



Assessment quality and practices in secondary PE in the Netherlands

Lars B. Borghouts, Menno Slingerland & Leen Haerens

To cite this article: Lars B. Borghouts, Menno Slingerland & Leen Haerens (2017) Assessment quality and practices in secondary PE in the Netherlands, *Physical Education and Sport Pedagogy*, 22:5, 473-489, DOI: [10.1080/17408989.2016.1241226](https://doi.org/10.1080/17408989.2016.1241226)

To link to this article: <https://doi.org/10.1080/17408989.2016.1241226>



Published online: 04 Oct 2016.



[Submit your article to this journal](#)



Article views: 1019



[View Crossmark data](#)



Citing articles: 4 [View citing articles](#)



Assessment quality and practices in secondary PE in the Netherlands

Lars B. Borghouts^a, Menno Slingerland^a and Leen Haerens^b

^aSchool of Sport Studies, Fontys University of Applied Sciences, Eindhoven, the Netherlands; ^bDepartment of Movement and Sport Sciences, Ghent University, Ghent, Belgium

ABSTRACT

Background: Assessment can have various functions, and is an important impetus for student learning. For assessment to be effective, it should be aligned with curriculum goals and of sufficient quality. Although it has been suggested that assessment quality in physical education (PE) is suboptimal, research into actual assessment practices has been relatively scarce.

Purpose: The goals of the present study were to determine the quality of assessment, teachers' views on the functions of assessment, the alignment of assessment with learning goals, and the actual assessment practices in secondary PE in the Netherlands.

Participants and setting: A total of 260 PE teachers from different schools in the Netherlands filled out an online Physical Education Assessment Questionnaire (PEAQ) on behalf of their school.

Data collection: The online questionnaire (PEAQ) contained the following sections: quality of assessment, intended functions of assessment, assessment practices, and intended goals of PE.

Data analysis: Percentages of agreement were calculated for all items. In addition, assessment quality items were recoded into a numerical value between 1 and 5 (mean \pm SD). Cronbach's alpha was calculated for each predefined quality aspect of the PEAQ, and for assessment quality as a whole.

Findings: Mean assessment quality (\pm SD) was 3.6 ± 0.6 . With regard to the function of assessment, most PE teachers indicated that they intended using assessment as a means of supporting the students' learning process (formative function). At the same time, the majority of schools take PE grades into account for determining whether a student may enter the next year (summative function). With regard to assessment practices, a large variety of factors are included when grading, and observation is by far the assessment technique most widely applied. A minority of PE teachers grade students without predetermined assessment criteria, and usually criteria are identical for all students. There is an apparent discrepancy between reported PE goals and assessment practices; although increasing students' fitness levels is the least important goal of PE lessons according to the PE teachers, 81% reports that fitness is one of the factors being judged. Conversely, while 94% considers gaining knowledge about physical activity and sports as one of the goals of PE, only 34% actually assesses knowledge.

Conclusions: Assessment in Dutch PE is of moderate quality. The findings further suggest that PE teachers consider assessment *for* learning important but that their assessment practices are not generally in line

ARTICLE HISTORY

Received 24 March 2015
Accepted 8 September 2016

KEYWORDS

Assessment; constructive alignment; assessment for learning; quality

with this view. Furthermore, there seems to be a lack of alignment between intended learning outcomes and what is actually being valued and assessed. We believe that these results call for a concerted effort from PE departments, school boards, and the education inspectorate to scrutinise existing assessment practices, and work together to optimise PE assessment.

Introduction

Curriculum, pedagogy, and assessment have been described as three interrelated, fundamental dimensions of quality physical education (PE) (Penney et al. 2009). However, multiple researchers have suggested that assessment quality in PE is worrisome (Hay and Penney 2009; Thorburn 2007; Veal 1988), and that physical educators struggle to meet the demands for a reliable and valid grading system (Annerstedt and Larsson 2010; Dinan-Thompson and Penney 2015). Indeed, assessment has been referred to as ‘one of the most fraught and troublesome issues physical educators have had to deal with over the past 40 years or so’ (López-Pastor et al. 2013, 1). These concerns coincide with a growing emphasis on assessment in education, due to the increasing global prominence of discourses of accountability and standardisation within education (Hursh 2005; Roberts-Holmes and Bradbury 2016).

Assessment quality

For teachers and students to be well informed and be able to make valid judgements about the learning process and its outcomes, assessment quality is paramount. Knowledge of assessment quality and efficacy is considered part of ‘assessment literacy’ (Stiggins 1999), which has long been viewed as an important characteristic of effective teachers. A need for investment in assessment literacy for PE teachers has recently been advocated (Collier 2011; Dinan-Thompson and Penney 2015). When considering whether assessment quality in PE is indeed suboptimal (Hay and Penney 2009; Thorburn 2007; Veal 1988), it seems prudent to consider what determines quality assessment.

According to Stiggins et al. (2007), assessment quality can be defined according to five central quality aspects: *clear purpose*, *clear targets*, *sound design*, *effective communication*, and *student involvement*. These quality aspects have subsequently been applied to PE by other scholars (Collier 2011; Melograno 2007). *Clear purpose* means that the *why* of the assessment being conducted is apparent to all involved, and that this purpose fits into the larger educational picture. *Clear targets* ensure that achievement expectations are transparent and completely defined. A *sound design* of the assessment refers to the validity and reliability of measurement, as well as its alignment with learning goals. *Effective communication* is taking place when communication is planned as an integral part of the assessment, and when results can be easily accessed and understood by the intended users. Finally, *student involvement* in assessment incorporates diverse elements such as explanation of learning targets and assessment procedures to students, offering descriptive feedback and applying self- or peer-assessment.

Assessment functions

As a consequence of neo-liberal political reforms, standardised tests and national curricula have been introduced in many countries over the last three decades (Dinan-Thompson and Penney 2015; Hursh 2005). In secondary education, final exams are increasingly used to determine the students’ readiness for (different levels of) tertiary education (Berliner 2011). This has led to a debate about the possible negative side effects of high-stakes testing (Nichols and Berliner 2007), and the place of assessment within the curriculum. It is well established that assessment can serve various functions (Newton 2007).

Firstly, when assessment is used for certification and selection, it is referred to as assessment of learning (AoL) or summative assessment (Stiggins et al. 2007). Summative assessment typically takes place at the end of a unit, and often results in grades that contribute to determining a student's performance, for example, via a report card. This function of assessment is in line with a prominent discourse of educational accountability.

Secondly, assessment can be used to *support the students' learning process*. In order to achieve this, it should be constructed as a part of the learning process instead of merely as its conclusion. In that case, assessment *for* learning (AfL) or formative assessment is taking place (Black and Wiliam 1998; Wiliam 2011). According to Broadfoot et al. AfL constitutes

[...] 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. (2002, 2)

This is in line with Hattie and Timperley's (2007) notion of the purpose of feedback, namely to reduce discrepancies between current understandings and performance, and a learning goal. Feedback is therefore intricately related to AfL, encompassing not only teacher feedback, but also active involvement from students (e.g. through self- and peer-assessment) (Black and Wiliam 2009; Pat El et al. 2013). AfL is considered to be a powerful tool for teachers to enhance students' learning (Black and Wiliam 2009; Cauley and McMillan 2010) and has been widely advocated not only within education in general, but also specifically in PE (Collier 2011; Hay 2006; Hay and Penney 2009; Ní Chróinín and Cosgrave 2012). Unfortunately, it has been noted at the same time that research into AfL is rather scarce in this field and that its implementation is perceived as challenging by teachers (Hay 2006; MacPhail and Halbert 2010). Thirdly, assessment can be used as a means for *evaluation of curriculum effectiveness* (Newton 2007). To be able to do so, it is imperative that curriculum goals, learning activities, and assessment are in accordance.

Alignment of PE goals and assessment

Constructive alignment (Biggs 1996) is considered essential for a coherent curriculum and quality (physical) education (Lund 1992; Penney et al. 2009). In an aligned curriculum, learning goals are clearly described and teaching and learning activities are chosen that are likely to realise those goals. Furthermore, assessment tasks are designed in such a way that they allow teachers and learners to see to what extent the goals have been reached. It has previously been suggested that in PE practice, intended goals and what is actually assessed may not be very well aligned (Chan, Hay, and Tinning 2011; Redelius and Hay 2012). This renders assessment ineffective (Collier 2011; Hay and Macdonald 2008; Hay and Penney 2009). Matanin and Tannehill (1994) concluded from their early research with 11 US PE teachers that teachers gained little knowledge about what students accomplished and that they graded students on attendance, dress, participation, and effort rather than knowledge and skills. A descriptive case study in elementary PE in MA, USA (James, Griffin, and Dodds 2008) showed teachers shifted their espoused agendas (focus on student learning) to an enacted agenda that focused on safety and completing tasks. As a result of this shift, students were not assessed in the manner that the teachers had planned. Consequently, there was no alignment between the teachers' espoused agenda, lesson tasks, and assessments. More recently, research suggested poor constructive alignment in Australian PE and School Sport (Georgakis and Wilson 2012).

Given this apparent lack of constructive alignment it is not surprising that research has shown that students may seem confused or ill-informed about PE goals and what its assessment is based on (Erdmann, Chatzopoulos, and Tsormbatzoudis 2006; Redelius and Hay 2012; Zhu 2015). Students in these studies did not perceive the official standards and criteria as the predominant basis for assessment, and their perspectives of grading were inconsistent with their own conception of achievement in PE. Insufficient constructive alignment has previously been suggested to arise as a result from low levels of accountability (Dinan-Thompson and Penney 2015).

PE in the Netherlands

Dinan-Thompson and Penney (2015) investigated assessment literacy in primary PE with 18 teachers in a regional area in Australia and argued that PE assessment is an ‘entrenched discourse in taken-for-granted practice’ and suggested that:

In contrast to the ways in which global discourses of testing and performativity can be seen to dominate thinking about assessment in literacy and numeracy [...] in PE, teachers appear to have far more relative autonomy. (12)

This relative autonomy of teachers is equally evident in secondary PE in the Netherlands, where the current study took place. Within the framework of attainment targets and examination requirements set by central government, schools in the Netherlands govern with a high level of autonomy. This implies that schools are fully responsible for the organisation of teaching and learning, and deployment of personnel and materials (EP-Nuffic 2015). Although there are national, uniform examinations for most subjects, both PE and the Arts form a notable exception to this. Schools are free to determine how PE is assessed and whether it is taken into account for yearly grade advancement, although all students must pass PE at a ‘satisfactory’ or ‘good’ level in order to graduate. At the same time, accountability within PE can be considered low in the Netherlands. The Dutch Inspectorate of Education periodically assesses potential problems that could affect the quality of education and a school’s capacity to assure and improve quality. It performs site visits to schools and publishes national evaluations about the educational system as a whole as well as at subject level.¹ It has been noted however that PE receives little attention from the Inspectorate (Brouwer et al. 2015), with the last national evaluation of secondary PE dating back to 1999.

Assessment practices

Given the relative autonomy of PE and its assessment in Dutch secondary schools, practices can be expected to vary widely between schools. Schools, and even teachers, may differ in the aspects taken into consideration when grading (e.g. motor skills, knowledge, game tactics), and in the modes of assessment applied (e.g. observation, written tests, portfolio). Furthermore, since secondary schools are free to determine whether PE is taken into account for yearly grade advancement, they may differ in the extent to which its assessment can be considered summative or formative.

Although the subject of assessment in PE has received ample attention in the international research literature, studies into actual assessment practices have been relatively scarce. Most of these studies were relatively small and/or date from more than a decade ago (Desrosiers, GenetVolet, and Godbout 1997; Imwold, Rider, and Johnson 1982; Kneer 1986; Matanin and Tannehill 1994; Mintah 2003; Veal 1988). Moreover, very few studies have been undertaken in Europe (Cassady, Clarke, and Latham 2004; Redelius and Hay 2012), and none in the Netherlands. Earlier studies have noted a predominance of assessment based on the subjective evaluation of aspects such as effort, preparedness, and sportsmanship (Imwold, Rider, and Johnson 1982; Matanin and Tannehill 1994; Veal 1988) and a low prevalence of knowledge testing and written assignments (Imwold, Rider, and Johnson 1982; Mintah 2003; Veal 1988). In addition, López-Pastor et al. (2013) have described that up to the early 1990s, Physical Fitness Tests (PFTs) were a popular form of assessment in PE.

It has been suggested that in PE, there is a high prevalence of standardised, product-oriented assessment practices such as fitness testing and the assessment of isolated technical skills (Loriente-Catalán and Kirk 2016; Penney et al. 2009). It has been argued that these forms of assessment lack meaningfulness to students because they do not relate to the world outside the school building (López-Pastor et al. 2013); in other words, they are not authentic. Although López-Pastor et al. (2013) have suggested that over the last three decades, more authentic forms of assessment have emerged, their review of alternative assessment practices concluded that it remains to be elucidated to what extent these approaches have become standard practice.

Although the quality of assessment and its alignment with learning goals in the Netherlands have been questioned (Borghouts and Slingerland 2014), to date there have been no empirical data to

substantiate this expectation. Therefore, the present research project aimed to determine assessment quality in secondary PE in the Netherlands in relation to each of the five quality aspects for assessment (Collier 2011; Melograno 2007; Stiggins et al. 2007). Since a teacher's or school's view on the function of assessment can be expected to impact its translation into practice, we also investigated how secondary schools view the function of assessment in PE. Finally, we examined the coherence of the intended curriculum and assessment goals and actual assessment practices in the Netherlands, as an indication for constructive alignment.

Methods

National context

Secondary education in the Netherlands is intended for children in the age group 12–18 yrs. *General secondary education* prepares for higher education and is compulsory up to the age of 16. There are two types of general education, HAVO (5 years) or VWO (6 years). *Preparatory secondary vocational education* (VMBO) is vocationally oriented and lasts 4 years (EP-Nuffic 2015). PE is an obligatory subject at all types of secondary education. The general aim of secondary school PE has been described and published by the national PE-society (KVLO) in cooperation with the schools for PE teacher education, as: '... developing skills and attitudes in youth in order to facilitate their participation in different sports- and movement activities, from a pedagogical perspective and as part of a healthy and physically active lifestyle' (Brouwer et al. 2011). More specifically, PE should introduce young people within a diversity of movement activities as well as develop various other skills such as organisation, social behaviour, and knowledge and understanding of sport and movement activities.

There is no formal national or regional curriculum for PE. Instead, PE goals are expressed in a set of broadly defined achievement goals (5–9 depending on the educational track). Examples of these are: 'Students are able to participate in at least two activities within the domains of gymnastics, athletics, dance and self-defence' and 'Students can make a well-informed choice from physical activity-opportunities in contemporary society, based on insights into their own possibilities and preferences'. As a result, schools and PE teachers have considerable freedom in their design and practice of PE.

Population sampling

Contact information of PE departments or individual PE teachers were obtained through combining an address list of all secondary schools, provided by the Dutch Ministry of Education, Culture and Science and an internet search. A total of 412 unique schools were approached through email. Additionally, to further enlarge the sample, PE teachers were invited to participate in the study through a short text on the website and in the newsletters of the national Dutch PE association (KVLO). In 12 instances, this led to more than one response from the same school. In these cases, the first response was retained while the additional responses were deleted. Participants had to be employed by their school as a PE teacher for at least two years. They were informed that data would be analysed and reported anonymously. In total, 260 schools returned a complete questionnaire (63% response rate). The spread of schools over the regions north, middle, and south of the Netherlands was 22%, 48%, and 30%, which roughly reflects the spread of the total population over these regions (26%, 56%, and 19%, respectively). Distribution over the three nationally existing educational tracks VMBO, HAVO, and VWO was reasonably representative with 42%, 36%, and 22% compared to a national distribution of 47%, 27%, and 26%, respectively (DUO 2015). Teachers had on average 16 ± 10 yrs. of experience in teaching secondary PE. Their average (\pm SD) age was 40.0 ± 10.9 yrs and 196 were males, 64 were females. There are no data on gender distribution of PE-teachers in general to compare this to, yet the general impression is that female PE teachers are indeed underrepresented in the population.

Construction of the PE assessment questionnaire

We developed a questionnaire for determining actual assessment practices and assessment quality in secondary PE in the Netherlands, the Physical Education Assessment Questionnaire (PEAQ). After the literature on assessment in PE was reviewed, themes were identified that should be contained in a comprehensive questionnaire aimed at secondary PE. These themes were presented to a group of 18 Dutch experts in the field of PE and/or assessment. The themes were revised and a first draft of the questionnaire was constructed. Aforementioned experts again reviewed the questionnaire and gave written and verbal feedback. The resulting version of the questionnaire was then filled out by 25 PE teachers as a pilot test. They were asked to write down comments on the clarity and content of the questionnaires, and were given the opportunity to verbally discuss their opinions with three researchers present. From this feedback, a third version of the questionnaire was devised and again presented to the expert group. This resulted in several smaller revisions. An online version was then created using the Survey Monkey-web tool and this was filled out by 25 PE teachers (different from the first group) from three schools. Teachers were asked to note the time it took them to complete the questionnaire and email their comments. Since several teachers reported the questionnaire to be overly long (exceeding 20 minutes to complete), it was decided by the researchers together with the expert group to remove a number of items regarding the feasibility of certain assessment approaches. Eventually, the final questionnaire contained the following sections: (1) Quality of assessment, consisting of 15 items, for example, 'All PE teachers in my school use identical criteria for assessment'. (2) Intended functions of assessment, consisting of four items, for example, 'I assess students in PE in order to inform adjustments to the content and delivery of my lessons'. (3) Intended goals of PE, consisting of nine items, for example, 'I believe an important goal of my PE lessons is to provide students with the opportunity to develop motor/technical skills'. This section was added to be able to investigate the alignment of intended goals and the assessment of PE. (4) Actual assessment practices consisting of six items with questions asked on: the various factors being judged during assessment (e.g. knowledge, motor skills); what modes of assessment were being used (e.g. written tests, observation); how criteria were used; whether assessment is a topic during PE department meetings; the form in which PE grades are communicated; and if PE grades are taken into account for determining whether a student may enter the next year.

Apart from the section on general information and assessment practices, all of the items were answered on a five-point scale from strongly disagree to strongly agree. Cronbach's alpha for all of the 'quality of assessment' items of the PEAQ was satisfactory, at .80. The reliability analysis showed that removal of individual items resulted in a lower Cronbach's alpha. This means that the questionnaire would become less reliable by the removal of any item, thus indicating all items should be retained (Field 2009). However, Cronbach's alpha for the five different predefined aspects of quality was lower: clear purpose .66; clear targets .65; sound design .54; effective communication .57; and student involvement .50. Therefore, it was decided not to report on separate aspects of assessment quality, but only on individual items and quality as a whole.

Data analysis

First, incomplete responses were removed. Then, questionnaire data were exported from Survey Monkey to Microsoft Excel for further analysis. In the event of multiple responses from the same school (which happened with six schools), the first complete response was used and the remaining removed. Data are presented as mean \pm standard deviation, where applicable. For analytical and clarification purposes, data on assessment quality are also transformed to numerical values (1 = strongly disagree, 5 = strongly agree), acknowledging the limitations of this approach (Norman 2010). All data were analysed using IBM SPSS statistics 22.

Results

Assessment quality

Table 1 shows the outcomes on the ‘quality of assessment in PE’-items. Depicted are the percentages of respondents within the various categories. When transformed into numerical values, mean assessment quality (\pm SD) was 3.6 ± 0.6 , corresponding to a response roughly between ‘neutral’ and

Table 1. PE teachers’ agreement on items regarding the five quality aspects for assessment (Stiggins et al. 2007).

	%Strongly disagree (1)	%Somewhat disagree (2)	%Neutral (3)	%Somewhat agree (4)	%Strongly agree (5)	Mean score \pm SD
Clear purpose:						
A document is available at my school which describes how PE assessment practices are aligned with its learning outcomes. ¹	5.8	12.1	9.7	33.1	39.3	3.9 \pm 1.2
A document is available at my school identifying what the learning outcomes for PE are, and how these are assessed. ²	5.8	5.4	6.6	32.9	49.2	4.1 \pm 1.1
A document is available at my school in which it is described how the results of student assessments are used. ³	12.8	13.6	17.9	21.0	34.6	3.5 \pm 1.4
Clear targets:						
The criteria for every assessment are described and fully defined in a school document. ⁴	16.6	19.7	11.2	32.8	19.7	3.2 \pm 1.4
All PE teachers in my school use identical criteria for assessment.	5.8	15.9	7.8	39.9	30.6	3.7 \pm 1.2
A document is available at my school in which it is described how the evaluation (grade or feedback) of a student is determined for any given time period.	9.7	12.0	11.6	29.8	36.8	3.7 \pm 1.3
Sound design:						
I succeed in carrying out every planned instance of assessment for every student.	5.4	8.5	9.3	41.1	35.7	3.9 \pm 1.1
The grade a student receives is <u>not</u> dependent on which teacher is doing the evaluation.	8.5	20.5	12.4	35.5	23.2	3.4 \pm 1.3
Assessments are developed in a way that relevant differences between students are captured.	3.1	7.3	13.9	40.9	34.7	4.0 \pm 1.0
Effective communication:						
Assessment criteria are shared with my students prior to assessment.	2.7	5.4	5.8	30.9	55.2	4.3 \pm 1.0
My students can access or find out all of their self-received grades/feedback.	3.1	8.1	4.6	12.4	71.8	4.4 \pm 1.1
The PE department has a document that states which grades/feedback are communicated when and to whom (students, parents, school).	25.1	19.6	20.0	16.5	18.8	2.8 \pm 1.5
Student involvement:						
My students are provided with opportunity to self - and/or peer assess in PE.	9.2	11.2	20.4	46.5	12.7	3.4 \pm 1.1
My students receive interim grades/feedback when working towards a final assessment.	10.0	18.1	21.9	34.6	15.4	3.3 \pm 1.2
My students are involved in developing the criteria by which they are assessed.	35.4	33.1	17.7	11.5	2.3	2.1 \pm 1.1

Footnotes contained within the questionnaire: ¹An example of this might be: “The development of motor skills are deemed important and therefore they are assessed in the following manner ... , an important goal is that students get acquainted with a range of sports and this is assessed in the following manner ... , etc”.

²An example of this might be: an assessment is aimed at measuring certain motor skills, attitudes, sport skills; a written test is taken to evaluate particular knowledge, etc. ³An example of this might be: interim assessments/evaluations are carried out to provide feedback to students, PE grades are taken into account when writing student report cards, etc. ⁴An example of this might be: in order to attain a certain grade a student should demonstrate a number of specific skills/behaviours/competencies, etc.

‘somewhat agree’. Sixteen percent of schools scored ≤ 3.0 (corresponding to neutral). When inspecting each of the individual items four items scored equal to or higher than 4.0 on a five-point scale, these were: ‘Assessment criteria are shared with my students prior to assessment’, ‘My students can access or find out all of their self-received grades/feedback’ (both concerning effective communication about assessment to students), ‘A document is available at school identifying what the learning outcomes for PE are, and how these are assessed’ and ‘Assessments are developed in a way that relevant differences between students are captured’. On the contrary, two items scored particularly low: ‘The PE department has a document that states which grades/feedback are communicated when and to whom (students, parents, school)’ and ‘My students are involved in developing the criteria by which they are assessed’.

Assessment functions

Table 2 shows the results of the intended functions of assessment according to the PE teachers. The selection of students (determining readiness for progress) was viewed as the least important function of assessment, while most PE teachers agreed that they intended using assessment as a means of supporting the student’s learning process.

PE goals

PE teachers rated their agreement with statements regarding the intended goals of their PE curriculum (Table 3). It can be observed that teachers rate all but one of the intended goals highly, with scores equal to or above 4.3. The exception to this was ‘I believe an important goal of my PE lessons is to increase students’ fitness levels’, which scored 3.6 (± 1.0). ‘Teaching adequate and respectful interaction in physical activity’ and ‘allowing students to experience fun in physical activity’ displayed the highest score.

Assessment practices

A wide variety of factors are taken into account when grading (Table 4), of which ‘effort’ is the aspect most often reported, closely followed by game tactics, active participation, social behaviour, and motor/technical skills. ‘Knowledge and reflection’ scored lowest, being assessed in PE by roughly one-third of the schools.

Table 5 shows the modes of assessment that are applied to judge these various aspects. Observation is by far the mode of assessment most widely applied. Table 5 also indicates that only a minority of PE teachers (very) often grades students without predetermined assessment criteria, and that criteria are not widely adapted to suit different (groups of) students; usually criteria are identical

Table 2. Intended functions of assessment according to PE teachers.

Assessment function	%Strongly disagree (1)	%Somewhat disagree (2)	%Neutral (3)	%Somewhat agree (4)	%Strongly agree (5)	Mean score \pm SD
I assess students in PE in order to determine their readiness to progress (to the next year, ...)	32.3	24.6	16.2	18.5	7.7	2.4 \pm 1.3
I assess students in PE in order to inform adjustments to the content and delivery of my lessons.	8.1	20.8	19.2	39.6	11.9	3.3 \pm 1.2
I assess students in PE as a means of justifying the subject within my school.	14.6	14.6	11.2	39.6	20.0	3.4 \pm 1.3
I assess students in PE as a means of supporting their learning process.	2.3	6.2	7.3	35.4	48.8	4.2 \pm 1.0

Table 3. Intended goals of PE.

I believe an important goal of my PE lessons is to ...	%Strongly disagree (1)	%Somewhat disagree (2)	%Neutral (3)	%Somewhat agree (4)	%Strongly agree (5)	Mean score ± SD
Provide students with the opportunity to develop motor/technical skills	3.1	1.2	1.9	27.3	66.5	4.5 ± 0.9
Prepare students with skills to participate in physical activity in a variety of roles	2.3	1.9	7.7	35.8	52.3	4.3 ± 0.9
Enable students to interact adequately and respectfully with others in physical activity	3.5	0	0	7.3	89.2	4.8 ± 0.8
Prepare students with knowledge and understanding with regards to physical activity	2.7	0	3.1	38.8	55.4	4.4 ± 0.8
Allow students to experience fun through physical activity	2.7	0	0.8	5.0	91.5	4.8 ± 0.7
Increase students' fitness levels	4.2	8.1	25.0	46.5	16.2	3.6 ± 1.0
Encourage students to find personally appropriate physical activities	2.3	2.7	8.1	40.4	46.5	4.3 ± 0.9
Get students physically active during school hours	3.1	0.8	5.4	30.4	60.4	4.4 ± 0.9
Support self-development through physical activity and sports	2.3	1.5	2.3	28.8	65.0	4.5 ± 0.8

Table 4. Assessment practices; factors judged in PE assessment (percentages; multiple answers allowed).

Factor	Factor is being judged?
Effort	96.2
Game tactics	90.4
Active participation	90.0
Social behaviour	89.7
Motor/technical skills	89.7
Fitness	81.2
Supporting skills	74.7
Safety assistance	71.6
Attendance	53.6
Appropriate attire	47.5
Knowledge and reflection	34.1
Other	2.7

Table 5. Assessment practices; modes of assessment in PE, and the use of criteria (percentages).

	Never	Seldom	Sometimes	Often	Very often
Modes of assessment					
Observation (in the gymnasium, on the pitch, etc.)	3.1	1.9	16.7	32.7	45.4
Outcome measurement (time, height, number of scores, etc.)	0.8	6.2	39.2	40.4	13.5
Written test (e.g. open or closed questions exam)	49.2	31.9	17.7	1.2	0
Written assignment (e.g. reflection, essay)	27.7	36.9	31.5	3.1	0.4
Exercise book with assignments filled out by the student	84.6	9.6	4.2	1.2	0.4
Student portfolio	80.4	9.2	6.5	3.1	0.8
Other	48.5	9.6	15.0	3.8	1.2
Use of criteria					
Grading without predetermined assessment criteria	25.8	36.9	18.8	16.2	1.9
Grading based on predetermined assessment criteria identical for all students	3.5	4.2	23.1	41.9	26.9
Grading taking into account the student's individual level and progress	3.8	8.8	28.8	39.6	18.8
Other	35.4	11.5	20.0	2.3	0.4

Table 6. Assessment practices; miscellaneous (percentages).

	Yes	No	Don't know	
Is assessment a regular topic at PE department meetings	62.7	36.2	1.2	
	Numerical	Fail/Pass/Excellent	Written appraisal	Other
How is the grade for PE communicated on the school report? (multiple answers allowed)	85.0	30.0	6.5	2.7
	Yes	No	Sometimes	Don't know
Are PE grades taken into account for determining whether a student may enter the next year?	77.7	15.8	6.2	0.4

for all students. Most teachers take the individual progress of students into account at least sometimes.

Table 6 shows that a little over one-third of the respondents indicated that assessment is not a regular topic at PE department meetings. Most schools use numerical grades for communicating PE grades. Finally, at a large majority of the schools (77.7%), PE grades are taken into account for determining whether a student may enter the next year.

Discussion

The present study aimed to determine assessment quality and actual assessment practices (i.e. what is evaluated through which mode) in secondary PE in the Netherlands. We also examined how PE teachers in secondary schools view the function of assessment in PE, and whether the intended curriculum and assessment goals are aligned with actual assessment practices. Since it has been suggested that assessment quality in PE is suboptimal (Hay and Penney 2009; Thorburn 2007; Veal 1988), we devised a questionnaire (PEAQ) that enabled us to examine assessment quality as defined previously (Collier 2011; Melograno 2007; Stiggins et al. 2007).

Assessment quality

In general, PE assessment quality in the Netherlands seems to be moderate. In our sample, assessment quality as a whole scored between 'neutral' and 'somewhat agree'; converted to a numerical five-point scale this corresponds to 3.6. The questionnaire items were constructed in such a manner that a school with an optimal approach to assessment would score 'totally agree' on all items. This indicates that on average, there is room for improvement. Especially schools scoring below 'neutral' (16%) should probably re-evaluate their assessment practices. The relevance of this is underlined by our finding that in the majority of schools in the Netherlands, PE grades are taken into account when determining whether a student may enter the next year. PE assessment can therefore be considered 'high-stakes' testing. The fact that, for example, only 31% of PE teachers strongly agree that all of their colleagues use identical assessment criteria, is therefore cause for concern.

Collier (2011) has stated that it is important to provide PE teachers with sound strategies to improve assessment, that can be implemented relatively easily. He proposed that an understanding of the five aspects of quality assessment (Stiggins et al. 2007) could increase what has been referred to by Melograno (2007) as *assessment literacy*. According to Dinan-Thompson and Penney (2015), building on Hay and Penney (2013), assessment literacy encompasses four components: assessment comprehension, application, interpretation, and critical engagement with assessment. Although the PEAQ used in this study was developed for research purposes, we propose that it might be used by PE departments to evaluate the strengths and weaknesses of their system of assessment, and thereby assist in critical engagement with assessment.

Assessment functions

Examination regulations in the Netherlands state that students must pass PE in order to obtain a diploma. It is therefore interesting to see that teachers do not consider selection an important assessment function, even though at 77.7% of the schools, PE grades are taken into account for determining whether a student may enter the next year. With regard to other assessment functions, both assessment as a means for informing lesson adjustments, and as a means of justifying the subject within school, are only moderately valued. Instead, PE teachers consider ‘supporting the student’s learning process’ as the most important function of assessment.

Using assessment to guide the learning process is in accordance with the tenets of assessment for learning (AFL) (Black and Wiliam 1998; Wiliam 2011). Klenowski (2009, 264) states that AfL is part of everyday practice by students, teachers, and peers that seeks, reflects upon, and responds to information from dialogue, demonstration, and observation in ways that enhance ongoing learning. It can be derived from this that student involvement is key to the concept of AfL. Our findings suggest that although (perhaps implicitly), PE teachers in the Netherlands strive to apply AfL, assessment practices are often not in line with this. Only 15.4% of PE teachers strongly agree that students receive interim grades or (formative) feedback, self- and peer-assessment are not very widespread, and students do not often participate in the development of the criteria by which they are assessed.

It has been suggested that AfL and assessment of learning can very well exist together, as long as the assessment programme consists of an arrangement of assessment methods that are planned to optimise its suitability for purpose (Van der Vleuten et al. 2012). AfL has been widely advocated within PE (Collier 2011; Hay 2006; Hay and Penney 2009; Ní Chróinín and Cosgrave 2012), but it has also been noted that research into the impact of AfL is rather limited in this field (Hay 2006; MacPhail and Halbert 2010). AfL is believed to support student achievement and motivation for education in general (Black and Wiliam 2009; Cauley and McMillan 2010). Since it has been shown that a favourable motivation for PE is associated with various positive behavioural outcomes both in- and outside of the PE class, such as higher student engagement and physical activity levels (Haerens et al. 2010; Ntoumanis and Standage 2009), we argue that it would be of interest to establish the effects of AfL in PE on student motivation in future research.

Alignment of PE goals and assessment

Learning objectives for PE should be meaningful in relation to the overall goals of PE, and the learning activities and assessment activities should be aligned with the learning objectives. This ‘constructive alignment’ (Biggs 1996) is important for achieving a coherent curriculum, and quality PE (Lund 1992; Penney et al. 2009). In addition, constructive alignment is considered a requirement for AfL (Georgakis and Wilson 2012). By comparing the intended goals of the PE curriculum with the aspects judged in assessment in our sample it was possible to gain insight into the degree of constructive alignment in Dutch PE. Although increasing students’ fitness levels is considered the least important goal of PE lessons by the PE teachers, 81% of them report fitness is one of the aspects being judged. Conversely, while according to a national PE consensus report, gaining knowledge about physical activity and sports is one of the goals of PE (Brouwer et al. 2011), and 94% of the PE teachers in our sample agrees with this, only 34% actually assesses knowledge.

Furthermore, it is notable that in line with findings from previous research (James, Griffin, and Dodds 2008; Matanin and Tannehill 1994; Redelius and Hay 2012), several aspects that are judged are not so much linked to learning objectives, but rather are prerequisites for learning. Examples of this are effort, active participation, or attendance. Our study indicates that in the Netherlands, as in other countries, there may be a lack of alignment between intended learning outcomes and what is actually being valued and assessed. Acknowledging the importance of achieving a coherent, aligned curriculum for quality PE, we recommend PE teachers to give closer attention to the alignment of learning goals, learning activities and assessment practices.

Assessment practices

Secondary education is divided into separate cognitive levels in the Netherlands. Within these cognitive levels, PE and Arts are an exception to the rule of applying norm-referenced criteria to grading. Examination regulations by the Dutch government state that for PE and Arts, the individual student's 'capabilities' must be taken into consideration. Our findings reflect this situation, since most teachers in our study indicated that they (very) often take into account students' individual level and progress when assessing practical skills (such as game tactics and technical skills).

Observation seems by far the instrument of assessment most widely applied within PE, followed by 'objective' outcome measurements. This is not surprising, since team games form the most prevalent lesson subject in the Netherlands (Slingerland, Oomen, and Borghouts 2011) and observation lends itself for judging both technical and tactical criteria during game play. For assessing gymnastics and dance, observation is equally suitable, while athletics is most likely to be assessed through quantitative measurements (time, distance, etc.). A previous study in the US showed a comparably high prevalence of teacher observations as an instrument of assessment (Mintah 2003). However, written forms of assessment seem to be reported to a higher extend in several US-based studies compared to our data from the Netherlands (Imwold, Rider, and Johnson 1982; Kneer 1986; Matanin and Tannehill 1994; Mintah 2003; Veal 1988). This again confirms our finding that a minority of our population actually assesses knowledge, even though knowledge transfer is considered a goal for PE.

According to Hay and Penney (2009), assessment efficacy in PE is enhanced through a focus on authentic tasks. In their definition, authentic assessment

[...] pursues tasks and foci that are meaningful to students and that have value and meaning beyond the instructional context. (394)

However, a review of PE practices in Scotland, England and Australia, identified the creation of authentic assessment arrangements as a significant professional challenge, at least in high-stakes school examination (Thorburn 2007). Practical demonstration only represented a modest part of the total of assessments in comparison to written coursework, investigations, and examinations. The present study shows that in secondary education in the Netherlands, practical demonstration is by far the most prevalent form of assessment, with 78% of schools reporting to use observation as an instrument for assessment often to very often. In contrast, written tests and assignments are seldom used. Although this can be considered a positive outcome from the perspective of authenticity, it also suggests that although the majority of respondents agree that preparing students with knowledge and understanding is an important goal of PE, this is often not explicitly assessed.

Most teachers in our study reported frequent use of assessment criteria, but it is of note that these criteria are not always identical to those of other PE teachers within the same school. At the same time, assessment is not a regular topic at meetings in more than a third of the PE departments. This, together with the apparent suboptimal assessment quality, leads us to suggest PE departments, school boards, and the education inspectorate should scrutinise existing assessment practices, and work together to optimise the systemic conditions (Redelius and Hay 2012) needed for quality PE assessment.

Study strengths and limitations

The schools that responded to our questionnaire were geographically spread throughout the country, and distribution over the three nationally existing educational tracks was reasonably representative. Nevertheless, whether our sample is representative for the whole of the country cannot be said with any certainty.

Reliability for the questionnaire as a whole, as determined by Cronbach's alpha, was satisfactory but this was not the case for the separate quality aspects (clear purpose, clear targets, sound design, effective communication, and student involvement). A possible explanation for this could be that

there is some conceptual overlap between aspects, for example, between *effective communication* and *student involvement*. In addition, each aspect was represented by just three items and it is known that Cronbach's alpha tends to deteriorate with smaller numbers of items (Field 2009). Expanding the questionnaire by adding appropriate items to each aspect can be expected to increase the reliability, but might also compromise its practical applicability; the total number of items of the PEAQ was already reduced after pilot testing due to PE teachers considering it too long.

Our study was purposefully designed to give a large-scale, quantitative indication of assessment quality and practices in PE in the Netherlands, yet we acknowledge that for a more in-depth, contextualised analysis, qualitative data from PE teachers, school leaders and students would be of great value.

Conclusions and implications

This study provided insight into assessment in PE in the Netherlands. Assessment is of moderate quality according to the five quality aspects for assessment (Collier 2011; Melograno 2007; Stiggins et al. 2007). It is therefore of note that PE assessment can be considered 'high stakes' since most schools take PE grades into account when determining whether a student may enter the next year. We also investigated how secondary schools view the purpose of assessment in PE. Our findings suggest that PE teachers consider assessment *for learning* the most important function of assessment, but practices are not generally in line with this. Furthermore, there may be a lack of alignment between intended learning outcomes and what is actually being valued and assessed.

We believe our results call for a concerted effort from PE departments, school boards and the education inspectorate to scrutinise existing assessment practices, and work together to optimise PE assessment. We suggest that this should include attention for the concepts of *constructive alignment* and *assessment for learning*, and the various aspects of assessment quality.

It would be of value to identify 'good practices' from the present research, and conduct more intensive depth-based qualitative research in these schools in order to better inform policy and practice revisions. Future research might also address the subject of constructive alignment in PE in the Netherlands in a more explicit manner. Mixed methods designs could provide valuable insights into the alignment of learning goals, learning activities, and assessments within specific courses or lesson series, preferably contextualised by interviewing the involved PE teachers about their considerations regarding curriculum design, their beliefs and attitudes toward assessment, etc. Furthermore, it would be interesting to compare our findings to similar data from other countries, since large-scale data on PE assessment are scarce. Finally, we believe professional development could also greatly benefit from insights gained from PE assessment intervention studies, for example, into the effects of assessment for learning.

Note

1. For more information on the Dutch school system and its inspection, see <http://www.onderwijsinspectie.nl/english>).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek under [grant number 2012-14-15P].

References

- Annerstedt, C., and S. Larsson. 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.
- Berliner, D. 2011. "Rational Responses to High Stakes Testing: The Case of Curriculum Narrowing and the Harm that Follows." *Cambridge Journal of Education* 41 (3): 287–302.
- Biggs, J. 1996. "Enhancing Teaching Through Constructive Alignment." *Higher Education* 32 (3): 347–364.
- Black, K. P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education* 5 (1): 7–74.
- Black, P., and D. Wiliam. 2009. "Developing the Theory of Formative Assessment." *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)* 21 (1): 5–31.
- Borghouts, L., and M. Slingerland. 2014. "Kwaliteit van Beoordeling binnen LO." *Lichamelijke Opvoeding* 4: 12–13.
- Broadfoot, P., R. Daugherty, J. Gardner, W. Harlen, M. James, and G. Stobart. 2002. *Assessment for Learning: 10 Principles*. Cambridge: University of Cambridge School of Education.
- Brouwer, B., A. Aldershof, H. Bax, M. Van Berkel, G. J. Van Dokkum, M. J. Mulder, and J. Nienhuis. 2011. *Human Movement en Sports in 2028*. Enschede: SLO.
- Brouwer, B., M. Van Berkel, G. Van Mossel, and E. Swinkels. 2015. *Bewegingsonderwijs en sport; vakspecifieke trendanalyse 2015*. Enschede: SLO.
- Cassady, H., G. Clarke, and A.-M. Latham. 2004. "Experiencing Evaluation: A Case Study of Girls' Dance." *Physical Education and Sport Pedagogy* 9 (1): 23–36.
- Cauley, K. M., and J. H. McMillan. 2010. "Formative Assessment Techniques to Support Student Motivation and Achievement." *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* 83 (1): 1–6.
- Chan, K., P. J. Hay, and R. Tinning. 2011. "Understanding the Pedagogic Discourse of Assessment in Physical Education." *Asia-Pacific Journal of Health, Sport and Physical Education* 2 (1): 3–18.
- Collier, D. 2011. "Increasing the Value of Physical Education: The Role of Assessment." *Journal of Physical Education, Recreation & Dance* 82 (7): 38–41.
- Desrosiers, P., Y. GenetVolet, and P. Godbout. 1997. "Teachers' Assessment Practices Viewed Through the Instruments Used in Physical Education Classes." *Journal of Teaching in Physical Education* 16 (2): 211–228.
- Dinan-Thompson, M., and D. Penney. 2015. "Assessment Literacy in Primary Physical Education." *European Physical Education Review* 21 (4): 485–503.
- DUO. 2015. *Inschrijvingsgegevens*. Groningen: Dienst Uitvoering Onderwijs. <https://duo.nl/zakelijk/VO/Inschrijvingsgegevens/inschrijvingsgegevens.asp>.
- EP-Nuffic. 2015. *Education System; The Netherlands*. Den Haag: Nuffic.
- Erdmann, R., D. Chatzopoulos, and H. Tsormbatzoudis. 2006. "Pupils Grading: Do Teachers Grade According to the Way They Report?" *International Journal of Physical Education* 43 (1): 4–10.
- Field, A. 2009. *Discovering Statistics Using SPSS*. Thousand Oaks, CA: Sage.
- Georgakis, S., and R. Wilson. 2012. "Australian Physical Education and School Sport: An Exploration into Contemporary Assessment." *Asian Journal of Exercise and Sports Science* 9 (1): 37–52.
- Haerens, L., D. Kirk, G. Cardon, I. De Bourdeaudhuij, and M. Vansteenkiste. 2010. "Motivational Profiles for Secondary School Physical Education and its Relationship to the Adoption of a Physically Active Lifestyle among University Students." *European Physical Education Review* 16 (2): 117–139.
- Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112.
- Hay, P. J. 2006. "Assessment for Learning in Physical Education." In *The Handbook of Physical Education*, edited by D. Kirk, D. Macdonald, and M. O'Sullivan, 312–325. London: Sage.
- Hay, P. J., and D. Macdonald. 2008. "(Mis)appropriations of Criteria and Standards-referenced Assessment in a Performance-based Subject." *Assessment in Education: Principles, Policy & Practice* 15 (2): 153–168.
- Hay, P. J., and D. Penney. 2009. "Proposing Conditions for Assessment Efficacy in Physical Education." *European Physical Education Review* 15 (3): 389–405.
- Hay, P. J., and D. Penney. 2013. *Assessment in Physical Education: A Sociocultural Perspective*. London: Routledge.
- Hursh, D. 2005. "Neo-liberalism, Markets and Accountability: Transforming Education and Undermining Democracy in the United States and England." *Policy Futures in Education* 3 (1): 3–15.
- Imwold, C. H., R. A. Rider, and D. J. Johnson. 1982. "The Use of Evaluation in Public School Physical Education Programs." *Journal of Teaching in Physical Education* 2 (1): 13–18.
- James, A., L. L. Griffin, and P. Dodds. 2008. "The Relationship between Instructional Alignment and the Ecology of Physical Education." *Journal of Teaching in Physical Education* 27: 308–326.
- Klenowski, V. 2009. "Assessment for Learning Revisited: An Asia-Pacific Perspective." *Assessment in Education: Principles, Policy & Practice* 16 (3): 263–268.
- Kneer, M. E. 1986. "Description of Physical Education Instruction Theory/Practice Gap in Selected Secondary Schools." *Journal of Teaching in Physical Education* 5: 91–106.
- López-Pastor, V. M., D. Kirk, E. Lorente-Catalán, A. MacPhail, and D. Macdonald. 2013. "Alternative Assessment in Physical Education: A Review of International Literature." *Sport, Education and Society* 18 (1): 57–76.

- Lorente-Catalán, E., and D. Kirk. 2016. "Student Teachers' Understanding and Application of Assessment for Learning During a Physical Education Teacher Education Course." *European Physical Education Review* 22 (1): 65–81.
- Lund, J. 1992. "Assessment and Accountability in Secondary Physical Education." *Quest* 44 (3): 352–360.
- MacPhail, A., and J. Halbert. 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 (1): 23–39.
- Matanin, M., and D. Tannehill. 1994. "Assessment and Grading in Physical Education." *Journal of Teaching in Physical Education* 13 (4): 395–405.
- Melograno, V. J. 2007. "Grading and Report Cards for Standards-based Physical Education." *Journal of Physical Education, Recreation & Dance* 78 (6): 1–58.
- Mintah, J. K. 2003. "Authentic Assessment in Physical Education: Prevalence of Use and Perceived Impact on Students' Self-concept, Motivation, and Skill Achievement." *Measurement in Physical Education and Exercise Science* 7 (3): 161–174.
- Newton, P. E. 2007. "Clarifying the Purposes of Educational Assessment." *Assessment in Education: Principles, Policy & Practice* 14 (2): 149–170.
- Ní Chróinín, D., and C. Cosgrave. 2012. "Implementing Formative Assessment in Primary Physical Education: Teacher Perspectives and Experiences." *Physical Education and Sport Pedagogy* 18 (2): 219–233.
- Nichols, S. L., and D. C. Berliner. 2007. *Collateral Damage: How High-stakes Testing Corrupts America's Schools*. Cambridge, MA: Harvard Education Press.
- Norman, G. 2010. "Likert Scales, Levels of Measurement and the 'Laws' of Statistics." *Advances in Health Sciences Education* 15 (5): 625–632.
- Ntoumanis, N., and M. Standage. 2009. "Motivation in Physical Education Classes: A Self-determination Theory Perspective." *Theory and Research in Education* 7 (2): 194–202.
- Pat El, R. J., H. Tillema, M. Segers, and P. Vedder. 2013. "Validation of Assessment for Learning Questionnaires for Teachers and Students." *British Journal of Educational Psychology* 83 (1): 98–113.
- Penney, D., R. Brooker, P. J. Hay, and L. Gillespie. 2009. "Curriculum, Pedagogy and Assessment: Three Message Systems of Schooling and Dimensions of Quality Physical Education." *Sport, Education and Society* 14 (4): 421–442.
- Redelius, K., and P. J. Hay. 2012. "Student Views on Criterion-referenced Assessment and Grading in Swedish Physical Education." *Physical Education and Sport Pedagogy* 17 (2): 211–225.
- Roberts-Holmes, G. and Bradbury, A. 2016. "Governance, Accountability and the Datafication of Early Years Education in England." *British Educational Research Journal*. doi:10.1002/berj.3221.
- Slingerland, M., J. Oomen, and L. Borghouts. 2011. "Physical Activity Levels During Dutch Primary and Secondary School Physical Education." *European Journal of Sport Science* 11 (4): 249–257.
- Stiggins, R. J. 1999. "Assessment, Student Confidence, and School Success." *Phi Delta Kappan* 81 (3): 191–198.
- Stiggins, R. J., J. A. Arter, J. Chappuis, and S. Chappuis. 2007. *Classroom Assessment for Student Learning*. New York, NY: Pearson.
- Thorburn, M. 2007. "Achieving Conceptual and Curriculum Coherence in High-stakes School Examinations in Physical Education." *Physical Education and Sport Pedagogy* 12 (2): 163–184.
- Van der Vleuten, C. P. M., L. W. T. Schuwirth, E. W. Driessen, J. Dijkstra, D. Tigelaar, L. K. J. Baartman, and J. van Tartwijk. 2012. "A Model for Programmatic Assessment Fit for Purpose." *Medical Teacher* 34 (3): 205–214.
- Veal, M. L. 1988. "Pupil Assessment Perceptions and Practices of Secondary Teachers." *Journal of Teaching in Physical Education* 7 (4): 327–342.
- Wiliam, D. 2011. "What is Assessment for Learning?" *Studies in Educational Evaluation* 37 (1): 3–14.
- Zhu, X. 2015. "Student Perspectives of Grading in Physical Education." *European Physical Education Review* 21 (4): 409–420.

Appendix: Items of the PEAQ-questionnaire

Quality of assessment in PE (5-point scale from strongly disagree to strongly agree)

Clear purpose:

- A document is available at my school which describes how PE assessment practices are aligned with its learning outcomes.¹
- A document is available at my school identifying what the learning outcomes for PE are, and how these are assessed.²
- A document is available at my school in which it is described how the results of student assessments are used.³

Clear targets:

- The criteria for every assessment are described and fully defined in a school document.⁴
- All PE teachers in my school use identical criteria for assessment.
- A document is available at my school which describes how the evaluation (grade or feedback) of a student is determined for any given time period.

Sound design:

- I succeed in carrying out every planned instance of assessment for every student.
- The grade a student receives is **not** dependent on which teacher is doing the evaluation.
- Assessments are developed in a way that relevant differences between students are captured.

Effective communication:

- Assessment criteria are shared with my students prior to assessment.
- My students can access or find out all of their self-received grades/feedback.
- The PE department has a document that states which grades/feedback are communicated when and to whom (students, parents, school).

Student involvement:

- My students are provided with opportunity to self - and/or peer assess in PE.
- My students receive interim grades/feedback when working towards a final assessment.
- My students are involved in developing the criteria by which they are assessed.

Footnotes contained within the digital questionnaire: ¹An example of this might be: "The development of motor skills are deemed important and therefore they are assessed in the following manner ... , an important goal is that students get acquainted with a range of sports and this is assessed in the following manner ... , etc".

²An example of this might be: an assessment is aimed at measuring certain motor skills, attitudes, sport skills; a written test is taken to evaluate particular knowledge, etc. ³An example of this might be: interim assessments/evaluations are carried out to provide feedback to students, PE grades are taken into account when writing student report cards, etc. ⁴An example of this might be: in order to attain a certain grade a student should demonstrate a number of specific skills/behaviours/competencies, etc.

Intended functions of assessment (5-point scale from strongly disagree to strongly agree)

- I assess students in PE in order to determine their readiness to progress (to the next year, to the level of 'Leaving Certificate', etc.).
 - I assess students in PE in order to inform adjustments to the content and delivery of my lessons.
 - I assess students in PE as a means of justifying the subject within my school.
 - I assess students in PE as a means of supporting their learning process.
-

Intended PE goals (5-point scale from strongly disagree to strongly agree) I believe an important goal of my PE lessons is to ...

- Provide students with the opportunity to develop motor/technical skills
 - Prepare students with skills to participate in physical activity in a variety of roles
 - Enable students to interact adequately and respectfully with others in physical activity
 - Prepare students with knowledge and understanding with regards to physical activity
 - Allow students to experience fun through physical activity
 - Increase students' fitness levels
 - Encourage students to find personally appropriate physical activities
 - Get students physically active during school hours
 - Support self-development through physical activity and sports
-

Assessment practices

Factors that are judged when assessing PE (yes/no, more than one answer could be given)

- Effort
- Game tactics
- Active participation
- Social behaviour
- Motor/technical skills
- Fitness
- Supporting skills
- Safety assistance
- Attendance
- Appropriate attire
- Knowledge and reflection
- Other

Modes of assessment (5-point scale from never to very often)

- Observation (in the gymnasium, on the pitch, etc.)
- Outcome measurement (time, height, number of scores, etc.)
- Written test (e.g. open or closed questions exam)
- Written assignment (e.g. reflection, essay, etc.)
- Exercise book with assignments filled out by the student
- Student portfolio
- Other

Use of criteria (5-point scale from never to very often)

- Grading without predetermined assessment criteria
- Grading based on predetermined assessment criteria identical for all students
- Grading taking into account the student's individual level and progress

Miscellaneous

- Is assessment a regular topic at PE department meetings (Yes/No/Don't know)
- Are PE grades taken into account for determining whether a student may enter the next year? (yes/no/sometimes/don't know)

How is the grade for PE communicated on the school report? (Numerical grade/Fail;Pass;Excellent (or similar)/Written appraisal/Other)
