

Quality assurance in assessment: An introduction to this special issue



Keywords:

Quality of assessment
Formative assessment
Summative assessment
Quality framework

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 (William, 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

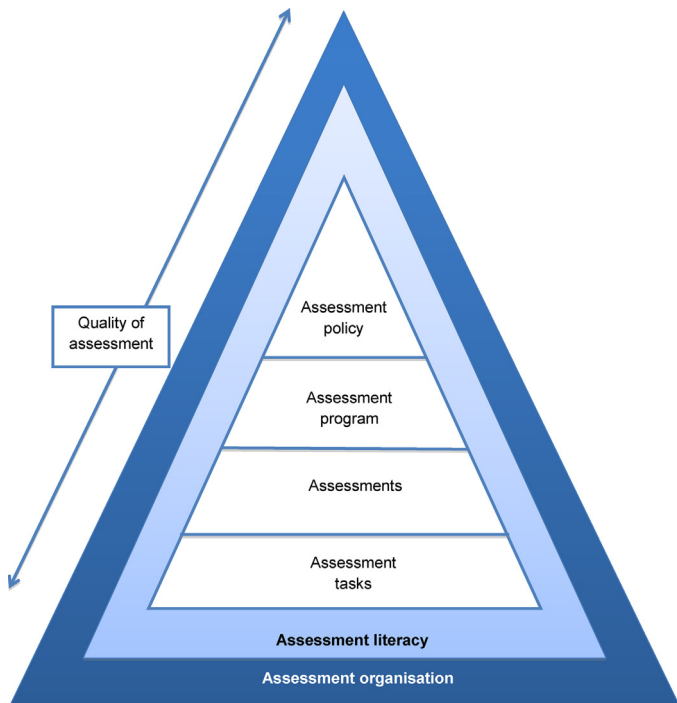


Fig. 1. Quality pyramid of assessment (Joosten-ten Brinke, 2011; Sluismans et al., 2012).

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.

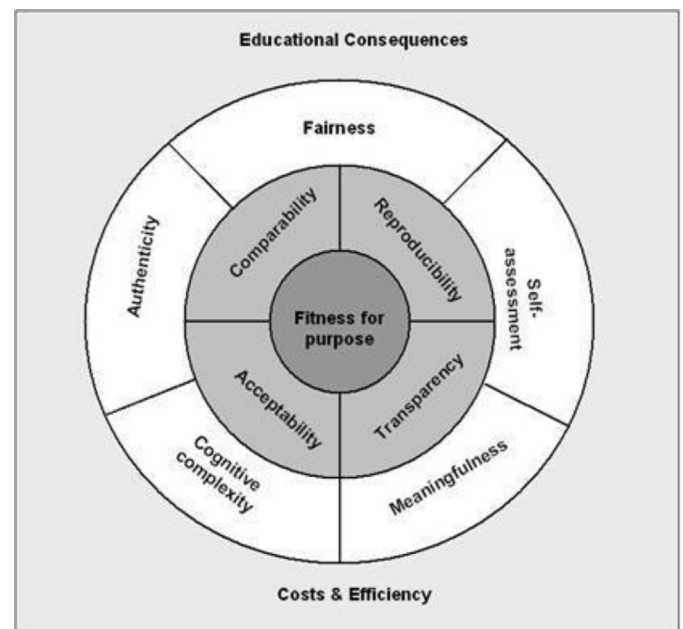


Fig. 2. The wheel of competency assessment (Baartman et al., 2006).

Assessment literacy

Assessment literacy refers to the understanding and appropriate use of assessment practices including the knowledge of the theoretical and philosophical underpinnings involved in the measuring of students' learning (DeLuca & Klinger, 2010). Assessment literacy of teachers highly impacts the quality of assessment. In general, teachers are poorly educated in assessment – especially how to practice assessment for and as learning – and as a consequence lack the fundamentals in how to design assessments that yield sound decisions about learning and foster student learning (Popham, 2009; Stiggins, 2009). Research showed that teachers demonstrate a low level of knowledge in educational assessment (for example, Alkharusi, Aldhafri, Alnabhani, & Alkalbani, 2012; Mertler & Campbell, 2005). Balancing assessment of, for and as learning requires a complex set of new knowledge and skills and can therefore be regarded as a complex task for teacher educators. Moreover, the possibilities of e-assessment will pose new demands on teachers, for example the use of digital testing and the use of e-portfolios.

Assessment organisation

Assessment organisation refers to the definition of all roles, tasks and responsibilities of all who are involved in assessment, and the way that they collaborate and trained for these tasks. In the Netherlands for example, examination committees in higher education do not only have procedural responsibility for the assessment programme, but also have the responsibility to assure the quality of assessments. Assessment organisation also refers to the rules and procedures regarding the design and evaluation of assessment tasks, assessments and assessment programmes.

The presented quality framework can be helpful in determining the strengths and weakness in curricula. For this special issue, it will be used to frame each contribution.

Overview of the contributions to this special issue

This special issue presents six studies in which quality assurance in assessment is operationalised in multiple variables that relate to the presented quality framework (Fig. 1).

The special issue starts with the paper of Smit and Birri on rubrics as assessment tools of and for learning. They conducted a study in which mathematical reasoning competencies in primary school were assessed by means of using standards-based rubric. The study aims to evaluate the use of this tool in supporting teachers to teach and to help them assess student learning aligned to standards. Results show that working with the rubric fostered the teachers' and the students' understanding of the standard, enabled the students to self- and peer-assess and allowed teachers to provide effective feedback.

The second paper, written by Sheard and Chambers, builds further on the assessment for learning purpose in primary schools and investigates how using an online resource with learner response devices meets quality assurance criteria for formative assessment. Grammar learning of pupils was central in the technology-enhanced formative assessment tool which provided instantaneous feedback to learners and teachers about student's understanding of the grammatical concepts.

Van der Kleij, Eggen and Engelen also focus their research on technology-enhanced assessments. Serving also summative functions of assessment, the pupil-monitoring Computer Program LOVS supports users when interpreting test results. This validity issue is usually underexposed in research on assessment. The authors take on the challenge to redesign and evaluate the score reports that are generated by the software.

The involvement of users and experts in the design process tended to be essential.

Still in the area of technology-enhanced assessments, Struyven, Blicke and De Roeck, evaluated the use of e-portfolio to develop and assess pre-service teachers' teaching competences during teaching practice internships in schools. The results show how diverse stakeholders can perceive the effects and use of the e-portfolio in different (even contradictory) ways, and how assessment literacy of teacher trainers may interfere with the intended implementation and learning processes associated with the tool.

In the fifth paper, Tierney further elaborates on teachers' assessment literacy, in particular on the concept of 'fairness'. By means of a multiple case study, teachers responded to fairness issues in written vignettes, and then discussed their concerns and gave recommendations for fair assessment. The most prominent issues of fairness involve students' opportunities to learn and demonstrate learning, transparency, the classroom environment, critical reflection, and the tension between equal and equitable treatment in assessment.

Finally, Leong focuses on a nation-wide initiative in Singapore 'holistic and balanced assessment'. The findings from one of the 'high-achieving' case-study Singaporean teachers reveals that any quality assurance framework or guideline for evaluating teachers' assessment practices needs to be sensitive to their intentions, meaning and context of teaching.

Overviewing the six contributions, it can be noticed that most articles focus on one of the entities of the model, with the assessment entity being the most prevalent one. The contributions in this special issue demonstrate the complexity of capturing a definition of quality assurance in assessment. This implies that more research is needed what quality assurance implies form a more holistic perspective, that is, taking into account each entity in the quality pyramid. Interestingly, every contribution pinpoints the importance of assessment literacy of teachers, but also emphasises the importance of assessment literacy of students. When teachers and students are not trained in assessment, this impairs the quality of assessment in terms of sound decisions and benefits for learning. Effective and contemporary models of teachers' professional learning appear to be essential in assuring quality in assessment (DeLuca, Luu, Sun, & Klinger, 2012).

In conclusion, assessment is an essential task of any professional who has to make sound decision about students' learning. Assessments without any quality control, as it is often the case in educational practice, risk poor quality leading to invalid assessment and unintended learning processes and outcomes. Awareness about what constitutes assessment quality is an essential condition to assure assessment quality on the long term. The quality pyramid of assessment can provide schools and institutions a helpful tool to analyse the quality of assessment on each of the six entities. In practice, it can be observed that quality assurance is often translated into criteria, procedures and checklists that are administered before, during and after an assessment. This special issue intends to contribute to the awareness that quality assurance in assessment is not only a matter of control (using checklists and procedures), but mainly guaranteed by the quality of the users (students and teachers) and the way this is reflected in the spirit instead of the letter of assessment.

Brussels, August 29–31, 2012

Starring the Special Interest Group SIG 1 Assessment and Evaluation biennial conference of EARLI, the idea of this special issue was born. 'Linking multiple perspectives on assessment' had been the conference theme and outcome of this 3-day gathering of assessment and evaluation researchers from around the world. One of the 'perspectives' that linked many contributions to one

another was the concept of 'quality' in assessment. Participants of the SIG1 conference were invited to submit contributions to a special issue that would have this focus as a 'linkage'. Proposals were received, feedback let to first drafts, reviewer comments to revisions and re-evaluations to final acceptations for publication in this special issue that we are proud to present today.

Contributions

1. Editorial: Sluijsmans and Struyven
2. Smit and Birri
3. Sheard and Chambers
4. Van der Kleij, Eggen and Engelen
5. Struyven, Blicek and De Roeck
6. Tierney
7. Leong

Acknowledgements

The guest editors, Katrien Struyven and Dominique Sluijsmans, would like to thank the authors, the reviewers and Prof. Dr. Peter van Petegem, the editor of Studies in Educational Evaluation, for their contribution and support to this special issue.

References

- Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning*, 1(2), 217–232.
- Baartman, L. K. J., Bastiaens, T., Kirschner, P. A., & Van der Vleuten, C. P. M. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programmes. *Studies in Educational Evaluation*, 32, 153–177.
- Baartman, L. K. J., Prins, F. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2011). Self-evaluation of assessment programs: A cross-case analysis. *Evaluation and Program Planning*, 34, 206–216.
- Biggs, J. (1996). Enhancing learning through constructive alignment. *Higher Education*, 32, 347–364.
- Birenbaum, M., Kimron, H., & Shilton, H. (2011). Nested contexts that shape assessment for learning: School-based professional learning community and classroom culture. *Studies in Educational Evaluation*, 37(1), 35–48.
- Boud, D., & Associates (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.
- Brookhart, S. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12.
- Cizek, G. J. (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205–249.
- DeLuca, C., Luu, K., Sun, Y., & Klinger, D. A. (2012). Assessment for learning in the classroom: Barriers to implementation and possibilities for teacher professional learning. *Assessment Matters*, 4, 5.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419–438.
- Downing, S. M., & Haladyna, T. M. (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education*, 10(1), 61–82.
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *Review of American Psychologist*, 39, 193–202.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Joosten-ten Brinke, D. (2011). *Eigentijds toetsen en beoordelen*. [Contemporary assessment]. Lectorale rede. Tilburg, Netherlands Fontys Lerarenopleiding Tilburg.
- Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment knowledge inventory. Paper presented at the meeting of the American Educational Research Association.
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- NVAO (2011). *Assessment frameworks for the higher education accreditation system*. NVAO http://www.nvao.net/page/downloads/DEFINITIEVE_KADERS_22_november_2011_English.pdf.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4–11.
- Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2005). A plea for new psychometric models in educational assessment. *Medical Education*, 40, 296–300.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
- Sluijsmans, D. M. A., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self- peer- and co-assessment. *Learning Environments Research*, 1, 293–319.
- Sluijsmans, D. M. A., Peeters, A., Jakobs, L., & Weijzen, S. (2012). De kwaliteit van toetsing onder de loep [A closer look at the quality of assessment]. *Onderwijsinnovatie*, 4, 17–25.
- Stiggins, R. J. (2009). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment*. New York, NY: Routledge.
- Van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessment of professional competence: from methods to programmes. *Medical Education*, 39, 309–317.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3–14.

Dominique M.A. Sluijsmans^{b,c}
Katrien Struyven^{a,*}

^aVrije Universiteit Brussel (VUB), Educational Sciences Department,
1050 Elsene, Brussel, Belgium

^bZuyd University of Applied Sciences, Heerlen, The Netherlands

^cSchool of Health Professions Education, Maastricht University,
Maastricht, The Netherlands

*Corresponding author

E-mail address: katrien.struyven@vub.ac.be (K. Struyven).

Received 26 August 2014

Accepted 26 August 2014

Available online 16 September 2014