



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

Postprint

This is the accepted version of a paper published in *Sport, Education and Society*. 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. (2018)

Swedish PE teachers struggle with assessment in a criterion-referenced grading system

*Sport, Education and Society*, 23(4): 381-393

<https://doi.org/10.1080/13573322.2016.1200025>

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-4491>

## **Swedish PE teachers struggle with assessment in a criterion-referenced grading system**

Lena Svennberg<sup>ab\*</sup>, Jane Meckbach<sup>b</sup> and Karin Redelius<sup>b</sup>

<sup>a</sup> Faculty of Health and Occupational Studies, University of Gävle, Gävle, Sweden

<sup>b</sup> Department of Sport and Health Sciences. The Swedish School of Sport and Health Sciences, Stockholm, Sweden

---

\* Corresponding author. The Swedish School of sport and health sciences (GIH), Box 5626, 114 86 Stockholm, Sweden. Email: lsb@hig.se

## Swedish PE teachers struggle with assessment in a criterion-referenced grading system

### Abstract

In the field of education, the international trend is to turn to criterion-referenced grading in the hope of achieving accountable and consistent grades. Despite a national criterion-referenced grading system emphasising knowledge as the only base for grading, Swedish physical education (PE) grades have been shown to value non-knowledge factors, such as students' characteristics and behaviour. In 2011 a new national curriculum was implemented which attempts to deal with the problem by prescribing specific knowledge requirements with a clear progression as the only basis for different grades. The aim of the present study is to explore the impact of the new knowledge requirements on what teachers consider important when assigning grades. It is also to discuss what non-knowledge-related aspects (if any) teachers continue to look for and why these seem to remain resilient to the reform. The Repertory Grid technique was employed to interview the teachers before (2009) and after the implementation (2013). During the interviews the grading of 45 students was discussed, which generated 125 constructs. After the implementation there was a near doubling of knowledge constructs, half as many motivation constructs and an almost total elimination of constructs based on confidence and social skills. While motivational factors were still considered valuable for the award of a higher grade, clear criteria seemed to be important, but too limited for the teachers' needs. In order to understand the persistence of motivational factors, we discuss the results in relation to Bernstein's interrelated message systems of curriculum, pedagogy and assessment. We emphasise the need to discuss how valid grades can be achieved and, at the same time, give value to the regulative discourse in order to realise the overarching national goals of values and norms in education and PE.

**Keywords:** *Bernstein; Curriculum regulation; Interrelated message systems; Motivation; Pedagogic discourse; Physical education; Regulative discourse; Repertory Grid; Standards-based grading; Teachers grading practice*

Assessment sends a powerful message about what counts as legitimate knowledge (Bernstein, 2003; Hay and Penney, 2013). Therefore it is important to study the base for teachers' assessment of students, and how it aligns with official criteria. Penney, Brooker, Hay, and

Gillespie (2009) consider alignment in the message systems of assessment, pedagogy and curriculum essential for high quality physical education (PE). Alignment is also important for validity and accountability. As a result of today's neoliberal agenda, schools are under increasing pressure to adopt educational models that move toward accountability and standardisation of education (Connell, 2013; Evans & Davies, 2014). An ambition for evidence of achievement, is thus visible in many countries' neoliberal educational reforms (Klenowski & Wyatt-Smith, 2012). The international trend is to turn to criterion-referenced grading in the hope of achieving accountable and consistent grades. Countries have arrived at different solutions in an attempt to balance curriculum regulation with adjustments to local context and student populations (Kuiper & Berkvens, 2013).

In line with neoliberal ideas the Swedish school system has gone from one of the most centralised in the western world to one of the most decentralised systems (Daun, 2006; von Greiff, 2009). Even if the responsibility has shifted from the state to local schools, the state still decides the goals and the grading criteria that all schools in the country are obliged to accomplish. A goal-oriented grading system, allowing great freedom to make adjustments to local context as long as the goals are achieved, was introduced in 1994 (Swedish National Agency for Education [SNAE], 1994). However, the goals that were to be achieved were difficult to interpret and not clear enough for teachers. Although the goals emphasised knowledge as the only base for grading, teachers valued non-knowledge factors such as students' characteristics and behaviour (Klapp Lekholm & Cliffordson, 2009; Tholin, 2006).

The gap between national grading criteria and teachers' grading practices has also been recognised in PE, where motivation and effort are often used as a basis for grading decisions in addition to the intended goals (Redelius, Fagrell & Larsson, 2009; SNAE, 2004; Tholin, 2006). Validity is especially important in the Swedish school system, where grades (including grades in physical education [PE]) are high-stakes and used as selection instruments to higher

education. However validity and reliability of PE grades have been strongly questioned by Annerstedt and Larsson (2010). Teachers sometimes also refer to internalised grading, which can jeopardise demands of transparency (Annerstedt & Larsson, 2010; Svennberg, Meckbach & Redelius, 2014).

To address the problem, the Swedish government tried to increase regulation by implementing a new curriculum in 2011 (SNAE, 2011). National knowledge requirements, identifying the expected achievements for the different grades, form a new part in the curriculum text (SNAE 2011) with evidence and knowledge in focus (Sivesind, 2012). In addition to this, Swedish curriculum also refers to general purposes, principles and guidelines and thereby, as Sivesind (2012, p. 52) points out, ‘merge professional semantics on the purposes, and content of schooling with a language formed by evidence-based policy on outcomes’. Wyatt-Smith, Klenowski and Gunn (2010) emphasise how ‘professional judgement may not be readily circumscribed by any given set of standards and how judgement may remain responsive to the influence of other knowledge and skills’ (p. 60). If so, does it matter what message is mediated in the knowledge requirements?

Against this background, the aim of the study is to explore the impact of the new knowledge requirements on what teachers consider important when assigning grades. The aim is also to discuss what non-knowledge-related aspects (if any) teachers continue to look for and why these seem to remain resilient to the reform. The results are discussed in the light of the three interrelated message systems of curriculum, pedagogy and assessment (Bernstein, 2003).

To provide a context, previous research about the gap between policy and practice is presented first and thereafter the new curriculum (SNAE, 2011), initiated by this gap is introduced. Finally the new curriculum including national knowledge requirements is

elaborated in light of Bernstein's concept of the interrelated message system under the heading Theoretical framework.

### **The 'gap'**

In the Swedish debate, the gap between the national grading criteria and teachers' grading practices has been explained in terms of vaguely worded criteria and a lack of support for teachers (e.g. Tholin 2006). The vagueness of the criteria have also been criticised for endangering students' rights to be equally assessed, regardless of which teacher they have or which school they attend (Swedish National Audit Office, 2011). Internationally teachers' attitudes and interest in student management are common explanations for the inclusion of affective characteristics when grading students' work or performances (e.g. Brookhart, 2011; McMillan, 2003; Penney et al., 2009). Factors influencing the interpretation and implementation of curriculum have been studied. Reasons for teachers' decision-making are found to be embedded in the context in which they work (MacPhail, 2004) and influenced by occupational socialisation (Curtner-Smith, 1999).

A review of international literature shows that subjective assessment criteria, such as effort and clothing, are often included in traditional assessment (López-Pastor, Kirk, Lorente-Catalán, MacPhail, & Macdonald, 2013). Regardless of subject, the habit of grading students' affective characteristics is shown in countries such as Greece (Ikonopoulou, Tzetzis, Kioumourtzoglou, & Tsorbatzoudis, 2006), England (Biddle & Goudas, 1997), Israel (Biberman-Shalev, Sabbagh, Resh, & Kramarski, 2011; Resh, 2009), Canada (Tierney, Simon, & Charland, 2011), Australia (Chan, Hay, & Tinning, 2011), China (Sun & Chen, 2014) and the United States (Brookhart, 2013; Cross & Frary, 1999; Weiyun, 2005; Young, 2011). Brookhart calls teachers' grading 'a hodgepodge grade of attitude, effort and

achievement' (1991, p. 36) and is concerned about teachers' attitudes. She challenges teachers to consider the position: 'grades are not about what students earn; they are about what students learn' (2011, p. 12). The new curriculum implemented in Sweden in 2011 is an attempt to address the problem described above by giving clearer criteria (SNAE, 2011).

### **Swedish curriculum regulation**

The Swedish curriculum consists of three parts. The first part is fundamental values and tasks of the school, the second part is overall goals and guidelines and finally the third part consist of syllabuses for each subject. The first two parts are the same in the new and the former curriculum, only the syllabuses have been rewritten. In the first two parts, both the new and the former Swedish curriculum declare the overarching goals of education to be the learning of knowledge, values and norms. Here, the concept of knowledge is understood in its broadest sense and includes facts, understanding, skills, familiarity and accumulated experiences. Goals for values and norms are introduced as: 'The school should actively and consciously influence and stimulate pupils into embracing the common values of our society, and their expression in practical daily action' (SNAE, 2006, 2011, p.14). Equity is an important goal for education where every individual's specific circumstances and needs should be taken into account and different ways to reach the goals are encouraged. Teaching shall stimulate each pupil towards self-development and personal growth and to develop increasingly greater responsibility for their studies (SNAE, 2006, 2011).

The goals in both the new and former PE syllabus also include knowledge, values and norms. In the 2011 PE syllabus, the content knowledge is basically the same as in the former curriculum, but in the present document it has been organised in three knowledge areas: *movement, health and lifestyle* and *outdoor life and activities*. Students should also develop

values and norms such as: ‘an interest in being physically active and spending time outdoors in nature [...] a healthy lifestyle [...] their interpersonal skills and respect for others [...] and a belief in their own physical capacity’ (SNAE, 2011, p. 50). The stated knowledge requirements for PE reflect what are often referred to as knowledge- or achievement-based criteria. Typical criteria when grading are that students should be able to demonstrate, describe or in other ways reproduce the acquired knowledge. Grades are expected to express only the extent to which the national knowledge requirements for PE have been attained.

However, values and norms are not represented in the new knowledge requirements and in this context the grading of, for example, students’ characteristics and behaviour is considered criterion-irrelevant and a threat to validity and accountability. The knowledge requirements are organised in the three knowledge areas described above and clarified with a progression that allows a matrix to be used. The number of steps on the grading scale has increased from three to six, A-F, where A is the highest grade and F represents a fail grade. The requirements are to be measured qualitatively and expressions such as: to some extent, relatively well and well, are intended to guide the teachers in differing between the grades E, C and A. Guidelines for teachers about how to use the knowledge requirements have been produced and are available on the SNAE website. Municipals and schools should assist teachers in the implementation of the new knowledge requirements but conditions differ between schools.

### **Theoretical framework**

To better understand curriculum’s impact on teachers’ grading practice our starting point is the interrelation between the three message systems identified by Bernstein (2003):



Formal educational knowledge can be considered to be realized through three message systems: curriculum, pedagogy and evaluation. Curriculum defines what counts as valid knowledge, pedagogy defines what counts as valid transmission of knowledge, and evaluation defines what counts as valid realization of this knowledge on the part of the taught. (p. 85)

Like Hay and Penney (2013), we identify evaluation (assessment in our current vernacular [Hay & Penney, 2013, p. 5]) as sending the most powerful message of what counts as legitimate knowledge. Chan et al. (2011) argue that assessment influences the teaching and learning process and defines the educational product. High quality curriculum is thought to be the product of alignment between curriculum, pedagogy and assessment, which together send a consistent message of what knowledge ‘counts’ (Penney et al., 2009). Some studies have explored the influence of assessment on teachers’ and students’ perceptions of learning and important knowledge (Chan et al., 2011; James, Griffin & France, 2005; Redelius & Hay, 2009; Thorburn, 2007) and found them related. Bernstein’s concept of a pedagogic discourse and how the interrelated message systems are applied in a Swedish context are described in the following sections.

Bernstein’s (1996) concept of a pedagogic discourse includes an instructional discourse, which creates specialised skills, and a regulative discourse, which creates order and relations. In this paper, we refer to the instructional discourse as knowledge and the regulative discourse as values and norms. According to Bernstein, the two discourses cannot be separated:

Often people in schools and classroom make a distinction between what they call the transmission of skills and the transmission of values. These are kept apart as there was a conspiracy to disguise the fact that there is only one discourse (Bernstein 1996, p. 46).

The instructional discourse is embedded in the regulative discourse, which means that the regulative discourse is dominant (Bernstein, 1996). Since the instructional discourse cannot be separated from the regulative discourse within the pedagogic discourse, we can assume that knowledge, as well as values and norms, are mediated in pedagogy.

The interrelated message systems can be understood as communicating different messages about what students are expected to achieve in the Swedish context. On the one hand, national goals for education and for PE identify knowledge, values and norms as equally important. On the other hand, the grading criteria only assess whether a student has acquired the stipulated knowledge. In order to better understand teachers' grading practices, we are interested in how the goals in curriculum influence assessment and grading. If the grading criteria are applied, knowledge stands out as more important than values and norms, and thus sends a different message than the rest of the curriculum. This inconsistency can also be compared with the lack of alignment between teachers' espoused agendas, lesson tasks and assessment suggested by James, Griffin & Dodd (2008) and the lack of constructive alignment between what teachers consider to be the goals for PE, important knowledge and what they use as grounds for assessment (Redelius et al., 2009).

Grades have a component of socially valued knowledge (Evans, 2004; Hay & MacDonald, 2010). Both the instructional and the regulative discourse reflect what is considered valid in society and both are expected to be transferred in school. Both discourses can either be visible and measurable, what Bernstein (2003) calls visible pedagogy, or implicit, referred to as invisible pedagogy. The Swedish curriculum suggests a visible pedagogy for the knowledge dimension. In this context, the grading criteria are transparent and available to students and parents alike. Values and norms, on the other hand, have no transparent and measurable criteria and in this respect an invisible pedagogy is applied.

## **Methodology**

Teachers sometimes use internalised criteria and seldom verbalised criteria, also described as gut feelings (Annerstedt & Larson, 2010; Hay & MacDonald, 2008; Svennberg et al., 2014).

In our study, we have employed the RG interview technique that helps people verbalise their perceptions of a specific subject matter that they have experienced and are familiar with (Fransella, Bell, & Bannister, 2004). Even if teachers find it difficult to formulate what is necessary in order to be awarded a high grade, they are often able to describe the differences between a student with a high grade and another student if they know their own students well enough. These differences are called constructs. George Kelly (1955), who first developed the technique, believed that constructs facilitate the prediction of one's surroundings and enable people to choose which direction their behaviour should take. Constructs can be 'formulated or implicitly acted out, verbally expressed or utterly inarticulate' (Kelly, 1955, p. 9). The method not only helps us to reveal the presence of criterion-irrelevant constructs, but also their content. As the teachers generate their own constructs around a topic they are familiar with, the risk of directing the interview with questions based on a precondition other than their own is minimised (Borg, 2008). It is difficult to make any statement about the method's validity and reliability, because there is no standard-grid and the method can be used in many different areas. If, as in this paper, RG is regarded as a technique rather than a method, the definition of the validity is the usability of the technique. According to Fransella et al. (2004), the validity is high in order to differentiate between groups and predict behaviour.

The three teachers participating in the study had received a teacher education degree in PE and were grading students aged between 15 and 16 in a compulsory school setting, where final grades are used as instruments for selection to upper secondary school. Grades were only assigned the last two years of compulsory school at that time. A requirement was that a full distribution of grades was assigned in the same class to facilitate the RG-technique. One of the teachers, Adam, had 12 years' experience of grading and was teaching PE at the time of the first interview. In his school he worked with several PE teacher colleagues. The second teacher, Bill, had 4 years' experience of grading at the time of the first interview and taught

PE in a small school. In contrast to Adam, Bill had no colleagues with whom he could discuss his grading. The third teacher, Gloria, had 6 years' experience of grading at the time of the first interview. She taught PE and English and had several PE colleagues. The teachers were told that their identities would not be revealed (pseudonyms are used in the text), that participation was voluntary and that they could choose to withdraw from the study at any time. The RG interviews took place in private rooms of their choice and lasted for about 90 minutes each. The teachers were interviewed twice. The first round of interviews was carried out in the spring of 2009, before the implementation of the new curriculum, and the second round was performed in the spring of 2013, when the implementation had been completed in all the year cohorts. The RG technique was used at both occasions in order to explore what they considered important for awarding a high grade.

The RG interviews were conducted in two steps. In the first step, the teacher was asked to select seven to eight students from the same class that he or she taught and graded in PE (generating elements – in this case students). The selected students had to represent all the possible grades (Fransella et al., 2004). The names of the students under consideration were written on pieces of paper. In the second step (generating constructs) the names of three of the students were presented to the teacher, who was then asked to describe in what ways two of the students were similar and different from the third (Fransella et al., 2004) in the aspects that were regarded as important for the awarded grade. The similarities and differences expressed by the teachers constituted a bipolar construct, for instance, *always did his best – did not want to try new things*. The teachers were presented with different combinations of students until they were no longer able to think of any new constructs. The constructs represent what the teachers consider to be important when awarding grades. In total, the empirical data consists of 125 constructs about 45 students derived from six interviews with three teachers.

The constructs in 2009 were regarded as key thoughts and were coded and categorised using an inductive approach. After becoming familiar with the contextualisation of the constructs, four themes were identified. A follow-up meeting with each of the three teachers was held in order to confirm the identification of the themes, which allowed us to discuss the constructs and the categories and to correct errors and challenge interpretations. An experienced peer researcher independently categorised the constructs and arrived at the same categories. The two constructs that differed were categorised after checking with the teachers who had generated them during the follow-up meetings. In 2013, the same themes were used to categorise the teachers' new constructs. The interviews were recorded and transcribed, which made it possible to compare the teachers' contextualisations of their constructs.

## **Results**

In the first interview setting, when the teachers were expected to grade using the previous grading criteria, they generated in total 66 constructs: 25 constructs about knowledge and 41 constructs about values and norms. The constructs relating to values and norms focused on three themes: motivation, confidence and social skills, including leadership skills. In the second interview, after the implementation of the new curriculum, the teachers had made major alterations to the things they looked for in assessment.

### Table 1

As shown in Table 1 there were noticeable changes from 2009 to 2013 in what teachers considered important, a near doubling of knowledge constructs, half as many motivation constructs and an almost total elimination of constructs based on confidence and social skills.

## *Knowledge*

A near doubling of constructs relating to knowledge indicates that a discourse equating assessment of knowledge with best practice have made an impact on the teachers' grading practices. Knowledge comes more easily to mind when they reflect on the similarities and differences between students with different grades. Their constructs in 2009 address specific skills, general skills, all-roundness, theoretical knowledge, physical qualities and sporting experience on both interview occasions. Constructs that exemplifies these sub-themes are: *good at ball-games – not so good at ballgames, has practical skills – have less practical skills, all-rounder – less versatile, got strength – weaker and have experience of club sport – has no experience of club sport (2009).*

In 2013 the same sub-themes are mentioned but the teachers talk about theoretical knowledge in new ways and more often (increasing from three constructs in 2009 to ten in 2013). Planning and reflection on the activities are now in focus, for instance: *can plan and evaluate – can plan but not evaluate, can reflect on activities – participate but have difficulties to reflect on activities and has theoretical knowledge when planning – has no depth when planning.*

Planning of activities and evaluation of health effects are national knowledge requirements for the knowledge area of health and lifestyle in both the former and the new grading criteria. However, a national evaluation of PE carried out during the former curriculum show that discussions and reflections about activities were seldom practiced. (SNAE 2004). Progressive knowledge requirements seem to help teachers to include planning and evaluation in their assessments. Despite the alterations made in teachers' practices, values and norms are still part of their constructs and we want to explore what non-knowledge aspects they continue to look for and discuss why these remain resilient to the reforms in the

national criteria. Below we present the teachers' constructs categorised in the three themes motivation, confidence and social skills.

### *Values and norms*

The 2011 curriculum document (SNAE, 2011) state that only the specified knowledge requirements are to be used as a basis for grading students and thus values and norms should not be used as a base for grading in the national knowledge requirements. Constructs about confidence and social skills almost disappeared in teachers' talk after the implementation of clearer progressive knowledge requirements but motivation is still considered important (Table 2).

#### Table 2

Both in 2009 and in 2013 the teachers value high activity and ambition to learn (Table 2).

Although motivation is not a national knowledge requirement, the quotations from the teachers, below, indicate that grading motivation can be interpreted as helping students to reach the goals and facilitating assessment. When Adam ponders how to differentiate two of his students from the third in the RG interview, he says:

In general, S and V [two students with the grade A, our comment] are better in all the activities because they are more motivated to learn and show me what they can do than D are [the third student, our comment]. That's also why he has a B (2013).

Bill gives a similar explanation for his constructs *challenge themselves – does what he has to*: 'You often give those wanting a higher grade the opportunity to try other things... Those who are happy with a pass only do what is required to pass, whereas the others like a challenge' (2013).

Another explanation for the persistence of the motivation constructs can be the national goal to develop students' interest in physical activity and a healthy lifestyle and for the students to develop responsibility for their own studies. Constructs about interest and joy in

movement and to strive to improve, (Table 2), can be interpreted to facilitate the development. In line with this, Gloria comments that attendance is necessary for assessment and learning. ‘[I]f you are absent a lot, you can’t participate in all activities.(...) I am not allowed to grade just ... attendance, but if you are absent a lot there is not much to assess’ (2013). Gloria finds it strange that you cannot grade attendance but it is still reflected in the grades. The teachers’ concern about the learning process and students’ opportunities to be assessed (Table 2) can be explanations for why they value motivation.

While all three teachers still value the fact that students are motivated the two themes of confidence and social skills are almost eliminated. When the teachers talked about what they valued before the reform, they all mentioned confidence in different contexts (Table 2). The goal in the PE syllabus for teaching to develop students’ belief in their own physical capacity can be tracked for instance in the constructs about being brave and daring to try new activities. Constructs about confidence to express something orally is an advantage when assessing theoretical knowledge. None of the teachers generated any constructs about students’ confidence after the reform. It would thus seem as though motivation is more persistent than confidence to changes in official grading criteria.

In 2013, social skills constitute a new theme in Bill’s constructs, represented by the construct *work for the best of the group – egotistic*. In 2009, every time Bill said anything associated with social skills, in contrast to the other two teachers he rejected it as a construct since he knew it was not relevant to the grades. In 2013 after struggling with the expectancy of not using social skills as base for grading, he said the following:

I don’t know, maybe it’s not relevant for the grades, but these two are good role models, good mates and work...but it has nothing to do with grades. It is supposed to be something that affects the grade. Ehh ... but that’s just it, if you can take that as an example... There’s a reason for this one not being given an A. This one..., the other two work hard and are good in the group as well, they are team players, they are



really good at working with others and focusing on what is best for the group, whereas this one is very egotistic, finds it difficult to work with others, with team games, collective exercises and so on.

Even though the teachers seem to abandon constructs concerning confidence and social skills including leadership, it is still possible that these constructs impact the grade in the way that Bill explains it above. In 2009, Gloria and Adam generated constructs relating to skills in leadership, which also matched the grades awarded to the group of students (Authors, 2014). In 2013, none of the three teachers mentioned leadership, which was an important factor for getting a high grade in 2009 (cf. Redelius et al., 2009). In order to assess leadership in 2009, the teachers allowed the students to lead some of the lessons. Gloria commented that this was no longer the case in 2013:

They used to have a theory assignment to plan for others. Leading a lesson has been removed, and instead I sometimes let them plan their own training schedules. At present they are planning a fitness lesson or a strength training lesson and sometimes they can choose what to plan.

To sum up the results, there have been changes in the ways in which the teachers talk about grading, in that values and norms are no longer as important and knowledge has gained ground. The teachers still take students' motivation into account when grading, but confidence and social skills now seem to be of lesser importance.

## **Discussion and conclusion**

The aim of the study presented in this article was to explore the impact of new knowledge requirements on teachers' grading practice. It was also to explore what non-knowledge-related aspects (if any) teachers continue to look for, and to discuss why these seem to remain resilient to the reform. The RG interview technique was used to enable us to examine all the grading criteria, including those seldom verbalised. The results illustrate that more specific criteria seem to direct teachers' attention to factors that should be graded and away from,

what Hay and Penney call, ‘irrelevant factors such as students’ dispositional and behavioural characteristics’ (2009, p. 398). The teachers in the study focus more on knowledge which is in line with the prescribed requirements. Clear grading criteria, or knowledge requirements, and the provision of support to help teachers interpret them, are necessary in order to achieve comparable grades in a high-stakes grading system. However, the results also indicate that clearer knowledge requirements alone do not lead to teachers ceasing to grade values and norms. The results point to the impact of the regulative discourse and how teachers sometimes use values and norms as additional criteria. Our interpretation is that teachers struggle to find a balance between more specific criterion-referenced assessment and more process-oriented non-criterion referenced learning. The neoliberal influences advocating a performative culture that emphasises standardised learning outcomes (Evans & Davies, 2014; Ferry, Meckbach & Larsson, 2013) seem to be too limited for the teachers’ needs.

Bernstein’s (1996) concept of the interrelated message systems of curriculum, pedagogy and assessment can be employed to better understand what teachers value. The official message in the Swedish curriculum is that values and norms, as well as knowledge, are important in order to reach the overarching goals of education and the goals for PE (SNAE, 2006, 2011) but only specified knowledge is to be used as criteria for grading. In the concept of the pedagogic discourse, Bernstein (1996) claims that the instructional discourse is deeply embedded in the regulative discourse. As the regulative discourse is always present in the pedagogic discourse, and values and norms are important goals of education and PE in the curriculum, teachers sometimes compensate by creating an alignment between curriculum and assessment in a grading system in which the regulative discourse should not be graded. Penney (2013) emphasises that the dynamic between curriculum, pedagogy and assessment is inseparable from the knowledge structure in PE.

How can we understand that motivation is more prevalent than confidence and social skills as a base for assigning grades? One possible explanation for the prevalence is that the three teachers prioritise an interest in physical activity and a healthy lifestyle when grading, rather than goals concerning ‘interpersonal skills’ and ‘a belief in their own capacity’ in the PE syllabus (SNAE, 2011, p.50). Part of the legitimacy of PE is to encourage young people to live an active life and have a healthy lifestyle. Different activity theories claim that a positive attitude or motivation is an important prerequisite for activity and that knowledge alone does not motivate people to be active. Biddle and Goudas’ (1997) found that PE teachers prefer to grade their students on effort and performance, because effort is controllable and perceived autonomy is a predictor of intrinsic motivation in PE. If experienced teachers are convinced that effort and motivation will result in a more active lifestyle for their students, they are more likely to include them in their grading. We know from earlier studies that teachers sometimes tend to adjust their assessment to strengthen the students’ motivation (Biddle & Goudas, 1997; McMillan, 2003).

Another possible interpretation in the Swedish context is the impact of the national goal for students to develop a greater responsibility for their studies, (SNAE, 2011). This goal can be tracked in many of the motivation constructs mentioned by the teachers as being important in the grading process, such as *always participating*, *striving to develop* and *being motivated to learn*. Motivation can be considered a facilitator to learn and be assessed. Since it is considered an important ingredient in the process of learning Gloria finds it strange not to assess.

The use of values and norms has previously been found to be more common when grading younger students (Chan et al., 2011), lower-achieving students (Korp, 2006) and students from lower socio-economic backgrounds (Klapp Lekholm & Cliffordson, 2008). This suggests that different classroom demands influence the balance between the

instructional and regulative discourse and, therefore, the grading. If the students are already familiar with the prevalent values and norms there is less need to signal their value by using the message system of grading. Even though the teachers are left with an invisible pedagogy without transparent criteria for values and norms it seems like students are aware that behavioural characteristics are used as basis for grading decisions in PE (Redelius & Hay, 2009). Cliffordson (2008) suggests that teachers' grading of these qualities could explain why the success of students in higher education is predicted by grades rather than scholastic aptitude tests. If the definition of validity for selection to higher education is to measure the qualities that will lead to fewer dropouts and failures, the grading of motivation as a facilitator for learning does not seem to be a problem.

Inconsistency in the interrelated message systems can have a number of consequences. Here, we have discussed the confusion that can arise in teachers' grading. As that which is considered valid in the school situation reflects what is considered valid in society (Bernstein 2003), the grading of values and norms is accepted by students (Redelius & Hay, 2009; Cross & Frary, 1999) and probably also by parents (Cross & Frary, 1999). When students and parents accept the grading of values and norms there is little incentive to change the practice. On the other hand, through the grading criteria, society communicates that no-one should have access to higher education simply because they are responsible or interested, but because they have the required knowledge. The alignment between knowledge requirements and grading practice is essential for the validity and reliability of grades. In order to achieve this alignment, it is important to acknowledge that the instructional discourse is embedded in the regulative discourse (Bernstein, 1996). It is therefore important to discuss how the struggle between the pedagogic discourse, with values and norms embedded in it, and the need for grades to align with the knowledge-based criteria should be dealt with. Could the learning of values and norms be promoted in other ways? Could teachers find other ways of motivating

students and endorsing values and norms? If norms and values are regarded as something to be learned, and not as personal traits, what kind of opportunities can be created to promote such learning? More research is needed about how to achieve valid grades and, at the same time, acknowledge the need to give value to the regulative discourse in order to accomplish the goals of the curriculum.

We are aware that many factors have influenced the results of this study and that these need to be discussed. Apart from clearer knowledge requirements, the teachers also experienced an implementation process and were given support that was likely to enhance the effect of the criteria. They also had time to adapt to the idea of criterion-referenced grades when the previous curriculum was in use. It is also possible that the RG technique used in the first interview helped the three teachers to become more aware of their own grading practices. In addition, this insight could have helped to develop their grading and may have affected the constructs generated in the second interview. Nevertheless, despite any additional factors that might have contributed to the three teachers' diminished use of values and norms in their grading, they still find motivation important when awarding a high grade. The results verify earlier research about values and norms being present, regardless of national criteria. They also contribute to knowledge about which characteristics that are resistant to changes in official reforms and point to teachers' struggle with the difference between more specific criterion-referenced assessment and more process-oriented non-criterion referenced learning,

## **References**

- Annerstedt, C., & Larsson, S. (2010). 'I have my own picture of what the demands are ... ': Grading in Swedish PEH - problems of validity, comparability and fairness. *European Physical Education Review*, 16(2), 97-115. doi:10.1177/1356336X10381299

- Bernstein, B. (1996). *Pedagogy, Symbolic Control and Identity: Theory, Research, Critique*. London, England: Taylor & Francis Ltd.
- Bernstein, B. (2003). *Class, codes and control*. (Vol. 3) *Towards a theory of educational transmission*. London, England: Routledge & Kegan Paul.
- Biberman-Shalev, L., Sabbagh, C., Resh, N., & Kramarski, B. (2011). Grading styles and disciplinary expertise: The mediating role of the teacher's perception of the subject matter. *Teaching and Teacher Education*, 27(5), 831-840.  
doi:10.1016/j.tate.2011.01.007
- Biddle, S., & Goudas, M. (1997). Effort is virtuous: Teacher preferences of pupil effort, ability and grading in physical education. *Educational Research*, 39(3), 350-355.  
Retrieved from Web Of Science: 00071059800010
- Borg, K. (2008). 'Repertory grid som forskningsmetod' [Repertory Grid as research method]. In V. Lindberg & K. Borg (Eds.), *Kunskapande, kommunikation och bedömning i gestaltande utbildning*, Stockholm: Stockholms universitets förlag.
- Brookhart, S. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35-36. doi: 10.1111/j.1745-3992.1991.tb00182.x
- Brookhart, S. (2011). Starting the Conversation About Grading. *Educational Leadership*, 69(3), 10-14. Retrieved from ERIC: EJ963092
- Brookhart, S. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment In Education: Principles, Policy & Practice*, 20(1), 69-90.  
doi:10.1080/0969594X.2012.703170
- Chan, K., Hay, P., & Tinning, R. (2011). Understanding the pedagogic discourse of assessment in Physical Education. *Asia-Pacific Journal Of Health, Sport & Physical Education*, 2(1), 3-18. doi:10.1080/18377122.2011.9730340

- 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. doi:10.1080/10627190801968240
- Connell, R. (2013). Why do market 'reforms' persistently increase inequality?. *Discourse: Studies In The Cultural Politics Of Education, 34*(2), 279-285.  
doi:10.1080/01596306.2013.770253
- Cross, L., & Frary, R. (1999). Hodgepodge Grading: Endorsed by Students and Teachers Alike. *Applied Measurement In Education, 12*(1) 53-73.  
doi:10.1207/s15324818ame1201\_4
- Curtner-Smith, M. D. (1999). The more things change the more they stay the same: Factors Influencing Teachers' Interpretations and Delivery of National Curriculum Physical Education. *Sport, Education & Society, 4*(1), 75-97. doi:10.1080/1357332990040106
- Daun, H. (2006). Privatisation, Decentralisation and Governance in Education in the Czech Republic, England, France, Germany and Sweden. In J. Zajda (Ed.), *Decentralisation and Privatisation in Education: The Role of the State*, (pp75–96). Dordechts: Springer.
- Evans, J. (2004). Making a difference? Education and 'ability' in physical education. *European Physical Education Review, 10*(1), 95-108. doi:10.1177/1356336X04042158
- Evans, J., & Davies, B. (2014). Physical Education PLC: neoliberalism, curriculum and gouvernance. New directions for PESP research. *Sport, Education % sSociety, 19*(7), 869-884. doi: 10.1080/13573322.2013.850072
- Ferry, M., Meckbach, J., & Larsson, H. (2013) School Sport in Sweden: what is it, and how did it come to be? *Sport in Society*, vol.16 (6), pp.805-818.
- Fransella, F., Bell, R., & Bannister, D. (2004). *A manual for repertory grid technique*. (2nd. ed.). Chichester, West Sussex: Wiley.

- Hay, P., & 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. doi:10.1080/09695940802164184
- Hay, P., & MacDonald, D. (2010). Evidence for the social construction of ability in Physical education. *Sport, Education and Society*, 15(1), 1-18. doi:10.1080/13573320903217075
- Hay, P., & Penney, D. (2009). Proposing Conditions for Assessment Efficacy in Physical Education. *European Physical Education Review*, 15(3), 389-405. doi:10.1177/1356336X09364294
- Hay, P., & Penney, D. (2013). *Assessment in Physical Education: a sociocultural perspective*. London: Routledge
- Ikonomopoulos, G., Tzetzis, G., Kioumourtzoglou, E., & Tsorbatzoudis, C.(2006). Attitudes and Grading Practices of Physical Educators in Greece. *International Journal of Physical Education*, 43(1), 23-31. Retrieved from SPORTDiscus:SPHS-1022947
- James, L., Griffin, L., & Dodds, P. (2008). The Relationship Between Instructional Alignment and the Ecology of Physical Education. *Journal of Teaching In Physical Education*, 27(3), 308-326. Retrieved from SPORTDiscus: 33018632
- James, A., Griffin, L., & France, T. (2005). Perceptions of Assessment in Elementary Physical Education: A Case Study. *Physical Educator*, 62(2), 85-95. Retrieved from SPORTDiscus: 17500794
- Kelly, G. (1955). *The Psychology of Personal Construct: A Theory of Personality* (Vol. 1). New York: W.W. Norton & Company Inc.
- Klapp Lekholm, A., & 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. doi:10.1080/13803610801956663



- Klapp Lekholm, A., & Cliffordson, C. (2009). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation*, 15(1), 1-23.  
doi:10.1080/13803610802470425
- Klenowski, V., & Wyatt-Smith, C. (2012). The Impact of High Stakes Testing: The Australian Story. *Assessment In Education: Principles, Policy & Practice*, 19(1), 65-79.  
doi: 10.1080/0969594X.2011.592972
- Korp, H. (2006). *Lika chanser i gymnasiet?: en studie om betyg, nationella prov och social reproduktion.*[ Same chances in upper secondary school?: a study of grades, national tests and social reproduction]. Dissertation, University of Lund, Malmö, Sweden.  
Retrieved from <http://libris.kb.se/bib/10166306>
- Kuiper, W., & Berkvens, J. (Eds.). (2013). *Balancing curriculum regulation and freedom across Europe.* CIDREE Yearbook 2013. Enschede, the Netherlands: SLO.
- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: a review of international literature. *Sport, Education & Society*, 18(1), 57-76. doi:10.1080/13573322.2012.713860
- MacPhail, A. (2004). The Social Construction of Higher Grade Physical Education: The Impact on Teacher Curriculum Decision- making. *Sport, Education & Society*, 9(1), 53-73. doi:10.1080/1357332042000175818
- McMillan, J. H. (2003). Understanding and Improving Teachers' Classroom Assessment Decision Making: Implications for Theory and Practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43. doi:10.1111/j.1745-3992.2003.tb00142.x
- Penney, D. (2013). Points of tension and possibility: boundaries in and of physical education. *Sport, Education & Society* 18(1), 6-20. doi: 10.1080/13573322.2012.713862
- Penney, D., Brooker, R., Hay, P., & 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,  
doi:10.1080/13573320903217125
- Redelius, K., & Hay, P. (2009). Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review*, 15(3), 275-294. doi:10.1177/1356336X09364719
- Redelius, K., Fagrell, B., & 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. doi:10.1080/13573320902809195
- Resh, N. (2009.) Justice in grades allocation: teachers' perspective. *Social Psychology of Education*, 12(3), 315-325. doi:10.1007/s11218-008-9073-z
- Sun, Y., & Cheng, L. (2014). Teachers' Grading Practices: Meaning and Values Assigned. *Assessment In Education: Principles, Policy & Practice*, 21(3), 326-343. doi:10.1080/0969594X.2013.768207
- Swedish National Agency for Education (1994). *Läroplaner för det obligatoriska skolväsendet och de frivilliga skolformerna: Lpo 94: Lpf 94 [Curriculum for the compulsory school system, the pre-school class and the leisure-time centre Lpo 94: Lpf 94]*. Stockholm: Utbildningsdepartementet.
- Swedish National Agency for Education. (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.
- Swedish National Agency for Education. (2006). *Curriculum for the compulsory school system, the pre-school class and the leisure-time centre Lpo 94*. Stockholm: Author.
- Swedish National Agency for Education. (2011). *Curriculum for the compulsory school system, the pre-school class and the leisure-time centre 2011*. Stockholm: Author.

- Swedish National Audit Office. (2011). *Equal grades, equal ability? Follow-up on the Government's efforts to achieve equivalent grading in compulsory school* (Report 2011:23). Retrieved from <http://www.riksrevisionen.se/en/Start/Audit-reports/?year=2011>
- Svennberg, L., Meckbach, J., & Redelius, K. (2014). Exploring PE teachers' 'gut feelings': An attempt to verbalise and discuss teachers' internalised grading criteria. *European Physical Education Review*, 20(2), 199-214. doi: 10.1177/1356336X13517437
- 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]*. (dissertation, University College of Borås, Sweden). Retrieved from <http://libris.kb.se/bib/15874339>
- Thorburn, M. (2007). Achieving conceptual and curriculum coherence in high-stakes school examinations in Physical Education. *Physical Education & Sport Pedagogy*, 12(2), 163-184. doi:10.1080/17408980701282076
- Tierney, R. D., Simon, M., & Charland, J.(2011). Being Fair: Teachers' Interpretations of Principles for Standards-Based Grading. *Educational Forum*, 75(3), 210-227. doi: 10.1080/00131725.2011.577669
- von Greiff, C. (2009). *Lika skola med olika resurser? En ESO-rapport om likvärdighet och resursfördelning [Equal School with Unequal Resources? An ESO Report on Equality and Resource Distribution]*. Stockholm: Finansdepartementet, 2009.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The Centrality of Teachers' Judgement Practice in Assessment: A Study of Standards in Moderation. *Assessment In Education: Principles, Policy & Practice*, 17(1), 59-75. doi:10.1080/09695940903565610

Weiyun, C. (2005). Examination of curricula, teaching practices, and assessment through National Standards. *Physical Education & Sport Pedagogy*, 10(2), 159-180.  
doi:10.1080/17408980500105056

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. doi: 10.1080/08924562.2011.10590959