Initial Reliability of the Standardized Orthopedic Assessment Tool (SOAT)

Mark R. Lafave, PhD, CAT(C)*; Larry Katz, PhD†; Tyrone Donnon, PhD†; Dale J. Butterwick, MSc, CAT(C)†

*Mount Royal College, Calgary, AB, Canada; †The University of Calgary, Calgary, AB, Canada

Context: Orthopaedic assessment skills are critical to the success of athletic therapists and trainers. The Standardized Orthopedic Assessment Tool (SOAT) has been content validated.

Objective: To establish interrater reliability of the SOAT.

Patients or Other Participants: Thirty-two college students, 10 raters, and 2 standardized patients (SPs) from Calgary, Alberta, Canada.

Design: Randomized observational study.

Intervention(s): Students were allowed 30 minutes to complete a mock orthopaedic assessment of an SP with an injury specific to a region of the body (shoulder, knee, or ankle). Using the region-specific SOAT, raters and SPs evaluated students' orthopaedic assessment skills.

Main Outcome Measure(s): The sum totals of the SOAT for 2 raters and 1 SP were used to calculate each student's performance scores for respective scenarios. Scale reliability analysis (Cronbach α) was completed on the SOAT for each of the 3 body-region examinations.

Results: The mean overall reliability of 3 SOATs (ie, ankle, knee, and shoulder) was positive: $\alpha = .85$ with the SP scores factored into the equation and $\alpha = .86$ without the SP scores factored into the equation. Reliability for the ankle region was highest ($\alpha = .91$), followed by the knee ($\alpha = .83$) and the shoulder ($\alpha = .82$).

original research

Conclusions: The study sample size was small, but the results will enable further study with generalization to a broader audience of athletic therapists and athletic trainers. Because a baseline measure of reliability was established using a robust statistical analysis, future researchers can employ more stringent statistical analysis and focus on the effects of various pedagogical techniques to teach and learn the underlying construct of clinical competence in orthopaedic assessment.

Key Words: musculoskeletal assessment, validity, Objective Structured Clinical Examinations, measurement and evaluation, education

Key Points

- The Standardized Orthopedic Assessment Tool had good interrater reliability for measuring athletic therapy student performance in assessing the shoulder, knee, and ankle.
- Providing rater judgment and decision making may add to the overall validity of the Standardized Orthopedic Assessment Tool to measure the construct of clinical competence in orthopaedic assessment.
- Using both detailed checklists and global rating scales, initially taking the extensive content validation steps, and requiring
 extensive training for each rater and standardized patient before the examination could have contributed to the reliability of
 the tool.
- The tool can be implemented into a larger-scale study to better evaluate its generalizability and effectiveness in measuring the construct of clinical competence in orthopaedic assessment.

S kills for assessing patients with musculoskeletal conditions are critical in the athletic therapy, athletic training,¹ and physical therapy professions.² They are also important for primary care physicians, who use these skills in 10% to 15% of all examinations.³ Teaching assessment skills is equally important but challenging, particularly if no educational standards are associated with them.¹ Student evaluation can yield clear signs of successful learning and teaching. A procedure that athletic therapists (ATs) in Canada employ to assess orthopaedic conditions was content validated⁴ and resulted in the presentation of the Standardized Orthopedic Assessment Tool (SOAT). Essentially, the SOAT was based on evaluation protocols originally described by Cyriax⁵ and further defined by Magee.⁶

Before the SOAT, no tool to measure orthopaedic assessment skills had been described in the literature.⁴ In brief, the SOAT was content validated using a modified Ebel

procedure as described by Butterwick et al.⁷ This procedure consisted of experts reviewing all tasks associated with a complete orthopaedic assessment. Each task was classified by major category commonly employed in an orthopaedic assessment: history, observation, scanning examination, clearing joints above and below the lesion site, examination (active, passive, and isometric resisted testing), special testing, palpation, and conclusion (diagnosis). The SOAT consists of identical history tasks for all regions of the body; its remaining components have unique tasks that are specific to each region. Experts reviewed all tasks associated with each body part (approximately 200 to 250 per SOAT for a specific body region) 3 times with different scenarios or conditions. The target was 80% expert agreement for tasks on a 3-level importance scale (essential, important, and not as important) to accomplish content validation of the tool. The standardized tool was finalized after further discussion

and electronic communication, and it met the goal of 80% expert agreement. The product was a tool that could measure the underlying construct of clinical competence in orthopaedic assessment.

A standardized tool to measure orthopaedic assessment skills is important in educational settings for several reasons. Classroom-based assessment rarely goes beyond face validation, which limits its generalizability beyond a specific school.8 Furthermore, the athletic therapy and athletic training professions, like most medical and paramedical professions, are accountable for providing a standardized accredited curriculum. Professional standards are in place to protect the public.9,10 Educational institutions must work in concert with professional governing bodies, licensing bodies (where applicable), and certifying agencies to ensure standards are maintained. A standardized assessment tool provides a mechanism to maintain standards through constructive alignment with professional competencies that tie all stakeholders to each other.¹¹ Assessment completes the teaching and learning cycle, and the assessment tool is merely the mechanism that facilitates the cycle. The process of tool development is critical and must be planned carefully to build an accurate measurement of the underlying construct.

To establish an instrument's validity and reliability, content validation is considered a crucial step in its evolution.¹² The next logical step in the development of such an instrument involves reliability testing. Hodges¹³ pleaded for qualified validity of assessment tools. Moreover, qualified validity indicates that a tool can be valid in certain circumstances, but those circumstances must be clearly defined so that the reader can make judgements about its generalizability. In that regard, the target audience for the SOAT is the athletic therapy profession in Canada. It takes several studies to test a measurement instrument's construct validity.8,14 We have established content validity of the SOAT, but we need to continue establishing the overall construct validity of the tool. Therefore, the purpose of our study was to establish baseline, and more specifically interrater, reliability of the SOAT.

METHODS

The Standardized Orthopedic Assessment Tool

The SOAT was originally developed to evaluate athletic therapy students' clinical skills or competence in orthopaedic assessment during performance-based examinations. The SOAT was designed to assess athletic therapy students during a 30-minute period through expert evaluation by 2 raters and 1 standardized patient (SP). Components of the SOAT are not unique because most athletic therapy, physiotherapy, and medical professions employ similar assessment protocols.5,6,15 However, the SOAT was the first documented assessment tool that underwent content validation to measure the orthopaedic assessment construct in a performance-based examination. A sample of the SOAT for a specific body region is provided in the Appendix of the online publication of this article, which is available at http:// www.nata.org/jat. Its major categories comprise history, observation, scanning examination, clearing joints above and below the lesion site, examination (including active range of motion, passive range of motion, and isometric resisted testing), special testing, palpation, and conclusion. Each category includes a checklist and a global rating scale. The history category also includes subsections with checklists and associated global rating scales.

Research Design

The original SOAT validation study focused on 6 regions of the body (shoulder, elbow, wrist and hand, hip, knee, and foot and ankle).⁴ In our study, 3 regions of the body were the foci: the shoulder, knee, and ankle. These regions are frequently injured and, thus, assessed by ATs, physiotherapists, and sports medicine specialists.¹⁶ We employed 1 scenario or diagnosis for each body region (shoulder: subacromial bursitis; knee: infrapatellar bursitis; ankle: lateral ankle sprain). A convenience-block randomized sampling method was used to ensure that at least 10 students were tested for each region of the body. Each student was randomly assigned 1 of the 3 regions of the body and was assessed by 2 independent expert raters and 1 SP using the SOAT (Table 1).

Sampling

Raters. Raters were solicited from the Alberta Athletic Therapists Association through an e-mail distributed to all certified members in Alberta (approximately 100 at the time of the study). The 10 raters who participated were clinicians and educators from educational institutions in the region (University of Calgary and Southern Alberta Institute of Technology), as well as community-based clinicians. The primary factor associated with choosing raters was their availability for specific time blocks. Raters were placed in time blocks most convenient for them. A time block was set according to rater availability for examining a minimum of 5 students at 45 minutes per student. We chose the raters and set the time blocks before student or scenario assignment. Raters were trained according to the protocol outlined in the Rater Training subsection.

Standardized Patients. The 2 clinical practicum course instructors from Mount Royal College acted as the SPs for the examinations. One SP rated students during 3 time blocks, whereas the other rated students during 2 time blocks (Table 1). The SPs were assigned to time blocks in a way that would limit fatigue. The SPs were trained according to the protocol described in the Rater Training subsection of this article.

Students. Participants consisted of 32 athletic therapy students at Mount Royal College who were enrolled in their first clinical practicum class, which was held during the winter semester of 2006. These students were selected as participants for this study because they were familiar with the performance-based testing procedures and the SOAT from the clinical practicum class. Participation in the study was voluntary. This study received ethics approval from the Mount Royal College Human Research Ethics Board, and strict student recruitment practices were followed. All students signed a voluntary consent form.

Rater Training

All raters and SPs attended a 3-hour orientation and training session (with breaks) for the SOAT (ie, the grading scheme and rating scales) that they would be using

Table 1. Research Design							
		Body Region					
5	<u>.</u>	and	_		Standardized		
Block	Student	Diagnosis	Exar	niner	Patient		
Block 1: April 19, 8 ам–12:15 рм			1	2	1		
	1	Shoulder					
	2	Knee					
	3	Knee					
	4	Knee					
	5	Shoulder					
	6	Ankle					
Block 2: April 20, 8 ам–12:15 рм			3	4	1		
	7	Shoulder					
	8	Shoulder					
	9	Ankle					
	10	Shoulder					
	11	Knee					
	12	Ankle					
Block 3: April 20,			5	6	2		
12:15 рм-5:15 рм							
	13	Knee					
	14	Knee					
	15	Ankle					
	16	Ankle					
	17	Shoulder					
	18	Knee					
	19	Ankie					
Block 4: April 21, 8 ам–12:15 рм			7	8	2		
	20	Ankle					
	21	Knee					
	22	Ankle					
	23	Ankle					
	24	Knee					
	25	Shoulder					
Block 5: April 21, 12:15 рм–5:15 рм			9	10	1		
	26	Shoulder					
	27	Shoulder					
	28	Knee					
	29	Knee					
	30	Shoulder					
	31	Ankle					
	32	Snoulder					

throughout the evaluation process, the purpose of the assessment, and the targeted audience. Moreover, great attention was paid to the rules for marking the iterative nature of the SOAT and, more specifically, the special testing section and the palpation component of the orthopaedic assessment. A complete description of this iterative nature of marking was outlined in Lafave et al.4 In brief, the SOAT permits examinees to examine a patient (or SP) using tasks that they see as important and relevant to the specific condition. Concomitantly, the rater must judge whether the examinee has made the appropriate choice or choices based on the answers to the questions posed in the history aspect of the orthopaedic assessment or testing within their overall SP assessment. As a result, 2 examinees may take markedly different pathways to a diagnosis (or conclusion), but the methods used in both pathways may be considered correct. Because raters must carefully scrutinize flexibility in an examinee's decision-making

process, the SOAT is more than just a dichotomous checklist, which was the original intent of the Objective Structured Clinical Examination (OSCE).^{17,18} This level of flexibility was designed to address some of the validity concerns of the OSCEs.^{19,20}

After the basic orientation to the rules of using the SOAT was complete, each SOAT for the shoulder, knee, and ankle was reviewed. The answer key was reviewed to train the SPs how to act and what to say when the students asked certain questions or performed specific tasks. Each task across the entire SOAT (approximately 200 to 250 tasks that were specific to a region of the body) was reviewed with both SPs and raters present. The interactive training included many opportunities to stop the presenter and ask for clarification.

Data Analysis

Scale reliability analysis (Cronbach α) was completed on the SOAT for each of the 3 body-region examinations. When applied in this way, the Cronbach α reliability coefficient is an intraclass correlation coefficient (ICC) or measure of interrater reliability.^{12,21,22} We acknowledge that generalizability theory testing is superior to ICC,^{23,24} but employment of the Cronbach α reliability coefficient is considered a 1-facet special case of generalizability theory and is most appropriate for this study, considering all limitations, which we outline in the Discussion.

The sum total of the SOAT for the 2 raters and 1 SP was used to calculate the students' performance scores on respective scenarios. The SPs completed a global rating scale for each major section and for each of the history subsections, whereas the raters completed the detailed task checklists in the SOAT and then completed the global rating scales. A sample of the detailed checklist from the history subsection on the nature of injury and a corresponding global rating scale are illustrated in the Figure. Global rating scales consisted of an overall impression for each of the major categories that composed the orthopaedic assessment. The raters' scores were actually a sum total of the detailed tasks combined with the global ratings for each category. Each task and global rating score was weighted equally (ie, each was scored as 1 point). In contrast, the SP data consisted of only the global ratings for each major section and the history subsection within the SOAT. Each global rating scale also was weighted equally for the SP grading sheet (ie, each was scored as 1 point).

Two separate reliability analyses were completed: 1 with the SP's scores included in the analysis and 1 with only the scores of the 2 raters.

RESULTS

As shown in Table 2, the mean overall reliability of all 3 SOAT regions of the body was similar whether the SP's scores were ($\alpha = .85$) or were not ($\alpha = .86$) included in the analysis. In other words, the difference between reliability scores with or without the SP's scores included produced an overall agreement rate of 99%.

DISCUSSION

The initial results indicate good interrater reliability for the SOAT for the specific scenarios that were used. The

Nature of Injury	Chief Complaint	0	1					
	Date of Injury/When Did Injury Occur	0	1					
	MOI/Recreation of MOI	0	1					
	Previous Injury to This Site	0	1					
	Previous Injury to Joint Above/Below	0	1					
	Functional Limitations	0	1					
	History Back or Neck Injury/Pain	0	1					
	Subjective/Patient SHARP	0	1					
	Unusual Sounds/Sensations	0	1					
Nature of Injury Global Rating Scale		0	1	2	3	4	5	

Figure. Sample of the Nature of Injury subsection of the History section from the Standardized Orthopedic Assessment Tool (SOAT). MOI indicates mechanism of injury; SHARP, swelling, heat-altered function, redness, pain.

ankle assessments had the highest reliability coefficients (α = .91), followed by the shoulder and knee when the scores of the SPs were included in the grading process. The reliability of the ankle assessment may have been highest because it had the least number of special tests and likely had the least complex scenario compared with the shoulder and knee assessments. Future testing with more scenarios is needed to compare SOAT reliability of 1 joint with another.

During the content validation study,⁴ some athletic therapists anecdotally questioned whether an assessment tool that permitted such great flexibility by examinee and examiner would allow consistent grading. As our study shows, the initial scale reliability results of the SOAT support the need for further research into its use as a standard protocol for orthopaedic assessment.

Traditional OSCEs commonly employed in medical education to measure a clinical construct are limited by binary tools.^{17,18} In contrast, the SOAT employs both binary and global ratings of performance. Some authors have criticized traditional OSCEs, stating that they have a tendency to trivialize the underlying construct that they attempt to measure and, thus, call into question the overall validity.^{19,20,25} The trivialization seems to have 2 reasons:

Table 2. Reliability Statistics by Body Region Described With the Cronbach $\boldsymbol{\alpha}$

	Relia	bility
Body Region and Diagnosis	Standardized Patient Data Included	Standardized Patient Data Not Included
Ankle	.91	.91
Knee	.83	.91
Shoulder	.82	.76

(1) all scales are binary, thus removing the expert opinion of the performance.²⁵ and (2) scales are traditionally a list of detailed tasks that represent the underlying construct.^{19,20} In contrast, a rater using the SOAT is grading more than just a dichotomous scale. Raters are making judgements based on the individual's performance rather than on a predetermined set of answers. In addition, tasks and grades are associated with the examinee's ability to draw connections between categories, thus permitting complete flexibility in decision making throughout the orthopaedic assessment. In other words, the SOAT requires raters to use expert judgement when grading examinees and changes how raters evaluate each examinee relative to another. For example, if an examinee decided to use the Lachman test for anterior cruciate ligament stability, but another examinee chose to use the anterior drawer test, both choices could be considered correct if applied appropriately. Furthermore, after the examinee chooses a test, the rater can choose not only to grade whether he or she thinks that the examinee selected the correct test for anterior cruciate ligament stability but also to grade the examinee's performance of the test using a continuous scale, rather than just stating whether the examinee did the test or did not do the test. Provision of rater judgement and decision making may add to the overall validity of the SOAT to measure the underlying construct, particularly as it relates to traditional OSCEs.

In addition to rating students on detailed tasks within each major category and history subcategory, raters also can rate students on global scales. Researchers originally thought that dichotomous scales would increase reliability, but authors of many studies with a global rating scale have shown that this theory is false.^{25–27} As a result, the SOAT was designed as a hybrid of detailed checklists and global rating scales. Both the addition of global rating scales to the detailed checklist and the iterative nature of grading examinees with the SOAT may have helped address concerns of validity with OSCEs (or practical, performance-based examinations).

Conceptually, the SOAT has a slightly different approach from the one Denegar and Fraser¹ proposed. They recommended that evidence-based decisions for special tests be employed during the physical examination. In support of this approach, some authors of meta-analyses have proposed the superiority of the sensitivity and specificity of some special tests compared with others.^{28,29} However, even if 1 special test demonstrates superior diagnostic power compared with another, how each special test integrates into an overall orthopaedic assessment has not been shown. Ideally, a standardized assessment protocol should be developed for the evaluation of all patients with knee injuries. Yet those protocols will likely be limited by the many factors associated with each case. Based on these limitations, the SOAT was designed to address concerns that OSCEs tend to trivialize content, bringing into question the overall validity of the measurements.19,20,25

Strong reliability of the scales may be attributed to the extensive content validation steps taken initially in the process of developing the rating scale. In addition, a thorough rater training session may also contribute to the tool's overall reliability.⁴ Some investigators have not reported the validation process with performance-based examinations, leading readers to infer that this step in the overall validation may not have been established before measuring the reliability of the tool employed.^{30,31} Ignoring the initial content validation phase in the development of competency measures may result in lower overall scale reliability.¹² In contrast, good reliability does not ensure good validity. A balance and constant interaction between validity and reliability is critical to an instrument's evolution toward construct validity.³²

A final explanation of the SOAT's reliability could be the extensive training required for each rater and SP before the examination. Part of the training process included a review of the rules associated with the SOAT. The SOAT rules and assumptions for use originally were published by Lafave et al.⁴ The SP and rater training session was standardized through a common PowerPoint (version 2003; Microsoft Corp, Redmond, WA) presentation and set of detailed instructions on how the tool was to be used during the examination. The training session was interactive, permitting rater trainees to ask questions and gain clarification based on the specific scenarios. Although the training session took approximately 3 hours to complete, the explicit training on the procedures for using the SOAT may have contributed to the resulting strong reliability coefficients.

Limitations

One limitation of our study was the research design. Ideally, a fully crossed generalizability design, in which each examinee is tested by the same raters across multiple cases, is warranted. However, the SOAT is unlike traditional OSCEs that have a station length of approximately 10 to 15 minutes and multiple stations to measure the same underlying construct. Rather, the SOAT is designed so that the history, physical examination, and

interpretation are not only part of the same station over a 30-minute period but also rely on the subsequent section for rater scoring. Two 30-minute stations multiplied by 30 students in 45-minute time slots equates to 45 hours of examining. Practicality was one of the main psychometric issues that Harden and Gleeson¹⁸ raised. Even if ample rest is provided between testing time blocks, testing students for 45 hours does not seem practical. Thus, for practical reasons, we involved multiple examiners in our study, limiting the research design and the statistical analysis employed (ie, ICC or Cronbach α).

Another limitation of our study was that participants were limited to the student volunteers from the clinical practicum class at Mount Royal College in the winter semester of 2006. This convenience sample may not permit the results to be generalized to other athletic therapy student populations in Canada or elsewhere. Although generalizability theory may have been a solution to this limitation, the rationale for not employing that technique is explained in the preceding paragraph. Another solution may include further testing of a broader population outside the Mount Royal College environment. In future studies, investigators need to focus on using the SOAT with a broader audience beyond athletic therapy students at Mount Royal College.

CONCLUSIONS

The SOAT has shown promise as an effective instrument in measuring performances of athletic therapy students in OSCE-type examinations. Our study was a critical step in determining whether the SOAT could be implemented into a larger-scale study with a larger group of participants across Canada to better evaluate its generalizability and ultimate effectiveness in measuring the orthopaedic assessment clinical construct. In future studies, we will integrate the SOAT into the curricula of 4 athletic therapy programs across Canada (University of Manitoba, University of Winnipeg, Concordia University [Montreal], and the University of Calgary/Mount Royal) and include various exposures for both students and instructors to measure its effectiveness as a teaching and learning tool. This future study will have a larger number of participants so that reliability testing and generalizability analysis (where feasible) can occur, and stronger construct validity for the SOAT can be built. After overall reliability and validity have been established with the athletic therapy population, the SOAT can be tested in other medical and allied health professional environments.

ACKNOWLEDGMENTS

We thank the volunteers of the Alberta Athletic Therapists Association and its members for their support of our study. Specifically, we thank Ms Meryl Wheeler and Dr Khatija Westbrook for their assistance in examining students. Finally, we thank the University of Calgary Athletic Therapy Education Group for its guidance throughout this process.

REFERENCES

 Denegar CR, Fraser M. How useful are physical examination procedures? Understanding and applying likelihood ratios. J Athl Train. 2006;41(2):201–206.

- 2. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther.* 2001;81(9):1546–1564.
- Lawry GV 2nd, Schuldt SS, Kreiter CD, Denson P, Albanese MA. Teaching a screening musculoskeletal examination: a randomized, control trial of different instructional methods. *Acad Med.* 1999;74(2): 199–201.
- Lafave M, Katz L, Butterwick D. Development of a content-valid standardized orthopedic assessment tool (SOAT) [published online ahead of print January 3, 2007]. *Adv Health Sci Educ Theory Pract.* doi: 10.1007/s10459-006-9050-2.
- Cyriax JH. Textbook of Orthopaedic Medicine. 8th ed. London, England: Bailliere Tindall; 1982. Diagnosis of Soft Tissue Lesions; vol 1.
- Magee D. Orthopedic Physical Assessment. 4th ed. Toronto, ON: WB Saunders; 2002.
- Butterwick DJ, Paskevich DM, Lagumen NG, Vallevand AL, Lafave MR. The development of content-valid technical skill assessment instruments for athletic taping skills. J Allied Health. 2006;35(3):147–155.
- 8. Violato C, Marini A, McDougall D. Assessment of Classroom Learning. Calgary, AB: Detselig Enterprises Ltd; 1998.
- 9. Haladyna TM. A research agenda for licensing and certification testing validation studies. *Eval Health Prof.* 1994;17(2):242–256.
- 10. Cavanaugh SH. Response to a legal challenge: five steps to defensible credentialing examinations. *Eval Health Prof.* 1991;14(1):13–40.
- 11. Biggs J. Enhancing teaching through constructive alignment. *J Higher Educ.* 1996;32(3):347–364.
- 12. Hopkins KD. Educational and Psychological Measurement and Evaluation. 8th ed. Toronto, ON: Allyn & Bacon; 1998.
- 13. Hodges B. Validity and the OSCE. Med Teach. 2003;25(3):250-254.
- Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. 3rd ed. New York, NY: Oxford University Press; 2003.
- 15. Shultz SJ, Houglum PA, Perrin DH. *Examination of Musculosketetal Injuries*. 2nd ed. Champaign, IL: Human Kinetics; 2005.
- 16. Caine DJ, Caine CG, Lindner KJ. *Epidemiology of Sports Injuries*. Champaign, IL: Human Kinetics; 1996.
- Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *Br Med J*. 1975;1(5955):447–451.
- Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ*. 1979;13(1):41–54.

- Norman GR, Van der Vleuten CP, De Graaff E. Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Med Educ.* 1991;25(2):119–126.
- 20. Van der Vleuten CP, Norman G, De Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ*. 1991;25(2):110–118.
- Armstrong GD. The intraclass correlation coefficient as a measure of interrater reliability of subjective judgments. *Nurs Res.* 1981;30(5): 314–320.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.
- Brennan RL. Generalizability theory. *Educ Measure*. 2003;40(1): 105–107.
- Brennan RL. Performance assessment from the perspective of generalizability theory. *Appl Psychol Measure*. 2000;24(4):339–353.
- 25. Van Luijk S, Van der Vleuten C. A comparison of checklists and rating scales in performance based testing. In: Hart IR, Harden RM, Des Marchais J, eds. *Current Developments in Assessing Clinical Competence*. Montreal, QC: Can-Heal; 1992:357–382.
- Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med.* 1998;73(9):993–997.
- Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med.* 1999;74(10):1129–1134.
- Solomon DH, Simel DL, Bates DW, Katz JN, Schaffer JL. The rational clinical examination: does this patient have a torn meniscus or ligament of the knee? Value of the physical examination. *JAMA*. 2001;286(13):1610–1620.
- Benjaminse A, Gokeler A, van der Schans CP. Clinical diagnosis of an anterior cruciate ligament rupture: a meta-analysis. J Orthop Sports Phys Ther. 2006;36(5):267–288.
- Cohen R, Reznick RK, Taylor BR, Provan J, Rothman A. Reliability and validity of the objective structured clinical examination in assessing surgical residents. *Am J Surg.* 1990;160(3):302–305.
- Winckel CP, Reznick RK, Cohen R, Taylor B. Reliability and construct validity of a structured technical skills assessment form. *Am J Surg.* 1994;167(4):423–427.
- 32. Messick S. Standards of validity and the validity of standards in performance assessment. *Educ Measure*. 1995;14(4):5.

Mark R. Lafave, PhD, CAT(C), contributed to conception and design; acquisition and analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Larry Katz, PhD, contributed to conception and design and critical revision and final approval of the article. Tyrone Donnon, PhD, contributed to analysis and interpretation of the data and drafting, critical revision, and final approval of the article. Dale J. Butterwick, MSc, CAT(C), contributed to conception and design; acquisition and analysis and interpretation of the data; and drafting, critical revision, and final approval of the article.

Address correspondence to Mark R. Lafave, PhD, CAT(C), Department of Physical Education and Recreation Studies, Mount Royal College, 4825 Mount Royal Gate SW, Calgary, Alberta, Canada T3E 6K6. Address e-mail to mlafave@mtroyal.ca.