	9	Ziel-2	387	33,3	24	0,992	0,989	0,026
3	CFA-2 2. Ord.	9	Ziel-2	387	33,3	24	0,992	0,989	0,032
<b>Attraktive Aufgaben stellen und Rückmeldungen geben (AUFGABE)</b>									
1	CFA-1	14	Aufgabe-1	682	201,8	71	0,942	0,925	0,052
2	CFA-2	12	Aufgabe-2	513	91,8	48	0,973	0,962	0,042
3	CFA-2 2. Ord.	12	Aufgabe-2	513	95,3	50	0,972	0,963	0,049

Berichtet sind Modellfits. *N* Stichprobenumfang,  $\chi^2$  chi-quadrat Wert, *df* Anzahl der Freiheitsgrade, *CFI* comparative fit index, *TLI* Tucker-Lewis index, *RMSEA* Root Mean Square Error of Approximation

**Tab. 3** Korrelationen der J + S-Handlungsempfehlungen zum „Lernförderlichen Klima ermöglichen“

Lernförderliches Klima ermöglichen (KLIMA)		WERT	EINBE	FEHL	FREU	KLIMA FW 2. Ord.
1	Teilnehmende wertschätzen (WERT)	1	0,552*	0,771*	0,710*	<b>0,847*</b>
2	Teilnehmende einbeziehen (EINBE)	0,385*	1	0,587*	0,595*	<b>0,669*</b>
3	Fehler nutzen (FEHL)	0,618*	0,396*	1	0,754*	<b>0,897*</b>
4	Freude wecken (FREU)	0,440*	0,354*	0,557*	1	<b>0,847*</b>

Letzte Spalte: *FW 2. Ord.* standardisierte Faktorwerte zweiter Ordnung

\*  $p < 0,01$ ; oberhalb der Diagonalen: latente Korrelationen; unterhalb der Diagonalen: Pearson-Korrelationen

erfolgte auf Basis der in der Literatur vorgeschlagenen Fit-Indices (Weiber und Mühlhaus 2014).

Zur konfirmatorischen Prüfung der faktoriellen Validität wurden für jedes der vier Handlungsfelder konfirmatorische Faktoranalysen (CFA) in drei Analyseschritten durchgeführt:

1. **CFA mit allen Items (CFA-1):** Im ersten Schritt wurden die Stichproben des ersten Messzeitpunkts analysiert. Für jedes der vier J + S-Handlungsfelder wurden separate CFAs berechnet. Die latenten Faktoren bildeten die J + S-Handlungsempfehlungen ab, welche konzeptionell die Zuordnung der Items zu einem Faktor bestimmten (Anhang [Tab. 8](#), [9](#), [10](#) und [11](#)). Die Faktorladungen der Items auf die latenten Faktoren wurden frei

geschätzt, Nebenladungen wurden nicht zugelassen.

2. **CFA mit drei Items pro Merkmal (CFA-2):** Im zweiten Schritt wurde die Itemzahl pro latentem Faktor auf drei Items reduziert. Hierfür wurden die Items mit den geringsten Faktorladungen der CFA-1 inhaltlich geprüft und folglich eine Auswahlentscheidung zur Itemreduzierung getroffen. In einer CFA mit den Stichproben der zweiten Erhebung wurde das Modell mit reduzierter Itemzahl erneut geprüft und die Interkorrelationen der latenten Faktoren berechnet.
3. **CFA mit drei Items pro Merkmal und latenten Faktoren zweiter Ordnung:** Im dritten Schritt wurde geprüft, ob die latenten Faktoren erster Ordnung (im Sinne der Handlungsempfehlungen) die übergeordneten J + S-

Handlungsfelder abbilden. Dazu wurden Faktoren zweiter Ordnung in die CFA-2 einbezogen.

Zur Überprüfung der Reliabilität der erfassten Handlungsempfehlungen wurden auf Basis der Stichproben der zweiten Erhebung die Cronbach's  $\alpha$ -Werte (mit gekürzter Itemzahl) berechnet. Idealerweise sollten die  $\alpha$ -Werte über 0,70 liegen, wobei sie abhängig von der Itemanzahl sind (Weiber und Mühlhaus 2014). Die Unterrichtsqualität in einer Klasse oder Gruppe gilt als ein Maß für eine geteilte Umwelt, die in Beziehung zu anderen Konstrukten (z. B. Verhalten der Lehrenden) steht. Hierfür wurden individuelle Einschätzungen der Teilnehmenden gruppenweise gemittelt. Die Unterschiede in den Einschätzungen der Teilnehmenden zwischen den Gruppen werden durch den ICC charakterisiert, welcher ebenfalls berechnet wurde (Lüdtke et al. 2006). Abschließend wurden auf manifester Ebene Mittelwerte (*M*), Standardabweichung (*SD*) sowie Pearson-Korrelationen der Handlungsempfehlungen innerhalb eines Handlungsfelds ermittelt.

## Ergebnisse

Nachfolgend werden die Ergebnisse der quantitativen Validierungsstudie dargestellt. [Tab. 2](#) liefert einen Überblick über die Modellanpassung der durchgeführten konfirmatorischen Faktorenanalysen. Anschließend werden die Ergebnisse für die vier Handlungsfelder separat dargestellt.

### Faktorielle Validität „Lernförderliches Klima ermöglichen“

Das Handlungsfeld „Lernförderliches Klima ermöglichen“ wurde mittels vier Faktoren und 15 Items erfasst. Die CFA-1 zeigte eine gute Anpassung der vierfaktoriellen Struktur auf die vorliegenden Daten ([Tab. 2](#)). Die resultierenden Faktorwerte lagen bei  $0,655 \leq \beta \leq 0,819$  (Anhang [Tab. 8](#)). Bei den Faktoren mit vier Items wurde das Item mit der geringsten Faktorladung gelöscht. Die verbleibenden zwölf Items wurden mit der CFA-2 erneut geprüft, wobei die

**Tab. 4** Korrelationen der J + S-Handlungsempfehlungen zu „Gruppe sicher und effizient führen“

Gruppe sicher und effizient führen (GRUPPE)	FUEHR	ALLG	ZEIT	STOER	REGEL	SICH	GRUPPE FW 2. Ord.
1 Angepasst führen (FUEHR)	1	0,682*	0,720*	0,531*	0,509*	0,625*	0,768*
2 Allgegenwärtigkeit (ALLG)	0,415*	1	0,838*	0,656*	0,645*	0,642*	0,887*
3 Hohe Lern- oder Bewegungszeit ermöglichen (ZEIT)	0,493*	0,476*	1	0,640*	0,605*	0,745*	0,922*
4 Angepasst mit Störungen umgehen (STOER)	0,297*	0,397*	0,377*	1	0,411	0,504*	0,678*
5 Regelklarheit (REGEL)	0,369*	0,491*	0,429*	0,401*	1	0,705*	0,713*
6 Sicherheit gewährleisten (SICH)	0,410*	0,432*	0,411*	0,311*	0,519*	1	0,806*

Letzte Spalte: FW 2. Ord. standardisierte Faktorwerte zweiter Ordnung  
 \*  $p < 0,01$ ; oberhalb der Diagonalen: latente Korrelationen; unterhalb der Diagonalen: Pearson-Korrelationen

**Tab. 5** Korrelationen der J + S-Handlungsempfehlungen zu „Lernziele priorisieren und Struktur schaffen“

Lernziele priorisieren und Struktur schaffen (ZIELE)	KOMM	STRU	LEFO	ZIELE FW 2. Ord.
1 Ziele priorisieren und transparent kommunizieren (KOMM)	1	0,737*	0,744*	0,834*
2 Inhalte sinnvoll strukturieren (STRU)	0,611*	1	0,789*	0,884*
3 Ziele überprüfen und Lernfortschritte aufzeigen (LEFO)	0,645*	0,669*	1	0,892*

Letzte Spalte: FW 2. Ord. standardisierte Faktorwerte zweiter Ordnung  
 \*  $p < 0,01$ ; oberhalb der Diagonalen: latente Korrelationen; unterhalb der Diagonalen: Pearson-Korrelationen

Modellanpassung wiederum gute Werte erreichte (Tab. 2) mit Faktorwerten von  $0,705 \leq \beta \leq 0,900$  (Anhang Tab. 8).

Die interne Konsistenz der vier Handlungsempfehlungen war mit  $0,835 \leq \alpha \leq 0,868$  sehr gut. Auf Basis der Modellanpassung und gegebenen Reliabilität konnte die faktorielle Validität angenommen werden, sodass die Bildung von (manifesten) Faktormittelwerten zulässig war. Die Mittelwerte lagen zwischen  $2,72 \leq M \leq 3,56$ . Die mittleren ICC-Werte zeigten auf, dass 9,6–23,7% der Varianz in den Faktoren durch die Zugehörigkeit zu einer Trainingsgruppe erklärt werden kann (Tab. 7). Entsprechend gibt es bedeutsame Unterschiede in der Wahrnehmung des Klimas zwischen den Gruppen.

Die Interkorrelationen der Handlungsempfehlungen zeigten mittlere manifeste und hohe latente Zusammenhänge (Tab. 3). Entsprechend konnte gezeigt werden, dass die Faktoren zwar verbunden sind, jedoch ausreichend trennscharf abgebildet werden können.

Die vier Faktoren, welche aus den theoretischen Überlegungen heraus dem Handlungsfeld „Lernförderliches Klima ermöglichen“ zugeordnet wurden, bildeten ebenfalls mit hohen Faktorwerten einen latenten Faktor zweiter Ordnung ab (Tab. 3), womit die Zuordnung der Handlungsempfehlungen zum übergeordneten Handlungsbereich empirisch bestätigt werden konnte.

**Faktorielle Validität „Gruppe sicher und effizient führen“**

Das Handlungsfeld „Gruppe sicher und effizient führen“ wurde mittels sechs Faktoren und 23 Items erfasst. Die CFA-1 (Tab. 2) zeigte eine ausreichende Passung der sechsfaktoriellen Struktur auf die vorliegenden Daten. Niedrige CFI-/TLI-Werte wiesen auf unpassende Itemzuweisungen und potenzielle Nebenladungen hin. Entsprechend lagen teils Faktorwerte ( $0,373 \leq \beta \leq 0,782$ ) mit niedrigen Werten vor. Basierend auf den Faktorladungen wurden alle

Handlungsempfehlungen auf drei Items gekürzt. Ausnahme bildeten die Items zur Handlungsempfehlung „Sicherheit gewährleisten“. Hier zeichneten sich drei Items (sich1–sich3) ab, welche eher eine wahrgenommene Sicherheit abbildeten sowie drei Items (sich4–sich6), welche auf Sicherheitsregeln abzielen. Auf Basis der Faktorwerte und inhaltlichen Überlegungen wurden die drei Items zu den Sicherheitsregeln im Modell belassen (Anhang Tab. 9).

Die verbleibenden 18 Items wurden in der CFA-2 erneut geprüft. Die Modellanpassung erreichte nun zufriedenstellende Werte (Tab. 2) mit mittleren bis hohen Faktorwerten (Anhang Tab. 9). Die interne Konsistenz der sechs Handlungsempfehlungen fiel zufriedenstellend aus. Zwar lagen die Faktoren „Allgegenwärtigkeit“ und „Lern- und Bewegungszeit“ unter dem Richtwert von  $\alpha = 0,700$ , was aber für Skalen mit nur drei Items als noch akzeptabel gewertet werden kann. Insgesamt konnte auch für die Handlungsempfehlungen dieses Handlungsfelds die faktorielle Validität angenommen werden. Die Faktormittelwerte betragen  $3,33 \leq M \leq 3,65$ . Die niedrigen bis mittleren ICC-Werte zeigten, dass 5,3–15,4% der Varianz in den Faktoren durch die Zugehörigkeit zu einer Trainingsgruppe erklärt werden kann (Tab. 7).

Die Interkorrelationen der Handlungsempfehlungen zeigten mit mittleren manifesten und mittleren bis hohen latenten Zusammenhängen (Tab. 4), dass die Faktoren ausreichend trennscharf sind. Weiterhin bildeten die sechs Faktoren mit hohen Faktorwerten einen latenten Faktor zweiter Ordnung ab, welcher das übergeordnete Handlungsfeld „Gruppe sicher und effizient führen“ abbildet.

**Faktorielle Validität „Lernziele priorisieren und Struktur schaffen“**

„Lernziele priorisieren und Struktur schaffen“ wurde mittels drei Faktoren und zwölf Items erfasst. Die CFA-1 (Tab. 2) zeigte eine nicht zufriedenstellende Modellanpassung mit drei Faktoren, wenngleich die Faktorwerte mit  $0,575 \leq \beta \leq 0,872$  mittel bis hoch und



**Tab. 6** Korrelationen der J + S-Handlungsempfehlungen zu „Attraktive Aufgaben stellen und Rückmeldungen geben“

Attraktive Aufgaben stellen und Rückmeldungen geben (AUFGABE)	RUCK	VIEL	HERA	ERFO	AUFGABE FW 2. Ord.
1 Zeitnahe und konstruktive Rückmeldungen geben (RUCK)	1	0,402*	0,698*	0,688*	0,874*
2 Vielseitige Lernarrangements planen und umsetzen (VIEL)	0,369*	1	0,445*	0,356*	0,486*
3 Herausfordernde Aufgaben stellen (HERA)	0,540*	0,384*	1	0,608*	0,806*
4 Erfolgserlebnisse ermöglichen (ERFO)	0,575*	0,339*	0,500*	1	0,769*

Letzte Spalte: *FW 2. Ord.* standardisierte Faktorwerte zweiter Ordnung

\*  $p < 0,01$ ; oberhalb der Diagonalen: latente Korrelationen; unterhalb der Diagonalen: Pearson-Korrelationen

weitestgehend homogen ausfielen (Anhang [Tab. 10](#)). Zur Reduzierung auf drei Items pro Handlungsempfehlung wurden insgesamt drei Items gelöscht.

Die verbleibenden neun Items erreichten in der CFA-2 ([Tab. 2](#)) eine gute Modellanpassung mit durchgängig hohen Faktorwerten (Anhang [Tab. 10](#)). Die interne Konsistenz fiel bei allen drei Handlungsempfehlungen hoch aus. Entsprechend konnte die faktorielle Validität angenommen werden. Die Faktormittelwerte lagen zwischen  $3,13 \leq M \leq 3,51$ , die ICC-Werte bei  $13,5\text{--}21,2\%$ , bedeutend, dass die Ausprägung dieser Handlungsempfehlungen stärker von der Zugehörigkeit zu einer bestimmten Trainingsgruppe erklärt wird ([Tab. 7](#)).

Die drei Faktoren zeigten hohe manifeste und latente Interkorrelationen und konnten mit hohen und homogenen Faktorwerten den latenten Faktor zweiter Ordnung des übergeordneten Handlungsfeldes abbilden ([Tab. 5](#)).

### Faktorielle Validität „Attraktive Aufgaben stellen und Rückmeldungen geben“

Das Handlungsfeld „Attraktive Aufgaben stellen und Rückmeldungen geben“ wurde mittels vier Faktoren und 14 Items erfasst. Die Modellanpassung der CFA-1 ([Tab. 2](#)) fiel mit Faktorwerten von  $0,619 \leq \beta \leq 0,867$  gut aus (Anhang [Tab. 11](#)). Da die Faktorwerte keine klaren Hinweise zur Itemreduktion boten, wurden die zwei Items gelöscht, welche sprachlich schwieriger verständlich erschienen. Die Modellanpassung der verbleibenden zwölf Items erreich-

te in der CFA-2 ([Tab. 2](#)) sehr gute Werte. Die internen Konsistenzen fielen ebenfalls gut aus, so dass die faktorielle Validität angenommen werden konnte. Die Faktormittelwerte lagen zwischen  $2,85 \leq M \leq 3,62$ , die ICC-Werte zwischen  $1,9$  und  $27,2\%$  ([Tab. 7](#)).

Die vier Faktoren zeigten mittlere manifeste und latente Interkorrelationen, wobei die Korrelationen mit „Vielseitige Lernarrangements planen und umsetzen“ etwas geringer ausfielen als die Korrelationen zwischen den weiteren drei Handlungsempfehlungen. Auch der Faktorwert zur Bildung eines latenten Faktors zweiter Ordnung fiel mit  $\beta \leq 0,486$  tiefer aus als die anderen drei Faktorwerte ( $0,769 \leq \beta \leq 0,874$ ) ([Tab. 6](#)). Dies resultierte in einer geringeren Anpassung der CFA zweiter Ordnung, wenngleich die Werte immer noch gut ausfielen ([Tab. 2](#)).

### Zusammenfassung, Diskussion und Ausblick

Der vorliegende Beitrag stellt die Entwicklung und Validierung des J+S-Evaluationsinstruments zur Erfassung der Lehr- und Lernqualität bei J+S-Aktivitäten dar. Auf theoretisch-konzeptioneller Ebene zeigten sich große Überschneidungen zwischen den von J+S definierten Handlungsfeldern und -empfehlungen und den in der Unterrichtsqualitätsforschung definierten Unterrichtsdimensionen und -merkmalen. Folglich wurde basierend auf dem bereits validierten QUALLIS-Evaluationsinstrument für den Sportunterricht (Herrmann 2019; Herrmann et al. 2023a,

2015) das J+S-Evaluationsinstrument für J+S-Aktivitäten (weiter-)entwickelt.

Die qualitativen Pilotstudien ergaben, dass die J+S-Handlungsfelder und -empfehlungen inhaltlich valide operationalisiert werden konnten. Durch systematische Item-Weiterentwicklung entstand ein verständlicher Fragebogen, der von Kindern, Jugendlichen und J+S-Leitenden verstanden, akzeptiert und als nützlich anerkannt wurde.

Die umfangreiche quantitative Validierungsstudie bestätigte die psychometrischen Gütekriterien des J+S-Evaluationsinstruments. Alle vier Handlungsbereiche mit ihren 17 Handlungsempfehlungen konnten mittels 51 Items valide und reliabel abgebildet werden, wodurch das Evaluationsinstrument aus Perspektive der an J+S-Aktivitäten teilnehmenden Kinder und Jugendlichen im Alter zwischen zehn und 20 Jahren empfohlen werden kann.

Eine Stärke des J+S-Evaluationsinstruments liegt auch darin, dass bereits eine deutsch-, französisch- und italienischsprachige Version vorliegt (vgl. online verfügbare Itemdokumentation Herrmann et al. 2023b). Zukünftig sollte mit einer größeren Stichprobe die Konstruktäquivalenz der drei Sprachversionen geprüft werden, um zu klären, ob die faktorielle Struktur über die Sprachen hinweg gleich (invariant) ist und damit das gleiche Konstrukt misst.

Einschränkungen bestehen in der Generalisierbarkeit hinsichtlich Alter und Perspektive. So muss an dieser Stelle offen bleiben, ob die Nutzung des Evaluationsinstruments bei Kindern unter zehn Jahren zu validen Ergebnissen führt. In der Trainingspraxis muss auch geprüft werden, ob und unter welchen Bedingungen die Kinder und Jugendlichen den gesamten Fragebogen mit 51 Items ausfüllen können, oder ob – wie in der Validierungsstudie – eine weitere Auswahl an J+S-Handlungsfeldern und Items in Abhängigkeit von der Erhebungssituation getroffen werden muss. Weiterhin wird zu prüfen sein, ob die mittels des J+S-Evaluationsinstruments vorgenommenen Einschätzungen der J+S-Leitenden und externer Beobachter\*innen ebenfalls zu reliablen und validen Ergebnissen führen.

**Tab. 7** J+ S-Handlungsempfehlungen mit Beispielitems und psychometrischen Kennwerten

		<b>Beispielitem</b>	<b>FW</b>	<b>C's α</b>	<b>M</b>	<b>SD</b>	<b>ICC</b>
<b>KLIMA</b>	<b>Lernförderliches Klima ermöglichen</b>						
WERT	Teilnehmende wertschätzen	„Unser/e Leiter/in beachtet mich.“	$0,750 \leq \beta \leq 0,859$	0,835	3,43	0,85	0,096
EINBE	Teilnehmende einbeziehen	„Unser/e Leiter/in geht auf unsere Vorschläge zum Inhalt des Trainings ein.“	$0,793 \leq \beta \leq 0,862$	0,866	2,72	1,14	0,237
FEHL	Fehler nutzen	„Unser/e Leiter/in unterstützt mich, wenn ich Fehler mache.“	$0,705 \leq \beta \leq 0,868$	0,843	3,55	0,74	0,123
FREU	Freude wecken	„Ich freue mich auf das Training.“	$0,790 \leq \beta \leq 0,900$	0,868	3,56	0,74	0,124
<b>GRUPPE</b>	<b>Gruppe sicher und effizient führen</b>						
FUEHR	Angepasst führen	„Im Training gibt der/die Leiter/in klare Anweisungen.“	$0,676 \leq \beta \leq 0,763$	0,778	3,57	0,61	0,053
ALLG	Allgegenwärtigkeit	„Unser/e Leiter/in bekommt mit, was in der Gruppe passiert.“	$0,603 \leq \beta \leq 0,729$	0,655	3,33	0,71	0,103
ZEIT	Hohe Lern- oder Bewegungszeit	„Im Training sind wir jederzeit beschäftigt.“	$0,520 \leq \beta \leq 0,724$	0,638	3,46	0,66	0,064
STOER	Angepasst mit Störungen umgehen	„Im Training hören wir auf die/den Leiter/in.“	$0,680 \leq \beta \leq 0,771$	0,753	3,38	0,68	0,154
REGEL	Regelklarheit	„Im Training ist klar, was man machen darf und was nicht.“	$0,669 \leq \beta \leq 0,860$	0,728	3,49	0,78	0,104
SICH	Sicherheit gewährleisten	„Im Training achtet unser/e Leiter/in auf die Einhaltung der Sicherheitsregeln.“	$0,614 \leq \beta \leq 0,783$	0,731	3,65	0,62	0,075
<b>ZIELE</b>	<b>Lernziele priorisieren und Struktur schaffen</b>						
KOMM	Ziele priorisieren und transparent kommunizieren	„Unser/e Leiter/in erklärt uns die Ziele des Trainings.“	$0,808 \leq \beta \leq 0,876$	0,885	3,22	1,00	0,212
STRU	Inhalte sinnvoll strukturieren	„Unser/e Leiter/in gibt uns hilfreiche Übungen, die mich weiterbringen.“	$0,687 \leq \beta \leq 0,820$	0,794	3,51	0,72	0,135
LEFO	Ziele überprüfen und Lernfortschritte aufzeigen	„Unser/e Leiter/in sagt uns, welche Fortschritte wir gemacht haben.“	$0,816 \leq \beta \leq 0,853$	0,876	3,13	0,97	0,145
<b>AUFGABE</b>	<b>Attraktive Aufgaben stellen und Rückmeldungen geben</b>						
RUCK	Zeitnahe und konstruktive Rückmeldungen geben	„Unser/e Leiter/in zeigt mir, wie ich die Übungen besser machen kann.“	$0,677 \leq \beta \leq 0,782$	0,752	3,62	0,61	0,019
VIEL	Vielseitige Lernarrangements planen und umsetzen	„Unser/e Leiter/in ... bietet besseren Teilnehmende anspruchsvollere Übungen an.“	$0,682 \leq \beta \leq 0,924$	0,877	2,85	1,20	0,272
HERA	Herausfordernde Aufgaben stellen	„Unser/e Leiter/in lässt uns Übungen machen, bei denen ich mich richtig anstrengen muss.“	$0,692 \leq \beta \leq 0,750$	0,765	3,48	0,69	0,021
ERFO	Erfolgserlebnisse ermöglichen	„Unser/e Leiter/in lobt mich, wenn ich meine Leistung verbessere.“	$0,817 \leq \beta \leq 0,889$	0,874	3,58	0,71	0,057

Kodierung: (0) trifft nicht zu; (1) trifft eher nicht zu; (2) trifft teils teils zu; (3) trifft eher zu; (4) trifft zu  
 FW Faktorwerte, C's Cronbach's alpha, M Mittelwert, SD Standardabweichung, ICC Intraklassenkorrelation

Aus einer wissenschaftlichen Perspektive wäre ein Vergleich der theoretischen Fundierung wie auch der empirischen Ergebnisse zwischen dem niederschwellig einsetzbaren J+ S-Evaluationsinstrument und dem hoch inferenten CLASS-Beobachtungstool (Classroom Assessment Scoring System; Pianta et al. 2008) interessant. So konnte das CLASS-Beobachtungstool für den Leistungssport (Richartz und Anders 2016) wie auch für den Breitensport (Kohake et al. 2022) adaptiert und validiert sowie in der Trainer\*innenausbildung gewinnbringend eingesetzt werden (Richartz

et al. 2021a). Dieser Vergleich könnte weitere Hinweise zur Kriteriumsvalidität der beiden Evaluationsinstrumente sowie zu deren Anwendung und Weiterentwicklung in der Trainingspraxis bieten.

Mit dem vorliegenden Evaluationsinstrument wird es nun möglich sein, die Qualität von Sportangeboten für Kinder und Jugendliche zu beurteilen. Es erscheint unverzichtbar, dass der schulische und außerschulische Sport an der Qualität seiner Maßnahmen arbeitet und darüber Rechenschaft ablegt. Da in dieser Altersgruppe die Weichen für lebenslan-

ges Sporttreiben gelegt werden, obliegt es dem organisierten Kinder- und Jugendsport, qualitätsvolle Sportaktivitäten anzubieten und sicherzustellen. Das entwickelte J+ S-Evaluationsinstrument kann einen Beitrag dazu leisten, die Angebote gezielt zu optimieren.

In der Schweiz wird das J+ S-Evaluationsinstrument breit implementiert werden. Das Ziel ist, jährlich ca. 1200 Evaluationen von J+ S-Aktivitäten durchzuführen, was etwa 10 % aller J+ S-Aktivitäten in den nächsten fünf Jahren entspricht. Dabei liegt der Fokus darauf, die Qualität der J+ S-Leitenden in ih-

ren J + S-Aktivitäten zu verbessern und die J + S-Ausbildung bedarfsgerecht weiterzuentwickeln. Hierfür sind folgende Maßnahmen vorgesehen: Die Evaluationsdaten werden auf Kursebene unmittelbar nach dem Ausfüllen grafisch aufbereitet. Auf dieser Basis sollen externe Beobachter\*innen Feedbackgespräche mit den Leitenden führen, um mögliche Verbesserungen aufzuzeigen. Zudem erfolgt mindestens einmal jährlich eine Auswertung aggregierter Daten, um Handlungsbereiche zur generellen Optimierung der J + S-Ausbildungsangeboten zu identifizieren. Die Daten dienen als Grundlage für neue Inhalte der J + S-Aus- und Weiterbildung.

Die breite Einführung des J + S-Evaluationsinstruments in der Schweiz könnte international als Modell für die systematische und breit angelegte Evaluation ähnlicher Jugend- und Sportprogramme dienen. Dabei könnte vergleichend herausgearbeitet werden, wie das Instrument in unterschiedlichen Sport- und Bildungssystemen adaptiert werden könnte. Dies würde eine Diskussion über Trainingsqualität und deren Evaluation in vielfältigen kulturellen und sprachlichen Kontexten eröffnen. Durch die Verfügbarkeit des Evaluationsinstruments in deutscher, italienischer und französischer Sprache wird die Anpassung an verschiedene Länder erleichtert. Dies könnte zu einem internationalen Austausch bewährter Praktiken führen, der die Qualität des organisierten Kinder- und Jugendsports verbessern könnte.

### Korrespondenzadresse

**Prof. Dr. C. Herrmann**  
 Forschungsgruppe Didaktik Bewegung und Sport, Pädagogische Hochschule Zürich  
 Lagerstrasse 2, 8090 Zürich, Schweiz  
 christian.herrmann@phzh.ch

**Funding.** Open access funding provided by Zurich University of Teacher Education

## Einhaltung ethischer Richtlinien

**Interessenkonflikt.** C. Herrmann, S. Seiler, A. Siffert, L.C. Dapp, V. Gashaj und F. Studer geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autor/-innen keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

## Anhang

**Tab. 8** J + S-Handlungsempfehlungen und Items zum „Lernförderlichen Klima ermöglichen“

KLIMA	Lernförderliches Klima ermöglichen	CFA-1	CFA-2
<i>WERT</i>	<i>Teilnehmende wertschätzen</i> Unser/e Leiter/in ...		
wert1	Beachtet mich	0,726	0,806
wert2	Nimmt sich ausreichend Zeit für mich	0,739	0,859
wert3	Hört mir zu	0,655	–
wert4	Behandelt alle Teilnehmende gleich	0,686	0,750
<i>EINBE</i>	<i>Teilnehmende einbeziehen</i> Unser/e Leiter/in ...		
einbe1	Lässt uns die Übungen mitgestalten	0,787	0,793
einbe2	Geht auf unsere Vorschläge zum Inhalt des Trainings ein	0,758	0,862
einbe3	Gibt uns Möglichkeiten, bei der Gestaltung des Trainings mitzuwirken	0,720	0,802
einbe4	Bietet uns bei den Übungen Auswahlmöglichkeiten an	0,685	–
<i>FEHL</i>	<i>Fehler nutzen</i> Unser/e Leiter/in ...		
fehl1	Macht mir Mut, Übungen noch besser auszuführen	0,662	0,705
fehl2	Unterstützt mich, wenn ich Fehler mache	0,791	0,848
fehl3	Bietet mir seine Unterstützung an	0,788	0,868
<i>FREU</i>	<i>Freude wecken</i>		
freu1	Ich bin mit dem Ablauf des Trainings zufrieden	0,819	0,900
freu2	Ich finde die Gestaltung des Trainings durch unsere/n Leiter/in gut	0,775	0,815
freu3	Ich freue mich auf das Training	0,675	–
freu4	Mir macht das Training Spass	0,744	0,790

Kodierung: (0) trifft nicht zu; (1) trifft eher nicht zu; (2) trifft teils teils zu; (3) trifft eher zu; (4) trifft zu  
 Berichtet werden die standardisierten Faktorwerte. Striche zeigen die ausgeschlossenen Items

**Tab. 9** J+S-Handlungsempfehlungen und Items zu „Gruppe sicher und effizient führen“

GRUPPE	Gruppe sicher und effizient führen	CFA-1	CFA-2
<i>FUEHR</i>	<i>Angepasst führen</i> Im Training ...		
fuehr1	Verstehe ich die Erklärungen unserer Leiterin/unsere Leiters	0,667	0,676
fuehr2	Kann ich die Rückmeldungen unserer Leiterin/unsere Leiters nachvollziehen	0,720	0,763
fuehr3	Gibt der/die Leiter/in klare Anweisungen	0,735	0,741
<i>ALLG</i>	<i>Allgegenwärtigkeit</i> Unser/e Leiter/in ...		
allg1	Merkt sofort, wenn jemand in der Gruppe beginnt, etwas anderes zu tun	0,619	0,633
allg2	Bekommt mit, was in der Gruppe passiert	0,625	0,603
allg3	Sorgt dafür, dass wir das ganze Training konzentriert bleiben	0,656	0,729
<i>ZEIT</i>	<i>Hohe Lern- oder Bewegungszeit</i> Im Training ...		
zeit1	Starten wir pünktlich	0,478	–
zeit2	Sind wir jederzeit beschäftigt	0,652	0,724
zeit3	Nutzen wir die Zeit voll aus	0,753	0,707
zeit4	Sind die Erklärungen kurz	0,492	0,520
<i>STOER</i>	<i>Angepasst mit Störungen umgehen</i> Im Training ...		
stoer1	Verursachen wir wenige Störungen	0,648	0,680
stoer2	Hören wir auf die/den Leiter/in	0,762	0,771
stoer3	Verhalten wir uns angemessen	0,782	0,757
stoer4	Reagiert unser/e Leiter/in auf Störungen	0,448	–
<i>REGEL</i>	<i>Regelklarheit</i> Im Training ...		
regel1	Sind uns die Verhaltensregeln bekannt, die wir einhalten müssen	0,717	0,703
regel2	Hat unser/e Leiter/in die Konsequenzen erklärt, wenn man Verhaltensregeln verletzt	0,574	0,669
regel3	Ist klar, was man machen darf und was nicht	0,756	0,860
<i>SICH</i>	<i>Sicherheit gewährleisten</i> Im Training ...		
sich1	Fühle ich mich sicher	0,596	–
sich2	Habe ich keine Angst mich zu verletzen	0,373	–
sich3	Habe ich keine Angst mich zu blamieren	0,379	–
sich4	Achtet unser/e Leiter/in auf die Einhaltung der Sicherheitsregeln	0,682	0,783
sich5	Weist uns der/die Leiter/in darauf hin, mit Material richtig umzugehen	0,548	0,748
sich6	Zeigt uns unser/e Leiter/in, wie man die Sportart sicher betreibt (z. B. Aufwärmen, Schmuck ablegen)	0,556	0,614

Kodierung: (0) trifft nicht zu; (1) trifft eher nicht zu; (2) trifft teils teils zu; (3) trifft eher zu; (4) trifft zu  
Berichtet werden die standardisierten Faktorwerte. Striche zeigen die ausgeschlossenen Items

**Tab. 10** J + S-Handlungsempfehlungen und Items zu „Lernziele priorisieren und Struktur schaffen“

ZIELE	Lernziele priorisieren und Struktur schaffen	CFA-1	CFA-2
<i>KOMM</i>	<i>Ziele priorisieren und transparent kommunizieren</i> Unser/e Leiter/in ...		
komm1	Sagt uns am Anfang des Trainings, was geübt werden soll	0,776	0,808
komm2	Sagt uns, was wir im Training lernen sollen	0,872	0,876
komm3	Erklärt uns die Ziele des Trainings	0,835	0,871
<i>STRU</i>	<i>Inhalte sinnvoll strukturieren</i> Unser/e Leiter/in ...		
stru1	Baut die Übungen Schritt für Schritt auf	0,525	–
stru2	Zeigt uns die wichtigen Punkte der Übungen	0,691	0,741
stru3	Erklärt uns die Ziele der Übungen	0,794	0,820
stru4	Gibt uns hilfreiche Übungen, die mich weiterbringen	0,575	0,687
<i>LEFO</i>	<i>Ziele überprüfen und Lernfortschritte aufzeigen</i> Unser/e Leiter/in ...		
lefo1	Sagt uns, welche Fortschritte wir gemacht haben	0,785	0,816
lefo2	Bespricht mit uns Ziele	0,761	0,850
lefo3	Überprüft mit uns, ob wir die gesetzten Ziele erreicht haben	0,794	0,853
lefo4	Sagt mir, ob ich mich verbessert habe	0,676	–
lefo5	Macht mit mir persönliche Ziele aus	0,634	–

Kodierung: (0) trifft nicht zu; (1) trifft eher nicht zu; (2) trifft teils teils zu; (3) trifft eher zu; (4) trifft zu  
Berichtet werden die standardisierten Faktorwerte. Striche zeigen die ausgeschlossenen Items

**Tab. 11** J + S-Handlungsempfehlungen und Items zu „Attraktive Aufgaben stellen und Rückmeldungen geben“

AUFGABE	Attraktive Aufgaben stellen und Rückmeldungen geben	CFA-1	CFA-2
<i>RUCK</i>	<i>Zeitnahe und konstruktive Rückmeldungen geben</i> Unser/e Leiter/in ...		
ruck1	Gibt mir Rückmeldungen, wie ich die Übungen richtig ausführen kann	0,704	0,747
ruck2	Zeigt mir, wie ich die Übungen besser machen kann	0,678	0,782
ruck3	Gibt mir nach Übungen zeitnah eine Rückmeldung	0,703	0,677
ruck4	Gibt mir hilfreiche Rückmeldungen	0,676	–
<i>VIEL</i>	<i>Vielseitige Lernarrangements planen und umsetzen</i> Unser/e Leiter/in ...		
viel1	Bietet uns, je nach unserem Können, unterschiedliche Übungen an	0,718	–
viel2	Bietet besseren Teilnehmende anspruchsvollere Übungen an	0,867	0,923
viel3	Bietet schwächeren Teilnehmenden einfachere Übungen an	0,850	0,924
viel4	Bietet allen Teilnehmenden nach ihrem Können Übungen an	0,718	0,682
<i>HERA</i>	<i>Herausfordernde Aufgaben stellen</i> Unser/e Leiter/in lässt uns Übungen machen, ...		
hera1	Bei denen ich mein ganzes Können zeigen kann	0,619	0,692
hera2	Bei denen ich mich richtig anstrengen muss	0,634	0,750
hera3	Die mich herausfordern	0,715	0,739
<i>ERFO</i>	<i>Erfolgslebnisse ermöglichen</i> Unser/e Leiter/in ...		
erfo1	Lobt mich, wenn ich meine Leistung verbessere	0,830	0,889
erfo2	Lobt schwächere Teilnehmende, wenn sie sich verbessern	0,755	0,817
erfo3	Lobt mich, wenn ich eine gute Leistung erzielt habe	0,845	0,819

Kodierung: (0) trifft nicht zu; (1) trifft eher nicht zu; (2) trifft teils teils zu; (3) trifft eher zu; (4) trifft zu  
Berichtet werden die standardisierten Faktorwerte. Striche zeigen die ausgeschlossenen Items

**Open Access.** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Bundesamt für Sport (Hrsg.). (2023). *J+S-Manual Grundlagen*. BASPO. <https://tool.jugendundsport.ch/asset/6398859d8f0eb283a557723/download?lang=de>
- Bundesamt für Sport. (2019). *Das Magglinger Ausbildungsmodell*. BASPO. [https://www.baspo.admin.ch/content/baspo-internet/de/bildung-und-forschung/magglinger-ausbildungsmodell/\\_jcr\\_content/contentPar/downloadlist/downloadItems/121\\_1581073454438.download/Bro\\_Magglinger\\_Ausbildungsmodell\\_d\\_screen.pdf](https://www.baspo.admin.ch/content/baspo-internet/de/bildung-und-forschung/magglinger-ausbildungsmodell/_jcr_content/contentPar/downloadlist/downloadItems/121_1581073454438.download/Bro_Magglinger_Ausbildungsmodell_d_screen.pdf)
- Charalambous, C. Y., & Praetorius, A.-K. (2018). Studying mathematics instruction through different lenses: setting the ground for understanding instructional quality more comprehensively. *ZDM*, 50(3), 355–366. <https://doi.org/10.1007/s11858-018-0914-8>.
- Collins, D. (2003). Pretesting survey instruments: an overview of cognitive methods. *Quality of Life Research*, 12(3), 229–238. <https://doi.org/10.1023/a:1023254226592>.
- Helmke, A. (2017). *Unterrichtsqualität und Lehrprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (7. Aufl.). Seelze: Klett.
- Helmke, A., & Lenske, L. (2013). Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung. *Beiträge zur Lehrerbildung*, 31(2), 214–233.
- Herrmann, C. (2019). Evaluation der Unterrichtsqualität im Sportunterricht mit dem QUALLIS-Instrument. *Bewegung und Sport*, 73(2), 12–17.
- Herrmann, C., & Gerlach, E. (2017). Unterrichtsqualität im Schneesport. In I. Bach (Hrsg.), *Skilauf und Snowboard in Lehre und Forschung*. Schriftenreihe der ASH, (Bd. 24, S. 118–129). Hamburg: Feldhaus.
- Herrmann, C., & Gerlach, E. (2020). Unterrichtsqualität im Fach Sport – Ein Überblicksbeitrag zum Forschungsstand in Theorie und Empirie. *Unterrichtswissenschaft*, 48, 361–384. <https://doi.org/10.1007/s42010-020-00080-w>.
- Herrmann, C., Seiler, S., Pühse, U., & Gerlach, E. (2015). „Wie misst man guten Sportunterricht?“ – Erfassung der Dimensionen von Unterrichtsqualität im Schulfach Sport. *Zeitschrift für Sportpädagogische Forschung*, 3(1), 6–26.
- Herrmann, C., Seiler, S., Gerlach, E., & Sohnsmeier, J. (2018). Unterrichtsqualität im Fach Sport – Dimensionen, Merkmale und Evaluation. In E. Balz & D. Kuhlmann (Hrsg.), *Sportwissenschaft in pädagogischem Interesse*. Schriften der Deutschen Vereinigung für Sportwissenschaft, (Bd. 269, S. 44–46). Hamburg: Feldhaus.
- Herrmann, C., Gogoll, A., & Gerlach, E. (2020). Unterrichtsqualität im Fach Sport. In E. Balz, C. Krieger, W.-D. Miethling & P. Wolters (Hrsg.), *Empirie des Sportsports 3*. Aufl. Edition Schulsport, (Bd. 20, S. 61–81). Aachen: Meyer & Meyer.
- Herrmann, C., Niederkofler, B., & Seelig, H. (2023). Evaluation der Qualität des Lehrens und Lernens im Fach Sport – Validierung des QUALLIS-Instruments. *Swiss Journal of Educational Research*, 45(2), 202–219. <https://doi.org/10.24452/sjer.45.2.10>
- Herrmann, C., Studer, F., Siffert, A. & Seiler, S. (2023). *Y+S Evaluation Instrument for Assessing the Quality of Teaching and Learning in Swiss Children's and Youth Sports: Documentation of Items in German, French, Italian and English*. <https://doi.org/10.5281/zenodo.8359244>
- IBM Corp. (2021). *IBM SPSS statistics for Windows, version 28.0*. IBM Corp.
- J+S-Ausbildungsteam (2019). *Gute J+S Aktivitäten – Vermitteln*. BASPO. [https://lm.baspomedia.ch/mf\\_gute\\_aktivitaeten\\_vermitteln\\_d/index.html](https://lm.baspomedia.ch/mf_gute_aktivitaeten_vermitteln_d/index.html)
- Klieme, E. (Hrsg.). (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie*. Weinheim: Beltz.
- Klieme, E., & Baumert, J. (Hrsg.). (2001). *TIMSS – Impulse für Schule und Unterricht: Forschungsbefunde, Reforminitiativen Praxisberichte und Video-Dokumente*. BMBF publik.
- Kohake, K., Richartz, A., & Maier, J. (2022). Measuring pedagogical quality in children's sports: validity and reliability of the classroom assessment scoring system K-3 in extracurricular sports training. *German Journal of Exercise and Sport Research*, 53(1), 47–58. <https://doi.org/10.1007/s12662-022-00836-9>.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers: results from the COACTIV project*. New York: Springer.
- Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Analyse von Lernumwelten: Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 20, 85–96.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (13. Aufl.). Weinheim: Beltz.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus: statistical analysis with latent variables: user's guide (version 8)*
- Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics*, 7(3), 349–365.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS) manual, pre-K*. Baltimore: Paul Brookes.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of three basic dimensions. *ZDM*, 47(Suppl. 1), 97. <https://doi.org/10.1007/s11858-018-0918-4>.
- Raubenheimer, J. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology*. <https://doi.org/10.4102/sajip.v30i4.168>.
- Richartz, A., & Anders, D. (2016). Pädagogische Qualität des Trainings in der Talentförderung. Entwicklung, Implementation und Evaluation von videobasierten Lehr-Lern-Umgebungen für das Grundagentraining in den Sportarten Turnen und Judo. In E. Hillenbach (Hrsg.), *BISp-Jahrbuch: Forschungsförderung 2014/15* (1. Aufl. S. 193–204). Köln: Sportverlag Strauß.
- Richartz, A., Kohake, K., & Maier, J. (2021a). Die Förderung pädagogischer Qualität im Kinder- und Jugendsport – vier videogestützte Module für die Trainerbildung. In N. Neuber (Hrsg.), *Kinder- und Jugendsportforschung in Deutschland – Bilanz und Perspektive*. Bildung und Sport, (Bd. 26, S. 365–385). Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-30776-9\\_19](https://doi.org/10.1007/978-3-658-30776-9_19).
- Richartz, A., Maier, J., & Kohake, K. (2021b). Pädagogische Qualität des Trainings im Kinder- und Jugendsport – normative und wirksamkeitsorientierte Kriterien. In N. Neuber (Hrsg.), *Kinder- und Jugendsportforschung in Deutschland – Bilanz und Perspektive*. Bildung und Sport, (Bd. 26, S. 171–201). Springer VS. [https://doi.org/10.1007/978-3-658-30776-9\\_9](https://doi.org/10.1007/978-3-658-30776-9_9).
- Weiber, R., & Mühlhaus, D. (2014). Güteprüfung reflektiver Messmodelle. In R. Weiber & D. Mühlhaus (Hrsg.), *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl. S. 127–172). Berlin: Gabler.

**Hinweis des Verlags.** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.