Quality assurance in assessment: An introduction to this special issue

Introduction

Since assessment steers student's learning, it is important to design quality assessments that are well-aligned with curricular goals. If this is not the case, the form of assessment generally dominates and can lead to undesirable learning strategies by students (Cizek, 2001; Frederiksen, 1984). For a long time assessment has – both in research and practice – been approached as a psychometric issue, where reliability and validity were regarded as the most important quality criteria. In the last decade, however, assessment is increasingly approached as an educational design issue (Schuwirth and Van der Vleuten, 2005). In this design perspective, assessment is seen as the backbone in the design of learning environments, where the constructive alignment between learning, instructional approaches and assessment needs to be assured (Biggs, 1996). As a consequence, the sole function of assessment is not anymore to measure cognitive learning outcomes, but also to enhance students' metacognitive learning. This enriched perspective on assessment implies that assessment has multiple purposes.

The first purpose is to optimise sound decisions of students and to determine if learners achieved certain curricular goals. This perspective is referred to as *assessment of learning.* The second purpose is to use assessment results for subsequent learning. Because the focus of assessment is its effect on learning, this perspective is referred to as *assessment for learning* or formative assessment (Wiliam, 2011). In assessment for learning, learners and teachers interact by means of for example self-, peer- and co-assessments (Sluijsmans, Dochy, & Moerkerke, 1999). Providing effective feedback that is helpful for the learner is essential in assessment for learning (Shute, 2008). The third purpose is to use assessment as learning activities. In this *assessment as learning* perspective, the decision function and the learning function merge. In both assessment for and as learning, learners are encouraged to become self-regulated learners, who are able to plan, monitor and evaluate their own learning (process and outcomes) (Clark, 2012). To guide learners in the process of self-regulation, they are provided with feed-up (Where am I going?), feed-back (How did it go?) and feed-forward (Where to next?) (Hattie & Timperley, 2007).

The transition from a testing culture where assessment solely focuses on psychometric measurement of learning, to an assessment culture where assessments are used to stimulate (self-regulated) learning, requires an updated view to the question of quality in assessment. In the following, a framework to define assessment quality is presented.

A framework to define assessment quality

The traditional approach to assessment quality mainly focuses on the psychometric criteria validity and reliability as leading criteria. In this view, assessment quality is fully related to the design of assessments of learning, which can be labelled as an instrumental approach to quality. This concept of assessment quality needs reconsideration according to contemporary views on learning and assessment where assessment also is used to enhance learning. This means that assessment quality is a multifaceted concept, which requires careful analysis. In the search to define quality, the quality pyramid of assessment can serve as a useful framework. This pyramid was originally developed by Joosten-ten Brinke (2011) and extended by Sluijsmans, Peeters, Jakobs, and Weijzen (2012). The pyramid is developed based on literature on quality criteria for assessments (i.e., Baartman, Bastiaens, Kirschner, & Van der Vleuten, 2006; Boud & Associates, 2010; Brookhart, 2011; Downing & Haladyna, 1997; Messick, 1995; Schuwirth and Van der Vleuten, 2005; Stiggins, 2009) and distinguishes six quality entities: assessment tasks, assessments, assessment programme, assessment policy, assessment literacy and assessment organisation (see Fig. 1). The purpose of the quality pyramid is to approach assessment quality from a holistic perspective, where the quality of assessment is determined by the weakest link. This means that if the assessment programme for example is poorly designed, this also affects the quality of the other entities of the pyramid (this is illustrated by the bi-directional arrow on the left side of the pyramid in Fig. 1).

Assessment tasks

The entity assessment tasks refers to every task, assignment or question used in an assessment, whether this is a multiple choice
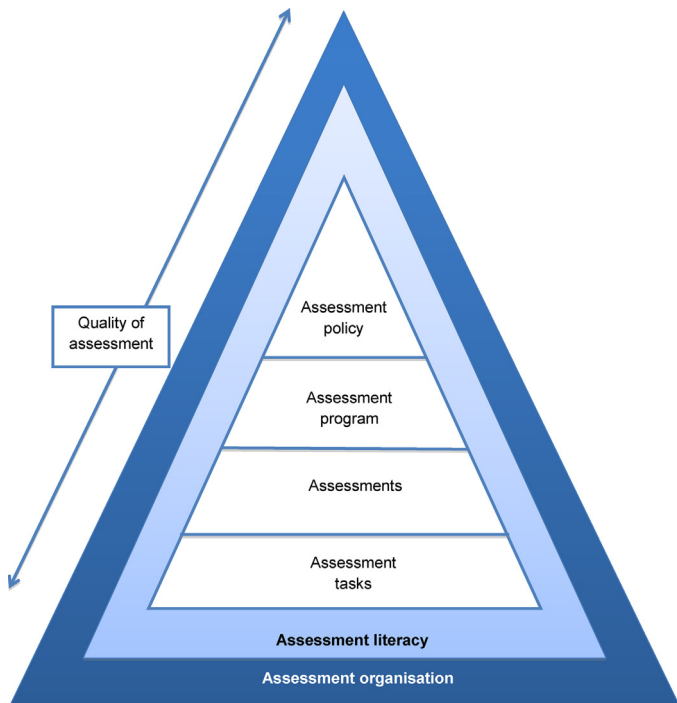
Fig. 1. Quality pyramid of assessment (Joosten-ten Brinke, 2011; Sluijsmans et al., 2012).

question, an essay question or an assignment to write a paper. Important quality criteria of assessment tasks are relevance, objectivity, efficiency and difficulty. Relevance means that the intended measure should correspond with the observed measure. Objectivity means that the interpretation of items is independent of the item constructor or assessor. Efficiency means that the most efficient item form should be selected if there are more equal options to choose from. Finally, the difficulty level of an item should be acceptable, given the required educational level. Quality of assessment tasks is assured when assessment tasks are designed in a systematic manner according to iterative process of preparation, implementation and evaluation (Downing & Haladyna, 1997).

Assessments

The entity 'assessments' includes all assessment methods that are used to measure if and to what extent the learners reach curricular goals. Outcomes of the design of assessments are blueprints, the assessment tasks, model answers, scoring rubrics, instructions, etcetera. Examples of assessments are simulations, portfolios and performance assessments. Important quality criteria of these assessments are the utility, validity and reliability of the assessment mode. This means that assessments should be efficient and fair, that they measure what is intended to be measured and that the assessment results are consistent. Quality of assessments is assured when assessments are designed in a systematic manner according to a cycle. An example of a cycle for the design of assessments is provided by Birenbaum, Kimron, and Shilton (2011). This cycle consists of the following steps: (1) planning (setting goals, defining objectives); (2) designing tools to elicit learner's understanding; (3) evidence collection (including provision of accommodations when needed); (4) interpretation (estimating the gaps between intended and obtained outcomes and generating feedback to the learners and the teacher); (5) utilisation (taking measures, where needed, to close the gaps); and (6) evaluation (assessing the effectiveness of those measures in

closing the gaps). Other activities in this cycle could be peer reviewing of assessment tasks, piloting, developing scoring standards, checklists, or scoring rubrics, training of assessors, and choosing an appropriate standard setting method (Cizek, 2001).

Assessment programme

Because every single assessment task has its limitations, it is preferable not to optimise individual assessment tasks, but to optimise all assessment tasks in a curriculum in an assessment programme (Van der Vleuten and Schuwirth, 2005). In a programme of assessment, methods of assessment are purposefully and carefully selected and organised, aiming at an optimal positive effect on learning (Schuwirth and Van der Vleuten, 2005). Baartman et al. (2006) developed a set of twelve criteria of assessment programmes and a self-evaluation instrument to judge the quality of these programmes (Baartman, Prins, Kirschner, & Van der Vleuten, 2011). These criteria are presented in Fig. 2. An elaboration on these criteria can be found in Baartman et al. (2006).

Assessment policy

Assessment policy includes the agreements – both on content and procedures – concerning assessment quality. The agreements are partly set by the government. The National Accreditation system of the Netherlands and Flanders for example, which is used as a standard to assess the quality of educational institutions, formulated three standards strongly related to assessment quality: (1) the intended learning outcomes of the programme are translated to the curriculum design, (2) the curriculum, staff and programme-specific services and facilities enable students to master the intended learning outcomes, and (3) there is an adequate assessment system that proves how the intended learning outcomes are achieved (NVAO, 2011). The latter standard is a so-called 'knock-out criterion', meaning that a low score on this standard implies a negative judgment for the whole programme. Institutions increasingly describe their agreements regarding assessment quality for internal and external purposes.
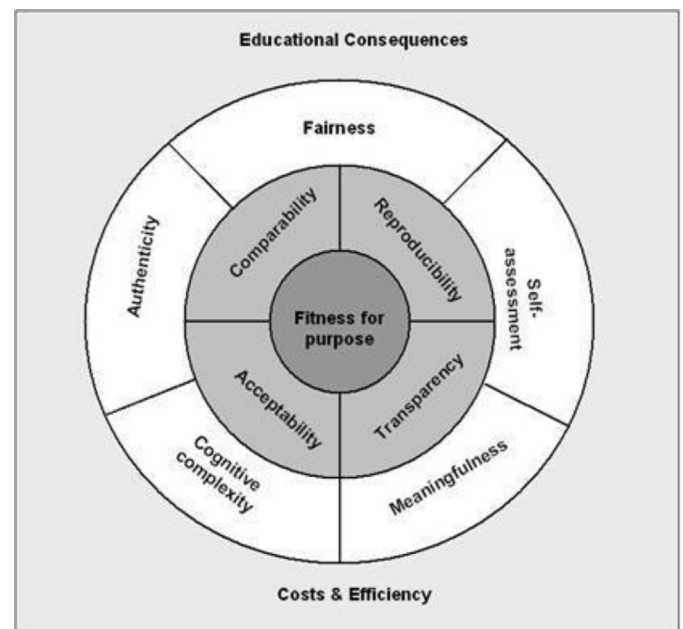

Fig. 2. The wheel of competency assessment (Baartman et al., 2006).