Philosophical Considerations in Educational Assessment

5

Richard B. Gunderman

Assessment is one of the three basic components of an educational program. Curriculum addresses the issue of what gets taught. Instructional method addresses how it gets taught. And assessment addresses how effectively it has been taught and learned. Overall, educational effectiveness is like a chain made up of these three links, and a chain can only be as strong as its weakest link. If we are teaching the wrong things, it does not matter how effectively we are teaching them. If our method of instruction is poor, the subject matter is unlikely to be grasped by learners. And if we fail to assess, or assess poorly, the effectiveness of teaching and learning, then there is a good chance that curriculum and instructional methods may not meet the needs of learners.

There are a number of issues worth considering in educational assessment. One concerns what we are attempting to assess. The easiest thing to assess is whether or not learners have memorized particular facts or skills. For example, does a learner know the differential diagnosis for multiple osteoclastic lesions in the skull? But there are multiple levels of understanding, as indicated by widely employed assessment models such as Bloom's taxonomy of learning objectives (Bloom et al. 1956). First created in 1956, Bloom and colleagues ranked various learning objectives, placing recall at the bottom, understanding in the middle, and application in practice near the top. Associated with application is the ability to evaluate and create. To promote the highest forms of learning, it is insufficient to focus assessment on whether or not learners can recall facts. Educators need to determine whether or not they can put that knowledge to use.

Yet knowledge and skills do not exhaust the range of educational objectives that need to be assessed. If learners were robots, this might be sufficient – they would

R.B. Gunderman
Departments of Radiology, Pediatrics, Medical Education, Philosophy, Liberal Arts, Philanthropy, Indiana University School of Medicine, Indianapolis, IN, USA
e-mail: rbgunder@iupui.edu

K.M. Hibbert et al. (eds.), *Radiology Education*,
DOI 10.1007/978-3-642-27600-2_5, © Springer-Verlag Berlin Heidelberg 2012

know how to carry out the tasks we prescribe to them, like a computer programmer writing software. But our learners are human beings, who need to be able to develop a set of professional commitments, a style of practice, and the ability to weigh alternative points of view and approaches to practical decisions in their daily practice (Gunderman 2011). This is an area in which radiology has not performed as well as it might, because most of the standardized tests that have dominated the learning horizons of radiology residents have focused primarily on knowledge and skills, neglecting issues such as professionalism and how effectively learners interact with patients and other health professionals.

We are moving toward more computer-based testing, in which learners have less face-to-face interaction with educators in real time. This will tend to deemphasize the assessment of noncognitive and nontechnical learning objectives. This shift is driven by a variety of considerations, including finances, convenience, and control (Alderson and Becker 2008). Computer-based testing is less expensive, generally requires less effort in travel on the part of candidates and examiners, and provides those who design the tests more control over how the tests are administered and scored. It also appears to be fairer. The scoring is carried out by a computer, thus reducing the potential for scores to be skewed by individual biases on the part of examiners. On the other hand, it tends to devalue more subjective learning objectives, including attitudes and beliefs.

As outlined in Chap. 2, there are important distinctions to be drawn in assessment. One is between summative and formative types of assessment (Bloom 1971). Summative assessment generally takes place at the end of a course or period of training and attempts to determine what score or grade a learner should receive. Most board exams are an example of summative assessment. By contrast, formative assessment is carried out during learning and aims to help learners learn more effectively. Historically, radiology education has tended to emphasize summative assessment, although recent efforts to ensure that learners receive periodic performance appraisals during their training have mitigated this imbalance somewhat. If enhancing learning is our primary educational objective, then we need to focus more attention on helping learners improve, as opposed to merely grading them.

Another important distinction in educational assessment is between objective and subjective assessment (Gipps 1994). Questions formulated in an objective format have only one correct answer, while subjective assessments may have multiple more or less correct possible responses. Again, educational assessment in radiology has tended to favor the objective approach, which accustoms learners to looking for the one correct answer, as though all others are wrong. This makes scoring examinations relatively straightforward, but it may undermine habits of mind that are crucial for the future advancement of the field, such as critical thinking, innovation, and creativity. To discover new ideas and ways of doing things, you need to focus more on whether an idea is interesting than whether it is correct or incorrect.

Another important distinction is between formal and informal assessment (Nitko 2001). Formal assessment involves the use of standardized assessment instruments, such as paper or computer-based examinations, or a standard form that is used to evaluate all learners. Formal assessments tend to ask educators to assign numerical

scores. By contrast, informal assessments can be based on observations of practice, discussion, or learning portfolios, where each learner can be evaluated somewhat differently from others. From some perspectives, the downside of informal assessment is that it requires trust that the evaluator will avoid bias and adapt the assessment to the particular learner at hand. The big drawback of formal assessment is the fact that it tends to homogenize learners and educators, by implicitly encouraging each one to conform to the same pattern.

Another important distinction is between centralized and decentralized assessment. Centralized assessments are generated by organizations outside the institution where learning is taking place. One example would be the American Board of Radiology, which produces a single examination that is taken by learners who train in hundreds of different institutions. Decentralized assessment, by contrast, is designed and implemented locally, at the institution where learning is taking place. Again, centralized assessment tends to homogenize educators and learners, since both want learners to perform well on the standardized exam. The development of distinctive local educational approaches will tend to be stunted, since all learners take the same exam. This, in turn, may limit the diversity of interests and experiences and thereby undermine creativity.

One of the best examples of the pitfalls of educational assessment is the 2001 US No Child Left Behind Act, which mandated the use of standardized testing throughout the United States (American Psychological Association 2011). The goal of the legislation was to increase the accountability of teachers, administrators, and schools for the achievement of objective educational standards. By obtaining objective, nationwide data on learning outcomes, underperforming teachers and schools could be fired or closed. In fact, however, the act resulted in the disappearance of subjects not covered on the test, such as art and music, from the curricula of many schools. Cognitive horizons became narrower and more superficial. Teachers spent more and more time teaching to the test. In some well-publicized cases, teachers and schools actually began cheating to avoid the sanctions associated with poor performance (National Center for Fair and Open Testing 2011).

To gain a deeper appreciation of the importance of how educational assessment is designed and implemented, let us consider in more depth the important differences between centralized and decentralized approaches to assessment. In centralized approaches, a few high-level educators, such as a board of radiology, make the decisions and policies about how educational assessment will take place. In decentralized approaches, such decisions are made by educators at individual institutions. In centralized approaches, knowledge and authority are located centrally, and decisions spread in a central-to-local direction. In decentralized approaches, knowledge and authority reside locally, and ideas and information flow from the local to the central. Both approaches exhibit characteristic advantages and disadvantages.

What are the advantages of centralized educational assessment? First, it makes it easier to compare the performance of different educators, institutions, and learners, all of whom are being assessed by the same objective criteria. Underperformers can be sanctioned and ultimately closed down if they do not meet minimum standards. Moreover, centralized approaches appear to be fairer than decentralized ones, since

everyone is assessed by the same instrument, reducing the potential for individual bias. In addition, centralized approaches tend to be efficient. Only one test needs to be designed, rather than each institution "reinventing the wheel," and all the scoring can be performed by a single organization. Finally, centralization generally makes it possible to introduce changes in assessment over a wider area in a shorter period of time, because a single body makes the decisions.

On the other hand, centralization also exhibits characteristic drawbacks, while decentralization offers important advantages. Since current trends seem to be in the direction of centralization, let us explore these in somewhat more depth. One major disadvantage of centralization is the fact that it tends to involve fewer people, and thus draw on fewer perspectives, in its decision-making process. To avoid this pitfall, centralized educational assessors such as boards of radiology need to ensure that their membership reflects diverse points of view and work extra hard to listen to new ideas and criticisms from individual educators and programs. Human beings have a natural tendency to want to be in control and to avoid criticism, which can lead the administrators and staff of such organizations to become more isolated from the educators and learners over whom they exercise authority.

A related pitfall in centralization of educational assessment is for the organization, its survival, and its growth to come to be seen as ends in themselves, as opposed to means by which to promote educational quality and professional flourishing. For example, job security and compensation may both increase as the organization grows, and this may contribute to a natural tendency for such centralized assessment organizations to expand their staffs, budgets, and programs, even when doing so is not raising educational quality. In some cases, the growth of the organization itself may, in the eyes of executives and board members, begin to be regarded as a proxy for organizational success. Yet, the growth in the size and complexity of bureaucracies does not necessarily benefit the constituencies they serve. To avoid such pitfalls, centralized assessment organizations need to work extra hard to be mindful of the needs and missions of local constituencies.

Another pitfall of centralized educational assessment is the fact that those making the rules often do not live under the rules they promulgate. For example, many executives and staff in educational assessment organizations have never served as a program director in one of the programs required to follow their rules. Moreover, current program directors may occupy relatively few seats on the boards of such organizations. As a result, the costs of new programs of educational assessment may not be well proportioned to their benefits. Operating under the presumption that you cannot manage what you cannot measure, centralized authorities may impose onerous requirements for data collection and analysis that hamper important work of educators, unaware of such imbalances because they do not live with the requirements day to day. To avoid this, educators actually doing the daily work of education should play a major role in the development of new programs of educational assessment.

One of the advantages of decentralization of educational assessment is its tendency to promote the engagement of educators at the local level. People are much more likely to play an active and enthusiastic role in programs that they had a hand

in designing. When they are merely told what to do by a distant authority, they are less likely to feel committed to the program they are charged with implementing. This represents one of the geniuses of democratic and republican forms of government, which promote the participation of citizens, as opposed to despotisms and tyrannies, where the governed play little or no role in self-government. This, in turn, promotes greater creativity, resourcefulness, and personal and professional development of educators, who are able to think and act responsibly for themselves. One important contribution centralized assessment organizations can make is to focus less on what they can require local educators to do and more on how they can help such educators to acquire the knowledge, skills, and time and resources they need to do their jobs well.

Decentralization can contribute not only to engagement but also to morale. One of the surest ways to undermine morale is to saddle educators with responsibility for achieving objectives for which they lack resources. Another is to saddle them with objectives that they do not believe in. Another is to saddle them with objectives they do not understand. Instead of telling educators how they must assess their learners, centralized assessment organizations may be able to provide more benefit by encouraging innovation and then helping to disseminate new ideas and best practices among different programs. There is a danger that initiatives designed to identify, remediate, and weed out underperformers will stifle the enthusiasm and fulfillment of educators and program directors who are doing a good job and could do even better. Centralized assessment organizations can avoid this pitfall in part by taking the time to visit educational programs, first and foremost to learn about what they are doing, second to explain new initiatives of the centralized authority, and third to ensure compliance.

Decentralization also enables people on the local scene to play a greater role in adapting their approach to assessment to their local circumstances. Institutions differ from one another in all sorts of important respects, including size, mission, personnel, and culture. Small institutions and large institutions face different challenges. Some institutions are focused on educating clinicians, while others have a greater research focus. Some institutions can draw on particular kinds of educational expertise, such as simulation-based methods of learning, which others lack. And institutional cultures vary widely, with varying levels of support for education, different degrees of emphasis between one-on-one and large-group instruction, and divergent attitudes toward the scholarship of teaching and learning. If educators enjoy a free hand, they can adapt assessment to their distinctive needs. To help with this, centralized assessment organizations can help to support the professional development of local educational leaders.

Decentralization also promotes the professional development of individual educators and program directors, by encouraging them to think, innovate, and assess for themselves. As physicians and human beings, these individuals have an innate need to grow and develop as fully as possible in their professional roles. For an educator, this means playing an active role in helping to develop not only curriculum and instructional methods but also assessment programs. When someone at a central assessment organization does the work of designing the assessment program and

merely requires local educators to implement it, it tends to stunt the development of the local educators. They may spend all their time and energy logging, documenting, counting, and checking off someone else's boxes rather than looking at and engaging their learners for themselves.

When it comes to educational assessment, a balance needs to be struck between centralized and decentralized approaches. Centralization promotes fairness and efficiency, while decentralization promotes engagement and creativity. Letting the pendulum swing too far in either direction is fraught with peril. For the moment, however, the pendulum is swinging in the direction of centralization, and it may already have passed the point at which risks and costs are outweighing any additional benefits. Radiology educators need to promote thoughtful and vigorous discussion around these issues, focused on deepening our understanding of the purposes and approaches of educational assessment and finding the sweet spot along the continuums between summative and formative, objective and subjective, formal and informal, and centralized and decentralized approaches to assessment.

References

Alderson PO, Becker GJ (2008) The new requirements and testing for American Board of Radiology certification in diagnostic radiology. Radiology 248:707–709

American Psychological Association (2011) Appropriate use of high-stakes testing in our nation's schools. Retrieved from http://www.apa.org/pubs/info/brochures/testing.aspx. Accessed 1 Nov 2011

Bloom BS (1971) Handbook on formative and summative evaluation of student learning. McGraw Hill, New York

Bloom BS, Englehart MD, Furst EJ, Hill WH, Krathwohl DR (1956) Taxonomy of educational objectives; Handbook I. Longmans, New York

Gipps CV (1994) Beyond testing: towards a theory of educational assessment. Routledge, London

Gunderman RB (2011) Achieving excellence in medical education, 2nd edn. Springer, New York

National Center for Fair and Open Testing (2011) Tests, cheating, and educational corruption. Retrieved from http://fairtest.org/sites/default/files/Cheating_Fact_Sheet_8–17–11.pdf. Accessed 1 Nov 2011

Nitko AJ (2001) Educational assessment of students, 3rd edn. Prentice-Hall, Des Moines