

Atle Mjåtveit

Universitetet i Stavanger

Rune Giske

Universitetet i Stavanger

DOI: <http://dx.doi.org/10.5617.adno.7910>

Formative assessment in Physical Education: An analysis of teaching and mentoring in preservice training practice.

Abstract

This study investigates the prevalence and quality of formative assessment in Physical Education pre-service teachers' (PSTs') teaching lessons. Moreover, we draw attention to how mentor teachers emphasise formative assessment in mentoring sessions. Data collection was conducted during the PSTs' field practice. A multi-case design was chosen and the study included 15 PSTs and 4 mentors. The PSTs were in their second year (fourth semester) of the ITE-program. The participants were in four different groups (School 1, 2, 3 and 4) each consisting of three/four PSTs and one mentor. Results revealed brief instances of formative assessment in PSTs' teaching and in mentors' comments during mentoring sessions. The findings indicated PSTs' limited capability to identify critical elements in an exercise, and partly a lack of ability to differentiate their teaching. Furthermore, we found little evidence of feedback loops between PSTs and students that stimulated learning, and/or promoted students' understanding and participation. The mentoring sessions had a strong focus on organising the lessons and effective time utilisation. There were no discussions of how the PSTs could maximise students' engagement for learning. A possible explanation for the PSTs' modest use of formative assessment may be that campus teaching has failed to introduce the topic in a proper way. Moreover, there appears to be a need to strengthen cooperation between campus and partner schools to stimulate the importance of formative assessment.

Keywords: assessment for Learning, teacher education, physical education, mentoring

Vurdering for læring i kroppsøving: En analyse av undervisning og veiledning i praksisopplæring.

Sammendrag

Denne studien undersøker forekomsten av, og kvaliteten på kroppsøvingslærerstudenters bruk av formativ tilbakemelding i sin egen undervisning. Dessuten ønsker en å avdekke i hvilken grad praksislærerne vektlegger betydningen av formativ tilbakemelding i veiledningssamtaler. Datainnsamlingen ble gjort i løpet av lærerstudentenes praksisperiode. Det ble benyttet et multi-case design og studien omfatter 15 lærerstudenter og 4 praksislærere. Lærer-studentene var i sitt 2. studieår (4.

semester), og hadde valgt kroppsøving som fordypningsfag. Deltakerne var inndelt i fire ulike grupper (Skole 1, 2, 3 og 4) som var sammensatt av tre/fire lærerstudenter og en praksislærer. Resultatene avdekket et beskjedent fokus på formativ tilbakemelding både i lærerstudentenes undervisningsøkter og praksislærernes veiledningsøkter. Funnene indikerer lærerstudentenes begrensede evne til å identifisere kritiske elementer i en aktivitet, samt evne til å øke elevenes læring gjennom å tilrettelegge for differensiering. Videre fant en lite bruk av tilbakemeldingssløyer mellom lærerstudentene og elevene som stimulerte læring, eller fremmet elevens forståelse og deltakelse. Veiledningsøktene hadde et sterkt fokus på organisering og effektiv tidsutnyttelse. Det var ingen diskusjon om hvordan lærerstudentene kunne stimulere elevenes engasjement for læring gjennom tydelig presentasjon av læringsmål. Den mest åpenbare forklaringen er at formativ tilbakemelding ikke er presentert på en tilstrekkelig kontekstualisert måte i kroppsøving-lærerutdanningen. Videre ser det ut til å være et behov for å styrke samarbeidet mellom campus og praksisfelt for å stimulere betydningen av formativ tilbakemelding.

Nøkkelord: vurdering for læring, lærerutdanning, kroppsøving, veiledning

Introduction

Learning to teach is a complex and challenging task for pre-service teachers (PSTs), and designing programmes that support such learning through instruction, practice, and mentoring is also challenging for Initial Teacher Education (ITE) programmes (Buscà, Ambròs, & Bursset, 2017; Darling-Hammond, 2017; Korthagen, 2010; Livingston & Flores, 2017).

Many ITE programmes offer more than ten possible choices of school subjects (e.g. mathematics, languages, physical education (PE), arts). This makes ITE programmes especially complex. Teacher educators are diverse, representing many subject areas, and the need for collaboration is especially daunting in light of researchers' recommendations to construct coherent programmes across subject areas and across campus and partner schools for field practice (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Darling-Hammond, 2006; Hammerness, 2013). The importance of extensive and intensely collaboration with schools is also emphasized (Darling-Hammond, 2006), and this means that colleges and universities must collaborate closely with field practice schools and provide follow-up for all PSTs.

The quality of mentoring that goes on during field practice has emerged as a key aspect of teacher education (Eriksson, 2017; Orland-Barak, 2014; Sundli, 2007). A mentor, in our study, is a school teacher who works at a partner school where PSTs have their field practice during ITE. Behets and Vergauwen (2006) argue that mentors are the most influential factor shaping PSTs' behaviour in teacher education, and one important way that mentors influence PSTs' learning is through the feedback they give.

The Norwegian ITE programmes are nationally regulated, with concurrent BA degrees that qualify for teaching in either grades 1-7 or grades 5-10. Physical education can be chosen as an elective, and can be either 30 ECTS or 60 ECTS. Since 2015, the four-year programmes have become five-year integrated and concurrent MA degrees (no students have graduated from the 5-year programmes yet). The majority of teaching and learning (90%) takes place on campus (Sundli, 2007). Field practice is a total of 100 days spread out over four years. Partner schools are paid to be partners in teacher education and to provide qualified mentors. During the field practice, PSTs are organised in groups of three or four and each group is assigned a mentor. This is a common form of organization in Norwegian field practice, but not a national regulation. National guidelines also open up for individual field practice.

Studies indicate that Norwegian PE teacher educators play a very distant role during field practice, and that it is mostly left up to the PSTs to make connections between theory and practice (Moen & Standal, 2016; Standal, Moen, & Moe, 2014). Consequently, researchers call for improved cooperation between the parties involved to support PSTs' learning.

Assessment and feedback in PE

Qualitative feedback is vital in a learning process (Sadler, 2010; Silverman, Tyson, & Krampitz, 1992). Whereas assessment is mostly related to purpose and provides a foundation for evaluating and documenting students' performance, feedback intends to help students recognize quality and to lead learning forward (Torkildsen & Erickson, 2016). However, Havnes, Smith, Dysthe, and Ludvigsen (2012) state that systematic use of feedback to support students' learning is a weak element in school. They also point out that teachers lack systematic strategies to use the feedback they have given students in ways that influence their future teaching. Students can also react differently to the same type of feedback (Cloes, Denève, & Piéron, 1995). Feedback must be tailored to the individual student and problem faced (Silverman, et al., 1992). In a study of Physical Education Teacher Education (PETE) students' feedback practice, Metzler, Tjeerdsma, and Mozen (2000) find that students learn more from feedback that includes some indication of which part of an observed practice contributed to its success or failure. This can be illustrated by an example from the high jump. In order to succeed, *run-up, takeoff, flight and landing*, must be mastered. It is important for the teacher to give feedback directly related to these phases, so that the student gets to know which part of the exercise is mastered / not mastered.

Assessment for Learning (AfL) is defined as the process of seeking and interpreting evidence to decide where the students are in their learning, where they need to go and how best to go there (Hattie & Timperley, 2007; Lorente-Catalán & Kirk, 2016). According to Chng and Lund (2018) the term "formative assessment" is often embedded in the term "assessment for learning". The major principles in formative assessment is to identify students' weakness and strengths,

to enhance students' motivation and metacognition, and to provide feedback to enhance teaching and learning (William, Lee, Harrison, & Black, 2004). In educational literature, the focus on formative assessment has grown dramatically, partly because research indicates that good AfL is one of the most effective instructional tools to drive student learning forward (Leirhaug, 2016). Hamodi, López-Pastor, and López-Pastor's study (2017) finds that formative assessment is not used much during ITE, but is highly valued because it enables deeper learning and makes it possible to reproduce practices that were experienced as valuable during ITE.

In many ways, formative assessment represents a break with the established assessment practice in physical education (Lorente-Catalán & Kirk, 2016). The assessment mission in PE has been regarded as problematic for some time (Tolgfors & Öhman, 2016). When the stated criteria are inconsistent with how the grading is done, this affects the learning-teaching process since the assessment sends out a different message regarding what is important to learn (Svennberg, Meckback, & Redelius, 2014).

Teachers' feedback practices, according to Crum (2013), are closely related to their ideology. He presents two prevailing ideologies in European PE, *education through the physical* and *training of the physical*. In *education through the physical*, the subject is considered to be a tool to achieve health benefits, social learning and recreation. In *training through the physical*, the human body is seen as a machine that must be kept in good shape through physical exercise. According to Crum (2013) legitimisation characterised by pretentious claims for outcomes, mind-body dualism and non-teaching practices is a threat for good teaching and relevant learning outcomes.

In order to reverse this trend, Crum (2013) recommends providing students with a solid basis of movement vocabulary (*movement culture*) and creating situations that provide possibilities to reach goals on different performance levels. Furthermore, he strongly argues for the importance of strengthening PETE. This can be done by abandoning the old ideologies with their non-teaching practices (Crum, 1985; Fishburne & Borys, 1987) and that the faculty reinforces the idea that PE should be a teaching – learning enterprise (Crum, 2013).

For teachers linked to the *education through the physical* and *training of the physical* ideologies, their beliefs will manifest themselves through non-teaching perspectives (Crum, 1985), focusing on intensity and effort, and the importance of being physically active, while largely ignoring how the activity is performed. On the other hand, the learning-oriented teacher connected to a *movement culture* ideology, will focus on what the student learns, and how the student solves the task (Crum, 2013).

In Hattie and Timperley's (2007) model for developing skills and understanding, the teacher's role is to support and facilitate the students' development processes by presenting clear, achievable and specific learning goals. Furthermore, the teacher supports the students through effective teaching

strategies and feedback (Gamlem & Munthe, 2014). Hattie and Timperley's model (2007) must be considered in relation to four major levels: task level (whether work is correct or incorrect), process level (directly aimed at the processing of information), self-regulated level (including greater skill in self-evaluation) and self-level (often unrelated to performance on the task). Feedback on task level, process level and self-regulated level are highlighted as the most effective because they do not draw attention away from the task (Hattie & Timperley, 2007).

Leirhaug and Annerstedt (2015) examined the implementation of Assessment for Learning in Norwegian PE. Their main findings indicate that the students find that formative assessment is used to a limited extent and that the teachers conduct assessment in very different ways. They highlight engaging students as an important development area, and their study revealed challenges regarding feedback that moves learners forward. This is in line with conclusions drawn by Hammerness (2013) who reports that PSTs in Norwegian ITE programmes have few opportunities to learn in the context of field practice.

The present study will concentrate on illuminating formative assessment in PSTs' teaching and in the mentoring sessions during their field practice. By doing so, we will contribute with a systematic analysis of an ITE program from a new perspective, and will add to our understanding of the complexity of teacher education such as Flores (2017) calls for. Understanding the complexity is a prerequisite for future development. The aim of this study is to explore formative assessment in physical education within ITE. In this study, formative assessment has two different meanings; 1) Information provided by the PSTs to the students in 5th to 10th grade, which is intended to modify their thinking or behaviour to improve learning outcomes, e.g. reinforce or change performance (Hargreaves, 2012). 2) Information, provided by the mentor teachers to the PSTs, intended to review and adjust their teaching (Chng & Lund, 2018).

On this basis, we have posed the following research questions:

1. Is formative assessment present in the PSTs' teaching sessions? If so, how?
2. Do mentor teachers promote PSTs' formative assessment during mentoring sessions? If so, how?

Method

This study is part of a major multidisciplinary project, Teachers as Students, which focuses on teacher education. The project is supported by the Norwegian Research Council (project 212276), and conducted in accordance with national ethical guidelines. The study has used a multiple-case design, which is especially expedient for in-depth investigation of a phenomenon in a real-world context, when the boundaries between the phenomenon and the context may not be evident

(Yin, 2013, p.16). Yin (2013, p.59) argues that a multiple-case research design should follow a replication logic and not a sampling logic used in surveys. A case study covers the phenomena in its context which often leads up to a large number of potentially relevant variables. The complexity in the teaching and mentoring sessions are high and convergent evidence is therefore sought regarding the facts and conclusions in each individual case.

The influence of important contextual conditions in both PSTs' teaching and the following mentoring sessions makes a case study design especially suitable. Context-dependent knowledge lies at the centre of the case study as a research method, and it is important for development of a nuanced view of reality (Flyvbjerg, 2006). According to Yin (2013, p. 57) having more than two cases can blunt criticism like fears about the uniqueness surrounding the case and we have therefore chosen four cases. The purpose of the present study is exploratory, and the replication of cases was based on typical PSTs and mentors in the ITE-program. It relies on both qualitative - and quantitative data, and the analyses are guided by previous research on formative assessment in teaching (Yin, 2013, p. 103). The participants were assured of confidentiality, and informed of their right to withdraw from the study at any time. The study procedures were in accordance with the ethical standards of the authors' university.

Participants

Participants included 15 PSTs and 4 mentors, 11 female students aged 21-25, and 4 male students aged 22-25 years old. None had teaching practice beyond the field practice in teacher education. The PSTs were in their second year (fourth semester) of the ITE-program, and had chosen PE as a specialization. The mentors included three female teachers and one male teacher, between 34 and 42 years old, with 6 to 12 years of teacher experience. One mentor lacked credits in PE, two mentors had 30 ECTS credits in physical education and one mentor had 60 ECTS credits. They were employed in four different public elementary schools, where the activities are conducted in line with the current curriculum. All the schools have been used as practice schools for several years.

Before the PSTs began their field practice they participated in eight weeks PE-teaching on campus. During this period, we had two theory sessions where we presented current theory related to formative assessment. This included an analysis of MacPhail and Halbert's (2010) article, as well as presentation of key features of Hattie and Timperleys' (2007) model for developing skills and understanding. We organized group work where the PSTs discussed their own understanding. Through practical work, the PSTs were challenged to identify key elements in common activities, such as, high jump and forward roll etc. Then they were challenged to guide fellow students in line with the principles of formative assessment. This was repeated four times in advance of the field practice, and the experiences were discussed in plenary.

Procedure

Data collection was conducted during the PSTs' field practice. The participants were in four different groups (School 1, 2, 3 and 4) each consisting of three/four PSTs and one mentor. Each group performed two sessions (Lesson 1 and 2), i.e. the first lesson at school 3 will be referenced as "School 3, Lesson 1". The class size ranged between 21 and 26 students, and they were in 5th to 10th grade. The following activities were represented; ballgames, team athletics, gymnastics, ice skating and dancing. Data collection consisted of two parts:

1. Video recording of two teaching sessions in each practice group, a total of eight teaching sessions.
2. Video recording of the mentoring sessions.

Mentoring was performed before and after each of the eight teaching sessions. This made a total of 16 mentoring sessions. Each session was video recorded and transcribed verbatim.

Analyses

Analyses of the teaching sessions were based on the Classroom Assessment Scoring System (CLASS) (Pianta, Hamre, & Mintz, 2012). CLASS is a standardised and validated method for identifying key factors in classroom interaction (Munthe, 2005), and has recently been validated for use in a Norwegian context (Westergård, Ertesvåg, & Rafaelsen, 2018).

CLASS identifies three domains that are of particular importance when supporting learning and social development in school: Emotional Support, Classroom Organisation, and Instructional Support. Our study focused on the Instructional Support domain, closely related to formative assessment. The underlying subcategories included in the Instructional Support domain are: Instructional learning formats, Content understanding, Analysis and enquiry, Quality of feedback, and Instructional dialogue. Each sub category is scored on a seven-point Likert-type scale (1-2: low-range, 3-5: mid-range and 6-7: high-range).

We used conventional content analysis (Hsieh & Shannon, 2005), in which the coding categories are derived directly from the text data, i.e the CLASS manual. We have therefore categorized the teaching sessions with respect to subcategories (Instructional learning formats, Content understanding, Analysis and enquiry, Quality of feedback, and Instructional dialogue) within the domain Instructional Support, from low (1) to high (7). Results and discussion will be presented related to these subcategories.

The criteria for determining low/middle or high degree (for each subcategories) is based on the CLASS manual (Pianta et. al., 2012). The criteria for "Quality of feedback" are defined in the following ways: Low Quality of feedback is characterized as perfunctory or non-existent of feedback. Students are not provided with assistance, hints or prompting from the teacher or peers when

participating in class work, but are left to complete work without such assistance. For the mid-range group there are occasional feedback loops between the teacher and students or among students. The teacher and/or peers sometimes scaffold student learning, but these interactions are brief or not of sufficient depth to allow students to fully perform at a higher level. High-range feedback is described as frequent feedback loops between the teacher and students and among students, which lead students to obtain a deeper understanding of material and concepts. Moreover, the teacher and/or peers often scaffold student learning, allowing them to perform at a higher level than they would be able to perform independently.

To avoid subjective bias (Yin, 2013) two researchers watched the videos together, but coded independently. The sessions were split into 20-minute sections. Previous studies have indicated that 15-25 minutes are an appropriate period (Jamison, Cabell, LoCasale-Crouch, Hamre, & Pianta, 2013; La Paro, Pianta, & Stuhlman, 2004). Based on scores in the different sequences, we developed a common score for the entire lesson. In case of discrepancies, we discussed and agreed on a common understanding. This is in line with the procedures recommended by Yin (2013, p.199) and applied in another study related to Norwegian teacher education (Helgevold, Næsheim-Bjørkvik, & Østrem, 2015). To ensure rater reliability we conducted a reliability test, ICC (Inter class reliability). The Cronbach's alpha value ranged between 0.89 and 1 for all lessons.

The video transcriptions from the mentoring sessions were also analyzed using conventional content analysis (Hsieh & Shannon, 2005) and excerpts from the transcription will be cited in the results section to illustrate the mentors' focus. The unit of analysis was the mentor teachers' questions as posed in the mentoring sessions. The questions were categorised in relation to the various dimensions within the Instructional Support domain.

Results and comments

Results and discussion will be presented together, while the research questions will be dealt with separately.

The results for research question 1 were based on analysis of video from the PSTs' eight different teaching sessions, and the results for research question 2 were based on video recording of mentoring sessions related to these teaching sessions.

Research question 1: Is formative assessment present in the PSTs' teaching sessions? If so, how?

Table 1. Scores for PSTs' teaching lessons (Instructional support) (Pianta, Hamre, and Mintz, 2012)

	Sch1, L1	Sch1, L2	Sch2, L1	Sch2, L2	Sch3, L1	Sch3, L2	Sch4, L1	Sch4, L2	Mean score
Instructional Learning Formats	4	6	4	4	3	4	6	6	4.625
Content Understanding	2	5	3	5	2	2	5	4	3.500
Analysis and Inquiry	1	3	1	1	1	1	1	1	1.250
Quality of Feedback	2	3	2	3	2	2	4	3	2.625
Instructional Dialogue	2	4	2	2	2	2	2	2	2.250

Table 1 reveals the scores, and mean scores, for all sub categories included in the dimension "Instructional Support" for all eight sessions. Results show significant variations between the different dimensions.

The sub categories need to be explained further, and in the following we will also present our reasoning related to the scores given:

Instructional Learning Formats focuses on the ways in which the teacher maximises student engagement in learning through clear presentation of material, active facilitation and the provision of interesting and engaging lessons and materials.

For the dimension *Instructional Learning Formats* the score was 4.625 (table 1). According to the CLASS manual (Pianta, et. al., 2012) such a score indicates the upper part of the middle layer. In our study, this is reflected in the fact that the PSTs were presenting learning objectives and were keen to explain what would happen in the class, but did not succeed communicating the goals properly. The PSTs in general did not effectively draw students' attention to the objectives.

This can be illustrated by an example; One PST presented objectives for the lesson; 1) collaborate in teams and 2) become familiar with the activities of team athletic (A composite activity that stimulates skills in speed, agility, strength, coordination and endurance). In addition, the PST challenged the class to describe the extent to which they felt they had achieved the objectives:

Pre-service teacher (PST): Was this fun to do in PE?

Student (S): Yes. Good fun.

PST: Did you manage to jump or throw further, or keep your balance over time?

S: My results worsened.

PST: Were you worse? That's probably because you were more tired towards the end.

S: My performance in javelin was worse at the end of the lesson.

- PST: Do you remember the objective for the lesson?
S: To work as a group and become familiar with team athletic.
PST: How do you think that you managed to achieve that goal?
S: Very well.
PST: I agree, you did well.

This PST tried to maximise the students' engagement in learning, through presentation of and reflection on learning objectives. However, the goals were only partially operationalised; this made it difficult for students to know what they should learn. More specific learning objectives could have been related to such things as flexibility, speed or technique. Students were not informed what was expected in terms of cooperation (e.g. give tips, offer corrections, encourage each other). By verifying the students' statements, confirming that the objective was reached, we interpret PSTs' focus in line with Crum's (2013) non-teaching perspective, focusing on being physically active, largely ignoring how the activity is performed or how to support the students' learning. However, there are evidence of multiple instructional strategies, such as switching between whole group and small group exercises combined with individual tasks/exercises, and the PSTs have put some effort into creating lessons which encourage and stimulate students' interest and exploration.

Content understanding refers to the depth of lesson content and the approaches used to help students comprehend the framework, key ideas, and procedures in a discipline.

For the dimension *Content understanding* the score was 3.500 (table 1). According to the CLASS manual (Pianta, et. al., 2012) such a score indicates the lower part of the middle layer. Class discussion and materials communicate a few of the essential attributes of the exercise/task, but the PSTs often fail to explain how and when the particular skills should be used.

To teach a motor exercise (such as jumping, catching a ball, gymnastics etc.) or to develop a physical quality (e.g. endurance, strength, speed etc.) among students requires content understanding. This means that the PSTs must be able to (Pianta, et al., 2012):

- Introduce appropriate exercises
- Deconstruct an exercise in appropriate stages
- Identify critical elements in an exercise
- Develop an exercise, when students need new challenges (e.g. from forward roll to dive roll)
- Build on previous skills (throwing, receiving, make use of the learned techniques) in ball games, for example.
- Interact with the students that lead to an integrated understanding of skills and concepts

Our study revealed that PSTs chose appropriate exercises, but were less able to deconstruct them in appropriate stages and identify critical elements in an exercise. This was manifested through records from the gymnastic lessons:

The PST introduces the "forward roll" by demonstrating the exercise and highlighting the importance of bending the neck. Thereafter, students work in groups while the PSTs supervise their performance. Many students have trouble with the rolls; some are crooked, while others fail to come up on their legs. Nevertheless, the PSTs only provide a limited amount of individual feedback. Instead, they repeat the collective messages for the whole class. Nor are they able to differentiate tasks in relation to students' different skill levels.

Those who struggled could have been given instructions on the arm/head position and the kick process, while the students who could perform well on the assigned task, could have been given tasks of greater complexity, like the "diving roll".

These findings indicated PSTs' limited capability to identify critical elements in an exercise and to improve learning through facilitating differentiated tasks. Moreover, they largely failed to clarify why they worked with different exercises, and often failed to present the meaning and purpose of the activities. This contrasts with key aspects of AfL which are concerned with clear learning goals (Lorente-Catalán & Kirk, 2016).

Analysis and Inquiry refers to how the PSTs maximise students' higher-level thinking skills using knowledge and skills on novel/or open-ended problems, tasks and questions.

For the dimension *Analysis and Inquiry* the score was 1.250 (table 1). According to the CLASS manual (Pianta, et. al., 2012) such a score indicates the lower part of the scale. Analysis and enquiry are not evident in our data. The PSTs focused on implementation. There was no evidence of problem solving, experimentation or exploring new movement solutions in the data collected.

PSTs did not succeed to develop students' skills through challenging questions/tasks, such as:

Why do you think you jumped further this time?
Why do you think you fell backwards in landing?
How will your head/arm/feet position affect the result?

By not using these types of questions, the teachers deprived students of the opportunity to engage in metacognition, i.e. challenging their own thinking. Instead, they asked questions like:

Was it fun?
How far did you jump?
Are you tired?

These questions did not invite analysis and enquiry. There was little focus on students' learning experiences. The PSTs failed to provide a proper summary,

focusing on achievement of learning target. In order to assess PSTs' facilitation for analysis and enquiry, it was important to consider the quality of questions they asked the students. We noticed that the PSTs were concerned about results, and we interpret this as a product focus at the expense of the process. This violates central aspects of AfL which emphasize the meaning of involving the students in their own learning process (Lorente-Catalán & Kirk, 2016).

Quality of feedback refers to how the PSTs extend students' learning through their responses to students' ideas, comments and work.

For the dimension *Quality of feedback* the score was 2.625 (table 1). According to the CLASS manual (Pianta, et. al., 2012) such a score indicates the higher part of the lower layer. We found little evidence of feedback loops between PSTs and students that expanded and stimulated learning, or promoted students' understanding and participation. Some general comments on self-level (good, fine, great) may have encouraged students, but we could not identify individual feedback on the task, process, or self-regulated levels. Some general feedback was given in plenary, such as:

Bend your neck.
Make a triangle with your hands.
Push with your legs.

However, these comments were not observed in the individual feedback. PSTs' feedback practices can be summarised as:

- Little feedback on task -, process - and self-regulation level, majority of the feedback at the self-level. E.g., "*Great, good work*" or "*After this session you deserve a warm shower*"
- Appropriate error correction is absent; the focus is on the result. E.g., "*Great, now you jumped 2 meters!*" or "*Yes, now you threw longer than John.*"

This is not consistent with key aspects of AfL which points out the importance of guiding the students where to go (Lorente-Catalán & Kirk, 2016).

Instructional dialogue refers to the purposeful use of content-focused discussion among PSTs and students that is cumulative, with the PST supporting students to chain ideas together in ways that lead to a deeper understanding of content.

For the dimension *Instructional dialogue* the score was 2.250 (table 1). According to the CLASS manual (Pianta, et. al., 2012) such a score indicates the middle part of the lower layer. The PSTs did not capture the purposeful use of content-focused discussion among students and teachers, discussions contained only simply question-response exchanges between PSTs and students. Nor did they invite students actively into the discussions, and there was little opportunity for students to express their own ideas and points of view. The class was

dominated by PSTs talking, characterised by extensive use of closed questions. This breaks with central aspects of AfL which emphasizes the importance of giving students greater skills in self-evaluation (Hattie & Timperley, 2007).

PSTs' modest use of formative assessment highlights the importance of charting how mentors treat this during field practice.

Research question 2: Do mentor teachers promote PSTs' formative assessment during mentoring sessions? If so, how?

Instructional Learning Formats

Conversations about Instructional Learning Formats are limited during the mentoring sessions. The mentors had a strong focus on organising and effective time utilisation. There was no discussion of how the PSTs maximised students' engagement in learning through clear presentation of learning targets. This is illustrated in the following excerpt from a mentoring session prior to teaching a ballgame lesson (School 4, Lesson 1);

Pre-service teacher (PST): The class starts with an introduction to the students, what to do.

Mentor (M): Great, and what is the plan for this session?

PST: We start with different ball games and then we finish off by playing basketball.

M: That's a good choice.

Through this statement, the mentor shows that he is most concerned with the content of the session. Students' learning outcomes are not discussed, but during the mentoring session, the mentor comments favorably that the session will probably engage the students.

Content understanding

We found modest traces of content understanding in mentoring sessions. One mentor, however, pointed out some occasions where the exercises demanded motor skills and challenged the PST.

M: Did you notice the main challenge in long jump?

PST: No, but it became a little chaotic.

M: Can you split the exercise into different phases?

PST: Yes, takeoff and landing.

M: Are there other key phases?

PST: Run-up.....

M: True, it is common to use the following phases: run-up – takeoff – flight – landing.

Then we can focus on some of these, depending on the students' level. Therefore, it is important to have a thorough understanding of an activity before we teach it.

This mentor demonstrated content understanding and was able to show the relevance of such understanding when teaching PE. However, we noted few

statements in this category. In contrast, feedback from the mentors was mainly aimed at organising the activities and students, and students' wellbeing.

Analysis and enquiry

Focus on analysis and enquiry was not prominent during the mentoring sessions. The mentors were mostly concerned with efficient organisation and high activity levels, and paid little attention to what PSTs could do to give students mastery experiences and increased motor skills.

However, the mentor who emphasised content understanding also challenged the PSTs to reflect on how the students could be helped further in their learning process. She referred to a situation where a student was struggling to balance on one foot:

M: What advice would you give this student?

PST: The student can be encouraged to lean against a wall or a mate.

M: Yes, that's a possibility, do you see other solutions?

PST: You could ask the student to concentrate better.

M: Is it possible to make the activity easier by regulating the body's position?

PST: Yes, you can ask the student to bend the leg a bit.

M: That could be good - What will happen if the arms are abducted?

PST: That makes it easier to keep your balance.

M: Great, that would be a good strategy.

Here we met a mentor who through active, conscious use of questions manages to stimulate the PSTs' own reflection about the consequences of alternative approaches in the teaching process, in line with the principle of AfL (Lorente-Catalán & Kirk, 2016). Unfortunately, this was a unique case in our study.

Quality of feedback

There were few instances of use of feedback intended to promote students' learning. We noted, however, that the mentoring related to the gymnastics session focused on the PSTs' lack of professional feedback;

M: In the exercise, forward roll, you asked the student to bend the chin against the chest. Great! If you detect inappropriate use of arms, how can you help them?

PST: It's so hard, there are so many students.

M: Yes, I understand. However, you can focus on some specific students who receive feedback this time and then you focus on others next time. Which feedback would you give in relation to the position of the arms?

PST: I don't know.

M: You could say; make sure you have parallel arms, bend the elbow as the arms touch the ground, finish the arm's movement by stretching your arms.

PST: Oh, yes, but it's hard to detect errors.

The mentors focused to a small extent on the PSTs' use of feedback in the teaching sessions. Moreover, we note that the mentor who challenged in relation to

feedback failed to stimulate any reflections in the PST. It could be that the PST has little content knowledge to tap into, or it be due to the fact that PST is unfamiliar with this type of questions, that the focus in the mentoring traditionally has been directed at organization and activity level (Helgevold, et.al., 2015).

Instructional dialogue

Finally, we noted the absence of discussions related to learning outcomes in teaching sessions. The PSTs did not capture the purposeful use of content-focused discussion among students and teachers. None of the mentors highlighted the significance of instructional dialogue in the mentoring sessions.

Discussion

Previous research in PE has done little to examine formative assessment in general (Leirhaug & Annerstedt, 2015), and in PSTs teaching sessions in particular. Furthermore, we have limited scientific content knowledge of mentoring sessions in connection with PST teaching (Helgevold, et. al., 2015) and if mentor teachers promote PSTs' formative assessment during these sessions. This knowledge gap is surprising because mentor sessions is considered a core educational initiative during field practice and mentors are considered the most important factor shaping PSTs' behaviour (Behets & Vergauwen, 2006).

The purpose of this explorative multiple case study was therefore to investigate formative assessment in PSTs teaching and in the mentoring discussions related to teaching sessions. Observational data using the CLASS manual to map the observed PSTs' Instructional Support of Teaching (table 1) reveals small traces of formative assessment in all four cases. Low level of equivalent dimensions that are all expected to have a major impact on how teachers can support students' learning is perhaps not surprising since PSTs still are in their educational phase. Compared with national and international research in other school subjects among teachers (Hayes, 2018; Westergård, et. al., 2018) using the CLASS model, shows that the values for *Content understanding*, *Quality of feedback* and *Instructional dialog* are considerably lower in the present study.

Previous observational research on PSTs' teaching sessions shows that they are high on social support but low on learning support (Mjåtveit & Giske, 2017), and the findings in the present study underpin that this core teaching competency is not completely acquired during teacher education. The results displayed in table 1 shows large variations between the different dimensions, e.g. between *Instructional learning formats* (4.625) and *Analysis and inquiry* (1.250). This variation between the scores in different learnings support dimensions does emerge more distinctively in the present study compared with the findings of Westergård et. al. (2018). This discrepancy may be explained by research design

differences, or by the CLASS manual not having been used specifically in PE. Furthermore, Westergård, et. al. (2018) problematizes the values on the Instructional Support dimensions because it is challenging to rate.

Another possibly worrisome explanation of the low scores on some of the learning support dimensions such as *Analysis and inquiry*, *Quality of feedback* and *Instructional Dialog* are that they are not given sufficient attention in the teacher educational program. The score on *Instructional learning formats* in the present study was almost identical with Westergård et. al. (2018) findings (4.6). This observational dimension is about teacher behavior categorized as clear learning targets and clear presentation of information, indicating that the *Instructional learning formats* category doesn't distinguish between what to do and what to learn. When students met PSTs, it was common to ask questions such as; what are we going to do today? Can't we play volleyball? The focus was on content, not the purpose of the activities. Through such questions, the PSTs' attention was more easily directed towards justifying activity choices, and it was thus irrelevant to challenge students' ability to reflect on their own learning and the learning process. In cases where the PSTs presented specific objectives for the lesson, we saw examples of these being misunderstood or simply not perceived by students. In several sessions, the PSTs' final summary revealed that the students had not captured the goal of the lesson. In other school subjects there are perhaps a more obvious implicit connection between what to do and what to learn indicating that *Instructional learning formats* category in CLASS-manual in the PE context should be further elaborated.

We expected that the teacher education, with its focus on the AfL, would contribute to a higher degree of AfL than the modest elements found in the present study. This minor outcome is however in line with Lortie and Clement (1975) who claims that PSTs are shaped by their own experiences. PSTs' prior experiences and the nature of those experiences have been identified as an influential factor in the process of learning how to teach (Hammerness, 2013; Korthagen & Verkuyl, 2007; Lortie & Clement, 1975). Backman and Larsson (2016) and Moen (2011) point out that PETE has only limited impact on how PE is taught in schools and their findings indicate that PETE partly lacks an integrated perspective on knowledge.

The mentors' modest focus on AfL in the mentoring sessions can be explained by ambiguous and unclear expectations from the university side. Two weeks before the field practice, a planning session was arranged between the PSTs, mentors and teacher educators. In retrospect, we recognize that the main focus was on practical clarifications, related to the number of teaching lessons, planning form, laws and regulations, and equipment available. In addition, information was provided about the students' level and the relevant teaching topics. These are important prerequisites for being able to provide good teaching, but we acknowledge that we should be more distinct in relation to our expectations related to mentoring. There was no discussion about the purpose and content of

the mentoring sessions. Traditionally, the mentor's focus has been on PSTs' doings and classroom organization in the teaching session (Helgevold et al., 2015; Ottesen, 2007; Sundli, 2007) and to a lesser extent on students' learning. Therefore, even though the mentors may use AfL in their own teaching, their established perception of the mentor role will mean focusing on actions rather than students' learning. Changing such a tradition requires an active process of structured meetings prior to the field practice, where the parties' involved (PST, mentor and teacher educator) also clarify professional expectations towards PSTs' teaching.

Another explanation of the low prevalence of AfL in the mentor session is PE's historical traditions with a focus on fitness training and/or entertainment, implying that the mentors also are unfamiliar with AfL in physical education. This interpretation is supported by previous research within the field of formative assessment (Lorente-Catalan & Kirk, 2016; Leirhaug & Annerstedt, 2015), where it is argued that AfL represents a break with the established assessment practice in PE and are used in a limited way.

Behets and Veergauwen (2006) findings show that they consider field practice to be the most important arena in teacher education. Field practice can be considered as a primary experience which according to Reed (1996) is gained through our senses and are the most basic source for understanding the reality and learning for ourselves in contrast to secondary experiences where information is selected, modified, packaged and presented to us by others underpins this assumption. Dewey (1938) argue that not all experience is good, and it might be mis-educative if stimulation is without information. Low level of AfL in the mentoring sessions may give the impression that the field practice is cementing old non-teaching traditions in the subject and may deprive students from nurturing experience. Dewey (1938) claims that mis-education has the effect of arresting the growth of further experience and in physical education this can be illustrated throughout Crum's vicious circle (Crum, 2013) which emphasizes that even though a PETE program succeeds in accomplishing the desired PE teaching perspective, there is a great chance that these changes will appear to be cosmetic once the PST confronts the real life in schools. Because many mentors and students hold non-teaching perspectives and expectations concerning PE, the old perspective will be reinforced.

Limitations

The present study is not without its limitations and these issues should be considered when the findings are interpreted. The relatively small number of cases recruited from one teacher education institution means that care is required drawing general conclusions. The four cases are demarcated into two teaching sessions in each practice group and connected pre- and post-mentoring sessions indicating that other sessions might have given another result. Furthermore, professional dialogues between mentor and PETE-student could have occurred

outside these pre-structured pedagogical sections and they are therefore not included in the data. The video recording could also have evoked anxiety among the participants and prevented them from authentic behaviour.

Practical implication and future research

The findings in present study shows modest elements of AfL both in the PSTs' teaching and in the mentors' mentoring and it reveal that PETE second year students have not receive the AfL knowledge they need neither on campus nor during field practice. The absence of this knowledge is probably due to three different reasons; first, it may indicate that the PE-educators have not succeeded in implementing AfL in a proper way. To remedy this, we may spend more time developing a common understanding of the term AfL among the PE-educators at campus, and furthermore, a strategy to meet the need for AfL. Secondly, it may indicate that the PSTs have not received adequate training in AfL. This can be improved by providing PSTs with more and better opportunities to explore the AfL in their practical teaching sessions at campus. We also recommend that the PSTs are given assignments related to the CLASS manual through the teaching taking place on campus. By challenging the PSTs to observe and evaluate each others teaching, we hope to increase the focus on Instructional support in general and promote awareness of the importance of analysis and inquiry in special.

Thirdly, in addition to unclear expectations about the content of the mentoring, it may appear that the mentors lack competence related to AfL. To address this, it may be appropriate to provide mentors mandatory training related to AfL in advance of the field practice.

A longitudinal research design following the PSTs' teaching development during the educational phase and as newly educated teachers is warranted. This might give us valid information about the efficiency of the PETE program and eventually reveal weaknesses in the progression of teacher education. Future research should also direct the attention towards learning outcome of field practice in general and mentoring session in particular. Furthermore, knowledge about teacher educators' ideology and teaching in practical activities involving considerations about transfer to the school context is also needed. Last, but not least there is a need for more research that explores the CLASS manual in a PE context.

About the authors

Atle Mjåtveit is assistant professor at the Department of Education and Sports Science, University of Stavanger (UiS). He teaches physical education and sports didactics.

Institutional Affiliation: Department of Education and Sports Science, University of Stavanger (UiS), Pb. 8600 Forus, 4036 Stavanger
E-mail: atle.mjaatveit@uis.no

Rune Giske is professor at the Department of Education and Sports Science, University of Stavanger (UiS). His main research topics are team sport performance (shared cognition) and coaching. Furthermore, topics related to learning environment in physical education and teacher education. He has long experience as a handball coach and youth coach in football.

Institutional Affiliation: Department of Education and Sports Science, University of Stavanger (UiS), Pb. 8600 Forus, 4036 Stavanger
E-mail: rune.giske@uis.no

References

- Backman, E., & Larsson, H. (2016). What should a physical education teacher know? An analysis of learning outcomes for future physical education teachers in Sweden. *Physical Education and Sport Pedagogy*, 21, 185-200.
- Behets, D., & Vergauwen, L. (2006). Learning to teach in the field. In D. Kirk (Ed.), *The handbook of physical education* (pp. 407-424). London: Sage Publications.
- Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, 31, 416-440.
- Buscà, F., Ambròs, A., & Burset, S. (2017). Bibliometric characteristics of articles on key competences indexed in ERIC from 1990 to 2013. *European Journal of Teacher Education*, 40, 144-156.
- Chng, L. S., & Lund, J. (2018). Assessment for Learning in Physical Education: The What, Why and How. *Journal of Physical Education, Recreation & Dance*, 89, 29-34.
- Cloes, M., Denève, A., & Piéron, M. (1995). Inter-individual variability of teachers' feedback. Study in simulated teaching conditions. *European Physical Education Review*, 1, 83-93.

- Crum, B. (2013). How to pave the road to a better future for physical education. *Journal of Physical Education & Health-Social Perspective*, 2, 53-64.
- Crum, B. J. (1985). The use of learner reports for exploring teaching effectiveness in physical education. *Sport pedagogy*, 97-102.
- Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education*, 57, 300-314.
- Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education*, 1-19.
- Dewey, J. (1938) *Experience and education*. (Reprinted in *The Later Works of John Dewey vol 13. Carbondale: Southern Illinois University Press, 1988*)
- Eriksson, A. (2017). Pre-service teachers' questions about the profession during mentoring group conversations. *European Journal of Teacher Education*, 40, 76-90.
- Fishburne, G. J., & Borys, A. H. (1987). A comparison between elementary school teachers' and student teachers' conceptions of successful teaching. In ICHPER/CAHPER Conference, University of British Columbia, Vancouver, Canada.
- Flores, M. A. (2017). Practice, theory and research in initial teacher education: international perspectives. In: Taylor & Francis.
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry*, 12, 219-245.
- Gamlem, S. M., & Munthe, E. (2014). Mapping the quality of feedback to support students' learning in lower secondary classrooms. *Cambridge Journal of Education*, 44, 75-92.
- Hammerness, K. (2013). Examining features of teacher education in Norway. *Scandinavian Journal of Educational Research*, 57, 400-419.
- Hamodi, C., López-Pastor, V. M., & López-Pastor, A. T. (2017). If I experience formative assessment whilst studying at university, will I put it into practice later as a teacher? Formative and shared assessment in Initial Teacher Education (ITE). *European Journal of Teacher Education*, 40, 171-190.
- Hargreaves, E. (2012). Teachers' classroom feedback: still trying to get it right. *Pedagogies: An International Journal*, 7, 1-15.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77, 81-112.
- Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in educational evaluation*, 38, 21-27

- Hayes, S. A. (2018). Observing Classroom Practices with the Classroom Assessment Scoring System–Secondary (CLASS-S): A Scoping Review, and Preliminary Reliability and Validity Study in Alberta, Canada.
- Helgevold, N., Næsheim-Bjørkvik, G., & Østrem, S. (2015). Key focus areas and use of tools in mentoring conversations during field practice in initial teacher education. *Teaching and Teacher Education*, 49, 128-137.
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15, 1277-1288.
- Jamison, K. R., Cabell, S. Q., LoCasale-Crouch, J., Hamre, B. K., & Pianta, R. C. (2013). CLASS–Infant: An Observational Measure for Assessing Teacher–Infant Interactions in Center-Based Child Care. *Early Education and Development*, 25, 553-572.
- Korthagen, F., & Verkuyl, H. S. (2007). Do you encounter your students or yourself? Enacting a pedagogy of teacher education: Values, relationships and practices, 106.
- Korthagen, F. A. (2010). How teacher education can make a difference. *Journal of education for teaching*, 36, 407-423.
- 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, 409-426.
- Leirhaug, P. E. (2016). Exploring the relationship between student grades and assessment for learning in Norwegian physical education. *European Physical Education Review*, 22, 298-314.
- Leirhaug, P. E., & Annerstedt, C. (2015). Assessing with new eyes? Assessment for learning in Norwegian physical education. *Physical Education and Sport Pedagogy*, 1-16.
- Livingston, K., & Flores, M. A. (2017). Trends in teacher education: a review of papers published in the European journal of teacher education over 40 years. *European Journal of Teacher Education*, 40, 551-560.
- Lorente-Catalán, E., & Kirk, D. (2016). Student teachers' understanding and application of assessment for learning during a physical education teacher education course. *European Physical Education Review*, 22, 65-81.
- Lortie, D. C., & Clement, D. (1975). Schoolteacher: A sociological study: JSTOR.
- MacPhail, A., & Halbert, J. (2010). 'We had to do intelligent thinking during recent PE': students' and teachers' experiences of assessment for learning in post-primary physical education. *Assessment in Education: Principles, Policy & Practice*, 17, 23-39.

- Metzler, M. W., Tjeerdsma, B. L., & Mozen, D. M. (2000). Chapter 7: Assessing Pedagogical Knowledge. *Journal of Teaching in Physical Education*, 19, 487-507.
- Mjåtveit, A., & Giske, R. (2017). Learning climate in physical education: Analysis of pre-service teachers' reflections about, and facilitation of, learning climate in their own teaching. *Journal of Physical Education and Sport*, 17(1), 224-232. Retrieved from <http://www.efsupit.ro/>. doi:<http://dx.doi.org/10.7752/jpes.2017.01034>
- Moen, K. M. (2011). " Shaking or stirring?": a case-study of physical education teacher education in Norway.
- Moen, K. M., & Standal, Ø. (2016). Practicum in Physical Education Teacher Education: An Educational Partnership? *SAGE Open*, 6, 2158244016635715.
- Munthe, E. (2005). Innholdsanalyse av klasseromsvideoer: med CLASS som et eksempel *Norsk pedagogisk tidsskrift*, 89 (2), 159-174).
- Orland-Barak, L. (2014). Mediation in mentoring: A synthesis of studies in Teaching and Teacher Education. *Teaching and Teacher Education*, 44, 180-188.
- Ottesen, E. (2007). Reflection in teacher education. *Reflective practice*, 8, 31-46.
- Pianta, R. C., Hamre, B. K., & Mintz, S. L. (2012). Classroom Assessment Scoring System: Secondary Class: Teachstone.
- Reed, E. S. (1996). *The necessity of experience*: Yale University Press.
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*, 35, 535-550.
doi:10.1080/02602930903541015
- Silverman, S., Tyson, L., & Krampitz, J. (1992). Teacher feedback and achievement in physical education: Interaction with student practice. *Teaching and Teacher Education*, 8, 333-344.
- Standal, Ø. F., Moen, K. M., & Moe, V. F. (2014). Theory and practice in the context of practicum The perspectives of Norwegian physical education student teachers. *European Physical Education Review*, 20, 165-178.
- Sundli, L. (2007). Mentoring—A new mantra for education? *Teaching and Teacher Education*, 23, 201-214.
- Svennberg, L., Meckbach, J., & Redelius, K. (2014). Exploring PE teachers' 'gut feelings' An attempt to verbalise and discuss teachers' internalised grading criteria. *European Physical Education Review*, 20, 199-214.
- Tolgfors, B., & Öhman, M. (2016). The implications of assessment for learning in physical education and health. *European Physical Education Review*, 22, 150-166.

- Torkildsen, L. G., & Erickson, G. (2016). 'If they'd written more...'—On students' perceptions of assessment and assessment practices. *Education Inquiry*, 7, 27416.
- Westergård, E., Ertesvåg, S. K., & Rafaelsen, F. (2018). A Preliminary Validity of the Classroom Assessment Scoring System in Norwegian Lower-Secondary Schools. *Scandinavian Journal of Educational Research*, 1-19. doi: 10.1080/00313831.2017.1415964.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11, 49-65
- Yin, R. K. (2013). *Case study research: Design and methods*: Sage publications.
- .