# DEFINING AND DISTINGUISHING BETWEEN THE CONCEPTS OF ASSESSMENT AND EVALUATION

## Conceptual Paper

## Yuraisha Chetty and Tshepho Mokwele, Research Directorate

## June 2020

# 1. Introduction and context

The South African Qualifications Authority (SAQA) provides on its website, a Standard Glossary of Terms document that clarifies and defines the key terminology relating to the development and implementation of the National Qualifications Framework (NQF). SAQA has updated this Glossary since it was first published in 2014. SAQA will continue to review and refine these definitions as necessary, to strengthen a common understanding and use of such terminology and reduce ambiguity and confusion.

The purpose of this paper is to foreground two terms used in the context of the NQF, namely, 'assessment' and 'evaluation', and to define and distinguish between these terms. This desktop exercise was precipitated by inputs received from the Quality and Standards (Q&S) Committee of SAQA, which sought clarity on how these terms are understood and defined by SAQA. Feedback pointed to the conflation of the terms 'assessment' and 'evaluation', particularly in the Glossary.

The research found that the concepts of assessment and evaluation are indeed conflated. In the Glossary, these concepts are used interchangeably. The following extracts from the Glossary illustrate this conflation:

> **Assessment criteria** - *the standards used to guide learning and to assess learner achievement and/or to <u>evaluate</u> and certify competence* (2017:14).

> **Summative assessment** - *assessment conducted at the end of sections of learning, at the end of a whole learning programme, or at any point in the learning programme, to <u>evaluate</u> learning related to a particular qualification, part-qualification, or professional designation* (2017:57)

It is evident that the concept of evaluation appears in both the assessment-related definitions above.

The desktop research (discussed in Section 2) revealed that some literature on assessment and evaluation adds to the confusion and ambiguity that exists in understanding and using these concepts, by not making a clear enough distinction between these two concepts. Overall, the research found that the concepts of assessment and evaluation are conceptually distinct (they do not hold the same meaning). This outcome points to the need for a refinement of both assessment-related definitions in the Glossary and any other documents (e.g. policies) where these terms may appear. In essence, it is recommended that the definitions of assessment provided above should not include the term 'evaluate'. Furthermore, the discourse at SAQA will need to cater for the evident distinction between these terms.

Section 2 draws from various literature sources and attempts to define and distinguish between the concepts of assessment and evaluation to a) provide support for the recommendation to refine the definitions of assessment in the Glossary and other documents (e.g. policies), and b) enrich our understanding of both concepts.

The desktop research was exploratory in nature. The literature search was not exhaustive, and further research may provide additional insights.

## 2. Defining the concepts of assessment and evaluation

The Oxford Dictionary and Thesaurus (2007) defines the concepts of assessment and evaluation similarly and uses the terms interchangeably, leading to an understanding that the two concepts have the same meaning.

> **Assess (dictionary)** – *to make a judgement about the value or quality of something.*
>
> **Assess (thesaurus)** – *evaluate, judge, rate, estimate, appraise, weigh-up.*
>
> **Evaluate (dictionary)** - *to form an idea of the amount or value of.*
>
> **Evaluate (thesaurus) -** *assess, judge, gauge, rate, estimate, appraise, weigh-up.*

The above definitions add to the confusion that exists in trying to clarify these two terms and makes it challenging for stakeholders to see each term as conceptually distinct.

In distinguishing 'assessment' from 'evaluation', Harmer (2001, as cited in Azizifar et al., 2010: 54), writes that the assessment of a coursebook, for instance, is an out-of-class judgment as to how well a new book will perform in class. Coursebook evaluation, on the other hand, is a judgment on how well a book has performed in fact.

Writing on assessment and evaluation in the context of Higher Education, Edith Kealey (2010) notes that, traditionally, assessment and evaluation have been the means through which feedback is provided to both teachers and students. However, the two activities generally take place in isolation. Kealey (2010: 66) further notes that "while the literature does not clearly distinguish between the two terms," the term 'assessment' refers to the measurement of student learning while 'evaluation' refers to the measurement of instructor teaching. In turn, assessment and evaluation can each be considered as the measurement of a process (formative) or as the measurement of a product (summative), respectively. A basic distinction between assessment and evaluation lies in the orientation: assessment is process-oriented, while evaluation is product-oriented (Key Differences Website, 2016).

Kelter (2018) defines evaluation as a process of data collection to determine the worth of a program (current value), and assessment as the collection of data to determine how to improve a program (looking toward the future). In his paper, Kelter (2018) uses the term 'assessment' broadly to mean the process by which we determine if program goals and objectives have been met, and how we might change our program to meet these goals and objectives better.

Scriven (2007:2) synthesised the definition of evaluation which appears in most dictionaries and the professional literature, and defined evaluation as "the process of determining merit, worth, or significance; an evaluation is a product of that process." This is akin to Kelter's (2018) definition, which includes the terms 'worth' and 'current value'. Scriven (2017:8) points out that the doctrine of 'value-free social science' initially led to evaluation being perceived as 'taboo', however, it was taken seriously in disciplines such as mainstream sociology, political science and social psychology. He refers to the subsequent development of evaluation models such as the CIPP (Context, Input, Process, Product), and thereafter the logic model as it was referred to in the United States (U.S.) or the realistic or realist approach, as it was referred to in the United Kingdom (U.K.). He provides a critique of this model, which may be of interest to readers. However, due to the scope of the desktop research, this paper will not delve into his critique of the logic model or other aspects of evaluation.

Scriven (2007:2) furthermore stated that the logic of evaluation is concerned with "(i) how, if at all, professional evaluation is possible; (ii) its nature and its location in the organisation of knowledge, and (iii) the logical structure of its inferences." He explains the 'extensive territory' of evaluation, in which evaluation occupies every discourse, from "proposing, attacking, and defending evaluative claims about food products, football teams, human behavior [sic], global warming, and almost everything else". With regard to what Scriven refers to as 'professional evaluation', he distinguishes seven standard categories or sub-divisions, namely, performance evaluation, product evaluation, personnel evaluation, proposal evaluation, program evaluation (which he states includes "practice, procedural, and process evaluation"), policy analysis, and portfolio evaluation – and he adds that there is some overlap between some of these (Ibid:3).

In expanding the understanding of 'value' which is key to the definition of evaluation, Scriven (2007:11-12), describes and explains the different types of value claim. These include personal preferences (wants); market value; real, true or essential value (which Scriven states is the sense of value that professional evaluation seeks to uncover, and is the one that "evaluation as a discipline is all about, just as the 'real truth' is what the professional journalist or scientist seeks"); public value; standards and requirements; contextual values; and illustrative and exemplary value [See Scriven, 2017, for more detail].

Focusing on informal learning, Savenye (2014) asks *what is assessment?* Like Kelter, Savenye (2014: 260) writes that assessment is typically defined as measuring learning – "the process of measuring, documenting, and interpreting behaviours related to learning". What is measured here is individual learning, both for the benefit of the learner and to improve instruction. *How does assessment relate to evaluation?* In making a distinction between evaluation and assessment, educators and researchers examine two aspects: the learner's performance (assessment) and the instruction or learning environment (evaluation) (Ibid.). Evaluation is thus usually focused on a program or course or at an organisational level. It also includes assessing learning,

but also the activities involved in making judgements and decisions about the quality of the program or initiative as a whole.

Savenye (2014) states that *formative evaluation* is conducted to collect data to aid in improving a program, product or approach during development, ideally on an ongoing basis. *Summative evaluation*, by comparison, is done typically at the end of development for reporting purposes, or to make a final determination about retaining a program, or choosing an alternative (Savenye, 2014). Specifically with regard to evaluation, Patton (2009) suggests that the purpose of evaluation informs the type of evaluation used. He distinguishes between developmental, formative and summative evaluation. He describes developmental evaluation as evaluation used for developing or emerging initiatives, while summative and formative evaluation are typically used to examine established programs. Summative evaluation measures outcomes against pre-determined goals and frameworks while formative evaluation can assist in continuous improvement. Furthermore, Patton (Ibid) states that participatory evaluation is cross-cutting and refers to the element of participation in all types of evaluation.

Lemhouer's paper (1982) on *Monitoring in the Classroom* defines what assessment is: the testing of the performance of an individual to find out whether s/he has reached the level intended to achieve. Assessment may thus serve as a monitoring instrument to make decisions about the progress of the individual assessed. Kay (1975, as cited in Lemhouer, 1982:50) notes that "assessment may be carried out on a sampling basis as a tool of evaluation in order to make decisions about a system".

Christ-Janer (1973) understands assessment to imply knowledge about the individual and performance and contrasts this to evaluation, which implies a judgment about the individual's performance. Assessment and evaluation within a social context are hardly new ideas.

MacDonald (1976) also distinguishes 'evaluation' from 'assessment'. The purpose of assessment, he writes, is to make statements about the recipients of educational service, and statements about their actual and potential accomplishments in relation to the opportunities for learning by that service. Assessment is the basis for decisions about what students will get in the way of further provision, and for predictions of their future accomplishments. The purpose of evaluation, on the other hand, is not to make statements about the recipients, but to make statements about the educational service.

The terms evaluation and assessment have been used differently in different countries and contexts (Sadler, 2012: 202). In the 1970s, the term evaluation in the United States (U.S.) covered appraisals of student learning as well as curriculum reforms and educational projects and programmes. On the other hand, assessment in U.S. higher education included everything to do with appraising the effectiveness, worth or value of institutional characteristics and functions – teaching, research, facilities, services, student support, organisational systems, and student learning. In referring to

assessment and evaluation in international discourse, meanings have varied across countries and time and may continue to do so (Ibid.).

In an article on *Assessment and evaluation in medical education*, Gibbs, Brigdon & Hellenberg (2006) write that 'assessment' is derived from a Latin word that means "to sit beside and judge" – appropriately used to describe the systemic gathering of information about what the learner should know, be able to do or to work towards. The word is usually associated with some measurement, marks or percentages, but could be related to specific descriptors: excellent, good, average, or poor. If an assessment is a measurement, Gibbs et al. (2006) suggest that we need to know what it is measuring. Competency is thus one such measurement and refers to the specific skill that has been taught, and is now being measured. Competency and objectives or outcomes are often used interchangeably in describing what has to be measured. This form of measurement has to be judged against a particular standard or benchmark to be achieved.

According to Gibbs et al. (2006:5), "in describing the students' achievement, it is common to refer to these standards as well as describing its reference point. If we measure and compare student performance against a large number of his [sic] peers and judge how or where he [sic] fits into the distribution of marks, this is known as norm referencing".

Norm referencing is used if one wishes only a certain percentage of students to pass or fail. However, if we assess how students perform against a specific objective or criterion, this is known as criterion referencing. With this latter form, all students can pass an assessment if they have all reached the required standard or competency (Gibbs et al., 2006).

Like many scholars, Gibbs et al. (Ibid), refer to commonly used forms of assessment:

- Formative assessment, whereby assessment is often informal, continuous or ongoing and is a two-way communicative process between teacher and learner/s, with a large degree of discussion regarding strengths, weaknesses and opportunities to improve.
- Summative or end-point assessment is much more formal, and judgmental, (the traditional examination system). It tends to be more numeric and quantitative, whereas formative assessment tends to be more descriptive and qualitative.
- Self or ipsative assessment concerns measuring oneself – you learn about yourself, you test yourself against standards or specific parameters, giving you an idea of how you are performing. This form of assessment becomes very important in teaching methods such as problem-based learning, where self-directed learning often leads to students finding difficulty in knowing or understanding to what level of learning, they need to work towards (Gibbs et al., 2006).

On the other hand, 'evaluation' is a much broader concept than assessment, because it is used to measure the value of educational activities, programmes, curricula etc. While assessment relates to individuals or groups and their performance, "evaluation relates more to process, content and appropriateness of activities" (Gibbs et al., 2006: 6). For example, an evaluation into a specific course/program could look into whether it had the required effect or has addressed the correct issues.

Assessment is a process that involves documenting knowledge, skills, attitudes and beliefs, in measurable terms, with a view to making improvements, not just to make judgement. It is a process of describing, collecting, recording, scoring and interpreting information about learning (Pennsylvania State University, 2017).

Focusing on Higher Education, Pereira, Flores and Niklasson (2016) importantly note that assessment methods conventionally used in higher education are examinations and written tests, which encourage surface, rather than deep, learning. Learner-centred assessment methods such as self- and peer-assessment are seen as alternatives or complementary because they encourage problem-based learning and stimulate deep learning and critical thinking.

In their article focusing on assessment in post-apartheid South African schooling, Kanjee and Sayed (2013) note that, in the formal education sector, the assessment system comprised school-based testing and the national matriculation examinations, with the primary purpose being selection to the next level of education. They further argue that "policy intention of implementing an effective classroom *assessment* system is, in reality, a classroom *measurement* system" (Kanjee & Sayed, 2013: 443).

Assessment is defined by Nitko and Brookhart (2011 as cited in Kanjee & Sayed, 2013:Ibid) as, "a broad process for obtaining information that is used for making decisions about students, curricular programmes, and schools and educational policy", and importantly, these authors argued that assessment should be distinguished from evaluation, measurement and testing.

The *NQFPedia* (SAQA, 2017: 13) defines assessment as "the process used to identify, gather and interpret information against the required competencies in a qualification or part-qualification in order to make a judgement about a learner's achievement". This definition also appears in SAQA's *National Policy and Criteria for Designing and Implementing Assessment for NQF Qualifications and Part-Qualifications and Professional Designations in South Africa* (2014). This definition of assessment does not include the term 'evaluate' – correctly so. The policy expands the definition to include the statement that assessment can be formal, non-formal or informal, and that assessment can be of learning already done, or towards learning to inform and shape teaching and learning still to be done.

# 3. Conclusion

This paper has attempted to explore definitions of the concepts of assessment and evaluation. While definitions and understandings of assessment and evaluation are not always clear-cut, generally, there is evidence to support notions that assessment and evaluation are conceptually distinct. It is recommended that SAQA review its definitions of 'assessment criteria' and 'summative assessment' and remove references to 'evaluate'. Further research might lend additional insights.

*Paper prepared by Ms Yuraisha Chetty, Deputy Director: Research and Mr Tshepho Mokwele, Assistant Director: Research, with inputs from Ms Charmaine Lebooa: Admin Officer: Research.*

*Research Directorate, SAQA – June 2020.*

# References

Azizifar, A., Koosha, M. & Lotfi, A.R. 2010. An analytical evaluation of Iranian high school ELT textbooks from 1970 to the present. *Procedia - Social and Behavioral Sciences*, 3: 36–44.

Christ-Janer, A.F. 1973. Postsecondary Options. *NASSP Bulletin*: 68–78.

Gibbs, T., Brigden, D. & Hellenberg, D. 2006. Assessment and evaluation in medical education. *South African Family Practice*, 48(1): 5–7.

Kanjee, A. & Sayed, Y. 2013. Assessment policy in post-apartheid South Africa: challenges for improving education quality and learning. *Assessment in Education: Principles, Policy & Practice*, 20(4): 442–469.

Kealey, E. 2010. Assessment and Evaluation in Social Work Education: Formative and Summative Approaches. *Journal of Teaching in Social Work*, 30(1): 64–74.

Kelter, P. 2018. Assessment in Academia. The good, the bad and the ugly. *Educación Química*, 15(2): 113-119–119.

Key Differences Website. 2016. Difference Between Assessment and Evaluation (with Comparison Chart). *Key Differences*. Accessed 20 June 2020 from https://keydifferences.com/difference-between-assessment-and-evaluation.html.

Lemhouer, A. 1982. Monitoring the Classroom. In Second MATE Conference. 49–65. Accessed 15 May 2020 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.1581&rep=rep1&type=pdf#page=49.

MacDonald, B. 1976. Evaluation and Assessment: Who's Afraid of Evaluation? *Education 3-13*.

Oxford Paperback Dictionary and Thesaurus. 2007. Oxford University Press. United Kingdom.

Patton, M. Q. (2009). Developmental evaluation as alternative to formative assessment. [Web Video]. Accessed 28 June 2020 from http://www.youtube.com/watch?feature=player_embedded&v=Wg3IL-XjmuM

Pennsylvania State University. 2017. Differences between Testing, Assessment, and Evaluation. *iStudy for Success!* Accessed 20 June 2020 from http://tutorials.istudy.psu.edu/testing/testing2.html.

Pereira, D., Flores, M.A. & Niklasson, L. 2016. Assessment revisited: a review of research in Assessment and Evaluation in Higher Education. *Assessment & Evaluation in Higher Education*, 41(7): 1008–1032.

Sadler, D.R. 2012. Assessment, evaluation and quality assurance: Implications for integrity in reporting academic achievement in higher education. *Education Inquiry*, 3(2): 201–216.

SAQA. 2014. *National Policy and Criteria for Designing and Implementing Assessment for NQF Qualifications and Part-Qualifications and Professional Designations in South Africa*. Pretoria: South African Qualifications Authority. Accessed 15 May 2020 from https://www.saqa.org.za/sites/default/files/2019-11/National%20Policy%20for%20Assessment%28PrintReady%29.pdf.

SAQA. 2017. *NQFpedia: Standard Glossary of Terms*. Pretoria: SAQA. Accessed 15 May 2020 https://hr.saqa.co.za/glossary/pdf/NQFPedia.pdf#.

Savenye, W.C. 2014. Perspectives on Assessment of Educational Technologies for Informal Learning. In M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop, eds.

*Handbook of Research on Educational Communications and Technology*. New York: Springer-Verlag: 257–267.

Scriven, M. 2007.The Logic of Evaluation. OSSA Conference Archive, 138. Accessed 26 June 2020 from https://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=1390&context=ossaarc hive.