

1-1-2022

Enhancing quality and equity? Performance assessment validation in examination physical education in Western Australia

Dawn Penney
Edith Cowan University, d.penney@ecu.edu.au

Eibhlish O'Hara
Edith Cowan University, e.ohara@ecu.edu.au

Rob Lund
Edith Cowan University, r.lund@ecu.edu.au

Follow this and additional works at: <https://ro.ecu.edu.au/ecuworks2022-2026>



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

[10.1080/25742981.2022.2136007](https://doi.org/10.1080/25742981.2022.2136007)

Penney, D., O'Hara, E., & Lund, R. (2022). Enhancing quality and equity? Performance assessment validation in examination physical education in Western Australia. *Curriculum Studies in Health and Physical Education*.

Advance online publication. <https://doi.org/10.1080/25742981.2022.2136007>

This Journal Article is posted at Research Online.

<https://ro.ecu.edu.au/ecuworks2022-2026/1439>



Enhancing quality and equity? Performance assessment validation in examination physical education in Western Australia

Dawn Penney, Eibhlish O'Hara & Rob Lund

To cite this article: Dawn Penney, Eibhlish O'Hara & Rob Lund (2022): Enhancing quality and equity? Performance assessment validation in examination physical education in Western Australia, Curriculum Studies in Health and Physical Education, DOI: [10.1080/25742981.2022.2136007](https://doi.org/10.1080/25742981.2022.2136007)

To link to this article: <https://doi.org/10.1080/25742981.2022.2136007>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 21 Oct 2022.



Submit your article to this journal [↗](#)



Article views: 242



View related articles [↗](#)



View Crossmark data [↗](#)

Enhancing quality and equity? Performance assessment validation in examination physical education in Western Australia

Dawn Penney , Eibhlish O'Hara  and Rob Lund 

School of Education, Edith Cowan University, Australia

ABSTRACT



The positioning and format of performance assessment in examination physical education varies between courses across Australia and internationally. This paper centres on developments in performance assessment in the Physical Education Studies (PES) course in Western Australia (WA). In 2021 The School Curriculum and Standards Authority (SCSA) undertook an assessment validation trial of school-based assessment of students participating in modified format competitive game play in the 10 PES sports. This contrasted to existing centralised examination arrangements. The paper reports on findings from observation of the trial in nine of the 10 sport contexts, and semi-structured interviews with teachers, validators and SCSA staff. Analysis drew on conceptualisations of quality assessment to critically examine features of assessment information collection and judgement processes in the trial and the inter-relationships between these two elements of assessment. Discussion highlights issues of quality and equity in performance assessment for future policy and research to consider.

KEYWORDS

Performance assessment;
senior secondary physical
education; equity;
assessment policy

Introduction

In 1997 Macdonald and Brooker identified assessment as central to ongoing debates about 'the educative worth of performance-oriented subjects in schools' (p. 83), including physical education. Their investigation of the early development of the Board of Senior Secondary School Studies (BSSSS) physical education syllabus in Queensland highlighted the need for further research addressing issues of quality and equity in performance assessment in physical education. Since then, studies in Australia and internationally have affirmed that the form that performance assessment should take in examination and/or senior secondary physical education courses, and what standing it should be accorded in course specifications and grade calculations, remain matters of debate and considerable variation between courses (Brown & Penney, 2018; Scanlon et al., 2019;

CONTACT Dawn Penney  d.penney@ecu.edu.au  School of Education, Edith Cowan University, 2 Bradford Street, Mount Lawley, Western Australia 6050

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Whittle et al., 2017). In short, performance assessment is a contentious aspect of examination and senior secondary physical education.

To clarify terminology, reference to ‘examination and senior secondary physical education’ at many points in this paper reflects that internationally, *examination physical education courses* (i.e. studies in physical education that are linked to certification and involve an external examination component) exist at secondary and senior secondary levels of education. While the nomenclature and the age associated with years of schooling may vary across jurisdictions, here *secondary* education refers to years 7–10 in schooling (for ages 12/13–15/16 years) and *senior secondary* refers to years 11 and 12 (for ages 16/17–17/18). *Senior secondary physical education courses* may or may not be examination level courses, with several jurisdictions offering examination and non-examination pathways. *Performance assessment* relates specifically to assessment of ‘students’ performance of a physical activity or movement skill’ (Whittle et al., 2017, p. 610) in these courses. Research affirms that performance assessment is often an element of the internal (school-based) assessment rather than external assessment (Brown & Penney, 2018; Whittle et al., 2017), with the latter typically deemed to be ‘high stakes’.

This paper reflects that how performance assessment is positioned and what form it takes in courses, are matters that will be highly influential in shaping teachers’ enactment of course specifications (stipulating required content and assessment elements for a course) and hence, students’ learning experiences and opportunities in secondary and senior secondary physical education. From this perspective, assessment specifications in particular have a pivotal role in shaping curriculum enactment and pedagogy in schools and as such, are worthy of research attention. Mandated high-stakes assessment – and changes to this – are recognised as key drivers and/or impediments in curriculum reform (Barnes et al., 2000), opening up or closing down the pedagogic possibilities that teachers will recognise and explore in the units of work, lessons, and assessment tasks that constitute students’ experience of examination or senior secondary physical education courses.

In focusing on developments in senior secondary physical education in Western Australia (WA), the paper directs attention to a jurisdiction that in 2008 established performance assessment as an examination assessment component in a new Physical Education Studies (PES) course that could count towards students’ tertiary entrance score (Jones & Penney, 2019; Paveling et al., 2019; Penney et al., 2012). As explained below, this development privileged performance assessment in a way that was unparalleled in Australia. The research centres on the trial of notable changes to the performance assessment in PES in WA and was commissioned by the School Curriculum and Standards Authority (SCSA), the organisation responsible for curriculum, assessment and reporting for all schools in WA. The paper draws on research-based conceptualisations of quality assessment (Hay & Penney, 2009; 2013) to critically examine the trial and the research findings arising from it. Implications for assessment policy and research relating to senior secondary and examination physical education internationally, are discussed.

Performance assessment in examination and senior secondary physical education: national and international insights

Over several decades research has identified high-stakes assessment requirements as a pivotal issue in conceptual and pedagogical tensions arising in examination and senior

secondary physical education course developments (Bowes, 2010; Brown & Penney, 2017; Jones & Penney, 2019; Thorburn, 2007, 2008). Past research affirms that attempts to reach decisions about assessment specifications and more specifically, the way in which performance assessment will feature (if at all), are likely to be challenging, as those charged with curriculum development seek to address concerns for authenticity in physical education assessment (Hay & Penney, 2009, 2013; Thorburn, 2008), respond to demands for rigour in assessment to meet the expectations of high-stakes environments (Brown & Penney, 2018; Hay & Macdonald, 2008; Penney et al., 2012), and remain cognisant of the pedagogical significance of assessment decisions.

Bowes (2010) has previously noted that in Aotearoa New Zealand ‘practical learning as assessment continues to be marginalised in SSPE [Senior Secondary Physical Education] and does not legitimize physical education as a senior subject in the same way theory does’ (p. 23). In Australia, Brown and Penney (2017, 2018) echoed international calls for the conceptual coherence between curriculum texts and assessment frameworks to be strengthened (MacPhail, 2007; Thorburn, 2007) and highlighted the pedagogical impact of assessment arrangements (associated with the Victorian Certificate of Education Physical Education (VCEPE) particularly) that marginalise embodied learning in physical education. Brown and Penney (2017) identified the written (only) examination in the VCEPE, as one of several inter-related factors inhibiting the exploration of ‘original and creative’ pedagogy in teachers’ enactment of course specifications (p. 134).

In the recent development of specifications for the Leaving Certificate Physical Education (LCPE) in Ireland, Scanlon et al. (2019) identified assessment as a highly contentious issue, with those involved recognising the distance between ‘what was desirable [i.e. more weighting on the practical aspect] and what is acceptable [i.e. more weighting on the theoretical aspect to keep in line with other Leaving Certificate subjects]’ (Scanlon et al., 2019, p. 82). Scanlon et al. (2019) explained that ultimately, the use of digital technologies was pivotal in enabling 50% of the examination marks (relating to a physical activity project (20%) and performance assessment (30%)), ‘to be assessed in a formative manner and to be facilitated by the teacher for external assessment’ by the State Examinations Commission (p. 87). The accompanying details for the performance assessment further conveyed expectations for teaching and learning relating to this component of assessment. The National Council for Curriculum and Assessment ([NCCA], 2017) clarified that digital capture would occur ‘in sessions designed to demonstrate the learner’s best personal performance *in fully competitive and/or conditioned practices or performance settings*’ (p. 47, emphasis added), with learners expected to demonstrate ‘their capacity to select, apply and perform the skills and techniques’, ‘ability to apply and adapt different tactics, strategies and compositional knowledge in response to different scenarios’ and their ‘knowledge and application of relevant rules, regulations and codes of practice’ in these settings (NCCA, 2017, p. 47). The NCCA (2017)’s stipulated format and accompanying expectations for performance assessment have notable similarities to those featuring in the trial validation in WA and described below.

Looking across jurisdictions, Whittle et al.’s (2017) finding that only 6 senior secondary physical education courses of 15 analysed, incorporated performance assessment in the exit (final) year of courses, is particularly pertinent to this study. Whittle et al. (2017) reported that in all instances, ‘physical performance is assessed independently of the assessment of other areas of the course’ (p. 618), and in only two cases (Western Australia

and the Caribbean Islands) was it an element of external assessment. The backdrop for this research is thus, the relative marginality of performance assessment in examination and senior secondary physical education, particularly in external assessment, and the parallel sustained privileging of propositional knowledge (Brown & Penney, 2017, 2019; Whittle et al., 2017). The focus on a context and course developments that have openly sought to value performance assessment, is therefore significant. Additionally, this research drew upon and sought to contribute to scholarly insights and professional debates centring on matters of quality and equity in physical education assessment (See e.g. AIESEP, 2020; Borghouts et al., 2017; Hay & Penney, 2009, 2013).

Quality and equity in physical education assessment

Following Hay and Penney (2013), this study recognised two foundational elements of assessment, ‘collecting information’ and ‘making interpretations’ (p. 7), and their inter-relationship, as key considerations for quality and equity in assessment. Hay and Penney (2013) emphasised that any exploration of these elements needs to engage with the prime purpose and context of assessment. In a high-stakes and/or examination context, the need for assessment to provide ‘opportunity for an account of learning’ (Hay & Penney, 2013, p. 7, original emphasis) sets a particular frame for many assessment decisions. In the case of performance assessment in examination or senior secondary physical education, these decisions include:

- what information will be collected, by whom, via what means, in what physical activity and task conditions, and
- what criteria, standards and weightings will be employed, how and by whom, in making an interpretive judgement in relation to the information gathered.

Our focus locates such decisions in an environment characterised by high accountability. Following Hay and Penney (2013) we propose that accountability discourses should not displace concerns that a performance assessment experience delivers on its potential to support valued, authentic learning. That is, ‘the learning experiences that form the medium for information gathering have application and meaning for students’ lives and are not abstract or dissociated’ (Hay & Penney, 2013, p. 9). From this perspective, quality assessment in a high-stakes environment is fundamentally oriented towards facilitating, as well as evidencing and communicating, quality learning (Hay & Penney, 2013). Furthermore, claims for assessment efficacy rest on ‘the satisfaction of this learning intent through *the authentic, socially just alignment of assessment, curriculum and pedagogy*’ (Hay & Penney, 2009, p. 390, emphasis added). The conceptualisation of quality assessment advanced by Hay and Penney (2009, 2013) thus centres on quality learning, foregrounds the alignment of assessment, curriculum, and pedagogy, and is inherently tied to the need to pursue questions of equity in any exploration of assessment. This conceptualisation informed our investigation of changes to performance assessment in senior secondary physical education in WA.

Performance assessment in PES in Western Australia

Since 2008, performance assessment has been an examination assessment component within the PES course in Western Australia (SCSA, 2016), that may count towards

students' tertiary entrance score (Jones & Penney, 2019; Paveling et al., 2019; Penney et al., 2012). The existing format of the performance assessment examination within what is termed the 'ATAR' (Australian Tertiary Admissions Rank) version of PES in WA, and several associated examination procedures, constituted an important backdrop to this research. Table 1 identifies key characteristics of the existing performance

Table 1. PES performance assessment in WA (information gathered from SCSA (2011, 2016, 2020, 2021a, 2021b)).

Existing performance assessment examination	Trial performance assessment validation
Designated day at centralised examination centres (Perth metropolitan and some larger regional centres). Students travel to venues (may include air travel to Perth). Students choose to be examined in one of 10 sports:	School-based on a date agreed between SCSA and the school. Validators travel (road and/or air) to schools. School chooses the sport context for the performance assessment validation from the 10 sports listed. All PES students at a school undertake the assessment in the sport selected by the school.
<ul style="list-style-type: none"> • Australian football • Badminton • Basketball • Cricket • Hockey • Netball • Soccer • Tennis • Touch football • Volleyball 	
This is an individual student choice and may or may not be a sport that has been studied at school. Approximately 30 minutes of actual examination time for each candidate. Two equally weighted components:	Approximately 30 minutes. Some variation depending on number of students to be assessed. No distinct skill performance component. Conditioned performance in a modified format competitive game play environment addressing three criteria: skill execution, spatial awareness, and tactical application.
<ul style="list-style-type: none"> • Skill performance in 5 skills selected by the examination panel from a published range of skills for each sport (e.g. lob pass, 3-feet (0.9 m) recovery, outside foot land and pivot, split and re-offer, and shadowing, for netball in the 2020 examination). • Conditioned performance (application of skills, decision making, and tactics in a modified competitive environment). 	
Numeric scale used to assess and allocate marks relating to proficiency and application of individual skills. Detailed marking guides for each sport articulate the mark allocation applied for each skill performance and in each element of the conditioned performance in that sport (e.g. offensive and defensive play).	Marking guide for each sport provided descriptors for allocation of marks using 2-point bands (i.e. 1–2, 3–4, 5–6 etc.) for marks associated with each criterion. 'Skill execution' – 20 marks comprised of 10 marks for skill proficiency and 10 for selection and application of skills. 'Spatial awareness' – 10 marks. 'Tactical application' – 20 marks comprised of 10 marks for offence and 10 for defence.
Marking undertaken by two specifically trained and independent markers who reconcile any differences with the assistance of a Chief marker if required. Students only identifiable by a unique student number. Support materials published for each sport provide detailed specifications for all skills and associated drills that may be included in the performance examination in any given year, and guidance relating to the foci that may be selected for inclusion in the conditioned performance component.	Marking undertaken independently by the teacher and a validator. Both sets of marks submitted to SCSA. Students identifiable by their teacher. The assessment guides provided by SCSA for each sport included the criteria and marking guide to be used by teachers and validators and information relating to the space, team composition, game format, rules and equipment required.

assessment examination specifications and procedures and provides comparable information about the 2021 performance assessment validation trial.

Previous research engaging with the existing PES performance assessment examination has affirmed a flow-on influence for pedagogy in year 11 and 12 PES classes, and in preceding secondary years (Jones, 2017; Jones & Penney, 2019; Paveling, 2016; Paveling et al., 2019). While the PES performance examination can be regarded as an important prompt for practically based teaching and learning in PES classes, the skill performance component particularly has been associated with ‘teaching to the test’ approaches. These focus on preparing students to replicate the ‘de-contextualised’ performance of specific skills in the manner stipulated in examination support materials for each sport, and as represented in examination marking keys. Students’ scope to choose to be assessed in a sport that is not covered in the PES curriculum at the school, has also led some schools to minimise practical teaching within PES curriculum time (Jones, 2017). Meanwhile, from an equity perspective, teachers who have direct knowledge and experience of examination marking have been identified as advantaged in their capacity to support students (Paveling, 2016), and students in regional and/or rural locations are acknowledged as often facing considerable logistical challenges to participate in the performance examination.

These insights from research reflect tensions arising amidst efforts to privilege performance assessment in a high-stakes examination context and in a vast state. Matters of quality and equity, together with feasibility and manageability considerations, were all reflected in SCSA’s decision to undertake a trial of a different format for PES performance assessment, with a new set of arrangements.

The PES performance assessment validation trial

In 2021 SCSA designed and implemented a performance assessment validation trial that involved school-based assessment of students participating in modified format competitive game play, in each of the 10 sports offered in the PES ATAR examination (see Table 1). Undertaking the assessment at schools and focusing on modified game play were marked contrasts to the existing examination. For the trial, SCSA recruited 33 schools across metropolitan and regional locations, appointed validators for the 10 sports, and produced assessment guidance for each sport. All trial validation sessions were scheduled by SCSA in negotiation with schools, teachers, and validators, to occur within normal lesson time. Schools selected the sport that they would undertake the trial in.

The assessment guidance provided by SCSA included the criteria and marking guide to be used by teachers and validators and information relating to the space, team composition, game format, rules and equipment required. Appendix provides an example of the assessment overview provided in all sport-specific assessment guides. Students undertaking the assessment were assigned bib colours and numbers for unique identification and rotation systems (positions and teams) were applied to provide all students with sufficient opportunity to demonstrate their skills, spatial awareness, and tactical application in the sport. The lead teacher at each school and the validator in attendance were tasked to independently assign marks for all students, using the marking guide provided. SCSA’s guidance relating to the criteria is also pertinent to note:

- Competence in demonstrating the individual skills required in the selected sport must be assessed holistically rather than by focusing on a detailed analysis of their individual parts. The final mark for ‘Skill execution’ must also take into account the timing and appropriateness of skills being used in specific competitive conditions.
- The assessment of ‘Spatial awareness’ includes observations made with respect to the use of space through movement, positioning, shot accuracy and placement. These must be demonstrated in offence and defence as well as in various positions and roles.
- When allocating a mark for ‘Tactical application’, shot selection and placement in relation to teammates and/or opponent, possession and scoring opportunities must be taken into consideration. In this section, students will be rewarded for their demonstration of deception, creativity and/or anticipation. These must be demonstrated in offence and defence as well as in various positions and roles. (SCSA, 2021b, p. 3)

SCSA’s (2021b) emphasis was that the final mark awarded to a student for each criterion ‘must be that which reflects the student’s performance consistently during the assessment and not intermittent occurrences at either end of the continuum’ (p. 3). The performance assessment thus emphasised holistic and contextualised judgements.

Researching the PES performance assessment validation trial

SCSA commissioned this research to complement their internal evaluation of the trial and to inform future policy and practice pertaining to performance assessment in PES and prospectively, other courses with a performance assessment component. The remit for the research was to (i) explore issues of feasibility and manageability in the performance assessment validation trial process, (ii) investigate strengths and limitations of the trial from the perspective of teachers, validators, and SCSA staff managing the trial, and (iii) in the light of data, identify implications and considerations for SCSA. As explained below, this paper reflects a particular focus that has been pursued in expanded analysis of the dataset generated from the research undertaken for SCSA.

Ethical approval for the research was gained from the researchers’ institution (Edith Cowan University, Approval 2021-02302-PENNEY). The project combined observation, interview, and documentary methods to investigate the above issues. The research did not extend to analysis of teacher and validator marks, nor did it explore the financial costs of the trial in comparison to current PES performance assessment arrangements. The scope and timeframe for the project also precluded data collection from students. Documentary data comprised the guidance materials produced for the trial by SCSA. Observation of the performance assessment validation trial was conducted for nine of the 10 sports. Each observation involved a different school. Observation in soccer was not possible due to two cancellations, firstly because of inclement weather and secondly, due to an insufficient number of students able to attend the assessment. External facilities were utilised in the trials observed for hockey (a specialist hockey centre), badminton (a leisure centre) and cricket (a specialist indoor cricket facility). All other trials observed were on school sites. For all observations, researchers arrived 30 minutes prior to the scheduled start time to observe set-up and attended until pack-up was completed. Data was collected via detailed fieldnotes, using an observation guide with the intent of fully documenting: the processes followed in the performance assessment validation,

issues pertinent to feasibility and manageability (including equipment, facilities, grouping/organisation of students, time involved in all aspects of the trial), and issues arising associated with strengths and limitations of the trial (generic, sport-specific and school-specific issues).

Semi-structured interviews with teachers, validators and the SCSA staff managing the trial, explored the three main areas of research interest identified above. An interview guide provided prompts for each part of the interview. In total 25 interviews were conducted, either in-person or via zoom, involving 11 validators, teachers from 13 schools and SCSA staff. All interviews were audio-recorded, professionally transcribed, checked and edited by the researchers to remove all identifying information, and member-checked by participants.

Initial analysis of data explicitly explored each of the three main lines of inquiry, relating to the trial as a whole and to also pursue sport-specific findings. Themes and sub-themes pertaining to the three foci (feasibility/manageability; strengths/limitations; and implications) were generated and progressively refined through multiple readings, coding, and re-coding of data. This paper reports on extended analysis of the project dataset that was conducted with the intent of critically examining issues of quality and equity associated with the performance assessment validation trial. The two foundational elements of assessment identified by Hay and Penney (2013, p. 7) ‘collecting information’ and ‘making interpretations’, and their inter-relationship, were employed as a guiding framework to further explore the findings arising from the initial analysis with this specific focus on quality and equity in performance assessment. The sections that follow foreground the empirical and conceptual insights arising from this framework and focus being applied.

Findings: assessment processes and practices

Collecting performance assessment information

This section probes the ways in which the trial specifications and arrangements described above, enacted in different sport and school contexts, variously shaped the collection of assessment information by teachers and validators. ‘Collecting information’ as explained by Hay and Penney (2013), encompasses what information is collected, how, by whom, and with what intent. From the outset, it is pertinent to acknowledge that the SCSA performance assessment validation trial set parameters for the collection of assessment information, in particular school contexts and game conditions. The prime intent of the performance assessment was to provide ‘opportunity for *an account of learning*. That is, the information collected through assessment is used to inform others of learning and learning quality’ (Hay & Penney, 2013, p. 7, original emphasis). Data arising needs to be viewed with this purpose and orientation in mind. In several instances, attention is drawn to aspects of interpretation evidenced in teachers’ and validators’ collection of assessment information. The inter-relationship between the two foundational elements of assessment is thus emphasised as a critical facet of processes and practices being employed with the intent of generating the required ‘account of learning’ for all students being assessed in the trial.

As indicated, one of the distinct features of the assessment was that it was school-based, with external validators travelling to schools. The assessment setting thus varied

in quality. One teacher noted, for example, that ‘we don’t have a full-time groundskeeper who can make it brilliant’, adding that the oval being used for the performance assessment was ‘in better condition last week than it was today’. While playing areas were clearly defined and marked for many of the sessions observed, the use of temporary markers and/or spray paint was required in some instances. In one case it was noted that some students were getting confused because of apparent unfamiliarity with the temporary area markings being used.

Validators also recognised the prospective influence that type and quality of playing surface could have in the performance assessment and pointed to differences arising with indoor and outdoor playing conditions.

Is it grass or hard? You’re on different surfaces, the length of your point shortens if it’s on grass compared to hard because of the variation of the bounce. (Validator)

I think playing indoors as opposed to outdoors is quite different. So, if some kids are examined indoors and others are outdoors, you need to try to make sure that it’s fair. So, I think it should be indoors and even if the school hasn’t got a full-size court/gym, I’m sure they can hire somewhere for that. (Validator)

In one of the observed trials, a school chose to hire an external facility for the performance assessment in badminton, explaining that they felt the lighting and quality of space was preferable to that in their school gym. A validator also expressed the view that hockey was a sport that ‘would typically have to be offsite at a community facility’ because of the need for astroturf, which very few schools have on-site.

Variation in the type and quality of equipment (net height, ball size and pressure) being used at different schools was similarly recognised as potentially impacting the quality of the assessment opportunity for students, and in turn, the assessment information able to be collected from the performance assessment. Several validators also highlighted that formal sport specifications called for differences in equipment (e.g. ball size) and/or set up (e.g. net height) for males and females, that in their view, needed to be reflected in the performance assessment conditions. The performance assessment validation trial thus drew attention to several physical resource factors impacting quality and equity in the school-based assessment process. In doing so, it prompted consideration of enhanced standardisation and/or further specification of the assessment conditions.

With students being assessed in a modified game format competitive activity, and information collection and interpretation required to be undertaken in real time (rather than, for example, retrospective judgements being made with reference to recorded assessment information) group size and composition was also shown to be important. The total number of students to be assessed at any school, the composition of the student group (particularly in ability range and gender balance), and aspects of the arrangements for performance assessment, variously impacted game ‘quality’ and the quality of the assessment opportunity arising for different students. The performance assessment format, student cohort being assessed, and the precise arrangements employed in relation to groupings and rotations, were thus recognised as inter-related influences on quality and equity in the collection of assessment information. The following points expand upon these influences as seen in different sport and school contexts.

Overall group size was identified as impacting the opportunity for teachers and validators to gather quality assessment information about all students and avoid assessment

being rushed within the lesson time available. ‘It was a very big group ... the size of the group makes a difference. A big difference’ (Validator). A smaller overall group size was associated with less pressured conditions for teachers and validators to collect sufficient assessment information and reach a judgement in relation to each criterion. The emphasis here is that these were necessarily inter-related activities, with teachers and validators progressively looking to collect additional information that would enable them to fill gaps in the accounts of learning they were producing for all the students being assessed. The pressures of the ‘point in time’ nature of the performance assessment were thus also evident, especially with larger student cohorts.

Combining girls and boys in the performance assessment was regarded as problematic, with girls frequently identified as disadvantaged, particularly in instances where they were very outnumbered by boys. For example, a teacher reflected ‘my one female student was heavily disadvantaged in that game’. Another teacher commented, ‘Ideally, it’s boys only, and girls only. And the reality is, you probably do mark them slightly differently, because they play a different way, the girls and the boys’.

It was also rare for the class size (or the number of boys, or girls) to exactly align with the required team sizes for the modified game. In these instances, recruitment of additional students, and/or substitutions as well as rotations (from e.g. offence to defence) were required, with validators determining what changes were made and when, during the assessment. Decisions about groupings, substitutions and rotations were thus central to ensuring that all students had equitable opportunities to demonstrate their learning – and that teachers and validators could collect adequate assessment information about all students being assessed. One validator explained that,

They [rotations or interchanges] come at a point where you don’t see enough from players ... I’m trying to create the challenge, so that they can show me whether they are up to the standard, or whether they have the capability to perform the shots ... (Validator)

Groups also typically featured a considerable range in students’ abilities. Teachers were asked to pre-rank students and arrange groups comprising students of similar ability. This was more or less feasible in the light of overall student numbers, gender balance and range of abilities within the group. It also highlighted that teachers’ anticipation of students’ performance in relation to the criteria and thus, aspects of interpretation, came into play in shaping the context for the collection of assessment information. In practice, a large disparity in ability levels within a group was seen to be problematic given the intent that the performance assessment will afford *all* students appropriate opportunity to demonstrate their skill execution, spatial awareness, and tactical application in a modified game play context. Sole or a small number of ‘high level’ students within a mixed ability were recognised as potentially disadvantaged by the absence of players around them of comparable skill level and/or understanding.

If you’ve got an elite player in with kids that have never played before, it’s really hard because they throw the ball behind, or they don’t go to the spaces they should. And no matter how hard that girl or boy works, they can’t find a space on court to demonstrate their repertoire of skills. (Validator)

In another setting, a sole lower ability student was recognised as potentially disadvantaged for very similar reasons. Talking about one student, a teacher reflected that they

were ‘just not up to the level’ and consequently they ‘just don’t get the ability ... to demonstrate their skills ... not saying they had high skills, but they don’t run to the right spots ... so they disadvantage themselves’.

Rotations to enable students to be observed playing with and against different peers, and in different playing positions, were again a crucial strategy in seeking to minimise any disadvantage individual students may experience in the mixed-ability assessment context and to ensure quality assessment information could be gathered for all students. The modified game format was recognised as minimising, but not entirely precluding the potential for lower ability students to ‘hide’ or remain at the periphery of play. At the same time, it was noted that the presence of some higher ability students could serve to raise the performance of lower ability students, with some students observed actively supporting less able students by creating opportunities for them. Such cooperation between students was positively associated with the performance assessment being school-based, with students assessed with their usual classmates.

The provision of a referee or umpire and their capacity to maintain appropriate flow in play also emerged as an important consideration in the assessment. Both teachers and validators recognised that officiating could influence the quality of the performance assessment experience for students and hence, teachers’ and validators’ capacity to collect appropriate and adequate assessment information.

Someone who can control the game is definitely going to make the experience and the game format much, much better. (Validator)

You can’t run any game without officiators to drive that competitiveness and get the kids the best marks they can get in gameplay. (Teacher)

Teachers’ and validators’ experience in undertaking PES performance assessment previously, and their familiarity with the criteria and marking guide for the trial assessment, were also associated with the perceived ease or challenge of producing the required account of learning for all students. Some validators clearly drew on their previous experience of marking in the current PES performance examination to inform their approach to collecting assessment information and making interpretive judgements.

The better athletes were the ones that had the arms moving left and right, you could see them communicating, ‘my ball’, ‘this ball’, and they just lift to a different level. And I think that’s always been something in [sport name] in the exams, whoever communicates and dictates and tells everyone what they’re doing is always going to get a better mark ... (Validator)

The next section turns to findings that centred more overtly on ‘making interpretations’; that is, the processes that teachers and validators engaged in to arrive at an account of students’ learning. Data illustrates challenges that teachers and validators experienced in working with the criteria, marking guides and mark recording forms used in the trial. It also further speaks to complex inter-relationships between the two foundational elements of assessment being evidenced in teachers’ and validators’ assessment practices. We suggest that ‘collecting information’ and ‘making interpretations’ are inter-twined in the processes that teachers and validators employed for the purposes of making and presenting their judgements about students in the format required by SCSA.

Making (and recording) interpretive judgements

Teachers' and validators' descriptions of their experiences working with the marking criteria, the associated marking guide and marks recording form, highlighted the extent to which their understanding of the criteria shaped their approach to information collection, as well as the interpretation of assessment information. It was clear that in many instances, teachers and validators found addressing the criteria challenging, with both the number of criteria and some perceived lack of distinction between criteria identified as sources of concern. With respect to the number of criteria, as teachers and validators recognised, the marking scheme and reporting form called for five judgements for every student: skill execution – proficiency; skill execution – selection and application of skills; spatial awareness; tactical application – offence; and tactical application – defence.

I just found there was quite a lot, like, there was two, the section one and the section three had two different marks, and so ... it was looking at quite a lot of different things. But I think all the things we were looking at were worthwhile. (Teacher)

One validator reflected that 'there's a lot of crossover, you're marking the same things in a lot of areas' (Validator). A teacher expanded on one criterion that several participants singled out as, from their perspective, problematic; 'I feel that [spatial awareness] should be part of the strategies, the like strategies and tactics kind of mark ... We always talk about strategies and tactics, we talk about identifying space, creating space, using space'. Another validator's comment indicated that from their perspective the quality and/or extent of game play observed meant that it was extremely difficult for them to gather information to inform a judgement about this criterion:

Spatial awareness was the one that really did my head in, I'm thinking, how the hell do we assess this when you've got kids just sort of running around within a 25-metre arc and just kind of passing the ball around. (Validator)

In another setting, however, a validator's response and approach to this criterion was quite different; 'the easiest to mark is the spatial awareness, that's the one I start with. Are they moving to the right place in the right time, and is it the right thing to do?'

The indication was that understandings of criteria and in turn, the processes employed to gather and interpret assessment information, varied. The necessary focus on generating a particular account of learning, framed by criteria that were relatively unfamiliar, gave rise to challenges and dilemmas that further affirmed that interpretive judgements were inherent in information collection.

I've got offence, defence, I need to see that. I'm seeing some offence, but other than serve, receive, I'm not seeing any defence ... So does that mean already I'm down to 50 per cent? Theoretically yes, but that's not how we're marking. (Validator)

Some teachers and validators were seeking more structure or specification to inform both information collection and their interpretive judgements, particularly for skill execution.

I think it needs to be ... these are the group of skills we're looking for. (Validator)

Imagine if every sport has this spreadsheet, and then you've got your skills and – however it's set up, which I'm sure they can so, then we're all on the same page. (Teacher)

One validator's comment that 'I just think it's got to be simpler' seemed to capture a shared sentiment, particularly given the unfamiliarity of the process. Another validator reflected that 'for me, it's not a straightforward process'.

An added challenge for validators was that they did not know the students and hence, were reliant upon bib colours and numbers to then link their assessment information with a particular student for the purposes of informing and recording a judgement.

Because we're pressed for time ... you know, you're hurrying. And if two kids look alike and their numbers are a three and a five, it's a little bit ... you know I try really hard to make sure I've got the right kid with the right mark, but sometimes I'm not a hundred percent. (Validator)

Teachers had the benefit of readily being able to recognise individual students and hence, more quickly associate observations with individual students. At the same time, teachers' familiarity with the students presented them with a different challenge, of seeking to ensure that their interpretative judgements were made with reference to the assessment information generated in the performance assessment, on the day – rather than on prior occasions. In talking about differences between teacher and validator views, one validator reflected 'perhaps it's because they see them every single day, every single session, every single class, and therefore they're saying that no, I have it in my mind you are there'.

Observations and interviews with teachers and validators gave further insight into their interpretation of assessment information during the performance assessment. They frequently referred to a ranking approach in which they, in essence, made a holistic comparative judgement, to then guide their allocation of marks in relation to the criteria.

I think the first thing we did was rank the students ... and speaking to a lot of markers from a lot of different sports, that is a pretty standard process for how marks are allocated on exam day, it's just ranking your best student all the way through to your worst student. (Validator)

As this comment reflected, this practice was associated with marking processes employed in the current performance assessment examination. Thus, we saw the transfer of existing assessment practices to the new performance assessment context. Validators and teachers also identified that they tended to establish a mark range (e.g. 5–6) to be allocated to a student for a specific criteria and used subsequent observations to confirm the mark to be awarded. Again, therefore, interpretation and data collection were inter-related processes. 'Scribble sheets' and/or the space to the side or below the mark recording cells, was also used to progressively record observations about individual students and/or highlight a need for a further observation to confirm a mark. Again, time pressures and the perceived complexity of the assessment process were highlighted;

I can see two, three, sometimes four things happen, and I wanted to put a mark for those kids in that regard. And then you have to find it on the sheet because you've got five columns. And then you're thinking, well okay, I've got two or three or four assessment here, where's the other five? And I've got nothing for them ... (Validator)

It is difficult to get through the marking key with these amounts of kids in that amount of time. That is the single most difficult thing. To make accurate judgements. (Validator)

Invariably, full completion of the marks recording came after the performance assessment was finished, with teachers and validators then using their notes and provisional marks to inform their allocation of a full set of marks for all students. The process of

making interpretive judgements was thus iterative and inter-twined with the collection of assessment information. It was a process in which teachers and validators clearly sought to ensure both quality and equity in their assessment, while recognising challenges in doing so.

A final characteristic to note about the marking process was teachers' and validators' appreciation of the opportunity the validation trial provided for collegial discussions about assessment. There was shared recognition of the critical role that professional development could play in enhancing understandings of performance assessment and consistency in judgements.

[The trial was] such a valuable exercise as professional development for teachers. The collaboration between us and the validator was invaluable. Having the chance to talk to a fellow colleague and examiner about specific sport and skill related aspects is vital to ensure we are on the right page ... (Teacher)

Actually sitting with teachers to say, 'No, for me, that's a five out of ten' ... [and talk about] 'What does it look like?' I think the PD is going to be the most important part of this ... (Validator)

Comments such as this affirmed the value of the trial in facilitating professional learning amongst the teachers and validators involved, while also highlighting the investment needed to support further progress towards quality and equity in performance assessment between sports and across schools in the state.

Conclusion: progressing quality and equity in performance assessment

The conceptualisation of quality assessment advanced by Hay and Penney (2009, 2013) and expanded upon in the AIESEP (2020) position statement on PE assessment centres on quality learning, foregrounds the alignment of assessment, curriculum, and pedagogy, and is inherently tied to the need to pursue questions of equity in assessment. It is a conceptualisation that calls for developments such as the validation trial in WA to be understood and researched as a pedagogical process, the success of which will ultimately be measured by advances in both teachers' and students' learning. The AIESEP (2020) position statement affirms that extending teachers' and students' knowledge and understandings of assessment is essential in policy developments directed towards advancing assessment quality and equity. We suggest that the preceding analysis provides an important foundation for addressing this need and for the strengthening of policy and guidance relating to the PES performance assessment validation. More specifically, our data vividly illustrates that both foundational elements of assessment explored – collecting assessment information and making interpretive judgements – and the ways in which these elements are inter-related in policy and practice, will variously impact assessment quality and equity. How both elements and their inter-relationships are addressed in assessment policy and/or guidance, and how they are enacted in practice, emerge as important considerations for future policy development in PES in WA and in other jurisdictions engaged in performance assessment.

In considering strengths of the performance assessment validation trial, the game play format and the criteria employed in the assessment were identified with increased authenticity in assessment. The school-based setting saw the enhanced opportunity for

students to demonstrate communication skills and tactical skills in game play with known peers, while also avoiding logistical challenges and stress for regional students particularly. Further, the trial validation was regarded as an important professional learning opportunity for teachers and validators involved. The prospect of such a validation process being extended to all schools offering PES at examinable level was widely welcomed as a move that would support state-wide advances in performance assessment practices. While we affirm that such a development extends opportunities for teachers across the state to engage in high-stakes assessment processes, at this time it is not possible to comment on the impact that such engagement may have on teachers' assessment knowledge and practices.

Some limitations of the trial performance assessment validation were also revealed through this research. Quality game play that could be deemed authentic and appropriate to create adequate opportunities for all students to demonstrate their skills, knowledge and understanding pertinent to the criteria in this 'point in time' assessment, was by no means assured. Within school-based assessment environments, students' varied skill proficiency and tactical understanding impacted game quality and in turn, assessment opportunities. Gender equity was also an issue highlighted in the trial as requiring further consideration.

The assessment process itself, requiring five judgements to be made and recorded for all students in real time, was challenging for teachers and validators. Ultimately, all teachers and validators drew on varied professional knowledge and experience in striving to achieve quality and equitable outcomes for all students in the trial, while engaging in an assessment process that for all involved, was necessarily a learning process. In revealing these challenges, the research also drew attention to the ways in which overarching features of the performance assessment process – namely, that it is a point in time assessment process with judgements made in real time – set parameters within which efforts to achieve quality and equity need to be understood. Processes associated with a performance assessment that are employed in other jurisdictions (such as the use of portfolios in the LCPE as reported by Scanlon et al., 2019) illustrate alternative approaches that it was beyond the scope of this research to consider.

The research reported in this paper was commissioned to inform the next steps in shared learning and SCSA's development of performance assessment processes. It has shaped a number of modifications to the format, arrangements, criteria, and guidance materials for the performance assessment that are being implemented by SCSA in a state-wide pilot in 2022. For the physical education research and policy community more broadly, this work has extended insight into the complexities inherent in designing and enacting high-stakes performance assessment and the associated challenges in delivering on expectations that such assessment will be characterised by quality and equity.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the School Curriculum and Standards Authority of Western Australia.

Notes on contributors

Dawn Penney is a professorial research fellow in the School of Education at Edith Cowan University. Dawn's research in senior secondary education and assessment in physical education has spanned three decades and involved collaboration with jurisdictions nationally and internationally. Her work across all sectors of education and in sport settings brings sociological insight to the exploration of developments in policy and pedagogy, with a focus on issues of equity.

Eibhlísh O'Hara is a lecturer of Primary Health and Physical Education in the School of Education at Edith Cowan University. She is an experienced kindergarten – year 7 generalist teacher who pursued graduate studies in psychology. Eibhlísh has considerable interdisciplinary and mixed methods research experience centred on Health and Physical Education. She has been involved in projects directed towards building excellence in learning and teaching within the field of Health and Physical Education.

Rob Lund is a lecturer in the School of Education at Edith Cowan University. He is an experienced teacher and has worked in education in both the UK and Australia, across a range of sectors, learning areas and age ranges, from primary through to tertiary contexts although his learning area specialism is Health and Physical Education. Building on his contribution to the project reported in this paper, Rob is commencing PhD research to further explore developments in pedagogy that are associated with changes in assessment specifications in senior secondary physical education.

ORCID

Dawn Penney  <http://orcid.org/0000-0002-2000-8953>

Eibhlísh O'Hara  <http://orcid.org/0000-0002-9093-2746>

Rob Lund  <http://orcid.org/0000-0002-7645-0863>

References

- AIESEP. (2020). *Position statement on physical education assessment*. <https://aiesep.org/scientific-meetings/position-statements/>
- Barnes, M., Clarke, D., & Stephens, M. (2000). Assessment: The engine of systematic curricular reform? *Journal of Curriculum Studies*, 32(5), 623–650. <https://doi.org/10.1080/00220270050116923>
- Borghouts, L. B., Slingerland, M., & Haerens, L. (2017). Assessment quality and practices in secondary PE in the Netherlands. *Physical Education and Sport Pedagogy*, 22(5), 473–489. <https://doi.org/10.1080/17408989.2016.1241226>
- Bowes, M. (2010). Teaching as inquiry. What has influenced the development of senior school physical education in New Zealand? *Journal of Physical Education New Zealand*, 43(2), 20–24.
- Brown, T. D., & Penney, D. (2017). Interpretation and enactment of senior secondary physical education: Pedagogic realities and the expression of Arnoldian dimensions of movement. *Physical Education and Sport Pedagogy*, 22(2), 121–136. <https://doi.org/10.1080/17408989.2015.1123239>
- Brown, T. D., & Penney, D. (2018). *Examination physical education. Policy, practice and possibilities*. Routledge.
- Hay, P. J., & Penney, D. (2013). *Assessment in physical education. A socio-cultural perspective*. Routledge.
- 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. <https://doi.org/10.1080/09695940802164184>
- Hay, P., & Penney, D. (2009). Proposing conditions for assessment efficacy in physical education. *European Physical Education Review*, 15(3), 389–405. <https://doi.org/10.1177/1356336X09364294>

- Jones, A. (2017). *Lost in translation? – The “integration of theory and practice” as a central focus for senior schooling physical education studies* [Doctoral dissertation]. Edith Cowan University. <https://ro.ecu.edu.au/theses/1950>
- Jones, A., & Penney, D. (2019). Investigating the ‘integration of theory and practice’ in examination physical education. *European Physical Education Review*, 25(4), 1036–1055. <https://doi.org/10.1177/1356336X18791195>
- Macdonald, D., & Brooker, R. (1997). Assessment issues in a performance-based subject: A case study of physical education. *Studies in Educational Evaluation*, 23(1), 83–102. [https://doi.org/10.1016/S0191-491X\(97\)00006-0](https://doi.org/10.1016/S0191-491X(97)00006-0)
- MacPhail, A. (2007). Teachers’ views on the construction, management and delivery of an externally prescribed physical education curriculum: Higher grade physical education. *Physical Education and Sport Pedagogy*, 12(1), 43–60. <https://doi.org/10.1080/17408980601060267>
- National Council for Curriculum and Assessment. (2017). *Physical education curriculum specification. Leaving certificate ordinary and higher level*. https://curriculumonline.ie/getmedia/41817053-8f40-4365-8893-dba1a68508f3/LCPE_Specification_en.pdf
- Paveling, B. J. (2016). *Senior school physical education curriculum policy reforms in an Australian context: A focus on Western Australia 2005 to 2015* [Doctoral dissertation]. University of Western Australia. <https://doi.org/10.4225/23/59b7656784d5a>
- Paveling, B., Vidovich, L., & Oakley, G. (2019). Global to local tensions in the production and enactment of physical education curriculum policy reforms. *Curriculum Studies in Health and Physical Education*, 10(2), 141–155. <https://doi.org/10.1080/25742981.2019.1583066>
- Penney, D., Jones, A., Newhouse, P., & Cambell, A. (2012). Developing a digital assessment in senior secondary physical education. *Physical Education and Sport Pedagogy*, 17(4), 383–410. <https://doi.org/10.1080/17408989.2011.582490>
- Scanlon, D., MacPhail, A., & Calderón, A. (2019). Original intentions and unintended consequences: The ‘contentious’ role of assessment in the development of leaving certificate physical education in Ireland. *Curriculum Studies in Health and Physical Education*, 10(1), 71–90. <https://doi.org/10.1080/25742981.2018.1552500>
- SCSA. (2011). *Physical education studies. Support materials for practical examinations. Netball*. https://senior-secondary.scsa.wa.edu.au/__data/assets/pdf_file/0017/131255/Physical-Education-Studies-practical-examination-support-material-Netball.pdf
- SCSA. (2016). *Physical education studies. ATAR course. Year 12 syllabus*. https://senior-secondary.scsa.wa.edu.au/__data/assets/pdf_file/0011/10172/Physical_Education_Studies_Y12_Syllabus_ATAR_GD.pdf
- SCSA. (2020). *Physical education studies. Practical (performance) examination 2020 netball. Marking key*. Retrieved from: https://senior-secondary.scsa.wa.edu.au/__data/assets/pdf_file/0011/649487/2020_PES_Practical_Marking_Key_Netball.PDF
- SCSA. (2021a). *Physical education studies ATAR course practical (performance) examination requirements*. https://senior-secondary.scsa.wa.edu.au/__data/assets/pdf_file/0005/652262/2021-PES-ATAR-Practical-performance-examination-requirements.pdf
- SCSA. (2021b). *Physical education studies. Support materials for school-based practical assessment. Basketball*.
- Thorburn, M. (2007). Achieving conceptual and curriculum coherence in high-stakes school examinations in physical education. *Physical Education and Sport Pedagogy*, 12(2), 163–184. <https://doi.org/10.1080/17408980701282076>
- Thorburn, M. (2008). Articulating a merleau-pontian phenomenology of physical education: The quest for active student engagement and authentic assessment in high-stakes examination awards. *European Physical Education Review*, 4(2), 263–280. <https://doi.org/10.1177/1356336X08090709>
- Whittle, R., Benson, C., & Telford, A. (2017). Enrolment, content and assessment: A review of examinable senior secondary (16–19 year olds) physical education courses: An international perspective. *The Curriculum Journal*, 28(4), 598–625. <https://doi.org/10.1080/09585176.2017.1318770>

Appendix. Overview of performance assessment for basketball (SCSA, 2021b, p. 4)

Section	Basketball observations
<p>Section one Skill execution – 20 marks (30% of the practical assessment) The student is required to demonstrate individual, sport specific skills within a competitive environment. The assessment will include:</p> <ul style="list-style-type: none"> • proficiency of performance in individual skills • selection and application of appropriate skills in specific situations. 	<ul style="list-style-type: none"> • Execution consistency • control fluency • precision • kinematics of arms during throwing/catching including sequencing and energy transfer • footwork • skill selection and timing^a
<p>Section two Spatial awareness – 10 marks (15% of the practical assessment) The student is required to demonstrate individual and/or team, sport specific tactics within a competitive environment. The assessment will include:</p> <ul style="list-style-type: none"> • use of space • positioning. 	<ul style="list-style-type: none"> • pass distribution^a • selection of shot/pass for optimum advantage • appropriateness of pass distance • successful passes to target • passes intercepted • dribbling with the ball into space^a • support of ball carrier^a • support of other defenders^a • positioning/movement to dispossess or intercept^a
<p>Section three Tactical application – 20 marks (55% of the practical assessment) The student is required to demonstrate individual and/or team, sport specific tactics within a competitive environment. The assessment will include the use of team/individual tactics.</p>	<p><u>Offence</u></p> <ul style="list-style-type: none"> • maintaining possession • setting up a play • ball projection (angle), velocity and spin^a • pace of attack^a • creating scoring opportunities • communication with teammates <p><u>Defence</u></p> <ul style="list-style-type: none"> • regaining possession • preventing scoring opportunities^a • communication with teammates

^a indicates skills recognised as involving a significant decision-making process.