<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *European Physical Education Review*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Svennberg, L., Meckbach, J., Redelius, K. (2014)

Exploring PE teachers' 'gut feeling': An attempt to verbalise and discuss teachers' internalised grading criteria.

European Physical Education Review, 20(2): 199-214

<http://dx.doi.org/10.1177/1356336X13517437>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:gih:diva-3295>

Exploring PE teachers' 'gut feeling': An attempt to verbalise and discuss teachers' internalised grading criteria

Lena Svennberg

The Swedish School of Sport and Health Sciences, Sweden; University of Gävle, Sweden

Jane Meckbach

The Swedish School of Sport and Health Sciences, Sweden

Karin Redelius

The Swedish School of Sport and Health Sciences, Sweden

Abstract

Research shows that teachers' grading is influenced by non-achievement factors in addition to official criteria, such as knowledge and skills. Some grading criteria are internalised by the teacher, who is sometimes unable to verbalise the criteria used and refers to what is called a 'gut feeling'. Therefore, transparency, validity and reliability are problematic. The aim of this study was to explore which criteria physical education teachers consider important when grading. Such an exploration makes it possible to discuss how the verbalised criteria and the value they are given by the teachers can be understood. Four Year 9 teachers at different Swedish compulsory schools were interviewed using Kelly's Repertory Grid technique. Among the verbalised criteria, four themes were identified: motivation, knowledge and skills, self-confidence and interaction with others. The teachers sometimes had difficulties predicting which criteria had relevance to the grades given, and the criteria considered important by the teachers were not always reflected in the grade. The verbalised criteria revealed teachers using grades to encourage such student behaviours that helped them to handle the classroom

situation and to facilitate students learning. To become cognisant of and develop their grading, methods to verbalise their individual grading criteria were needed, and Kelly's Repertory Grid technique is one possible option. The results provide discussion points about reasons for the way teachers are grading.

Keywords

Assessment, grading practice, physical education, Repertory Grid

Introduction

Grading non-achievement factors, such as behaviour and attitudes, has been identified in different subjects and school systems (Brookhart, 1994; Cizek, Fitzgerald and Rachor, 1996; Cross and Frary, 1999; McMillan, 2003). Early research shows that teachers tend to use a 'hodgepodge grade of attitude, effort, and achievement' (Brookhart, 1991: 36) and recent research indicates that little has changed. Grades still reflect teachers' contexture of knowledge and skills with other aspects like attendance, motivation and attitude (Cox, 2011; Klapp Lekholm and Cliffordson, 2009; Young, 2011). Several studies indicate that assessment and grading in physical education (PE) are no exception (Chan, Hay and Tinning, 2011; Redelius, Fagrell and Larsson, 2009; Tholin, 2006) and non-achievement factors sometimes make up half of the grade or more (James, Griffin and France, 2005; Young, 2011).

When the stated criteria are inconsistent with how the grading is done, it affects the learning-teaching process since the assessment is sending out a different message regarding what is important to learn (Chan, Hay and Tinning, 2011; Hay and Penney, 2012; James, Griffin and

France, 2005; Redelius and Hay, 2009). What causes this ‘hodgepodge grading’? Stiggins (1986) suggests three possible explanations for the discrepancy between the recommended grading practice and the actual practice: differences of opinion of best practices; classroom realities make recommended practice inappropriate; and teachers lack sufficient knowledge and skills to meet the recommended standards.

Another phenomenon research deals with is that teachers often seem sure which grade a certain student deserves, but when asked to explain why, they find it hard to express the grading criteria in words and refer to a ‘gut feeling’ or internalised criteria (Annerstedt and Larsson, 2010; Hay and MacDonald, 2008). The internalised criteria are considered to be the teacher’s own criteria that have an impact on the grades regardless if they are consistent with the official grading criteria or not (cp. Penney et al., 2009). The facts that teachers assess and grade non-criteria-related factors and that some teachers are unable to express which criteria they use are problematic. In a high-stakes grading system, the grades are important for the students’ access to higher education and have to be fair and comparable. Students have the right to know what is being assessed and graded. When the criteria used by teachers are internalised, they are not transparent. Therefore, fairness, comparability and validity are questionable (Annerstedt and Larsson, 2010; Hay and McDonald, 2008).

The aim of this study is to explore what four PE teachers consider important when grading. This is done by studying what they consider important when talking about grading and by analysing the relevance the expressed criteria have to the grades they have given their students. By using the Repertory Grid technique (Fransella, Bell and Bannister, 2004) our attempt is to examine seldom-verbalised criteria. Only when we know more about the

spectrum of existing internalised grading criteria is it possible to analyse and increase the understanding of why teachers tend to use other criteria than the official ones when grading their students. An orientation of the Swedish grading system and a discussion about the impact of the curriculum are displayed in the next section.

The impact of the curriculum

The Swedish school system is goal-oriented, and in the national curriculum (Skolverket, 2000), goals to attain are stated for each subject. The curriculum is designed to make clear what all pupils should learn whilst providing great scope for teachers and pupils to choose their own materials and working methods. In other words, ways of working, organisation and methods are not determined, but the learning outcomes are. A criterion-referenced grading system is in use, and grades should be awarded on the basis of how well the student meets the stated knowledge criteria or learning outcomes. There is a three-point grading system: Pass (P), Pass with Distinction (PD) and Pass with Special Distinction (PSD)¹. The grading criteria include no other aspects than what knowledge students should attain. The PE curriculum stresses the importance of knowledge about health, how the body works and a healthy lifestyle. Students should also, for example, be able to participate in games, dance, sports and other activities, and be able to perform movements appropriate to a task. However, several studies carried out on Swedish PE indicate that how the student behaves is just as important as knowledge and skills (Annerstedt and Larsson, 2010; Redelius, Fagrell and Larsson, 2009; Tholin, 2006). Nevertheless, research also reveals that criterion-referenced grades are as good as, or even better, predictors of success in higher education than the former norm-referenced grades (Cliffordson, 2008). Klapp Lekholm and Cliffordson (2008) interpret this as a possible result of the teachers' grading of curriculum-irrelevant qualities that facilitate learning.

Bernstein (2003: 85) points out the importance of the curriculum: 'Curriculum defines what counts as valid knowledge, pedagogy defines what counts as valid transmission of knowledge, and evaluation defines what counts as a valid realisation of the knowledge on the part of the taught'. Linde (2012) discusses Bernstein's thesis that curriculum defines what counts as valid knowledge and raises the question as to whether it is the written official curriculum that counts, or the mediated curriculum that results from the teachers' transformations. He then points to the fact that the content and subject matters taught by teachers or learnt by students are not always the content expressed in the written official curriculum. Linde (2012) is interested in how this process can be explained. He uses the concept of arena, or scene, to describe the levels and processes that lead to the realisation of educational practice. According to Linde, steering documents are drawn up in the formulation arena, and local schools and teachers reformulate them in the transformation arena into a content that is then constituted in the realisation arena of educational practice. These arenas can also be applied to grading. In the formulation arena, the official grading criteria are devised. Thereafter, they are interpreted and reformulated in the transformation arena by the school and the teachers in order to make them more concrete and applicable in the classroom context. Finally, in the realisation arena, they are actualised in the grading criteria used to assess the students (see Figure 1). In this article, we explore the teachers' transformation and realisation of the grading criteria and discuss the influences involved. In the next section, different influences on grading identified in earlier research are presented.

Figure 1.

Influences on the transformation and realisation arenas

There are several factors involved in the transformation of the curriculum, such as teachers' repertoire, students, school management, teachers' desire to make it work and goals of representatives from the private and public sector (Linde, 2012). How curriculum content, or, in this case, grading criteria, are interpreted by teachers in the transformation arena and used in the realisation arena depends on a variety of factors. Some of them are external, such as the students' family background (Evans, 2004; Klapp Lekholm and Cliffordson, 2008), gender (Klapp Lekholm and Cliffordson, 2008; Redelius, Fagrell and Larsson, 2009), political decisions (Klapp Lekholm and Cliffordson, 2008; Korp, 2006), parents' and principals' influence (Mickwitz, 2011), the classroom situation (Brookhart, 1994; McMillan, 2003) and school culture and school facilities (Linde 2012). Other factors are internal, such as the teachers' knowledge of the subject and pedagogy (Selghed, 2004; Stiggins, 1986), their expectations and evaluation of students' achievements (Redelius, Fagrell and Larsson, 2009; Stiggins, 1986) and their beliefs and values (McMillan and Nash, 2000; McMillan, 2003; Penney et al., 2009). This map of influencing factors is not complete but gives a picture of the complexity and variety of factors impacting teachers' grading practice (see Figure 1). In the following section, an introduction to the theory behind the Repertory Grid technique, Personal Construct Theory (PCT), and the nature of personal constructs are outlined. Thereafter, the methodological application, the Repertory Grid technique used in the interviews, and how the interviews are carried out in this study are described.

Research design

George Kelly, the father of PCT, believes that our behaviour can be understood in terms of personally constructed explanations of how the world works: 'Man looks at his world through

transparent patterns or templets which he creates and then attempts to fit over the realities of which the world is composed' (Kelly, 1955: 8–9). These patterns are called constructs and enable us to predict our surroundings and choose a direction of behaviour. They can be 'formulated or implicitly acted out, verbally expressed or utterly inarticulate, intellectually reasoned or vegetatively sensed' (Kelly, 1955: 9). Constructs are sometimes described as the intuition, gut feeling or perception that guides our actions without necessarily being verbalised (Björklund, 2008a). In this vein we can surmise that teachers will consider all of their experiences and hypotheses when transforming the curriculum content in order to grade their students. The outcome, their constructs, are sometimes articulated, but sometimes vaguely sensed and, therefore, occasionally described as a gut feeling.

All of our present interpretations of the universe are continuously subject to revision and replacement. We try to improve our constructs to better fit with new experiences by making new constructs, altering and subsuming them with superordinate constructs or systems (Kelly, 1955). A new curriculum, a student with functional limitations or a new political decision to include grades in the evaluation of the schools are examples of such experiences. The impact the experiences have on the teachers' system of constructs depends on how they influence the teachers' realisation arena and how central the constructs are to the teachers (Walker and Winter, 2007).

The Repertory Grid technique

Accessing teachers' tacit knowledge of grading criteria is possible using the Repertory Grid technique (Björklund, 2008b). The technique used in interviews is employed to map and find patterns in the individual constructs, the ones a person is both aware and unaware of, in a

given area. An advantage of the technique is that it helps people to express differences and similarities between familiar elements when having difficulty giving a clear verbal description or definition of the subject (Björklund, 2008a; Borell and Brenner, 1997). For example, even if it is hard to find the words to explain the characteristics of a student with the highest grade, most teachers are still able to describe the distinction between that student and another student if it is their own students, whom they know well. The respondents generate their own constructs, and the bias by the interviewer's preconceptions and language is minimised since few questions are asked. 'The technique can draw on the qualitative methods' strength, openness for what the respondents actually mean – at the same time as it tries to cope with the weakness of these methods, namely a relatively unwieldy, diffuse form of data review and processing' (Borell and Brenner, 1997: 58; authors' translation). In the field of PE and health the Repertory Grid technique has been employed, for example, by Rossi and Hooper (2001).

How the study was carried out

The study involved four Year 9 (the year when the final grades are awarded in compulsory school²) PE teachers. The respondents were two men and two women, one of each gender with a long experience of grading and one of each gender with only a few years' experience. A condition was that all grades should be represented in the class they were teaching. The teachers were informed of the research problem, that they would be anonymous (pseudonyms are used) and participation was voluntary. The four Repertory Grid interviews took place in private rooms of their choice and lasted for about 90 minutes. The Repertory Grid technique was carried out using the three first steps outlined below. The last step, the analysis, was performed by the researchers after the interviews.

Generating elements. In the first step, each of the teachers picked eight of their own students (elements), two of whom had received the highest grade, PSD, three a PD, and three students a P. The number of students with different grades was based on Swedish grade statistics (Skolverket, 2009), where the selected group can be seen as a representation of a class. The students were named by grade, number and gender. For instance, PSD1f represents Passed with Special Distinction, number one, female, and P3m denotes Pass, number three, male. The teachers were instructed not to choose students whom they thought of as in between two grades. It was important that the teachers knew the students well (here coming from their own class) so that it was natural for the teacher to compare them and also that the students represented the variation in the category (here reflecting all grades) (Kelly, 1955; Fransella, Bell and Bannister, 2004). The students were inserted as columns in the grid.

Comparing elements and generating constructs. The constructs are bipolar and based on differences. To understand the meaning of the first pole, it is important to know the opposite pole (Fransella, Bell and Bannister, 2004). In the second step, the interviewer picked three different students, and asked the teacher to describe in what way two of them were similar to each other and different from the third concerning aspects relevant to grading. The expressions the teacher used for the similarity between the two students made up one pole of the construct, and the expressions used for describing the difference of the third student the other pole. For instance, if two students were considered to *strive to develop* in contrast to the third who only *does what she has to do*, the construct is *strive to develop – only does what she has to do*. All differences the teacher came to think about are referred to as constructs. Each construct is inserted into a separate row in the grid. The eight students were presented in different triads, first with all possible combinations of grades and then randomly until the

teacher could not think of any new constructs. Triading is one of the original ways Kelly used to generate constructs (Fransella, Bell and Bannister, 2004).

Rating the elements. In the third step, the teachers were asked to rate the eight students on a five-point Likert scale for every construct they had generated in the grid. On the scale, one represents the negative pole in the construct, for instance *doesn't care*, and five the positive pole, for example *takes responsibility*. When all eight of the students were rated between one and five on every construct, the results were composed of a grid. An example of one of the teacher's grids where the students represent the columns and the constructs the rows is presented in Figure 2. Every student has a value between one and five in every construct.

Figure 2

The analysis. In the last step, the four teachers' constructs were summarised and categorised. Two researchers did the categorisation independently and the results were compared. Out of 86 constructs, only two differed in the two researchers' categorisations. They were placed in different categories after consulting the teachers. The data in the grids were analysed quantitatively using WebGrid 5, a software program developed for Repertory Grid interviews. The resulting PrinGrid maps were analysed to explore how well the constructs matched the grades given.

After the Repertory Grid interviews, the teachers were asked to rank how important they considered each of their own generated construct to be when grading students. A five-point Likert scale was used, with one being not important and five the most important. The value the teacher had given the construct was compared with the relevance the construct had to the grade given when analysing the PrinGrid map. The teachers were also requested to make a list of the documentation they used for the grading of the class. The constructs were compared with the documentation to see how they matched each other. Six months later, there was a follow-up meeting, where the teachers could comment on the analyses.

Themes in teachers' constructs about what to grade

The constructs that the teachers generated in the Repertory Grid interviews can be considered the outcome of the transformation arena and are influenced by both the curriculum and a number of other factors. In the interviews, the four teachers generated 86 constructs and a wide spectrum of aspects that they considered important when grading became evident. These 86 constructs have been categorised into four different themes. They are motivation, knowledge and skills, self-confidence and interaction with others. Three of the themes, motivation, self-confidence and interaction with others, deal with non-achievement factors, not present in the official grading criteria. The fourth theme, knowledge and skills, is the theme closest to the official criteria. The constructs in the theme do not, however, cover all of the learning outcomes in the official criteria. Instead they include some aspects not mentioned, for instance sporting experience and physical qualities. Different areas that emerge in the four themes are revealed and a few examples of the bipolar constructs generated are given in Table 1.

Table 1.

Missing constructs

The Swedish national grading criteria include several abilities missing in the constructs. Students' reflection and analysis of the value of different physical activities, knowledge about health, outdoor education in all seasons and swimming are examples of such. There are no constructs generated by the teachers of students' ability in these learning outcomes even though they are stressed in the curriculum and were taught and documented by the teachers. Only one teacher mentioned anything to do with outdoor activity: map-reading. The missing constructs have in common that they refer to activities that usually take place outside the gymnasium. The interpretation of one teachers whole system of constructs, presented in Figure 2, will be discussed in the next section.

Criteria relevant to the grades

In a PrinGrid map, 'the students are plotted as dots in a n-dimensional space defined by the constructs as axes centred on the mean of the elements' (RepGrid Manual, 2009). Figure 3 shows a PrinGrid map of the grid of one of the teachers, Gloria. The large number of constructs represented by lines in the PrinGrid map is reduced to a few components (the axis in the figure) that can explain most of the variance. In the PrinGrid map, the students are represented by dots. The students with the highest grades are placed to the left in the PrinGrid map, close to the positive poles of the first component. The students with the lowest grades are placed to the right, towards the negative poles. Students with a grade PD are in the middle, which shows that grades are positioned in relation to the first component, henceforward called

the grade-relevant component. All four of the teachers have one such component that explains most of the variance. There is no positioning of the students' grades in relation to the second (vertical) component. Nor do the students' grades relate to components three, four and five, which are not visible in Figure 3. The constructs that are close to and have a high loading on the grade-relevant component can be understood as constructs with greatest relevance to the grades. This means that the students rated high by the teacher on these constructs have also been given a high grade and students rated low by the teacher have been given a low grade. In Figure 3, this is exemplified by the constructs that deal with being *good at ball games* and being an *all-rounder*. The constructs that are far from and have a low loading on the grade-relevant component and closer to another component can be understood to reflect another dimension of grading that does not have great relevance to the grade (e.g. even if the teacher rates the student high on these constructs, it is not necessarily followed by a high grade and the other way around). Constructs about *theoretical knowledge* and *dance* are examples of this (see Figure 3). Therefore, only the grade-relevant component and the five constructs closest to and furthest from that component are discussed in relation to the teachers' rating of their importance.

Figure 3.

Sometimes there is a difference between what the teachers consider important according to their own rating and what relevance it has to the grades according to their PrinGrid maps (see Table 2). This indicates that the teachers are not always aware of what constructs have greatest relevance to the grades. The teacher in the PrinGrid map in Figure 3 is described below to illustrate this.

One teacher's expressed criteria

Gloria is a teacher with six years' experience of grading. She teaches a mixed class (typical for the Swedish school system). Her constructs closest to the grade-relevant component in the PrinGrid map are shown in the left column in Table 2. Gloria agrees that these constructs are important (her ratings are enclosed in the parentheses in Table 2), except for being *good at ball games*, which she does not think should be of such major importance. The five constructs furthest from the grade-relevant component are shown in the right column of Table 2. Three of them, which concern *good theoretical knowledge*, *control of the body* and *technically proficient at movement activities*, are valued as the most important or important by Gloria, but, according to the PrinGrid map, they do not have great relevance to the awarded grades (Table 2). The rest of the constructs that Gloria has assessed as the most important when grading all have a fairly high loading on the grade-relevant component, namely constructs about being *motivated*, *always does his best*, *strives to develop*, *brings clothes*, *active presence*, *involved in the lesson*, *strives to improve*, *motivates others*, *leadership* and being *braver*. Gloria believes that these are the most important qualities for the grades, but, instead, being good at *ball games* turns out to be the most relevant in the PrinGrid map.

Table 2.

Differences and similarities between the teachers

The remaining three teachers have different backgrounds. Bill has four years' experience of grading. He teaches a boys' class. Doris has 15 years' experience of grading. She has only known her current students for one year and tries to adapt to the grading culture of the other

teachers since they teach each other's students. Her class is mixed. Finally, Adam has 12 years' experience of grading. The boys and girls in his class have chosen football as a special interest and, on top of their PE lessons, have additional football training. There are many similarities in what the teachers consider important when grading. All four teachers have generated constructs they consider important in three of the categories in Table 1 and three of them also in the last category: interaction with others. An example of individual differences in the value they give their constructs is an interesting contradiction when one teacher rates motor skills as the most important, while another thinks that motor skills are of no importance. A possible explanation can be seen in the opposite pole. The first teacher's opposite pole is *undeveloped motor skills* and the other's is *motor difficulties*. How the teacher defines and understands motor skills is important for the value the teacher gives it. If the skills can be developed, it is something important to grade, but if it is interpreted as something that you are born with and cannot be blamed for, it is not. The need to pay attention to the teachers' beliefs and values and their influence on professional practice has been stressed by Penney et al. (2009). In Doris's constructs, there is an example of how a construct can relate to something else. She explains the construct *being heavy* in contrast to *having physical prerequisites* and points out that it is not being heavy that is graded, but her experience tells her it causes difficulties in certain activities. This implies that she sees it as an indirect part of her internalised criteria.

In their PrinGrid maps, as expected, most of the constructs are explained by the grade-relevant component. Noteworthy is that all the constructs that relate to individual sports (*racket sports, dance, map-reading*) have little relevance to the grades regardless of how important they are considered by the teachers, while constructs concerning *ball games* are most relevant to the grades. Adam is an exception since he attaches importance to *ball games*,

but it is not reflected in the grades that he has given to the students. Doris is the teacher most aware of her grading practice, in the sense that what she thinks is important is close to the grade-relevant component. *Responsibility* is something that she rates as having a great impact on her grading practice, which it does. Bill ranks *excellent map-reading, does his best and good at everything* as the most important, but, instead, having *sporting experience* and being *good at team games* are closest to the grade-relevant component followed by constructs regarding confidence. Bill rates the construct about having *sporting experience* as unimportant. Some of the qualities that Adam thinks are the most important for the grades (*rhythmic, good at ball games, practical skills* and *developed motor skills*) are among the five constructs with the lowest relevance to the grades. Instead, the ability to interact with the group has the highest relevance according to his PrinGrid map. In Adam's group, no one has a low score in ball games and only one student is considered not so fit: a boy who had been ill and is therefore behind in his motor development. He is rated low on *practical skills, self-confidence* and *easy to get on with his peers* but high on *motivation* and *easy to get on with the teacher*. He was awarded a P at the time of the interview, but in the follow-up meeting Adam says that he has raised the final grade to a PD. Adam is well aware that the skills are low but calls it a 'grade for hard work'. He states that students can't get the highest grade without showing an effort in the lessons. He gives the example that he has had students who compete at an elite level and they could not get a PSD since they did not work hard enough in the lessons. Adam believes he is capable of grading based on a 'gut feeling' and says in the follow up meeting that the grade he feels is right often corresponds to his colleagues' opinion of the student. Since they cannot justify the grades with a 'gut feeling' to the students, they have introduced a number of practical tests (ball games, swimming, orienteering, gymnastics, track and field sports, outdoor activities and leadership) and a written health-awareness test.. The test could have been created to validate the official curriculum but since not all of the

tests are supported by the national grading criteria they can also be interpreted as an alibi for the ‘gut feeling’ when justifying the grades.

Discussion

The aim of this study was to explore what PE teachers consider important when grading students, and to discuss how it can be understood. In particular we were interested in why teachers tend to use internalised criteria for their assessment. First, the spectrum of constructs the teachers regard as important in their grading expressed in the Repertory Grid interviews was examined. Thereafter, the importance the teachers attached to different parts of the spectrum and how this relates to the grades they award the students according to the PrinGrid map were exposed. By means of the generated constructs, we will discuss why the teachers use curriculum-irrelevant criteria, why criteria are missing and how to understand teachers’ inability to predict some constructs’ relevance to the grades.

The Repertory Grid technique makes it possible to explore the spectrum of seldom-verbalised criteria that teachers value. Apart from criteria that are in line with the official ones, the teachers in this study also used other criteria when grading their students. We hope to contribute to the discussion of why they use other criteria. Linde (2012) argues that the selection of content in the transformation arena is influenced by a number of factors. For instance, the classroom situation in PE is special in the sense that all students, regardless of performance level, are supposed to act together to make the lessons work. This can be an explanation for why teachers give cooperation value in their grading even if it does not exist in the official grading criteria. Drawing on the work of Bernstein, Redelius and Hay (2009)

point out that assessment is a powerful message system. Our way of understanding teachers' grading is that they are using the criteria as a message to their students in order to make the lesson work. We would contend that the grading criteria used by the teachers reflect the classroom situation in PE and teachers may gain advantages by using them. When analysing the teachers' non-official constructs three intentions that can help the teachers to manage the situation and facilitate the students' learning can be identified. The teachers want the students to be active, take responsibility for their learning, and to create a stimulating learning environment for all students (cp. teachers beliefs and value in Penney et al. 2009). The intention for the students to be active is represented in the constructs in different ways. On a basic level, it is important that the students *attend*, change into *suitable clothes* and are *active*, preferably with some *intensity*. On a higher level, the teachers also want their students to *challenge themselves*, *try new activities*, and *not to give up*. Sometimes trying hard has a greater impact on the grade than achieving the learning outcomes, exemplified by a 'hard-working grade' as one of the teachers called it. Other constructs that facilitate being active reflect the students' *confidence to be seen*, for instance by their peers while learning new things, or being keen to attract the teacher's attention to show their skills. All the constructs about being active and confident in being active help the students' learning by giving them the opportunity to practise; they also help the teachers to manage the classroom and enable them to assess their students.

Another intention reflected in the constructs is the teachers' ambition for the students to be responsible for their own learning. It can be sensed in the constructs about *respecting deadlines*, *being interested* and *striving to improve their results*. Besides showing responsibility, those constructs also facilitate the lessons, where the teacher has to attend to about 30 students in a limited amount of time. A third intention is to create a learning

environment that is enabling and stimulating for all students. The constructs about *working together* and *leadership* facilitate this. The teachers need the students to work together with all their peers, not just their friends or the ones who have the same interests and skills. In Swedish PE the groups are heterogeneous. Students competing at an elite level in a sport work with beginners in the same sport. To make the lesson work, it is essential that the students with experience assume a leadership role vis-à-vis their inexperienced peers.

A pattern can be detected in the goals that the teachers omit from the official grading criteria when they construe what is important for the grades: they are all performed outside the gymnasium (i.e. swimming, orienteering, winter sports and outdoor life). Only one construct associated with this is generated: the ability to read a map, but it has little relevance to the grade in the PrinGrid map. Swedish schools often have limited access to swimming pools. The lessons last about one hour with short breaks before and after; therefore, transportation limits the opportunities to reach nearby nature. The school facilities and access to learning environments other than the gymnasium seem to be factors that influence the teachers' interpretation of the curriculum in the transformation arena (cp. Linde 2012). The low impact on the grades is supported by the Swedish national evaluation (Skolverket, 2004), where in every grade, there are students who can't swim, dance or do orienteering, and only one-third of those with a PSD claim that they had acquired knowledge of outdoor life. Time and access are restricting factors. In Sweden, the most common PE activity is ball games (Skolverket, 2004), where it is easy for many students to be active at the same time in the time and space available. This is reflected in the PrinGrid maps where ball games have great relevance to the grades and individual sports little relevance. Another possible explanation for the missing constructs could be the teachers' ability and pedagogical content knowledge of the activities outside the gymnasium (cp. Selghed 2004). For sport activities outside the gymnasium, to be

attached importance in the grades, there is a need to reflect on how the teachers can spend enough time on these activities to provide opportunities for the students to achieve the goals in the curriculum and for the teachers to assess them. To achieve this both access and teachers' repertoire must be considered and given adequate support.

The results of this study confirm the findings of Brookhart (1994) and Stiggins (1986), that teachers' grading reflects classroom realities that are not addressed in the grading criteria. In addition, the Repertory Grid technique gives a diversified picture of the teachers' pedagogical aspirations, beliefs and the restrictions that have influenced the transformation of the official grading criteria. We suggest that, from the results of these four teachers, there is a need for further research to find out what the influences are in other contexts.

Some constructs have more relevance to the grades in the PrinGrid map and the teachers have difficulties predicting which. Sometimes constructs that a teacher consider being most important for the grades, like dancing, have little relevance to the grades awarded to the students. On other occasions, constructs that a teacher does not think should influence the grade, like having experience of club sport, have great relevance to the grade awarded. This indicates that the teachers are only aware of part of their grading practice, while some of the grading is tacit and influences the grading taking place in the realisation arena. Another possible explanation for the teachers' difficulty predicting which constructs lack importance is that even if the construct itself has no value for the grade, it matches other constructs that are important. This can be sensed in the construct *physical prerequisites* in contrast to being *heavy*, where the teacher says that it is not being heavy that is graded but the problems it creates in certain activities. Another example is when Bill remarks that previous *sporting*

experience should not matter for the grades, but it matches another construct that is important: being *good at team games*. Both are equally important for the grades in the PrinGrid map. It is the students with previous sporting experience who are good at team games. These indirect ways of assessing the desired knowledge can help the teachers to identify qualities that are easy to detect in a situation with a lot of input. According to PCT, patterns of matching constructs help them to understand and predict their surroundings. However, there is an obvious risk that if Bill is unaware of his system of constructs based on his experiences it makes it harder for him to identify students that are good at ball games without having sporting experience or the other way around.

Conclusion

The concept of criterion-referenced grades as a fair grading system is based on the condition that the criteria are clear enough and teachers are well educated in the grading process. For fair grades to be awarded, it is necessary to accomplish this, but this study indicates that it might not be sufficient. As long as the criteria do not address classroom realities and the restricting factors are not considered, there is an obvious risk that the grades will continue to be inconsistent with the criteria. A discussion is needed on the options the teachers have to meet their pedagogical and disciplinary intentions in other ways than encouraging students' behaviour with high grades. To develop their grading practice, teachers require methods to become aware of what their 'gut feeling' consists of. The Repertory Grid technique is one conceivable option to make their individual constructs in their context visible and possible to discuss. Only when the constructs are verbalised is it possible to discuss how they meet the official claims and to identify needs for support and professional development for sound

grading. We need further studies about how to understand teachers' grading and possible pedagogical and disciplinary implications for PE practice.

Note

¹ Sweden has introduced a new curriculum (Skolverket, 2011) that stresses knowledge and learning outcomes even more. A criterion-referenced system is still used, but the grades range from A to F. The first students to leave the Swedish compulsory school system with A–F grades will finish school in the spring of 2013. This study was conducted before the new curriculum was implemented.

² The Swedish compulsory school system consists of nine years of schooling from the ages of seven to sixteen.

References

- Annerstedt C and 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.
- Bernstein B (2003) *Class, codes and control. Vol. 3, Towards a theory of educational transmission*. London: Routledge.
- Björklund L-E (2008a) The repertory grid technique, Making Tacit Knowledge Explicit: Assessing Creative work and Problem solving skills. In: Middleton H (ed) *Researching Technology Education: Methods and techniques*. Rotterdam: Sense Publishers, pp. 46-69.

- Björklund L-E (2008b) *Från novis till expert: förtrogenhetskunskap i kognitiv och didaktisk belysning* (From Novice to Expert: Intuition in a Cognitive and Educational Perspective). PhD Thesis, Norrköping: University of Linköping, Sweden.
- Borell K and Brenner S-O (1997) *Att spegla verkligheten* (To reflect reality). Lund: Studentlitteratur.
- Brookhart SM (1991) Grading practices and validity. *Educational Measurement: Issues and Practice* 10(1): 35-36.
- Brookhart SM (1994) Teachers' Grading: Practice and Theory. *Applied Measurement In Education* 7(4): 279-302.
- Chan K, Hay P and Tinning R (2011) Understanding the pedagogic discourse of assessment in Physical Education. *Asia-Pacific Journal Of Health, Sport & Physical Education* 2(1): 3-18.
- Cizek G, Fitzgerald S and Rachor R (1996) Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment* 3(2): 159-179.
- Cliffordson C (2008) Differential prediction of study success across academic programs in the Swedish context: The validity of grades and tests as selection instruments for higher education. *Educational Assessment* 13(1): 56-75.
- Cox K (2011) Putting Classroom Grading on the Table: A Reform in Progress. *American Secondary Education* 40(1): 67-87.
- Cross L and Frary R (1999) Hodgepodge Grading: Endorsed by Students and Teachers Alike. *Applied Measurement In Education* 12(1): 53-73.
- Evans J (2004) Making a difference? Education and Ability in Physical Education. *European Physical Education Review* 10(1): 95-108.
- Fransella F, Bell R and Bannister D (2004) *A manual for repertory grid technique*. 2. ed. Chichester, West Sussex: Wiley.

- Hay P and MacDonald D (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 and Penney D (2012) *Assessment in Physical Education: a sociocultural perspective*, London: Routledge
- James A, Griffin L and France T (2005) Perceptions of Assessment in Elementary Physical Education: A Case Study. *Physical Educator* 62(2): 85-95.
- Kelly GA (1955) *The psychology of personal constructs vol. 1. A theory of personality*. New York: W.W. Norton & Company Inc.
- Klapp Lekholm A and Cliffordson C (2008) Discrepancies between school grades and test scores at individual and school level: effects of gender and family background. *Educational Research and Evaluation* 14(2): 181-199.
- Klapp Lekholm A and Cliffordson C (2009) Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation* 15(1): 1-23.
- Korp H (2006) *Lika chanser i gymnasiet?: en studie om betyg, nationella prov och social reproduction* (Same chances in upper secondary school?: a study of grades, national tests and social reproduction). PhD Thesis. University of Lund, Sweden.
- Linde G (2012) *Det ska ni veta!: En introduktion till läroplans teori* (This you should know!:An introduction to the theory of curriculum). 3rd ed. Lund: Studentlitteratur.
- McMillan J and Nash S (2000) Teacher Classroom Assessment and Grading Practices Decision Making. In: *Annual Meeting of the National Council on Measurement in Education*. New Orleans, LA, 25-27 April 2000.
- McMillan JH (2003) Understanding and Improving Teachers' Classroom Assessment Decision Making: Implications for Theory and Practice. *Educational Measurement: Issues And Practice* 22(4): 34-43.

- Mickwitz L (2011) *Rätt betyg för vem? betygsättning som institutionaliserad praktik (The right grade for who? grading as an institutionalised practice)* Stockholm: University of Stockholm.
- Penney D, Brooker R, Hay P and Gillespie L (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 Hay P (2009) Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review* 15(3): 275-294.
- Redelius K, Fagrell B and Larsson H (2009) Symbolic capital in physical education and health: to be, to do or to know? That is the gendered question. *Sport, Education & Society* 14(2): 245-260.
- RepGrid manual (2009) Available at: <http://repgrid.com> (access 7 January 2013).
- Rossi T and Hooper T (2001) Using personal construct theory and narrative methods to facilitate reflexive constructions of teaching physical education. *The Australian Educational Researcher* 28(3): 87-116.
- Selghed B (2004) *Ännu icke godkänt: lärares sätt att erfara betygssystemet och dess tillämpning i yrkesutövningen (Not yet passed: How Teachers Experience a Criterion-Referenced Grading System and what they say about its use in Swedish Secondary School)*. PhD Thesis, University of Lund, Sweden.
- Skolverket (2000) *Grundskolan: kursplaner och betygskriterier (Compulsory school: curriculum and grading criteria)*. Stockholm: Author.
- Skolverket (2004) *Nationella utvärderingen av grundskolan 2003: huvudrapport- bild, hem- och konsumentkunskap, idrott och hälsa, musik och slöjd* (National Evaluation of

- compulsory school 2003: main report- art, domestic science, physical education and health, music and craft). Stockholm: Author.
- Skolverket (2009) Statistics about schools in Sweden. Available at www.skolverket.se/statistik (accessed 3 January 2011).
- Skolverket (2011) *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011* (Curriculum for compulsory school, preschool and youth leisure centre 2011). Stockholm: Author.
- Stiggins R et al. (1986) *Inside High School Grading Practices*. The Northwest Regional Educational Laboratory Program Report no.143.
- Tholin J (2006) *Att kunna klara sig i okänd natur: en studie av betyg och betygskriterier - historiska betingelser och implementering av ett nytt system (Being able to survive in an unknown environment: a study of grades and grading criteria- historical factors and the implication of a new system)*. PhD Thesis, University College of Borås, Sweden.
- Walker BM & Winter DA (2007) The Elaboration of Personal Construct Psychology. *The Annual Review of Psychology* 58(1): 453-477.
- Young S (2011) A Survey of Student Assessment Practice in Physical Education: Recommendations for Grading. *Strategies: A Journal for Physical and Sport Educators* 24(6): 24-26.

Figure 1. A model combining Linde's arenas (2012) with factors that influence teachers when transforming the grading criteria from the formulation arena to the realisation arena based on previous research.

1

2

3 Figure 2. Bill's grid, that shows how he has rated his eight male students on the different
 4 constructs.

5

1

2 **Table 1.** Themes in the four teachers' constructs

Themes	Examples of areas mentioned in the themes	Examples of constructs
Motivation	Strive to perform/improve	<i>strive to develop — only does what she has to performance focused — content</i>
	Responsible	<i>respects deadlines — sloppy</i>
	Involved	<i>Involved in the lesson — little involvement</i>
	Interested	<i>interested — not fun grade focused — choices based on interests</i>
	Attends/participates/clothes	<i>present — absent participates in everything — doesn't participate in everything brings clothes — doesn't bring clothes</i>
	Motivation/does one's best	<i>keeps at it— gives up always does his best — does not want to try new things</i>
Knowledge and skills	Sporting experience	<i>experience of club sport — no experience</i>
	Physical qualities	<i>control of the body — no coordination fit — less fit physical prerequisites — being heavy</i>
	Theoretical knowledge	<i>knowledge of the rules — no knowledge of the rules knowledge of injuries — no knowledge of injuries excellent map-reading — difficulty reading maps</i>
	Specific skills	<i>good at ball games — not so good at ball games good at dancing — not so good at dancing good at racket sports — not so good at racket sports good at individual sports — not so good at individual sports</i>
	General skills	<i>practical skills — less skilled</i>
	All-round	<i>all-rounder — less versatile</i>
Confidence	Confident	<i>confident — unsure</i>
	Brave/has a go	<i>has a go — doesn't have a go</i>
	Gets attention/verbally communicative	<i>wants to be the focus of attention — doesn't care about being the focus of attention gives an opinion — quiet verbally communicative — difficult to express oneself</i>
Interaction with others	Relations with peers	<i>team player — keeps to himself works together — only wants to be with friends</i>
	Relations with teacher	<i>accepts instructions — doesn't accept instructions</i>
	Leadership	<i>motivates others — doesn't motivate others has leadership qualities — follows</i>

3

4

Figure 3. Gloria's PrinGrid map showing her system of constructs, the two components that explains most of the variance and their relation to the students with different grades.

1 **Table 2.** Gloria's five constructs closest to and furthest from the grade-relevant component.

Constructs closest to the grade-relevant component (how Gloria rates their importance for the grades)	Constructs furthest from the grade-relevant component (how Gloria rates their importance for the grades)
<i>Good at ball games – not so good at ball games (3)</i>	<i>Good at dancing – not so good at dancing (3)</i>
<i>All-rounder – less versatile (4)</i>	<i>Theoretical knowledge) – less theoretical knowledge (5)</i>
<i>Team player – keeps to himself (4)</i>	<i>Control of the body – no coordination (4)</i>
<i>Appropriate clothing – inappropriate clothing (5)</i>	<i>Technically proficient at movement activities – less technically proficient at movement activities (4)</i>
<i>Has a go – doesn't have a go (5)</i>	<i>Respects deadlines – sloppy (3)</i>

2

3

4