Diagnostic Competence and Judgement Accuracy of Primary School PE Teachers in Assessing Basic Motor Competencies

von Simone Storni, Johanna Kress, Kathrin Bretz, Christian Herrmann, Benoît Lenzen, Ilaria Ferrari

Abstract: This study analyses the judgement accuracy of primary school physical education teachers in assessing their pupils' basic motor competencies, comparing their subjective evaluations (LEMOK: Lehrpersoneneinschätzung motorischer Kompetenzen) to the competence levels measured by the MOBAK (Motorische Basiskompetenzen) test, using a sample of 689 pupils (339 pupils in year four of primary school aged: *M* = 9.94 years, *SD* = 0.37, 51.0% girls; and 350 pupils in year four of primary school aged: *M* = 10.94 years, *SD* = 0.38, 50.9% girls). Results show a significant overestimation, especially with regard to Self Movement. Structural Equation Modelling analysis suggests that neither pupils' gender nor BMI directly influence the teachers' judgements, but that they have an indirect effect, mediated by the pupils' effective basic motor competencies. The findings highlight the importance of promoting initial and continuing professional development that integrates objective diagnostic tools and metacognitive awareness to support fairer, more informed and conscious evaluation practices.

Keywords: Basic motor competencies, diagnostic competence, judgement accuracy, physical education teachers.

Diagnostische Kompetenzen und Urteilsgenauigkeit von Sportlehrpersonen bei der Einschätzung motorischer Basiskompetenzen in der Grundschule

Zusammenfassung: Die Studie analysierte die Urteilsgenauigkeit von Sportlehrpersonen an Primarschulen, indem die subjektiven Lehrpersoneneinschätzungen motorischer Kompetenzen (LEMOK) mit den objektiven Ergebnissen der motorischen Basiskompetenzen (MOBAK) von 689 Kindern (339 Viertklässler:innen im Alter *M* = 9,94 Jahre, *SD* = 0,37, 51,0% Mädchen; und 350 Fünftklässler:innen im Alter *M* = 10,94 Jahre, *SD* = 0,38, 50,9% Mädchen) verglichen wurden. Die Daten zeigten eine signifikante Überschätzung der motorischen Basiskompetenzen durch die Lehrpersonen, insbesondere im Bereich des Sich-Bewegens. Mediationsanalysen zeigten, dass weder das Geschlecht noch der BMI die Einschätzungen der Lehrpersonen beeinflussen, sondern indirekt über die tatsächlichen motorischen Basiskompetenzen mediiert werden. Die Ergebnisse betonen die Relevanz der Förderung der beruflichen Aus- und Weiterbildung, wobei die Integration objektiver Diagnoseinstrumente und metakognitiver Reflexionsfähigkeit zur Unterstützung fairerer, fundierterer und bewussterer Bewertungsverfahren von entscheidender Bedeutung ist.

Schlagwörter: Motorische Basiskompetenzen, Diagnostische Kompetenz, Genauigkeit diagnostischer Urteile, Sportlehrpersonen.

Diagnostic Competence and Judgement Accuracy of Primary School PE Teachers in Assessing Basic Motor Competencies

1 Introduction

As early as the 1960s, Swiss pedagogue Moor emphasised that education requires understanding the child before attempting to teach (Moor, 1965). This early pedagogical view anticipated later developments in educational diagnostics, which conceptualise teachers' capacity to assess pupils' cognitive, emotional, and motivational states as a central aspect of professional competence (Shavelson, 1978).

In everyday school life, teachers need a wide range of assessment literacy to correctly assess pupils' characteristics, learning, and task demands (Herppich et al., 2018). In this regard, diagnostic competence has long been recognised as a fundamental component of professional competence (Weinert, 2001). In the German research literature, this construct corresponds to «*Diagnostische Kompetenz*«, which is conceptually close to assessment competence but more explicitly grounded in the pedagogical perspective (Hascher, 2008; Ingenkamp & Lissmann, 2008). This view naturally leads to a pedagogical-didactical conception of teachers' professional work, one that emphasises how diagnostic insights inform and adapt instruction. Following this tradition, diagnostic competence serves as a bridge between understanding and promoting learning, enabling teachers to translate diagnostic observations into adaptive teaching actions aligned with pupils' developmental needs.

However, despite growing attention to diagnostic competence, little is known about how it manifests in physical education (PE). Only few studies have examined how accurately PE teachers assess their pupils' basic motor competencies (BMC) (Ferrari et al., 2022; Herrmann & Seelig, 2020; Niederkofler et al., 2024).

In this study, the focus is on BMC, which constitute an essential prerequisite for pupils' participation in sport and physical activity within the school context (Herrmann & Seelig, 2020). Against this background, the study examines how accurately primary school PE teachers assess their pupils' BMC by comparing teachers' anticipatory subjective judgements (LEMOK) with pupils' objectively measured performances (MOBAK). Additionally, it investigates whether pupils' gender and body mass index (BMI) are associated with systematic biases in teachers' judgements. This dual focus allows for an examination of both judgement accuracy and potential sources of diagnostic bias.

1.1 Diagnostic competence and teachers' judgement accuracy

Following Weinert (2001), competencies are conceived as the dynamic interplay of cognitive, motivational, volitional, and social abilities that enable individuals to

DOI: 10.5771/2196-5218-2026-1-30 ZSF, 14. Jg., 1/2026, S. 30 – 53

act successfully in domain-specific contexts. This definition provides a broad and widely recognised theoretical foundation for understanding teacher competence.

In the field of teacher professional competence, Blömeke et al. (2015) conceptualise in their PID model competence as a continuum linking latent dispositions with situation-specific skills (perception, interpretation and decision-making), which result in observable performance. The PID model describes how teachers transform professional knowledge and underlying attitudes into action by perceiving, interpreting, and deciding in response to specific situations. High-quality PID processes make teaching more adaptive, increasing learning time and supporting pupils' motivation and learning outcomes.

Diagnostic competence is understood as a broader construct of skills and knowledge that includes the collection, interpretation, and use this information in authentic classroom situations (Blömeke et al., 2015; Voss et al., 2011). In this study, diagnostic competence refers to the underlying professional capacity that enables teachers to make accurate, reflective, and pedagogically meaningful judgements, whereas judgement accuracy represents the observable outcome of these diagnostic processes (Blömeke et al., 2015; Südkamp et al., 2012), operationalised as the degree of congruence between teachers' judgements and pupils' objectively measured BMC. This distinction aligns with Artelt and Gräsels' (2009) view that judgement accuracy constitutes a central expression of teachers' professional competence.

In this framework, «diagnostic competence" refers to the overarching professional capacity to make accurate and pedagogically meaningful judgements, while «diagnostic competencies« refer to the specific, observable skills like observing motor performance and adapting instruction. The singular form (competence) designates the integrated construct, whereas the plural (competencies) refers to its observable subcomponents (Barnett et al., 2010; Weinert, 2001). This conceptual distinction is particularly relevant in PE, where diagnostic work involves both underlying dispositions (e.g., perception and interpretation) and concrete, observable actions such as analysing motor performance, identifying errors and adjusting tasks as part of situation-specific professional decision making (Blömeke et al., 2015).

In this study, judgement accuracy serves as an empirical indicator of how effectively teachers apply their diagnostic competence in practice. It is operationalised as the degree of correspondence between teachers' anticipatory LEMOK judgement and pupils' objectively measured MOBAK performance, representing a measurable manifestation of competence in action (Blömeke et al., 2015; Herppich et al., 2018; Herrmann & Seelig, 2020; Südkamp et al., 2012). Through this lens, judgement accuracy represents the performance end of the competence continuum introduced by Blömeke et al. (2015), linking diagnostic dispositions and PID processes to observable evaluative outcomes relevant for effective instructional decision-making.

In PE, this framework has specific relevance, as diagnostic competence is exercised through the observation of embodied motor actions rather than purely cognitive outputs. Consequently, judgement accuracy offers a valid and context-sensitive indicator of diagnostic processes in practice (Südkamp et al., 2012). This perspective underscores the situated, performance-based nature of diagnostic work in PE, where perceptual sensitivity and interpretation are tightly coupled with teachers' ability to make adaptive and evidence-informed decisions that support both learning and instructional alignment.

1.2 Diagnostic judgement as a situated and accuracy-oriented process

Diagnostic judgement is best understood as a situated, accuracy-oriented process that integrates cognitive, motivational, and situational dimensions of teachers' professional action (Blömeke et al., 2015; Brühwiler, 2017; Helmke, 2017). It represents the applied dimension of diagnostic competence, where teachers' professional knowledge and situational skills are transformed into observable evaluative behaviour and thus become accessible through measures of judgement accuracy.

These interrelated components enable teachers to analyse pupils' learning processes, identify difficulties, and design targeted interventions (Brühwiler, 2017; Grotegut & Klingsieck, 2025). Accordingly, diagnostic competence goes beyond mere observation: it involves analytical reasoning and pedagogical interpretation that lead to adaptive and evidence-based decisions (Grotegut & Klingsieck, 2025; McElvany et al., 2009). Observation remains a cornerstone of diagnostic practice. Hascher (2008), Klug et al. (2013) and Kunter et al. (2013) highlight that classroom observation allows teachers to collect rich, situated information essential for informed pedagogical decisions. Moreover, semi-formal learning situations, less structured than standardised assessment situations but remaining pedagogically meaningful, enable teachers to collect complementary data that individualise instructional strategies (Südkamp et al., 2012), supporting more calibrated and accurate judgements.

Recent models (Baumert & Kunter, 2013; Loibl et al., 2020) emphasise that diagnostic judgement is cyclical and self-regulative, requiring teachers to perceive, interpret, and evaluate learning cues while actively managing cognitive biases.

1.3 Diagnostic judgement accuracy in PE

In PE research, diagnostic competence assumes specific characteristics due to the inherently performative, visual, and often subjective nature of classroom observations (Ferrari et al., 2022; Herrmann & Seelig, 2020). These conditions make PE an ideal domain for studying how diagnostic competence manifests through perception and decision-making in complex, dynamic contexts. Within this context, the focus

lies on BMC, which represent an important foundation for children's participation in sport and physical activity within the school setting (Herrmann & Seelig, 2020).

While Niederkofler et al. (2018) conceptualise PE teachers' diagnostic work as involving the observation of motor behaviour, the assessment of BMC, and the adaptation of instruction, Baumgartner (2022) emphasises the reflective and interpretative dimensions of teachers' evaluative practice, highlighting the role of subjectivity and pedagogical meaning in diagnostic processes. Their perspectives are complementary in that both underline the interplay between perception, interpretation, and pedagogical decision-making. In this framework, judgement accuracy represents the critical link between observation and pedagogical adaptation, as only sufficiently accurate judgement can support instructional decisions and learners' motor development (Hulteen et al., 2018; Stodden et al., 2008).

These interconnected processes operationalise diagnostic competence in PE, requiring not only theoretical and practical knowledge but also the ability to make accurate and context-sensitive judgements of pupils' motor performance. Effective diagnostic reasoning also requires teachers to attend to the affective and motivational factors that influence motor behaviour (Hascher, 2008; Spinath, 2005). Moreover, teachers' diagnostic work should remain aligned with the broader educational aims of PE as articulated in official curricula (Deutschschweizer Erziehungsdirektoren-Konferenz [D-EDK], 2017; Divisione della Scuola [DECS], 2015).

Unlike other school subjects assessed through written or standardised tests, PE judgements rely mainly on visual observation, making them more susceptible to bias and perceptual distortion (Herrmann & Seelig, 2020). The accuracy and pedagogical value of such judgements depend largely on teachers' diagnostic competence, specifically their capacity to observe motor behaviour systematically, assess physical abilities, and interpret the psychosocial dynamics influencing engagement (Ferrari et al., 2022; Spinath, 2005). This complexity underscores why judgement accuracy is particularly crucial in PE: it minimises interpretative errors, mitigates bias, and ensures equitable learning opportunities (Südkamp et al., 2012). Accordingly, judgement accuracy serves as a key indicator of diagnostic competence in PE, enabling teachers to identify pupils' motor strengths and difficulties based on a judgement that is sufficiently accurate for educational practice, rather than perfectly precise (Südkamp et al., 2012).

Among the several factors influencing judgement accuracy in PE, gender and body mass index (BMI) warrant particular attention, as both may shape teachers' expectations and perceptions and thus introduce systematic bias (Barnett et al., 2010). Empirical evidence shows that teachers often associate boys with higher competence in Object Movement (OM) tasks (e.g., throwing, catching, bouncing and dribbling) and girls with better performance in Self Movement (SM) tasks (e.g., balancing, rolling, jumping and running), reflecting enduring gender stereotypes that can bias

evaluations prior to observation (Barnett et al., 2010; Seyda & Langer, 2020). Similarly, pupils with higher BMI are often perceived as less competent regardless of their actual motor performance — a bias more closely linked to assumptions about body size than to objective ability (Baumgartner, 2022; Ferrari et al., 2022; Niederkofler et al., 2018).

Research consistently shows that teachers often struggle to assess pupils' performance accurately (Helmke, 2017; Südkamp et al., 2012), a finding that is also well documented in sport- and PE-specific research (Niederkofler et al., 2018). Many tend to overestimate both pupils' competence levels and their own judgement accuracy, thereby limiting the precision of their evaluations. Even experienced teachers are subject to cognitive distortions such as the halo effect or confirmation bias, which may compromise objectivity (Ferrari & Baumgartner, 2021; Van Ophuysen & Behrmann, 2015). In PE, where evaluation is often based on spontaneous and largely unstructured visual observation of pupils' motor behaviour during ongoing instructional activities (Urhahne & Wijnia, 2021), the predominance of perceptual impressions may increase the risk of biased or inconsistent judgements, potentially affecting diagnostic reliability. Overconfidence in teachers perceived judgement accuracy poses a significant risk, as it can lead to distorted assessments and expectations that may negatively influence pupils' motivation and development through self-fulfilling processes (Ferrari et al., 2022; Good et al., 2018). To minimise such biases, teachers must develop metacognitive awareness of the subjective elements inherent in diagnostic decision-making. In this context, self-efficacy — understood as the belief in one's ability to accurately observe, assess, and interpret pupils' competencies — emerges as a central determinant of judgement accuracy (Bandura, 1997). While contextual factors such as experience, pedagogical training, and class size may influence diagnostic competence (Baumert & Kunter, 2013; Klug et al., 2013; Südkamp et al., 2012), they fall beyond the scope of this study, which focuses primarily on judgement accuracy.

1.4 Research questions and Hypotheses

From a conceptual standpoint, diagnostic competence in teachers and motor competence in pupils can be seen as parallel and mutually informing constructs: both integrate cognitive, motivational, and performance-related dimensions. While pupils demonstrate their motor competence through observable performance, teachers express their diagnostic competence through the accuracy of their professional judgements, captured in this study through anticipatory LEMOK ratings completed before the MOBAK tests. This dual perspective reflects the reciprocal nature of teaching and learning processes in PE, where diagnostic and motor competences interact dynamically.

Building on this theoretical and empirical foundation, the present study investigates how accurately primary school PE teachers assess their pupils' BMC by comparing their anticipatory LEMOK judgements with objectively measured MOBAK performance. The analyses draw on a large dataset from pupils in year four and five of primary school in the Swiss Canton of Ticino, combining the MOBAK test (Herrmann, 2018) with teachers' LEMOK ratings for comparison.

It also examines whether judgement accuracy varies across the two MOBAK domains of OM and SM, which reflect children's ability to handle objects and to perform movements with their own body. Finally, the study explores the extent to which gender and BMI shape teachers' judgements, considering both potential direct influences and indirect effects mediated through pupils' actual motor performance.

Drawing on the theoretical framework and research questions, three hypotheses guide the present study.

1) Teachers' judgements (LEMOK) are expected to show a positive correlation with objectively measured MOBAK scores, indicating diagnostic competence through judgement accuracy (Blömeke et al., 2015; Brühwiler, 2017).
2) Teachers are assumed to demonstrate greater judgement accuracy in the OM domain than in the SM domain, as OM tasks are typically more structured and observable (Ferrari et al., 2022; Herrmann & Seelig, 2020).
3) Gender and BMI are not expected to directly influence teachers' judgements, but rather to exert indirect effects mediated by pupils' actual motor performance (Barnett et al., 2010; Südkamp et al., 2012).

Together, these hypotheses translate the conceptual model into empirically testable propositions, linking diagnostic competence to measurable indicators of judgement accuracy.

2 Material and Methods

The presented cross-sectional study focuses on data from the fourth measurement point (2024) of the longitudinal study «Development of Basic Motor Competencies in Children (EMOKK: Entwicklung Motorischer Kompetenzen in der Kindheit)« which took place in the Swiss Canton of Ticino. The Switzerland-wide study was funded by the Swiss National Science Foundation (SNSF; Grant no. 200840). The study was conducted in accordance with the Declaration of Helsinki and was approved by the Directorate of Education, Culture and Sport of the Canton of Ticino (No. 2021-00252, Rif CE Ti 3819) and by the University Commission for Ethical Research in Geneva (CUREG-2023-01-05). Prior to the testing, both teachers and parents or legal guardians of the children provided written consent for participation. The children verbally agreed to participate.

The school system in the Canton of Ticino has a peculiar feature: PE teachers in primary schools are specialised professionals, with a university degree (Bachelor's or Master's) and pedagogical preparation at the University for Teacher Education aimed at teaching PE exclusively. Each PE teacher supervises several classes, often in different schools, delivering three teaching units (45min lessons) per week in each class (2h15min per class in total).

2.1 Data collection

Data were collected from three different sources: (1) pupils' MOBAK scores, assessed using the MOBAK test instrument (Herrmann, 2018; Herrmann & Seelig, 2020); (2) the LEMOK, based on a questionnaire completed by the PE teacher for each child; and (3) teacher-related information obtained through a separate teacher questionnaire (Herrmann et al., 2024).

BMC (MOBAK test)

The MOBAK test instrument assesses children's BMC. In this study, BMC are operationalised through the validated MOBAK test battery. In the MOBAK test, these competencies are assessed based on eight items that can be divided into two competency areas: SM tested through the four items balancing, rolling, jumping, and running; and OM tested with the items throwing, catching, dribbling and bouncing.

The MOBAK test instruments were developed for preschool (Herrmann et al., 2020), as well as year one to six of primary school (Herrmann, 2018; Herrmann & Seelig, 2020). The test is designed with an appropriate age level of difficulty. The tasks were developed to assess motor competencies in accordance with the objectives defined in the official school curricula (D-EDK, 2017; DECS, 2015). The MOBAK tests were validated in several studies with confirmatory factor analyses (Herrmann et al., 2016; Herrmann & Seelig, 2020).

The assessment of the MOBAK test items is easily implemented due to a dichotomous coding (1 = pass or 0 = fail) for each test item and trial based on clear standardised criteria. No practice trial was allowed. Children performed each test item twice except for the items *throwing* and *catching*. Therefore, for each test item, a maximum of 2 points could be reached. The number of successful attempts was recorded, ranging between 0 = no successful attempts, 1 = one successful attempt and 2 = two successful attempts. For the items *throwing* and *catching* children had six attempts and the number of hits or caught balls was recorded. 0-2 hits or caught balls were given 0 points, 3-4 hits or caught balls were coded as 1 point, 5-6 hits were coded as 2 points. For the test evaluation, the sum of the scores of the

individual items in the OM and SM competence areas was calculated, resulting in a score of 0-8 points.

Data was collected between February and March 2024, during regular PE classes. Pupils were divided into groups of 3-4 children and performed the MOBAK 3-4 or the MOBAK 5-6 tests depending on the class level, under the supervision of trained staff.

Teachers' judgement (LEMOK)

Teachers made individual judgements about pupils' motor performance approximately one week before the MOBAK testing sessions. Each PE teacher received the LEMOK questionnaire in advance and was asked to complete one form for each child based on their prior knowledge and observations during regular PE lessons. To assess teachers' judgement (LEMOK), teachers were asked to indicate, for each child in the class, how they judged the child's motor competence in eight tasks based on the MOBAK test items (Herrmann et al., 2024). Since this teacher-based LEMOK version is an internal judgement instrument developed specifically within the EMOKK project and is not yet published as an independent tool, no external source is currently available[1]. In the questionnaire, the eight test items that the teachers had to evaluate were described based on the required competence. For example, one item asked: «*Can the child hit a target with a ball?*«

All participating PE teachers had access to detailed descriptions of the MOBAK tasks to ensure a consistent understanding of the required competences. Moreover, they were already familiar with the MOBAK test procedures, as they had participated in previous EMOKK data collections. The answer for each item was given on a trichotomous scale with the options "No", "Maybe" and "Yes".

The assumption behind this form of questioning children's motor competencies is that teachers also evaluate the children's competencies within the construct of the competence areas of SM and OM, each represented by four test items, resulting in a total of eight items. In previous studies, teachers carried out the judgement at class level (Ferrari et al., 2022). To align the child-centred judgement with the MOBAK test, the teachers' ratings were scored on a 3-point scale. The answers were recorded with 0 («No"), 1 («Maybe"), and 2 («Yes«) analogous to the MOBAK test coding, which is assessed with no, one and two successful attempts, resulting in a maximum of 8 points per competence area.

1 For further information on LEMOK, the authors can be contacted directly.

2.2 Sample

In total, N = 339 children[2] (age: M = 9.94 years, SD = 0.37, 51.0% girls) in year four of primary school from 21 classes and N = 350 (age: M = 10.94 years, SD = 0.38, 50.9% girls) children in year five of primary school from 22 classes were included. The average class size was n = 16.10 children in year four of primary school and n = 15.91 children in year five of primary school. The assessed participants came from ten different schools, evenly distributed across the Canton of Ticino. The 43 classes were taught by 14 different PE teachers (age: M = 42.48 years, SD = 9.17, 71.4% female). Data about the MOBAK tests of children in year four and five of primary school, the teachers' judgements of BMC (LEMOK) and further information about the teachers were collected. Table 1 illustrates the extent of data available for the LEMOK and MOBAK tests, divided by school years and the two competence areas.

Table 1. Overview of the sample characteristics of classes in year four and five with respect to availability. It reports the number of children from whom basic motor competencies (MOBAK) were assessed, the number of children with teacher judgements of basic motor competencies (LEMOK), and the number of children from whom both data sources are available.

		MOBAK	LEMOK	MOBAK and LEMOK
		n	n	n
Year 4	Object Movement	315	336	313
	Self Movement	315	337	314
Year 5	Object Movement	329	345	324
	Self Movement	329	342	321

2.3 Data analysis

After data collection and processing, the datasets of the MOBAK test, the teacher's judgement (LEMOK) and the information about the teachers were merged. Multivariate analysis with Mplus 8.8 (Muthen & Muthen, 2017) was then conducted as follows. The modelling of latent structures was carried out using a three-step approach. First, confirmatory factor analyses (CFA) were conducted to assess the factorial validity of the MOBAK and LEMOK instruments separately. Second, the relationship between MOBAK scores and teachers' judgements in the two motor competence areas was examined via a structural equation model (SEM) which serve as the basis for assessing judgement accuracy. Third, a structural equation

2 Number of children from whom at least one of the test forms was collected.

modelling was used to examine the associations between judgement accuracy and the covariates gender and BMI, in order to determine whether gender and BMI influence teachers' judgements directly or indirectly through their relationship with BMC. The motor competence areas SM and OM differ regarding the covariates gender and BMI (Ferrari et al., 2022; Herrmann & Seelig, 2020). Based on this knowledge, an analysis was conducted to determine whether gender and BMI represent a type of bias in the LEMOK assessment by teachers. A mediation analysis was carried out between the LEMOK (dependent variable (DV)), gender and BMI (both as independent variables (IV)) with the mediator variable MOBAK. These analyses were calculated separately for the two competence areas SM and OM.

The following recommended criteria were used to evaluate a good model fit: a comparative fit index of CFI > .95, although values of CFI > .90 are considered acceptable; and a root mean square error of approximation RMSEA < .06 (Hu & Bentler, 1999). Significance tests were conducted and reported two-sided and p values below .05 were regarded as statistically significant. The multi-level structure was considered by correcting the standard errors using the «type = complex« function.

Further descriptive and correlational analyses were conducted using SPSS 28 (IBM Corp., 2021). On a manifest level, descriptive statistics regarding MOBAK and LEMOK were calculated. For this, the sum values of MOBAK OM and SM as well as the LEMOK OM and SM were calculated. The teachers' judgement accuracy was calculated by the difference between the LEMOK and the MOBAK value in the respective area. Differences in mean values were analysed using independent samples t-tests. In addition to the mean values, standard deviations (SD) and effect sizes[3] (Cohen's d) were calculated for the teachers' judgement accuracy. This comparison between teacher ratings and performance scores reflects the level component of judgement accuracy, as conceptualised by Urhahne and Wijnia (2021), and serves as an indicator of diagnostic realism rather than discriminatory precision. Due to the change in the MOBAK test instrument between pupils in year four and five of primary school, all mean values and analyses were done separately for each year group.

Missing values occurred due to the use of different data sources (MOBAK assessment, LEMOK assessment) and variations in survey days (see Table 1). In some cases, children participated in the MOBAK assessment, but the teachers did not assess the LEMOK, and vice versa. Among children who took part in the MOBAK assessment, the proportion of missing values ranged from 7.1% (MOBAK catching) to 9.1% (MOBAK rolling) in year four of primary school and from 6.0% (MOBAK throwing) to 7.4% (MOBAK dribbling) in year five of primary school. In the

3 The interpretation of Cohen's d followed Cohen (1988) guidelines, categorising effect sizes as small (d = 0.10), medium (d = 0.50), and large (d = 0.80).

LEMOK assessment, missing values were observed in all items ranging from 0.6% (LEMOK jumping and running) to 1.5% (LEMOK catching) in year four of primary school and 1.4% (LEMOK bouncing) to 2.9% (LEMOK running) in year five of primary school.

The full information maximum likelihood (FIML) algorithm was used for latent analysis to estimate missing values. This procedure minimises bias in the sample composition by preserving the full sample size and is widely used in educational research (Urban & Mayerl, 2014). In contrast, for descriptive manifest statistics, missing data were handled using listwise deletion meaning that only children who presented MOBAK and LEMOK data in both competence areas were included (see Table 1).

3 Results

3.1 Correlation between LEMOK and MOBAK

The confirmatory factor analysis (CFA) confirmed an acceptable factorial validity of the LEMOK instrument for the two factors SM and OM in the classes in year four of primary school ($\chi^2 = 50.312$, $df = 19$, $p < .001$, CFI = .977, RMSEA = .070, $N = 337$) with excellent factor reliabilities (FR SM = .87; FR OM = .93). In classes in year five of primary school, the CFA also confirmed the factorial validity of the LEMOK instrument ($\chi^2 = 37.599$, $df = 19$, $p = .007$, CFI = .993, RMSEA = .053, $N = 345$), with excellent factor reliabilities (FR SM = .88; FR OM = .93).

The test of the two-factor structure of the MOBAK 3-4 instrument showed good model fit ($\chi^2 = 21.04$, $df = 19$, $p = .33$, CFI = .988, RMSEA = .018, $N = 315$) in year four of primary school. Both latent factors demonstrated satisfactory factor reliabilities (FR), which is notable given the small number of items per factor (FR SM = .58; FR OM = .66; see Weiber & Mühlhaus, 2014). The model fits and factor reliabilities of the selected subsample are consistent with those of the overall EMOKK sample (Herrmann et al., 2025).

In year five of primary school, the factorial validity ($\chi^2 = 22.84$, $df = 19$, $p = .24$, CFI = .977, RMSEA = .025, $N = 329$) and factor reliabilities (FR SM = .65; FR OM = .63) of the MOBAK 5-6 test instrument were also confirmed.

To investigate correlations between pupils' MOBAK scores and the teachers' judgements (LEMOK), as indicators of judgement accuracy, SEM was performed between the two latent factors of MOBAK (OM and SM) and the two factors of LEMOK (OM and SM) for both year groups separately (see Appendix figures 1 and 2). The four-factor SEM resulted for both year groups in a good model fit (year four of primary school: $\chi^2 = 137.350$, $df = 98$, $p = .005$, CFI = .980, RMSEA = .034, $N = 350$; year five of primary school: $\chi^2 = 138.285$, $df = 98$, $p = .0046$, CFI = .980, RMSEA = .035, $N = 338$).

The correlation between the factors MOBAK OM and LEMOK OM was $r = .66$ ($p < .001$) in year four of primary school and $r = .72$ ($p < .001$) in year five of primary school (Table 2). For the factors MOBAK SM and LEMOK SM the correlations were higher in year four of primary school $r = .76$ ($p < .001$) than in year five of primary school $r = .46$ ($p < .001$). The correlation between MOBAK OM and LEMOK SM was significant in both year groups (year four of primary school: $r = .65$, $p < .001$; year five of primary school: $r = .39$, $p < .001$), whereas the correlation between MOBAK SM and LEMOK OM was moderate but only significant in year four of primary school (year four of primary school: $r = .51$, $p < .001$; year five of primary school: $r = .18$, $p = .092$). The correlations within the MOBAK construct between OM and SM were $r = .48$ ($p < .001$) in the year four of primary school and $r = .34$ ($p < .001$) in the year five of primary school. The correlations between LEMOK SM and OM resulted as follows: $r = .89$ ($p < .001$) in year four of primary school and $r = .81$ ($p < .001$) in year five of primary school.

Table 2. Latent correlations of the SEM of the two basic motor competence areas (MOBAK OM and MOBAK SM) and the teachers' judgements of the motor competencies in the two competence areas (LEMOK OM and LEMOK SM) for year four and five.

	(1)	(2)	(3)	(4)
(1) LEMOK OM		.89**	.66**	.51**
(2) LEMOK SM	.81**		.65**	.76**
(3) MOBAK OM	.72**	.39**		.48**
(4) MOBAK SM	.18	.46**	.34**	

Note. *Above the diagonal values from the year four of primary school, below from the year five of primary school.*
*** p < .01*

3.2 Judgement accuracy and the role of gender, BMI, and BMC

Figure 1 shows SEM with LEMOK OM as the outcome (DV), while the standardised path coefficients (β) and their significance for both year four and five of primary school are presented. The analysis showed that the two independent variables gender (boys = 1 and girls = 2) and BMI were not directly significantly associated with LEMOK OM with the only exception in year five of primary school where the BMI had a small significant effect ($\beta = .12$, $p = .038$) on LEMOK OM. However, gender had a moderate to large significant effect on the MOBAK OM (year four of primary school $\beta = -.38$, $p < .001$; year five of primary school $\beta = -.51$, $p < .001$) with boys performing better. Also, the MOBAK OM had a significant association with the LEMOK OM (year four of primary school $\beta = .62$, $p < .001$; year five of primary school $\beta = .70$, $p < .001$). Gender and BMI seem not to be directly

associated with LEMOK OM. However, gender seems to be important for the MOBAK OM assessment, and this is in turn important for the teacher's judgement accuracy LEMOK OM. The association between gender and LEMOK OM was partially and significantly mediated by MOBAK OM. This results in an indirect effect of gender over MOBAK OM on LEMOK OM of β = -.234**, p < .001 in year four of primary school and β = -.354**, p < .001 in year five of primary school. The indirect effects of BMI over the MOBAK OM on the LEMOK OM were for both years small and not significant (year four of primary school β = -.062, p = .121; year five of primary school β = -.040, p = .377).

The model had acceptable model fits for both school classes (year four of primary school: χ^2 = 68.111, df = 31, p < .001, CFI = .967, RMSEA = .059, N = 339; year five of primary school: χ^2 = 65.590, df = 31, p < .001, CFI = .977, RMSEA = .056, N = 350). Overall, the model with gender, BMI and MOBAK OM accounted for 43% in year four of primary school and 47% in year five of primary school of the variance in LEMOK OM.

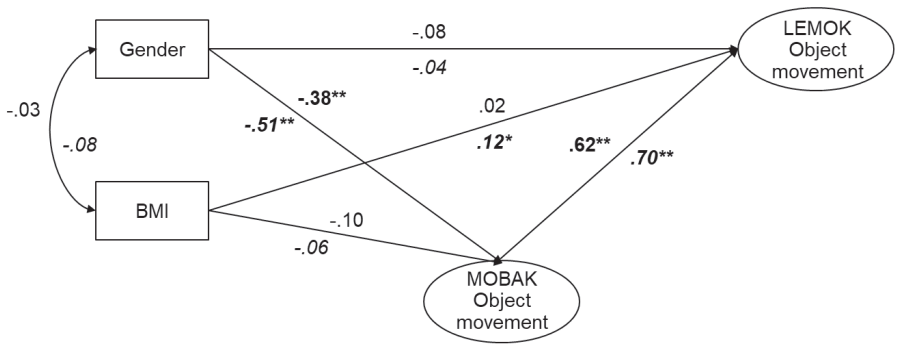

Figure 1. Mediation model of gender and BMI (IV), MOBAK OM (mediator), and LEMOK OM (DV).

Note. *Above the arrows are the standardised regression weights (Std B) of year four of primary school, below in italics those of year five of primary school; significance is indicated by p < .05 (*), p < .001 (**).*

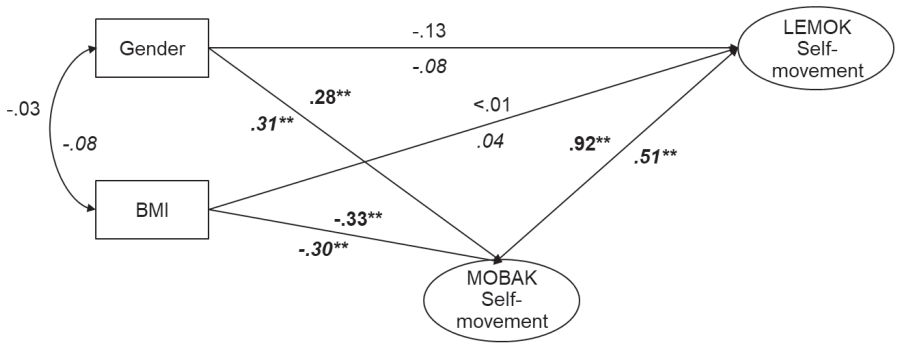

Figure 2. Mediation model of gender and BMI (IV), MOBAK SM (mediator), and LEMOK SM (DV).

Note. *Above the arrows are the standardised regression weights (Std B) of year four of primary school, below in italics those of year five of primary school; significance is indicated by p < .05 (*), p < .001 (**).*

The SEM with LEMOK SM as outcome is shown in Figure 2. Analogue to the calculation in Figure 1 the path coefficients indicate that gender (boys = 1 and girls = 2) and BMI are not directly associated with LEMOK SM. However, gender (year four of primary school β = .28, p < .001; year five of primary school β = .31, p < .001) and BMI (year four of primary school β = -.33, p < .001; year five of primary school β = -.30, p < .001) are significantly associated with MOBAK SM. In turn, the direct effects of MOBAK SM show an important significant association with LEMOK SM (year four of primary school β = .92, p < .001; year five of primary school β = .51, p < .001). The indirect effects of the mediation show, that the association between gender and LEMOK SM was mediated partially by MOBAK SM (year four of primary school β = .26, p < .001; year five of primary school β = .16, p = .001). This finding is consistent with the association of BMI and LEMOK SM, which is also significantly mediated by MOBAK SM (year four of primary school β = -.31, p < .001; year five of primary school β = -.15, p < .001).

The model presented a good model fit (year four of primary school: χ^2 = 50.870, df = 31, p < .001, CFI = .925, RMSEA = .053, N = 339; year five of primary school: χ^2 = 70.512, df = 31, p < .001, CFI = .934, RMSEA = .060, N = 350). Overall, the model with gender, BMI and MOBAK SM accounted for 79% in year four and 23% in year five of primary school of the variance in LEMOK SM.

3.3 Description of the sample at student level

The teachers generally estimated the children's BMC in both SM and OM to be high, with an average score of more than 6 points on a scale from 0 to 8. This trend was consistent across both class levels and competence areas (Table 3). In

comparison, the MOBAK scores of the pupils' were lower in both competence areas for both classes. Additionally, the difference between the teachers' judgements and pupils' MOBAK scores - for each child is indicated as the teachers' judgement accuracy.

The results indicate that teachers significantly overestimate the competence of the children in both year groups and in both competence domains OM (year four of primary school d = 1.09; year five of primary school d = 1.61) and SM (year four of primary school d = 1.45; year five of primary school d = .95).

Table 3. Mean values of the teachers' judgement (LEMOK) and children's basic motor competencies (MOBAK test), as well as the mean differences between the two measures across the competence areas self-movement and object movement in year four and five of primary school.

			Teacher Judgement (LEMOK)			MOBAK test		Teachers' judgement accuracy (difference between the teacher's assessment and the children's actual motor competence)		
		n	M	*SD*		M	*SD*	M	95%-CI	d
Year 4	Object Movement	313	6.66	1.69		4.50	1.97	**2.16****	[1.94; 2.38]	**1.09**
	Self-Movement	314	7.18	1.28		4.62	1.85	**2.56****	[2.37; 2.76]	**1.45**
Year 5	Object Movement	324	6.78	1.64		3.58	2.04	**3.20****	[2.99; 3.42]	**1.61**
	Self-Movement	321	6.14	1.90		3.87	2.15	**2.26****	[2.00; 2.52]	**.95**

Note. *Children included in this table (n) have a value for LEMOK and MOBAK.*
*** The t-test is significant at the 0.01 level (two-sided).*

4 Discussion

4.1 Interpreting main findings and hypotheses

The results of this study contribute to a clearer understanding of the judgement accuracy of primary school PE teachers, i.e. their capacity to make accurate and reliable assessments of children's BMC measured by the MOBAK test. The following discussion interprets the main findings with regard to the research questions and theoretical assumptions.

A central finding of this study is the significant correlation between pupils' MOBAK scores and teachers' judgements (LEMOK). This can be interpreted as an indicator of teachers' diagnostic competence, insofar as it suggests that teachers' judgements

are at least partially grounded in systematic observation and reflective processing. This aligns with international research on diagnostic competence (Herppich et al., 2018; Südkamp et al., 2012), which emphasises the importance of systematic observation and reflexive information processing. The correspondence also reflects the idea that judgement accuracy represents the observable outcome of teachers' diagnostic competence.

However, the analysis also reveals a systematic tendency for PE teachers to overestimate motor competencies: mean LEMOK scores exceed MOBAK scores across domains and in both year groups. This significant overestimation (see Table 3), already observed in the literature at class level (Ferrari et al., 2022), may reflect overconfidence in one's evaluative capacity, or an evaluation influenced by general impressions not verified through structured observation possibly shaped by cognitive biases such as the halo effect or confirmation bias (Seyda & Langer, 2020; Urhahne & Wijnia, 2021). These evaluative distortions can affect expectations, planning, and pupil support (Good et al., 2018). This underscores the importance of professional learning opportunities that help teachers calibrate their perceptions based on empirical evidence rather than intuition. This tendency is particularly evident in the OM domain, where correlations of $r = .66$ in year four of primary school and $r = .72$ in year five of primary school suggest a moderate but consistent level of judgement accuracy. A possible explanatory factor is the greater exposure to structured and clearly defined OM tasks in upper year groups, such as throwing, catching, or dribbling, which offer teachers more consistent and comparable opportunities for observation. These repeated and observable performances make it easier for teachers to judge pupils' competence accurately (Herrmann & Seelig, 2020).

In the SM domain, a decrease in judgement accuracy appears from year four of primary school ($r = .76$) to year five of primary school ($r = .46$). While this may suggest a reduced ability to assess BMC, it should be interpreted cautiously. Two pedagogical aspects may have contributed to this trend: (1) fewer opportunities to observe SM activities in higher classes/year groups (Herrmann et al., 2021), and (2) a shift from structured observation to more global impressions as pupils grow older (Hascher, 2008), which increases the risk of cognitive bias in teacher judgements (Ferrari et al., 2022). Additionally, the increased difficulty of the MOBAK 5-6 test may have made the BMC in pupils in year five of primary school less predictable for teachers and thus harder to anticipate accurately.

Another important finding is the high correlation between teachers' LEMOK ratings in the OM and SM domains. This indicates teachers' tendency to rely on a general impression of pupils' overall motor ability rather than forming distinct judgements for OM and SM. In practice, this suggests that teachers integrate multiple aspects of motor performance into a single global evaluation, reflecting a holistic but potentially less differentiated diagnostic approach. Such an approach is common in PE

contexts (Seyda & Langer, 2020) but may compromise judgement accuracy. Recent studies show that even PE teachers need structured tools and targeted training to accurately differentiate between the two domains (Gonzalez-Huenulef et al., 2023; Heemsoth & Wibowo, 2020).

A particularly encouraging finding concerns the mediation analysis, which shows no direct influence of pupils' gender or BMI on teachers' judgement accuracy. Instead, both variables act indirectly through pupils' MOBAK scores. The results confirm well-known trends: boys score higher on MOBAK OM tasks, while girls perform better in MOBAK SM tasks. This gap is reflected in teachers' ratings which tend to assess boys' motor competencies more positively, particularly in the domain of OM (Barnett et al., 2010; Herrmann & Seelig, 2020). However, SEM analysis showed that this relationship is entirely mediated by pupils' MOBAK scores: teachers do not appear to base their judgements directly on gender but rather on pupil's observed performance. This finding is encouraging from an educational perspective, as it highlights an assessment practice based more on observation and empirical data than on gender stereotypes (Gonzalez-Huenulef et al., 2023; Seyda, 2018).

Overall, the judgement accuracy of teachers is moderate and domain-specific. However, the data reveals two main challenges, including a general tendency to overestimate pupils' motor competencies and difficulties in distinguishing between the two motor domains. These aspects highlight an ongoing need to strengthen diagnostic training in both initial and continuing professional development, focusing on the use of structured observation tools (e.g., forms, rubrics, checklists) and fostering critical reflection regarding one's own assessment practices (Klug et al., 2013). To mitigate the effects of bias and subjectivity, teachers should be equipped with validated diagnostic tools and methodological training that foster objective, evidence-based assessment (Tschannen-Moran & Hoy, 2001). Therefore, the education and continuing professional development of PE teachers should not only strengthen their diagnostic and pedagogical professional knowledge but also develop a critical awareness of the cognitive processes that influence diagnostic judgement. Consequently, the development of diagnostic competence should be a central component of PE teacher training, with a focus on both structured observation and the use of tools and strategies that support fair and accurate assessment grounded in professional criteria as well as objective and scientifically validated data.

4.2 Methodological considerations and limitations

One methodological strength of this study is the combined use of objective data (MOBAK) and teachers' judgements (LEMOK), analysed by means of SEM models that made it possible to verify not only direct correlations but also mediating mechanisms between variables. Furthermore, the level of analysis is that of the

individual pupil, which gives the results greater precision and internal validity. The use of a large and representative sample of pupils from the school context in the Canton of Ticino increases the applicative relevance of the results. However, the relatively small number of participating teachers invites caution when generalising these findings to broader populations.

However, this study has some limitations. The cross-sectional design does not allow causal conclusions to be drawn, nor does it allow an assessment of how judgement accuracy develops over time. In addition, the use of the LEMOK instrument, which is based on a subjective rating scale, may have been influenced by implicit biases affecting teachers' judgements. This concern is consistent with findings from recent literature highlighting how subjective ratings can be influenced by factors unrelated to pupils' actual performance (Urhahne & Wijnia, 2021). A further limitation is that anticipatory ratings may reinforce expectation-based biases.

While the comparison between teachers' evaluations (LEMOK) and pupils' performances (MOBAK) offers a meaningful measure of judgement accuracy, its scope must be interpreted with caution. The LEMOK reflects teachers' anticipatory judgements before the MOBAK test, whereas the MOBAK tests captures only performance outcomes, not the underlying learning processes. Thus, judgement accuracy reflects a correspondence between subjective and objective indicators rather than the dynamic reasoning through which teachers interpret and promote learning.

A further methodological consideration concerns the MOBAK instrument itself. Although this is a validated and standardised measure of pupils' BMC, its structure focuses primarily on observable performance outcomes rather than on the diagnostic thought processes underlying motor learning and adaptation.

Consequently, the comparison between LEMOK and MOBAK captures only one dimension of teachers' professional judgement: the measurable correspondence between teachers' subjective, anticipatory evaluations and pupils' objective, performance-based results. In line with critical reflections (Herrmann & Seelig, 2020; Niederkofler et al., 2018), the validity of MOBAK as a diagnostic reference should be understood within the boundaries of what it measures. Future studies could integrate more process-oriented indicators to complement performance-based measures.

4.3 Practical and educational implications

From a practical standpoint, the results highlight the value of encouraging teachers to reflect critically on the accuracy of their own judgement. Beyond accuracy, diagnostic competence also involves understanding pupils' emotional and motivational states, which shape their engagement and learning in PE (Spinath, 2005). Supporting teachers in integrating both performance and contextual cues may strengthen diagnostic practice.

Furthermore, the tendency toward overestimation observed in this study deserves careful consideration. Although it may appear to be a simple quantitative bias, it can have a significant impact on teachers' expectations and educational planning. From a diagnostic perspective, overestimating pupils' motor competencies may lead to the selection of tasks whose demands are not aligned with pupils' actual competence levels, potentially resulting in overload and reduced opportunities for targeted support. Conversely, underestimation may lead to overly simplified tasks that limit challenge and learning opportunities. From this perspective, improving diagnostic competence means not only refining observational accuracy but also maintaining a balanced perception of pupils' actual learning needs. Simple structured tools (e.g., rubrics) may help reduce reliance on global impressions.

5 Conclusion

This study analysed the judgement accuracy of primary school PE teachers in assessing pupils' BMC, also considering the indirect influence of gender and BMI. Results reveal moderately accurate judgements, with significant correlations between teachers' judgements of pupils' BMC (LEMOK) and pupils' MOBAK performance, but also a systematic overestimation across domains (OM, SM) and school levels (year four and five of primary school).

At the same time, it is important to acknowledge that the MOBAK test, while validated, captures only pupils' BMC, operationalised through eight standardised performance tasks. However, diagnostic competence in PE also involves interpretative, reflective and context-sensitive processes (e.g. consideration of situational and motivational factors) that are not directly measurable through performance tests. Therefore, while judgement accuracy constitutes a valid empirical indicator, it captures only the observable outcome of teachers' diagnostic reasoning, which results from latent processes of perception, interpretation and decision making, without providing insight into how teachers reflect on their judgements or adapt them to situational and pedagogical contexts (Blömeke et al., 2015). Moreover, as Baumgartner (2022) notes, even standardised diagnostic instruments can only partially reflect the situational and interpretative nature of professional judgement. Future research should therefore integrate complementary multimodal and qualitative approaches to better capture the dynamic interplay between perception, interpretation, and pedagogical decision-making.

From a broader epistemological perspective, these findings reaffirm the reciprocal link between diagnostic and motor competencies. Teachers' ability to make accurate judgements not only reflects their professional expertise but also contributes to pupils' competence development through adaptive teaching. Strengthening diagnostic competence in PE thus represents both a pedagogical and an epistemic commitment to connect evidence-based assessment with meaningful learning in con-

texts of movement. Beyond its empirical contribution, this study also highlights an epistemological challenge: diagnostic competence, as a professional construct, lies at the intersection between measurable performance and interpretative understanding. Its assessment through judgement accuracy provides valuable insights, yet it cannot capture the reflective and situational nature of teachers' diagnostic reasoning. Recognising this tension is essential for advancing both theoretical models and empirical approaches in research on teacher competence.

6 References

Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift Für Pädagogische Psychologie*, *23*(3–4), 157–160. https://doi.org/10.1024/1010-0652.23.34.157

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Barnett, L. M., Van Beurden, E., Morgan, P. J., Brooks, L. O., & Beard, J. R. (2010). Gender differences in motor skill proficiency from childhood to adolescence: A longitudinal study. *Research Quarterly for Exercise and Sport*, *81*(2), 162–170. https://doi.org/10.1080/02701367.2010.10599663

Baumert, J., & Kunter, M. (2013). Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV. *Zeitschrift Für Erziehungswissenschaft*, *16*(2), 233–253.

Baumgartner, M. (2022). Professionelle Kompetenz(en) von Sportlehrkräften: Begriffe, Traditionen, Modellierungen und Perspektiven. In *R. Sygusch, J. Hapke, S. Liebl, & C. Töpfer (Hrsg.), Kompetenzorientierung im Sport: Grundlagen, Modellentwurf und Anwendungsbeispiele* (pp. 27–41). Hofmann.

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift Für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Brühwiler, C. (2017). Diagnostische und didaktische Kompetenz als Kern adaptiver Lehrkompetenz. In *A. Südkamp & A.-K. Praetorius (Hrsg.), Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen*. Waxmann.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.

Deutschschweizer Erziehungsdirektoren-Konferenz (D-EDK). (2017). *Lehrplan 21*. https://zh.lehrplan.ch/

Divisione della Scuola. (2015). *Piano di studio della scuola dell'obbligo del Canton Ticino*. Repubblica e Cantone Ticino, Dipartimento dell'Educazione, della Cultura e dello Sport.

Ferrari, I., & Baumgartner, M. (2021). Der Kompetenzansatz im Bewegungs- und Sportunterricht. In *Ch. Hayoz, N. Lanthemann, G. Patelli, & G. Grossrieder (Hrsg.), Kompetenzorientiertes Lernen und Lehren im Bewegungs- und Sportunterricht: Fachdidaktisches Referenzmodell*. HEP Verlag.

Ferrari, I., Kühnis, J., Bretz, K., & Herrmann, C. (2022). Diagnostische Kompetenz der Lehrpersonen und deren Bedeutung für die Förderung motorischer Basiskompetenzen. In R. Messmer & C. Krieger (Eds), *Narrative zwischen Wissen und Können* (pp. 175–194). Nomos. https://doi.org/10.5771/9783985720118-175

Gonzalez-Huenulef, Y., Martino Fuentealba, P., Bretz, K., Ferbol, C., & Carcamo-Oyarzun, J. (2023). Competencia motriz real y percibida en niños y niñas de la Patagonia chilena: La percepción del estudiantado y del profesorado de educación física. *Retos*, *50*, 290–297. https://doi.org/10.47197/retos.v50.99781

Good, T. L., Sterzinger, N., & Lavigne, A. (2018). Expectation effects: Pygmalion and the initial 20 years of research. *Educational Research and Evaluation*, *24*(3–5), 99–123. https://doi.org/10.1080/13803611.2018.1548817

Grotegut, L., & Klingsieck, K. B. (2025). Wie können unterschiedliche Aspekte diagnostischer Kompetenz gefördert werden? Drei Maßnahmen im Vergleich. *Zeitschrift Für Pädagogische Psychologie*, *39*(1–2), 95-109.

Hascher, T. (2008). Diagnostische Kompetenz im Lehrberuf. In *R. F. B. Schulte, T. Hascher, & K. O. Bauer (Hrsg.), Wissen erwerben, Kompetenzen entwickeln: Modelle zur kompetenzorientierten Lehrerbildung* (pp. 71–86). Waxmann.

Heemsoth, T., & Wibowo, J. (2020). Fachdidaktisches Wissen von angehenden Sportlehrkräften messen. *German Journal of Exercise and Sport Research*, *50*(2), 308–319. https://doi.org/10.1007/s12662-020-00643-0

Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Klett-Kallmeyer.

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*, 181–193. https://doi.org/10.1016/j.tate.2017.12.001

Herrmann, C. (2018). *MOBAK 1-4: Test zur Erfassung motorischer Basiskompetenzen für die Klassen 1–4*. Hogrefe.

Herrmann, C., Bretz, K., Kress, J., & Seelig, H. (2024). *Development of basic motor competencies during childhood (EMOKK): Documentation of items and scales. Survey 2024* [Technical report].

Herrmann, C., Bretz, K., Kühnis, J., Keller, R., Seelig, H., Kress, J., & Ferrari, I. (2021). *MOBAK – Monitoring of Basic Motor Competencies of 4- to 8-year-old Children in Switzerland: Documentation of Items and Scales*. Pädagogische Hochschule Zürich. https://doi.org/10.5281/zenodo.5726943

Herrmann, C., Ferrari Ehrensberger, I., & Bretz, K. (2025). *Assessing basic motor competencies in 3rd- and 4th-grade primary school children: The MOBAK-3-4 test* [Technical report]. https://doi.org/10.5281/zenodo.17140288

Herrmann, C., Ferrari Ehrensberger, I., Wälti, M., Wacker, S., & Kühnis, J. (2020). *MOBAK-KG: Motorische Basiskompetenzen im Kindergarten – Testmanual (3. Aufl.)*. https://doi.org/10.5281/ZENODO.3774435

Herrmann, C., Gerlach, E., & Seelig, H. (2016). Motorische Basiskompetenzen in der Grundschule. *Sportwissenschaft*, *46*(2), 60–73. https://doi.org/10.1007/s12662-015-0378-8

Herrmann, C., & Seelig, H. (2020). *MOBAK-5-6: Motorische Basiskompetenzen in der 5. und 6. Klasse – Testmanual (3. Aufl.)*. https://doi.org/10.5281/ZENODO.3774445

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hulteen, R., Morgan, P., Barnett, L., Stodden, D., & Lubans, D. (2018). Development of Foundational Movement Skills: A Conceptual Model for Physical Activity Across the Lifespan. *Sports Medicine*, *48*(7), 1533–1540. https://doi.org/10.1007/s40279-018-0892-6

Ingenkamp, K., & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik (6. Aufl.)*. Beltz.

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. https://doi.org/10.1016/j.tate.2012.10.004

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, *105*, 805–820. https://doi.org/10.1037/a0032583

Loibl, K., Leuders, T., & Dörfler, T. (2020). A Framework for Explaining Teachers' Diagnostic Judgements by Cognitive Modeling (DiaCoM). *Teaching and Teacher Education*, *91*, 103059. https://doi.org/10.1016/j.tate.2020.103059

McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H., & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift Für Pädagogische Psychologie*, *23*(3–4), 223–235.

Moor, P. (1965). *Heilpädagogik: Ein pädagogisches Lehrbuch*. Huber. https://books.google.ch/books?id=0DbnGAAACAAJ

Muthen, L. K., & Muthen, B. (2017). *Mplus (Version 8) user's guide*. Muthén & Muthén.

Niederkofler, B., Ferrari, I., Kühnis, J., & Herrmann, C. (2024). Diagnostische Kompetenz von Lehrpersonen in der Einschätzung der motorischen Basiskompetenzen von Kindern. In *Motorische Basiskompetenzen: Konstrukt, Erfassung und Anwendung* (pp. 209–225). Springer. https://doi.org/10.1007/978-3-658-45759-4_11

Niederkofler, B., Herrmann, C., & Amesberger, G. (2018). Diagnosekompetenz von Sportlehrkräften: Semiformelle Diagnose von motorischen Basiskompetenzen. *Zeitschrift für sportpädagogische Forschung*, *6*(2), 72–96. https://doi.org/10.5771/2196-5218-2018-2-72

Seyda, M. (2018). Können Sportlehrkräfte die Perspektive ihrer Schülerinnen und Schüler einnehmen? Eine Untersuchung über die Akkuratheit von Beurteilungen physischer Fähigkeitsselbstwahrnehmungen. *Unterrichtswissenschaft*, *46*, 215–231. https://doi.org/10.1007/s42010-017-0006-4

Seyda, M., & Langer, A. (2020). *Die Orchestrierung des diagnostischen und didaktischen Handelns im Sportunterricht—Eine Frage der Kompetenz! 69*, 353–357. https://doi.org/10.30426/SU-2020-08-3

Shavelson, R. J. (1978). Teachers' Estimates of Student ‹States of Mind' and Behavior. *Journal of Teacher Education*, *29*(5), 37–40. https://doi.org/10.1177/002248717802900511

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz: Accuracy of teacher judgments on student characteristics and the construct of diagnostic competence. *Zeitschrift Für Pädagogische Psychologie*, *19*(1/2), 85–95.

Stodden, D. F., Goodway, J. D., Langendorfer, S. J., Roberton, M. A., Rudisill, M. E., Garcia, C., & Garcia, L. E. (2008). A Developmental Perspective on the Role of Motor Skill Competence in Physical Activity: An Emergent Relationship. *Quest*, *60*(2), 290–306. https://doi.org/10.1080/00336297.2008.10483582

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgements of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762. https://doi.org/10.1037/a0027627

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Urban, D., & Mayerl, J. (2014). *Strukturgleichungsmodellierung: Ein Ratgeber für die Praxis. [Structural Equation Modeling. A Best Practice Guide]*. Springer VS. https://doi.org/10.1007/978-3-658-01919-8

Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgements. *Educational Research Review*, *32*, 100374. https://doi.org/10.1016/j.edurev.2020.100374

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*, 82–98.

Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, *103*(4), 952–969. https://doi.org/10.1037/a0025125

Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung*. Springer Gabler. https://doi.org/10.1007/978-3-642-35012-2

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds), *Defining and selecting key competencies* (pp. 45–65). Hogrefe & Huber.

Appendix

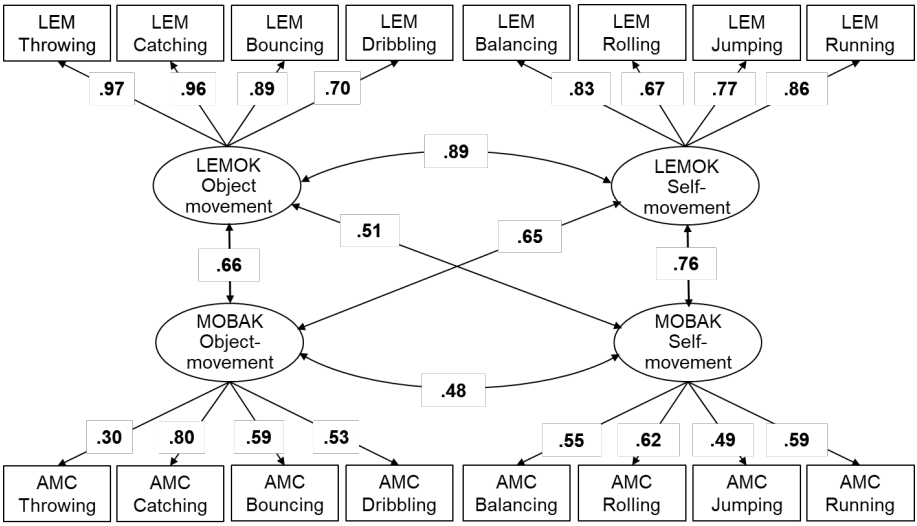

Figure 1. Structural equation model with the factor's pupils' MOBAK-3-4 performance (MOBAK OM and SM) and teachers' judgement (LEMOK OM and SM) for year four of primary school. Bold factors p < .05.

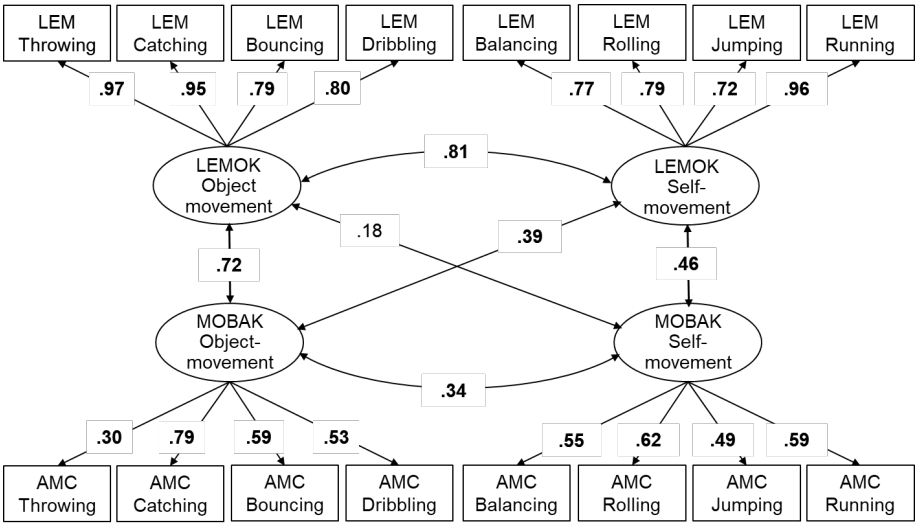

Figure 2. Structural equation model with the factor's pupils' MOBAK-5-6 performance (MOBAK OM and SM) and teachers' judgement (LEMOK OM and SM) for year five of primary school. Bold factors p < .05.