

Ger J Exerc Sport Res  
<https://doi.org/10.1007/s12662-024-01004-x>  
Eingegangen: 22. Dezember 2023  
Angenommen: 25. Oktober 2024

© The Author(s) 2024



Bastian Schittkowski<sup>1</sup> · Alexander Woll<sup>2</sup> · Ingo Wagner<sup>3</sup>

<sup>1</sup> Institut für Sport und Sportwissenschaft (IfSS), Karlsruher Institut für Technologie (KIT), Karlsruhe, Deutschland

<sup>2</sup> Institut für Sport und Sportwissenschaft (IfSS), Karlsruher Institut für Technologie (KIT), Karlsruhe, Deutschland

<sup>3</sup> Institut für Sport und Sportwissenschaft, Universität Freiburg, Freiburg, Deutschland

# Multidimensionale Herausforderungen und Strategien der Leistungsbewertung im Sportunterricht: eine qualitative Analyse

## Einleitung

Die Leistungsbewertung im Sportunterricht spielt eine wichtige Rolle für Schüler:innen sowie Lehrpersonen (Deutscher Sportbund [DSB], 2006) und wird auch international zunehmend in den Fokus gerückt (Hay & Penney, 2009; Tomolyius & Surkamin, 2017). Im Schulfach Sport ist die Bewertung dabei andersartig bzw. hinsichtlich der zusätzlichen Bewegungskomponente sogar komplexer als im Klassenzimmerunterricht (Siedentop, Hastie, & van der Mars, 2011). Diese Komplexität ergibt sich u.a. aus den unterschiedlichen Anforderungen eines kompetenzorientierten Sportunterrichts (Guardiera, 2019), in dem neben kognitiven auch soziale, affektive und sportmotorisch-physische Kompetenzen eine Rolle spielen. Die verschiedenenartigen Kompetenzen werden nicht nur vermittelt und erworben, sondern auch benotet (Almreiter et al., 2020). Dafür stehen Bewertungsdimensionen zur Verfügung, die je nach Schwerpunktsetzung unterschiedliche Rollen spielen können (Gehrman & Kastrup, 2023).

Die Sportnote stellt sich also als ein Konglomerat einer multidimensionalen Bewertung von Leistungen über ein gesamtes Schuljahr hinweg dar. Die Leistungen in den einzelnen Bereichen werden teilweise als messbare

Leistungen, oft jedoch auch durch reine Lehrpersonenbeurteilungen bewertet (Gülich & Krüger, 2013). Dass fachdidaktische Ansprüche an die Bewertung des Sportunterrichts laut Scheid und Prohl (2022) nicht immer deckungsgleich mit bildungspolitischen Vorgaben sind, verschärft die Herausforderungen für Sportlehrpersonen bei der Bewertung. Die bundesland- sowie schulspezifischen Vorgaben zur Notenbildung sind einzuhalten, gleichwohl diese oftmals keine expliziten Handlungsanweisungen enthalten.

Somit verbleibt in letzter Konsequenz die Verantwortung der Bildung einer pädagogisch legitimierten Sportnote einzig und allein bei der Lehrperson. Dies mündet bei den Lehrpersonen oftmals in großer Unsicherheit (Feth, 2014), die in der Wahrnehmung umso größer werden kann, je mehr Dimensionen zur Leistungsbewertung herangezogen werden. Es liegt dabei nahe, dass unter der Pflicht Noten zu bilden, sich bei allen Sportlehrpersonen über die Zeit Handlungsmuster herausbilden, diese Unsicherheiten abzuschwächen und somit Fehler bei der Leistungsbewertung zu reduzieren. Angesichts dieser zu bewältigenden Komplexität richtet sich unser Fokus auf die spezifischen Strategien, die Sportlehrpersonen anwenden, um Fehlern bei der Bewertung vorzubeugen

und eine faire, transparente Leistungsbewertung zu gewährleisten. Diese Studie untersucht daher die Frage:

Welche spezifischen Strategien verwenden Sportlehrpersonen, um die Herausforderungen der Leistungsbewertung im Sportunterricht zu bewältigen und Bewertungsfehler zu minimieren? Durch die Beleuchtung dieser Fragestellung mittels einer Interviewstudie hoffen wir, einen tieferen Einblick in die Praxis der Leistungsbewertung im Sportunterricht zu geben und zugleich praktische Ansätze für Lehrkräfte zu identifizieren, die zu einer verbesserten Bewertungspraxis beitragen können.

## Theoretischer Hintergrund

### Der Leistungsbegriff

Bevor die Leistung einer Schülerin oder eines Schülers durch eine Lehrperson bewertet werden kann, ist es erforderlich, dass sich die Lehrperson des Begriffs „Leistung“ und dessen Bedeutung bewusst ist. Es existiert keine allgemeingültige Definition des Leistungsbegriffs, was bereits zu Beginn des Bewertungsprozesses zu unterschiedlichen Ansätzen führen kann. Diese Variabilität im Verständnis und in der Anwendung des Leistungsbegriffs kann dazu führen, dass die

## Hauptbeitrag

Bewertungen durch verschiedene Lehrpersonen nicht einheitlich sind.

Während Erdmann (1993) von einer „absichtsvollen Handlung“ spricht, differenzieren Grupe und Mieth (2001) in vier „Sichtweisen der Leistung“:

1. sachlich orientierte Leistung (eine gute Ausführung wird betont),
2. individuumszentrierte Leistung (der Fokus liegt darauf, besser zu sein als zuvor),
3. sozialbezogene Leistung (die Leistung im Vergleich zu einem Partner bewertet) und
4. normorientierte Leistung (basierend auf externen Gütekriterien).

Diese bilden die Basis der Bezugsnormen nach Rheinberg (2006), die als Standards fungieren, um eine Leistung zu vergleichen und daraufhin zu beurteilen. Bei jeder Bewertung ist der Bezug zu einer Norm zwingend (Rheinberg, 2014), da diese sonst nicht eingeordnet werden kann. Diese Bezugsnormorientierung geschieht jedoch häufig nicht bewusst (Kleiner, 2019), ist als überdauerndes Persönlichkeitsmerkmal in der Lehrperson selbst schon angelegt.

Svennberg, Meckbach und Redelius (2014) bezeichnen diese Internalisierung der Bewertungsrichtlinien als „Bauchgefühl“. Sie kann allerdings situativ variabel angepasst werden (Sülz, 2014). Rheinberg (2014) unterscheidet 3 Bezugsnormen, die (1) soziale (der Vergleich mit anderen), die (2) sachliche bzw. kriteriumsorientierte (der Vergleich mit vorgegebenen Standards) und die (3) individuelle (der Vergleich mit sich selbst) Bezugsnorm. Diese können von den Lehrpersonen einzeln oder gemischt in unterschiedlicher Gewichtung bei der Bewertung einer Leistung verwendet werden.

### Bewertungsdimensionen

Im Rahmen der Kompetenzorientierung im Sportunterricht, bei dem es nicht nur um fachliche Inhalte geht, sind die Unterrichtsziele auch darauf ausgerichtet, überfachliche Fähigkeiten und Fertigkeiten zu entwickeln (Guardiera, 2019). Dies bedeutet, dass Wissen und Können durch zusätzliche Leistungsdimensionen erweitert werden, die bei der

Leistungsbewertung ebenfalls berücksichtigt werden müssen (ebd.). Darüber hinaus wird der Kompetenzerwerb als ein prozesshafter Vorgang verstanden, bei dem die individuelle Entwicklung der Schüler:innen in den verschiedenen Leistungsdimensionen im Rahmen der kompetenzorientierten Leistungsbewertung einbezogen wird (ebd.). Diese verschiedenen Aspekte der Leistung im Sportunterricht spiegeln die Komplexität der Bewertung in diesem Fach wider. Sie zeigen, dass nicht nur die physische Leistungsfähigkeit, sondern auch kognitive, soziale und emotionale Fähigkeiten eine wichtige Rolle spielen. Lehrpersonen müssen daher ein breites Spektrum an Bewertungskriterien anwenden, um ein vollständiges Bild der Leistungen ihrer Schüler:innen zu erhalten. Eine reine Sportartenorientierung genügt bei Weitem nicht den Bildungsansprüchen des Schulfaches Sport (Prohl, 2010).

Um dieser Multidimensionalität in Erweiterung der Bezugsnormen besser gerecht werden zu können, haben Gehrman und Kastrup (2023) aus der aktuellen fachdidaktischen Literatur 5 Bewertungsdimensionen extrahiert:

1. die sportmotorische Dimension (messbare Leistungen),
2. den Leistungsfortschritt (individuelle Fortentwicklung der Fähigkeit der Person),
3. die Leistungsbereitschaft,
4. die soziale Dimension (bspw. Fairness, Kooperationsbereitschaft) und
5. die kognitive Dimension (schriftliche Leitungen).

Je nach fachlicher oder pädagogischer Schwerpunktsetzung können die Bewertungsdimensionen alle, einzeln oder in unterschiedliche Rollen mit unterschiedlicher Gewichtung eingesetzt werden. Damit ergibt sich für die Lehrpersonen über die Bezugsnormorientierung hinaus ein flexiblerer Umgang bei der Bewertung von Leistungen, v.a. vor dem Hintergrund der Bewertung von zeitlich andauernden Prozessen oder Prozessen der sozialen Interaktion.

### Bewertungsfehler

Die unterschiedlichen Definitionen von Leistung und die variierenden Gewichtungen der Bewertungsdimensionen führen zu einer hohen Subjektivität und potenziellen Fehlern in der Bewertung von Schüler:innenleistungen. Gepaart mit der immanenten Subjektivität des reinen Beobachtungsvorgangs bei der Bewertung selbst, stellt dies die Objektivität und Gerechtigkeit der Sportnoten infrage und kann zu ungerechten Bildungsentscheidungen führen. Brookhart (2013) betont, dass die Leistungsbewertung nicht nur ein technischer, sondern auch ein interpretativer Prozess ist, in dem Lehrpersonen ihre subjektiven Wahrnehmungen und Erfahrungen einbringen. Dies unterstreicht die Notwendigkeit, dass Lehrpersonen sich ihrer eigenen Bewertungsvorlieben und -tendenzen bewusst sein müssen, um eine gerechtere und objektivere Leistungsbewertung zu gewährleisten. Um dies zu verbessern, ist eine Sensibilisierung der Lehrpersonen für die Komplexität der Leistungsbewertung essenziell, da nur so etwaige Fehler minimiert werden können (Jürgens, 2005).

Die Herausforderungen bei der Leistungsbewertung sind vielfältig. Neben der Neigung von Lehrpersonen, die Schülerleistungen stets nach der gleichen Gauß'schen Verteilung zu bewerten (von Saldern, 2011), spielt die persönliche Expertise eine Rolle. Zu den persönlichen Erfahrungen der Lehrperson kommen stets kontext- und situationsabhängige Beobachtungen (Tageszeit, Gruppensammensetzung, ...), die das Verhalten und die Leistungen der Schüler:innen beeinflussen können (Miceli, 2021). Miceli (2021) spricht neben Bewertungsfehlern von Beobachtungsverzerrungen und zeigt damit die Schwierigkeit der Objektivität erneut auf. Häufige Verzerrungen, die in Fehlern münden, sind nach Miceli und Fiebeler (2020):

- Pygmalion-Effekt: Dieser Effekt beschreibt die Wechselwirkung zwischen den Erwartungen einer Lehrperson und ihrem Verhalten gegenüber den Schüler:innen sowie den darauf folgenden Reaktionen der Schüler:innen. Diese Dynamik kann

	<b>Zusammenfassung · Abstract</b>
<p>das Selbstkonzept und die Motivation der Schüler:innen signifikant beeinflussen.</p> <ul style="list-style-type: none"> <li>— Halo-Effekt: Hierbei wird eine Eigenschaft der Schülerin oder des Schülers auf andere, unzusammenhängende Bereiche/Sportarten/Kompetenzen übertragen, was zu einer verzerrten Wahrnehmung führen kann.</li> <li>— Reihungseffekt: Dieser Effekt bezieht sich auf die Verzerrung der Bewertung, die durch die Reihenfolge beeinflusst wird, in der Informationen präsentiert werden. Er tritt auf, wenn die Position einer Person in einer Reihe von Bewertungen die Beurteilung dieser Person beeinflusst.</li> <li>— Stereotype: Das Heranziehen von Stereotypen kann dazu führen, dass bestimmte Eigenschaften fälschlicherweise als typisch für eine bestimmte Gruppe angesehen werden, was die objektive Beurteilung beeinträchtigen kann.</li> <li>— Tendenz zur Mitte: Diese Tendenz zeigt sich in der Zurückhaltung von Lehrpersonen, extreme Bewertungen vorzunehmen.</li> <li>— Kontrastfehler: Dieser Fehler tritt auf, wenn die Beurteilung eines Schülers/ einer Schülerin im Vergleich zu vorher bewerteten Schüler:innen erfolgt. Ein durchschnittliche/r Schüler:in könnte z. B. nach einer Gruppe von überdurchschnittlichen Schüler:innen schlechter bewertet werden als gerechtfertigt.</li> <li>— Voreingenommenheit durch physische Präsenz: Lehrpersonen tendieren dazu die Leistungen von Schüler:innen, die physisch präsenter und sichtbarer im Unterricht sind, überbewerten, während zurückhaltendere Schüler:innen unterbewertet werden könnten.</li> <li>— Überbewertung von quantifizierbaren Ergebnissen: In Sportarten, die leicht messbare Ergebnisse liefern (wie Laufzeiten oder Weitsprünge), könnten qualitative Aspekte wie Teamarbeit, Anstrengung und technische Fähigkeiten unterbewertet werden.</li> </ul> <p>Bewertungsfehler setzen die Existenz einer objektiven Bewertung voraus, von der</p>	<p>Ger J Exerc Sport Res <a href="https://doi.org/10.1007/s12662-024-01004-x">https://doi.org/10.1007/s12662-024-01004-x</a>      © The Author(s) 2024</p> <p><b>B. Schittkowski · A. Woll · I. Wagner</b></p> <p><b>Multidimensionale Herausforderungen und Strategien der Leistungsbewertung im Sportunterricht: eine qualitative Analyse</b></p> <p><b>Zusammenfassung</b></p> <p>Die vorliegende Interviewstudie untersucht die Herausforderungen und Strategien der Leistungsbewertung im Sportunterricht. Hauptaugenmerk liegt auf den spezifischen Herausforderungen des kompetenzorientierten Unterrichts und den Methoden zur Fehlervermeidung. Insgesamt 17 Sportlehrpersonen aus Baden-Württemberg wurden hierzu interviewt. Die Ergebnisse zeigen, dass Lehrpersonen eine Vielzahl von Strategien anwenden, um Transparenz zu schaffen und Bewertungsfehler zu minimieren. Dazu gehören die Verwendung von Bewertungstabellen, die Einbindung der Schüler:innen in die</p> <p>Kriterientwicklung und die systematische Planung der Bewertung über das Schuljahr hinweg. Auf Basis der Studienergebnisse werden die vielfältigen Bewertungsdimensionen und Normenorientierungen diskutiert, die im Sportunterricht zur Anwendung kommen sowie praktische Ansätze zur Verbesserung der Bewertungspraxis deduziert.</p> <p><b>Schlüsselwörter</b></p> <p>Sportunterricht · Leistungsbewertung · Benotung · Noten · Bewertungsfehler · Bezugsnormen · Bewertungsdimensionen</p> <p><b>Multidimensional challenges and strategies of performance assessment in physical education: a qualitative analysis</b></p> <p><b>Abstract</b></p> <p>This interview study examines the challenges and strategies of performance assessment in physical education. The focus is on the specific challenges of competency-based teaching and the methods used to avoid errors in assessment. A total of 17 physical education teachers from Baden-Württemberg were interviewed. The results show that teachers employ a variety of strategies to create transparency and minimize assessment errors. These include the use of evaluation tables, involving students in the development</p> <p>of criteria, and systematically planning assessments throughout the school year. The study highlights the diverse assessment dimensions and norm orientations used in physical education and offers practical approaches to improve assessment practices.</p> <p><b>Keywords</b></p> <p>Physical education · Assessment · Evaluation · Grading · Assessment errors · Reference standards · Assessment dimensions</p>
	<p>abgewichen werden kann. Diese normative Vorstellung gilt es in vielen Fällen zu vermeiden, um die Herausforderungen und Unsicherheiten bei der Leistungsbewertung nicht übermäßig aufzublähen, da Bewertung immer ein gewisses Maß an Subjektivität und Interpretation beinhaltet (Moskal, 2000; Wegis, 2008).</p> <p>Die Reduktion dieser Fehler sollte dennoch stets oberste Priorität haben, um einerseits die Fairness gegenüber den Schüler:innen zu wahren, andererseits aber auch den Bildungsauftrag bestmöglich zu erfüllen. Den Schüler:innen transparent gemachte Bewertungskriterien können helfen, den Bewertungsprozess zu objektiveren (Miceli, 2021). Wiggins spricht schon 1998 von den Prinzipien der fairen Leistungsbewertung, die fordern, dass Bewertungskriterien und -standards klar kommuniziert werden sollten, um den Lernenden eine Orientierung zu geben. Bei der Erarbeitung der Kriterien können die Schüler:innen eingebunden werden, was zu mehr Verständnis und Akzeptanz führt (Bohl, 2015), die Subjektivität weiter verringert (Moskal, 2000) sowie zur Entwicklung von Selbstregulationsfähigkeiten beiträgt (Hattie &amp; Timperley, 2007). Dabei muss nach Wegis (2008) auf die Anzahl und Komplexität geachtet werden, damit der Bewertungsprozess nicht behindert wird. Dies muss jedoch vorher gut in einer Unterrichtseinheit geplant sein,</p>

## Hauptbeitrag

damit genügend Zeit und Raum dafür zur Verfügung steht.

Miceli und Fiebeler (2020) empfehlen des Weiteren einen regelmäßigen Austausch mit Kolleg:innen. Wie Zaborowski, Meier, & Breidenstein (2011) zeigen, können die Gütekriterien der allgemeinen Testtheorie – Reliabilität, Validität, Objektivität in der Schule kaum erfüllt werden.

### Exemplarische Leistungsbewertung im Sportunterricht

Der Sportunterricht ist im schulischen Fächerkanon nicht nur sehr beliebt (Götz, 2010; Kurz & Schulz, 2010), sondern nimmt eine besondere Stellung ein (Balz & Kuhlmann, 2015). Neben den herkömmlichen kognitiven und sozialen Kriterien (Wagner, Bartsch, & Rulofs, 2021) der schulischen Leistungsbewertung kommt die sportmotorische Dimension hinzu (Feth, 2014), was eine ganzheitliche Bewertung der Schüler:innenleistungen erfordert. Während bspw. in Mathematik eine exakte Bewertung durch klar definierte Aufgaben- und Lösungsformate erwartet und in Deutsch ein fachimmanenter Bewertungsspielraum toleriert wird, befindet sich das Fach Sport in einer Doppelrolle. Einerseits gibt es exakte Messungen von Leistungen, andererseits soziale Interaktionen, die es ebenso zu bewerten gilt. Die Leistungen werden als „körperliche Fähigkeit und Fertigkeit (Kompetenz) individuell vorgezeigt und [...] qualitativ bzw. quantitativ [...] bewertet und schließlich in Ziffernnoten umgerechnet (transformiert) [werden]“ (Kleiner, 2019). Die einzelnen Bezugsnormen bieten dabei jeweils unterschiedliche Perspektiven und sind je nach pädagogischem Kontext auch im Sportunterricht unterschiedlich sinnvoll einsetzbar (Paradies, Wester, & Greving, 2009). Beispielsweise könnte bei einem 800-Meter-Lauf die Leistung einer Schülerin anhand verschiedener Normen bewertet werden, oder auch mit Hilfe verschiedener Bewertungsdimensionen bewertet werden:

— Soziale Bezugsnorm: Hier wird die Laufzeit im Vergleich zu den Mitschüler:innen bewertet. Eine Zeit

von 3:30 min, die schneller ist als 75 % der Klasse, könnte zu einer guten Note führen.

- Individuelle Bezugsnorm: Diese berücksichtigt den persönlichen Fortschritt. Eine Verbesserung von 4 min auf 3:30 min kann eine sehr gute Note rechtfertigen.
- Sachliche Bezugsnorm: Die Bewertung erfolgt im Vergleich zu festgelegten Standards, offiziellen Bewertungstabellen oder standardisierten Test wie dem Sportabzeichen, wobei eine Zeit unter 3:45 min eine sehr gute Bewertung bedeuten könnte.

Zusätzlich können verschiedene Bewertungsdimensionen herangezogen werden, um eine umfassende Beurteilung zu gewährleisten:

- Sportmotorische Dimension: Eine gute Laufzeit und/oder korrekte Lauftechnik führen zu einer guten Note.
- Leistungsfortschritt: Wenn die Schülerin ihre Zeit von 4 min auf 3:30 min verbessert hat, wird dies positiv bewertet.
- Leistungsbereitschaft: Zeigt die Schülerin vollen Einsatz und hält durch, auch wenn sie erschöpft ist, kann dies zu einer guten Note führen.
- Soziale Dimension: Unterstützt die Schülerin ihre Mitschüler:innen und zeigt Teamgeist, wird dies positiv bewertet.
- Kognitive Dimension: Kennt die Schülerin die Bedeutung von Aufwärmübungen, Lauftechniken und Regeneration, oder zeigt Wissen über Trainingsmethodik, wird dies positiv in die Bewertung einfließen.

Im Kontext des Sportunterrichts spielt die Auswahl der geeigneten Bezugsnormen und Bewertungsdimensionen eine entscheidende Rolle. Dies wird besonders deutlich, wenn man sich die Vielfalt sportlicher Disziplinen und die damit verbundenen unterschiedlichen Leistungsaspekte und erforderlichen Kompetenzen vor Augen führt. So erfordert beispielsweise die Bewertung in einem Mannschaftssport wie Basketball nicht nur die Berücksichtigung der

technischen Fähigkeiten und der Spielstatistiken, sondern auch des taktischen Verständnisses und der Teamdynamik. Eine Lehrperson könnte hier die individuelle Bezugsnorm verwenden (Möddinger, Woll, & Wagner, 2022), um den Fortschritt der einzelnen Schüler:innen in Bezug auf die persönlichen Fähigkeiten und dem Verständnis des Spiels zu bewerten. Dies würde es ermöglichen, die Fortschritte von Schüler:innen zu würdigen, die vielleicht nicht die meisten Punkte erzielen, dafür erhebliche Verbesserungen in der Spielübersicht und Zusammenarbeit im Team zeigen (Gülich & Krüger, 2013).

Bei der Anwendung der Bewertungsdimensionen könnten die sportmotorische Dimension (technische Fähigkeiten im Dribbeln, Passen und Werfen), der Leistungsfortschritt (Verbesserungen in der Treffsicherheit oder im Ballhandling über einen bestimmten Zeitraum), die Leistungsbereitschaft (der Einsatz und die Anstrengung während des Spiels, insbesondere in kritischen Spielsituationen), die soziale Dimension (das Verhalten im Team, die Kommunikation und der Zusammenhalt mit den Mitspielern) und die kognitive Dimension (das taktische Verständnis des Spiels und die Fähigkeit, Spielzüge zu erkennen und umzusetzen) berücksichtigt werden.

Ein weiteres Beispiel könnte die Bewertung im Turnunterricht sein. In dieser Disziplin sind technische Präzision und die Ausführung von Bewegungsabläufen zentral (Gutbrod & Wagner, 2023). Hier könnte eine sachliche Bezugsnorm angewandt werden, bei der die Leistung jedes Schülers anhand klar definierter Kriterien bewertet wird, wie z. B. die korrekte Ausführung eines Handstands oder einer Rolle vorwärts. Auf diese Weise kann eine objektive Bewertung der technischen Ausführung erfolgen, ohne dass der Vergleich mit anderen Schüler:innen die Bewertung beeinflusst (Weinert, 2014). Auch hier könnten die Bewertungsdimensionen sportmotorische Dimension (Präzision und Qualität der ausgeführten Bewegungen), der Leistungsfortschritt (Verbesserung der Bewegungsabläufe und Techniken über die Zeit), die Leistungsbereitschaft (Anstrengungsbereitschaft und Wille, schwierige Übungen zu

erlernen und durchzuführen), die soziale Dimension (Unterstützung und Zusammenarbeit mit anderen Schüler:innen, beispielsweise beim Hilfestellen) und die kognitive Dimension (Verständnis der Techniken und Fähigkeit, Bewegungsabläufe theoretisch zu erklären) eine Rolle spielen.

Diese Beispiele verdeutlichen, wie die Auswahl der Bezugsnormen und Bewertungsdimensionen je nach sportlicher Aktivität und Lernziel variiert werden kann, um eine faire und umfassende Bewertung der Schüler:innenleistungen zu gewährleisten. Es zeigt sich, dass die Flexibilität in der Anwendung es Lehrpersonen ermöglicht, den individuellen Fertigkeiten und Fähigkeiten der Schüler:innen innerhalb der Kompetenzorientierung gerecht zu werden, während gleichzeitig die Ziele des Sportunterrichts erfüllt werden. Dies erfordert von den Lehrpersonen ein tiefes Verständnis für die Komplexität der Leistungsbewertung in verschiedenen sportlichen Kontexten sowie die Fähigkeit, diese Erkenntnisse praktisch umzusetzen.

## Empirischer Forschungsstand

International wird zur Leistungsbewertung im Sportunterricht wenig geforscht (Zaborowski et al., 2011). Ein überblickgebendes Scoping-Review von Forschungen zur Bewertungspraxis des Sportunterrichts in den USA (Killian & Mays Woods, 2021) identifizierte für die letzten beiden Dekaden insgesamt (nur) 17 vorhandene Artikel.

Empirische Studien zur Leistungsbewertung allgemein zeigen einerseits die Bedeutung der Leistungsbewertung für den Lernfortschritt der Schüler:innen (Kirkwood & Price, 2008), andererseits die Fehleranfälligkeit: Zwar konnten Lehrpersonen in einer Studie klasseninterne Leistungsunterschiede beim Bewerten gut abbilden, scheiterten jedoch im Vergleich mit der Parallelklasse (Schrader & Helmke, 2014). Auch wurde eine verschiedenartige Bewertung gleicher Leistungen durch unterschiedliche Lehrpersonen aufgezeigt (ebd.). Grundlage hierfür ist die Anwendung unterschiedlicher Vergleichsstandards bei der Bewertung einer Leistung (Rheinberg,

2008). Denn schon allein die Veränderung des angelegten Standards kann bei exakt gleichen Ergebnissen die Anzahl besserer bzw. schlechterer Leistungen beeinflussen (Rheinberg, 2006).

Zum Sportunterricht zeigen Studien, dass sich viele Lehrpersonen aufgrund der Komplexität auf einzelne Bewertungsfaktoren fokussieren, um einen Orientierungspunkt zu schaffen (Dassel, 1971). Sie orientieren sich bevorzugt an Leistungstabellen (Meier & Ruin, 2015), was jedoch den Zielen eines kompetenzorientierten Sportunterrichts widerspricht (Gehrman & Kastrup, 2023). Teilweise wird dabei sogar auf reine Trefferanzahlen als Leistungskriterium reduziert (Peiffer, 1998). In einer qualitativen Untersuchung zeigen Gehrman und Kastrup (2023), dass die Gewichtung der verschiedenen Dimensionen in der Praxis stark variiert. Weiter stellen sie fest, dass die sportmotorische Dimension zwar als die wichtigste angesehen wird, aber die soziale Dimension und die Leistungsbereitschaft einen großen Einfluss auf die Gesamtnote haben können.

Während früher detaillierte Punkt- und Wertungstabellen existierten, geben aktuelle Bildungspläne den Lehrpersonen oft einen großen Handlungsspielraum (Wolters, Ehni, Kretschmer, Scherler, & Weichert, 2000), der in allen Facetten ausgefüllt werden muss. Dies wird häufig intuitiv per Bauchgefühl durchgeführt (Svennberg et al., 2014).

Die Organisation und Durchführung der Leistungsbewertung kostet viele Sportlehrpersonen erhebliche Ressourcen (Lavay, Sakai, Ortiz, & Roth, 2015), wobei zeitliche Beschränkungen und aufwändige Aufzeichnungen als Hauptbarrieren einer effektiven Bewertung identifiziert wurden (Gallo, Sheehy, Patton, & Griffin, 2006) und die Planung aufwändiger machen. Dadurch wird der Bewertungsprozess, der das Lernen der Schüler:innen fördern soll, teilweise erschwert (Hay & Penney, 2009). Feth (2014) geht auf Benotungsstrategien ein und unterteilt in drei Strategien: vielfältige Beurteilungsmaßnahmen, sportmotorische Beurteilungsverfahren und Relativierung von Einzelnoten. Die vielfältigen Beurteilungsmaßnahmen

umfassen die Anwendung verschiedener Bewertungsmethoden, um ein umfassendes Bild der Schüler:innenleistungen zu erhalten. Sportmotorische Beurteilungsverfahren konzentrieren sich auf die messbare physische Leistungsfähigkeit der Schüler:innen. Die Relativierung von Einzelnoten erfolgt durch die Erhöhung der Anzahl an Bewertungen, wodurch einzelne schlechtere Leistungen die Gesamtnote weniger stark beeinflussen. Hierbei wird deutlich, dass Sportlehrpersonen bei der Benotung klar erkennbare Strategien ausbilden, um die in der Literatur beschriebenen Herausforderungen aufzufangen.

Weiter treten bei den Sportlehrpersonen teilweise auch Schwierigkeiten auf, sich im Spannungsgelecht unterschiedlicher Leistungsvorstellungen in Sport, Gesellschaft und Schule zurecht zu finden (Meier, 2023), wie auch eine Interviewstudie von Kastrup (2009) auf Basis eines systemtheoretischen Fundaments verdeutlicht.

Über alle Studien hinweg fällt auf, dass die Lehrpersonen bei der Leistungsbewertung mit vielen Problemen konfrontiert sind und die Leistungsbewertung dadurch in vielerlei Hinsicht fehleranfällig ist. Für die einzelnen Herausforderungen werden auch Lösungsvorschläge gemacht, dabei fehlt jedoch der Blickwinkel, wie die Lehrpersonen im Alltag damit umgehen. Das führt uns auf folgende Forschungsfrage (FF) zur Leistungsbewertung im Sportunterricht in Deutschland: Welche spezifischen Strategien verwenden Sportlehrpersonen, um die Herausforderungen der Leistungsbewertung im Sportunterricht zu bewältigen und Bewertungsfehler zu minimieren?

## Methodik

Um diese Fragestellungen angemessen zu bearbeiten, wurde eine leitfadengestützte Interviewstudie in Einzelinterviews in Baden-Württemberg durchgeführt. Diese Methodik macht die Sportlehrpersonen zu Beteiligten des Forschungsprozesses und zeichnet sich durch die erforderliche Subjektivität und Offenheit der Befragten zu ihrem Arbeitsalltag und der Flexibilität in der Durchführung aus (Mey & Mruck, 2010).

### Stichprobe

Um Informationen über die Leistungsbewertung im Sportunterricht zu erhalten, sind 17 Sportlehrpersonen ( $m = 9$ ;  $w = 8$ ) aus Baden-Württemberg befragt worden. Die Auswahl der Teilnehmer:innen erfolgte auf Basis des Geschlechts, der Berufserfahrung und der Fächerkombination, um ein gemischtes Abbild innerhalb der Stichprobe sicherzustellen. Die Teilnehmer:innen waren zwischen 31 und 61 Jahren alt ( $M = 37$ ;  $SD = 7,2$ ) und hatten eine Lehrerfahrung zwischen 3 und 23 Jahren ( $M = 9$ ;  $SD = 4,6$ ). Für alle gilt die in Baden-Württemberg normativ gültige Verordnung des Kultusministeriums über die Notenbildung (NVO), gemäß der die Bildung der Note eine pädagogisch-fachliche Gesamtbewertung aller vom Schüler bzw. von der Schülerin im Beurteilungszeitraum erbrachten Leistungen ist. Weiter gilt das Transparenzgebot (§ 7 NVO), nach dem den Schüler:innen u. a. die maßgebenden Kriterien der Leistungsbewertung dargelegt werden müssen.

### Interviewleitfaden

Um theoretische Annahmen zu berücksichtigen und zugleich die Interviewsituation offen zu halten, wurde ein halbstrukturierter Interviewleitfaden entwickelt. Die Inhalte des Leitfadens wurden auf Basis bisheriger Erkenntnisse zur Leistungsbewertung im Sportunterricht (u. a. Feth, 2014) generiert. Im Sinne eines offenen Einstiegs wurde zunächst mit einer Frage zur persönlichen Einstellung zu Sportnoten begonnen. Für die Hauptphase des Interviews wurden Fragen zur Bewertungspraxis der Lehrperson gestellt. Angesichts der zuvor beschriebenen Erkenntnisse aus der Forschung umfassten diese die Bezugsgrößen, die Planung, die Durchführung (deduktiv) sowie den Umgang mit Herausforderungen und Fehlern und die Transparenz bei der Leistungsbewertung (induktiv).

### Untersuchungsablauf

Die Interviews fanden zwischen September und November 2021 statt. Die Sport-

lehrpersonen wurden einheitlich über den Hintergrund der Studie informiert und über Freiwilligkeit, Anonymität, rechtliche Rahmenbedingungen, Aufnahme des Interviews und die Verarbeitung der Daten aufgeklärt und willigten anschließend ein. Die Dauer der Interviews lag zwischen 25 und 39 min ( $M = 32$ ;  $SD = 4,4$ ). Alle Interviews wurden auf Deutsch geführt. Nach Beendigung des Interviews wurden per Fragebogen soziodemografische Angaben der Interviewpartner:innen (Alter, Geschlecht, Familienstand, Fächer, Berufsjahre, Deputat, Funktionen an der Schule) erfasst.

### Datenauswertung

Alle Interviews wurden aufgezeichnet, transkribiert sowie anonymisiert. Die Transkripte wurden wiederholt gelesen und passende Segmente mit MAXQDA Analytics Pro 2020 nach den deduktiven Kategorien sowie den induktiven Kategorien kodiert, die sich aus dem transkribierten Interviewmaterial herauskristallisierten. Die qualitative Datenanalyse ergab 114 Codes und 452 kodierte Segmente. Diese wurden im Rahmen der qualitativen Inhaltsanalyse nach Mayring ausgewertet (Mayring & Fenzl, 2014).

Es wurden Zitate aus den Interviewtranskripten entnommen, um die Ergebnisse und Interpretationen in Bezug auf die Forschungsfragen zu veranschaulichen. Sowohl das Kategoriensystem als auch die Kodierungen und Interpretationen von Zitaten wurden mehrfach in einer institutsinternen Arbeitsgruppe geprüft und diskutiert.

### Ergebnisse

Die Analyse der 17 Interviews ergab zur übergeordneten Forschungsfrage, welche spezifischen Strategien Sportlehrpersonen verwenden, um die Herausforderungen der Leistungsbewertung im Sportunterricht zu bewältigen, die folgenden Ergebnisse:

### Bezugsgrößen

Alle Lehrpersonen nutzen in der Leistungsbewertung sowohl Wertungsta-

bellen als auch spezifische Kriterien, um die sachliche Bezugsnorm aktiv einzubeziehen. Bemerkenswerterweise berichten die Lehrpersonen von keinerlei Notendiskussionen oder einem erhöhten Rechtfertigungsdruck seitens der Schüler:innen.

Das Bewerten von Spielsportarten wird oft als eine der größten Herausforderungen im Sportunterricht angesehen. Hier setzen Lehrpersonen gezielt Kriterien ein, um die Leistungsbewertung greifbarer zu gestalten und entwickeln diese gemeinsam mit den Schülern während der Unterrichtseinheiten. Viele Lehrpersonen verteilen die Kriterien zu Beginn der Einheit in schriftlicher Form, die oft tabellarisch organisiert werden, um trotz der hohen Anzahl an Schüler:innen und Kriterien den Überblick zu bewahren. Dies ermöglicht es ihnen, Beobachtungen effizient festzuhalten:

„Da habe ich dann eine [Schüler-]Tabelle, wo ich dann auch mehrere Plus eintragen kann.“ (W2, 17)

Die Verwendung selbst gewählter Kriterien wird zudem als eine Möglichkeit gesehen, den Schüler:innen zu ermöglichen, ihre Leistungen im Laufe des Schuljahres zu verbessern. Die Bewertung nach vordefinierten Leistungstabellen wird aufgrund ihrer Einfachheit bevorzugt, da diese selten in Frage gestellt werden.

Viele Lehrpersonen streben danach, die Bewertungen so objektiv wie möglich zu gestalten, indem sie auch in allen Spielsportarten versuchen, über einen Technikparkour messbare Kriterien wie Meter, Sekunde und Gewichte für die Notenerhebung zu nutzen.

Die soziale Bezugsnorm wird in vielfältiger Weise, jedoch unterschiedlich konsequent und oft nur indirekt verwendet. Ein Hauptproblem dabei ist die potenzielle Störung des Klassenrangings. Im Gegensatz dazu wird die Leistungsbereitschaft aktiv bewertet, indem Wertungstabellen zurückgestellt werden, wenn Schüler:innen versuchen, über sich hinauszuwachsen. Dies umfasst auch die Bewertung ihrer Einstellung zum Sport, um eine Grundlage für lebenslanges Sporttreiben zu schaffen. Manchmal werden die Standards angepasst, um die

Schüler positiv zu überraschen und zu motivieren:

*„Das heißt, wenn bei einer Ausdauerleistung bspw. alle schon ganz schlimm traurig gucken, weil alles so schrecklich ist, dann setze ich die Skala so, damit vielleicht doch der eine oder andere noch überrascht ist und denkt: Ach ja, cool, das hat ja jetzt doch geklappt.“ (M5, 5)*

Die soziale Dimension der Bewertung, wie Fairness und Kooperationsbereitschaft, wird ebenfalls berücksichtigt. Einige Lehrpersonen nutzen eine soziale Note, die in die Gesamtbewertung einfließt und Aspekte wie die Unterstützung von Mitschüler:innen und die Beteiligung am Sportunterricht bewertet:

*„Ich mache auch noch eine soziale Note. So eine RSP-Note also für Rahmenbedingungen und soziale Kompetenz und personale Kompetenz und lasse die aber immer bei den einzelnen Sportarten schon mit einfließen.“ (W8, 7)*

Zudem wird die Transparenz der Bewertungsmaßstäbe als entscheidend angesehen, um Fairness sicherzustellen. Durch die Verteilung von Rückmeldungsbögen und klar kommunizierten Bewertungsmaßstäben wird sowohl die Transparenz als auch die kognitive Dimension unterstützt, um den Schülern eine genaue Vorstellung ihrer Leistungen und Verbesserungsmöglichkeiten zu geben. Schriftliche Leistungen hingegen werden nicht eingefordert.

## Bewertungsfehler und Strategien

Die Analyse der Bewertungsfehler verdeutlicht die Komplexität und Vielschichtigkeit der Leistungsbeurteilung. Insbesondere wurden zwei zentrale Punkte identifiziert, die fehleranfällig sind: die objektive Messbarkeit von Leistungen und die subjektive Wahrnehmung von Leistungsverhalten.

**Objektive Messbarkeit.** Die Tendenz, messbare Leistungen zu überbewerten, führt häufig zu einer Vernachlässigung qualitativer Aspekte wie Anstrengung und Teamfähigkeit: „*Da gibt es natürlich auch Noten-Tabellen und die kann ich wunderbar einfließen lassen, ohne dass*

*mir jemand was kann*“ (M9, 7), was die starke Abhängigkeit von quantifizierbaren Bewertungskriterien verdeutlicht.

**Subjektive Wahrnehmung.** Die Bewertung ist häufig von persönlichen Voreingenommenheiten geprägt: „*Manchmal weich ich dann auch ab, weil ja, je nachdem denke ich, ich rede auch mit der Klassenlehrerin: wie sieht es denn aus?*“ (W7, 15)

Die Notenverteilung innerhalb einer Klasse wird von fast allen Lehrpersonen automatisch für alle Klassen ähnlich angenommen, dadurch gibt es eine Tendenz zu einer „schönen Notenverteilung“. Diese Tendenz ist sehr präsent, wird jedoch nicht aktiv unterbunden, denn „wenn man eine starke und schwache Klasse hat, dann ist mysteriöserweise der Durchschnitt am Ende ja trotzdem gleich.“ (M5, 15)

Einen Reihungseffekt gaben sehr viele Lehrpersonen als bewussten Bewertungsfehler an. Bei der Reihenfolge in der die Schüler:innen bewertet werden, zeigen sich individuelle Lösungsansätze, die sehr zielorientiert sind. Manche Lehrpersonen halten sich stets an die Klassensumme und variieren dabei nur den Anfang (W5, 29), andere lassen die vermeintlich besten Schüler:innen beginnen, um ihre Erwartungen mit der Leistbarkeit erneut zu vergleichen (W7, 21). Wieder andere mischen mehrfach durch. Vereinzelt werden bei Choreografien Videoaufnahmen gemacht, teilweise um überhaupt eine faire Bewertung zu ermöglichen, obwohl diese verboten sind.

Als weitere wiederkehrende Fehlerquelle wird das Nachmachen von Noten von fehlenden Schüler:innen genannt, denn ihnen fehlt dann plötzlich doch der direkte Vergleich zum Rest der Klasse, und sie setzen an dem Tag ein bisschen einen anderen Maßstab oder achten auf etwas anderes ein bisschen mehr (M3, 23). Eine Beeinflussung durch Schüler:innen bei der Leistungsbewertung kommt selten vor. Nur in Ausnahmefällen wird im positiven Sinne für die Schüler:innen dadurch eine Note indirekt verbessert (M4, 19).

Die primäre Strategie, um Bewertungsfehler zu vermeiden oder zumindest zu vermindern, ist bei allen Lehr-

personen sich Zeit lassen bei der Leistungsbewertung und Noten erst einmal zurückzuhalten, denn

*„wenn man sich viel Zeit lässt, habe ich das Gefühl [...], dass ich dann weniger Fehler mache. Das heißt, dass ich eben nicht direkt die Note sage, sondern immer sage ich schaue erst mal. Und dann kann ich vergleichen und mir auch noch mal daheim Gedanken machen.“ (W2, 29)*

Die meisten Lehrpersonen sehen sich als fachlich sehr gut ausgebildet und vertrauen dadurch auf ihre eigene Expertise um ohne weitere Hilfsmittel zu fairen Bewertungen zu kommen (M3, 23). Einige Lehrpersonen nehmen gerne die Hilfe von Kolleg:innen in Anspruch, um sich eine weitere Meinung einzuholen und so die Bewertungsfehler nach dem Vier-Augen-Prinzip zu minimieren (M2, 20).

## Planung

Alle Lehrpersonen geben an, die Leistungsbewertung in ihren Unterrichtseinheiten über das ganze Schuljahr hinweg im Vorhinein mit einzuplanen und Einzelnoten in den Sportartenteilelementen zu machen, um diese dann ins Verhältnis zu den übergreifenden Sportartennoten zu setzen (M3, 5).

Bei der Jahresplanung wird vorrangig auf äußere Einflüsse wie Schultermine oder das Wetter bei Nutzung von Außensportanlagen geachtet (W2, 13). Jedoch spielt auch die Art der Leistungsbewertung bei einigen Lehrpersonen in die Planung innerhalb eines Schuljahres, da

*„natürlich Sportarten wie zum Beispiel die Leichtathletik leichter zu bewerten [...] und auch dadurch sehr objektiv sind. Ich versuche dann solche Sportarten zum Beispiel auch ein bisschen zu Beginn des Schuljahres zu machen [...] Das gibt mir auch immer so ein bisschen die Möglichkeit zu gucken, wo steht denn so ein Schüler?“ (W8, 7)*

Die systematische Planung der Leistungsbewertung dient auch der Fehlervermeidung, indem sie eine strukturierte und konsistente Bewertungspraxis über das gesamte Schuljahr hinweg ermöglicht. Innerhalb einer Unterrichts-

## Hauptbeitrag

einheit erfolgt die Leistungsbewertung bei allen Lehrpersonen als Abschluss der Einheit. Je nach Schwerpunkt werden einzelne Teilnoten schon während der Unterrichtseinheit erfasst (W7, 9). Manche (4) Lehrpersonen sind durch Schulcurricula an engere Rahmenbedingungen gebunden. Bei diesen wird in der Schule vorgegeben, welche Sportarten wann gemacht werden (W5, 9). Diese Planung hilft dabei, Bewertungsfehler zu minimieren, indem sie die Bewertung systematisiert und den Lehrpersonen ermöglicht, objektive Kriterien frühzeitig zu berücksichtigen.

### Transparenz

Notentransparenz ist sowohl den Lehrpersonen als auch den Schüler:innen wichtig. Für die Lehrpersonen dient die Transparenz vorrangig der Erwartungskommunikation, hat aber auch Rechtfertigungsfunktion (M1, 23). Bei den Schüler:innen ist die Verteilung heterogener, v.a. leistungsstärkere Schüler:innen fordern mehr Transparenz ein (W5, 21). Die Lehrpersonen führen während der Unterrichtseinheiten häufig detaillierte Notizen oder nutzen Bewertungsraster und verwenden diese zur späteren Reflexion auch mit den Schüler:innen. Dies hilft ihnen dabei, einen umfassenderen Überblick über die Leistungen der Schüler:innen zu behalten.

Ob eine hohe Transparenz zu fairen Noten führt, ist umstritten (W4, 19). Frühe Kommunikation der Erwartungen wird dabei jedoch von allen als wichtiger Punkt für eine gute Notentransparenz angegeben (M4, 21). Die meisten Lehrpersonen teilen hierfür ihre Bewertungskriterien im Vorhinein (schriftlich) aus (W5, 19) und beziehen sich im Laufe einer Einheit immer wieder darauf, bzw. arbeiten damit weiter:

*„Die bekommen vorher genau gesagt, was die Kriterien sind für die Leistungsbewertung. Und ich versuche auch während der Einheit natürlich schon zu sagen: Das und das sollten Sie verbessern, damit sie dann auch sich darauf vorbereiten können und verbessern können.“ (W2, 17)*

### Diskussion

Die Ergebnisse dieser Interviewstudie zeigen deutlich, dass Lehrpersonen verschiedene Strategien entwickeln, um die Komplexität und Herausforderungen der Leistungsbewertung zu bewältigen. Diese Strategien müssten der Ansicht der Autoren nach auch in der universitären Ausbildung thematisiert werden. Die Strategien spiegeln nämlich die im theoretischen Teil beschriebenen Forschungsergebnisse wider und vertiefen das Verständnis für die praktische Umsetzung dieser Erkenntnisse im schulischen Alltag, haben also durchaus praktische Relevanz. Im Gegensatz zu den Ergebnissen Feths (2014) konnte nur wenig Rechtfertigungsdruck bzgl. der Noten auf der Seite der Lehrpersonen festgestellt werden. Dies könnte daran liegen, dass es sich bei der Stichprobe aufgrund der Freiwilligkeit um eine positiv verzerrte Auswahl handelt oder daran, dass im Fach Sport kaum schlechte Noten vergeben werden (Kastrup, 2009; Götz, 2010). Letztgenanntes kann auf Basis der Interviews erklärt werden als Teil einer zugrundeliegenden normativen Setzung, wonach zum lebenslangen Sporttreiben motiviert werden soll, wofür wiederum die Vergabe überwiegend guter Noten den Lehrkräften ein geeignetes Hilfsmittel darstellt.

Durch diese – nicht offenliegende – Handlungsnorm erfährt die Leistungsbewertung im Sportunterricht eine fachspezifische Verschiebung der Gauß'schen Normalverteilung (von Saldern, 2011). Diese fachspezifische Norm scheint das Handeln der Lehrpersonen bei der Leistungsbewertung zumindest implizit oder unbewusst zu beeinflussen, sei es auch nur durch eine angepasste Festlegung der Bewertungsdimensionen. Zukünftig wäre zu eruieren, inwiefern diese aufgedeckte Handlungsnorm in der Fachkultur tabuisiert wird, da bei Offenlegung eine Abwertung des Faches mit einhergehendem Statusverlust befürchtet wird (Kurz & Schulz, 2010). Die potenzielle Beeinflussung stellt in diesem Sinne jedoch keine Verzerrung der Beobachtung (Miceli, 2021) dar und ist somit nicht als Bewertungsfehler zu werten, was die Sportnote nicht legitimiert. Zudem sollte dazu

die Sicht der Schüler:innen zu dieser Bezugsnorm erforscht werden, da vielleicht schon Noten im Bereich von befriedigend als sehr schlecht und damit doch demotivierend wahrgenommen werden – die Bestrebungen der Lehrpersonen somit absurdum geführt würden.

Die Ergebnisse zeigen, dass die befragten Lehrpersonen die Aussicht auf eine weitere Professionalisierung (Helsper, 2021) ihres Handelns antriebt, Bewertungsfehler zu reduzieren. Hierfür werden Kriterien in vielerlei Hinsicht als eine Leitstrategie verwendet, die sich gut in die universitäre Ausbildung integrieren ließen, hier allerdings oftmals zu wenig berücksichtigt werden: Mit geeigneten Kriterien können Hilfsmittel wie Bewertungsschemata verwendet werden, Schüler:innen können in den Erstellungs- und Bewertungsprozess eingebunden werden und es können je nach Bedarf verschiedene Bewertungsdimensionen betrachtet werden. Einzelne Lehrpersonen entwickeln ihre Kriterien im Lernkontinuum, was nach Moskal (2000) die Subjektivität weiter verringert. Um den Bewertungsprozess nicht zu behindern, muss dabei nach Wegis (2008) auf die Anzahl und Komplexität geachtet werden, was die hier durchgeführte Interviewstudie ebenso bestärkt. Bei der kriteriengeleiteten Leistungsbewertung bewerten die Sportlehrkräfte oftmals anhand von vorgefertigten Wertungstabellen. Die Zeitersparnis, die komplikationslose Rückmeldung und die Unanfechtbarkeit bei der Leistungsbewertung sind genannte Vorteile. Entgegen der Kompetenzorientierung zeichnet sich hier eine Ausrichtung am fachdidaktischen Konzept der Sportartenorientierung ab, die im Hinblick auf Bildungsansprüche jedoch als reduziert eingeordnet wird (Prohl, 2010).

Transparenz ist sowohl Lehrpersonen als auch Schüler:innen sehr wichtig. Dies steht in Einklang mit den Prinzipien der fairen Leistungsbewertung, die fordern, dass Bewertungskriterien und -standards klar kommuniziert werden sollten, um den Lernenden eine Orientierung zu geben (Wiggins, 1998). In diesem Sinne werden die Leistungserwartungen sehr früh im Unterrichtsgang kommuniziert und mit diesen im Unterricht gearbeitet, was das Verständnis der Schüler:innen

für die Bewertungskriterien fördert und zur Entwicklung von Selbstregulationsfähigkeiten beiträgt (Hattie & Timperley, 2007).

Technikparkoure sind aufgrund ihrer Einfachheit sehr beliebt, was sich mit den Erkenntnissen von Meier und Ruin (2015) deckt und angesichts der sportmotorischen Dimension auch wenig Raum für Bewertungsfehler lässt. Der von Gehrmann und Kastrup (2023) eruierte Widerspruch zur Kompetenzorientierung wird dabei unter der Betrachtung der Vorteile in Kauf genommen.

Eine weitere Leitstrategie ist der Faktor Zeit. Aufgrund der knappen zeitlichen Ressourcen (Gallo et al., 2006) bleibt je nach Klassengröße für die Bewertung einer einzelnen Leistung nur wenig Zeit. Um diesen Zeitdruck, der fehleranfällig wirken kann, zu entlasten, geben die Lehrpersonen ihre Bewertungen nicht sofort preis. Dies ermöglicht es die Leistungen wirken zu lassen und auf genannte Effekte hin zu überdenken.

Die Strategien der Lehrpersonen den Reihungseffekt abzuschwächen, dürfen ebenso einen etwaigen Kontrastfehler reduzieren. Das legt nahe, dass die in der Praxis entwickelten Strategien multifunktional einsetzbar sind. Die Verschiedenartigkeit der Strategien zeigen deutlich, wie bemüht die Lehrpersonen sind, im Rahmen ihres Alltags, Bewertungsfehler zu minimieren.

Der von Miceli (2021) empfohlene Austausch mit Kolleg:innen wird rege genutzt, und es wird oftmalig eine Art schulinternes Peer-Review angewendet, indem Kolleg:innen bei Leistungsbewertungen mit hinzugenommen werden. Ein solches 4-Augen-Prinzip vermindert die Anzahl der Störgrößen und erhöht somit die Bewertungsqualität.

Die Planung der Leistungsbewertung über das gesamte Schuljahr hinweg stellt eine weitere wichtige Strategie dar, um Bewertungsfehler zu minimieren und eine konsistente Bewertungspraxis zu gewährleisten. Lehrpersonen planen die Leistungsbewertung im Voraus und passen sie an die äußereren Rahmenbedingungen sowie an die spezifischen Anforderungen der verschiedenen Sportarten an. Diese systematische Planung entspricht den Empfehlungen von Feth (2014),

der die Bedeutung einer strukturierten und durchdachten Benotungsstrategie im Sportunterricht betont.

Ferner erkennen die Lehrpersonen die Komplexität und Subjektivität der Leistungsbewertung, da sie um die Bewertungsfehler wissen. Schließlich ist Bewertung ein von menschlichen Wahrnehmungen und Interpretationen beeinflusster Prozess (Brookhart, 2013), gegen dessen Fehler Lehrpersonen gleichzeitig strategische Ansätze zur Minimierung entwickeln, wie etwa das Zurückhalten von Noten zur späteren Reflexion und Vergleich. Dies deutet darauf hin, dass Lehrpersonen im Sportunterricht bestrebt sind, die Bewertungsgerechtigkeit zu verbessern, obwohl Fehlerquellen im Alltag kaum vermieden werden können, wie Bohl (2015) beschreibt.

Auf praktischer Ebene kann die Minimierung von Bewertungsfehlern bereits in der universitären Ausbildung der Sportlehrpersonen implementiert werden. Eine vertiefte Schulung im Bereich der Bewertungsmethoden und -strategien ist essentiell, um zukünftige Lehrpersonen adäquat auf die Herausforderungen der Leistungsbewertung vorzubereiten. Dabei sollten praxisnahe Übungen und Simulationen integriert werden, um die Reduktion subjektiver Einflüsse zu fördern und die Anwendung objektiver Kriterien zu stärken. Zusätzlich könnten regelmäßige Fortbildungen und ein intensiver Austausch mit Kolleg:innen die kontinuierliche Weiterentwicklung der Bewertungskompetenz unterstützen. Die Einführung von Peer-Review-Verfahren, bei denen Lehrpersonen gegenseitig ihre Bewertungspraxis überprüfen und reflektieren, könnte ebenfalls dazu beitragen, Bewertungsfehler zu minimieren. Solch ein Evaluationssystem kann den Lehrpersonen regelmäßig Feedback zu ihrer Bewertungspraxis geben, um die Reliabilität und Validität der Bewertungen zu überprüfen.

## Grenzen der Studie

Die in dieser Studie enthaltenen Forschungsergebnisse zeigen subjektive Erfahrungen einzelner Lehrpersonen in Bezug auf die Leistungsbewertung von

Schüler:innen auf. Diese können je nach Bundesland, Schulcurriculum, Klassenzusammensetzung und Lehrperson anders ausfallen und sind deshalb nicht repräsentativ. Alle Interviews unterliegen der verbalen Kompetenz des/der Interviewten und den Fähigkeiten des Interviewers. Die zeitliche Begrenzung der Interviews könnte die Ergebnisse beeinflussen, da Veränderungen in der Bewertungspraxis über längere Zeiträume nicht berücksichtigt wurden. Darüber hinaus bietet die Studie keine quantitativen Daten, die eine breitere Verallgemeinerung ermöglichen würden. Schließlich wurden spezifische Schulkulturen und politische Rahmenbedingungen nicht detailliert berücksichtigt, obwohl sie die Bewertungspraktiken erheblich beeinflussen könnten. Diese Grenzen unterstreichen die Notwendigkeit weiterer Forschung mit größeren und diverseren Stichproben, einer Mischung aus qualitativen und quantitativen Methoden.

## Ausblick

Die vorliegende Studie legt nahe, dass kontinuierliche Anpassungen und Verbesserungen der Bewertungspraktiken im Sportunterricht notwendig sind. Insbesondere die Entwicklung klarer und transparenter Bewertungskriterien sowie die Schulung der Lehrpersonen im Umgang mit Bewertungsfehlern könnten dazu beitragen, die Fairness und Objektivität der Leistungsbewertung zu erhöhen. Zukünftige Forschungen sollten auch die Perspektive der Schüler:innen einbeziehen, um ein umfassenderes Bild der Bewertungspraxis im Sportunterricht zu erhalten. Diese könnten direkte Beobachtungen der Bewertungspraxis mit einbeziehen, um die Erkenntnisse der tatsächlichen Umsetzung der beschriebenen Strategien zu erweitern.

### Korrespondenzadresse



**Bastian Schittkowski**  
Institut für Sport und  
Sportwissenschaft (IfSS),  
Karlsruher Institut für  
Technologie (KIT)  
Engler-Bunte-Ring 15,  
76131 Karlsruhe, Deutschland  
bastian.schittkowski@kit.edu

**Funding.** Open Access funding enabled and organized by Projekt DEAL.

### Einhaltung ethischer Richtlinien

**Interessenkonflikt.** B. Schittkowski, A. Woll und I. Wagner geben an, dass kein Interessenkonflikt besteht.

Alle beschriebenen Untersuchungen am Menschen oder an menschlichem Gewebe wurden mit Zustimmung der zuständigen Ethikkommission, im Einklang mit nationalem Recht sowie gemäß der Deklaration von Helsinki von 1975 (in der aktuellen, überarbeiteten Fassung) durchgeführt. Von allen beteiligten Patient/-innen liegt eine Einverständniserklärung vor.

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

- Almreiter, B., Krieger, C., Riecke-Baulecke, T., Bähr, I., Bott, L., Deister, D., vom Endt, M., Gaum, C., Graf, T., Gruitrooy, A., Hamann, F., Heemsoth, T., Klimmek, A., Kröger, C., Löck, C., Lüdike, F., Miehlke, M., Ramm, G., Rantzow, O., et al. (2020). Basiswissen Lehrerbildung: Sport unterrichten. [http://bvbr.bib-bvb.de:8991/exlibris/aleph/a23\\_1/apache\\_media/3FTUIDHM15HR6SCL9FVAT5DLKB3RCL.pdf](http://bvbr.bib-bvb.de:8991/exlibris/aleph/a23_1/apache_media/3FTUIDHM15HR6SCL9FVAT5DLKB3RCL.pdf)
- Balz, E., & Kuhlmann, D. (2015). *Sportpädagogik. Ein Lehrbuch in 14 Lektionen*. Aachen: Meyer & Meyer.
- Bohl, T. (2015). *Prüfen und Bewerten im offenen Unterricht* (4. Aufl.). Weinheim: Beltz.
- Brookhart, S.M. (2013). How to create and use rubrics for formative assessment and grading. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=1123215>
- Dassel, H. (1971). Mixtum compositum. Zur Frage der Benotung im Schulsport. *Die Leibeserziehung*, 20(3), 87–89.
- DSB (2006). Die Sprint Studie. Eine Untersuchung zur Situation des Schulsports in Deutschland. <https://kinderrechte.rlp.de/fileadmin/kinderrechte/Materialien/Entwicklung-und-Gesundheit/Bewegung/Sprint-Studie.pdf>
- Erdmann, R. (1993). Leisten – Leistung – Sportunterricht. *sportpädagogik*, 17(3), 11–17.
- Feth, C. (2014). *Benotungspraxis von Sportlehrkräften – Rekonstruktion von Benotungsstrategien im Sportunterricht*. Bochum: Ruhr-Universität.
- Gallo, A. M., Sheehy, D. A., Patton, K., & Griffin, L. (2006). Assessment benefits and barriers. *Journal of Physical Education, Recreation & Dance*, 77(8), 46–50. <https://doi.org/10.1080/07303084.2006.10597926>.
- Gehrmann, S., & Kastrup, V. (2023). Benotungspraxis im Sportunterricht an Grundschulen und in der Sekundarstufe II – Eine qualitative Untersuchung zu Benotungskriterien und deren Gewichtung. *Zeitschrift für Bildungsforschung*, 13(3), 365–383. <https://doi.org/10.1007/s35834-023-00401-w>.
- Götz, T. (2010). Lehrerhandeln im Sportunterricht der Schule. Beiträge zur Lehre und Forschung im Sport. In A. Woll, F. Mess & H. Haag (Hrsg.), *Handbuch Evaluation im Sport* (Bd. 177, S. 271–284). Schorndorf: Hofmann. Abruf unter <http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-129545>.
- Grupe, O., & Mieth, D. (Hrsg.). (2001). Schriftenreihe des Bundesinstituts für Sportwissenschaft: Bd. 99. <https://www.gbv.de/dms/faz-rez/F19980528ETHIK--100.pdf>
- Guardiera, P. (2019). Leistungen bewerten: Herausforderungen und Perspektiven der Leistungsbeurteilung im kompetenzorientierten Sportunterricht. *Sportpädagogik: Zeitschrift für Sport-, Spiel- u. Bewegungserziehung*, 43(2), 2–6. Abruf unter <https://fis.dshs-koeln.de/de/publications/leistungen-bewerten-herausforderungen-und-perspektiven-der-leistu>.
- Güllich, A., & Krüger, M. (Hrsg.). (2013). *Bachelor. Sport. Das Lehrbuch für das Sportstudium*. Berlin Heidelberg: Springer. Abruf unter <http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1526824>
- Gutbrod, J., & Wagner, I. (2023). Verbale Kommunikation im Sportunterricht. *sportunterricht*, 72(5), 206–211. <https://doi.org/10.30426/SU-2023-05-3>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Hay, P., & Penney, D. (2009). Proposing conditions for assessment efficacy in physical education. *European Physical Education Review*, 15(3), 389–405. <https://doi.org/10.1177/1356336X09364294>.
- Helsper, W. (2021). *Professionalität und Professionalisierung pädagogischen Handelns: Eine Einführung*. Opladen, Toronto: Barbara Budrich.
- Jürgens, E. (2005). *Leistung und Beurteilung in der Schule. Eine Einführung in Leistungs- und Bewertungsfragen aus pädagogischer Sicht* (6. Aufl.). Sankt Augustin: Academia. aktualisierte und starkerw. Aufl.
- Kastrup, V. (2009). *Der Sportlehrerberuf als Profession*. Göttingen: GBV Gemeinsamer Bibliotheksverband. Dissertation
- Killian, C. M., & Mays Woods, A. (2021). Assessment practices in K-12 physical education in the United States: a scoping review of research, 2000–2020. *Research Quarterly for Exercise and Sport*, 92(2), 248–258. <https://doi.org/10.1080/02701367.2021.1894315>.
- Kirkwood, A., & Price, L. (2008). Assessment and student learning: a fundamental relationship and the role of information and communication technologies. *Open Learning: The Journal of Open, Distance and e-Learning*, 23(1), 5–16. <https://doi.org/10.1080/02680510701815160>.
- Kleiner, K. N. (2019). Leistung – oder: Wenn zwei das Gleiche tun, ist es noch lange nicht dasselbe. *bewegung und sport*, 73(3), 3–9.
- Kurz, D., & Schulz, N. (Hrsg.). (2010). *Sport im Abitur. Ein Schulfach auf dem Prüfstand* (1. Aufl.). Aachen: Meyer & Meyer. <https://doi.org/10.5771/9783840305955>.
- Lavay, B., Sakai, J., Ortiz, C., & Roth, K. (2015). Tablet technology to monitor physical education IEP goals and benchmarks. *Journal of Physical Education, Recreation & Dance*, 86(6), 16–23. <https://doi.org/10.1080/07303084.2015.1053633>.
- Mayring, P., & Fenzl, T. (2014). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 543–556). Wiesbaden: Springer VS.
- Meier, S. (2023). Leisten, Leistung und Leistungsbeurteilung – Auslegungen in pädagogischem Interesse. In S. Ruin & G. Stibbe (Hrsg.), *Sportdidaktik und Schulsport. Zentrale Themen einer diversitätssensiblen Fachdidaktik* (S. 171–186). Schorndorf: Hofmann.
- Meier, S., & Ruin, S. (Hrsg.). (2015). *Inklusion als Herausforderung, Aufgabe und Chance für den Schulsport* (2. Aufl.). Schulsportforschung, Bd. 6. Berlin: Logos. Abruf unter <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5247185>
- Mey, G., & Mruck, K. (2010). *Handbuch qualitative Forschung in der Psychologie*. Wiesbaden: VS.
- Miceli, N. (2021). Beobachtungs- und Beurteilungsfehler im Unterrichtsalltag. <https://www.fachportal-hochbegabung.de/oid/10056/#article-3>
- Miceli, N., & Fiebeler, A. (Hrsg.). (2020). *Schulische Begabtenförderung – Potenziale erkennen und fördern*. Stuttgart: Dr. Josef Raabe. Sonderheft
- Mödinger, M., Woll, A., & Wagner, I. (2022). Video-based visual feedback to enhance motor learning in physical education—a systematic review. *German Journal of Exercise and Sport Research*, 52(3), 447–460. <https://doi.org/10.1007/s12662-021-00782-y>.
- Moskal, B. M. (2000). Scoring rubrics: what, when and how? *Practical Assessment, Research, and Evaluation*, 7(1), 3. <https://doi.org/10.7275/A5VQ-7Q66>. Vorab-Onlinepublikation.
- Paradies, L., Wester, F., & Greving, J. (2009). *Leistungsmessung und -bewertung* (3. Aufl.). Berlin: Cornelsen.
- Peiffer, L. (1998). Historische Aspekte zur aktuellen Diskussion über Bewerten und Zensieren im Sportunterricht. In N. Gissel (Hrsg.), *Sportliche Leistung im Wandel. Jahrestagung der Dvs-Sektion Sportgeschichte vom 22.–24.9.1997 in Bayreuth*. Schriften der Deutschen Vereinigung für Sportwissenschaft, (Bd. 94, S. 111–124). Hamburg: Czwalina.

- Prohl, R. (2010). *Grundriss der Sportpädagogik*. Bd. 3. Wiebelsheim: Limpert. korrigierte Auflage
- Rheinberg, F. (2006). Bezugsnorm-Orientierung. In H.-K. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 643–648). Bad Heilbrunn: Klinkhardt. Abruf unter [https://www.researchgate.net/publication/316128414\\_Bezugsnorm-Orientierung](https://www.researchgate.net/publication/316128414_Bezugsnorm-Orientierung).
- Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der pädagogischen Psychologie* (S. 178–186). Göttingen: Hogrefe.
- Rheinberg, F. (2014). Bezugsnormen und schulische Leistungsbeurteilung. In F.E. Weinert (Hrsg.), *Ciando library. Leistungsmessungen in Schulen* (3. Aufl. S. 59–72). Weinheim: Beltz.
- von Saldern, M. (2011). *Schulleistung 2.0. Von der Note zum Kompetenzraster*. Halberstedt: Books on Demand.
- Scheid, V., & Prohl, R. (Hrsg.). (2022). *Sportdidaktik. Grundlagen – Vermittlungsformen – Bewegungsfelder*. Bd. 3. Wiebelsheim: Limpert. durchgesehene und korrigierte Auflage
- Schrader, F.-W., & Helmke, A. (2014). Alltägliche Leistungsbeurteilung durch Lehrer. In F.E. Weinert (Hrsg.), *Ciando library. Leistungsmessungen in Schulen* (3. Aufl. S. 45–58). Weinheim: Beltz.
- Siedentop, D., Hastie, P.A., & van der Mars, H. (2011). *Complete guide to sport education* (2. Aufl.). Champaign: Human Kinetics.
- Sülz, C. (2014). Bezugsnormen in der Leistungsmotivation. [https://epub.ub.uni-greifswald.de/frontdoor/deliver/index/docId/1549/file/Diss\\_Suelz\\_Christoph.pdf](https://epub.ub.uni-greifswald.de/frontdoor/deliver/index/docId/1549/file/Diss_Suelz_Christoph.pdf)
- Svennberg, L., Meckbach, J., & Redelius, K. (2014). Exploring PE teachers' 'gut feelings': An attempt to verbalise and discuss teachers' internalised grading criteria. *European Physical Education Review*, 20(2), 199–214. <https://doi.org/10.1177/1356336X13517437>.
- Tomolyius, T., & Surkamin, Y. (2017). Performance assessment for physical education. In *2nd international conference on sports science, health and physical education* (Bd. 1, S. 59–64).
- Wagner, I., Bartsch, F., & Rulofs, B. (2021). Unterschiede zwischen Schüler\*innen im Sportunterricht in der Wahrnehmung von Lehrkräften. *German Journal of Exercise and Sport Research*, 51(3), 277–289. <https://doi.org/10.1007/s12662-021-00716-8>.
- Wegis, H. M. (2008). *Pre-service physical education teachers' attitudes toward, and use of handheld technology*
- Weinert, F.E. (Hrsg.). (2014). *Ciando library* (3. Aufl.). Weinheim: Beltz. Leistungsmessungen in Schulen
- Wiggins, G.P. (1998). *Educative assessment. Designing assessments to inform and improve student performance* (1. Aufl.). San Francisco: Jossey-Bass. Abruf unter <http://www.loc.gov/catdir/bios/wiley043/97049935.html>
- Wolters, P., Ehni, H., Kretschmer, J., Scherler, K., & Weichert, W. (Hrsg.). (2000). *Didaktik des Schulsports*. Schorndorf: Hofmann.
- Zaborowski, K.U., Meier, M., & Breidenstein, G. (2011). *Leistungsbewertung und Unterricht. Ethnographische Studien zur Bewertungspraxis in Gymnasium und Sekundarschule*. Wiesbaden: VS.

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