OPEN ACCESS

GAVIN PUBLISHERS

Commentary

Programming Best Assessment

Francesco Carelli*

Department of Clinical Medicine, University of Milan, Italy

It seems there is no ideal instrument to get a gold assessment. Any single assessment implies a compromise on quality criteria and the choice on which criterion to compromise should be based on a well-considered decision as to which quality element is to be optimised on the specific assessment context. A programme of assessment, combining different assessments, can alleviate the compromises on individual methods, thereby rendering the total more than the sum of its parts.

One single method can only assess a part of Miller's pyramid and there is no magic bullet that can do it all in one go. A complete assessment programme will inevitably also to employ non-standardised methods. Particularly, if we wish to assess in real practice, i.e. at the top of Miller's pyramid (the "does" level), standardisation is out of reach. The real world is non-standardised and haphazard, and, more importantly, any attempt at standardisation will only trivialise the assessment.

The users, i.e. the assessors, learners and patients, are more important than the instrument. Their expertise in using the instrument, the extent to which they take the assessment seriously and the time they can spend on it, these aspects together determine whether or not the assessment is performed well. While extensive training is not required for someone handing out multiple choice test booklets to students, with non-standardised observational assessment it is of crucial importance that all those involved in the assessment process should receive extensive training.

Since an assessment programme without non-standardised methods is unthinkable, we need to develop a "technology" to help users to function appropriately in their assessment role. In doing so, we need to realise that someone who learns is a learner, even if most of the time, they are assessors, teachers or supervisors. All people learn in the same way, preferably by training, practice and feedback. It will not suffice to simply provide assessors with information or instruments. If the users, assessors and assesses do not fully understand the meaning and purpose of the assessment, the assessment is doomed to be trivialised.

From the perspective of a conceptual framework of programmatic assessment, the formative-summative distinction is not a very useful one, considering that the framework predicates that any assessment should be both formative and summative, only to varying degrees. Therefore, conceptualizing the stakes of the assessment as a continuum from low to high stakes seems more useful. In low-stake assessment, the results have limited consequences for the learner in terms of promotion, selection or certification, whereas high-stake assessment can have far-reaching and dramatic consequences. In a programme of assessment, only low-stake decisions can be based on single data points, whereas all high-stake decisions require input from many. However, when high-stake decision making is informed by many data points, it would be foolish to ignore the information from the rich material derived from all the single data points. Information from combined low-stake assessments should therefore feed into high-stake information.

There is strong evidence that formative feedback can enhance learning. If assessment is to drive learning, it is imperative that it should produce meaningful information to the learner. In other words, assessment information should be as rich as possible, in many different ways, both quantitatively and qualitatively.

Lack of meaningfulness leads to trivialisation, a serious and frequent hazard in assessment. If learners are required to memorise checklists for passing the Objective Structured Clinical Examination (OSCE) but have no connection with patients, their performance is trivial; if an assessor completes all items on a professional behaviour rating from one strike of the pen, the assessment loses all meaning and is trivialised. However, if the assessment information is meaningful, learning will be enhanced in a meaningful way: low-stake individual data points should be as meaningful as possible to foster learning, and high-stake decisions should be based on many individual data points. The aggregation of meaningful data points can result in a meaningful high-stake decision and in all elements of the assessment programme, we should be on our guard against trivialisation.

The purpose of an assessment programme is to maximise assessment for learning while at the same time arriving at robust decisions about learner's progress. We can start with a first period of training activities consisting on "learning tasks" as lectures, practical, patient encounter, a Problem Based Learning (PBL), tutorial, a project, a learning assignment or self-study.

Assessment drives learning: this principle requires that all assessment be maximally meaningful to learning and provide feedback on the learner's performance that is information-rich, whether quantitatively or qualitatively. We must be against passing or failing a learner based on one assessment point, as can be done in a mastery test. Each data point is but one element in a longitudinal array of data points. Although single data points are low stake, this does not preclude their use for progress decisions at a later point in the curriculum. With each single assessment, the assessor's principal task is to provide the learner with as rich and extensive feedback as possible. It is not useful to simply declare whether or not someone has achieved a certain standard.

Grades must not be the only feedback that is given, because they are poor feedback carriers and tend to have all kinds of adverse educational side effects: learners hunting for grades but ignoring what and how they have learned, teachers being content to use the supposed objectivity of grades as an excuse for not giving performance feedback.

We need a committee of examiners, trained and certified in that, because expert judgement is imperative for aggregating information across all data points. They learn as their experience accumulates and can change the procedures and supporting tools.

It is very important to have an intermediate developmental assessment as remediation oriented, offering information-rich recommendations for further learning, tailored to the individual learner and contingent on the diagnostic information.

The learner's logical longitudinal development through learning tasks, appropriate feedback and supported self-direction is of key importance. This is entirely the opposite of a purely mastery-oriented approach where passing an exam means being declared competent for life.

Ideally, the decisions should be motivated by a justification. The decisions may not be limited to a mere pass or fall, but also indicate distinctive excellence of performance. The committee may provide recommendations for further training or remediation. Overall, the final decision is robust and based on rich information and numerous data points and, if challenged, it should be accountable and defensible even in a court of law.

Key purpose is to evaluate the curriculum. Information from the supporting actors, such as mentors/ coaches, and information from the actors in the intermediate and final evaluation offer excellent data points for curriculum evaluation in terms of both the process and the outcomes of education and training.

As soon as an assessment procedure, an assessment strategy or an assessment procedure becomes more important than an original goal it was intended to accomplish, trivialisation rears its ugly head. We see it happening all the time learners perform tricks to pass exams, teachers complete forms with one stroke of the pen to complete administrative requirements strongly judgement meaningless, we stick to procedures for no other reason than that have always done it this way (we want grades because they are objective and accountable to society) or because of institutional policy. As soon as we notice the exchange of test materials on the black market or new internet resources peddling rafts of ready-made reflections, we can be sure that we have trivialised the assessment process. All actors in programmatic assessment should understand what they are doing, why they are doing it and why they are doing in this way. Otherwise, they are in danger of losing sight of the true purpose of assessment and will fall back on bureaucratic procedures and meaningless artefacts. Steering clear of trivialisation is the hardest yet most urgent task we have to tackle if we are to realise the best programmatic assessment.