



OPEN ACCESS

EDITED BY

Gavin T. L. Brown,
The University of Auckland, New Zealand

REVIEWED BY

Vanessa Scherman,
University of South Africa, South Africa
Ramashago Mphahlele,
University of South Africa, South Africa

*CORRESPONDENCE

Lonneke H. Schellekens
✉ l.h.schellekens@uu.nl

RECEIVED 18 January 2023

ACCEPTED 23 May 2023

PUBLISHED 15 June 2023

CITATION

Schellekens LH, Kremer WDJ, Van der Schaaf MF, Van der Vleuten CPM and Bok HGJ (2023) Between theory and practice: educators' perceptions on assessment quality criteria and its impact on student learning. *Front. Educ.* 8:1147213. doi: 10.3389/educ.2023.1147213

COPYRIGHT

© 2023 Schellekens, Kremer, Van der Schaaf, Van der Vleuten and Bok. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Between theory and practice: educators' perceptions on assessment quality criteria and its impact on student learning

Lonneke H. Schellekens^{1,2*}, Wim D. J. Kremer¹,
Marieke F. Van der Schaaf³, Cees P. M. Van der Vleuten⁴ and
Harold G. J. Bok¹

¹Department of Population Health Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands, ²Educational Consultancy and Professional Development, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, Netherlands, ³Utrecht Center for Research and Development of Health Professions Education, University Medical Centre Utrecht and Utrecht University, Utrecht, Netherlands, ⁴Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

Introduction: The shift toward an assessment *for* learning culture includes assessment quality criteria that emphasise the learning process, such as transparency and learning impact, in addition to the traditional validity and reliability criteria. In practice, the quality of the assessment depends on how the criteria are interpreted and applied. We explored how educators perceive and achieve assessment quality, as well as how they perceive assessment impact upon student learning.

Methods: We employed a qualitative research approach and conducted semi-structured interviews with 37 educators at one Dutch research university. The data were subsequently analysed using a template analysis.

Results: The findings indicate that educators predominantly perceive and achieve assessment quality through traditional criteria. The sampled curricular stakeholders largely perceived assessment quality at the course level, whilst few specified programme-level quality criteria. Furthermore, educators perceived the impact of assessment on student learning in two distinct ways: as a source of information to monitor and direct student learning, and as a tool to prompt student learning.

Discussion: The shift toward a culture of assessment *for* learning is not entirely reflected in educators' current perceptions. The study's findings set the stage for better assessment quality and alignment with an assessment for learning culture.

KEYWORDS

assessment quality, assessment quality criteria, assessment for learning, higher education, learning impact

Introduction

In higher education, a changed perspective on the function of assessment has evolved; this is reflected in the shift from the assessment *of* learning towards an assessment *for* learning culture (Shepard, 2000; Van der Vleuten et al., 2012). Such a shift implies that assessment is viewed as an integral part of students' learning process and not only as its endpoint (Biggs et al., 2022). If and how this shift occurs highly depends on educators' perceptions regarding assessment and assessment quality. The quality of the assessment is not just a characteristic of

the assessment (instrument) itself, but also depends on how it is interpreted and used in an educational setting (Kane, 2008). To understand more clearly how assessment for learning is achieved in practice, it is necessary to examine the perceptions of educators regarding assessment and the quality requirements they take into account. Perceptions refer to how individuals interpret and assign meaning to environmental information (Pickens, 2005).

The assessment for learning practices (that integrate student learning) involves assessing quality criteria that emphasise the learning process, such as “transparency” and “learning impact,” in addition to the traditional criteria of validity and reliability (Baartman et al., 2007a; Tillema et al., 2011; Maassen et al., 2015). Although relevant, it appears to be difficult to acquire insight into educators’ perceptions and practices regarding assessment quality criteria (Sridharan et al., 2015). Indeed, educators do not always perceive assessment criteria in the same way and thus can assign different meanings to the same assessment criteria (Van der Schaaf et al., 2012). Research on educators’ perceptions of assessment quality in higher education is scarce (Kleijnen et al., 2013; Opre, 2015) and most studies examining perceptions of assessment quality have instead focused on students (Nabaho et al., 2011; Gerritsen-van Leeuwenkamp et al., 2019; Ibarra-Sáiz et al., 2021). To fill this gap, it is essential to investigate how educators perceive and achieve assessment quality criteria. In addition, given the significance of the criterion learning impact for developing an assessment environment that places student learning at the centre, it is also essential to analyse how educators perceive that the criterion learning impact contributes to student learning.

This study focuses on two research questions: 1) What are educators’ current perceptions of assessment quality and how is assessment quality achieved? and 2) What are educators’ perceptions on how assessment impacts student learning?

We use the term educators to refer to stakeholders who are formally responsible for ensuring assessment quality at the course or curriculum level (i.e., teachers, course and programme leaders, and the Board of Examiners). Course-level stakeholders were asked how they perceived and achieved assessment quality at the course level. Curriculum-level stakeholders were asked how they perceived and achieved assessment quality at both the course and the curriculum level. These assessment stakeholders may have varying perceptions of assessment quality and its impact on student learning. By analysing their perceptions, a comprehensive picture of the current state of assessment quality that may influence assessment practice is provided (Kane, 2008). The acquired knowledge contributes to the literature’s understanding of how assessment quality criteria are applied in daily practice and what the consequences for an assessment *for* learning culture could be.

Theoretical framework

Assessment quality criteria

The shift towards assessment for learning has an impact on assessment practices, as well as on the choice of relevant criteria for assessment quality (Baartman et al., 2007c; Tillema et al., 2011). Assessment quality comprises the quality of all aspects of assessment

practices, including test items, tasks, the process of assessing, course assessments, the assessment programme, policies, and administration of the process (Gerritsen-van Leeuwenkamp et al., 2017). Assessment quality is essential to make meaningful judgments about students’ learning processes and performance (Borghouts et al., 2017). Review studies typically identify four assessment quality criteria for assessing assessment quality: 1) validity, 2) reliability, 3) transparency, and 4) learning impact (Van der Vleuten and Schuwirth, 2005; Baartman et al., 2007c; Poldner et al., 2012; Maassen et al., 2015; Gerritsen-van Leeuwenkamp et al., 2017). Traditionally, the goal of assessments was to provide a standardised and objective judgement when testing a student’s knowledge at a specific point in time (Benett, 1993; Brown, 2022); for example, as measured by a multiple choice knowledge test. To ensure assessment quality for the function of these tests, the quality criteria of validity and reliability were often used. Validity means that the test measures the intended construct and that administering the test produces the desired impact (Messick, 1989). Reliability is the extent to which a test measures consistency over time and across assessors (Dunbar et al., 1991). Current educational practices demand that students demonstrate not only a mastery of the required knowledge and understanding of a particular professional practice, but also proficiency in higher cognitive levels, such as the ability to think critically, apply knowledge in practice, and monitor and regulate their own learning (Boud and Soler, 2016). This requires alternative forms of assessment to the traditional knowledge tests, such as performance assessments or portfolio evaluations in which not only the outcome but also the learning process is emphasised. Consequently, besides validity and reliability, additional criteria of transparency and learning impact are often used. Transparency means that the purpose of the assessment, the assignment, and the evaluation criteria are clear, understandable, and feasible for those who administer them (Brown, 2005; Maassen et al., 2015). Learning impact (also known as educational consequences or educational impact in the literature) refers to the intended and unintended effects of an assessment on students’ learning (Dierick and Dochy, 2001). From an assessment for learning standpoint, learning impact aims to ensure that assessment contributes as much as possible to students’ learning and development (Baartman et al., 2007a; Dochy, 2009). When students gain knowledge through assessments or gain awareness of their strengths and weaknesses, assessments can become relevant and meaningful for them (Maassen et al., 2015).

Teacher perceptions on how assessment impacts student learning

The impact of an assessment on student learning focuses on the interpretation of the assessment score and the consequences associated with it (Baartman et al., 2007c). Teachers frequently use the score to evaluate student performance after a formal learning activity in order to determine whether students have gained the necessary knowledge, skills, and competencies to complete a course or learning trajectory; i.e., for the purposes of assessment of learning. In assessment cultures with a focus on the assessment of learning, teachers train students to pass rather than to improve their learning (Harlen, 2005). On the other hand, a score can be used for purposes of assessment for learning by using the score to indicate a student’s level of mastery and potential

areas for improvement (Heitink et al., 2016; Black and Wiliam, 2018). These two purposes are not mutually exclusive and can be classified into a continuum from reproductive (assessment of learning) conceptions with an emphasis on the measurement of students' reproduction of correct information, to more transformational (assessment for learning) conceptions with an emphasis on the development of students' thinking and understanding (Postareff et al., 2012).

Methods

Research design

To maximise the understanding of educators' perceptions in higher education on assessment quality criteria and its impact on student learning, we designed an explorative qualitative study using a directed content analysis approach (Stebbins, 2001; Hsieh and Shannon, 2005). With a directed approach, the analysis starts with a theory or relevant research findings to guide the creation of an initial coding template. Individual semi-structured interviews were conducted with educators. In general, the interview questions focus on how participants perceive and achieve assessment quality (what is assessment quality and how is it achieved in practice?), and on learning impact (how does assessment contribute to students' learning?).

Participants

Purposeful sampling (Teddle and Yu, 2007; Palinkas et al., 2015) was used to select undergraduate and graduate study programmes from three faculties of a Dutch university. The study programmes were chosen to reflect the wide range of disciplines offered at the university and to reflect distinct assessment programmes (i.e., varying in assessment methods) to analyse the university's breadth and heterogeneity. The educational philosophy of the university emphasises active student participation and values the implementation of continuous feedback to ensure that students stay on track and excel (Utrecht University Education Guideline, 2018). Each faculty adopts the university's educational philosophy in its own distinct way. The study programmes participating in the study have curriculums divided into four or five terms per academic year. Typically, each term consists of two or more courses, and each course is assessed by two or more summative high-stakes assessments (e.g., written exams, oral exams, papers, practical exams, portfolio assessments, etc).

We sampled participants using procedures for maximum variation sampling (Palinkas et al., 2015) determined by: (a) formal responsibility for assessment quality in the study programme (e.g., educators who are responsible for assessment quality at the course level, such as teachers or course coordinators, and educators who are responsible for assessment quality at the curriculum level, such as programme coordinators or Board of Examiners members), (b) years of experience in ensuring assessment quality (ranging from less than one year to more than 20 years of experience), and (c) variation in disciplines. A total of 37 educators participated in individual semi-structured interviews. The characteristics of the participants are shown in Table 1.

TABLE 1 Characteristics of participants (n=37).

Formal responsibility for assessment quality	
Participants with responsibilities at the course level	n = 22
Participants with responsibilities at the programme level	n = 15
Years of experience	
≤2 years	n = 10
>2 years and <10 years	n = 16
≥10 years	n = 11
Discipline	
Social and behavioural sciences	n = 11
Veterinary medicine	n = 12
Geosciences	n = 14

Data collection and analysis

The programme coordinators of the study programmes that participated were asked to select five to seven teachers responsible for course assessment with diverse experience in preparing and administering assessments. In addition, representatives of the study programme's Board of Examiners were invited to an interview. The first author (LS) approached all participants through email with information about the study and an invitation to participate voluntarily. The interviews took place at participants' workplace and lasted between 15 and 45 min. All interviews were recorded and transcribed verbatim.

To answer the first research question, the interview data were analysed using a template analysis that distinguished *a priori* themes (Brooks and King, 2014; Brooks et al., 2015). We chose a template analysis with *a priori* themes because it provided a framework that enabled us to consistently compare between the theoretical framework of quality criteria (i.e., validity, reliability, transparency, and learning impact) and the participants' actual experiences. The content of the *a priori* themes was developed using scientific literature review studies on the quality criteria for assessments that serve both formative and summative functions in higher education (Baartman et al., 2007a; Dochy, 2009; Tillema et al., 2011; Poldner et al., 2012; Maassen et al., 2015; Gerritsen-van Leeuwenkamp et al., 2017). The aim of the analysis was to provide a comprehensive and concise overview of how quality criteria were perceived and achieved in relation to the four identified criteria, as outlined in the theoretical framework. However, a potential limitation of a template analysis is the risk of missing themes that do not fit within the framework (Brooks et al., 2015). During data analysis, the researchers were aware of this limitation and regularly assessed whether more themes could be detected. Participants did not mention other major themes. The aim of this study was also to find variations in the participant descriptions. As a result, we identified subthemes within the *a priori* themes. To answer the second research question, we used a directive content analysis approach (Hsieh and Shannon, 2005). The text fragments of the interview data were coded inductively and the names of the codes were based on the content of the text fragments and thus bore a resemblance to the original data (Cohen et al., 2018). To ensure consistency and coverage of the codes, text fragments were re-read and codes were re-assigned several times.

To ensure the trustworthiness of the coding, two authors (LS and WK) independently analysed the first twelve interviews to define and agree on the coding template (Brooks et al., 2015). Appendix I provides the initial template. First, they coded the data and found sections in the transcripts where participants expressed their perceptions of both assessment quality and its impact on student learning (i.e., indicating what quality looks like and what they believe is essential). Second, they determined the sections in the transcripts where participants mentioned how assessment quality was achieved in practice. After each interview, LS and WK discussed the coding until they were in agreement that the template covered all sections of the transcripts (Brooks et al., 2015). LS coded the remainder of the interviews using qualitative data analysis software (NVivo 12 Pro). Minimal modifications were required to finalise the template. To answer research question 1, LS provided an overview of the subthemes within the *a priori* themes. Next, to answer research question 2, LS examined, compared, and conceptualised the codes on their content in order to categorise the codes into themes. The findings were discussed by the whole research team until a consensus was reached.

Ethical considerations

Ethical approval was given by the ethical board of the Faculty of Social and Behaviour Sciences of Utrecht University and registered under nr. UU-FETC19-022.

Results

The results for each research question are presented separately. First, results will be presented on how educators perceive assessment quality and examples are given of how the criteria are achieved in practice (Research Question 1). Then, results will be presented on how educators perceive the impact of assessment on student learning as denoted by the *learning impact* criterion (Research Question 2).

Participants' perceptions of assessment quality and its achievement in practice

Participants could give several responses to the question "what is assessment quality." More than three-quarters of participants' responses related to traditional assessment criteria of validity and reliability. Less than a quarter of participants' responses addressed criteria that emphasise the learning process: transparency and learning impact. Although curriculum stakeholders were asked how they perceived and achieved assessment quality at both the curriculum and course levels, few identified programme-level quality criteria, which were only highlighted within the theme of validity.

Validity

In relation to the validity criterion, participants' perceptions mainly referred to two subthemes:

Assessment is of high quality if there is alignment

Participants underlined the importance of alignment at both the curriculum level (i.e., alignment between course learning objectives and

curriculum learning outcomes) and the course level (i.e., alignment between course learning objectives, learning activities, and assessment):

"I believe that the most essential aspect of assessment quality is whether or not it is aligned with the learning objectives, that is, whether or not what you want students to learn is assessed." (Participant C9)

Assessment is of high quality if there is content coverage

Participants emphasised the need to ensure that the assessment measures what it intends to measure. At the course level, the assessment should accurately reflect the content that students have studied. At the curriculum level, the assessment programme should be designed in such a way that it is manageable for students; namely, not too difficult and not too easy:

"There should be a clear relationship and distribution of questions across the topics in the assessment. Therefore, it is inappropriate to make students read five chapters and just ask questions about chapter five." (Participant O10)

Examples of achievement. To achieve valid assessments in practice, most participants commonly use an assessment blueprint (i.e., table of specifications). The assessment blueprint provides information about alignment and content coverage. At the course level, an assessment blueprint is constructed to ensure that the assessment content corresponds to the learning activities and cognitive level specified. At the curriculum level, an assessment blueprint is used to determine whether and how the curriculum programme outcomes are assessed in courses.

Reliability

In relation to the reliability criterion, participants mainly referred to the subthemes:

Assessment is of high quality if there is consistency

Participants highlighted that assessments should be scored consistently both between various assessors and over time so that assessment results can be evaluated in the same manner. In addition, participants emphasised that assessments should be designed and administered consistently to ensure that questions are clear, assessments can be finished within the allotted time, and administration is well-organised:

"It should just be a proper assessment, with no typos, clear questions, and smooth running in the digital application where it is assessed, etc." (Participant C16)

Assessment is of high quality if it differentiates between students

Participants emphasised the importance of an assessment being unambiguous about whether the student has mastered the course material so that the decision about a student's performance is fair and accurate. The assessment should distinguish between more and less competent students:

"A quality assessment evaluates students' knowledge and skills... in a way that distinguishes between those who truly comprehend the subject or have mastered the skill and those who have not yet sufficiently done so." (Participant C17)

Examples of achievement. To achieve reliable assessments in practice, participants use a variety of methods. Participants most frequently mentioned that colleagues were involved (i.e., the four-eyes principle) to develop and evaluate test items and/or a scoring and grading scheme (such as a rubric). Next, participants mentioned that they analysed their written tests for psychometric measures (e.g., value of p , item-total correlation, Cronbach's alpha). Some participants stated that they organised a calibration session as a method to ensure reliable scoring. Here, the assessors discuss several cases and review the assessment scoring to verify if they are on the same page.

Transparency

A minority of participants' responses to the question "what is assessment quality" addressed the transparency and learning impact quality criteria. Participants indicated that a high-quality assessment is transparent if the assessment procedures, the content of the assessment, and the scoring are clear to students. Thus, transparency was viewed primarily through the eyes of the student, rather than the assessor:

"Students should have a clear understanding of how and what they will be assessed on in advance." (Participant C1)

Examples of achievement. To achieve transparent assessments in practice, participants communicated assessment procedures and criteria to students in advance; for example, via a student manual. Participants also reused questions or exams to allow students to prepare in advance and organised a perusing session to allow students to review their assessment results.

Learning impact

Participants highlighted that a high-quality assessment impacts learning if it provides students with knowledge of how they perform and where they stand:

"I believe that when your assessment has quality, you are helping someone to develop continuously. And that they understand: 'Yes, this is where I still fall short, this is where I need to do something more'. Instead of: 'Oh, I received a passing grade, so it is done.'" (Participant O6)

Examples of achievement. To achieve assessments with learning impact in practice, participants mentioned that they provided feedback on concept papers and provided students with opportunities to self-assess; for example, via practice exams and questions.

Participants' perceptions on how assessment impacts student learning

In the second part of the study, we continued with a focus on the learning impact criterion and asked participants to describe how assessment contributes to student learning. Curriculum stakeholders were

also asked how the assessment programme contributes to student learning. Just over a quarter of participants responded that they did not have a good understanding of how course assessments or the assessment programme impact student learning. A few participants perceived assessment as irrelevant to student learning:

"I do not really believe that making the test... [the students] do not really learn anything from that." (Participant C24)

Programme stakeholders, in particular, found it difficult to describe how the assessment programme contributes to student learning:

"For the programme as a whole, I really do not know. If there is, what is the general view on that? I simply do not know that. No doubt it has been described somewhere, but I missed it." (Participant O12)

In the responses of those participants who were able to answer the question, we identified two subthemes: (a) assessment as an information source to monitor learning; and (b) assessment as a tool for learning.

Assessment as an information source to monitor learning

About half of participants' responses indicated that assessment contributes to students' learning when the assessment can be used as an information source to monitor and track students' learning:

"I think it's much more important to monitor what students do and how they develop during a course as much as possible. So in my view, a learning process is something that not only requires students to actively engage, but also to know if they are doing well or not well, and what they can improve." (Participant O04)

In addition, students must be able to demonstrate how they comprehended the course material, as a form of reflection on what they have learnt. The younger teachers in the sample placed a greater emphasis on the learning process because they learnt these assessment practices in recent teacher training.

Participants highlighted three aspects to achieve to ensure that assessment contributes to students' learning. First, give **constructive feedback** to inform students where they stand and what can be improved and include an opportunity to act on it. For example, by giving feedback on concept papers and not on final papers, and by using a rubric to inform students about the standards and their progress in these standards. Second, offer **formative, low-stakes assessments** to enable students' self-assessment and to monitor students' development as a teacher. For example, by discussing concepts with students and asking questions in working groups, or by providing practice exams and progress exams. Third, design assessments that **integrate theory with practice** to engage students in assessment. For example, by offering placements, authentic assessments, or case studies wherein knowledge has to be applied.

Assessment as a tool to force learning

The other half of the participants' responses indicated that assessment contributes to students' learning because taking a test

“forces” the student to learn. In this perspective, the assessment is used as a *tool* to exert pressure on students to learn the course material:

“That [how assessments contribute to students’ learning] is a difficult question to answer. But I mostly see assessment as a form of pressure, that students are going to learn. I notice this when we cover a topic or when we assign a piece of reading. Then students often ask ‘Is this part of the test?’. And when we respond ‘Yes, that is part of the test’ they conclude ‘well, then we will learn this’. (Participant C9)

“The fact that you administer tests increases the urgency for students to pay close attention. I believe that students go the extra mile when they know it will be assessed.” (Participant O1)

Some participants feared losing control over the student learning process while implementing assessment for learning practices because they lack confidence in student autonomy. Participants mentioned that assessments contribute to students’ learning when many high-stake assessments are offered during a course. By providing numerous assessments, the study load is structured and distributed, allowing students to engage in learning.

Discussion

The purpose of this study was to explore how educators at one Dutch research university from diverse faculties and with varying years of experience perceived and achieved assessment quality, and how they perceived assessment impact upon student learning. In the first part of the study, we examined how educators perceive and achieve assessment (programme) quality. Two results were most notable: 1) educators primarily perceive and achieve assessment quality through traditional criteria, and 2) they primarily perceive assessment quality at the course level. First, educators primarily perceived assessment quality through traditional criteria of validity and reliability. Participants indicated that an assessment is of good quality if there is alignment and content coverage (validity) and if the test can be administered and scored consistently and can differentiate between student performance (reliability). Few of the participants perceived assessment quality through the transparency and learning impact criteria, which relate to the student learning process and development. There appears to be a gap between theory and practice; many educators in our study are not yet aware of quality assessment criteria that relate to assessment for learning. These results contradict previous research that investigated teachers’ perceptions of quality criteria for a competence assessment programme and concluded that teachers perceived traditional criteria and criteria that emphasise the learning process as being equally important (Baartman et al., 2007b). However, the participants in that study were given a questionnaire including the criteria, while the participants in the present study were not given a frame of reference with assessment criteria in advance.

Second, although the curriculum stakeholders in the sample were responsible for assessment quality at both the course and curriculum levels, they primarily perceived and achieved assessment quality at the course level. Some participants described the assessment programme’s quality as “the sum of the quality of the course assessments.” They did not consider criteria related to the overall quality of the assessment programme, even though assessment criteria for the programme as a

whole were developed (Baartman et al., 2006). For example, the criterion learning impact can be applied at the programme level by identifying learning pathways and continuous feedback loops therein (Schellekens et al., 2022). Focusing on the assessment programme provides opportunities to improve student learning (Jessop et al., 2012; Malecka et al., 2022) and is more consistent with an assessment for learning culture (Medland, 2016).

In the second part of the study, we examined how educators perceive the criterion of learning impact and identified two themes: 1) the assessment serves as a source of information to monitor and direct student learning and 2) the assessment serves as a tool to prompt student learning. These findings are consistent with earlier research on how teachers conceptualise the purpose of assessment, as classified in a continuum ranging from transformational ‘assessment for learning’ conceptions with an emphasis on the development of student’s learning, to more reproductive ‘assessment for learning’ conceptions with an emphasis on the measurement of students’ learning (Postareff et al., 2012; Schut et al., 2020; Fernandes and Flores, 2022). Approximately half of participants perceived learning impact with aspects related to the first theme wherein student learning is central. These findings are encouraging since previous research showed that teachers’ conceptions of assessment influence their assessment practices (Xu and Brown, 2016). The findings are also in line with the university’s current educational policy which stimulates teachers to practice formative assessments (Utrecht University Education Guideline, 2018). The other half of the participants perceived learning impact with aspects related to the assessment of learning. Research reveals that teachers’ perceptions of assessment move towards the reproductive end of the continuum when they perceive learning as preparing students for (graded) high-stake assessments to ensure student performance (Schut et al., 2020). In this study, teachers mentioned that they wished to control assessment processes because they did not have confidence in student autonomy without controlling mechanisms such as (high-stakes) assessments. This indicates that assessments with a summative function still dominate the learning process. Consequently, more focus is needed on how we can integrate and balance both formative and summative purposes of assessment to facilitate an assessment for learning culture.

While all of the participants could describe the criteria of assessment quality and how these criteria were achieved, some participants mentioned that they did not have a good understanding of the learning impact criterion, i.e., how assessment impacts student learning. This may indicate a lack of knowledge about assessment for learning practices, which potentially minimises the learning benefits of assessment (Van der Vleuten et al., 2012). Tillema et al. (2011) review study determined that the learning impact criterion (titled educational consequences) can be achieved in many ways and is relevant when designing, administering, and evaluating an assessment. In order to address educators’ knowledge of quality criteria that take student learning into account, faculty development should place a greater emphasis on assessment literacy, i.e., the understanding of how to use assessment to improve student learning and achievement (Stiggins, 1995). In addition, faculty policy should emphasise assessment for learning quality criteria more explicitly in order to generate a sense of urgency to change current practices. For example,

by addressing these criteria more explicitly in quality assurance procedures. Policies that promote the use of assessment to improve learning may conflict with those that emphasise the use of assessment to measure achievement. Indeed, teachers working under these conditions may experience tensions between these opposing purposes of assessment; this may subsequently influence how they perceive and achieve assessment (Yates and Johnston, 2018). The findings of this study should be considered in light of a number of limitations. First, we did not differentiate between formative (low-stakes) and summative (high-stakes) assessment practices when examining the participants' perceptions of quality criteria. While research indicates that quality criteria rarely differ between summative and formative assessment contexts (Ploegh et al., 2009), this could have biased our findings. Second, although we included participants with a variety of roles, disciplines, and years of experience, we did not distinguish these groups when presenting the results. This is because our preliminary analyses revealed no differences between these groups, and we were primarily interested in the underlying perceptions that educators generally have of assessment quality criteria. However, assessment perceptions and practices can be influenced by disciplinary traditions and years of experience (DeLuca et al., 2018; Simper et al., 2021). Future research on this topic should therefore include a larger sample size from diverse universities.

In conclusion, during the past two decades scholars have advocated for a paradigm shift from the assessment of learning to assessment for learning, in which assessment is viewed as a tool to enhance student learning and wherein formative low-stakes and summative high-stakes purposes are balanced. Synthesising the perspectives of various stakeholder groups who are responsible for ensuring assessment quality provides the literature with a good understanding of how assessment quality criteria and its impact on student learning are perceived. The findings of our study demonstrate that educators' current perceptions do not fully represent the shift toward an assessment for learning culture. There appears to be a gap between theory and practice; indeed, many educators in our study were not yet aware of quality assessment criteria that relate to assessment for learning. Furthermore, educators' perceptions on how assessment impacts learning indicate that summative high-stakes assessments dominate the learning process as a tool with which to force students to learn. The study's findings set the stage for better assessment quality and alignment with an assessment for learning culture.

References

- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., and Van der Vleuten, C. P. M. (2006). The wheel of competency assessment. Presenting quality criteria for competency assessment programmes. *Stud. Educ. Eval.* 32, 153–170. doi: 10.1016/j.stueduc.2006.04.006
- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., and Van der Vleuten, C. P. M. (2007a). Evaluating assessment quality in competence-based education: a qualitative comparison of two frameworks. *Educ. Res. Rev.* 2, 114–129. doi: 10.1016/j.edurev.2007.06.001
- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., and Van der Vleuten, C. P. M. (2007b). Teachers' opinions on quality criteria for competency assessment programs. *Teach. Teach. Educ.* 23, 857–867. doi: 10.1016/j.tate.2006.04.043
- Baartman, L. K. J., Prins, F. J., Kirschner, P. A., and Van der Vleuten, C. P. M. (2007c). Determining the quality of assessment programs: a self-evaluation procedure. *Stud. Educ. Eval.* 33, 258–281. doi: 10.1016/j.stueduc.2007.07.004
- Benett, Y. (1993). The validity and reliability of assessments and self-assessments of work-based learning. *Assess. Eval. High. Educ.* 18, 83–94. doi: 10.1080/0260293930180201
- Biggs, J., Tang, C., and Kennedy, G. (2022). *Teaching for Quality Learning at University*. McGraw-Hill Education, New York.
- Black, P., and Wiliam, D. (2018). Classroom assessment and pedagogy. *Assess. Educ. Princ. Policy Pract.* 25, 551–575. doi: 10.1080/0969594X.2018.1441807
- Borghouts, L. B., Slingerland, M., and Haerens, L. (2017). Assessment quality and practices in secondary PE in the Netherlands. *Phys. Educ. Sport Pedagog.* 22, 473–489. doi: 10.1080/17408989.2016.1241226
- Boud, D., and Soler, R. (2016). Sustainable assessment revisited. *Assess. Eval. High. Educ.* 41, 400–413. doi: 10.1080/02602938.2015.1018133
- Brooks, J., and King, N. (2014). *Doing Template Analysis: Evaluating an End of Life Care Service*. Sage, Thousand Oaks, CA
- Brooks, J., McCluskey, S., Turley, E., and King, N. (2015). The utility of template analysis in qualitative psychology research. *Qual. Res. Psychol.* 12, 202–222. doi: 10.1080/14780887.2014.955224
- Brown, S. (2005). Assessment for learning. *Teach. High. Educ.* 1, 81–89.

Data availability statement

The datasets presented in this article are not readily available because data are stored in a secure research drive. Participants have not been asked if the dataset can be shared with others. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Ethical board of the Faculty of Social and Behaviour Sciences of Utrecht University; registered under nr. UU-FETC19-022. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LS, WK, MS, CV, and HB have made a substantial contribution to the concept of the article and discussed the findings. LS carried out the interviews and coded the remaining interviews and wrote the initial manuscript. LS and WK performed the initial analysis of the results and created an a priori template for data analysis. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Brown, G. T. (2022). The past, present and future of educational assessment: a transdisciplinary perspective. *Front. Educ.* 7:855. doi: 10.3389/feduc.2022.1060633
- Cohen, L., Manion, L., and Morrison, K. (2018). *Research methods in education*. Eight edition. London: Routledge.
- DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., and Luhanga, U. (2018). Teachers' approaches to classroom assessment: a large-scale survey. *Assess. Educ. Princ. Policy Pract.* 25, 355–375. doi: 10.1080/0969594X.2016.1244514
- Dierick, S., and Dochy, F. (2001). New lines in edumetrics: new forms of assessment lead to new assessment criteria. *Stud. Educ. Eval.* 27, 307–329. doi: 10.1016/S0191-491X(01)00032-3
- Dochy, F. (2009). "The edumetric quality of new modes of assessment: some issues and prospects" in *Assessment, learning and judgement in higher education*. ed. G. Joughin (Berlin: Springer), 1–30.
- Dunbar, S. B., Koretz, D., and Hoover, H. D. (1991). Quality control in the development and use of performance assessments. *Appl. Meas. Educ.* 4, 289–303. doi: 10.1207/s15324818ame0404_3
- Fernandes, E. L., and Flores, M. A. (2022). Assessment in higher education: voices of programme directors. *Assess. Eval. High. Educ.* 47, 45–60. doi: 10.1080/02602938.2021.1888869
- Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., and Kester, L. (2017). Assessment quality in tertiary education: an integrative literature review. *Stud. Educ. Eval.* 55, 94–116. doi: 10.1016/j.stueduc.2017.08.001
- Gerritsen-van Leeuwenkamp, K. J., Joosten-Ten Brinke, D., and Kester, L. (2019). Students' perceptions of assessment quality related to their learning approaches and learning outcomes. *Stud. Educ. Eval.* 63, 72–82. doi: 10.1016/j.stueduc.2019.07.005
- Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *Curr. J.* 16, 207–223. doi: 10.1080/09585170500136093
- Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., and Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educ. Res. Rev.* 17, 50–62. doi: 10.1016/j.edurev.2015.12.002
- Hsieh, H.-F., and Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qual. Health Res.* 15, 1277–1288. doi: 10.1177/1049732305276687
- Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., and Boud, D. (2021). The quality of assessment tasks as a determinant of learning. *Assess. Eval. High. Educ.* 46, 943–955. doi: 10.1080/02602938.2020.1828268
- Jessop, T., McNab, N., and Gubby, L. (2012). Mind the gap: an analysis of how quality assurance processes influence programme assessment patterns. *Act. Learn. High. Educ.* 13, 143–154. doi: 10.1177/1469787412441285
- Kane, M. T. (2008). Terminology, emphasis, and utility in validation. *Educ. Res.* 37, 76–82. doi: 10.3102/0013189X08315390
- Kleijnen, J., Dolmans, D., Willems, J., and Hout, H. V. (2013). Teachers' conceptions of quality and organisational values in higher education: compliance or enhancement? *Assess. Eval. High. Educ.* 38, 152–166. doi: 10.1080/02602938.2011.611590
- Maassen, N., den Otter, D., Wools, S., Hemker, B., Straetmans, G., and Eggen, T. (2015). Kwaliteit van toetsen binnen handbereik: Reviewstudie van onderzoek en onderzoeksresultaten naar de kwaliteit van toetsen. *Pedagogische studiën* 92, 380–393.
- Malecka, B., Boud, D., and Carless, D. (2022). Eliciting, processing and enacting feedback: mechanisms for embedding student feedback literacy within the curriculum. *Teach. High. Educ.* 27, 908–922. doi: 10.1080/13562517.2020.1754784
- Medland, E. (2016). Assessment in higher education: drivers, barriers and directions for change in the UK. *Assess. Eval. High. Educ.* 41, 81–96. doi: 10.1080/02602938.2014.982072
- Messick, S. (1989). "Validity" in *Educational Measurement*. ed. R. L. Linn. 3rd ed (New York: Macmillan Publishing Company), 13–103.
- Nabaho, L., Aguti, J. N., and Oonyu, J. (2011). Making sense of an elusive concept: academics' perspectives of quality in higher education. *High. Learn. Res. Commun.* 7:2. doi: 10.18870/hlrc.v7i2.383
- Opre, D. (2015). Teachers' conceptions of assessment. *Procedia Soc. Behav. Sci.* 209, 229–233. doi: 10.1016/j.sbspro.2015.11.222
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., and Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Adm. Policy Ment. Health Ment. Health Serv. Res.* 42, 533–544. doi: 10.1007/s10488-013-0528-y
- Pickens, J. (2005). Attitudes and perceptions. *Organizational Behavior in Health Care*, N. Borkowski. 4, 43–76. Jones and Bartlett Publishers, Sudbury
- Ploegh, K., Tillema, H. H., and Segers, M. S. (2009). In search of quality criteria in peer assessment practices. *Stud. Educ. Eval.* 35, 102–109. doi: 10.1016/j.stueduc.2009.05.001
- Poldner, E., Simons, P., Wijngaards, G., and Van der Schaaf, M. (2012). Quantitative content analysis procedures to analyse students' reflective essays: a methodological review of psychometric and edumetric aspects. *Educ. Res. Rev.* 7, 19–37. doi: 10.1016/j.edurev.2011.11.002
- Postareff, L., Virtanen, V., Katajaviuori, N., and Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Stud. Educ. Eval.* 38, 84–92. doi: 10.1016/j.stueduc.2012.06.003
- Schellekens, L. H., van der Schaaf, M. F., van der Vleuten, C. P., Prins, F. J., Wools, S., and Bok, H. G. (2022). Developing a digital application for quality assurance of assessment programmes in higher education. *Qual. Assur. Educ.* 31, 346–366. doi: 10.1108/QAE-03-2022-0066
- Schut, S., Heeneman, S., Bierer, B., Driessen, E., van Tartwijk, J., and van Der Vleuten, C. (2020). Between trust and control: Teachers' assessment conceptualisations within programmatic assessment. *Med. Educ.* 54, 528–537. doi: 10.1111/medu.14075
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educ. Res.* 29, 4–14. doi: 10.3102/0013189X029007004
- Simper, N., Mårtensson, K., Berry, A., and Maynard, N. (2021). Assessment cultures in higher education: reducing barriers and enabling change. *Assess. Eval. High. Educ.* 47, 1016–1029. doi: 10.1080/02602938.2021.1983770
- Sridharan, B., Leitch, S., and Watty, K. (2015). Evidencing learning outcomes: a multi-level, multi-dimensional course alignment model. *Qual. High. Educ.* 21, 171–188.
- Stebbins, R. A. (2001). *Exploratory Research in the Social Sciences* (Vol. 48). Sage, Newbury Park.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan* 77:238.
- Teddle, C., and Yu, F. (2007). Mixed methods sampling: a typology with examples. *J. Mixed Methods Res.* 1, 77–100. doi: 10.1177/1558689806292430
- Tillema, H., Leenknecht, M., and Segers, M. (2011). Assessing assessment quality: criteria for quality assurance in design of (peer) assessment for learning—a review of research studies. *Stud. Educ. Eval.* 37, 25–34. doi: 10.1016/j.stueduc.2011.03.004
- Utrecht University Education Guideline. (2018). Available at: <https://www.uu.nl/en/education/education-at-uu/the-educational-model> (Accessed February 08, 2021).
- Van der Schaaf, M., Baartman, L., and Prins, F. (2012). Exploring the role of assessment criteria during teachers' collaborative judgement processes of students' portfolios. *Assess. Eval. High. Educ.* 37, 847–860. doi: 10.1080/02602938.2011.576312
- Van der Vleuten, C. P., and Schuwirth, L. W. (2005). Assessing professional competence: from methods to programmes. *Med. Educ.* 39, 309–317.
- Van der Vleuten, C. P., Schuwirth, L. W., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K., et al. (2012). A model for programmatic assessment fit for purpose. *Med. Teach.* 34, 205–214. doi: 10.3109/0142159X.2012.652239
- Xu, Y., and Brown, G. T. (2016). Teacher assessment literacy in practice: a reconceptualization. *Teach. Teach. Educ.* 58, 149–162. doi: 10.1016/j.tate.2016.05.010
- Yates, A., and Johnston, M. (2018). The impact of school-based assessment for qualifications on teachers' conceptions of assessment. *Assess. Educ. Princ. Policy Pract.* 25, 638–654. doi: 10.1080/0969594X.2017.1295020

Appendix I

Initial coding template

1) Quality criteria of assessment

The participant identifies a criterion used to establish the quality of a course assessment or of an assessment programme.

(a) Perception – how the participant interprets and assigns meaning to assessment quality, how the participant thinks about it.

Coding segment: Answer to the question: what is assessment (programme) quality?

(b) Achievement in practice – the participant clearly indicates how the aspect is addressed in practice, and how this is realised.

Coding segment: Answer to the question: how is assessment (programme) quality achieved in practice?

Coding template

(1) Validity; refers to whether an assessment (programme) is appropriate for achieving the intended learning outcomes, conclusions based on test results are justified.

(2) Reliability; refers to whether an assessment (programme) is administered accurately and consistently, independently judged, and replicable.

(3) Transparency; refers to whether the assessment (programme) has clear procedures and criteria for judging performance, and it is clear how the student will be assessed.

(4) Learning impact; refers to whether an assessment (programme) optimally supports and enhances students' learning processes and development.

2) Perception criterion learning impact

Participants' perception of how an assessment (programme) contributes to student learning processes.

Coding segment: Answer to the question: how does the assessment (programme) contribute to student learning?

Open coding (no initial template).