▲ A Model for Integrated Assessment of Clinical Competence

Karen J. Panzarella, PT, PhD
Andrea T. Manyon, MD

An interdisciplinary group of faculty from medicine, basic sciences, physical therapy, and education developed a performance assessment tool for evaluating clinical competence. This group was assembled following the revision to integrated systems-based curricula in the school of medicine and doctor of physical therapy program. The group was challenged to measure curricular outcomes through student assessment of clinical competence as defined through integration. The Integrated Standardized Patient Examination was developed as the assessment tool. This model utilizes standardized patients, who are trained to ask questions that require the students to integrate scientific knowledge and communicate this back to the patient. The student response is graded on a rubric and averaged with a history-taking portion of the examination. This model was administered to 140 first-year medical students who were randomly assigned to either an acute low back pain case or gastroesophageal reflux disease case. Clinical faculty scored the students in the treatment room, with additional faculty scoring in the monitor room. Interrater agreement was 87% for the low back pain case and 82% for the gastroesophageal reflux disease case. These results warrant further investigation of the Integrated Standardized Patient Examination for transfer to other health care professions, in particular physical therapy. J Allied Health 2007; 36:157–164.

FACULTY IS SPENDING substantial time and energy creating and implementing new curricula and is being held accountable to ensure students acquire a minimum level of clinical competence. While curricular reform and integrated curricula are well described and implemented in the health professions, meaningful measurements of competency to test the success of curricula have yet to be described, implemented, or validated.[1] Student assessment is an important aspect of any health profession curriculum. All assessment methods should be meaningful, applied consistently to all students, and linked to the intended learning outcomes.[2]

Traditional assessment methods have been perceived as distorting the learning process and favoring knowledge limited to recall, identification, or recognition. This report proposes a new model for assessment of clinical competence based on validation and pilot results with first-year medical students. The potential for application of this model to other health profession students is discussed.

Assessment, Integration, and Competence

Academic faculty must make conscientious decisions when choosing student assessments. Assessment drives teaching methods and learning and has powerful effects on student performance.[3,4] If students are queried only for their knowledge of facts, they tend to develop a learning strategy of rote memorization. If they are assessed for their capability to integrate their knowledge, students will develop a more deeply oriented learning strategy.[5] Students become more mindful and conscientious if exercises are observed and graded.[6] This "steering effect" of assessment pushes students to learn best those subjects on which they expect to be rigorously examined.[7]

During clinical encounters, health professionals integrate knowledge from basic science and clinical courses and apply these concepts during the process. For example, when a physical therapist examines a patient with a back injury, information acquired during courses on anatomy, physiology, biomechanics, musculoskeletal, and clinical foundations must be recalled and integrated.[8] Integration is a cognitive process that can be facilitated, but not guaranteed, by an integrated curriculum. The teacher provides a framework of learning opportunities, but integration of knowledge is done in the mind of the student, in the real world of clinical care.[9]

Clinical competence in health professions is an outcome of the curriculum and must include the integration of scientific knowledge with communication skills.[10] When students perform integration, they weed through their tacit knowledge, ruling in and ruling out information to make decisions they will communicate to their patients. If assessment of clinical competence evaluates the ability of students to apply and do something with their knowledge in a real or simulated clinical situation, they will likely improve at generalizing, integrating, applying, synthesizing, and interacting effectively.[11]

Dr. Panzarella is Director of Clinical Education for the Doctor of Physical Therapy Program at the University at Buffalo, Buffalo, NY; and **Dr. Manyon** is Chair of the Department of Family Medicine at Upstate Medical University, Syracuse, NY.

Received July 20, 2006; revision accepted March 1, 2007.

Address correspondence and reprint requests to: Karen J. Panzarella, PT, PhD, Program in Physical Therapy, Department of Rehabilitation Science, University at Buffalo, 515 Kimball Tower, Buffalo, NY 14214. Tel 716-829-3141 ext 151; fax 716-829-3217; e-mail kjp1@buffalo.edu.

In contrast, clinical performance in health professions educational programs is evaluated during the clinical portion of the curriculum at clinical sites away from, and outside the direct control of, the academic institution. The student's performance for these experiences is primarily determined by their clinical instructor (or preceptor) within their specific clinical site.[12] Clinical performance does not always predict clinical competence. Academic institutions often strive to incorporate real-life experiences into the classroom of their curricula to develop competencies that will translate to clinical performance.

Assessment Tools to Evaluate Clinical Competence

Academic programs have relied on several conventional methods for assessing clinical competence. Written examinations are most commonly used to assess clinical competency in medicine and health professional programs. Most popular is the multiple-choice question examination, which can demonstrate reliable results, is easy to administer, and has a low probability of guessing (commonly 0.20 to 0.25).[13] Multiple-choice question examinations were universally accepted for medical licensing examinations in the 1940s.[14] The Federation of State Boards of Physical Therapy currently uses multiple-choice question examinations for the National Physical Therapy Examination.[15]

Despite their popularity, multiple-choice question examinations are frequently poorly written, favoring knowledge limited to recall.[13] Other disadvantages include lack of content validity, the cueing effect of the options, and undesirable effects on student studying patterns.[3,16] A frequent criticism is that they may overestimate students' clinical problem solving and yield undesirable false-positive results.

Essay questions have a long history of use due to the advantage of measuring clinical judgment by asking an examinee to formulate problems and justify actions. Yet, the scoring is time consuming and shows poor reliability. In addition, the "halo effect" is a common error in scoring oral examinations where the rating of the student is high due to a global impression, causing the examiners to ignore the real data.[17] These limitations make them undesirable for assessing clinical competence.[18] Oral examinations were used in medical education for centuries until 1960, when a study by Hubbard found interrater reliability to be approximately 0.25.[19] Since then, the National Board of Medical Examiners has discontinued the use of oral examinations.

Practical examinations are the most prevalent performance tool for assessing competence in clinical skills within physical therapy education, despite the fact that they are not a component of the national licensing examination.[18] Typically, these examinations are generated from a single course and do not integrate concepts across the curriculum. Students perform various physical therapy assessment or treatment techniques learned that semester on a peer student or faculty member. The encounter is usually scored by one examiner using an arbitrary standard. Reliability or validity of practical examinations is usually not known or tested.

Since the 1980s, the Objective Structured Clinical Examination (OSCE) has been widely accepted in medical schools (>80% use) as a method of reliably assessing students' clinical skills.[20] The OSCE assesses what students can do in a structured fashion, with attention to the objectivity of the scoring.[21] The traditional OSCE consists of approximately 20 stations, and students spend 5 minutes with an examiner at each station.[22] Stations can be developed to test competencies such as history taking, patient education, communication, physical examination, diagnostic procedures, interpretation of findings, critical appraisal of literature, and even problem solving.[21] A score on the OSCE is usually determined by the percentage of correct checklist items performed within the given time frame.[23] Research demonstrates that multiple stations within the OSCE format generate scores with varied reliability (r = 0.20–0.95).[20]

Standardized patients (SPs) provide the greatest authenticity to an OSCE. SPs are lay people trained to represent a given diagnosis to ensure all students are presented with similar challenges. The first SP encounter occurred at the University of Southern California in 1963. For the purpose of teaching medical students, a faculty member trained a healthy artist's model to portray a paralyzed patient with multiple sclerosis.[24] The SP as a form of authentic testing has prompted other health professional programs to explore its use. The elegance of the SP examination is its ability to portray patient concerns of depression, anger, silence, cultural barriers, bad diagnostic news, death and dying, signs of domestic violence, or other ethical concerns that students may not always be exposed to in clinical practice.[25,26]

The content validity of the OSCE has been supported by a number of studies in which both the examiners and students rate it as appropriate and relevant to clinical practice.[27,28] The construct validity of the OSCE as a measure of competence has been demonstrated with students tested consecutively throughout their years of medical education.[28] Typically, the OSCE is administered as a summative examination in the last year of medical school, but recent research points out the benefits of introducing the OSCE in year 1.[10]

Nursing education has begun using SP encounters for experiences in assessing, diagnosing, and developing treatment plans for a variety of clinical problems. An advanced practice nursing program found that students' performance on one simulated clinical encounter did not reflect their performance on clinical evaluation measures or on national certifying examinations.[29] Of interest to note is that students who had poor performance on a nursing SP examination received satisfactory or excellent ratings from faculty or preceptors on clinical observations, demonstrating divergence of the SP examination from other forms of clinical assessment procedures. The effectiveness of using SPs for teaching in a nursing program was emphasized when com-

TABLE 1. Attributes of Integration, Diagnoses, and Patient Questions

Attributes of Integration	Diagnoses and Sample Patient Questions
To explain the relationship between structure and function	Herniated disc: "Why does it feel better when I tilt to the side?" Osteoporosis: "Will I have a hump in my back like my mother?"
To explain the role of injury in contributing to dysfunction	Ankle sprain: "Why do I continue to reinjure my ankle?" Herniated disc: "What specific movement caused this injury for me?"
To explain the scientific basis of the patient's symptoms	Atrial fibrillation: If I take a deep breath and then try to exhale without letting any air out, why does that sometimes stop the spell?" Herniated disc: "If this is a problem in my back, then why do I feel this sensation in my leg?"
To explain normal physiologic or biochemical mechanisms	Hypertension: "Why are two numbers given to describe my blood pressure?" Osteoporosis: "How does bone density change with age?"
To explain abnormal physiologic or biochemical mechanisms	Atrial fibrillation: "When I feel my pulse at my wrist, why is it so irregular—it appears to occur randomly in time and variably in strength?" Acute low back pain: "Why is sitting so uncomfortable?"
To explain the role of specific risk factors in the pathophysiology or prevention of a disease	Osteoporosis: "What sets me at risk for osteoporosis?" Hypertension: "What can I do to lower my blood pressure?"
To explain the rationale of a therapeutic intervention or diagnostic procedure and expected side effects	Atrial fibrillation: "A friend of mine has had a similar problem, and she takes Coumadin to prevent more serious problems. Will I have to do that, and what is the more serious problem?" Osteoporosis: "How does estrogen affect osteoporosis?"
To discuss prognosis; may include role of intervention, prevention, and/or second-degree complications	Hypertension: "What will happen if I do nothing at all for this problem?" Atrial fibrillation: "Can this be corrected and can I lead a normal life?"

pared with traditional lecture/laboratory experiences. Students who performed clinical skills with SPs had better clinical skills than the students who learned in the traditional lecture/laboratory approach.[30] The authors suggest the SP method should be used to evaluate competence in nursing students.

Use of the OSCE in physical therapy education is limited, and research on its outcomes is just beginning. One study of the OSCE with physical therapy students had poor internal consistency and did not predict clinical performance.[31] These OSCE stations were unlinked skill-related activities that did not mimic a real-life clinical encounter. Physical therapy educators have also used SPs for teaching purposes, not as a high-stakes integrative examination.[26,32,33] Recently, researchers have examined the effect of SP examinations on student confidence and attitudes.[34,35] To date, no studies in medicine or other health professions have examined the effect of the SP examination on students' ability to integrate the curriculum as a measurement of clinical competence.

Development of the Integrated Standardized Patient Examination

An interdisciplinary group of medicine, health-related, and education faculty assembled to develop an assessment tool that could measure clinical competence as defined through integration. The team's first step was to define integration in the context of a new integrated systems-based curriculum (in medicine and physical therapy). They defined integration in the context of medical education as "the ability to apply basic science concepts during each step of clinical practice" and concluded that clinical competency was defined as the student's ability to integrate scientific knowledge and clinical communication skills.[36] They concurred that when patients ask questions of physicians and physical therapists, they are testing the clinician's scientific knowledge. Similarly, when clinicians share this knowledge with patients, they are demonstrating their clinical communication skills.[36] A consensus was reached in the group that the key to measuring integration was to create commonly asked patient questions that could be used to measure a student's ability to integrate curriculum content.

The team took the supportive features of the OSCE and created the Integrated Standardized Patient Examination (ISPE), a performance assessment tool. They developed patient questions to prompt students to recall and integrate basic science knowledge and communicate back to the SP. The team worked to identify specific measurable attributes of integration for the question and answer portion of the case encounter, examples of which are listed in Table 1. Collectively, these attributes reflect the scope of possible clinical dialogue between a patient and a health care practitioner (Tables 1 and 2).

The ISPE is scored in three sections. Section I, the 15- to 20-item history portion, is a binary checklist (i.e., meets

TABLE 2. Rubric for Patient Assessment of Encounter with the Health Care Practitioner

Specific Measurable Attributes of Integration	3 Exceeds Expectations	2 Meets Expectations	1 Below Expectations	0 Detrimental
To explain the relationship between structure and function	Excellent explanation using a model or visual image, in a manner clearly understood by the patient	Good explanation may use model or visual image, in a manner likely to be understood by the patient	Explanation not clear or not supported by visuals, language does not match the patient's level of understanding	Confusing explanation, language confuses the patient
To explain the role of injury in contributing to dysfunction	Excellent explanation of dysfunction or injury using a model or visual image, in a manner clearly understood by the patient	Good explanation of dysfunction or injury may use model or visual image, in a manner likely to be understood by the patient	Explanation of dysfunction or injury not clear or not supported by visuals, language does not match the patient's level of understanding	Confusing explanation, language is confusing or insensitive to the patient
To explain the scientific basis of the patient's symptoms	Excellent explanation using a model or visual image, in a manner clearly understood by the patient	Good explanation may use model or visual image, in a manner likely to be understood by the patient	Explanation not clear for connection between symptoms and scientific basis; language does not match the patient's level of understanding	Explanation or information confuses the patient, may be unrelated to symptoms; language confuses the patient
To explain normal physiologic or biochemical mechanisms	Excellent explanation using a model or visual image, in a manner clearly understood by the patient	Good explanation may use model or visual image, in a manner likely to be understood by the patient	Explanation not clear or not supported by visuals, language does not match the patient's level of understanding	Confusing explanation, language confuses the patient
To explain abnormal physiologic or biochemical mechanisms	Excellent explanation using a model or visual image, in a manner clearly understood by the patient	Good explanation may use model or visual image, in a manner likely to be understood by the patient	Explanation not clear or not supported by visuals, language does not match the patient's level of understanding	Confusing explanation, language confuses the patient
To explain the role of specific risk factors in the pathophysiology or prevention of a disease	Excellent explanation, with education in prevention/wellness, in a manner clearly understood by the patient	Good explanation of risk and prevention, in a manner likely to be understood by the patient	Explanation not clear or not supported by visuals, language does not match the patient's level of understanding	Confusing explanation, language confuses the patient
To explain the rationale of a therapeutic intervention or diagnostic procedure and expected side effects	Excellent explanation of rationale for chosen intervention, in a manner clearly understood by the patient	Good explanation of rationale for chosen intervention, in a manner likely to be understood by the patient	Explanation of rationale not clear, language does not match the patient's level of understanding	Confusing explanation, language confuses the patient
To discuss prognosis; may include role of intervention, prevention, and/or second-degree complications	Excellent explanation using a model or visual image, in a manner clearly understood by the patient	Good explanation may use model or visual image, in a manner likely to be understood by the patient	Explanation not clear or not supported by visuals, language does not match the patient's level of understanding	Confusing explanation, language confuses the patient

expectations vs. below expectations).[21,37] Section II, the integration portion, consists of three to four patient questions focused on a particular disease or medical condition. Section III, the overall encounter, evaluates the quality of the student's entire interaction with the SP (Table 3). The overall encounter provides the rater four items to globally measure the student's performance. Both section II and III are scored using a four-point rubric. The total student score is a sum of the three sections, with a weighted emphasis on the integration and overall encounter sections. Students with scores falling below 70% reviewed their tapes for remediation with a faculty member and repeated the exercise with a new SP.

Reliability of the ISPE

The second phase of the tool development process was to establish the interrater reliability of the ISPE.

TABLE 3. Patient Assessment of Overall Encounter Quality with the Health Care Practitioner

	Please check box below statement.			
Student appeared comfortable conducting the interview; non-verbal communication and body language were appropriate, professional.	3 Exceeds expectations	2 Meets expectations	1 Below expectations	0 Detrimental
Student conducted the interview in a well-organized, smooth manner.	3 Exceeds expectations	2 Meets expectations	1 Below expectations	0 Detrimental
Student's speech was clear and easily heard.	3 Exceeds expectations	2 Meets expectations	1 Below expectations	0 Detrimental
Overall interface between knowledge and communication	3 The interface is fluent. Knowledge level and communication skills are consistently excellent.	2 The interface is reasonable. Knowledge level and communication skills are good.	1 Interface is awkward. Knowledge and communication skills are marginal.	0 The interface leads to confusion. Inadequate knowledge or communication skills.

METHODS

Twenty faculty members from the health-related professions (physical/occupational therapy and nursing), department of education, and school of medicine were recruited to participate in the reliability testing phase. Raters were given a 25-minute introduction regarding the nature of the tool and the assessment rubric. Two videotaped ISPE cases (low back pain and hypertension), contrasting student abilities, were presented in an auditorium to the faculty raters. Raters viewed each tape and scored the two students on each of the three sections of the ISPE (i.e., history, integration, and overall encounter). Focus groups were conducted following the scoring procedure to capture rater impressions of the ISPE.

RESULTS

Three different test statistics were used to test reliability. For the patient history (yes/no checklist), the percentage of agreement was 91% for the back pain case and 85% for the hypertension case. For the integration items and overall encounter sections, the within-group interrater reliability coefficient (r_{wg}) was calculated. This statistic assesses interrater agreement among several raters on a single stimulus object or person. Satisfactory reliability was found for both the integration items (r_{wg} = 0.74) and overall encounter quality (r_{wg} = 0.82) for the disc herniation checklist. Reliability coefficients for the hypertension case were lower for integration items (r_{wg} = 0.63) and overall encounter quality (r_{wg} = 0.66). Overall reliability of the coding instrument (calculated using Kendall's W) was 0.85 for the back pain case and 0.63 for the hypertension case.[38] The focus groups of experts concluded that the ISPE measured clinical competence as defined by the student's ability to integrate scientific knowledge and clinical communication skills through the use of the measurable attributes of integration.

Pilot Testing of the ISPE

Before pilot testing, the team created a gastroesophageal reflux disease (GERD) case to replace the hypertension case that demonstrated fair interrater reliability. The revised ISPE cases were administered to 140 first-year medical students enrolled in a clinical practice of medicine course.

METHODS

Twelve SPs were recruited to portray one of two cases. Six were trained to portray a person with a herniated lumbar disc, and the other six were trained to portray a person with GERD. The case scenarios for training the SPs were developed by clinical faculty who instructed the clinical practice of medicine course in the school of medicine. Training consisted of reviewing the written scenario, a demonstration of the 30-minute encounter with a faculty member playing a medical student, and each SP practicing their role with feedback from the trainer. All SPs were trained to score the students following the encounter on the overall encounter quality (Table 3).

The examination was conducted in a suite of patient rooms equipped with ceiling-mounted video cameras. Before entering the examination room, the student briefly reviewed the patient chart to note data confirming the patient's disease (e.g., a magnetic resonance imaging report confirming a disc herniation). The student then conducted a problem-focused history with the patient, summarized back to the patient, and asked if the SP had any questions. Here the SP asked the student four questions that corresponded to attributes of integration specific to the case. The student had a maximum of 15 minutes for the entire encounter.

Students were randomly assigned to one of the two clinical cases, and all encounters were videotaped. Ten faculty

members (in medicine and physical therapy) evaluated a median of 12 students each in the examining room. To estimate interrater reliability, 48 students were also observed by an additional four faculty members on a video monitor in an adjacent research room. A student could receive a maximum score of 45 points on the examination: 21 points for the history (section I), 12 points for the integration (section II), and 12 points for the overall encounter (section III).

Results

Students assigned to the low back pain case (n = 71) earned an average score of 37.58 (84%). Students assigned to the GERD case (n = 69) earned an average score of 35.36 (79%). The standard deviations for the low back pain and GERD cases were 3.92 and 5.41, respectively. A t-test yielded a statistically significant difference ($p < 0.01$) for the total examination scores between the cases. Further breakdown of the low back pain case and the GERD case show average section I (history) scores of 18.56 (SD 2.00) and 17.12 (SD 2.78), average section II (integration) scores of 9.17 (SD 1.62) and 9.00 (SD 1.79), and average section III (overall encounter) scores of 9.85 (SD 1.69) and 9.25 (SD 1.91), respectively, demonstrating that the students performed better on the low back pain case than the GERD case.

Interrater reliability was assessed for 22 of the low back pain cases and 26 of the GERD cases, randomly selected as a sample of total cases. Overall agreement between the examiners was 87% for low back pain and 82% for GERD. Cohen's ?, a measure of interrater agreement that adjusts for agreements by chance, was 0.75 and 0.68 for the two cases, respectively.[39] The interrater agreement for section I, the binary type responses (meets or below expectations) (low back pain, 0.94; GERD, 0.91), was significantly higher than that for sections II and III (integration questions and overall encounter) scored on the four-point rubric (low back pain, 0.66; GERD, 0.57). A matrix of agreement for the integration questions showed the biggest disparity in faculty agreement between students "meeting" or "exceeding" expectations (i.e., ratings of 2 or 3 on the rubric).

Discussion

The current research developed and evaluated a new performance assessment tool, the ISPE, for assessing clinical competence in medical and health professions education. The ISPE is distinctive from the OSCE in the manner in which the SP is used to require students to demonstrate their ability to integrate scientific knowledge with clinical communication skills, and it utilizes a scoring rubric that quantifies attributes of clinical competency.[21] The ISPE was designed by an interdisciplinary group of university faculty to ensure its applicability to a range of medical and health professions programs.

The results of this study show acceptable reliability via interrater agreement as well as interrater correlations on items that used a dichotomous scale, whereas the items requiring the use of the four-point rubric were somewhat less reliable. The rubric seems important, however, to capture the developmental component of competency and to display growth as students progress through their educational programs. Content validity was gathered in the process of developing cases and patient scenarios that were used in the ISPE and the focus groups.

The initial reliability results warrant further examination. Specifically, some faculty members had different understandings of the definition and characteristics of "meeting expectations" versus "exceeding expectations," resulting in 30% of variance attributable in part to rater bias or perhaps limited training/experience with the tool. In fact, one rater seldom rated students as meeting or exceeding his or her high expectations, whereas another rater consistently rated at exceeding expectations, representing the "hawk" and "dove" phenomena in rating.[17] In hindsight, the raters were not given sufficient training before the pilot and not all raters attended the validation session. Several raters left items blank and some even made up points on the rubric scale (e.g., 1.75), the effects of which were likely to attenuate the reliabilities. This variance was not attributable to the discipline of the rater because both were faculty in the school of medicine. Rater training should be considered to improve the reliability of the tool.[40]

This pilot test had medical students complete one of two cases: GERD or low back pain. The GERD case presented more difficulty for the students. Although the differences are not large, it is unfair to students to require cases with varying levels of difficulty when used for evaluation purposes. Standardization of case difficulty or utilizing only one case for all students is recommended to ensure fair testing.

The integration sections, the unique component of the ISPE, were judged to have a great deal of content validity from both the experts and students about questions directly relating to the ability to integrate. These questions were designed with great care to require the student to actually integrate and not just relay knowledge to the patient. The transdisciplinary team of experts that developed this instrument lends to its face validity, but future studies should be undertaken toward validation.

A concern often expressed about the use of SPs is their expense. SPs usually receive approximately $10 per hour for training and participation in the simulations.[32] Costs per student in 1994 for an OSCE-type examination were $34 per student.[41] This current pilot project cost approximately $50 per student for a 15-minute encounter. Costs beyond the SP include manpower to run the examination, make the necessary recordings, and document outcomes. Faculty time can be 10 to 20 hours, depending on the size of the faculty and involvement.

Future research with the ISPE should refine training

methods for raters to minimize potential biasing effects. Specifically, rater training should include practice trials with expert feedback to compare with a gold standard that averages ratings across three expert faculty raters. In addition, the use of the ISPE with year 2 and year 3 medical (or health professions) students should be examined to evaluate the ISPE for its ability to reflect increasing clinical competence that one would expect to occur over a three- or four-year professional program.

Conclusions

It is the social responsibility of academic institutions to judge program outcomes and promote only qualified students, and the current results suggest that the ISPE may be an avenue. One professor giving a practical examination in isolation versus the entire faculty being involved in an SP examination leads to greater validation and integration of curricular content.

Experience with the ISPE will validate the ability of the instrument to test students' integration of basic science knowledge with clinical skills and to shape the competent professional capable of functioning within today and tomorrow's health care society.

The essence of competence is the ability to do something under actual or simulated conditions. The evaluation of competence therefore is best achieved through performance-type assessments. "When the proof of the pudding is in the eating, there is no point in developing a multiple choice test; rather, have the students make pudding."[13] Performance assessments are not new to the fields of music, cooking, sport, driving, and even medicine. However, they are scantly used in the fields of health sciences: physical therapy, occupational therapy, nursing, and so on.[18,42]

Perhaps the most important implications of this research are the findings that resulted from cooperative efforts in different disciplines. This ISPE research made bridges between the schools of medicine and education and the department of physical therapy. It forced the faculty to look at "what really matters" when we send our graduates out to practice in our health care community.

References

1. Mavis B, Henry R, Ogle K, et al: The emperor's new clothes: the OSCE reassessed. *Acad Med* 1996; 71:447–453.
2. Vendrely A: Student assessment methods in physical therapy education: an overview and literature review. *J Phys Ther Educ* 2002; 16:64–69.
3. Frederiksen N: The real test bias. *Am Psychologist* 1984; 39:193–202.
4. McLachlan JC: Problem-based examining: a different approach to assessment. *Med Educ* 1997; 31:299–301.
5. Martenson D, Hakan A, Kerstin G: An integrated final examination in preclinical subjects for medical students: 10 years of experience. *Teach Learn Med* 1999; 11:26–33.
6. Epstein RM, Hundert EM: Defining and assessing professional competence. *JAMA* 2002; 287:226–235.
7. Kemahli S: Clinical teaching and OSCE in pediatrics. *Med Educ Online* 2001; 6(10).
8. Jafarey NA: Integrated curriculum. *J Pakistan Med Assoc* 2001; 51(2).
9. Harden RM: The integration ladder: a tool for curriculum planning and evaluation. *Med Educ* 2000; 34:551–557.
10. Feeley TH, Manyon AT, Panzarella KJ, et al: Toward validation of an assessment tool designed to measure medical students' integration of scientific knowledge and clinical communication skills. *Eval Health Prof* 2003; 26:222–233.
11. May W, Straker G, Foord-May L: *Opportunity Favors the Prepared: A Guide to Facilitating the Development of Professional Behavior*. Stoneham, MA: May & Associates Consulting; 2002.
12. Bridge PD, Ginsbrug KA: An integrated approach for evaluating students' achievement of clinical objectives. *Med Educ Online* 2001; 6:9.
13. Gentile JR: *Educational Psychology*, 2nd ed. Dubuque, IA: Kendall/Hunt Publishing Co; 1997.
14. Veloski J, Rabinowitz H, Robeson M, et al: Patients don't present with five choices: an alternative to multiple-choice tests in assessing physicians' competence. *Acad Med* 1999; 74:539–546.
15. Federation of State Boards of Physical Therapy: History of the Federation of State Boards of Physical Therapy. Alexandria, VA, FSBPT; 2005. www.fsbpt.org.
16. Newble DI, Baxter A, Elmslie RG: A comparison of multiple-choice tests and free-response tests in examinations of clinical competence. *Med Educ* 1979; 13:263–268.
17. Meier ST: *The Chronic Crisis in Psychological Measurement: A Historical Survey*. New York: Academic; 1994.
18. Nayer M: The assessment of clinical competency: an overview and preliminary report of Canadian physiotherapy programs. *Physiother Canada* 1995; 47:190–199.
19. Fowell SL, Bligh GH: Recent developments in assessing medical students. *Postgrad Med J* 1998; 74:18–24.
20. Doig C, Harasym P, Fick G, et al: An objective look at OSCE. *Acad Med* 2000; 75(10):96–98.
21. Harden RM: What is an OSCE? *Med Teach* 1988; 10(1):19–22.
22. Barrows H: An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med* 1993; 68:443–453.
23. Kahn M, Merrill W, Anderson D, et al: Residency program director evaluations do not correlate with performance on a required 4th-year objective structured clinical exam. *Teach Learn Med* 2001; 13(1):9–12.
24. Reteguiz J, Cornel-Avendano B: *Mastering the OSCE/CSA*. New York: McGraw-Hill; 1999.
25. Edinger W, Roberston J, Skeel J, et al: Using standardized patients to teach clinical ethics. *Med Educ Online* 1999; 4:4.
26. Ladyshewsky R, Gotjamanos E: Communication skill development in health professional education: the use of standardised patients in combination with a peer assessment strategy. *J Allied Health* 1997; 26:177–186.
27. Stillman P, Regan M, Swanson D, et al: An assessment of the clinical skills of fourth-year students at four New England medical schools. *Acad Med* 1990; 65:320–326.
28. Stillman P, Swanson D, Smee S, et al: Assessing clinical skills of residents with standardized patients. *Ann Intern Med* 1986; 105:762–771.
29. Vessey JA, Huss K: Using standardized patients in advanced practice nursing education. *J Prof Nurs* 2002; 18(1):29–35.
30. Yoo MS, Yoo IY: The effectiveness of standardized patients as a teaching method for nursing fundamentals. *J Nurs Educ* 2003; 42:444–448.
31. Wessel J, Williams R, Finch E, et al: Reliability and validity of an objective structured clinical examination for physical therapy students. *J Allied Health* 2003; 32:266–269.
32. Black B, Marcoux BC: Feasibility of using standardized patients in a physical therapist education program: a pilot study. *J Phys Ther Educ* 2002; 16(2):49–56.
33. Boissonnault W, Morgan B, Buelow J: A comparison of two strategies for teaching medical screening and patient referral in a physical therapist professional degree program. *J Phys Ther Educ* 2006; 20:28–35.

34. Hale L, Lewis D, Eckert R, et al: Standardized patients and multidisciplinary classroom instruction for physical therapist students to improve interviewing skills and attitudes about diabetes. *J Phys Ther Educ* 2006; 20:22–27.
35. Hayward L, Blackmer B, Markowski A: Standardized patients and communities of practice: a realistic strategy for integrating the core values in a physical therapist education program. *J Phys Ther Educ* 2006; 20:29–37.
36. Manyon A, Panzarella K, Feeley T, et al: Development of an assessment tool measuring medical students' integration of scientific knowledge and clinical communication skills. *Assessment Update* 2003; 15(1):1–14.
37. Hodges B, Reghr G, McNaughton N, et al: OSCE checklists do not capture increasing levels of expertise. *Acad Med* 1999; 74:1129–1134.
38. Siegel S: *Nonparametric Statistics for the Behavioral Sciences*. New York, McGraw-Hill; 1965.
39. Cohen J: A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20:37–46.
40. Martin JA, Rezknick RK, Rothman A, et al: Who should rate candidates in an objective structured clinical examination? *Acad Med* 1996; 71:447–453.
41. Cusimano M, Cohen R, Tucker W, et al: A comparative analysis of the costs of administration of an OSCE. *Acad Med* 1994; 69:571–576.
42. Bramble K: Nurse practitioner education: enhancing performance through the use of the objective structured clinical assessment. *J Nursing Educ* 1994; 33(2): 59–65.