

DIAGNOSEKOMPETENZ VON SPORLEHRKRÄFTEN – SEMIFORMELLE DIAGNOSE VON MOTORISCHEN BASISKOMPETENZEN

von Benjamin Niederkofler, Christian Herrmann und Günter Amesberger

ZUSAMMENFASSUNG | Die Diagnosekompetenz von Sportlehrkräften ist ein bisher kaum beachtetes Forschungsfeld. Dies ist erstaunlich, da sich hochwertige Diagnosen als wichtige Grundlage für eine adaptive Gestaltung des Sportunterrichts zeigen und sie damit potenziell zur Lernwirksamkeit beitragen können. Diese Studie untersucht daher die Ausprägung der drei Diagnosekomponenten Niveau, Differenzierung und Rangordnung nach Schrader (1989). Dazu wurden N = 15 Sportlehrkräfte gebeten, die motorischen Basiskompetenzen von N = 250 Kindern der 5. und 6. Schulstufe zu beurteilen. Die motorischen Basiskompetenzen der Kinder wurden mit dem validen MOBAK-5-Testinstrument (Herrmann & Seelig, 2017b) erfasst. Der empirische Abgleich der Lehrkräfteeinschätzung mit den motorischen Basiskompetenzen zeigte eine mittlere Urteilsakkuratheit, eine Überschätzung des absoluten motorischen Niveaus sowie eine Unterschätzung der Merkmalsstreuung der Kinder. Dies zeigt, dass Sportlehrkräfte selbst motorische Basiskompetenzen, die Lernende mehrfach in jeder Sportunterrichtseinheit in unterschiedlichen Kontexten zeigen, nur bedingt einschätzen können.

Schlüsselwörter: Diagnosekompetenz, Sportunterricht, motorische Basiskompetenzen, Sportlehrkräfte

DIAGNOSTIC COMPETENCE OF PHYSICAL EDUCATION TEACHERS – SEMIFORMAL DIAGNOSIS OF BASIC MOTOR COMPETENCES

ABSTRACT | There are rare investigations of diagnostic competence of physical education teachers. This is surprising due to the theoretical importance of diagnoses as a base for adaptive teaching and the consequently potential for learning effectiveness. Therefore, this study investigates three diagnosis components level, differentiation and rank (Schrader, 1989). Thus, N = 15 physical education teachers assessed the basic motor competences of N = 250 pupils from eight classes of 5th and 6th grade. Basic motor competences were tested with the valid MOBAK-5 test battery (Herrmann & Seelig, 2017b). Results showed an average accuracy, an overestimation of the level and an underestimation of the differentiation in teachers' judgments. These results show difficulties in teachers' judgement of basic motor competencies, although pupils show them constantly during every physical education class.

Key Words: Diagnostic competence, physical education, basic motor competencies

DIAGNOSEKOMPETENZ VON SPORLEHRKRÄFTEN – SEMIFORMELLE DIAGNOSE VON MOTORISCHEN BASISKOMPETENZEN

1 | EINLEITUNG

Diagnostische Kompetenz gilt als zentraler Aspekt der Lehrkraftkompetenz (Baumert & Kunter, 2013; Popham, 2004; Schrader, 2006). Als Basisaufgaben von Diagnostik gelten die Spezifikation des Förderbedarfs und die Verknüpfung mit pädagogisch-didaktischen Maßnahmen (Klauer, 1982). Im Unterricht sind Diagnosen zumeist in einem Prozess verhaftet, in dem das diagnostische Urteil eine bedeutsame Rolle einnimmt (Südkamp & Praetorius, 2017). Akkurate Urteile helfen der Lehrkraft im Unterrichtsprozess, Unterrichtshandlungen optimal an die Voraussetzungen von Lernenden anzupassen (Schrader, 2006; Spinath, 2005). Weiter zeigen empirische Daten, dass sich diagnostische Kompetenz fachspezifisch äußert (Anders, Kunter, Brunner, Krauss, & Baumert, 2010; Karst, 2012; Karst, Klug, & Ufer, 2017; Lorenz & Artelt, 2009; Spinath, 2005). Dementsprechend müssen Untersuchungen dazu im fachspezifischen Rahmen angelegt und diskutiert werden.

Im Sportunterricht erscheint Diagnostik im Kontext von motorischen Basiskompetenzen von Kindern (Herrmann, Gerlach, & Seelig, 2016) besonders relevant. Motorische Basiskompetenzen gelten nicht nur als Voraussetzung für die aktive Teilnahme am Sportunterricht sowie für die Teilhabe an der außerschulischen Bewegungs-, Spiel- und Sportkultur (Herrmann et al., 2016; Kurz, Fritz, & Tschepel, 2008), sondern sind auch in den schulischen Lehrplänen als zentrales Lernziel verankert (z. B. Ministerium für Bildung und Kultur Saarland, 2011). Motorische Basiskompetenzen weisen im Sportunterricht somit eine hohe curriculare Bedeutsamkeit auf und können einen unmittelbaren Beitrag zur Überprüfung der Lern- und Leistungsentwicklung leisten. Trotz der hohen Bedeutsamkeit von Diagnosen im Bereich von motorischen Basiskompetenzen ist die empirische Befundlage gering. An dieser Stelle setzt der vorliegende Beitrag an, indem er versucht, die diagnostische Kompetenz von Sportlehrkräften und die interindividuellen Unterschiede zu erfassen.

2 | DIAGNOSTISCHE KOMPETENZ VON LEHRKRÄFTEN

„In der gegenwärtigen Diskussion über Lehrerbildung und Lehrerfortbildung ist kaum eine Komponente des professionellen Wissens und Könnens so prominent wie die ‚diagnostische Kompetenz‘“ (Baumert & Kunter, 2013, S. 300). Diagnostische Kompetenz bezeichnet allgemein die Fähigkeit der Lehrkraft, Merkmale von Personen korrekt einzuschätzen (Schrader, 2006). Häufig müssen Fachlehrkräfte in Diagnosen allerdings auch Einschätzungen zur Aufgabenschwierigkeit einbeziehen (Karing, Matthäi, & Artelt, 2011; Südkamp, Kaiser, & Möller, 2012).

Es wird davon ausgegangen, dass das diagnostische Urteil essenziell für das adaptive Handeln der Lehrkraft im Unterricht ist (Ingenkamp & Lissmann, 2008; vgl. auch Beck et al., 2008; Brühwiler, 2014; Klug, Bruder, Kelava, Spiel & Schmitz, 2013). In diesem Beitrag wird von einem Diagnoseprozess im unterrichtlichen Kontext ausgegangen, den man sich wie folgt vorstellen kann. Lehrkräfte

treffen mit einer gewissen diagnostischen Fähigkeit auf eine Diagnosesituation. Neben anderen psychosozialen Situationen bilden Bewegungshandlungen der Lernenden zentrale Diagnosesituationen im Sportunterricht. Aus der professionellen Wahrnehmung der Diagnosesituation durch die Lehrkraft ergibt sich eine Diagnose, die dann als Wissen (über entsprechende Kompetenzen der Lernenden) aus Diagnosen bei Lehrkräften repräsentiert ist. Daraus kann ein diagnostisches Urteil abgeleitet werden, das im Unterricht in Verbindung mit einer didaktischen Entscheidung (Adaptivität, kognitive Aktivierung) potenziell lernwirksam werden kann (siehe Abb. 1). Zusammenfassend kann diagnostische Kompetenz sowohl wissensgenerierend (indem diagnostische Fähigkeiten in Situationen eine Diagnose ermöglichen) als auch bedingend (indem die Diagnose für eine didaktische Handlung nutzbar wird) beschrieben werden. Abb. 1 zeigt diese Verwobenheit von wissensgenerierenden und bedingenden Aspekten der diagnostischen Kompetenz. Auch Südkamp und Praetorius (2017) legen in einer aktuellen theoretischen Weiterentwicklung eine prozessorientierte Betrachtung nahe.

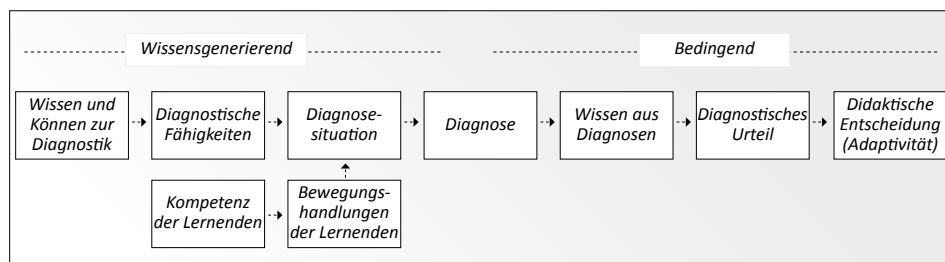


Abb. 1: Diagnostische Kompetenz als unterrichtlicher Diagnoseprozess

2.1 | DIAGNOSTISCHE FÄHIGKEITEN

Auf Basis der Forschung zur Expertiseentwicklung kann angenommen werden, dass Lehrkräfte die in der Ausbildung erworbenen *diagnostischen Fähigkeiten* mit wachsender Berufserfahrung weiterentwickeln (Berliner, 1994; Dreyfus & Dreyfus, 1987). So wird postuliert, dass sich situationspezifische und berufsbezogene Fähigkeiten und Fertigkeiten (wie etwa diagnostische Fähigkeiten) generisch und in fünf Stufen entwickeln (Berliner, 1994; Dreyfus & Dreyfus, 1987). Jede Stufe ist mit einer Fähigkeitserweiterung verbunden. Nach Berliner (1994) befinden sich angehende und berufseinsteigende Lehrkräfte in der ersten Stufe (*Novize*). Sobald Novizen umfangreiche praktische Erfahrungen gesammelt haben, wird in einer zweiten Stufe vom *fortgeschrittenen Beginner* gesprochen. In dieser Stufe lernen Lehrkräfte, ihr Regelwissen über etwa zwei bis drei Jahre in konkreten Situationen zu deuten, können aber Wesentliches noch nicht von Unwesentlichem unterscheiden. In der dritten Stufe (drei bis vier Jahre) werden Situationen im Beruf wiedererlebt, sodass demensprechend irrelevante Reize ausgeblendet werden können (*Kompetenz*). Die vierte Stufe (*Gewandtheit*) ist von automatischen Entscheidungen geprägt, da in Unterrichtssituationen intuitiv und ähnlich wie bei zurückliegenden Situationen gehandelt wird. Trotzdem ist diese Stufe von analytischen Phasen unterbrochen. Die fünfte Stufe (*Expertentum*) wird lediglich von wenigen erreicht. Durch Weiterentwicklung und intensives berufliches Engagement können Experten im

spezifischen Fachgebiet Herausforderungen lösen, ohne auf bewusste Lösungsstrategien zurückzugreifen (Berliner, 1994; Dreyfus & Dreyfus, 1987). Dies würde bedeuten, dass Lehrkräfte ab etwa 10 Jahren Berufserfahrung als *professionell* bezeichnet werden können (vgl. Baumgartner, 2016; Heinzer & Baumgartner, 2013).

Es wird weiter angenommen, dass diagnostische Fähigkeiten durch das *Wissen und Können zur Diagnostik* bestimmt sind. So sollten Lehrkräfte die Fertigkeit besitzen, verschiedene Diagnosemöglichkeiten (z. B. Beobachtung, Test, Selbsteinschätzung der Lernenden etc.) zu verwenden. Allerdings benötigt die Lehrkraft auch Wissensressourcen, die sie in eine *Diagnosesituation* mitbringt. So ist anzunehmen, dass das Wissen über Aufgaben sowie über spezifische Arten von *Wissen zur Diagnostik* wesentlich für die *diagnostische Fähigkeit* ist. Darunter fällt etwa das Wissen der Lehrkraft über die Anforderungen in einem Lerngebiet oder über die Schwierigkeiten von Aufgaben, bzw. was deren erfolgs- und entwicklungsbestimmende Voraussetzungen sind (Karst, 2012).

Bisherige Forschungsergebnisse zur Entwicklung von diagnostischen Fähigkeiten zeigen ein uneinheitliches Bild. Allgemein ist wenig zu den Bedingungen des Erwerbs diagnostischer Expertise bekannt. Van Ophuysen (2006) konnte aufzeigen, dass Lehrkräfte im Expertenstatus im Vergleich zu angehenden Lehrkräften in Schullaufbahnpfehlungen effizienter und weniger anfällig für Verzerrungen sind. Eine Untersuchung von Streit, Ruede und Weber (2015) zeigte markante Unterschiede zwischen Experten und Novizen in der Art und Weise des Diagnostizierens. So machen sich Experten in *Diagnosesituationen* deutlich häufiger Anmerkungen auf einer Metaebene. Offenbar können Experten im Unterschied zu Novizen besser einschätzen, welche Metaebene von Fähigkeiten der vorliegenden Informationen zugrunde liegt. Sie bemerken dementsprechend auch, wenn ihnen Informationen für eine präzise Aussage über Lernende fehlen (Streit et al., 2015). Praetorius, Karst, Dickhäuser und Lipowsky (2011) konnten allerdings in einer Untersuchung von 37 Grundschullehrkräften keinen Effekt der Berufserfahrung auf die diagnostische Kompetenz feststellen. Demaray und Elliot (1998) sowie Schrader (2017) schließen dementsprechend, dass bisher keine hinreichenden Belege für einen positiven Zusammenhang von Berufserfahrung und der Urteilsakkuratheit vorliegen.

2.2 | DIAGNOSESITUATIONEN

Ein Diagnostiker kann als kompetent bezeichnet werden, wenn er unterschiedliche, diagnostisch relevante Situationen erkennt sowie nutzt (Brühwiler, 2014; Südkamp & Praetorius, 2017). Im Kontext von Unterricht ergeben sich für Lehrkräfte vielfältige Diagnosesituationen. Sie sind zumeist dem pädagogischen Handeln vor- oder untergeordnet, sollen Informationen für eine anstehende pädagogische Entscheidung liefern, beinhalten Diagnosetätigkeiten, die sich auf Individuen beziehen, und befassen sich mit lernrelevanten Merkmalen als Urteilsgegenstand (Karst et al., 2017). Die konkreten Situationen unterscheiden sich anhand von mehreren Dimensionen. So konnten Karst et al. (2017) in einer Befragung von Lehrkräften folgende Klassifizierungsmerkmale von diagnostischen Situationen aufzeigen: Es können der Zweck der Diagnose (*assessment of learning, assesement for learning*), die Planbarkeit der Diagnosesituation (*langfristig, kurzfristig, nicht planbar*), die Verbind-

lichkeit und Konsequenz der Diagnose für Lernende sowie die Perspektive auf die Lernenden (*global, spezifisch*) unterschieden werden. Die Situationsanforderung an die Lehrkraft ergibt sich also aus der Einschätzung von einzelnen Lernenden im Hinblick auf lernrelevante Merkmale unter verschiedenen Rahmenbedingungen und kann mehr oder weniger gut bewältigt werden.

Karst (2012) beschreibt zudem drei idealtypische Situationen, in denen die diagnostischen Anforderungen im fachdidaktischen Rahmen bewältigt werden müssen. Sie werden durch die Perspektive in der Einschätzung (z. B. auf Aufgabe oder Person bezogen) der Lernenden sowie aus den spezifischen Arten von *Wissen zur Diagnostik* (Wissen über Aufgaben, Wissen über Lernende, Leistungsdispositionen) unterschieden. Gemeinsam ist den drei Situationen, dass bereits *Wissen aus Diagnosen* (siehe Abb. 1) über die Lernenden vorhanden sein muss. Die (1) *klassenbezogene Diagnosesituation* beschreibt die Phase, in der die Lehrkraft didaktische Entscheidungen zur Klasse trifft, z. B. wie sie in ein neues Thema einführen soll. In dieser Diagnose ist die gesamte Klasse im Fokus. Es geht also um die Auswahl von Aufgaben und Alternativaufgaben, die für möglichst alle Lernenden geeignet sind. In solchen Situationen benötigt die Lehrkraft *Wissen aus Diagnosen* über das allgemeine Niveau der Lernenden, sowie *Wissen zur Diagnostik* über verschiedene Aufgaben aus einem Lernbereich. Als (2) *schülerglobale Diagnosesituation* bezeichnet Karst (2012) modellhafte Situationen, in denen einzelne Lernende im Fokus des *diagnostischen Urteils* stehen. Dies wäre z. B. für Übungs- und Vertiefungsphasen notwendig, um binnendifferenzierte Maßnahmen zu ergreifen. In solchen Situationen ist das repräsentierte fachdidaktische *Wissen zur Diagnostik* über Aufgaben bzw. über Lernbereiche und das *Wissen aus Diagnosen* über einzelne Lernende leistungsbestimmend. Und schließlich werden (3) *schülerspezifische Situationen* skizziert, in denen einzelne Lernende spezifisch beurteilt werden. Individualdiagnostik ist wichtig für angemessene didaktische Maßnahmen zur inneren Differenzierung und individuellen Förderung. Dabei muss die Lehrkraft bereits über die Ausprägung der Merkmale informiert sein, um die Lernvoraussetzungen Einzelner bei bestimmten Lerninhalten fokussieren zu können.

Im Unterschied zur wissenschaftlichen Diagnostik findet unterrichtliche Diagnostik in der Regel zirkulär (in mehreren unterschiedlichen Situationen) statt, aus denen die Lehrkraft ihr Urteil über eine bestimmte Kompetenz fällt. Die drei beschriebenen Situationen können Lehrkräfte also erst nach Nutzung von unterschiedlichen Diagnosesituationen zielführend (im Sinne einer potenziellen Lernwirksamkeit) bearbeiten. Somit bedingt das *Wissen aus Diagnosen* jede weitere dieser idealtypischen Diagnosesituationen und zeigt damit eine zirkuläre Komponente im Diagnoseprozess auf. Diese Zirkularität wird auch im Ansatz der Adaptivität von Brühwiler (2014) aufgegriffen.

2.3 | DIAGNOSEN

Schulpädagogische Diagnose lässt sich in *formelle, semiformelle* und *informelle Diagnosen* differenzieren (Hascher, 2008; Ingenkamp & Lissmann, 2008; Jäger & Petermann, 1999). Der Unterschied zwischen den Formen liegt im Grad des Aufwandes und der wissenschaftlichen Fundierung. *Informelle Diagnosen* bezeichnen zufällige und unsystematische subjektive Urteile. Sie haben den Charakter von „Stegreif“-Urteilen und dienen dem schnellen Handeln im Unterrichtsalltag. *Semi-*

formelle diagnostische Tätigkeiten beschreiben ein Vorgehen im Rahmen des Unterrichts, das sich fokussiert um ein diagnostisches Urteil bemüht, allerdings den Qualitätskriterien einer formellen Diagnose nicht entspricht. Die formelle Diagnose beschreibt das Treffen inter- und intraindividuell Unterscheidungen auf der Basis testtheoretisch abgesicherter wissenschaftlicher Verfahren. Die Diagnoseformen resultieren dementsprechend in *Wissen aus Diagnosen* (siehe Abb. 1)

2.4 | WISSEN AUS DIAGNOSEN

Je nach Diagnoseform baut die Lehrkraft *Wissen aus Diagnosen* mit unterschiedlicher Güte auf. Dieses Wissen kann sich auf verschiedene Merkmale beziehen, die im Fokus der *Diagnosen* liegen. Karst (2012) beschreibt das individuelle *Wissen aus Diagnosen* als Teil des fachdidaktischen Wissens und führt dazu drei Wissensarten ein: *Niveauwissen*, *Heterogenitätswissen* und *Profilwissen*. Das *Niveauwissen* erfasst, wie genau Lehrkräfte die Ausprägungen der Leistungen einschätzen können. Das *Heterogenitätswissen* gibt an, ob Lehrkräfte die Unterschiedlichkeit bzw. Gleichheit der Leistungen bei Lernenden in ihrer Bandbreite wahrnehmen und die relative Position der Lernenden darin kennen. Das *Profilwissen* bezeichnet ein sehr spezifisches Wissen: „Es handelt sich dabei um die Kenntnis, die der Lehrer vom Leistungsprofil einzelner Schüler hinsichtlich unterschiedlicher Aufgaben hat“ (Karst, 2012, S. 92). Ein Beispiel aus dem Sportunterricht soll dies verdeutlichen. Fehlt einem Schüler die notwendige Stützkraft für das Schwingen am Barren, so kann er die Aufgabe „Abgang aus dem Schwingen mit Wende“ (noch) nicht bewältigen. Andererseits könnten bei diesem Schüler die koordinativen Voraussetzungen für die „Wende“ bereits vorhanden sein. Die Lehrkraft kennt damit das Profil, also die spezifische Ausprägung der Merkmale, dieses Schülers bei bestimmten Aufgaben.

Die Güte des *Wissens aus Diagnosen* hängt letztlich auch vom Wissen über Bewegungsaufgaben bzw. deren Schwierigkeit und Anforderung (z. B. Wissen zu Testmöglichkeiten, Kompensationsmechanismen, Erfolg versprechende Merkmale etc.) ab. Diese Wissensressourcen bringt die Lehrkraft als *Wissen zur Diagnostik* in die *Diagnosesituation* mit.

2.5 | DIAGNOSTISCHES URTEIL

Dem Handeln der Lehrkräfte liegen unter anderem *diagnostische Urteile* zugrunde. Entsprechend sollten diese eine möglichst hohe Güte aufweisen. Im Unterrichtsalltag laufen allerdings nicht alle Urteilsprozesse unter höchsten Güteanforderungen ab (siehe Kap. 2.3). Das führt zur Frage nach der Akkuratheit der *diagnostischen Urteile* (McElvany et al., 2009; Praetorius & Südkamp, 2017; Schrader, 1989). Sie ist ein Maß dafür, wie zutreffend die Lehrkraft bestimmte Merkmale bei Lernenden einschätzt. Solche Urteile können nach Schrader (1989) aufgabenbezogen („Wie viele Lernende werden diese Aufgabe lösen?“) oder personenbezogen („Wie viele Aufgaben werden vom Lernenden richtig gelöst?“) ausfallen und äußern sich in den drei Komponenten *Niveau*, *Differenzierung* und *Rangordnung*. Dabei entspricht die *Niveauelemente* dem *Niveauwissen*, die *Differenzierungs-* und *Rangkomponente* beziehen sich auf das *Heterogenitätswissen*. Die *Niveauelemente* beschreibt das allgemeine Leistungsniveau. Die *Differenzierungskomponente* gibt an, ob Lehrkräfte die Leistungen der Lernenden in ihrer Bandbreite wahrnehmen und ob sie

das Leistungsspektrum unter- bzw. überschätzen (Karing & Artelt, 2013). Die *Rangordnungskomponente* beschreibt die Kompetenz, die Rangordnung zwischen verschiedenen Lernenden für ein bestimmtes Merkmal genau wiederzugeben (Schrader & Helmke, 1987).

Untersuchungen zu *diagnostischen Urteilen* zeigen eine große Spannweite in der Akkuratheit (Südkamp et al., 2012). Im Mittel beobachtet die Metaanalyse von Südkamp et al. (2012) eine Korrelation von $p = .63$ für die Urteilsakkuratheit zwischen Lehrkräfteeinschätzung und der Leistung der Lernenden. Hoge und Coladarsi (1989) berechnen in einem früheren Review einen Korrelationswert von $p = .66$. Die Ergebnisse zum *Niveau* der Einschätzung zeigen ein uneinheitliches Bild. So beschreiben Begeny, Eckert, Montarello und Storie (2008) in der Untersuchung von Einschätzungen zu Leseleistungen eine Überschätzung. Feinberg und Shapiro (2009) konnten zum selben Merkmal allerdings eine Unterschätzung nachweisen. McElvany et al. (2009) folgerten für die Einschätzung des Umgangs mit einer Text-Bild-Integration, dass Lehrkräfte eine Unterschätzungstendenz aufweisen.

2.6 | ADAPTIVITÄT UND DIDAKTISCHE ENTSCHEIDUNG

Das Verständnis von diagnostischer Kompetenz als Produkt lässt einige Fragen offen. So wird von Schrader (2009) darauf verwiesen, dass die Forschung zur Urteilsakkuratheit und die prozessorientierte Forschung zumeist unverbunden bleiben. Auch wird die didaktische Relevanz von akkuraten Urteilen infrage gestellt (Van Ophuysen & Lintorf, 2014). Erste empirische Ergebnisse von Brühwiler (2014) konnten in diesem Zusammenhang zeigen, dass sich lediglich die Kopplung des *diagnostischen Urteils* mit einer didaktischen Unterrichtsmaßnahme als lernwirksam erweist. Auch für den Sportunterricht legen theoretische Überlegungen nahe, dass Diagnosen als Grundlage für kognitive Aktivierung genutzt werden können (Niederkofler & Amesberger, 2016). Erst damit kann eine potenziell lernwirksame kognitive Aktivität im Sportunterricht ausgelöst werden. Allerdings werden Details zur Verknüpfung von Diagnosen und adaptiven Unterrichtshandlungen hier nicht weiter verfolgt, denn der Beitrag konzentriert sich im empirischen Teil auf eine enge Definition von diagnostischer Kompetenz, die sich auf die akkurate Einschätzung von Merkmalen der Lernenden bezieht.

2.7 | MODERATOREN DER URTEILSAKKURATHEIT

Bereits Kelly (1955) verdeutlichte, dass Menschen konstruierende Wesen sind. Entsprechend sind auch Diagnosen Konstruktionen von Lehrkräften in einer je bestimmten Situation, die von der getesteten und lernenden Person je unterschiedlich konstruiert werden. Der Testwert ist damit nicht die Abbildung einer Fähigkeit, sondern eine soziale Rekonstruktion, die, wie oben dargestellt, in einem zirkulären Prozess entsteht.

Diagnostik meint immer einen Prozess, der unter Einbezug von vorhandenen Ressourcen Informationen erfasst und verarbeitet, ein Urteil ermöglicht und eine didaktische Entscheidung zur Folge haben kann (siehe Abb. 1). In diesem Prozess sind Moderatoren zu berücksichtigen, die das Urteil beeinflussen bzw. verändern. Nach Südkamp et al. (2012) können dies (1) Merkmale der Lehrkraft (z. B. Berufserfahrung) und (2) Merkmale des Urteils (z. B. Art der Skalierung im Rating) sein. Die

zu diagnostizierenden Schülerleistungen/Performanzen sind wiederum abhängig von (3) Schülermerkmalen (z. B. Motivation) und (4) Testmerkmalen (z. B. Länge des Tests). Alle vier Bereiche, so das Modell von Südkamp et al. (2012), können sich auf die Akkuratheit der Urteile auswirken. Nach Modellen der sozialen Kognition (Fiske, 1993) und der selektiven Zugänglichkeit (Mussweiler, 1997) kann das *Urteil* durch die Expertise der Lehrkraft¹ und das Verarbeitungsziel beeinflusst werden. Verarbeitungsziele bedingen in diesem Kontext, ob die Verarbeitung von diagnoserelevanten Informationen merkmals- oder kategoriegeleitet verläuft (Fiske, 1993; Krolak-Schwerdt, Böhmer, & Gräsel, 2009). Erfordert das Ziel eine hohe Genauigkeit, begünstigt dies die merkmalsgeleitete Informationsverarbeitung und präzisiert damit das Urteil. Verlangt das Ziel eine schnelle und effiziente Entscheidung, führt dies eher zu einer heuristischen Informationsverarbeitung und wird damit anfällig für sogenannte *Ankereffekte*² (Dünnebier, Gräsel, & Krolak-Schwerdt, 2009).

3 | DIAGNOSEN IM SPORTUNTERRICHT

3.1 | DIAGNOSEN VON MOTORISCHEN BASISKOMPETENZEN

Das Lernen im Sportunterricht beruht auf Bewegungs-, Lern- und Kontrollaufgaben (Frankhauser et al., 2014). Erstere dienen zur Entwicklung von Kompetenzen, Letztere zur Überprüfung des Kompetenzerwerbs. In diesem Kontext kann der MOBAK-Ansatz als Kontrollaufgabe verortet werden (MOBAK – MOtorische BASiskompetenzen, Herrmann & Gerlach, 2014; Herrmann et al., 2016). Er beschreibt motorische Basiskompetenzen als notwendige, jedoch nicht hinreichende Voraussetzung für eine aktive Teilnahme an der Bewegungs-, Spiel- und Sportkultur (Gogoll, 2012). Motorische Basiskompetenzen ermöglichen den Zugang zu weiteren Bildungsgegenständen im Bereich Bewegung, Spiel und Sport (Kurz et al., 2008) und bilden die Grundlage für die Entwicklung eines körperlich-aktiven Lebensstils (Lubans, Morgan, Cliff, Barnett & Okely, 2010; Stodden & Holfelder, 2013). Sie werden als nachhaltig erlernbare und latente Leistungsdispositionen verstanden und entwickeln sich aus aufgabenspezifischen Bewegungs- und Lernaufgaben (z. B. ein Ziel treffen).

Da die motorischen Basiskompetenzen nicht direkt beobachtet und erfasst werden können, erfolgt die Erfassung (indirekt) über die beobachtbaren motorischen Basisqualifikationen (MOBAQ; z. B. Werfen, Fangen). Diese werden mit entsprechenden Aufgaben getestet, um das Niveau der motorischen Basiskompetenzen abzuschätzen. Die theoretische Struktur von motorischen Basisqualifikationen und von motorischen Basiskompetenzen konnte für die Kompetenzbereiche (KB) *Sich-Bewegen* und *Etwas-Bewegen* in der ersten, dritten und fünften Schulstufe empirisch bestätigt werden (Herrmann, Gerlach; & Seelig, 2015; Herrmann & Seelig, 2017a; 2017d).

Den motorischen Basiskompetenzen kommt die Steuerungsfunktion der motorischen Fähigkeiten (Wie viel Kraft hat das Kind?) und Fertigkeiten (Ist die Wurftechnik des Kindes ausreichend?) bei

1 Dies wurde bereits weiter oben diskutiert.

2 Die Ankerheuristik wird als eine Verzerrung eines Urteils in Richtung eines vorher gesetzten Wertes definiert, des Ankers (Strack & Mussweiler, 1997).

konkreten motorischen Aufgaben (Trifft das Kind mit einem Ball die Zielscheibe?) zu. Das Konstrukt der motorischen Basiskompetenzen kann aus theoretischer Perspektive als Ergänzung zu den bislang in den Sportwissenschaften vorherrschenden Konstrukten der motorischen Fähigkeiten und Fertigkeiten verstanden werden, indem die Unterscheidung zwischen Fähigkeiten und Fertigkeiten aufgelöst wird und sich der Fokus ausschließlich auf die erfolgreiche Bewältigung motorischer Anforderungen bezieht.

Die beobachtbaren Performanzen (im Sinne der motorischen Basisqualifikationen; vgl. Herrmann, Gerlach & Seelig, 2016) der motorischen Basiskompetenzen (z. B. rollen, dribbeln) haben in der Primarstufe eine große Bedeutung. Sie sind in deutschsprachigen Lehrplänen als zentrale Lernziele verankert (z. B. Ministerium für Bildung und Kultur Saarland, 2011). Diagnosen von motorischen Basiskompetenzen weisen dementsprechend eine hohe curriculare Bedeutsamkeit (Karst, 2012; Madelaine & Wheldall, 2005) auf. Sie können einen unmittelbaren Beitrag zur Überprüfung der Lern- und Leistungsentwicklung der Schülerinnen und Schüler leisten. Mit ansteigender Schulstufe können motorische Basiskompetenzen jedoch als Voraussetzung für weitere Bewegungs- und Handlungsmöglichkeiten gelten. So kann das Erreichen eines motorischen Basisniveaus den Aufbau von weiteren Kompetenzen (z. B. Risiko und Wagnis in Relation zu den eigenen Fähigkeiten abzuschätzen) und das Lösen von anspruchsvolleren Bewegungsaufgaben erleichtern. Der MOBAK-Ansatz kann daher im erziehenden Sportunterricht (Prohl, 2012) verortet werden.

3.2 | DIAGNOSTISCHE KOMPETENZ VON SPORTEHRKRÄFTEN

Über die Diagnosekompetenz von Sportlehrkräften ist noch wenig bekannt (Seyda, 2016). Zwar sind normative Überlegungen zur Lernstanddiagnostik im Sportunterricht vorhanden (z. B. Oefener, Erlemeyer & Staack (o. J.), allerdings sind Studien dazu rar. Eine Studie von Seyda (2016) untersuchte die Komponenten *Niveau*, *Differenzierung* und *Rangordnung* bei 35 Sportlehrkräften in der Diagnose von physischer Fähigkeitsselbstwahrnehmung³. Die Ergebnisse zeigen im Mittel eine Unterschätzung des Niveaus der Fähigkeitsselbstwahrnehmung. Die Differenzierung zeigte eine Überschätzung der Streuung. Die Rangkomponente können Sportlehrkräfte, laut Studie, weniger gut einschätzen. Einzelergebnisse zu den Diagnosekomponenten liegen in der Studie allerdings nicht vor. Insgesamt zeigten sich bei allen Sportlehrkräften Schwierigkeiten in der Passung zwischen der Diagnose der Lernvoraussetzung und dem adaptiven Unterrichtsvorgehen.

Weitere Studien aus dem Sportunterricht liegen den Autoren nicht vor. Eine Untersuchung von Huemer und Amesberger (2016) in der Sekundarstufe I ließ Sportlehrkräfte im Zuge der Messung von fachspezifischen Kompetenzen motorische Fertigkeiten von Schülerinnen und Schülern im Gerätturnen einschätzen. Ein korrelativer Abgleich der Einschätzung der Lehrkraft mit den Selbsteinschätzungen der Schülerinnen und Schüler ergab eine breite Streuung der Korre-

3 Diese Lernvoraussetzung gilt neben der Lernstanddiagnostik als aufschlussreicher Faktor bei schulischen Diagnosen (Spinath, 2005)

lationswerte (Hüftaufschwung $r = .75$; Unterschwung $r = .14$; Rolle $r = .34$; Rad $r = .59$). Zudem zeigte die Untersuchung eine Stabilität der Einschätzung der Lehrkräfte zwischen Test und Retest von $r_{tt} = .84$. Weitere Ergebnisse zur Studie, z. B. zum Abgleich der Einschätzung der Lehrkraft mit einer formellen Diagnose der motorischen Fertigkeiten, liegen nicht vor. Aktuell wird an der Universität Münster unter der Leitung von Seyda das DFG-geförderte Forschungsprojekt „Diagnostische Kompetenz von Sportlehrkräften (DiKS) – Beurteilungsgüte der doppelten Diagnoseleistung zur Unterstützung realistischer physischer Fähigkeitswahrnehmungen von Grundschulkindern im Sportunterricht“ durchgeführt⁴. Dies unterstreicht die Bedeutung der Diagnosekompetenz bei Sportlehrkräften.

Zusammenfassend kann gefolgert werden, dass Diagnosen von motorischen Basiskompetenzen im Sportunterricht eine hohe pädagogische und didaktische Relevanz besitzen. Als Urteile über den Lern- und Leistungsstand können sich semiformelle Diagnosen in diesem Bereich entsprechend der vorhandenen diagnostischen Fähigkeiten von Sportlehrkräften in unterschiedlicher Güte auf eine didaktische Entscheidung auswirken. Aufgrund der theoretischen Überlegungen ist es erstaunlich, dass bisher kaum Untersuchungen zur Diagnosekompetenz von Sportlehrkräften über motorische Merkmale von Lernenden zu finden sind. Diese Lücke greift der vorliegende Beitrag auf und stellt folgende Fragen:

- (1) Wie akkurat schätzen Sportlehrkräfte motorische Basiskompetenzen ein?
- (2) Zeigen sich interindividuelle Unterschiede zwischen den Lehrkräften in der Diagnosekompetenz in Bezug auf die Kompetenzbereiche?
- (3) Zeigen sich Unterschiede in der Diagnosekompetenz in Abhängigkeit von der Berufserfahrung?
- (4) Zeigen sich Unterschiede in der Diagnosekompetenz in Abhängigkeit vom Schultyp?

4 | METHODE

Studien zur Diagnostik im Lehrberuf folgen zumeist einem fest etablierten methodischen Ansatz. Darin wird das diagnostische Urteil als Abgleich der Lehrkrafturteile mit den bei Schülerinnen und Schülern erhobenen Zielmerkmalen operationalisiert (Cronbach, 1955; Hoge & Coladarci, 1989; Schrader & Helmke, 1987; Seyda, 2016). Bei Merkmalen, für die es standardisierte, anerkannte Messverfahren gibt, ist dieses Untersuchungsparadigma gut anwendbar (Van Ophuysen & Behrmann, 2015). Dies trifft auch für die vorliegende Studie zu.

4.1 | MESSUNGEN

MOBAK-5-Testinstrument

Das MOBAK-5-Testinstrument (Herrmann & Seelig, 2017a) umfasst insgesamt acht Testitems. Die Testitems *werfen*, *fangen*, *prellen* und *dribbeln* werden aufsummiert sowie durch vier dividiert und bilden so die Basiskompetenz *Etwas-Bewegen* ab und die Testitems *variabel laufen*, *seilsprin-*

⁴ Eine Beschreibung des Forschungsprojekts ist unter folgendem Link zu finden: <http://gepris.dfg.de/gepris/projekt/315246907> (Zugriff am 16.04.2018).

gen, balancieren und rollen werden in gleicher Weise berechnet und operationalisieren die Basiskompetenz *Sich-Bewegen*. Beide KB zusammen ergeben die motorische Gesamtkompetenz. Alle Testitems wurden in drei Stufen skaliert (1 = nicht bestanden; 2 = bestanden; 3 = voll bestanden)⁵. MOBAK-5 gilt als valide und reliabel in der Erfassung der motorischen Basiskompetenzen *Sich-Bewegen* und *Etwas-Bewegen* in der fünften und sechsten Schulstufe (Herrmann & Seelig, 2017b; 2017d).

Lehrkraftfragebogen

Lehrkräfte wurden im Fragebogen gebeten, Angaben zu soziodemografischen Daten und ihrem Qualifikationsprofil zu geben. Zudem sollte die Lehrkraft das sportliche Können eines jeden Kindes einschätzen (siehe Tab. 1). Somit gab jede Lehrkraft für alle Kinder der Klasse eine Bewertung ab. Dabei dienten die MOBAK-5-Testitems als Referenzpunkt und Zielmerkmal der Einschätzung (Herrmann & Seelig, 2017c). Die Items (siehe Tab. 1) wurden nach einer fünfstufigen Likert-Skala (1 = überhaupt nicht; 1,5 = eher nicht; 2 = teils teils; 2,5 ziemlich sicher; 3 = ganz sicher) gestuft. Zusätzlich wurde das Antwortformat „weiß ich nicht“ angeboten. Es wurde jedoch von keiner Lehrkraft verwendet.

Tab. 1: Skalennennwerte und Stichprobengröße

		α	<i>M</i>	<i>SD</i>	<i>N</i>
Einschätzung der Kompetenzbereiche durch die Lehrkraft					
Etwas-Bewegen 4 Items	Beispielitem: „Er/sie kann einen Ball sicher fangen.“	.91	2.31	.43	237
Sich-Bewegen 4 Items	Beispielitem: „Er/sie kann eine gesprungene Rolle vorwärts turnen.“	.84	2.25	.45	237
Gesamteinschätzung 8 Items	4 Items <i>Etwas-Bewegen</i> und 4 Items <i>Sich-Bewegen</i>	.90	2.29	.40	237
Motorische Basiskompetenzen (MOBAK, Herrmann et al., 2017a-d)					
Etwas-Bewegen 4 Items	Items: <i>variabel laufen, seilspringen, balancieren, rollen</i>	.39	2.06	.45	253
Sich-Bewegen 4 Items	Items: <i>werfen, fangen, prellen, dribbeln</i>	.35	2.13	.47	252
Gesamtkompetenz 8 Items	4 Items <i>Etwas-Bewegen</i> und 4 Items <i>Sich-Bewegen</i>	.75	2.10	.38	252

4.2 | BERECHNUNG DER DIAGNOSEKOMPETENZ

Die *Niveau-*, *Differenzierungs-* und *Rangordnungskomponente* gelten als Indikatoren der Diagnosekompetenz (Schrader, 1989). Zur Beantwortung der ersten Forschungsfrage wurden die Komponenten aus den folgenden Abgleichen des Lehrerurteils (*y*) mit dem Testwert der Lernenden (*x*) mit den Softwareprogrammen *Microsoft Excel 2013* und *IBM SPSS Statistics 24* für jede Lehrkraft

5 Das Testmanual mit den Kriterien für die einzelnen Items ist unter folgendem Link zu finden: <http://www.dsbg4public.ch/custom/upload/docs/cdl64n2jbsd9ovn3o690nm7jiub7ned1272g.pdf> (Zugriff am 16.04.2018).

einzelnen berechnet und dann über alle Lehrkräfte zusammengefasst. Die Komponenten wurden für die Merkmale motorische *Gesamtkompetenz* und die KB *Sich-Bewegen* und *Etwas-Bewegen* berechnet (siehe Kap. 4.5).

Berechnung der Niveauelemente

Die Niveauelemente entsprechen bei Schrader (1989) der Differenz des mittleren Lehrkrafturteils und der mittleren Schülerleistung ($M_v - M_x$). Werte größer 0 kennzeichnen eine Überschätzung des allgemeinen Leistungsniveaus der Klasse, wohingegen Werte kleiner 0 eine Unterschätzung anzeigen. Als weiteres Maß gibt Schrader (1989) den Niveaufehler an; dieser Wert entspricht dem Betrag der Niveauelemente (Karst, 2012).

Berechnung der Differenzierungskomponenten

Nach Schrader (1989) gibt die Differenzierungskomponente das Verhältnis zwischen der Streuung der Einschätzung der Lehrkraft (s_v) und der Streuung der Schülerleistungen (s_x) an (σ_v/σ_x). Werte größer 1 kennzeichnen eine Überschätzung der leistungsbezogenen Heterogenität. Werte kleiner 1 zeigen an, dass die Lehrkraft die leistungsbezogene Heterogenität unterschätzt. Auch hier beschreibt Schrader (1989) ein weiteres Maß – den Differenzierungsfehler. Er ermöglicht gerichtete Aussagen über die Diagnosekompetenz und ergibt sich aus dem Betrag der Differenz der Streuung (vgl. Karst, 2012).

Berechnung der Rangkomponente

Die Rangordnungskomponente wird bei Schrader (1989) über die Produkt-Moment-Korrelation (r_{xy}) zwischen der Einschätzung der Lehrkraft und der gemessenen Schülerleistung operationalisiert. Sie entspricht also der relativen Urteilsakkuratheit der Lehrkraft. Zur Zusammenfassung der Einzelkorrelationen wurden diese zuerst in Fischers Z-Werte transformiert. Aufgrund der ungleich großen Anzahl an Schülerdiagnosen pro Lehrkraft wurden die Fischers Z-Werte nach der Summenformel (Bortz, 2005, S. 219) zu einem Z-Wert aufsummiert und anschließend rücktransformiert. Die Rangordnungskomponente bildet das Kernstück der Diagnosekompetenz ab und beschreibt die Kompetenz, die Rangordnung unter verschiedenen Lernenden in Bezug auf ein bestimmtes Merkmal (z. B. Leistung) wiederzugeben (Schrader & Helmke, 1987). Theoretisch können die Werte zwischen -1 und +1 variieren. Je höher die Übereinstimmung ist, also je näher der Wert bei +1 liegt, desto genauer fällt die Diagnose der Lehrkraft aus (Karing & Artelt, 2013; Karst, 2012). Praetorius et al. (2011) führen weitere normative Überlegungen ein. Sie gehen davon aus, dass Korrelationen von Variablen zum selben Merkmal grundsätzlich einen höheren Wert aufweisen sollten. Eine akkurate Beurteilung liegt, nach der Autorengruppe, also erst dann vor, wenn $r_{xy} > .55$ ausfällt.

4.3 | WEITERE BERECHNUNGEN

Zur Beantwortung der zweiten Forschungsfrage wurde der Intraclass-Correlation-Coefficient (ICC) als Modell mit Ein-Weg-Zufallseffekten im *Softwareprogramm IBM SPSS Statistics 24* berechnet. Für die Rangordnungskomponenten wurden die einzelnen Korrelationswerte (aus oben beschriebenen Gründen) vorab in Z-Werte transformiert.

Zur Beantwortung der dritten und vierten Forschungsfrage wurden *t*-Tests bei unabhängigen Stichproben im Softwareprogramm *IBM SPSS Statistics 24* und die Effektstärke (Cohen's *d*) online⁶ berechnet. Auch hierfür wurden die Rangordnungskomponenten (aus oben beschriebenen Gründen) vorab in *Z*-Werte transformiert und für die Ergebnistabelle nicht rücktransformiert. Für die dritte Forschungsfrage wurde die Gesamtstichprobe in die Erfahrungsklassen *< 10 Jahre* und *> 10 Jahre* Berufserfahrung unterteilt. Für die vierte Forschungsfrage wurde die Stichprobe in die Schultypen *neue Mittelschule* und *Gymnasium* unterteilt. Diese Unterteilung ist dadurch zu rechtfertigen, da die Ausbildungen für das Fach Bewegung und Sport bisher jeweils an unterschiedlichen Institutionen verortet waren⁷.

4.4 | DESIGN UND ABLAUF DER STUDIE

Im querschnittlichen Studiendesign wurden Lehrkräfte an österreichischen Sekundarstufenschulen am Ende des Schuljahrs gebeten, die motorischen Basiskompetenzen der Lernenden einzuschätzen. Im Anschluss an die Fragebogenerhebung durchliefen alle Lernenden das MOBAK-5-Testinstrument. Auch Eltern, Lehrpersonen, Lernende sowie Landesschulrat und Schulleitung wurden schriftlich und mündlich über den Zweck der Untersuchung informiert. Zudem wurde bei den Eltern eine Teilnahmeerlaubnis eingeholt. Die Lehrkräfte wurden in das MOBAK-5-Testinstrument eingeführt und beantworteten einen Fragebogen zur Einschätzung der motorischen Kompetenz der Lernenden. Dies erfolgte abseits der Testung. Sie konnten die motorische Performanz der Lernenden also nicht sehen.

4.5 | STICHPROBE

An der Erhebung nahmen 15 Lehrkräfte im Alter von $M = 44.2$ Jahren ($SD = 12.9$) teil, 55 % der Lehrkräfte waren weiblich. Alle Lehrkräfte unterrichteten das Fach Bewegung und Sport (Tätigkeitskriterium) und verfügten über eine abgeschlossene und anerkannte Ausbildung zur Sportlehrkraft (Professionskriterium). Diese Kriterien gewährleisteten nach Baumgartner (2013) das „Gefühl des Involviert-Seins“ und das Durchlaufen eines Professionalisierungsprozesses. Es nahmen zwanzig Klassen der fünften und sechsten Schulstufe mit $N = 252$ Kinder (48 % weiblich) im Alter von $M = 11.9$ Jahre ($SD = 0.7$) teil. Die Freiwilligkeit der Lehrkräfte zur Teilnahme an der Befragung führte dazu, dass die Einschätzung von einer männlichen Lehrkraft für 15 Lernende nicht erhoben werden konnte. Schulabsenzen am Testtag und die Freiwilligkeit den Lernenden an der Studie führten am Testtag zu einer Drop-out-Rate von etwa 20 %.

5 | ERGEBNISSE

Wie akkurat schätzen Sportlehrkräfte motorischen Basiskompetenzen ein?

Tab. 3 gibt einen Überblick über die Maße der Diagnosekomponenten von Sportlehrkräften. Dargestellt werden das Urteil über die *Gesamtkompetenz* von Schülerinnen und Schülern und die KB

6 www.psychometrica.de/effektstaerke.html

7 So wurden österreichische Sportlehrkräfte für das Gymnasium bisher an den Universitäten ausgebildet und jene für die neue Mittelschule an Pädagogischen Hochschulen. Die Unterscheidung ist somit ein weiterer Faktor der Berufserfahrung.

Etwas-Bewegen und *Sich-Bewegen*. Die Niveauelemente zeigt im Allgemeinen eine Überschätzung der motorischen Basiskompetenzen (.16; .17; .15). Betrachtet man den Niveaufehler, so fällt dieser im KB *Etwas-Bewegen* am geringsten aus (.15). In Bezug auf die Differenzierungskomponente lässt sich feststellen, dass Sportlehrkräfte im Mittel die Streuung der Merkmalsausprägung ihrer Schülerinnen und Schüler unterschätzen. Der absolute Differenzierungsfehler zeigt, dass dieser im Gesamturteil am geringsten ausfällt (-.03). Die Ausprägung der Rangordnungskomponente weist im Allgemeinen eine mittlere Urteilsakkuratheit auf ($r = .50$ für die *Gesamtkompetenz*, $r = .45$ für den KB *Sich-Bewegen*). Die Akkuratheit im KB *Etwas-Bewegen* ist mit $r = .39$ tendenziell etwas tiefer.

Des Weiteren weisen die Werte der Diagnosekomponenten zwischen den Lehrkräften eine große Bandbreite auf. Alle Werte finden sich hierzu in Tab. 3. Eine deskriptive Betrachtung der Bandbreite der Niveauelemente weist in allen Einschätzungen von einer Unterschätzung bis zu einer Überschätzung. Am auffallendsten ist diese Bandbreite im KB *Etwas-Bewegen*, hier erstreckt sich der Wertebereich von -.16 bis .68. Ähnlich sind auch die intraindividuellen Unterschiede in der Rangordnungskomponente. Die Einschätzung der *Gesamtkompetenz* weist hier z. B. eine Bandbreite von -.14 bis .90 auf. Für die Differenzierungskomponente zeigt sich z. B. im KB *Sich-Bewegen* eine Bandbreite von .02 bis 1.30. Die deskriptive Betrachtung legt offen, dass es Lehrkräfte gibt, die die Kompetenzen der Lernenden äußerst genau einschätzen (z. B. *Lehrkraft 3* in der Niveauelemente der *Gesamtkompetenz* und im KB *Etwas-Bewegen* oder in der Differenzierungskomponente im KB *Sich-Bewegen*; *Lehrkraft 4* alle Rangordnungskomponenten). Es zeigen sich aber auch Lehrkräfte mit großen Schwierigkeiten in der Einschätzung von motorischen Basiskompetenzen (z. B. *Lehrkraft 11* in der Niveauelemente der *Gesamtkompetenz* und im KB *Sich-Bewegen*; *Lehrkraft 5* in den Rangordnungskomponenten).

Tab. 2: Maße der Diagnosekomponenten in der Einschätzung von motorischen Basiskompetenzen (Mittelwert aus $N = 15$ Lehrkräfte)

	<i>Gesamtkompetenz</i>	Kompetenzbereiche	
		<i>Sich-Bewegen</i>	<i>Etwas-Bewegen</i>
Niveauelemente	.16	.17	.15
<i>Niveaufehler</i>	.16	.17	.15
Differenzierungskomponente	.92	.76	.82
<i>Differenzierungsfehler</i>	-.03	-.07	-.09
Rangordnungskomponente	.50	.45	.39
Intraclass-Correlation-Coefficient (ICC)	.68	.52	.65

Anmerkungen: Die Daten beziehen sich auf $N = 227$ Kinder; ICC: einzelne Maße, Modell mit Einweg-Zufallseffekten. **Niveau:** Akkuratheit der Einschätzung des absoluten Niveaus eines Merkmals. mittleres Lehrerurteil – mittlerer Testwert. > 0 = Überschätzung der Leistung. < 0 = Unterschätzung der Leistung. Wertebereich -3 bis 3; **Differenzierung:** Über- oder Unterschätzung der Streuung eines Merkmals. SD Lehrerurteils / SD Schülermerkmal. > 1 = Überschätzung der Heterogenität;

< 1 = Unterschätzung; **Rang**: Korrekte Rangfolge der Leistungen innerhalb einer Klasse. Bivariate Produkt-Moment-Korrelation. Wertebereich -1 bis 1.

Zeigen sich interindividuell Unterschiede zwischen den Lehrkräften in der Diagnosekompetenz in Bezug auf die Kompetenzbereiche?

Der Abgleich der Diagnosekomponenten zwischen den Lehrkräften in der Einschätzung der drei Merkmale zeigt für die *Gesamteinschätzung* einen ICC von .68. Etwas stärker fällt der interindividuelle Abgleich bei den Diagnosekomponenten im KB *Sich-Bewegen* aus (ICC = .52).

Zeigen sich Unterschiede in der Diagnosekompetenz in Abhängigkeit von der Berufserfahrung und vom Schultyp?

Tabelle 4 zeigt die Ergebnisse aus den Berechnungen der Mittelwertvergleiche der Diagnosekomponenten zwischen den Lehrkräften, die an einer neuen Mittelschule (NMS) und an einem Gymnasium (GYM) unterrichten, sowie den Mittelwertvergleich zwischen den Lehrkräften, die weniger als zehn Jahre und jenen, die mehr als zehn Jahre Unterrichtserfahrung aufweisen. Es wurden die Diagnosekomponenten von den drei Schülermerkmalen motorische *Gesamtkompetenz* und den beiden KB *Sich-Bewegen* und *Etwas-Bewegen* berechnet. Somit wurden Mittelwerte aus 15 Diagnosekennwerten verglichen.

Der *t*-Test im Schultyp zeigt bei neun NMS- und sechs Gymnasial-Lehrkräften – mit einer Ausnahme – keine bedeutsamen Unterschiede. Lediglich im Differenzierungsfehler im KB *Sich-Bewegen* zeigt der *t*-Test einen signifikanten Wert ($t = -3.05$; $p = .01$). Eine Überprüfung der Effektstärken legt für den Schultyp schwache, mittlere und einen starken Effekt offen. Diese zeigen je nach Merkmal und Diagnosekomponente in eine andere Richtung. Für die motorische *Gesamtkompetenz* ergibt sich in der Differenzierung ein mittlerer positiver Effekt ($d = .50$). Die Rangordnung weist einen negativen mittleren Effekt auf ($d = -.35$). Im KB *Sich-Bewegen* zeigt sich ein ähnliches Bild. Auch hier findet sich ein mittlerer positiver Effekt für die Differenzierung ($d = .45$) jedoch ein schwacher negativer Effekt für die Rangordnung ($d = -.13$). Zudem legen die Ergebnisse beim Differenzierungsfehler einen starken Effekt offen ($d = 1.52$). Im KB *Etwas-Bewegen* zeigt sich ein mittlerer Effekt beim Differenzierungsfehler ($d = -.43$).

Die Berechnungen zur Berufserfahrung ergeben keine signifikanten Mittelwertunterschiede. Die Einzelwerte sind in Tab. 4 aufgelistet. Es zeigt sich, dass die Effekte der Berufserfahrung auf die Diagnosekomponenten im schwachen, mittleren und starken Bereich liegen. Für die Rangordnung fallen diese am stärksten aus ($d = .91$ für die *Gesamtkompetenz*). Im KB *Sich-Bewegen* wird ein mittlerer negativer Effekt in der Differenzierung ($d = -.57$) erreicht und ein starker Effekt in der Rangordnung ($d = .77$).

Tab. 3: Individuelle Maße der Diagnosekomponenten für alle Lehrkräfte in der Einschätzung von motorische Basiskompetenzen

Lehrkraft	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Min	Max	M	SD	Mdn
Gesamtkompetenz																				
Niveau	.30	.06	.08	.22	.27	.17	.15	.15	.38	.51	.25	-.15	.04	.08	-.07	-.15	.51	.16	.17	.15
N-fehler (Be- trag)	.30	.06	.08	.22	.27	.17	.15	.15	.38	.51	.25	-.15	.04	.08	-.07	-.15	.51	.16	.17	.15
Differenzierung	1.24	.65	.88	1.63	.67	1.00	1.05	1.09	1.25	.39	1.01	.95	.66	.63	.63	.39	1.63	.92	.32	.95
D-fehler	.08	-.11	-.07	.12	-.12	.00	.01	.03	.06	-.14	.00	-.02	-.11	-.11	-.14	-.14	.12	-.03	.09	-.02
Rangordnung	.42	.07	.90	-.14	.61	.54	.76	.21	.72	.36	.81	.43	.50	.44	.78	-	-	-	-	-
Kompetenzbereich Sich-Bewegen																				
Niveau	.33	.11	-.08	-.14	.47	.33	-.38	.15	.48	.63	.63	-.14	-.25	.31	-.21	-.38	.63	.15	.33	.15
N-fehler (Be- trag)	.33	.11	-.08	-.14	.47	.33	-.38	.15	.48	.63	.63	-.14	-.25	.31	-.21	-.38	.63	.15	.33	.15
Differenzierung	.98	.02	1.05	1.30	.73	1.01	.87	1.07	.65	.41	.73	.87	.69	.52	.53	.02	1.30	.76	.32	.73
D-fehler	-.01	.00	.03	.06	-.11	.00	-.04	.03	-.12	-.18	-.13	-.06	-.11	-.18	-.20	-.20	.06	-.07	.09	-.06
Rangordnung	.16	.03	.92	.02	.52	.52	.55	.37	.25	.34	.78	.43	.51	.52	.61	-	-	-	-	-
Kompetenzbereich Etwas Bewegen																				
Niveau	.27	.01	.25	.57	.06	.00	.68	.15	.29	.38	-.13	-.16	.32	-.15	.06	-.16	.68	.17	.25	.15
N-fehler (Be- trag)	.27	.01	.25	.57	.06	.00	.68	.15	.29	.38	-.13	-.16	.32	-.15	.06	-.16	.68	.17	.25	.15
Differenzierung	.97	.31	.70	1.16	.52	.99	1.14	.93	1.30	.45	1.11	.85	.51	.77	.57	.31	1.30	.82	.30	.85
D-fehler	-.01	-.36	-.22	.06	-.23	.00	.04	-.03	.09	-.17	.04	-.08	-.23	-.08	-.22	-.36	.09	-.09	.14	-.08
Rangordnung	.56	.38	.78	.24	.32	.37	.58	.07	.58	.37	.76	.20	.42	.14	.63	-	-	-	-	-
N	15	8	6	11	29	17	13	33	15	21	11	11	25	13	11	6	33	16		14
Geschlecht	m	w	m	m	w	w	m	w	m	w	w	m	m	w	m					
Alter (Jahre)	30	53	63	36	34	27	57	60	58	31	49	34	38	62	62	27	63	45	14	44
Erfahrung (Jahre)	1	29	40	2	6	2	38	37	37	4	27	3	9	43	33	1	43	20	17	18
Schultyp	GYM	GYM	GYM	GYM	GYM	GYM	NMS	NMS	NMS	NMS	NMS	NMS	NMS	NMS	NMS					

Anmerkungen: GYM = Gymnasium; NMS = neue Mittelschule; Niveau: Akkuratheit der Einschätzung des absoluten Niveaus eines Merkmals. mittleres Lehrerurteil – mittlerer Testwert. > 0 = Überschätzung der Leistung. < 0 = Unterschätzung der Leistung. Wertebereich -3 bis 3; Differenzierung: Über- oder Unterschätzung der Streuung eines Merkmals. SD Lehrerurteils / SD Schülermerkmal. > 1 = Überschätzung der Heterogenität; < 1 = Unterschätzung; Rang: Korrekte Rangfolge der Leistungen innerhalb einer Klasse. Bivariate Produkt-Moment-Korrelation. Wertebereich -1 bis 1; - = nicht interpretierbar.

Tab. 4: Mittelwertvergleich der Diagnosekomponenten von Lehrkräften in Abhängigkeit vom Schultyp und von der Berufserfahrung

	Schultyp							Berufserfahrung						
		N	M	SD	t	p	d		N	M	SD	t	p	d
Gesamtkompetenz														
Niveau	NMS	9	.15	.21	-.23	.83	.12	< 10 Jahre	7	.18	.21	.49	.63	-.23
	GYM	6	.17	.09				> 10 Jahre	8	.14	.14			
N-fehler	NMS	9	.15	.21	-.23	.83	.12	< 10 Jahre	7	.18	.21	.49	.63	-.23
	GYM	6	.17	.09				> 10 Jahre	8	.14	.14			
Differenzierung	NMS	9	.85	.28	-.90	.40	.50	< 10 Jahre	7	.93	.41	.20	.85	-.09
	GYM	6	1.01	.37				> 10 Jahre	8	.90	.24			
D-fehler	NMS	9	-.05	.08	-.62	.55	.34	< 10 Jahre	7	-.03	.10	.30	.80	-.11
	GYM	6	-.02	.10				> 10 Jahre	8	-.04	.08			
Rangordnung	NMS	9	.68	.34	.62	.56	-.35	< 10 Jahre	7	.43	.27	-1.79	.10	.91
	GYM	6	.52	.56				> 10 Jahre	8	.79	.49			
Kompetenzbereich Sich-Bewegen														
Niveau	NMS	9	.14	.40	-.21	.84	.10	< 10 Jahre	7	.18	.35	.28	.79	-.15
	GYM	6	.17	.25				> 10 Jahre	8	.13	.34			
N-fehler	NMS	9	.14	.40	-.21	.84	.10	< 10 Jahre	7	.18	.35	.28	.79	-.15
	GYM	6	.17	.25				> 10 Jahre	8	.13	.34			
Differenzierung	NMS	9	.70	.21	-.74	.49	.45	< 10 Jahre	7	.86	.28	1.09	.30	-.57
	GYM	6	.85	.45				> 10 Jahre	8	.68	.34			
D-fehler	NMS	9	-.11	.08	-3.05	.01	1.52	< 10 Jahre	7	-.06	.08	.39	.70	-.23
	GYM	6	-.01	.06				> 10 Jahre	8	-.08	.09			
Rangordnung	NMS	9	.55	.24	.24	.82	-.13	< 10 Jahre	7	.39	.22	-1.39	.19	.69
	GYM	6	.49	.60				> 10 Jahre	8	.65	.49			
Kompetenzbereich Sich-Bewegen														
Niveau	NMS	9	.16	.29	-.26	.80	.11	< 10 Jahre	7	.21	.25	.45	.66	-.23
	GYM	6	.19	.22				> 10 Jahre	8	.15	.27			
N-fehler	NMS	9	.16	.29	-.26	.80	.11	< 10 Jahre	7	.21	.25	.45	.66	-.23
	GYM	6	.19	.22				> 10 Jahre	8	.15	.27			
Differenzierung	NMS	9	.85	.30	.44	.67	-.23	< 10 Jahre	7	.78	.28	-.48	.64	.23
	GYM	6	.78	.32				> 10 Jahre	8	.85	.33			
D-fehler	NMS	9	-.07	.12	.71	.50	-.43	< 10 Jahre	7	-.09	.12	-.03	.98	0
	GYM	6	-.13	.17				> 10 Jahre	8	-.09	.16			
Rangordnung	NMS	9	.48	.31	-.19	.86	.10	< 10 Jahre	7	.38	.14	-1.45	.17	.77
	GYM	6	.51	.29				> 10 Jahre	8	.59	.36			

Anmerkung: Die Ergebnisse zur Rangordnung beziehen sich auf Fischers Z-Werte.

6 | DISKUSSION

Die vorliegende Studie untersuchte die Diagnosekompetenz von Sportlehrkräften für motorische Basiskompetenzen. Zudem wurden Unterschiede in den Diagnosekomponenten zwischen den Lehrkräften thematisiert. Im theoretischen Teil wurde die Bedeutung eines akkuraten Urteils über die Ausprägung der motorischen Basiskompetenzen für fachdidaktische Entscheidungen herausgearbeitet. Dazu wurden die drei Diagnosekomponenten (Niveau, Differenzierung und Rangordnung) bei Sportlehrkräften erhoben und deren Zusammenhänge mit den Schülerergebnissen der MOBAK-Testbatterie geprüft. Die Ergebnisse zeigen, dass Sportlehrkräfte die Merkmale mit einer mittleren Akkuratheit einschätzen. Die Streuung der motorischen Merkmale (Differenzierung) wird dabei unterschätzt und die absolute motorische Kompetenz (Niveau) der Kinder überschätzt. Allerdings streuen die Werte zwischen den Lehrkräften stark, sodass einige Lehrkräfte die Kinder deutlich akkurater einschätzen. Auch gibt es Lehrkräfte, die motorische Merkmale stärker überschätzen, als auch solche, die diese unterschätzen bzw. beinahe zutreffend einschätzen.

Die beschriebenen Ergebnisse zur Akkuratheit der Urteile decken sich weitgehend mit bisherigen Ergebnissen. Auch Südkamp et al. (2012) beschreiben in ihrer Metaanalyse aus 75 Korrelationsstudien eine mittlere Urteilsakkuratheit von $r = .63^8$. Die mittlere Korrelation in einem früheren Review von Hoge und Coladarci (1989) lag bei $r = .66^9$. In der vorliegenden Studie muss die Urteilsakkuratheit von 15 Sportlehrkräften in der Einschätzung von motorischen Basiskompetenzen mit $M = .49$ (*Gesamtkompetenz*), $M = .44$ (*KB Etwas-Bewegen*) und $M = .43$ (*KB Sich-Bewegen*) als etwas geringer beschrieben werden. Die Medianwerte liegen zum Teil noch tiefer ($Mdn = .50$ für die *Gesamtkompetenz*; $Mdn = .38$ im *KB Etwas-Bewegen*; $Mdn = .51$ im *KB Sich-Bewegen*). Allerdings weisen die untersuchten Sportlehrkräfte eine massive individuelle Streuung auf. So liegen die Einzelwerte der Urteilsakkuratheit (r_{xy}) zwischen $-.13$ und $.90$, wobei das akkurateste Urteil bei einer sehr kleinen Klasse ($N = 6$) aufgefunden wurde. Im Mittel zeigt sich, dass Sportlehrkräfte selbst Kompetenzen, die Lernende mehrfach in jeder Sportunterrichtseinheit in unterschiedlichen Kontexten zeigen, nur bedingt einschätzen können.

Neben der Urteilsakkuratheit können die eingeschätzten Merkmale auch präzise bzw. unter- und überschätzt werden. In diesem Kontext zeigt der Abgleich mit anderen Studien uneinheitliche Ergebnisse. In der Einschätzung der physischen Fähigkeitsselbstwahrnehmung konnte Seyda (2016) eine Unterschätzung des Niveaus, jedoch eine Überschätzung der Merkmalsstreuung aufzeigen. Dies widerspricht unseren Ergebnissen, da sich die Diagnosekomponenten genau entgegengesetzt verhalten. Wird die Fähigkeitsselbstwahrnehmung von Kindern als Lern- und Leistungsvoraussetzung bezeichnet, ist es überraschend, dass Sportlehrkräfte das Niveau der Voraussetzung unterschätzen und die eigentliche motorische Kompetenz als Lernstand überschätzen. Dies würde be-

8 Südkamp et al. (2012) geben die Anzahl an Einschätzungen an, die in den einzelnen Studien abgegeben wurden. In den 75 Korrelationsstudien wurden je zwischen 16 und 9.650 Personen eingeschätzt. Allerdings finden sich keine Informationen zu der Anzahl an Personen, welche die Einschätzungen abgegeben haben.

9 Hoge und Coladarci (1989) geben im Literaturreview über 16 Korrelationsstudien keine Informationen zu einzelnen oder der durchschnittlichen Stichprobengröße an.

deuten, dass Sportlehrkräfte zwischen Lernvoraussetzung und Lernstand deutlich unterscheiden könnten. Im Bezug zur Differenzierungskomponente beobachtet Seyda (2016) eine Überschätzung der Streuung der Fähigkeitsselbstwahrnehmung, die vorliegende Studie allerdings eine Unterschätzung der Merkmalsstreuung. Aus Sicht der Sportlehrkraft erschwert dies die Passung der didaktischen Maßnahme mit Lernvoraussetzung, Lernstand und Lernziel. Daher darf dieser Befund als Anregung einer gezielteren Förderung von diagnostischen Kompetenzen interpretiert werden. Uneinheitlich zeigt sich diesbezüglich auch der Vergleich mit anderen Fächern. So wurde die Leseleistung in der Studie von Begeny et al. (2008) von Lehrkräften überschätzt, hingegen bei Feinberg und Shapiro (2009) unterschätzt. McElvany et al. (2009) beschreiben eine Unterschätzungstendenz im Umgang mit Text-Bild-Integrationen. Solche Unterschiede können laut Südkamp et al. (2012) einerseits messmethodisch auf die Validität der jeweiligen Testung, andererseits auch auf die unterschiedliche Zugänglichkeit der Merkmale zurückgeführt werden. Für die Ergebnisse der vorliegenden Arbeit sollte angemerkt werden, dass die Stichprobengröße Aussagen zur Stabilität der Ergebnisse offenlässt. Unabhängig davon, können Unterschiede zwischen den Untersuchungen auch mit der Komplexität der Aufgabe erklärt werden. So lässt sich für den Sportunterricht annehmen, dass die Lern- und Leistungsmerkmale der Lernenden für die Durchführung einer Rolle leichter zu diagnostizieren sind als jene bei einem Salto. Die damit verbundene Schwierigkeit in der akkuraten Bewertung kann eine nicht unwesentliche Rolle spielen.

Aufgrund der Daten können keine Aussagen zur Bedeutung der beobachteten Überschätzung des Niveaus und der Unterschätzung der Differenzierung für Schülerinnen und Schüler getroffen werden. Es wurden hierzu keine statistischen Grenzwerte zur Interpretation der Ergebnisse eingeführt. Der Niveau- und Differenzierungsfehler kann eine Deutungshilfe sein (Karst, 2012). So liegt der Niveaufehler für die *Gesamtkompetenz* – in einem Wertebereich von -3 bis 3 – bei $.16$ ($SD = .17$) und der Differenzierungsfehler für die *Gesamtkompetenz* bei $-.03$ ($SD = .09$). In diesem Zusammenhang konnte Brühwiler (2014) aufzeigen, dass erst die Kopplung des *diagnostischen Urteils* mit einer didaktischen Unterrichtsmaßnahme Auswirkungen für Lernende hat. Eine Überschätzung von Merkmalen kann sich dann durchaus positiv auf die Entwicklung der Lernenden auswirken, indem sie eine passende Anregung und Herausforderung für die kognitive Lernentwicklung ermöglicht (Hosenfeld, Helmke & Schrader, 2002). Sofern mit der Überschätzung der motorischen Basiskompetenzen also eine Aufgabenstellung verbunden wird, die leicht über den aktuellen Fähigkeiten der Lernenden liegt, könnte eine positive Auswirkung auf die Lern- und Leistungsentwicklung angenommen werden. Der Differenzierungsfehler ist vor allem im KB *Etwas-Bewegen* ($-.09$, $SD = .14$) auffallend. Somit könnte man schließen, dass sich Sportlehrkräfte in diesem Lernbereich nicht der Gesamtbreite des Lernstandes bewusst sind. Dies würde eine individuelle Förderung erschweren und könnte sich in einer Unter- bzw. Überforderung bei Lernenden äußern.

Die Ergebnisse zu den individuellen Unterschieden der Diagnosekompetenz und den Aufklärungsfaktoren Schultyp und Berufserfahrung zeigen einige Parallelen zum aktuellen Forschungsstand. Südkamp et al. (2012) beschreiben große Unterschiede in der Diagnosekompetenz zwischen Lehrkräften. Auch die vorliegende Untersuchung konnte eine deutliche Bandbreite zwischen den Sport-

Lehrkräften in der Qualität der Einschätzung von motorischen Merkmalen nachweisen. Am deutlichsten sind diese Unterschiede im KB *Etwas-Bewegen*. Woher diese Unterschiede kommen, bleibt allerdings unklar. Eine Annahme wäre aber, dass offene Bewegungen mit mehr Freiheitsgraden entsprechend schwieriger zu bewerten sind. Es konnten – mit einer Ausnahme – keine Unterschiede zwischen Lehrkräften in neuen Mittelschulen und an Gymnasien aufgefunden werden. Dem einzigen signifikanten Effekt (Differenzierungsfehler im KB *Sich-Bewegen*) wird aufgrund der kleinen Stichprobe und in Anbetracht der anderen Ergebnisse keine Bedeutung beigemessen. Auch die Berufserfahrung erklärt die individuellen Unterschiede in der Diagnosekompetenz nicht. Dies bestätigt die Ergebnisse von Demaray und Elliot (1998), welche ebenfalls keinen Einfluss der Berufserfahrung auf die Urteilsakkuratheit feststellen konnten. Zum Teil ist dieses Ergebnis aber auch überraschend, da sich berufliche Expertise üblicherweise durch konstante Konfrontation mit typischen Situationen im Arbeitsfeld entwickelt (Berliner, 1994; Dreyfus & Dreyfus, 1987). Zudem konnte Berliner (2004, zitiert nach Schrader, 2017) zeigen, dass erfahrene Lehrkräfte ihre Unterrichtskompetenz nur in Klassen mit vertrauten Schülerinnen und Schülern nutzen können. Auch in der vorliegenden Studie waren die Kinder der Sportlehrkraft mindestens über ein Schuljahr bekannt. Somit scheint die Entwicklung der Diagnosekompetenz (zumindest in der Ausprägung als Urteilsakkuratheit) anders zu verlaufen. Es lässt sich annehmen, dass den Lehrkräften im Regelfall kein lernrelevantes Feedback zu ihren Diagnoseleistungen im Unterricht zur Verfügung steht. Ganz ähnlich argumentieren Beck et al. (2008): Sie vermuten, dass in der komplexen Situation des Unterrichtens die Diagnosesituationen nicht erkannt oder hinreichend genutzt und verarbeitet werden können (vgl. auch Brühwiler, 2014). Dies würde nahelegen, dass Berufserfahrung nicht automatisch mit Expertise verbunden ist (Berliner, 1994; Dreyfus & Dreyfus, 1987). Erst ein intensives berufliches Engagement und das Bedürfnis nach Weiterentwicklung ermöglichen offenbar das Erreichen der Stufe der Expertise. Es bleibt also offen, woher die intraindividuellen Unterschiede zwischen den Sportlehrkräften in der Einschätzung von motorischen Merkmalen der Kinder kommen.

Limitationen

In dieser Studie konnten erste empirische Ergebnisse zur Diagnosekompetenz von 15 Sportlehrkräften in der Einschätzung von motorischen Basiskompetenzen bei 250 Kindern vorgelegt werden¹⁰. Trotzdem sind einige Limitationen anzuführen. Allen voran ist die Stichprobengröße zu erwähnen. So sind lediglich Kennwerte von 15 Sportlehrkräften vorhanden. Auch entstammen die Lehrkräfte keiner Zufallsstichprobe. Beides reduziert die Aussagekraft sowie die Stabilität der Ergebnisse. Entsprechend sollte die vorliegende Untersuchung als Pilotstudie verstanden werden, der weitere Erhebungen folgen sollten. Ferner sind bei der Erfassung der Diagnosekompetenz Längsschnittuntersuchungen anzustreben, um Wirkungen der Kompetenz zu erfassen. Dies hat den Vorteil, dass die Relevanz für den Fachunterricht steigt und Aussagen über die Professionalität der Lehrkraft möglich werden (Van Ophuysen & Behrmann, 2015). Diese Aspekte wurden lediglich in den theoretischen

¹⁰ Dies ist unter Beachtung der hohen Hürden bei schulischen Untersuchungen, des Aufwandes in der Durchführung von motorischen Testbatterien und der üblicherweise gering vorhandenen finanziellen Mittel für wissenschaftliche Zugänge der Sportpädagogik hervorzuheben.

Überlegungen berücksichtigt. Da für den Sportunterricht – im Gegensatz zu anderen Fächern – noch kaum Ergebnisse zur Ergebnisqualität von Diagnosen vorliegen, konzentrierte sich der empirische Zugang hierauf. Die Untersuchung der Diagnosekompetenz als Prozessfaktor bleibt für den Sportunterricht daher ein Forschungsdesiderat. Zudem ist anzumerken, dass diese Studie keine Grenzwerte (bzw. Normierungen) zur Interpretation der Diagnosekennwerte angibt. In einigen Studien (z. B. zur Diagnose des Fähigkeitsselbstkonzepts) werden die Komponenten des diagnostischen Urteils mit Berechnungen zu Grenzwerten ergänzt¹¹ (Praetorius et al., 2011; Spinath, 2005). Aufgrund der kleinen Stichprobe wurden die Diagnosekennwerte nicht mit Angaben zur Normierung erweitert. Die Diagnosekennwerte können somit nur in Relation zum möglichen Wertebereich oder ansatzweise über die Komponentenfehler interpretiert werden. Insbesondere für Untersuchungen der diagnostischen Kompetenz im Prozesskontext bleibt die zusätzliche Interpretationshilfe von Diagnosekennwerten durch Grenzwerte ein Desiderat. Auch die methodische Erhebung der diagnostischen Kompetenz sollte kritisch betrachtet werden. So wurden Lehrkräfte nicht aufgefordert, die Basiskompetenzen zu beobachten und dann zu bewerten, sondern sie sollten aus dem Gedächtnis eine Einschätzung vornehmen. Die Ergebnisse geben daher das Wissen aus summativen Diagnosen wieder. Damit kommen weitere Variablen, wie Motivation, Aufmerksamkeit oder Dauer des Unterrichts in der fokussierten Klasse, zur Wirkung. Es wurde also nicht direkt zwischen einer semiformalen und formellen Diagnose verglichen. Auch werden Berechnungen des diagnostischen Urteils häufig mit identischen Test- und Urteilsskalierungen vorgenommen (z. B. Spinath, 2005). Die Problematik der Äquidistanz in der Skalierung der Items wurde in der vorliegenden Studie mit einer Berechnung über die Skalennittelwerte etwas gemildert. Nicht zuletzt sind auch die Angaben zur Güte des MOBAK-5-Testinstruments weiter zu prüfen. So können Messfehler und Validität des Vergleichsinstruments als Störgröße in die Korrelationsberechnungen einfließen (Van Ophuysen & Behrmann, 2015). Vor diesem Hintergrund könnte die Akkuratheit des diagnostischen Urteils auch unterschätzt sein.

7 | FAZIT

Die vorliegende Studie konnte die semiformelle Diagnosekompetenz von Sportlehrkräften zum motorischen Lernstand von Schülerinnen und Schülern beschreiben. Dabei zeigt sich, dass Sportlehrkräfte selbst motorische Basiskompetenzen, die Lernende mehrfach in jeder Sportunterrichtseinheit in unterschiedlichen Kontexten zeigen, nur bedingt einschätzen können. In Anbetracht der Ausprägung der Diagnosen ergibt sich daher ein Bedarf, Sportlehrkräfte in ihrer Diagnosekompetenz gezielter als bisher zu fördern. Dieser Bedarf ist besonders in Kombination mit Erkenntnissen zur Einschätzung der physischen Fähigkeitsselbstwahrnehmung im Sportunterricht (Seyda, 2016) gegeben. Als bedeutsames Forschungsdesiderat kann, in Anlehnung an Schrader (2017), neben der Förderung der Diagnosekompetenz bei Sportlehrkräften, die Untersuchung der Entwicklung der Diagnosekompetenz in Aus- und Weiterbildung sowie im Beruf gelten.

11 Zusätzlich zur Unter-, Überschätzung bzw. präzisen Einschätzung (siehe Kap. 4.1) kann für die Niveauelemente mittels Ein-Stichproben-t-Test überprüft werden, ob sich der ermittelte Wert vom angestrebten Wert (0) unterscheidet (Praetorius et al., 2011). Zur Interpretation der Differenzierungskomponente kann nach Praetorius et al. (2011) der Pitman-Morgan-Test genutzt werden. Die Autorengruppe beschreiben weiter, dass es für die Rangkomponente keine grundsätzlichen Grenzwerte gibt. Sie prüfen allerdings, ob sich die ermittelten Werte signifikant von 0 unterscheiden.

LITERATUR

- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 3 (57), 175-193.
- Baumert, J. & Kunter, M. (2013). Professionelle Kompetenz von Lehrkräften. In I. Gogolin, H. Kuper, H.-H. Krüger, & J. Baumert (Hrsg.), *Stichwort: Zeitschrift für Erziehungswissenschaft* (S. 277-337): Springer Fachmedien Wiesbaden.
- Baumgartner, M. (2013). Kompetenzprofile von Sportlehrpersonen der Berufsfachschule. In F. Oser, T. Baurer, P. Salzmann & S. Heinzer (Hrsg.), *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzprofilen bei Lehrpersonen und Berufsbildungsverantwortlichen* (S. 97-126). Bad Heilbrunn: Klinkhardt.
- Baumgartner, M. (2016). *Performanzentwicklung in der Lehrerinnen- und Lehrerausbildung. Eine Interventionsstudie zur Entwicklung des Standards Feedback bei angehenden Sportlehrerinnen und Sportlehrern (Doktoratschrift)*. Europa-Universität Flensburg, Flensburg.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, C., Müller, P., . . . Vogt, F. (2008). *Adaptive Lehrkompetenz. Analyse und Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrerwissens*. Münster: Waxmann.
- Begeny, J. C., Eckert, T. L., Montarello, S. A. & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23 (1), 43-55.
- Berliner, D. C. (1994). Expertise: The wonders of exemplary performance. In J. N. Mangieri & C. Collin Blocks (Eds.), *Creating powerful thinking in teachers and students* (pp. 141-186). New York: Holt, Rinehart und Winston.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Brühwiler, C. (2014). *Adaptive Lehrkompetenz und schulisches Lernen. Effekte handlungssteuernder Kognitionen von Lehrpersonen auf Unterrichtsprozesse und Lernergebnisse der Schülerinnen und Schüler*. Münster: Waxmann.
- Cronbach, L. J. (1955). Processes affecting scores on understanding of others and assumed similarity. *Psychol Bull*, 52 (3), 177-193.
- Demaray, M. K. & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, 13 (1), 8-24.
- Dreyfus, H. L. & Dreyfus, S. E. (1987). *Künstliche Intelligenz. Von den Grenzen der Denkmaschine und dem Wert der Intuition*. Reinbek b. Hamburg: Rowohlt.
- Dünnebier, K., Gräsel, C. & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. *Zeitschrift für Pädagogische Psychologie*, 23 (34), 187-195.
- Feinberg, A. B. & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, 102 (6), 453-462.
- Fiske, S. T. (1993). Social cognition and social perception. *Annu Rev Psychol*, 44, 155-194.
- Frankhauser, D., Ferrari, I., Huber, C., Messmer, R., Moshfegh, Y., Plattner, M., ... Steinegger, A. (2014). Aufgaben im Sportunterricht. *Kompetenzorientierte Aufgaben für den Sportunterricht auf der Sekundarstufe I und II*. iTunes: Bundesamt für Sport (BASPO), Pädagogische Hochschule FHNW, Pädagogische Hochschule ZH.
- Gogoll, A. (2012). Sport- und bewegungskulturelle Kompetenz. Ein Modellentwurf für das Fach Sport. In A. C. Roth, E. Balz, J. Frohn & P. Neumann (Hrsg.), *Kompetenzorientiert Sport unterrichten* (S. 39-52). Herzogenrath: Shaker.
- Hascher, T. (2008). Diagnostische Kompetenzen im Lehrberuf. In C. Kraler & M. Schratz (Hrsg.), *Wissen erwerben, Kompetenzen entwickeln. Modelle zur kompetenzorientierten Lehrerbildung*. (S. 71-86). Münster: Waxmann.
- Heinzer, S. & Baumgartner, M. (2013). Bottom-up zur Qualität: Eine Quasi-Delphi-Studie zur Generierung von Kompetenzprofilen. In F. Oser, T. Baurer, P. Salzmann & S. Heinzer (Hrsg.), *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzprofilen bei Lehrpersonen und Berufsbildungsverantwortlichen* (S. 127-156). Bad Heilbrunn: Klinkhardt.

- mann & S. Heinzer (Hrsg.), *Ohne Kompetenz keine Qualität. Entwickeln und Einschätzen von Kompetenzprofilen bei Lehrpersonen und Berufsbildungsverantwortlichen* (S. 66-95). Bad Heilbrunn: Klinkhardt.
- Herrmann, C. & Gerlach, E. (2014). Motorische Basiskompetenzen in der Grundschule: pädagogische Zielentscheidung und Aufgabenentwicklung. *Sportunterricht*, 63 (11), 322-328.
- Herrmann, C., Gerlach, E. & Seelig, H. (2015). Development and validation of a test instrument for the assessment of basic motor competencies in primary school. *Measurement in Physical Education and Exercise Science*, 19 (2), 80-90.
- Herrmann, C., Gerlach, E. & Seelig, H. (2016). Motorische Basiskompetenzen in der Grundschule. Begründung, Erfassung und empirische Überprüfung eines Messinstruments. *Sportwissenschaft*, 46 (2), 60-73.
- Herrmann, C., & Seelig, H. (2017). „I can dribble!“. On the relationship between children's motor competencies and corresponding self-perceptions. *German Journal of Exercise and Sport Research*, 47(4), 324-334.
- Herrmann, C. & Seelig, H. (2017a). Basic motor competencies of fifth graders. *German Journal of Exercise and Sport Research*, 1-12.
- Herrmann, C. & Seelig, H. (2017b). Construct validity of the MOBAK-5 test instrument and determinants of basic motor competencies of fifth graders. *German Journal of Exercise and Sport Research*, 48 (1).
- Herrmann, C. & Seelig, H. (2017c). „I can dribble!“. On the relationship between children's motor competencies and corresponding self-perceptions. *German Journal of Exercise and Sport Research*.
- Herrmann, C. & Seelig, H. (2017d). Structure and profiles of basic motor competencies in the third grade—validation of the test instrument MOBAK-3. *Perceptual and Motor Skills*, 124 (1), 5-20.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59 (3), 297-313.
- Hosenfeld, I., Helmke, A. & Schrader, F.-W. (2002). Diagnostische Kompetenz. Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen*. (S. 65-82). Weinheim: Beltz.
- Huemer, M. & Amesberger, G. (2016). Evaluation von Schülerkompetenz im Unterrichtsfach Bewegung und Sport: Entwurf eines quantitativen Messverfahrens. In D. Wiesche, M. Fahlenbock & N. Gissel (Hrsg.), *Sportpädagogische Praxis – Ansatzpunkt und Prüfstein von Theorie* (S. 313-321). Hamburg: Feldhaus.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik*. Weinheim: Beltz.
- Jäger, R. S. & Petermann, F. (1999). *Psychologische Diagnostik*. Weinheim: Psychologie Verlags Union.
- Karing, C. & Artelt, C. (2013). Genauigkeit von Lehrpersonenurteilen und Ansatzpunkte ihrer Förderung in der Aus- und Weiterbildung von Lehrkräften. *Beiträge zur Lehrerbildung*, 31 (2), 166-173.
- Karing, C., Matthäi, J. & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I - Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie*, 25 (3), 159-172.
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern*. Münster: Waxmann.
- Karst, K., Klug, J. & Ufer, S. (2017). Strukturierung diagnostischer Situationen im inner- und außerunterrichtlichen Handeln von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen*. (S. 102-122). Münster: Waxmann.
- Kelly, G. A. (1955). *The psychology of personal constructs. A Theory of Personality (Bd. 1). Clinical diagnosis and psychotherapy (Bd. 2)*. New York, NY: Norton & Co.
- Klauer, K. J. (1982). *Handbuch der pädagogischen Diagnostik*. Düsseldorf: Schwann.
- Klug, J., Bruder, S., Kelava, A., Spiel, C., & Schmitz, B. (2013). Diagnostic Competence of Teachers: A Process Model That Accounts for Diagnosing Learning Behavior Tested by Means of a Case Scenario. *Teaching and Teacher Education*, 30, 38-46.
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielge-

- leiteter Prozess. Der Lehrer als „flexibler Denker“. *Zeitschrift für Pädagogische Psychologie*, 23 (34), 175-186.
- Kurz, D., Fritz, T. & Tscherpel, R. (2008). Der MOBAQ-Ansatz als Konzept für Mindeststandards für den Sportunterricht? In V. Oesterheld, J. Hofmann, M. Schimanski, M. Scholz & H. Altenberger (Hrsg.), *Sportpädagogik im Spannungsfeld gesellschaftlicher Erwartungen, wissenschaftlicher Ansprüche und empirischer Befunde* (S. 97-106). Hamburg: Czwalina.
- Lorenz, C. & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23 (34), 211-222.
- Lubans, D. R., Morgan, P. J., Cliff, D. P., Barnett, L. M. & Okely, A. D. (2010). Fundamental movement skills in children and adolescents: review of associated health benefits. *Sports Med*, 40 (12), 1019-1035.
- Madelaine, A. & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development and Education*, 52 (1), 33-42.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., ... Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23 (34), 223-235.
- Mussweiler, T. (1997). *A selective accessibility model of anchoring. Linking the anchoring heuristic to hypothesis-consistent testing and semantic priming*. Lengerich: Pabst Science Publisher.
- Niederkofler, B. & Amesberger, G. (2016). Kognitive Handlungsrepräsentationen als Strukturgrundlage zur Definition von kognitiver Aktivierung im Sportunterricht. *Sportwissenschaft*, 46 (3), 188-200.
- Oefner, J., Erlemeyer, R. & Staack, A. (o.J.). *Fördern und Fordern. Diagnostik und individuelle Förderung im Sportunterricht der Sekundarstufen I und II*. Arnberg: Bezirksregierung Arnberg Retrieved from https://www.schulsport-nrw.de/fileadmin/user_upload/schulsport-praxis_und_fortbildung/pdf/foerdern_fordern_klein.pdf
- Popham, W. J. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership*(82), 82-83.
- Praetorius, A.-K., Karst, K., Dickhäuser, O. & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, 58 (2), 81-91.
- Praetorius, A.-K., & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen*. (S. 13-18). Münster: Waxmann.
- Prohl, R. (2012). Der Doppelauftrag des Erziehenden Sportunterrichts. In V. Scheid & R. Prohl (Hrsg.), *Sportdidaktik. Grundlagen – Vermittlungsformen – Bewegungsfelder*. (S. 70-91). Wiebelsheim.
- Saarland, M. f. B. u. K. (2011). *Lehrplan Sport. Grundschule*. Retrieved from http://www.saarland.de/dokumente/thema_bildung/LPGSSport.pdf.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt am Main: Lang.
- Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 96-100). Weinheil: Beltz.
- Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23 (3-4), 237-245.
- Schrader, F.-W. (2017). Diagnostische Kompetenz von Lehrkräften – Anmerkungen zur Weiterentwicklung des Konstrukts. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (S. 245-255). Münster: Waxmann.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkung. *Empirische Pädagogik*, 1, 27-52.
- Seyda, M. (2016). Diagnostische Kompetenz von Sportlehrkräften als Voraussetzung für das adaptive Vorgehen im Sportunterricht. In C. Heim, R. Prohl, & H. Kaboth (Hrsg.), *Bildungsforschung im Sport. 29. Jahrestagung der dvs-Sektion Sportpädagogik vom 26.-28. Mai in Frankfurt/Main* (S. 91-92). Hamburg: Feldhaus.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19 (1/2), 85-95.

- Stodden, D. & Holfelder, B. (2013). Kein Kind bleibt zurück. *Zeitschrift für Sportpsychologie*, 20 (1), 10-17.
- Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, 73 (3), 437-446.
- Streit, C., Ruede, C. & Weber, C. (2015). *Diagnostische Kompetenz – Wie sich Experten und Novizen beim „Lesen“ von Schülerdokumenten unterscheiden*. Paper presented at the 49. Jahrestagung der „Gesellschaft für Didaktik der Mathematik“ GDM, Basel. https://www.researchgate.net/publication/296696367_Diagnostische_Kompetenz_-_Wie_sich_Experten_und_Novizen_beim_Lesen_von_Schülerdokumenten_unterscheiden
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *The Journal of educational psychology*, 104 (3), 743-762.
- Südkamp, A. & Praetorius, A.-K. (Hrsg.). (2017). *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen*. Münster: Waxmann.
- Van Ophuysen, S. (2006). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahneempfehlung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38 (4), 154-161.
- Van Ophuysen, S. & Behrmann, L. (2015). Die Qualität pädagogischer Diagnostik im Lehrerberuf: Anmerkungen zum Themenheft „Diagnostische Kompetenzen von Lehrkräften und ihre Handlungsrelevanz“. *Journal for educational research online*, 7 (2), 82-98.
- Van Ophuysen, S. & Lintorf, K. (2014). Unterschiede in der diagnostischen Praxis. *Empirische Pädagogik*, 28 (3), 211-228