

Educational Abstracts

Abstracted by
Dorthea Juul, Ph.D.

Veloski JJ, Rabinowitz HK, Robeson MR, et al: Patients don't present with five choices: an alternative to multiple-choice tests in assessing physicians' competence. *Academic Medicine* 1999; 74:539-546

Veloski and his colleagues argue that standard multiple-choice questions (MCQs) have two significant limitations in the context of medical testing. First, examinees can arrive at the correct answer by recognition or guessing because it is presented along with several incorrect or less correct choices. In real life, the answer to a patient's problem must be generated by the physician. Second, MCQs are not generally used to test core knowledge because such items are too easy, and relatively trivial or esoteric topics are covered instead because they tend to lead to more discriminating items. As an alternative, the authors propose the uncued (UnQ) format in which examinees are presented with a test question and then select the correct answer from a long list (500+) of options. This format preserves the advantage of computerized scoring but minimizes arriving at the correct answer by recognition or random guessing. As a trial, items in the UnQ format were included in the 1990 and 1991 in-training examinations in family practice.

In 1990, there were 40 such items, 32 of which were also administered in the MCQ format and eight of which were newly created. There were 562 options, including diagnoses, physical findings, infectious organisms, laboratory tests, medications, and other management options, from which to select the best answer. The test was administered to 7,036 family practice residents and to 35 board-certified family practitioners who were item writers for the American Board of Family Practice. As hypothesized, the 32 items were more difficult in the UnQ format than in the MCQ format (59% correct vs. 76% correct), and the difference was greatest for the first-year residents, presumably the least knowledgeable. The reliability coefficients were 0.74 for the UnQs and 0.60 for the MCQs. However, the validity coefficients for the two formats were about equal, with the UnQ score accounting for 18% of the variance in level of

training and the MCQ score accounting for 15%. In 1991, 40 UnQs were developed to assess essential content in family practice, and they were administered to 7,138 family practice residents. More than half of the residents got all of the items correct, and the validity coefficient was 0.20, similar to those obtained in 1990.

The authors reported that both the residents and experienced item writers perceived the UnQ format to be a better measure of their abilities. Other advantages included less time to write items in the UnQ format and less time to read the UnQs, although filling in the optical scan sheet took longer because of the four-digit code associated with each response. Veloski et al. conclude that using items in this relatively open-ended format can strengthen the validity of examinations in medicine.

Clauser BE, Clyman SG, Swanson DB: Components of rater error in a complex performance assessment. *Journal of Educational Measurement* 1999; 36:29-45

This study extends previous work by the authors on the scoring of computer-based simulations of physician-patient encounters. Each encounter results in a transaction list detailing the actions taken by the examinee and the time frame for these actions. Two studies were undertaken to investigate rater error in the scoring of these problems. The first examined the effect of raters working within committees, and the second examined intrarater reliability. For the first study, three committees of five raters each rated the transaction lists for 100 fourth-year medical students who each completed four cases. The raters were academic clinicians who met in their committees for training and rating. They used a nine-point rating scale with defined anchor points. After completing their ratings independently, discrepancies within the committee were discussed, and the raters could modify their ratings if they wished. Generalizability analyses indicated that there was fairly good agreement within committees but substantial variation across committees. There were modest improvements in

within-committee agreement after discussion of rating discrepancies. However, these served to increase the between-committee differences in ratings.

In the second study, a single committee of four clinicians rated the transaction lists from 200 fourth-year medical students who completed eight cases. They rated the same lists a second time 6 to 18 months later. The occasion-related variance components were small, indicating that these raters were consistent across time.

The authors conclude that if raters work within committees, then the committee must be a facet in generalizability analyses, or inflated reliability estimates are likely to be obtained. Although rating occasion had minimal impact here, the authors also recommend including rating occasion as a facet in generalizability studies because that finding could not have been predicted *a priori*.

Xu G, Hojat M, Veloski JJ, et al: The changing health care system: a research agenda for medical education. *Evaluation and the Health Professions* 1999; 22:152-168

Xu and his colleagues describe several changes that have occurred in the health care delivery system, including the dramatic expansion of managed care in the 1990s and changes in the practice environment from solo to group practice. Their impact on four aspects of medical education—admissions, faculty, curriculum, and educational outcomes—is reviewed, and areas needing additional research are identified. After a decade of fairly dramatic increases in the size of the applicant pool, the number of people applying for medical school admission decreased in 1997 and further decline is predicted for the class entering in 1999. The authors suggest that the medical profession has become less appealing. Other admissions-related issues that may be affected are the demographics of the applicant pool and noncognitive characteristics of applicants, such as motivation, character, and sensitivity. The composition of medical school faculties has changed, with increasing numbers of practicing physicians coming from community settings. Faculty may have less time for and less commitment to teaching because of increased pressure to generate practice revenue or obtain research funding, and there are survey

data suggesting that medical students are not receiving the one-to-one teaching that they did in the past.

Curriculum issues include the operational definition of competency and the quality of students' clinical experiences. Competencies that are desirable from the perspective of integrated managed care systems, and which are not yet in place in most medical school curricula, are utilization review and quality assurance, health promotion and disease prevention, team work and leadership, cost-effective and decision-making analysis, communication and management skills, evidence-based medicine, epidemiology, and statistics. There is concern about whether the increased amount of time spent in ambulatory settings provides sufficient opportunity for instruction and supervision because of high patient volume. The fourth area is the effect of these changes on educational outcomes, including students' academic performance, specialty choice, practice location, licensure, and board certification.

Xu et al. note that "salaried HMO models tend to have the most demanding requirements in recruiting participating physicians, based on board certification or eligibility."

The authors outline ways in which changes in the health care delivery system have already affected medical education and propose research questions that should be fruitful areas for study as the health care system evolves.

Topping K: Peer assessment between students in colleges and universities. *Review of Educational Research* 1998; 68:249-276

Perhaps the most widespread contemporary innovation in medical education is problem-based learning, an important component of which is peer assessment. Hence, Topping's review of 67 research studies done on peer assessment in higher education settings is a useful resource for medical educators. Peer assessment is defined as "an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status."

The effects of peer assessment are organized into four categories: cognition and meta-cognition, affect, social and transferable skills, and systemic benefits.

Examples of positive outcomes include improved learning, enhanced critical analysis skills, active involvement in the learning process, development of teamwork skills, and greater insight into institutional assessment processes. Difficulties in implementing peer assessment are discussed, as are validity and reliability, which have been found to be acceptable in a wide number of applications.

Acceptability of peer assessment has been variable and unrelated to actual reliability. Students have reported finding peer assessment to be demanding but anxiety-reducing. An area of particular interest to medical educators is the peer assessment of professional skills. Topping reports that adequate reliabili-

ties have been obtained, and there is some research indicating that peer assessments are highly correlated with teacher assessments. Factors important in the effective implementation of peer assessment systems included clarifying expectations, developing and clarifying assessment criteria, monitoring the process, and coaching. The author concludes that while not enough research has been done to fully assess the value of peer assessment, that which has been done suggests that if "organized, delivered, and monitored with care, it can yield gains in the cognitive, social, affective, transferable skill, and systemic domains that are at least as good as these from staff assessment."