

Proposing conditions for assessment efficacy in physical education

- **Peter Hay** University of Queensland, Australia
- and
- **Dawn Penney** University of Tasmania, Australia

Abstract

In arguing for more comprehensive practice, policy and research considerations of assessment in physical education (PE), this paper outlines and discusses four integrated conditions of assessment efficacy for the development and promotion of productive assessment in PE. These conditions are prefaced by the proposition that quality PE requires the concerted and considered alignment of curriculum, pedagogy and assessment and the inclusion of a primary focus on assessment for learning; authentic, integrated assessment; assurance of construct validity; and socially just approaches to assessment. The conditions themselves are not new, however their integration has been a notable omission from PE literature in the past. Implicit in these conditions is a call for a broadening of the curriculum content of PE and an employment of physical activity as site for learning in multiple domains in addition to its recognition as a learning focus.

Key-words: assessment efficacy • domain content • integration

Introduction

In recent years the importance of assessment in and for physical education (PE) has been increasingly recognized within the international PE community (e.g. Hay, 2006; Hardman and Marshall, 2000; Redelius et al., 2009; Rink and Mitchell, 2002; Thorburn, 2007). While a steady increase in the volume of assessment research is commendable, a coherent theoretical rationale for assessment practice, policy and research in the field has been surprisingly elusive. Certainly there has been solid consensus regarding the necessity for assessment that is learning oriented (e.g. Richard and Godbout, 2000; Veal, 1992a, 1992b, 1995) and authentic in nature (e.g. Mohnsen, 1997; Oslin, 2003; Oslin et al., 1998), however these respective emphases and other considerations such as validity have rarely been discussed or researched in an integrated manner. With this in mind, we believe it is timely to articulate an *integrated* set of conditions for the pursuit of *assessment efficacy* in PE, and to prompt further theoretical, critical and practical discussions to this end.

Considerations of assessment efficacy in PE draw our attention to the desired outcomes and effects of assessment and the factors that contribute to these outcomes. In this paper we argue that assessment should be viewed as a process through which learning can be promoted, and that the satisfaction of this learning intent through the authentic, valid and socially just alignment of assessment, curriculum and pedagogy is a basis for claims of efficacy. The assertions of this paper are not ignorant of the inherent contentions associated with attempting to define or describe the basis and content of learning in PE. Nor do they ignore the significance of assessment as a mechanism for accountability (Green, 2001; Hardman and Marshall, 2000; Rink and Mitchell, 2002) or as a potentially problematic contributor to the neo-liberal management processes of contemporary western education systems (Ranson, 2003) where the assessment outcomes of school cohorts are used for making judgements and comparisons for the purpose of regulation, control, attrition and change within the system (Ball, 2003; Meadmore and Meadmore, 2004). In proposing conditions for assessment efficacy we hope, rather, to bring such learning contentions to the fore and challenge PE stakeholders to 'think both critically and imaginatively about the values and logical basis' of learning for which an account can be meaningfully and validly given through assessment (Thorburn, 2007: 271).

Inherent in this endeavour is a query over the appropriateness and breadth of past content and assessment foci in PE. For example, while we acknowledge the veracity of Rink and Mitchell's assertions in light of community health concerns over increasing rates of lifestyle diseases that PE is in an excellent position to 'obtain support for its programs' (2002: 254), we argue that the PE community's capacity to account for the impact of PE on health is quite limited. Moreover, such a curriculum focus on health intervention raises questions about the types and educational worth of assessments that might be employed to provide such an account. We also note that the curriculum accounted for by assessment in many iterations of PE has tended to be overly narrow, typically focusing on either the individual execution of skills (both processes and products) and/or the strategic awareness of students in performance/game contexts (e.g. Nadeau et al., 2008; Oslin et al., 1998). In our view such narrow foci potentially stifle the possibilities for more sophisticated learning in the subject. The work of Macdonald and Brooker (1997a, 1997b) (Queensland) and Thorburn and Collins (2006) (Scotland) are examples of aberrations to this trend in that the curricula represented by the assessments proposed have been much broader, requiring the application of content from several subdisciplines of human movement such as motor learning, sport and exercise psychology, sociology, to movement itself.

Conditions of assessment efficacy in PE

The promotion of assessment efficacy requires a clear understanding of the desired effects of assessment as well as the conditions necessary for optimizing their realization. Before attempting to articulate an integrated statement of these conditions it is necessary to make some definitional clarifications about assessment. Sadler (2005)

noted that discussions about assessment were notoriously hampered by semantic differences in the definition of terms and their theoretical and practical employment. In this paper we define assessment as the collection and interpretation of information about students' learning in PE. This information and its consequences vary in scope and depth depending on the process used to collect the information and the purpose for that collection (Hay, 2006). Our definition recognizes the multiple forms and foci of assessment, including such purposes and practices as formative and summative assessment. It also deliberately returns to the original conceptions of these terms as proposed by Bloom et al. (1971), with the emphasis therefore that the terms apply to the uses of the collected information rather than the tasks themselves. Hence, we are of the opinion that establishing assessment efficacy necessitates the transcending of divisions in purpose, and that meaningful and considered assessment can achieve multiple purposes, a proposition previously attested to by Gipps (1996) and reaffirmed by Lingard et al. (2006).

In proposing conditions of assessment efficacy in PE we contend that 'quality assessment' can only be understood and realized in relation to quality curriculum and pedagogy, and the clear and enacted alignment of these three message systems. Our references to curriculum, pedagogy and assessment as message systems are cognizant of Tinning's (2009) definitional concerns about curriculum and pedagogy and are thus drawn from Bernstein's (1971) quite explicit explanation of message systems as the means of selection, classification, transmission and evaluation of educational knowledge. We propose that quality assessment tasks should provide students with opportunities to demonstrate the valued learnings defined by the curriculum 'in' and 'about' movement, without compromising the beneficial affective effects realized 'through' movement. Furthermore, assessment should be supported by, and an informant to, pedagogies that provide students with the requisite skills and knowledges to complete the tasks (Lingard et al., 2006) in relation to these dimensions.

The conditions of assessment efficacy that we propose here are informed by this overarching condition of quality and, if enacted effectively, may contribute to the consolidation of quality curriculum and pedagogy in PE. These conditions include:

- a primary focus on assessment for learning;
- authentic assessment and thus, where possible, assessment that is 'integrated';
- assurance of validity;
- socially just approaches to assessment.

It is important to note that we do not view these elements as hierarchical, but rather equally and conjointly significant in the construction and enactment of assessment in PE. It is the *interdependence* of these elements, more so than their individual logic, that underpins the pursuit of assessment efficacy.

The conditions themselves are not new from either a broad education or PE-specific perspective. Their stipulation in relation to each other has, however, been a glaring omission from PE literature. In discussing each condition we draw, in part, upon the 'productive assessment' framework proposed by Hayes et al. (2006).

Although not specifically focused towards PE, the work of Hayes et al. serves as a useful reference point because of the concerted attention it affords learning in relation to assessment and because it, too, recognizes the importance of aligning curriculum, pedagogy and assessment in the pursuit of productive student outcomes. Furthermore, the notions of 'intellectual rigour', 'connectedness', 'supportive classroom environment', and 'working with and valuing differences' that feature in the productive assessment framework *stimulate* important questions and discussions about the constitution of PE and the potential substance of its educative claims.

Assessment for learning

'Assessment for learning' (Black and Wiliam, 1998), 'educative assessment', (Wiggins, 1998), and 'formative assessment' (Pryor and Crossouard, 2008) are terms that have variously been employed to distinguish assessment as having a learning focus. Although there are differences in the formality and extent of assessment defined by these terms, all reflect the fundamental interest in learning that in our view is a necessary focus for assessment in PE, including for tasks that may also generate information for more summative purposes such as grading and reporting (Gipps, 1996). In proposing this view of assessment and its relationship to purposes other than student learning we are not unaware of their influence in schools or their potential impact on students as mechanisms for sorting and selection (see e.g. Hay and Macdonald, 2008; Penney and Hay, 2008). Rather, it is because of the well reported problematic outcomes of privileging grading and reporting in the classroom context (e.g. Hay and Macdonald, 2008; Redelius et al., 2009; Tholin, 2006) that we are promoting an engagement with assessment where teachers make a deliberate and explicit effort to promote assessment for learning and teach their students how to 'read' assessment information in like manner. While a grading-free PE is perhaps desirable, it is highly unlikely for the systemic imperatives we have already described. We propose then that it is necessary to both explicitly privilege the learning focus of assessment and work to limit the potential problematic consequences of grading and reporting through the assurance of construct validity and an associated tangible commitment to socially just assessment. We will say more on these elements in later sections of the paper.

Assessment for learning is fundamentally grounded in constructivist theories of learning (Shepard, 2000). The constructivist theory recognizes that learning occurs as a result of interactions between learners and within contexts, and that students actively appropriate and adapt new knowledge in relation to former understandings and cognitive structures. Such an approach recognizes that learning is not a passive process of knowledge transmission. Rather, it is a complex process dependent upon students' previous knowledge, the mode of learning (e.g. kinaesthetic, visual, auditory), the context and the task. Assessment that has a learning focus provides information for teachers on the progress of their students' learning so that appropriate adjustments in curriculum and pedagogy can be made to optimize future learning. Moreover, the information generated by assessment should also be utilized by

students to monitor and adequately plan for their own learning, a process requiring students' access to the information and their understanding of what to do with the information (Sadler, 1998). For example, in the movement performance domain of PE, students should have ready access to the evidence of and associated feedback on their performances. Such evidence might include video excerpts of their performances and spoken or written feedback. Students should also be provided with performance exemplars (through criteria sheets or video excerpts) to provide a reference point for their own reflections and learning adaptations.

Assessment for learning is not merely concerned with the contribution of assessment to the learning process. It is also concerned with the *nature* of learning being promoted and achieved. Here we consider the 'productive assessment' element of 'intellectual rigour' (Hayes et al., 2006) to focus on *what learning* (content and process) could be undertaken in PE and addressed in assessment design and enactment. The pursuit of intellectual rigour in assessment involves the construction of assessment tasks and contexts that promote high-quality academic outcomes based on higher order thinking, the consideration of alternative solutions or knowledges, the construction of knowledge, the articulation of and engagement with discipline knowledge in a sufficient depth (in relation to content and process), and the communication of this knowledge in multiple modes, including the physical. In most other curriculum domains, the absence of intellectual rigour precipitates questions of quality and the subject's contribution to the education of children (Lingard et al., 2006). In our view, it is timely for the PE community to *consider* whether this educational construct is worthwhile pursuing, and if so, to then articulate how it may be developed in the curriculum, pedagogy and assessment of PE.

The promotion of intellectual rigour as we view it is not a synonym for the scientization of PE subject matter (Whitson and Macintosh, 1990), nor a justification for more performance oriented iterations of PE. Rather, it represents an interest in the potential learning demands that are made of the students irrespective of the source or nature of the content. A broadening of the domain specifications of PE beyond the performance of physical activities to include the study of biophysical, sociocultural and health oriented concepts would allow for the promotion of intellectual rigour in PE. Obviously this could be achieved through the direct study of these subdisciplines. However, we advocate for a more integrated approach in which the learning of the subdiscipline content occurs through its application in the movement context and to the movement context. In this way the importance of physical activity is not diminished, but rather viewed as a context for learning beyond the psychomotor domain. We find that Arnold's (1985) seminal work articulating dimensions of movement continues to be useful in situating physical activity as a site for learning and assessment, and promoting integrated thinking about content and contexts of learning in PE. Arnold's dimensions of movement prompt more concerted attention to the possible depth and quality of learning and assessment in PE.

Arnold articulated three dimensions of movement – 'in', 'through' and 'about'. He described the dimension of 'in' as the knowledges and skills acquired to participate in the specific activity context itself. The dimension 'about' encompasses 'the

rational study of movement' (1985: 52) in which students develop knowledge and understanding of movement concepts. The final dimension, 'through', represents the use of movement as a context to promote the aesthetic appreciation of performance, moral and social responsibility, and dispositions to physical activity. These three dimensions provide a rationale for multiple foci and modes of assessment, as reflected in, for example, the Queensland Senior PE syllabus, where teachers are required to provide learning experiences and assessment tasks 'that allow students not only to understand the relationships between physical activity and the complexity of factors underlying performance, but also to experience such relationships themselves (that is, the close *integration* of learning experiences in, about and through physical activity)' (Queensland Studies Authority, 2004: 2).

Within an integrated approach to assessment in PE, tasks would thus require students to develop and utilize their understanding of a particular concept such as biomechanics or health promotion to improve their own performance or participation (or that of another person) in a focus physical activity. Such integrated applications demand higher order cognitive processes such as application, analysis, synthesis and evaluation (Bloom, 1956) and knowledge utilization, metacognition and self-system thinking across learning domains (Marzano, 2001), thereby optimizing the cognitive demands and expectations of learning in PE. In the next section, focusing on the closely aligned notion of authenticity, we provide an example of assessment that encompasses a focus on integrated learning in authentic contexts.

Authentic and integrated assessment

Authentic assessment has been promoted for some time as means of countering the educational limitations of traditional tests. Other terms such as 'performance-based assessment' have been used, particularly in the USA (Lund and Tannehill, 2005), to describe more meaningful approaches to assessment. We prefer, however, the term 'authentic' because of its international history and development, and its conceptual breadth, allowing for more overt opportunities to develop integration possibilities in PE assessment. Authenticity in assessment is concerned with the relationships between learning content and contexts and their connection with the world beyond the classroom. Such tasks acknowledge that solutions to real-world challenges will rarely come from a single discipline. Yet the notion of authenticity also goes beyond a multidisciplinary approach. In summarizing the original and key elements of authentic assessment (or as it was originally referred to, 'authentic achievement'), Cumming and Maxwell (1999: 179) explained that each assessment 'should involve constructive learning, disciplined enquiry, and higher-order thinking and problem-solving. It should also have a value dimension, of aesthetic development, personal development or usefulness in the wider world.' Recognizing that student learning and quality performances depend, in part, upon students' motivation, and that this motivation can be affected by the assessment contexts provided (Cumming and Maxwell, 1999), authentic assessment pursues tasks and foci that are meaningful to

students and that have value and meaning beyond the instructional context. This pursuit, in addition to the expectations of students to 'communicate their knowledge, present a product or performance, or take some action for an audience beyond the teacher, classroom or school building' (Hayes et al., 2006: 98) are elements of authentic assessment that Hayes et al. (2006) have referred to as 'connectedness'.

The importance of authentic assessment has previously been advocated in PE (e.g. Melograno, 1994; Mohnsen, 1997; Smith, 1997; Smith and Cestaro, 1998) and has been justified in relation to the employment of contextual and games-based curriculum approaches to PE such as Teaching Games for Understanding (Oslin, 2003; Oslin et al., 1998) and sport education (Siedentop et al., 2004). Veal (1992a) asserted that the 'role of authentic assessment in creating *real physical education* is to help students learn by providing a formal feedback loop that results in instructional strategies and achievement of skill and attitudinal goals' (p. 89; our emphasis). Although acknowledging the connection to student learning, Veal's definition is somewhat limited in its lack of reference to 'connectedness' consistently attended to in other definitions. Other descriptions of authentic assessment in PE have been similarly lacking in explicit connections with contexts and audiences beyond the movement context. Sport education provides, to some extent, an exception to this observation, given the explicit interest in the successful completion of tasks 'in a context that is relevant to how the task is done in the larger world' (Siedentop et al., 2004: 118). Nevertheless, we suggest that other more broad and sophisticated learning outcomes are possible.

Consistent with the broadening of the PE curriculum domain that we proposed in relation to assessment for learning, we agree with Thorburn (2007: 271) of the 'need to ensure that the learning aims associated with integrated learning are authentic rather than contrived and do genuinely contribute towards the achievement of high levels of attainment'. Furthermore, the definition of authentic experiences in PE should capture the possibilities for integrated learning. Such a definition was offered by Hay (2006: 317) who proposed that:

... authentic assessment in PE should be based in movement and capture the cognitive and psychomotor processes involved in the competent performance of physical activities. Furthermore, assessment should redress the mind/body dualism propagated by traditional approaches to assessment, curriculum and pedagogies in PE, through tasks that acknowledge and bring to the fore the interrelatedness of knowledge, process (cognitive and motor), skills and the affective domain.

This perspective aligns with our claims regarding the potential broadening of learning foci through assessment in PE, and the need for assessment to promote learning that is intellectually rigorous, and integrated. Yet Hay's proposition also falls short of mainstream understandings of authenticity because of its lack of reference to the connectedness of the task with potential experiences and audiences beyond the school context. To this end, we propose that authentic assessment in PE should be

based in movement and include the integration of movement-associated concepts in and to movement contexts in modes reflective of their appropriation beyond the classroom. Integrated tasks, for example, may be framed in relation to coaching, sports management, or public health interventions. Clearly the possible integrations are numerous and could serve multiple agendas while maintaining an initiative towards the educative rigour of the subject.

Senior PE, a 'high-stakes' subject offered in the Australian state of Queensland, is an example of an iteration of PE that has pursued a broader perspective on the nature and educative substance of PE (Macdonald and Brooker, 1997a, 1997b). This elective subject has been successfully delivered in Queensland schools since 1998 such that Senior PE has the fourth highest student enrolment of any subject in Queensland. Senior PE closely reflects the learning and authenticity conditions proposed in this section (Macdonald and Brooker, 1997a), and provides an example of the way in which movement concepts (such as motor control, exercise physiology or sport sociology, etc.) can be meaningfully engaged with 'in' and 'about' physical activity. A unit, for example, could involve the study of the psychology of golf (with psychology serving as the concept focus and golf the physical activity focus). The assessment within this unit would require the teacher's judgements of the students' capacities to acquire, apply and evaluate skills and strategies in golf. The second assessment element of the unit may involve an assignment in which students investigate the effects of arousal on their own golf performance, including the application and evaluation of their use of strategies to optimize their arousal levels and thus performance in golf. In this task, students learn the movement concepts in the authentic context of golf, satisfying the learning and authenticity conditions we have proposed. Although the high-stakes situation of Senior PE demands levels of content sophistication that would be above and beyond PE in earlier phases of learning, it demonstrates that integration is possible and that other more simple concepts such as 'simple physiological responses to exercise' or 'promoting positive relationships' could be conceivably engaged with in physical activities by students in earlier phases of learning.

Valid assessment

Validity could well serve as an overarching condition of assessment in PE in the same manner as the necessity for a concerted alignment of the three message systems of education. That is, unless the proposed assessment is valid (and by implication, reliable) its usefulness as a means of information collection and as an informant to both ongoing learning and reporting is questionable. Nevertheless, we address this efficacy condition at this point, recognizing that the substantiation of validity requires the definition of the domain construct (learning content and processes of a curriculum), and is necessarily instantiated in the construction and enactment of particular tasks.

In our opinion, validity has not been sufficiently attended to in PE assessment. In part this is due to historically shallow or menial assessment and reporting foci

(Siedentop et al., 2004). It is also the consequence of narrow domain definitions and thus validity applications. For example, while validity has been the focus of some research endeavours (Nadeau et al., 2008; Oslin et al., 1998), the focus is generally either on score validation rather than more encompassing and integrated notions of validity (Messick, 1995), or addresses the validity of assessment in relation to a narrowly defined domain construct. In fact, the question of what might *constitute* construct validity in PE assessment has been broached in but very few substantial research pieces (see e.g. Hay and Macdonald, 2008).

The international change in assessment focus from traditional style tests to authentic tasks, which we advocate for in this paper, has aligned with a change in conceptions of validity from measurement-oriented or psychometric approaches based on score theory to an emphasis on 'the appropriateness of assessment tasks as indicators of intended learning outcomes, and on the appropriateness of the interpretation of assessment outcomes as indicators of learning' (Cumming and Maxwell, 1999: 177). Furthermore, some have described authentic assessment, itself, as an intention towards broader notions of 'ecological validity' (Kirk and O'Flaherty, 2004; Wiggins, 1998). In a statement that encompassed traditional psychometric notions of validity as well as these alternative perspectives, Messick explained that 'validity is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of *inferences* and *actions* based on test scores or other modes of assessment' (1989: 13; our emphasis). In other words, what need to be valid are the information collected, the interpretation of the scores, and the implications and consequences for action that such meaning entails. Validity, therefore, is a principle relevant to *any* process or practice of observing or documenting behaviours or attributes (Kane, 2001) and is thus an important consideration for assessment in PE.

Messick (1995) argued that the validation of *any* assessment task should incorporate six aspects of construct validity: content, substantive, structural, generalizability, external, and consequential. These elements refer to complementary forms of evidence that are integrated to produce an overall judgement of construct validity. In relation to authentic or performance-oriented (Messick, 1994) PE assessment we summarize that, when planning for assessment, attention be given to:

- the content relevance and representativeness of the planned tasks to the PE domain content (*content*);
- a clear alignment of student response processes and content and the processes and content characteristic of the subject domain (*substantive*);
- the development of scoring or grading approaches (e.g. criteria and standards) that are reflective of task and domain structure (*structural*);
- the recognition of and response to factors such as time, context and assessors that may affect the *generalizability* of the information and particularly the reliability of assessment evidence and judgements;

- convergent and discriminate correlations of the assessment scores with other sources of capability evidence and the utility of the scores for applied purposes such as learning and/or accountability (*external*); and
- an investigation of the positive and negative, intended and unintended consequences of the task and its outcomes for PE students (*consequential*).

In attending to Messick's (1995) six elements, we reassert that a careful consideration and then articulation of the domain boundaries and curriculum detail of PE, including the attributed knowledge, skills, processes and integrations, should underpin considerations of construct validity in PE assessment. These domain specifications serve as a guide for the content of the tasks as well as the types of tasks or contexts that should be included in an assessment programme. Accordingly, PE teachers developing assessment tasks and defining assessment contexts need to ensure that the tasks and contexts, over a specified period, are *representative of* and demand the *demonstration by* the students of these knowledges, skills, processes and integrations. Furthermore, the tasks should generate defensible evidence on which to base judgements of performance, and teachers should ensure that this evidence is readily accessible and understandable for students. We believe that teachers should consider the permissibility of performance evidence generated from beyond a specified task (such as coaching, training or performing in community settings for example) to corroborate the evidence collected through the task. Attention should also be given to the processes by which performance evidence is referenced against criteria in the determination of students' grades to ensure that they are reliable, reflective of the task demands, and free from construct irrelevant factors such as students' dispositional and behavioural characteristics.

Socially just approaches to assessment

This fourth condition of assessment efficacy is concerned with the opportunities *all* students are given to engage in assessment, receive attention and recognition for demonstrations of performance, and learn as a consequence of their engagement in assessment. This could be best summarized as an intention towards 'inclusion' (Forlin, 2004) and, more broadly, 'social justice' (Lingard, 2005). While Lingard (2005) asserts that the provision of quality assessment opportunities supported by quality curriculum and pedagogy is in and of itself a social justice issue, Hayes et al. (2006) suggest that the pursuit of socially just assessment requires a supportive classroom environment and the working with and valuing of student differences. In planning assessment for a unit of work in PE, teachers should be considerate of the learning needs of their students and the importance of providing multiple response opportunities through a variety of assessment modes.

We argue that an inclusive, socially just approach to assessment in PE requires that all students are aware of how to engage in an assessment context or with an assessment task at the level valued by the assessors. This necessitates the provision of

adequate task scaffolding or instructional cues (Hayes et al., 2006) through communicative modes that are not socially or culturally exclusive. Moreover, students need to be let in on the secret of teachers' grading decisions through access to explicit and understandable criteria for the judging of quality performances (Hay and Macdonald, 2008; Sadler, 1998). Beyond social justice, these initiatives reflect a primary focus on assessment for learning rather than accountability-associated sorting and selecting purposes, and a shift away from what Bernstein (1996) described as performance modes – approaches to the education of students that are predicated on 'different from' relations (what sets students apart) and the establishment of hierarchical student distinctions. This does not mean that the generation of indicators of achievement quality (i.e. grades) is foregone, but rather that the teachers' interest is first on the learning and achievement of all students rather than the qualitative or quantitative distinctions between them.

In addition to supportive conditions for discrete tasks, inclusion and social justice in assessment requires that students be given multiple opportunities, in varying contexts, to display and receive recognition for the valued attributes of the domain construct. That is, decisions concerning students' learning in a unit of work should be based on the collection of information over a period of time and through multiple modes of assessment rather than as discrete or point-in-time exercises. Furthermore, students should have ready access to this evidence to increase their engagement in, and accountability for, the learning process. Notably, the provision of these opportunities is not only a social justice issue, it is an important factor in the negotiation of construct-irrelevant variance in students' assessment responses (William and Black, 1996). 'Point in time' and narrowly defined assessment moments do not adequately account for the variety of properties to a response at a particular point in time. These properties may include language (and other cultural codes), the context of the required response, the student's emotional and physical state at the time of assessment, etc. William and Black explain that the aim of ongoing and multiple modes of assessment is to 'average out all those effects which are not desired and so enhance the signal to noise ratio for the effect [or learning attribute] of interest' (1996: 539).

Messick's (1995) final element of construct validity (consequences as validity evidence) draws attention to the importance of taking into account the *effect* of assessment and its outcomes on students, and is thus also a consideration in relation to the promotion of social justice and, ultimately, efficacy in assessment. For example, if the intention of the task is for learning, it is necessary to establish what intended and unintended learning has occurred (Messick, 1995) and the aspects of the assessment approach employed that may have led to any unintended learnings. The key interest here is the negative consequences of assessment such as a student's diminished sense of capacity in the PE field and their disconnect with physical culture in and beyond the classroom, as well as the learning of undesirable ideologies such as elitism, sexism, racism, etc. (Tinning, 1990). If assessment in PE contributes to negative and unintended learnings such as these, then its construction and employment in the subject must be rigorously reviewed. All too often in mainstream education differences in

students' achievements are attributed to the meritocratic claims of student ability and effort rather than problems with the measures used to generate the achievement (Benjamin, 2003; Hart, 1998). We believe that the PE community should be wary of this tendency and give due care and attention to socially just assessment practices.

Conclusion

The purpose of this paper has been to challenge the international PE community to comprehensively consider assessment practices in PE. In particular we have drawn on theoretical and research-based principles of assessment to propose four interdependent conditions of assessment efficacy that reside within a concerted alignment of curriculum, pedagogy and assessment. Underpinning our assertions has been an argument for a broadening of the potential domain characteristics of PE to allow the subject to better optimize its educative substance as well as the unique learning context that it can provide. Central to the integrity of our efficacy proposal is the co-dependence of the four conditions. It is not sufficient, for example, to enrich and broaden the learning foci of PE assessment tasks if the tasks or method of generating grades are invalid. Moreover, if assessment outcomes are consumed by students in such a way as to highlight their limitations rather than promote their future learning and engagement in a movement culture, irrespective of the authenticity of the assessment, its effects are undesirable. Beyond the policy and practice implications of these co-dependent conditions, we suggest that research into assessment practices in PE should likewise assume a more integrated approach.

Admittedly, the conditions are ambitious, particularly given a history of at times trivial assessment in the subject (Siedentop et al., 2004) but are nevertheless possible. International systemic variation in the constitution of PE domains means that the provision of assessment exemplars is difficult. We have provided some examples of assessment that meet the proposed conditions from the Queensland context and encourage international readers to consider the logic and applicability of the conditions to their own curricular contexts. In so doing a concerted and directed dialogue can be established to consolidate and support effective practices and discuss possibilities for assessment reform where needed. Thus, we welcome further discussion, debate and critique to sharpen the understanding and practice of assessment in the PE community, and to optimize the educational rationale for the subject. Such discussions need to be underpinned by targeted, concerted and cohesive research approaches to assessment in PE and a broadening of our perspectives on the educative possibilities for PE.

Acknowledgement

We would like to thank the reviewers of this paper for their helpful reflections and suggestions for strengthening the paper, and for engaging with our work in a spirit of collegial discussion, debate and critique.

References

- Arnold, P. (1985) 'Movement, Physical Education and the Curriculum', *Bulletin of Physical Education* 16(1): 5–9.
- Ball, S. (2003) 'The Teacher's Soul and the Terrors of Performativity', *Journal of Education Policy* 18(2): 215–28.
- Benjamin, S. (2003) 'What Counts as "Success"? Hierarchical Discourses in a Girls' Comprehensive School', *Discourse: Studies in the Cultural Politics of Education* 24(1): 105–18.
- Bernstein, B. (1996) *Pedagogy, Symbolic Control and Identity: Theory, Research and Critique*. London: Taylor Francis.
- Bernstein, B. (1971) *Class Codes and Control*. London: Routledge.
- Black, P. and Wiliam, D. (1998) 'Assessment and Classroom Learning', *Assessment in Education* 5(1): 7–73.
- Bloom, B. (1956) *The Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York: David Mckay Co.
- Bloom, B. S., Hastings, J. T. and Madaus, G. F. (1971) *Handbook on the Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Cumming, J. J. and Maxwell, G. S. (1999) 'Contextualising Authentic Assessment', *Assessment in Education* 6(2): 177–94.
- Forlin, C. (2004) 'Promoting Inclusivity in Western Australian Schools', *International Journal of Inclusive Education* 8(2): 185–202.
- Gipps, C. (1996) 'Assessment for Learning', in A. Little and A. Wolf (eds) *Assessment in Transition: Learning, Monitoring and Selection in International Perspective*, pp. 251–62. Oxford: Elsevier Science.
- Green, K. (2001) 'Examinations in Physical Education: A Sociological Perspective on a New "Orthodoxy"', *British Journal of Sociology of Education* 22(1): 51–73.
- Hardman, K. and Marshall, J. J. (2000) *World-Wide Survey of the State and Status of School Physical Education*. Manchester: University of Manchester.
- Hart, S. (1998) 'A Sorry Tail: Ability, Pedagogy and Educational Reform', *British Journal of Educational Studies* 46(2): 153–68.
- Hay, P. J. (2006) 'Assessment for Learning in Physical Education', in D. Kirk, D. Macdonald and M. O'Sullivan (eds) *The Handbook of Physical Education*, pp. 312–25. London: SAGE.
- Hay, P. J. and Macdonald, D. (2008) '[Mis]appropriations of Criteria and Standards Referenced Assessment in a Performance-Based Subject', *Assessment in Education* 15(2): 153–68.
- Hayes, D., Mills, M. Christie, P. and Lingard, B. (2006) *Teachers and Schooling Making a Difference*. Sydney: Allen & Unwin.
- Kane, M. (2001) 'Current Concerns in Validity Theory', *Journal of Educational Measurement* 38(4): 319–42.
- Kirk, D. and O'Flaherty, M. (2004) 'Learning Theory and Authentic Assessment in Physical Education', paper presented at the Annual Conference of BERA, Edinburgh, Scotland, 10–13 Sept.
- Lingard, B. (2005) 'Socially Just Pedagogies in Changing Times', *International Studies in Sociology of Education* 15(2): 165–86.
- Lingard, B., Mills, M. and Hayes, D. (2006) 'Enabling and Aligning Assessment for Learning: Some Research and Policy Lessons from Queensland', *International Studies in Sociology of Education* 16(2): 83–103.
- Lund, J. and Tannehill, D. (2005) 'Assessment in Curriculum Development', in J. Lund and D. Tannehill (eds) *Standards-Based Physical Education Curriculum Development*, pp. 46–76. Boston, MA: Jones & Bartlett Publishers.
- Macdonald, D. and Brooker R. (1997a) 'Assessment Issues in a Performance-Based Subject: A Case Study of Physical Education', *Studies in Educational Evaluation* 23(1): 83–102.

- Macdonald, D. and Brooker, R. (1997b) 'Moving Beyond the Crises in secondary physical education: an Australian initiative', *Journal of Teaching in Physical Education*, 16(2): 155–175.
- Marzano, R. J. (2001) *Designing a New Taxonomy of Educational Objectives*. Thousand Oaks: Corwin Press.
- Meadmore, D. and Meadmore, P. (2004) 'The boundlessness of performativity in elite Australian schools', *Discourse: Studies in the Cultural Politics of Education*, 25(3): 375–387.
- Melograno, V. J. (1994) 'Portfolio assessment: documenting authentic student learning', *Journal of Physical Education, Recreation and Dance*, 65(8): 50–61.
- Messick, S. (1989) 'Validity', in R. L. Linn (ed) *Educational Measurement* pp. 13–103. Washington DC: American Council on Education and National Council on Measurement in Education.
- Messick, S. (1994) 'The interplay of evidence and consequences in the validation of performance assessments', *Educational Researcher*, 23(2): 13–23.
- Messick, S. (1995) 'Standards of validity and the validity of standards in Performance Assessment', *Educational Measurement: Issues and Practices*, 14(4): 5–8.
- Mohsen, B. (1997) 'Authentic Assessment in Physical Education', *Learning and Leading with Technology* 24(7): 30–3.
- Nadeau, L., Richard, J.-F. and Godbout, P. (2008) 'The Validity and Reliability of a Performance Assessment Procedure in Ice Hockey', *Physical Education and Sport Pedagogy* 13(1): 65–83.
- Oslin, J. L. (2003) 'The Role of Assessment in Teaching Games for Understanding', in L. L. Griffin and J. I. Butler (eds) *Teaching Games for Understanding: Theory, Research, and Practice*, pp. 125–36. Champaign, IL: Human Kinetics.
- Oslin, J. L., Mitchell, S. A. and Griffin, L. L. (1998) 'The Game Performance Assessment Inventory (GPAI): Development and Preliminary Validation', *Journal of Teaching in Physical Education* 17(2): 231–43.
- Penney, D. and Hay, P. J. (2008) 'Inclusivity and Senior Physical Education: Insights from Queensland and Western Australia', *Sport, Education and Society* 13(4): 431–52.
- Pryor, J. and Crossouard, B. (2008) 'A Socio-Cultural Theorisation of Formative Assessment', *Oxford Review of Education* 34(1): 1–20.
- Queensland Studies Authority (2004) *Senior Syllabus in Physical Education*. Brisbane: Queensland Studies Authority.
- Ranson, S. (2003) 'Public Accountability in the Age of Neo-Liberal Governance', *Journal of Education Policy* 18(5): 459–80.
- Redelius, K., Fagrell, B. and Larsson, H. (2009) 'Symbolic Capital in Physical Education: To Be, to Do or to Know? That is the Gendered Question', *Sport, Education and Society* 14(2): 245–60.
- Richard, J.-F. and Godbout, P. (2000) 'Formative Assessment as an Integral Part of the Teaching-Learning Process', *Physical and Health Education Journal* 66(3): 4–10.
- Rink, J. and Mitchell, M. (2002) 'High Stakes Assessment: A Journey into Unknown Territory', *Quest* 54(3): 205–23.
- Sadler, D. R. (1998) 'Formative Assessment: Revisiting the Territory', *Assessment in Education* 5(1): 77–84.
- Sadler, D. R. (2005) 'Interpretations of Criteria-Based Assessment and Grading in Higher Education', *Assessment and Evaluation in Higher Education* 30(2): 175–94.
- Shepard, L. (2000) 'The Role of Assessment in a Learning Culture', *Educational Researcher* 29(7): 4–14.
- Siedentop, D., Hastie, P. A. and van der Mars, H. (2004) *Complete Guide to Sport Education*. Champaign, IL: Human Kinetics.

- Smith, T. (1997) 'Authentic Assessment: Using a Portfolio Card in Physical Education', *Journal of Physical Education, Recreation and Dance* 68(4): 46–52.
- Smith, T. K. and Cestaro, N. G. (1998) *Student-Centered Physical Education: Strategies for Developing Middle School Fitness and Skills*. Champaign, IL: Human Kinetics.
- Tholin, J. (2006) *Att kunna klara sig i okänd natur: En studie av betyg och betygskriterier* (To Find the Way in Unknown Areas: A Study of Grades and Grading Criteria). Borås: Högskolan i Borås.
- Thorburn, M. (2007) 'Achieving Conceptual and Curriculum Coherence in High-Stakes School Examinations in Physical Education', *Physical Education and Sport Pedagogy* 12(2): 163–84.
- Thorburn, M. and Collins, D. (2006) 'The Effects of an Integrated Curriculum Model on Student Learning and Attainment', *European Physical Education Review* 12(1): 31–50.
- Tinning, R. (1990) *Ideology and Physical Education: Opening Pandora's Box*. Geelong: Deakin University Press.
- Tinning, R. (2009) 'Pedagogy, Sport Pedagogy, and the Field of Kinesiology', *Quest* 60(3): 405–24.
- Veal, M.-L. (1992a) 'The Role of Assessment in Secondary Physical Education: A Pedagogical View', *Journal of Physical Education, Recreation and Dance* 63(7): 88–92.
- Veal, M.-L. (1992b) 'School-Based Theories of Pupil Assessment: A Case Study', *Research Quarterly for Exercise and Sport* 63(1): 48–59.
- Veal, M.-L. (1995) 'Assessment as an Instructional Tool', *Strategies* 8(6): 10–15.
- Whitson, D. and Macintosh, D. (1990) 'The Scientization of Physical Education: Discourses of Performance', *Quest* 42(1): 40–51.
- Wiggins, G. (1998) *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco, CA: Jossey-Bass Inc.
- William, D. and Black, P. (1996) 'Meanings and Consequences: A Basis for Distinguishing Formative and Summative Functions of Assessment?', *British Educational Research Journal* 22(5): 537–48.

Résumé

Les conditions de l'efficacité de l'évaluation en éducation physique

Contribuant aux débats pour une meilleure compréhension des pratiques, des politiques et des recherches sur l'évaluation en Education Physique (EP), cet article décrit et discute quatre conditions d'efficacité de l'évaluation pour le développement et la promotion d'une évaluation formative en EP. Ces conditions sont sous-tendues à la proposition selon laquelle une EP de qualité tient à une conception concertée des programmes, des pratiques pédagogiques et de l'évaluation, et à l'intégration de l'évaluation parmi les principaux facteurs de l'apprentissage; à l'assurance de la validité de son élaboration; et à une approche socialement juste de l'évaluation. Ces conditions elles-mêmes ne sont pas nouvelles, cependant leur prise en compte a été, par le passé, largement occultée par la littérature spécifique à l'EP. De manière sous-jacente à ces propositions, nous lançons un appel à une meilleure prise en compte de l'évaluation dans la conception des programmes d'EP et à une conception de l'EP comme un lieu au service d'apprentissages dans des domaines multiples au-delà de sa reconnaissance comme un objet même d'apprentissage.

Resumen

Propuesta de condiciones para la eficacia de la evaluación en educación física

Al abogar por una práctica, una política y unas consideraciones de la evaluación de la investigación en educación física (PE) más amplias, este documento presenta y discute cuatro condiciones integradas de eficacia de la evaluación para el desarrollo y la promoción de una evaluación productiva en E.F. Estas condiciones están precedidas por la premisa de que la educación física de calidad requiere de la confluencia acordada y consensuada del currículo, la pedagogía y la evaluación, y la inclusión de un foco primario en la evaluación para el aprendizaje; una evaluación auténtica e integrada, la garantía de la validez de constructo; y socialmente meras aproximaciones a la evaluación. Las condiciones de por sí no son nuevas, sin embargo su integración ha sufrido una omisión notable de la literatura sobre Educación Física en el pasado. De manera implícita, con estas condiciones se reclama una ampliación del contenido del currículo de educación física y un empleo de la actividad física como contexto para el aprendizaje de varios dominios, además de su reconocimiento como un foco de aprendizaje.

Zusammenfassung

Vorschläge für wirksame Beurteilungsbedingungen im Sportunterricht

Der vorliegende Artikel veranschaulicht und diskutiert vier integrative Voraussetzungen für die Wirksamkeit von Beurteilung im Zusammenhang mit der Entwicklung und Förderung produktiver Bewertung im Sportunterricht. Dabei wird für nachvollziehbarere Überlegungen für eine Praxis, Politik und Forschung von Beurteilung im Sportunterricht argumentiert. Die Voraussetzungen werden eingeführt mit der Absicht das qualitativ hochwertiger Sportunterricht einen ausgewogenen und wohlüberlegten Abgleich von Curriculum, Pädagogik und Beurteilung braucht und folgende Aspekte integrieren sollte: die Berücksichtigung eines primären Augenmerks auf die Bewertung von Lernfortschritten, authentische und integrative Beurteilung, Sicherheit in der Aussagekraft der Bewertungskriterien, sowie sozial gerechte Bewertungsansätze. Diese Bedingungen ansich sind nicht neu, allerdings wurden sie in der Vergangenheit in Publikationen zum Sportunterricht deutlich vernachlässigt. Implizit liegt diesen Bedingungen der Aufruf zugrunde, die Curriculumsinhalte von Sportunterricht zu erweitern und Sport und Bewegung als einen Ort mit multiplen Lernbereichen in Ergänzung zu seiner Wahrnehmung als eigenes Lernfeld anzuerkennen.

Peter Hay is a Lecturer in the School of Human Movement Studies at the University of Queensland. His research focuses on assessment policy and practice in Health and Physical Education and the construction of students' abilities in learning contexts.

Dawn Penney is Associate Professor and Associate Dean (Research) in the Faculty of Education, University of Tasmania. Dawn has researched and published prolifically in the fields of physical education policy and curriculum.

Address for correspondence: Dr Peter Hay, School of Human Movement Studies, University of Queensland, St Lucia, Australia, 4072. [email: phay@hms.uq.edu.au]
