

Ger J Exerc Sport Res 2020 · 50:343–353
<https://doi.org/10.1007/s12662-020-00651-0>
 Received: 2 December 2019
 Accepted: 2 March 2020
 Published online: 18 March 2020
 © The Author(s) 2020



Steffen Greve¹ · Kira Elena Weber² · Björn Brandes³ · Jessica Maier⁴

¹ Institut für Bewegung, Sport und Gesundheit, Leuphana Universität Lüneburg, Lüneburg, Germany

² Institut für Schulpädagogik, Leuphana Universität Lüneburg, Lüneburg, Germany

³ Institut für Sport- und Bewegungswissenschaften, Universität Osnabrück, Osnabrück, Germany

⁴ Arbeitsbereich Bewegungs- und Sportpädagogik, Universität Hamburg, Hamburg, Germany

Development of pre-service teachers' teaching performance in physical education during a long-term internship

Analysis of classroom videos using the Classroom Assessment Scoring System K-3

Introduction

One of the most important components of pre-service teacher education programs is practical field experience. This is appreciated by both pre-service teachers and teacher educators (Arnold, Gröschner, & Hascher, 2014) and is therefore “a key aspect of a teacher education program” (Beck & Kosnik, 2002, p. 81). Accordingly, most teacher education programs in Germany have implemented long-term internships (Hascher & de Zordo, 2015). In line with this strategy, there is a growing body of research regarding the effects of internships on pre-service teachers' teaching competencies (e.g. Arnold et al., 2014; Cohen, Hoz, & Kaplan, 2013; Hascher, 2012). However, most studies focus on pre-service teachers' beliefs, attitudes, motivational variables, or perceived competency development and are mainly based on self-reports (Arnold et al., 2014; Cohen et al., 2013; Hascher, 2012). In order to measure the development of pre-service teachers' performance during a long-term internship, classroom observations are a less frequently used tool. Especially in physical education (PE), there is a lack of research regarding the effects of long-term internships on pre-service teachers' actual teaching performance.

This research gap needs to be closed, since PE with movement and play as its core content has specific challenges that are fundamentally different from other subjects like math or sciences, hence teaching situations differ from regular classrooms. Scherler (2008), for example, reported specific challenges for PE teachers with regard to content orientation, learning group management, and teacher–student communication. In order to contribute to a closure of the research gap in the effects of a long-term internship on pre-service teachers' teaching performance in PE classrooms, the authors conducted a video-based analysis using the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008), focusing on the quality of teacher–student interactions in PE classrooms.

Core domains of teacher–student interactions

Individual teachers have a great impact on students' learning (Hattie, 2009). Content knowledge and knowledge of methodical/didactic preparation of these contents can be regarded as the basis for good teaching. At this point, a distinction must be made between the structural quality of lesson preparation

by teachers and the dynamic quality of teacher–student interactions during the lesson itself. Good planning does not necessarily lead to good teaching. Highly qualified teachers should know what to teach and how to teach (pedagogical content knowledge) and should be able to apply this knowledge to lesson planning and in teacher–student interactions during their lessons. Ultimately, it is the classroom processes that students directly experience and primarily the quality of teacher–student interactions that facilitate student learning and future academic success (La Paro & Pianta, 2000; Pianta, La Paro, Payne, Cox, & Bradley, 2002). According to Pianta, Hamre, and Mintz (2012), the quality of social and instructional interactions between teachers and students is essential for promoting student learning and long-term school success. In the majority of the present work on the competence diagnostics of teachers, the actual performance is usually disregarded (Baumgartner, 2018). For ecological validity purposes, it is therefore currently required that the skill performance of teachers in the actual teaching situation should be given greater consideration. The concrete skill level of teachers in a teaching situation is described on the basis of Shavelson (2010, 2013)

by the term *teaching performance*. In German-speaking countries, three basic dimensions of teaching quality are used as a conceptual framework for teaching performance (e.g. Klieme, Pauli, & Reusser, 2009; Klieme & Rackoczy, 2008; Praetorius, Klieme, Herbert, & Pinger, 2018). These dimensions are classroom management, student-oriented classroom climate, and cognitive activation and therefore very similar to the domains suggested by Pianta and colleagues. They grouped these interactions into three domains: *Classroom Organization*, *Emotional Support*, and *Instructional Support* (Pianta et al., 2012; Pianta & Hamre, 2008). Especially for PE classrooms, the broader term of instructional support, which is suggested by Pianta et al. (2012), seems to grasp teacher–student activities better than the term cognitive activation. All CLASS dimensions are supposed to be subject-independent and are based on established theories and various empirical investigations in this field. Research suggests that students in classrooms with more emotional support have higher social competence and demonstrate higher performances in school (Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Curby et al., 2009; Curby & Chavez, 2013; Guo, Piasta, Justice, & Kaderavek, 2010; Mashburn et al., 2008). Moreover, effective classroom organization and instructional support are positively linked with behavior competence (Burchinal, Vernon-Feagans, Vitiello, Greenberg, & Family Life Project Key Investigators, 2014) and increased skills across different subjects (Hamre, Hatfield, Pianta, & Jamil, 2014; Maier, Vitiello, & Greenfield, 2012; Xu, Chin, Reed, & Hutchinson, 2014).

For PE, the three core domains of teacher–student interactions are also considered to be central to the quality of teaching (Herrmann, Seiler, Pühse, & Gerlach, 2015), taking into account the special features of the subject in terms of shape and form. This is especially important for the domain *Instructional Support*. For this domain, there is consensus in the scientific community that subject-specific supplementation is needed (Niederkofler & Amesberger, 2016). As this instrument does not ex-

ist—neither does a generally accepted competency model for PE—CLASS K-3 is used in this study without additions or modifications. A corresponding procedure is not considered necessary for the dimensions of *Classroom Organization* and *Emotional Support*; and corresponding subject-specific reading is considered sufficient (Herrmann et al., 2015). In the following, the three domains of teacher–student interactions are specified and empirical findings for PE presented.

Classroom organization

One of the most important components of teacher competence is classroom management (König, 2015). Classroom management or classroom organization refers to “the organization and management of students’ behavior, time, and attention in the classroom” (Pianta et al., 2012, p. 3). Teachers who can organize classrooms effectively minimize interruptions and maximize learning (Kounin, 1970). Therefore, in highly organized classrooms, students can spend more time on task (Lipowsky et al., 2009). Teachers display good classroom management skills when they establish consistent rules and routines and when their lessons are well-structured and efficiently organized (Praetorius, Pauli, Reusser, Rakoczy, & Klieme, 2014). These aspects can be defined as *Productivity* according to the CLASS framework (Pianta et al., 2012). Moreover, teachers should be able to stop inappropriate behavior immediately and encourage positive behavior (*Behavior Management*, Pianta et al., 2012). The way in which teachers utilize interesting learning materials is also important to engage students in active learning (Harris, 1998). The CLASS K-3 framework defines this as *Instructional Learning Formats*. Research has shown a correlation between student achievement and the ability of teachers to organize classrooms (for an overview, see Hattie, 2009). Moreover, research indicates that classroom organization is positively associated with children’s social skills, their behavioral competence, and their engagement in learning (Pianta et al., 2012). Furthermore, effective classroom

management helps to prevent anxiety disorders and burnout of teachers (Friedman, 2006; Lopez et al., 2008). To improve the ability of teachers to organize a classroom, not only is theoretical knowledge needed, but also teaching experiences in various classrooms. PE, in contrast to other subjects, does not take place in a regular classroom and PE teachers must perform in different teaching spaces (gymnasium, outdoor sports grounds, swimming pool) that are likely to suffer from high noise levels and poor acoustics, making effective teaching more demanding. Additionally, the children do not usually have a fixed seat in PE, but move around in a comparatively large area. Therefore, PE teachers must develop clear and consistent, yet flexible rules and routines that fit the different learning spaces (Cothran & Kulinna, 2014). Regarding the field of PE, Miethling and Krieger (2004) identified unsuccessful classroom management as a negative predictor for motivation. In addition, Seiler (2016) was able to show that effective classroom management has an impact on students’ motivation, while Heemsoth (2014) was able to prove that successful classroom management predicts performance motivation.

Emotional support

Another core domain of teacher–student interaction is emotional support. This refers to ways in which teachers help children develop warm and supportive relationships, experience enjoyment and excitement about learning, feel comfortable in the classroom, and experience appropriate levels of autonomy (Pianta & Hamre, 2008). The CLASS K-3 framework suggests four dimensions of emotional support (*Positive Climate*, *Negative Climate*, *Teacher Sensitivity*, and *Regard for Student Perspectives*). Research indicates that children display more social competence and engage more positively with their peers and teachers in classrooms with high levels of emotional support (Burchinal et al., 2010; Curby et al., 2009; Downer et al., 2011; Mashburn et al., 2008). Moreover, they demonstrate fewer behavior problems and therefore have fewer conflicts

with teachers (Hamre, Pianta, Downer, & Mashburn, 2008). There are also findings that emotional support is related to more engagement in school, academic achievements, and increased school motivation (e.g. Pianta et al., 2012). In comparison to other subjects, situations in PE are characterized by physical actions and are usually more emotionally charged (e.g., losing a match, performing in front of others). This makes emotional support more important and challenging and shapes teacher–student interactions in a special way (e.g., with regard to physical proximity and non-verbal or verbal cues). Regarding PE, Gerlach (2005) and Gerlach, Kussin, Brandl-Bredenbeck, & Brettschneider (2006) showed that positive teacher–student relationships have an impact on students’ exertion and well-being. Corresponding results are available for the intrinsic motivation of students (Jaakkola & Liukkonen, 2006). Heemsoth (2014) was able to show a connection between the performance motivation of students and teacher–student interactions as well as student–student relationships. There are also results on the importance of relationships in PE, mostly from qualitative studies (including Miethling & Krieger, 2004; about the students’ need for safety in PE lessons).

Instructional support

The conceptual foundation of instructional support is based on research into students’ cognitive and language development (Pianta et al., 2012). Teachers’ instructional support is associated with students’ development and learning, as well as their academic performance (Jamil, Sabol, Hamre, & Pianta, 2015). Teachers who demonstrate good instructional support provide their students with consistent, process-oriented feedback (*Quality of Feedback*; Pianta et al., 2008); they focus on higher-order thinking skills and help their students to understand things in a meaningful context (*Concept Development*; Pianta et al., 2008). Instructional support in preschool and elementary school also includes how teachers promote the language skills of their students (*Language Modeling*;

Ger J Exerc Sport Res 2020 · 50:343–353 <https://doi.org/10.1007/s12662-020-00651-0>
© The Author(s) 2020

S. Greve · K. E. Weber · B. Brandes · J. Maier

Development of pre-service teachers’ teaching performance in physical education during a long-term internship. Analysis of classroom videos using the Classroom Assessment Scoring System K-3

Abstract

Most teacher education programs in Germany have implemented long-term internships to foster pre-service teachers’ teaching skills. However, there is a lack of research to date regarding the effects of these internships on pre-service teachers’ actual teaching performance. This study attempts to offer insights into the performance development of physical education (PE) pre-service teachers during a 5-months internship. In total, 11 pre-service teachers were videotaped in the same class: three times at the beginning and three times at the end of the internship during PE lessons. All lessons were rated with the Classroom Assessment Scoring System (CLASS) by licensed CLASS observers. CLASS is a highly inferential rating system for the analysis of classroom observations and classroom videos. The CLASS for primary

schools (CLASS K-3) assesses 10 dimensions of teacher–student interactions, which are summarized in three core domains. These core domains comprise emotional support, classroom organization, and instructional support. Interrater reliability was assessed using the percent-within-one analysis. The development of teaching performance of pre-service teachers in PE showed no significant rise in the 10 dimensions/three core domains measured over the course of the internship. Interpretations of these findings and practical implications for teacher education are discussed.

Keywords

Classroom Assessment Scoring System · Teacher Education · Video analysis · Highly inferential rating system · Primary school

Pianta et al., 2008). Klieme, Pauli, & Reusser, (2009) refer to the dimension of cognitive activation, when teachers encourage higher-order thinking in their students. Teachers who are interested in their students’ thoughts, who provide stimulating questions, challenging tasks, and connect different concepts and ideas are more likely to create a cognitively activating classroom (Taut & Rakoczy, 2016). However, in contrast to classroom organization or emotional support, cognitive activation/instructional support “has to be defined for each academic subject based on specific findings from didactic research and cognitive psychology in that field” (Taut & Rakoczy, 2016, p. 47). In PE, many of these cognitive aspects seem less relevant at first glance. Students often imitate actions that were demonstrated by the teacher or other students. In addition, periods of conversation and verbal reflection are less frequent in PE, but not less important (Greve, 2013). In the field of PE research, the dimension of cognitive activation gained importance only recently (Niederkofler & Amesberger,

2016). This can be explained by the fact that there is currently no subject-specific competence model that could be transferred to a corresponding examination instrument (Herrmann et al., 2015).

Development of teacher–student interactions during long-term internships

Field experiences have “been widely recognized as a fundamental educational tool that allows for the integration of theoretical knowledge with real world practice in the professional field” (Tapp, Macke, & McLendon, 2012). Therefore, it is assumed that school internships offer a variety of learning opportunities for pre-service teachers and improve teaching qualities accordingly. In particular, the implementation of long-term internships in the first phase of teacher training is often justified by the fact that longer internships are superior to shorter ones (Weyland & Wittmann, 2015). In addition, Gröschner et al., (2015) point out that internships in teacher education can promote the acquisition of teaching skills.

Looking at previous research, it is noticeable that studies on subjective competence assessment by pre-service teachers themselves or by their mentors predominate (Gröschner, Schmidt, & Seidel, 2013; Weyland & Wittmann, 2015). For the field of PE, Linka and Gerlach (2019), for example, compare pre-service teachers' assessments of the dimension of class leadership with corresponding students' assessments.¹

However, restraint is called for, since only a handful of empirical studies examine the extent to which internships improve the concrete teaching activities of pre-service teachers (Gröschner, Klaß, & Dehne, 2018). Küster (2008) analyses classroom videos of eleven pre-service teachers during a long-term internship. He uses both low and highly inferential procedures of classroom observation to investigate the growth in teaching performance. Results of the highly inferential analysis display a significant increase over time, whereas the findings of the low inferential video analysis showed no increase. For the subject of PE, there are virtually no known findings regarding the analysis of pre-service teachers' performance using classroom videos. Only Baumgartner (2018) was able to reveal, in a pre-post design, that pre-service teachers for vocational school do not improve their feedback-related performance through a long-term internship. Baumgartner used the concept of performance as the observable behavior that emerges from professional competencies (Shavelson, 2010, 2013), which is also used in the present study. Schwarz (2009) was able to show that pre-service teachers remember their own teaching situations better if they analyzed them with the help of video recording.

Measuring teacher–student interactions and teaching performance

Classroom observation instruments have increasingly gained importance in recent years due to their potential to provide an evidence-based measure of teacher–stu-

dent interactions and teaching performance (White, 2018). In contrast to low inferential rating systems, highly inferential rating systems require a high degree of inferences on the part of the observer (Lotz, Gabriel, & Lipowsky, 2013). Therefore, observers must be carefully trained in order to provide accurate and consistent ratings (White, 2018). One instrument that was developed and published in 2008 and has since been widely adopted for research and evaluation purposes in over 3000 early childhood and elementary classrooms is the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008). CLASS is not restricted to a specific subject and can be used to analyze changes in different subdimensions and the aforementioned three domains of teacher–student interactions (Sandilos & DiPerna, 2011). Two previous US studies with an overall of 115 primary school teachers conducted with CLASS K-3 summarize the performance of the involved teachers in the domain of emotional support in an upper mid-range, classroom organization in an upper mid-range, and instructional support in a lower mid-range (Pianta et al., 2008). In their study of 332 Italian primary school teachers, Longobardi, Pasta, Marengo, Prino, & Settanni, (2018) found upper mid-range performances in all dimensions and domains.

CLASS observers need to complete a 2-day training program first and succeed in a video-based reliability certification. During the recertification process, trainees score five 20-min videotaped segments from classrooms on a seven-point scale, and their scores are compared with those of the master coders. To assess interrater reliability, CLASS uses the percent-within-one (PWO) analysis (Hamre et al., 2008; Pianta et al., 2008; La Paro, Pianta, & Stuhlman, 2004). When calculating PWO, scores are considered to be in agreement if they fall within ± 1 point of each other. Thus, for two raters to achieve 80% reliability on a CLASS cycle, eight out of 10 scores must fall within one point of each other. Trainees who pass with a PWO of 80% or higher in relation to the master coders are certified in CLASS. In field studies with different CLASS observers, most researchers

report adequate (between 0.78–0.96%) interrater reliability of scores on CLASS dimensions (Sandilos & DiPerna, 2011).

Aim of the study

Up until now, there have been no empirical findings on the development of PE pre-service teachers' teaching performance during long-term internships. The teaching performance of pre-service PE teachers was analyzed and evaluated with regard to the improvement of classroom organization, emotional support, and instructional support during a 5-month internship. In addition, the professional performance of pre-service teachers in the present study is compared to performance values from other CLASS studies (that do not refer to results of PE studies) to inspect possible subject-specificity or particularities of the sample.

Method

Participants and study design

The sample consisted of 11 (72% female; $M_{\text{age}} = 24$, $SD_{\text{age}} = 1.5$) pre-service PE teachers for primary school in their second masters semester of teacher education. All students had a bachelor's degree in teaching and learning, which they received from the same university. Their bachelor was set up for six semesters and two internships were part of the bachelor program. The first internship was a 3-week observation and teacher assistance in their second bachelor semester, while the second internship was a 4-week teaching position in their fourth bachelor semester. Since classroom management is considered one of the most important tasks for pre-service teachers (Wolff, Jarodzka, & Boshuizen, 2017), the first teaching internship focused on classroom management. A 5-month internship is integrated in the curriculum of the master program. The pre-service teachers have a subject-specific preparatory seminar for the internship, as well as an accompanying seminar during their internship. During the internship, pre-service teachers are obliged to teach 64 lessons in each subject. They

¹ This project is currently underway and cannot report any final findings yet.

Table 1 Definitions and examples from this study of CLASS K-3 domains and sub-dimensions

Domain/dimension		Indicators (markers)	Examples of observed classroom practices (high quality)
Classroom organization	Behavior management	Clear behavior expectations (clear expectations, consistency, clarity of rules), proactive (anticipates problem behavior or escalation, monitors low reactivity), redirection of misbehavior (effective reduction of misbehavior, attention to the positive, uses subtle cues to redirect, efficient redirection), student behavior (frequent compliance, little aggression and defiance)	At the beginning of a lesson on “throwing with different balls,” the pre-service teacher issues the rule that all balls must be placed in a ball box provided during the discussion phases. A whistle signal is used for this. When the whistle sounds, all the students place the balls in the ball box and come together in a circle
	Productivity	Maximizing learning time (provision of activities, choice when finished, few disruptions, effective completion of managerial tasks, pacing), routines (students know what to do, clear instructions, little wandering), transitions (brief, explicit follow-through, learning opportunities within), preparation (materials ready and accessible, knows lessons)	During a group work phase with the topic “jumping,” all students change stations at the pre-service teacher’s signal. The pre-service teacher uses a stopwatch and a whistle. At the stations there are task cards with pictures of the exercises. These are clear and unambiguous
	Instructional learning formats	Effective facilitation (teacher involvement, effective questioning, expanding children’s involvement), variety of modalities and materials (range of auditory, visual, and movement opportunities, interesting and creative materials, hands-on opportunities), student interest (active participation, listening, focused attention), clarity of learning objectives (advanced organizers, summaries, reorientation statements)	In a lesson on the subject of ‘parkour,’ the equipment in the gymnasium is already assembled when the students arrive. The pre-service teacher confronts the students with the task: “How can we get across the equipment to the other side of the gymnasium without touching the ground?” The students gather ideas using the ‘think-pair-share’ method and record these on a poster. Afterwards the students can put their ideas to practice
Emotional support	Positive climate	Relationships (physical proximity, shared activities, peer assistance, matched affect, social conversation), positive affect (smiling, laughter, enthusiasm), positive communication (verbal affection, physical affection, positive expectations), respect (eye contact, warm, calm voice, respectful language, cooperation and/or sharing)	During a passing exercise in soccer, the pre-service teacher approaches a group of students and says: “Elena, super! You did very well!”, while clapping her hands and pointing a thumb up to Elena
	Negative climate	Negative affect (irritability, anger, harsh voice, peer aggression, disconnected or escalating negativity), punitive control (yelling, threats, physical control, harsh punishment), sarcasm/disrespect (sarcastic voice/statement, teasing, humiliation), severe negativity (victimization, bullying, physical punishment)	In the sitting circle at the end of an hour, the pre-service teacher yells at a student: “Emre, you just got on my nerves!”
	Teacher sensitivity	Awareness (anticipates problems und plans appropriately, notices lack of understanding and/or difficulties), responsiveness (acknowledges emotions, provides comfort and assistance, provides individualized support), addresses problems (helps in an effective and timely manner, helps resolve problems), student comfort (seeks support and guidance, freely participates, takes risks)	During a dancing lesson a student has problems with the rhythm of the dance steps. The pre-service teacher recognizes this and immediately goes to the student. He counts the steps together with her and emphasizes the beat
	Regard for student perspectives	Flexibility and student focus (shows flexibility, incorporates students’ ideas, follows students’ lead), support for autonomy and leadership (allows choice, allows students to lead lessons, gives students responsibility), student expression (encourages student talk, elicits ideas and/or perspectives), restriction of movement (allows movement, is not rigid)	In an acrobatic lesson the students are allowed to think up freestyle and decide for themselves. The pre-service teacher advises the students on how to do this, but the decisions are made by the students in agreements with each other

Table 1 (Continued)

Domain/dimension		Indicators (markers)	Examples of observed classroom practices (high quality)
Instructional support	Concept development	Analysis and reasoning (why and/or how questions, problem solving, prediction/experimentation, classification/comparison, evaluation), creating (brainstorming, planning, producing), integration (connects concepts, integrations with previous knowledge), connections to the real world (real-world applications, related to student's lives)	At the beginning of a lesson on 'playing in teams,' the pre-service teacher asks who is doing a team sport. He then asks the question: "What do we need to play successfully as a team?" The students brainstorm and come up with different aspects
	Quality of feedback	Scaffolding (hints, assistance), feedback loops (back-and-forth exchanges, persistence by teacher, follow-up questions), prompting through processes (asks students to explain thinking, queries responses and actions), providing information (expansion, clarification, specific feedback), encouragement and affirmation (recognition, reinforcement, student persistence)	During a passing exercise in a hockey lesson, the pre-service teacher stays close to the practicing students. One student has difficulties stopping the ball with her hockey stick. The pre-service teacher asks her: "Why do you stay where you are?" The student shrugs her shoulders. The pre-service teacher gives her hint: "You have to run to the left or right side of the ball. And then you have to stop it with the flat side of the hockey stick!"
	Language modeling	Frequent conversation (back-and-forth exchanges, contingent responding, peer conversations), open-ended questions (questions require more than a one-word response, students respond), repetition and extension (repeats, extends/elaborates), self- and parallel talk (maps own actions with language, maps students action with language), advanced language (variety of words, connected to familiar words and/or ideas)	During a passing exercise in soccer, the teacher says: "Dustin, you have to shoot with the instep. And then stop with the instep." The teacher points to his own inner instep and lifts the foot slightly

Definitions and indicators were based on the CLASS K-3 manual (Pianta et al., 2008). Examples are based on the observation of classroom videos in this study. Negative Climate scores are reverse

had a school-mentor in each subject and received three visits from lecturers from the university seminars. Various topics were reflected on in the consultations, for example: organization in the gymnasium, the conversation between pre-service teachers and students, and the didactic-methodical preparation of lesson contents.

Video material and rating procedures

All pre-service teachers were videotaped three times at the beginning of the internship (cycles 1–3; first 3 weeks) and three times at the end (cycles 4–6; last 3 weeks), so that each participant had six videotaped lessons of his or her teaching. All six lessons per pre-service teacher were conducted in the same class (elementary classrooms, class level one, two, or three). Between the third and fourth filmed lesson, the pre-service teachers received three visits from their university supervisors. The supervisors observed the lessons and discussed them afterwards with the pre-service teachers. The discussions were generally about the con-

tent of the lessons and the performance of the pre-service teachers, but did not focus on CLASS criteria. In total, the data basis for this study consists of 66 filmed PE lessons. All the videos were rated by two licensed observers (independently from each other) using the CLASS K-3 (Pianta et al., 2008).

All 10 sub-dimensions were rated by using indicators and behavioral markers. See Table 1 for brief descriptions of the CLASS domains, the sub-dimensions, and examples of indicators and markers. All indicators and markers are described in detail in the CLASS K-3 Manual (Pianta et al., 2008). Moreover, the manual describes when processes in the classroom lead to a high, mid, or low rating in each dimension. All 10 sub-dimensions are rated from 1 = low to 7 = high, with 1 or 2 indicating low quality, 3, 4, or 5 indicating mid-range of quality, and 6 or 7 indicating high quality. To calculate the three domain scores (classroom organization, emotional support, and instructional support), the average dimension scores that fall into each domain are added up and divided by the number of dimensions in that domain.

Interrater reliability

Since PWO was the primary indicator of interrater reliability reported in the CLASS manual, it was also considered as the reliability indices in this study. Reliability indices for all subdimensions are reported in Table 2 and 3. The overall interrater reliability ranged from 82 to 100%.

CLASS observations were conducted in cycles. A cycle consisted of a 20-min observation period at the beginning of the lesson and a 10-min rating period. Agreement calculations were computed by cycle, as the six 20-min cycles were scored independently. All ratings were conducted based on the videos. Scores from the PWO analysis ranged from 82 to 100% for cycle 1, from 73 to 100% for cycle 2, from 91 to 100% for cycle 3, from 73 to 100% for cycle 4, from 82 to 100% for cycle 5 and from 91 to 100% for cycle 6.

Table 2 Interrater agreement for the CLASS dimensions calculated by dimensions at the beginning (T1) and at the end of the long-term internship (T2)

	T1	T2	T1	T2
	Exact		±1	
Positive Climate	33	45	91	97
Negative Climate	33	39	97	94
Teacher Sensitivity	30	30	97	91
Regard for Student Perspectives	03	18	97	85
Behavior Management	30	24	82	88
Productivity	39	33	94	91
Instructional Learning Format	24	48	82	94
Concept Development	39	48	94	94
Quality of Feedback	33	51	100	94
Language Modeling	42	42	97	94

All scores are reported in percent agreement

Table 3 Interrater agreement for the CLASS dimensions calculated by cycles

	Cycle 1		Cycle 2		Cycle 3		Cycle 4		Cycle 5		Cycle 6	
	Exact	±1	Exact	±1	Exact	±1	Exact	±1	Exact	±1	Exact	±1
Positive Climate	27	91	45	91	27	91	45	91	64	100	27	100
Negative Climate	36	91	45	100	18	100	45	100	45	91	27	91
Teacher Sensitivity	27	100	27	91	36	100	36	91	18	91	36	91
Regard for Student Perspectives	0	100	9	91	0	100	9	73	9	82	36	100
Behavior Management	27	91	27	73	36	91	27	82	18	91	27	91
Productivity	55	100	18	82	45	100	27	82	9	91	55	100
Instructional Learning Format	18	82	18	73	36	91	45	91	55	91	45	100
Concept Development	55	100	45	91	18	91	55	82	27	100	64	100
Quality of Feedback	45	100	36	100	18	100	36	91	45	91	73	100
Language Modeling	36	91	36	100	55	100	27	91	45	91	55	100

All scores are reported in percent agreement

Results

Development of physical education pre-service teachers' teaching performance

Analyses were conducted using the SPSS software (version 26; IBM Corp., Armonk, NY, USA). Means, standard deviations, and effect sizes are reported in **Table 4**. The results demonstrate that *emotional support* and *classroom organi-*

zation were rated as mid-range at the beginning of the long-term internship (cycles 1–3), whereas *instructional support* was rated low. The same pattern can be reported for the end of the internship (cycles 4–6). To investigate whether the change is significant, t-tests for paired samples were conducted and effect sizes for repeated measures calculated. In accordance with Cohen (1988), values <0.4 were regarded as a small effect, values between 0.5 and 0.7 as a medium effect,

and values 0.8 as a large effect. None of the core domains increased significantly over time. Regarding the sub-dimensions, results show that the PE pre-service teachers in this study indicate only rare or no negative affects, which resulted in low scores for *Negative Climate*, both at the beginning and at the end of the long-term internship. However, results show significantly less *Negative Climate* at the end of the long-term internship with a high effect size (see **Table 4**). *Positive Climate*, *Teacher Sensitivity*, *Behavior Management*, *Productivity* and *Instructional Learning Format* had an average score of four to five at both measurement points. In these dimensions, the pre-service teachers fit in the mid-range description of the CLASS K-3 manual, with one or two indicators in the high range. The ratings here show that the dimension *Regard for Student Perspectives* was rated somewhat lower than the aforementioned dimensions. Descriptively, participants had the highest increase in this dimension with a medium effect size (see **Table 4**). However, results of t-tests for paired samples indicated no significant effect. All dimensions of the domain *Instructional Support* were rated in the low range, both at the beginning and at the end of the long-term internship.

Discussion

This study aimed to offer some insights into the development of PE pre-service teachers' teaching performance during a 5-month internship. To measure teacher-student interactions of pre-service teachers, the CLASS K-3 observation instrument was applied. Results revealed no significant changes in the teaching performance of PE pre-service teachers over the course of 5 months, with one exception for the subdimension of *Negative Climate*. The significant change in this dimension could be due to the fact that pre-service teachers were stressed and nervous during their first teaching lessons, which could have resulted in slightly higher ratings. At the end of the internship, pre-service teachers could have been more used to interactions with their students. However, these interpretations need to be further investigated in

Table 4 Means (M), standard deviations (SD), and effect size for repeated measures (*d*) for CLASS dimension and domain scores at the beginning (T1) and at the end of the long-term internship (T2)

	T1 (cycle 1–3)		T2 (cycle 4–6)		Δ	t	Df	p	d
	M	SD	M	SD					
Classroom Organization	4.85	0.22	4.90	0.47	0.05	0.44	10	0.67	0.14
Behavior Management	4.97	0.34	4.99	0.56	0.02	0.11	10	0.91	0.04
Productivity	4.86	0.49	5.00	0.59	0.14	0.58	10	0.58	0.26
Instructional Learning Format	4.71	0.38	4.73	0.48	0.02	0.12	10	0.91	0.04
Emotional Support	4.84	0.27	4.97	0.47	0.13	1.08	10	0.31	0.34
Positive Climate	4.64	0.72	4.77	0.60	0.14	0.84	10	0.42	0.20
Negative Climate	2.06	0.24	1.40	0.37	-0.67	5.93	10	<0.01	2.21
Teacher Sensitivity	4.74	0.33	4.71	0.50	-0.03	-0.18	10	0.86	0.07
Regard for Student Perspectives	3.39	0.41	3.80	1.03	0.41	1.38	10	0.20	0.52
Instructional Support	1.85	0.38	1.81	0.37	0.05	0.41	10	0.69	0.11
Concept Development	1.76	0.37	1.67	0.32	-0.09	-0.80	10	0.44	0.26
Quality of Feedback	1.94	0.54	1.79	0.54	-0.15	-1.13	10	0.28	0.28
Language Modeling	1.86	0.46	1.97	0.46	0.11	0.63	10	0.54	0.24

1 or 2 = low, 3, 4, or 5 = mid-range, 6 or 7 = high (Pianta et al., 2008)

Negative Climate scores are reverse and were recoded for the overall score of emotional support

future research by using questionnaires or by conducting interviews with the pre-service teachers after each lesson. The recorded performance values of pre-service teachers in our study are similar to findings in other studies with experienced teachers in non-PE classrooms (Longobardi et al., 2018; Pianta et al., 2008). Like experienced teachers, the pre-service teachers in PE classrooms score in the upper mid-range in the domains of *Emotional Support* and *Classroom Organization*. Only in the domain *Instructional Support*, the pre-service teachers achieved lower values by comparison. They score in the low-range while experienced teachers in the studies mentioned achieved ratings in the middle-range. It is surprising that the pre-service teachers in our study performed equally well at the beginning of their internship as experienced teachers in the domains of *Emotional Support* and *Classroom Organization*. In the case of *Classroom Management*, one possible explanation is that this topic is discussed in detail during different university seminars in preparation for the internship and that the specific professional expertise is easily transferable

to actual classroom performance. Since the university seminars usually do not cover aspects of *Emotional Support*, it can be assumed that these performatively visible competences of PE pre-service teachers are acquired and consolidated in non-academic contexts ahead of time, e.g., in care and learning situations with children in private settings or through child-related coaching activities in sport clubs, for example. The fact that scores in *Emotional Support* and *Classroom Organization* were already relatively high at the beginning of the internship of pre-service PE teachers may explain why the impact of the internship in these areas of teacher performance is comparatively low. To increase performance quality in these areas, targeted and recurring feedback over longer periods of time would probably be necessary.

The low scores of PE pre-service teachers in the domain of *Instructional Support*, on the other hand, indicate that this is the most promising area of developmental growth for pre-service teachers at this stage. Since performative competences in this area are obviously acquired neither through university seminars nor in non-academic contexts beforehand,

they should be targeted in teacher training during the established long-term internships. Yet, the findings here demonstrate that no significant developmental changes in teaching performance occur during the internship of pre-service PE teachers. However, multiple possible explanations for these findings and implications for possible changes in these sequences of teacher training seem conceivable.

It is possible that developmental changes to the pre-service PE teachers teaching quality occur on the comparatively structural level of lesson planning and preparation that do not yet “trickle down” to the actual teacher–student interactions. Improved lesson planning as well as routines could lead to a situational cognitive discharge later on, which experienced teachers can use to focus their attention on individual learning processes of the students and give specific instructional support. Hence, it is possible that a 5-month internship is too short to see learning effects on a performance level of pre-service teachers. Further assessment of pre-service PE teachers’ performance during the second phase of teacher training and at the beginning of their teaching career could provide a deeper insight here. Another possibility is that pre-service PE teachers receive too little or non-specific feedback related to the domain of *Instructional Support* during their reflective talks with school and university mentors after their lessons. Vogler, Messmer, and Allemann (2017) come to similar conclusions regarding the pedagogical content knowledge development of pre-service PE teachers. In the case of the present study, these talks followed no systematic scheme, but focused in a situational manner on different aspects that stood out to the mentors in the observed lesson. Reflective talks that identify and select individual areas of growth and target these over longer periods of time during the internship could be more effective in this regard (Shavelson, 2010, 2013).

Furthermore, unfocused and non-systematic reflective talks in combination with a large extent of compulsory teaching in the internship could lead

to *learning by doing* or *trial and error* strategies. In the case of novices, these might only lead to success by chance and not to systematic development. Moreover, this would support existing critical views on the effects of internships in teacher education (Hascher, 2011). Whereas most research underlines positive effects on pre-service teachers' competency development, there are empirical findings that long-term internships with a high teaching obligation can have a deprofessionalizing effect. Due to the increasing and prolonged school practice during studies, currently existing (possibly problematic) school practice is reproduced without more critical reflection (Hascher, 2012; Vogler et al., 2017; Weyland & Wittmann, 2015). This self-reproduction of the school system contradicts the idea of a critical-constructive university education for teachers and can be regarded as very problematic in consideration of the discussion on the quality of teaching.

Limitations

This study faced some limitations that encourage future research in this field. First, the sample size is relatively small. Pre-service teachers needed to volunteer to participate in this study and, moreover, data privacy laws require parental consent of the school students. Therefore, video-based studies rarely involve large sample sizes (e.g., Major and Watson, 2018). Second, the CLASS K-3 was used, which is a highly implemented instrument to measure the quality of teacher-student interactions. CLASS K-3 is supposed to be subject-independent. However, as far as the authors know, there are no studies using CLASS K-3 in PE classrooms. As mentioned before, there is an ongoing discussion about a subject-specific supplementation for the domain of *Instructional Support* for PE classrooms, which has not resulted in an adapted observation instrument as yet. Therefore, the comparisons drawn here between scores in this domain of pre-service PE teachers and experienced teachers in other CLASS studies (in non-PE classrooms) may be limited in their validity. For future research, it would be of in-

terest to adapt and validate CLASS K-3 for the rating of PE lessons. Whereas most subjects in schools aim at higher-order thinking skills and cognitive activation of students, PE lessons have a different pivotal learning goal (Vogler et al., 2017). In PE, *hands on materials* and inductive physical-motor learning processes play a central role, where teacher feedback is not always necessary but provided by the reactions of materials themselves. Third, one has to take into consideration that a video-camera in a classroom may have an effect on the behavior of pre-service teachers during the lesson. Fourth, other factors related to pre-service teachers' teaching performance were not measured. For example, contextual factors, such as the number of informal feedback occasions in school, subject knowledge, motivation, or prior teaching experiences might have had an effect on the results of the study. Therefore, future research should include these variables.

Conclusions

This study is the first to measure the development of PE pre-service teachers' teaching performance during a long-term internship in Germany. The results showed no significant changes in the teaching performance of PE pre-service teachers over the course of 5 months (with one exception for the subdimension of *Negative Climate*). These results have several implications, both for teacher educators and for researchers in the field of PE and video-based measurement. School internships may need to be longer than 5-month periods to archive an increase in actual teaching performance and interactional teaching quality of pre-service teachers. Feedback sessions between mentors and pre-service teachers after lessons should be more structured and should focus on the quality of teacher-student interaction. Moreover, it is of interest for researchers to validate the CLASS instruments for use in PE lessons. In order to validate the CLASS instrument for PE with its specific features, studies should be carried out in which CLASS performances of pre-service teachers during internships

in PE and a regular subject are collected and compared. Thus, one could examine whether the interaction quality of teachers with CLASS can be captured subject-independent and whether the interaction quality of pre-service teacher does not rise over the course of a 5-month internship in any of the observed classes. Corresponding data could support the insights of this study that an extended practice phase and focused mentoring talks after the lessons are necessary to induce an increase in the interaction quality of pre-service PE teachers.

Corresponding address



Dr. Steffen Greve
Institut für Bewegung, Sport
und Gesundheit, Leuphana
Universität Lüneburg
Universitätsallee 1,
21335 Lüneburg, Germany
steffen.greve@leuphana.de

Funding. Open Access funding provided by Projekt DEAL.

Compliance with ethical guidelines

Conflict of interest. S. Greve, K.E. Weber, B. Brandes, and J. Maier declare that they have no competing interests.

All procedures performed in studies involving human participants or on human tissue were in accordance with the ethical standards of the institutional and/or national research committee and with the 1975 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arnold, K.-H., Gröschner, A., & Hascher, T. (Eds.). (2014). *Schulpraktika in der Lehrerbildung [Pedagogical field experiences in teacher education]*. Münster: Waxmann.
- Baumgartner, M. (2018). Performanzfortschritt in der Lehrerausbildung durch die Arbeit an eigenen video- und textbasierten Unterrichtsfällen? Eine Interventionsstudie zur Verbesserung des Feedbacks bei angehenden Sportlehrkräften [Progress in performance in teacher training through the work on own video- and text-based teaching cases? An intervention study in relation to the improvement of feedback for trainee PE teachers]. *Zeitschrift für Erziehungswissenschaft*, 21(6), 1135–1155.
- Beck, C., & Kosnik, C. (2002). Components of a good practicum placement: student teacher perceptions. *Teacher Education Quarterly*, 29, 81–98.
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, 25(2), 166–176.
- Burchinal, M., Vernon-Feagans, L., Vitiello, V., Greenberg, M., & Family Life Project Key Investigators (2014). Thresholds in the association between child care quality and child outcomes in rural preschool children. *Early Childhood Research Quarterly*, 29(1), 41–51.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn.). Hillsdale: Erlbaum.
- Cohen, E., Hoz, R., & Kaplan, H. (2013). The practicum in preservice teacher education: a review of empirical studies. *Teaching Education*, 24(4), 345–380.
- Cothran, D., & Kulinna, P. (2014). Classroom management in physical education. In E. Emmer & E. Sabornie (Eds.), *Handbook of classroom management* (2nd edn., pp. 239–260). New York: Routledge.
- Curby, T. W., & Chavez, C. (2013). Examining CLASS dimensions as predictors of pre-K children's development of language, literacy, and mathematics. *NHSA Dialog*, 16, 1–17.
- Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2009). The relations of observed pre-K classroom quality profiles to children's achievement and social competence. *Early Education and Development*, 20(2), 346–372.
- Downer, J. T., Lopez, M. L., Grimm, K. J., Hamagami, A., Pianta, R. C., & Howes, C. (2011). Observations of teacher child interactions in classrooms serving Latinos and dual language learners: applicability of the Classroom Assessment Scoring System in diverse settings. *Early Childhood Research Quarterly*, 27(1), 21–32.
- Friedman, I. A. (2006). Classroom management and teacher stress. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice, and contemporary issues* (pp. 925–945). Mahwah: Erlbaum.
- Gerlach, E. (2005). Prima Klima? Einflussgrößen und Effekte. [Great climate? Influencing factors and effects]. *Sportunterricht*, 54, 243–247.
- Gerlach, E., Kussin, U., Brandl-Bredenbeck, H. P., & Brettschneider, W.-D. (2006). Der Sportunterricht aus Schülerperspektive. [Physical education from a student's perspective. In Deutscher Sportbund (Ed.), *Die SPRINT-Studie. Eine Untersuchung zur Situation des Schulsports in Deutschland* (pp. 107–144). Aachen: Meyer & Meyer.
- Greve, S. (2013). *Lernen durch Reflektieren im Sportspiel – Möglichkeiten im Vermittlungsprozess im Sportunterricht am Beispiel Handball. [Learning by Reflecting in Sports Games—Possibilities in the Teaching Process in Physical Education using the Example of Team-Handball]*. Berlin: Logos.
- Gröschner, A., Klab, S., & Dehne, M. (2018). Lehrer-Schüler Interaktion im Praxissemester lernen? Effekte des videobasierten Peer Coaching auf die Kompetenzeinschätzung [Learning about teacher-student interaction during a teaching practicum? Effects of video-based peer coaching on self-assessment of competence]. *Zeitschrift für Hochschulentwicklung*, 13(1), 45–67.
- Gröschner, A., Müller, K., Bauer, J., Seidel, T., Prenzel, M., Kauper, T., & Möller, J. (2015). Praxisphasen in der Lehrerausbildung – Eine Strukturanalyse am Beispiel des gymnasialen Lehramtsstudiums in Deutschland [Field experiences in teacher training: a structural analysis of secondary teacher education in Germany]. *Zeitschrift für Erziehungswissenschaft*, 18(4), 639–665.
- Gröschner, A., Schmitt, C., & Seidel, T. (2013). Kompetenzentwicklung im Praktikum? Entwicklung eines Instruments zur Erfassung von Kompetenzen und Ergebnisse einer Befragung von Lehramtsstudierenden im betreuten Blockpraktikum. [Competence development during the internship? Development of an instrument to record competences and results of a survey of pre-service teachers in the supervised block internship]. *Zeitschrift für pädagogische Psychologie*, 27(1), 77–86.
- Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching and Teacher Education*, 26(4), 1094–1103.
- Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: associations with preschool children's development. *Child Development*, 85(3), 1257–1274.
- Hamre, B. K., Pianta, R. C., Downer, J. T., & Mashburn, A. J. (2008). Teachers' perceptions of conflict with young students: looking beyond problem behaviors. *Social Development*, 17(1), 115–136.
- Harris, A. (1998). Effective teaching: a review of the literature. *School Leadership & Management*, 18(2), 169–183.
- Hascher, T. (2011). Vom „Mythos Praktikum“ ... und der Gefahr verpasster Lerngelegenheiten. [The “internship myth” ... and the danger of missed learning opportunities]. *Journal für LehrerInnenbildung*, 11(3), 8–16.
- Hascher, T. (2012). Lernfeld Praktikum – Evidenzbasierte Entwicklungen in der Lehrer / innenbildung [Learning setting student teaching—evidence based developments in teacher education]. *Zeitschrift für Bildungsforschung*, 2(2), 109–129.
- Hascher, T., & De Zordo, L. (2015). Langformen von Praktika. Ein Blick auf Österreich und die Schweiz. [Long forms of internships. A look at Austria and Switzerland]. *Journal für LehrerInnenbildung*, 15(2), 22–32.
- Hattie, J. A. C. (2009). *Visible Learning. A synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.
- Heemsoth, T. (2014). Unterrichtsklima als Mediator des Zusammenhangs von Klassenführung und Motivation im Sportunterricht. [Teaching climate as mediator of the connection between classroom management and motivation in physical education]. *Psychologie in Erziehung und Unterricht*, 61, 203–215.
- Herrmann, C., Seiler, S., Pühse, U., & Gerlach, E. (2015). „Wie misst man guten Sportunterricht?“ – Erfassung der Dimensionen von Unterrichtsqualität im Schulfach Sport. [“How do you measure good gym class?”—To measure the dimensions of teaching quality in physical education]. *Zeitschrift für Sportpädagogische Forschung*, 3(1), 6–26.
- Jaakkola, T., & Liukkonen, J. (2006). Changes in students' self-determined motivation and goal orientation as a result of motivational climate intervention within high school physical education classes. *International Journal of Sport and Exercise Psychology*, 4, 302–324.
- Jamil, F. M., Sabol, T. J., Hamre, B. K., & Pianta, R. C. (2015). Assessing teachers' skills in detecting and identifying effective interactions in the classroom. Theory and Measurement. *The elementary school journal*, 115(3), 407–432.
- Klieme, E., & Rackoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. [Empirical teaching research and didactics. Outcome-oriented measurement and process quality of teaching]. *Zeitschrift für Pädagogik*, 54(2), 222–237.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster: Waxmann.
- König, J. (2015). Measuring classroom management expertise (CME) of teachers: a video-based assessment approach and statistical results. *Cogent Education*, 2(1), 991178.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart, Winston.
- Küster, O. (2008). *Praktika und ihre Lernpotentiale in der Lehrerbildung. Eine längsschnittliche Videostudie zur Untersuchung der Entwicklung unterrichtlicher Handlungskompetenzen in verlängerten Praxisphasen*. Weingarten: Universitätsdruck. Internships and their learning potential in teacher education. A longitudinal video study investigating the development of teaching competences in extended practice phases
- Linka, T. & Gerlach, E. (2019). Klassenmanagement im Sportunterricht – Ergebnisse zur Evaluation der ersten Phase eines videogestützten Seminarkonzepts für angehende Lehrkräfte. [Classroommanagement in Physical Education—Results of the evaluation of the first phase of a video-based seminar concept for pre-service teachers]. In M. Hartmann, R. Laging & C. Scheinert (Eds.), *Professionalisierung in der Sportlehrer*innenbildung. Konzepte und Forschungen im Rahmen der „Qualitätsoffensive Lehrerbildung“* (S. 97–107). Hohengehren: Schneider Verlag.
- Lipowsky, F., Rakoczy, K., Drollinger-Vetter, B., Klieme, E., Reusser, K., & Pauli, C. (2009). Quality of geometry instruction and its short-term impact on students' understanding of Pythagorean

- Theorem. *Learning and Instruction*, 19(6), 527–537.
- Longobardi, C., Pasta, T., Marengo, D., Prino, L.E., & Settanni, M. (2018). Measuring quality of classroom interactions in Italian primary school: structural validity of the CLASS K-3. *The Journal of Experimental Education*. <https://doi.org/10.1080/00220973.2018.1533795>.
- Lopez, J.M.O., Santiago, M.J., Godás, A., Castro, C., Villardefrancos, E., & Ponte, D. (2008). An Integrative Approach to burnout in secondary school teachers: Examining the role of student disruptive behavior and disciplinary issues. *International Journal of Psychology and Psychological Therapy*, 8(2), 259–270.
- Lotz, M., Gabriel, K., & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung [Lowly and Highly Inferential Procedures of Classroom Observation. Analyses on their reciprocal validation]. *Zeitschrift für Pädagogik*, 59(3), 357–380.
- Maier, M.F., Vitiello, V.E., & Greenfield, D.B. (2012). A multilevel model of child- and classroom-level psychosocial factors that support language and literacy resilience of children in Head Start. *Early Childhood Research Quarterly*, 27(1), 104–114.
- Major, L., & Watson, S. (2018). Using videos to support in-service teacher professional development: the state of the field, limitations and possibilities. *Technology, Pedagogy and Education*, 27(1), 49–68.
- Mashburn, A.J., Pianta, R.C., Hamre, B.K., Downer, J.T., Barbarin, O.A., Bryant, D., & Burchinal, M. (2008). Measures of classroom quality in pre-kindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732–749.
- Miethling, W.-D., & Krieger, C. (2004). *Schüler im Sportunterricht: die Rekonstruktion relevanter Themen und Situationen des Sportunterrichts aus Schülersicht (RETHESIS)*. [Students in physical education: the reconstruction of relevant topics and situations in physical education from the student's point of view]. Schorndorf: Hofmann.
- Niederkofler, B., & Amesberger, G. (2016). Kognitive Handlungsrepräsentationen als Strukturgrundlage zur Definition von kognitiver Aktivierung im Sportunterricht. [Cognitive action representations as structural basis for the definition of cognitive activation in physical education]. *Sportwissenschaft*, 46(3), 188–200.
- La Paro, K.M., & Pianta, R.C. (2000). Predicting children's competence in the early school years: a meta-analytic review. *Review of Educational Research*, 70, 443–484. <https://doi.org/10.3102/00346543070004443>.
- La Paro, K.M., Pianta, R.C., & Stuhlman, M. (2004). The classroom assessment scoring system: findings from the prekindergarten year. *The Elementary School Journal*, 104(5), 409–426. <https://doi.org/10.1086/499760>.
- Pianta, R.C., & Hamre, B.K. (2008). Conceptualization, measurement, and improvement of classroom processes: standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Pianta, R.C., Hamre, B.K., & Mintz, S.L. (2012). *The CLASS-secondary manual*. Charlottesville: University of Virginia.
- Pianta, R.C., La Paro, K., & Hamre, B.K. (2008). *Classroom assessment scoring system. Manual K-3*. Baltimore: Paul H. Brookes.
- Pianta, R.C., La Paro, K.M., Payne, C., Cox, M.J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal*, 102(3), 225–238. <https://doi.org/10.1086/499701>.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality. *ZDM Mathematics Education*, 50(3), 407–426.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Sandilos, L.E., & DiPerna, J.C. (2011). Interrater reliability of the classroom assessment scoring system—pre-K (CLASS pre-K). *Journal of Early Childhood and Infant Psychology*, 7, 65–85.
- Scherler, K. (2008). *Sportunterricht auswerten. [Evaluate physical education]* (2nd edn.). Hamburg: Czwalina.
- Schwarz, R. (2009). *(Neuro-)Kasuistische Sportlehrerbildung – Von Fällen des Sportunterrichts zum Modell pädagogischer Dissonanz: Lehrer werden von Fall zu Fall*. [(Neuro) casuistic physical education in teacher education—From cases of physical education to the model of educational dissonance: Become a teacher case by case]. Hamburg: Czwalina.
- Shavelson, R.J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training*, 2(1), 41–63.
- Shavelson, R.J. (2013). On an approach to testing and modeling competence. *Educational Psychologist*, 48(2), 73–86.
- Tapp, K., Mache, C., & McLendon, T. (2012). Assessing student performance in field education. *The Field Educator*, 2.2. <http://fieldeducator.simmons.edu/article/assessing-student-performance-in-field-education/#more-1058>. Accessed: 3 October 2019
- Taut, S., & Rakoczy, K. (2016). Observing instructional quality in the context of school evaluation. *Learning and Instruction*, 46, 45–60.
- Vogler, J., Messmer, R., & Allemann, D. (2017). Das fachdidaktische Wissen und Können von Sportlehrpersonen (PCK-Sport). *German Journal of Exercise and Sport Research*, 47, 335–347. <https://doi.org/10.1007/s12662-017-0461-4>.
- Weyland, U., & Wittmann, E. (2015). Langzeitpraktika in der Lehrerausbildung in Deutschland: Stand und Perspektiven. [Long-term Internships in Teacher Education in Germany: Status and Perspectives]. *Journal für LehrerInnenbildung*, 15(1), 8–21.
- White, M.C. (2018). Rater performance standards for classroom observation instruments. *Educational Researcher*, 47(8), 492–501.
- Wolff, C.E., Jarodzka, H., & Boshuizen, H. (2017). See and tell: differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education*, 66, 295–308.
- Xu, Y., Chin, C., Reed, E., & Hutchinson, C. (2014). The effects of a comprehensive early literacy project on preschoolers' language and literacy skills. *Early Childhood Education Journal*, 42(5), 295–304.
- Seiler, S. (2016). „What works in Physical Education?“ oder „Wer färbt das Ei gelb?“ Mehrperspektivische Betrachtung von Wirkungseinflüssen auf Lernleistungen im Sportunterricht. [„What works in physical education?“ or „Who turns the egg yellow?“ Multi-perspective consideration of the effects on learning performance in physical education]. Unveröffentlichte Dissertation. Potsdam: Universität Potsdam.