

# ‘It’s the other assessment that is the key’: three Norwegian physical education teachers’ engagement (or not) with assessment for learning

Petter E. Leirhaug<sup>a\*</sup> and Ann MacPhail<sup>b</sup>

<sup>a</sup>*Norwegian School of Sport Sciences, Oslo, Norway;* <sup>b</sup>*Department of Physical Education and Sport Sciences, University of Limerick, Limerick, Ireland*

The international agenda for assessment continues to convey a growing interest in assessment for learning (AfL) as a tool to support learning and enhance teaching. Complementing this, the recent literature on assessment in physical education acknowledges the need for physical educators to integrate AfL into their teaching and assessment practice as an important part of the future development of the subject. Appreciating that physical education must be recognized as part of the larger movement culture in society and is a place to learn about movement culture, this study explores how AfL is understood and enacted by physical education teachers and the extent to which such enactment complements or challenges learning movement cultures within physical education. This study shares how three Norwegian physical education teachers used AfL to term what they were practicing with respect to assessment in physical education. We follow the interactions of the selected teachers throughout focus groups, using the empirical data as our ‘dialogue partner’ in reconstructing and discussing their assessment stories. We conclude that the need of embedding AfL in learning theory may well be one of the strongest challenges to enacting AfL in physical education. We acknowledge that not only are most existing theories of learning defined cognitively, but also that learning connected to physical education and activity is, to a large extent, practical and embodied, and also linked to the powerful discourses of sport and related areas such as health.

**Keywords:** *Assessment for learning; Formative assessment; Physical education; Teacher stories; Norway*

## Introduction

Assessment for Learning (AfL), in contrast to the more traditional ‘assessment of learning’, has more recently gained considerable interest in research and development on assessment in physical education. AfL is viewed as fundamental to enhancing student learning and developing physical educators’ teaching and assessment practice, and as such is deemed to be an important element of physical education’s future (Chan, Hay, & Tinning, 2011; Georgakis & Wilson, 2012; Hay &

---

\*Corresponding author. Norwegian School of Sport Sciences, Oslo, Norway. Email: [petter.erik.leirhaug@nih.no](mailto:petter.erik.leirhaug@nih.no)

Penney, 2013; López-Pastor, Kirk, Lorente-Catalán, MacPhail, & Macdonald, 2013; MacPhail & Halbert, 2010; Ni Chróinín & Cosgrave, 2013). That said, assessment in physical education has been an area with little consensus, questionable validity and a slow ability to change (Annerstedt & Larsson, 2010; Arnesen, Nilsen, & Leirhaug, 2013; Hay & Penney, 2013). The historical legacy of physical education being closely linked to the practice of sport and performance culture does not necessarily correspond with the advances in educational theories of knowledge and learning (Kirk, 2010; Quennerstedt, Öhman, & Armour, 2014), and the need of embedding 'alternative assessment' in a learning culture (Black & Wiliam, 2009; Shepard, 2000).

The focus in school physical education is argued to be 'more on activity than on learning' (Annerstedt, 2008, p. 316), often linked to a taken-for-granted argument of how the subject contributes to a healthy and lifelong interest in sport and physical activity. Among others, Green (2014) concludes that there is little research to support the notion that physical education contributes to such a lifelong interest in sport and physical activity. Interestingly, Säfvenbom, Haugen, and Bulie (2014) view the mandatory aspects of physical education as a double-edged sword which also can 'prevent the adolescents from leaving a harmful context and thus produce learned helplessness and an overall negative attitude to movement contexts' (p. 6). Similar to many other countries, physical education in Norway 'seems to favour those who are already involved in movement activity and those who are involved in competitive youth sports in particular' (Säfvenbom et al., 2014, p. 15). It was partly to move away from such a 'meritocratic mode' that Bart Crum (1993), more than 20 years ago, introduced 'movement culture' as the core of physical education. Rather than making 'unrealistic effect claims' (Crum, 1993, p. 352) on behalf of physical education, Ward (2014) agrees that physical education must be recognized as part of the larger movement culture in society, and a place to learn about this, 'Physical Education should be a learning context in which pupils develop a personal movement identity which inherently involves utilising a range of Sporting activities to enable them to become critical and lifelong consumers of Movement Culture' (p. 577).

Calling for lifelong and life-wide learning in physical education, Hay and Penney (2013) recognize AfL:

as a tool that can be utilised to generate discussions and new thinking about learning and learning opportunities in physical education within and beyond schools. It is concerned with quality and equitable learning opportunities and experiences now and in the future, for all students. (p. 110)

In a recent review on 'alternative assessment' in physical education, López-Pastor et al. (2013) observed 'that such innovative practice is, however, far from regular, integral, widespread and educationally productive' (p. 73). Drawing attention directly to AfL, Georgakis and Wilson (2012) highlighted two main challenges for teachers wishing to engage with AfL; (1) a lack of practical examples and (2) a tendency in literature to simply reiterate to the reader that assessment in physical

education needed to be 'authentic'. In interviewing 17 teachers of physical education and school sport in Australia, Georgakis and Wilson (2012) reported that physical education and school sport teachers did not use the 'assessment for learning' term or discuss links between assessment and learning.

With this as a backdrop, the study in this paper shares instances where Norwegian physical education teachers referred to 'assessment for learning' (unprompted) as what they were doing, or attempting to do, in their engagement with assessment practice in school physical education. Taking this empirical finding as our starting point, the purpose of the study is to explore AfL in physical education through Norwegian physical education teachers' narratives. The main research question that directs this study is 'How is AfL motivated, understood and enacted by physical education teachers who are trying to integrate AfL into their teaching and assessment practice?'

Before sharing the assessment stories of three Norwegian physical education teachers, we map the development of AfL and research on AfL in physical education. We then present the Norwegian study context and discuss the methodological approach.

### **Assessment for learning**

'Formative assessment' has undergone considerable development since Scriven (1967) coined the term related to curriculum and school reforms. It was not before the late 1980s that 'the idea that classroom assessment practices could both afford and constrain student learning began to gain widespread acceptance' (Wiliam, 2011, p. 13), and during the 1990s that 'assessment for learning' became the preferred phrase to 'formative assessment'. The purpose of supporting student learning and learning opportunities defines AfL, allowing AfL to retain both summative and formative assessments. The validity and reliability questions are still relevant to examining the extent to which AfL practices support student learning and achievement (Gardner, 2012). Well-planned AfL informs the most effective, meaningful and worthwhile instruction strategies to improve teaching and the subsequent student learning experiences (Tannehill, van der Mars, & MacPhail, 2013).

Ongoing discourse surrounds how to theorize and understand AfL, and how to realize its considerable promises (Bennett, 2011; Gardner, 2012; Wiliam, 2011). We favour the view that AfL is embedded in a social-constructivist view of learning which places the learner at the centre of the learning process (Shepard, 2000; Hay, 2006). We subsequently adopt the much cited definition of AfL as, 'the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there' (Assessment Reform Group, 2002, p. 2).

In terms of classroom strategies and practical techniques that teachers can use to improve the quality of instructional decisions, the AfL literature agrees on some key

Table 1. Five key strategies of AfL, adapted from Wiliam and Thompson (2008)

	Where the learner is going?	Where the learner is right now?	How to get there?
Teacher	(1) Clarifying learning intentions and criteria for success	(2) Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	(3) Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	(4) Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	(5) Activating students as the owner of their own learning	

principles. Table 1 presents the five key strategies, numbered one to five, we use as a base for analysis.

Drawing on Shulman's (1986) distinction between subject knowledge and pedagogical content knowledge, Bennett (2011) argues that quality formative assessment depends on deep subject domain understanding and highlights two implications. First, the importance of interpreting assessment information meaningfully and being able to arrange suitable feedback. Second, and partly as a consequence of the first implication, 'the intellectual tools and instrumentation we give to teachers may differ significantly from one domain to the next because they ought to be specifically tuned for the domain in question' (pp. 15–16).

The limited number of AfL studies related to physical education reports positive learning movement experiences among students, and indicate that the implementation of AfL can be productive in teacher development and result in improved student learning. On completion of a physical education assessment-planning AfL framework to a number of schools in Ireland, MacPhail and Halbert (2010) reported that both the teachers and students believed that the AfL methodology had improved the quality of student learning. In implementing formative assessment in primary schools, Ni Chróinín and Cosgrave (2013) stated that AfL supported worthwhile experiences in primary physical education, recording a dramatic change:

Teachers' perspectives on assessment changed dramatically as the value of assessment in enhancing the learning process was recognised. As a result of using assessment strategies, the teachers believed that the children learned more in their physical education classes and that they themselves became better teachers of physical education. (p. 230)

While AfL has resulted in positive and welcome changes in student learning within school physical education, this has not been without challenges. Teachers have reported finding the implementation time consuming (MacPhail & Halbert, 2010). It has also been reported that students believe assessment to be boring, subsequently making it difficult for the physical education teacher to engage them 'authentically' in

the process (James, Griffin, & Dodds, 2009). Further, student views of expected outcomes of physical education, and the knowledge produced through participation in physical education (Quennerstedt, 2013), can be quite different from learning outcomes stated in the official curriculum. The lack of alignment and the connected problems for validity and equality in assessments, most often with a reference to Bernstein's (1975) three message systems (curriculum, pedagogy and assessment), is well documented in physical education (Annerstedt & Larsson, 2010; Hay & Macdonald, 2008; Redelius, Fagrell, & Larsson, 2009; Redelius & Hay, 2009). Conscious that the question of validity and the AfL strategies employed are dependent of both the actual curriculum, and the wider political, social and cultural contexts, the next section provides a brief presentation of educational assessment and physical education in Norway in a bid to set the context of the study.

### **The Norwegian context**

In 2006 the Norwegian school reform 'The Knowledge Promotion' introduced new national curriculums for all subjects, leading to an increased focus on assessment in part due to accountability perspectives and neoliberal ideas guiding educational reforms in many capitalist welfare states (Møller & Skedsmo, 2013). At the same time, the new assessment regulations were informed directly by the research of Black and Wiliam (1998), and there was an articulated goal to change assessment in schools from an over-emphasize on assessment of learning to an emphasize of AfL (Utdanningsdirektoratet, 2011). The 2006 reform resulted in student assessment becoming one of the main discussion areas among physical education teachers due to the challenging nature of reconsidering their practices (Arnesen et al., 2013).

The Norwegian physical education curriculum is characterized by a broad content area that can be described as a multi-activity model where the students are to learn about body and health through the use of their bodies. The experiences during physical education classes should prepare young students to know, and be able to participate in, the most significant part of Norwegian movement culture, and motivate lifelong interest in physical activity. The national curriculum's 'competence aims' describe the expected learning outcomes and constitute the basis for all assessment (summative and formative) in the subject.

Students obtain grades from year eight onwards (aged 12 years), and the grading scale is divided into six steps expressed from 1 to 6 (1 expresses very low competence and 6 expresses eminent competence in the subject). The consideration of educational assessment and how to organize teaching is highly decentralized in Norway, with schools expected to develop their own local work plan which includes the criteria for assessment and grading. To a large extent, student assessment is dependent on the assessment of the individual teacher.

## **Methodology**

The authors share the assessment stories of three physical education teachers, 'Geir', 'Ronald' and 'Leif' (pseudonyms). The selection of these teachers was as a result of analyses of focus groups conducted as part of a larger project which aims to study both students' and teachers' perceptions of assessment in physical education across six upper secondary schools in Norway.

The main project included six focus groups with a total of 23 physical education teachers. In two focus groups, three teachers started to talk unprompted about AfL. A decision was made to focus more closely on their interactions to explore their use of the term AfL to describe what they were doing in physical education.

Ronald and Leif worked in the same school and participated in a focus group with four other physical education teachers, while Geir and one other colleague constituted another focus group. Focus groups were facilitated by the first author, recorded with a digital voice recorder and transcribed verbatim. Due to the more active role of the researcher in the focus group Geir contributed to the focus group could be more appropriately termed a group interview (Parker & Tritter, 2006). The focus group (1 hr 7 min) including Geir was conducted in a meeting room at his school while a classroom was used at Ronald's and Leif's school for a focus group (1 hr 40 min).

The first author translated assessment-relevant sections of the transcripts into English. Informed by reflexive interpretation (Alvesson & Sköldberg, 2009), we have downplayed the importance of collected data as focus group material, and use transcripts as 'dialogue partners' in the following analytical approach. Both authors individually completed a thematic analysis in relation to the research question and AfL key strategies, before sharing their emerging understandings with each other.

Similar to Torrance and Pryor's (1998) investigation of formative assessment through interviews with teachers, we experienced that the individual teacher can represent ambiguous and sometimes contradictory views of assessment, as well as apply different rationales to the same practice. In a bid to portray some of the complexity and tensions, and how the three teachers constructed their use of the term AfL, we experimented with different ways of conveying the findings. In part inspired by the approaches in Dowling, Fitzgerald, and Flintoff (2012), we agreed to retell an assessment story for each teacher based on their interactions throughout the focus groups. The stories are specifically structured to reflect AfL strategies shared in [Table 1](#), and we recognize the constructed assessment stories as part of the analysis. They could be 'true' in the sense that they are generated from real events and teachers' conversations, and we quote passages from the respective focus groups and try to keep the vocabulary and syntax close to that of the physical education teacher. The intention is to move the focus from information collection and reporting of what the teachers share to encouraging the readers to revisit and rethink their understanding and beliefs about learning and assessment practices. It is anticipated that the constructed assessment stories will allow us to consider the extent to which teachers' references to AfL align with AfL theory and discuss how the stories relate to the wider contexts of physical education and movement culture.

**Leif's story – curious and prepared to experiment with AfL strategies**

Leif is in his late thirties and has been teaching physical education for eight years. Leif works with Ronald at a relatively new upper secondary school in Oslo and they have been colleagues for two years. With approximately 900 students at the school (boys and girls aged 15–19 years), there are four other physical education teachers at the school.

Leif reports the level of support from the school leadership of the new school as disappointing. Without a clear school policy for assessment, the physical education teachers have attempted to agree a common structure (four main teaching units of discrete content over the year) for physical education. Each unit ends with a 'test', a planned assessment situation where the students have to demonstrate knowledge and practical skills. In attempting to explain what this new and different discourse on assessment in physical education entailed, Leif commented:

One important aspect was that they [students] would meet these questions again at the end of the year, exactly the same. This way we tried to make it assessment for learning. They knew (...) at some point they would get the same questions again, so if I want to do better, I can check my old test and assure that I am capable of answering what I was supposed to back then. This way it was supposed to be assessment for learning, that was the thought behind it all.

Leif's explanation is an example of the formative use of a summative test, implying that the connection he makes to AfL is meaningful. Throughout the focus group, Leif communicated commitment to accommodating and arguing for what he referred to as 'skill-tests' (i.e. assessments that measure students' ability to perform a skill within a discrete drill). Due to the expectation of the skill tests informing the students where they are and where they are going next, Leif first not only presents them as an attempt to implement formative assessment practice but also states that 'the skill tests were carried out mainly because we were required to'. In this instance, the engagement in developing assessment becomes primarily a task of accountability and the requirements of documentation, rather than focusing towards how to use assessment to improve students' learning.

As the focus group develops, Leif extends the argument that there are more important aspects to assessment than solely assessing a student's ability at performing a particular skill. He explains how the physical education teachers try to include self-assessment, one of the main AfL strategies, as part of the skill tests and states that 'the fun part is that the students almost always have put down the same grade as me'. Leif admits that he struggles to be sufficiently precise to allow him and the students to share 'the same perception of what the grade 4 look like and when the student has arrived at 4'. The examples of self-assessment provided by Leif are all concerned with grading, resulting in him describing summative assessment practice.

Leif is also a qualified mathematics teacher, and makes a comparison of assessment in the two subjects he teaches:

I teach mathematics too, and I think the assessment contexts are two opposites. In mathematics I can sit down at the table and a cup of coffee and a sheet, so I can sit and consider; 'that's correct', 'that's wrong', and then I count up and get to 97% correct answers which give grade 6. (...) Physical education is the subject of assessing moments, or 'short-cuts', right? Twenty seconds, I observed Lisa who did some skilled and entertaining moves. Then something happens in the changing rooms which needed my attention. Later, I ask; 'hell, what just happened there?' What was it that Lisa did, was the short moment showing an important side of her movement competence?' Finding a system then, which systematises and records such short moments which in the end also is supposed to constitute the final grade.

Comparing assessment across the subjects, Leif highlights the complexity of the teaching and learning contexts and assessment challenges in physical education, noting that what is at the core of assessment in each school subject is presented differently. While ongoing assessment has a prominent role in physical education, the focus in mathematics appears to be concerned with arriving at a final grade. While Lisa 'doing some skilled and entertaining moves' during one lesson can influence assessment of Lisa in physical education, opportunities to convey learning in less-structured learning contexts are not accommodated in mathematics.

Leif is aware of the concerns related to skill tests as planned assessment situations and highlights the importance of understanding how continual assessment can be used to inform students' learning:

I just want to make a little comment to the on-going assessment, as it was also something I did make clear, that if you do not have a good day at the skill-tests, (...) and make a much worse impression than you are able to, then I will know. I will know from how you participate and perform on activities during the weekly lessons, and that overrides what you do on a test. I tried to be very clear on this issue, although students may not have seen it that way.

In concluding, while Leif expresses a continuous curiosity and will to continue experimenting with assessment, his practice also conveys a lack of systematic reflection, and perhaps a vocabulary, to allow him to interrogate effective assessment further.

### **Ronald's story – tensions between enacting different assessment cultures**

Ronald teaches in the same school as Leif, is in his late fifties and has been a full-time physical education teacher in upper secondary school for the last 20 years. Ronald is informed about AfL and positions himself somewhat differently from Leif in the ways in which he has attempted to implement AfL into practice. From the outset, Ronald appears to be an experienced physical education teacher who has recently found inspiration to explore new approaches to assessment. He admits that the previous year was the first time he did not complete a pen and paper test with the physical education students and, 'in relation to assessment which enhance learning, or assessment for learning, (...) I tried out some things I have not done before, for

example, that students are being engaged in self-assessment'. Ronald continues by providing an example of what this can look like:

After finishing a task, for example, they [students] as a group sit down and put down some points. What was good? What could have been better? Then I can do the same, and then we can talk about it. And so, not to agree on a grade, but we both put down a suggestion for a grade. So we'll see if it matches, (...) if I – who, of course always assesses correctly (laughter) – have graded a four and the student have six, so there we at least have something to talk about, and if we both have, for example five, then yes, we might suggest that it must be quite correct.

When it comes to clarifying criteria and providing feedback in accordance with the competence aims in the physical education curriculum, Ronald is supportive of the structure commenting that learning goals related to the competency aims heighten student learning. Ronald argues that it is easier to provide feedback and tutor students towards achieving competence aims when learning goals are more manageable and concrete. He also makes reference to the practical nature of physical education and how, rather than rely on providing oral feedback to the students, teachers will 'prepare a practical learning experience that makes them progress'. Ronald does admit to the inherent challenge in verbally communicating to students the connection between competence aims and enacted learning experiences.

Ronald is also conscious about assessment and grading in physical education being a way to communicate what is to be regarded as important in the subject, particularly when the 2006 school reform explicitly articulated that the level of student effort should not influence grading in the subject. Ronald does not agree with this and admits to accommodating level of effort in his assessments:

Halfway through the semester I noted down an assessment, which I called participation and activity level, on each student, and I did the same again at the end of the semester. This is to ensure that this aspect also would be considered, as well as pen and paper tests and workbook, how fast they were able to run and how they performed in basketball, floorball and stuff like that.

Reflecting the criterion-referenced and decentralized assessment system in Norway, Ronald explained how he also has rethought the concept of fairness:

Ronald: In terms of what can be difficult by student assessment (laughter) but at least, according to how my understanding gradually has developed, fairness is a challenge. Some might have the view that fair would be if everything was equal, a benchmark to assess for, some think that's fair, right?

Researcher: Yes?

Ronald: But the way I have come to understand it, eventually, is well that fairness could also be that students would get a clue of the overall aims of our practice and the assessment procedures and that this is prepared in advance so that they have something to relate to. Whether it is one or the other, it will vary between schools and classes, and it will, because that's the system now, it varies yes, because we here at our school determine what is fair and a teacher and student at another school

determine what they consider to be fair. In this way the crucial point is that they are aware of what is going to happen and what the criteria for success are.

Ronald has made space to engage students in assessment and explore, through his own practice, how to integrate AfL strategies into teaching and learning. He is an experienced teacher motivated to change assessment practice, believing that assessment can become more effective and make an important contribution to student learning in physical education. He does admit that he, students and some colleagues are still drawn at times to the traditional assessment culture where the allocation of a final grade has dominated practice. Admitting that he had not encouraged or facilitated peer-assessment in his classes, Ronald expresses this tension between the different assessment cultures or traditions:

I have not used peer-assessment, but I remember I asked [the students] about it, and the reaction to it was relatively negative. However, the frame was that they should put a grade on each other's performance or competence. It's very easy when we talk about assessment, to think grading, and start discussing grading. But if we really mean an assessment, more like guidance, then it, of course, becomes another matter. Then I think peer-assessment will be ok, in for example basketball and whatever else we do. It will be very good for students to tutor each other. But, you know, we do not often think about it that way, because when we talk about assessment we talk about numbers.

Laughing, he adds that 'it's really misunderstood, because it's the other assessment that is the key, really, to promote learning, right?'

### **Geir's story – an evolving disposition towards formative assessment**

Geir is in his late thirties and a full-time physical education teacher of 14 years who works in an upper secondary school in Oslo. The school has a relatively small student intake of approximately 300 boys and girls aged between 15 and 19 years. There are two physical education teachers in the school and consequently the focus group that Geir participated in included only his female colleague and the researcher. The physical education programme at the school entails a broad range of activities, mostly sport orientated. Geir and his colleague express that the level of support from the school leadership is good, and the school has participated in a national project for developing assessment practice (Utdanningsdirektoratet, 2011). They also note that they have good informal opportunities to discuss and reflect with each other on assessment practices.

Geir is quick to allude to the tension with the sole use of a 'skill track' (similar to what Leif had termed a 'skill-test') to encompass and demonstrate learning and the concept of ongoing assessment promoted by AfL:

We practice (...) the basic skills and then the student can test their skills in a skill track at the end (...) this is the formal assessment situation. (...) But I try to tell students that I also do a more comprehensive assessment [than that of a skill track/

skill test]. So that you may well have a bad day on test, if one has shown earlier in rehearsal lessons that he or she possesses the skills, I take that into account. It's a bit like that in relation to assessment for learning, that we are concerned with practice and rehearsal prior to testing.

While Geir appears to favour a skill track as a feasible assessment for physical education, he does provide instances where he engages with a more eloquent understanding of AfL when prompted on considering the relationship between the awarding of a grade and feedback:

What's important, also relating to feedback, is that the students know what is expected of them, that they know what they are going to be measured in relation to. So in advance, at the beginning of a period, we present the competence aim, and so, at least I try to break down the competence aim into learning objectives, then, it all becomes a little more concrete for students. (...) And so we present assessment criteria, what they must convey to get the different grades. Sometimes this is presented orally. Sometimes they get it on paper. (...) And then, during practice it is all along about giving them feedback on what they do well and things they can do better.

Geir explains how his teaching is somewhat guided by the 'two stars and a wish' concept. A literal translation of the concept is to acknowledge at least two things that the student is doing well before introducing a focus on something they can work on to improve their performance. Geir shares an evolving disposition towards formative assessment when he alludes to introducing self-assessment, admitting that he is still exploring the concept. He has used a 'thumbs up, thumbs down' technique to garner from students their assessment of their own experience in a task (i.e. thumbs up meaning they are pleased with their performance of the task and thumbs down that they are not pleased). Geir is conscious that student self-assessment is not solely for the benefit of the teacher to assess how individual students are reacting to a task but rather that it works to the benefit of the student, 'There is more about awareness for their own sake, and I think, that's a little of what self-assessment is all about'. Geir also admits to only recently introducing peer-assessment and notes the extent to which he is happy with how the students have reacted to such involvement.

The importance of sharing criteria for success is inadvertently mentioned by Geir when he comments that he believes that those students who are dissatisfied with their skill track grade are those students who did not appreciate what the assessment criteria was, believing they have invested sufficient effort to warrant a better grade. Geir goes on to consider ways in which physical education should assess effort more comprehensively, believing that 'effort, and being active and participating is central to promote more physical activity, and health and lifestyle'. He considers how assessment in physical education can best accommodate what the subject area is/should be concerned with promoting:

If we were to talk about what one would wish that physical education should be then. What role it could play in the school? So, communicating the importance of

being physically active is part of it. To understand how you can benefit from it, and also to motivate the interest of sports and activities. One idea is that the students can be exposed to some activities which they would like to participate in or do on their own. But certainly, there must be knowledge and skills as well.

Geir alludes to the tensions that arise when he explains the value of skill tracks to students who question the appropriateness of such assessments when, dependent on the skill, students can score highly in the assessment due to already being proficient performers in specific skills. Interesting, it is the students who appear to be requesting a reconsideration of the appropriateness of skill tracks in this instance and not the teacher. Somewhat contradictory to what Geir alluded to previously, with respect to what physical education should be acknowledged as promoting, Geir advises students that it is not enough for them to show up in class and participate but rather that they need to be able to perform a skill at a certain level to achieve a respectable grade, advising students:

So between lessons, even if you do not get it specified as homework you need to go home and [practice]. They must understand that according to skills, well, it's not enough to be present and to show much good effort.

Geir's focus on sharing the criteria for success, and few other components of AfL, appears to be somewhat responsible for his lack of engagement in considering assessments that encourage and acknowledge student investment and effort in a task. Geir believes that if students convey a level of effort it is likely that their skill set will improve.

Geir believes that the final assessment point in physical education in his school has improved due to a more systematic approach to formulating assessment criteria and clarifying how learning is to be assessed, both of which are identified as AfL strategies. However, he appears unable to believe that 'students are ready to deal with the abstract level of the competence aims' and subsequently continues to rely on the teaching of intricate and discrete skills.

## **Discussion**

Due to the selection criteria for the three teachers involved in this study (i.e. they had, unprompted, referred to AfL in focus groups concerned with assessment), it was not surprising that all three teachers revealed a high level of engagement and reflection with assessment, and the associated tensions in changing their own assessment preferences and practices. Throughout the focus groups, they provided descriptions of engagement with all five key strategies of AfL (Wiliam & Thompson, 2008). It is not clear whether they recognized a more general need for change in physical education (Kirk, 2010; Säfvenbom et al., 2014), or if their engagement with AfL stems mostly from what they are expected to do according to official requirements and new assessment regulations as set out in the previously mentioned Norwegian school reforms.

Recalling the double-edged sword metaphor about the mandatory aspects of physical education (Säfvenbom et al., 2014), it is somewhat surprising that analysis of the focus group interactions did not reveal critical reflection about how assessment practice can, to follow the metaphor, end up on the ‘wrong’ side of the sword. In other words, the teachers did not reflect upon how assessment can have unintended, and for some students harmful, consequences that may contribute to negative attitudes towards movement contexts in general. This indicates that the physical education teachers need to develop awareness of assessment as a process in which distribution of power is inherently disproportional, and challenge the ‘naturalness’ of assessment practices, performances and outcomes (Hay & Penney, 2013).

AfL strategies and the assessment stories of Leif, Ronald and Geir lead us to focus on engaging the students in a more active role as learners in physical education. This is one way of empowering students in regard to assessment and is potentially an effective way of linking movement and learning in physical education to their life as ‘movement consumers’ outside school. For example, teachers can facilitate learning experiences where students are invited to present and discuss different movement contexts that they already are part of. Geir, Ronald and Leif all alluded to the value of encouraging students to engage with where they currently reside in their learning and to pursue the goals that they are striving to achieve. However, conscious of the limited amount of self- and peer-assessment, it becomes clear that even Geir, who appeared to have engaged more fully with such strategies, admitted that he was still in an exploratory phase when it comes to promoting such forms of assessment. Leif and Ronald seemed to engage with self-assessment only at the end of the unit when they asked students how they would grade themselves. This could be construed with the popular notion that teachers consider assessment synonymous with providing a final grade.

Understanding the distinction between the processes of assessment and grading is an important part of teachers’ assessment literacy (Hay & Penney, 2013), but it also resonates an unsolved tension between the summative and formative purposes of assessment at the national policy level. All three teachers conveyed an appreciation of formative assessment’s contributions to summative assessment, acknowledging that it was not necessary for the final ‘skill-tests’ performance to be solely dependent on the enactment of a summative assessment. There were a number of instances where AfL strategies were compromised due to the external requirements of assessment. Leif explained that he sometimes completed the feedback component of AfL more from his obligation to address the national curriculum requirements than as a mechanism through which to encourage students’ own learning. The broader structures surrounding assessment become important if we are to understand such tensions and ambiguous descriptions regarding AfL in the assessment stories, appreciating that there was little support at the teacher level following the 2006 reform (Møller & Skedsmo, 2013). In addition, physical education teachers motivated to develop their formative assessment practices faced an almost complete lack of examples or descriptions of how to do AfL in movement contexts (Georgakis & Wilson, 2012). Teachers were, as conveyed in our assessment stories, predetermined

to explore, rather than implement, AfL in physical education. In this regard Geir, Ronald and Leif can be interpreted as reflective professionals trying to move away from physical education as 'a field where other abilities are valued than the ones to be found in the syllabus' (Redelius et al., 2009, p. 259). From an educational research perspective, we appreciate that they place learning at the foreground and will use assessment to clarify what is to be valued in the subject. At the same time, we are concerned to observe how certain practices developed in the name of AfL actually narrow the concept of learning, mostly under the influence of a still remaining 'meritocratic mode' (Crum, 1993), or in Kirk's (2010) terms 'physical education-as-sport-techniques' (p. 4).

The teachers' attempt to enact and interpret AfL appears to be constricted to the traditional emphasis in Norwegian physical education on linking assessment to classifying and grading students, as well as feeling pressurized to adjust practices to official requirements concerning documentation and accountability (Arnesen et al., 2013). In some cases, the teachers' commitment to clarifying and sharing criteria on behalf of the national curriculum seemed to be more important than planning and teaching worthwhile, meaningful and relevant student learning experiences. To a degree, it appears that the notion of 'learning' in physical education is confused with 'criteria compliance'. If so, the potential of AfL as being 'oriented towards greater equity within and beyond physical education' (Hay & Penney, 2013, p. 110) vanishes, replaced by what looks like assessment for assessment's sake.

### **Concluding remarks**

This paper has shared the stories of three Norwegian physical education teachers exploring 'alternative' assessment practices which they identify as AfL. All three teachers appeared to be well informed of the national curriculum, assessment regulations and the purposes of implementing AfL. However, it was evident that their understanding and enactment of AfL key strategies was somewhat constricted. If the physical education profession is unable to find ways in which to address this, we risk not supporting assessment reform with tragic consequences for student learning. It is therefore crucial that physical education teachers focus on the individual learner and appropriate learning experiences. There is a need of thorough teacher planning, preferable with systemic moderation and support for students to take on new roles and responsibilities in their own learning. Recalling Bennett (2011), we reiterate that intellectual and instructional tools 'ought to be specifically tuned for the domain in question' (p. 16), and hope more stories regarding physical education practices can help us figure out what this will look like.

We are concerned to observe in this instance that no real effort appears to have been made to create the supportive environment in physical education for successful implementation of the AfL strategies. Our concern is underpinned by research pointing to teacher professional development as a prerequisite for successful implementation of AfL (Smith, 2011), and the need of embedding AfL in learning theory (Gardner, 2012). This may well be a challenge, acknowledging

that not only are most existing theories of learning defined cognitively, but also that learning connected to physical education and activity is, to a large extent, practical and embodied, and also linked to the powerful discourses of sport and related areas such as health (Quennerstedt et al., 2014). Similar to Hay (2006) and López-Pastor et al. (2013), and acknowledging the limitation of a study on three physical education teachers, we advocate the need for more research on AfL related to physical education. It is only through further exploration that we will begin to track the extent to which AfL, as an integrated part of pedagogical practice, can contribute to an idea of physical education as a ‘planned introduction to movement culture’ (Crum, 1993, p. 352).

### Acknowledgements

We are very grateful to the two anonymous reviewers for focused and constructive comments on the original version of the paper.

### References

- Alvesson, M., & Sköldberg, K. (2009). *Reflexive methodology: New vistas for qualitative research*. London: SAGE.
- Annerstedt, C. (2008). Physical education in Scandinavia with a focus on Sweden: A comparative perspective. *Physical Education and Sport Pedagogy*, 13, 303–318. doi:10.1080/17408980802353347
- Annerstedt, C., & Larsson, S. (2010). ‘I have my own picture of what the demands are ...’: Grading in Swedish PEH – problems of validity, comparability and fairness. *European Physical Education Review*, 16(2), 97–115. doi:10.1177/1356336X10381299
- Arnesen, T. E., Nilsen, A. K., & Leirhaug, P. E. (2013). ‘Den læreplanen som ikkje kan tilpassast mi undervisning, finst ikkje.’ Vurdering og undervisning i kroppsøving etter Kunnskapsløftet [Assessment and teaching in physical education after the reform the Knowledge Promotion]. *Tidsskriftet FOU i praksis*, 7(3), 9–32.
- Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Retrieved from [http://assessmentreformgroup.files.wordpress.com/2012/01/10principles\\_english.pdf](http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf)
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5–25.
- Bernstein, B. (1975). *Class, codes and control, volume 3: Towards a theory of educational Transmission*. London: Routledge & Kegan Paul.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7–73.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Education, Assessment, Evaluation and Accountability*, 21, 5–31.
- Chan, K., Hay, P., & Tinning, R. (2011). Understanding the pedagogic discourse of assessment in physical education. *Asia-Pacific Journal of Health, Sport and Physical Education*, 2(1), 3–19.
- Crum, B. J. (1993). Conventional thought and practice in physical education: Problems of teaching and implications for change. *Quest*, 45, 339–356. doi:10.1080/00336297.1993.10484092
- Dowling, F., Fitzgerald, H., & Flintoff, A. (2012). *Equity and difference in physical education, youth sport and health: A narrative approach*. London: Routledge.
- Gardner, J. (Ed.). (2012). *Assessment and learning*. London: SAGE.
- Georgakis, S., & Wilson, R. (2012). Australian physical education and school sport: An exploration into contemporary assessment. *Asian Journal of Exercise & Sport Science*, 9(1), 37–52.

- Green, K. (2014). Mission impossible? Reflecting upon the relationship between physical education, youth sport and lifelong participation. *Sport, Education and Society*, 19, 357–375.
- Hay, P. (2006). Assessment for learning in physical education. In D. Kirk, D. Macdonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 312–325). London: SAGE.
- Hay, P., & Macdonald, D. (2008). (Mis)appropriations of criteria and standards-referenced assessment in a performance-based subject. *Assessment in Education*, 15, 153–168. doi:[10.1080/09695940802164184](https://doi.org/10.1080/09695940802164184)
- Hay, P., & Penney, D. (2013). *Assessment in physical education. A sociocultural perspective*. New York: Routledge.
- James, A. R., Griffin, L., & Dodd, P. (2009). Perceptions of middle school assessment: An ecological view. *Physical Education and Sport Pedagogy*, 14, 323–334. doi:[10.1080/17408980802225792](https://doi.org/10.1080/17408980802225792)
- Kirk, D. (2010). *Physical education futures*. London: Routledge.
- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: A review of international literature. *Sport, Education and Society*, 18, 57–76.
- 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 and Practice*, 17(1), 23–39.
- Møller, J., & Skedsmo, G. (2013). Modernising education: New public management reform in the Norwegian education system. *Journal of Education Administration and History*, 45, 336–352.
- Ni Chróinín, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, 18, 219–233.
- Parker, A., & Tritter, J. (2006). Focus group method and methodology: Current practice and recent debate. *International Journal of Research and Method in Education*, 29(1), 23–37. doi:[10.1080/01406720500537304](https://doi.org/10.1080/01406720500537304)
- Quennerstedt, M. (2013). Practical epistemologies in physical education practice. *Sport, Education and Society*, 18, 311–333.
- Quennerstedt, M., Öhman, M., & Armour, K. (2014). Sport and exercise pedagogy and questions about learning. *Sport, Education and Society*, 19, 885–898. doi:[10.1080/13573322.2013.847824](https://doi.org/10.1080/13573322.2013.847824)
- Redelius, K., Fagrell, B., & Larsson, H. (2009). Symbolic capital in physical education: To be, to do or to know? That is the gendered question. *Sport, Education and Society*, 14, 245–260.
- Redelius, K., & Hay, P. (2009). Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review*, 15, 275–294. doi:[10.1177/1356336X09364719](https://doi.org/10.1177/1356336X09364719)
- Säfvenbom, R., Haugen, T., & Bulie, M. (2014). Attitudes toward and motivation for PE. Who collects the benefits of the subject. *Physical Education and Sport Pedagogy*. Advance online publication. doi:[10.1080/17408989.2014.892063](https://doi.org/10.1080/17408989.2014.892063)
- Scriven, M. (1967). The methodology of evaluation. In P. A. Taylor & D. M. Cowley (Eds.), *Reading in curriculum evaluation* (pp. 39–83). Dubuque, IA: Wm. C. Brown.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. doi:[10.3102/0013189X029007004](https://doi.org/10.3102/0013189X029007004)
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. doi:[10.3102/0013189X015002004](https://doi.org/10.3102/0013189X015002004)
- Smith, K. (2011). Professional development of teachers – A prerequisite for AfL to be successfully implemented in the classroom. *Studies in Educational Evaluation*, 37, 55–61. doi:[10.1016/j.stueduc.2011.03.005](https://doi.org/10.1016/j.stueduc.2011.03.005)

- Tannehill, D., van der Mars, H., & MacPhail, A. (2013). *Building effective physical education programs*. Sudbury, MA: Jones & Bartlett.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Philadelphia, PA: Open University Press.
- Utdanningsdirektoratet. (2011). *Grunnlagsdokument: Satsingen Vurdering for læring 2010–2014* [Background document: Assessment for learning 2010–2014]. Oslo: Author. Retrieved from <http://www.udir.no/PageFiles/Vurdering%20for%20laring/Dokumenter/Nasjonalt%20satsing/2/Grunnlagsdokument%20for%20satsingen%20Vurdering%20for%201%3a6ring%20okt%20%2020111.pdf>
- Ward, G. (2014). Learning movement culture: Mapping the landscape between physical education and school sport. *Sport, Education and Society*, 19, 569–604. doi:10.1080/13573322.2012.690342
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3–14. doi:10.1016/j.stueduc.2011.03.001
- Wiliam, D., & Thompson, A. (2008). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum.