A Generalizability Theory Study of Athletic Taping Using the Technical Skill Assessment Instrument

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

*Department of Physical Education and Recreation Studies, Mount Royal University, Calgary, AB, Canada; †Faculty of Kinesiology, University of Calgary, AB, Canada

Context: Athletic taping skills are highly valued clinical competencies in the athletic therapy and training profession. The Technical Skill Assessment Instrument (TSAI) has been content validated and tested for intrarater reliability.

Objective: To test the reliability of the TSAI using a more robust measure of reliability, generalizability theory, and to hypothetically and mathematically project the optimal number of raters and scenarios to reliably measure athletic taping skills in the future.

Setting: Mount Royal University.

Design: Observational study.

Patients or Other Participants: A total of 29 university students (8 men, 21 women; age $= 20.79 \pm 1.59$ years) from the Athletic Therapy Program at Mount Royal University.

Intervention(s): Participants were allowed 10 minutes per scenario to complete prophylactic taping for a standardized patient presenting with (1) a 4-week-old second-degree ankle sprain and (2) a thumb that had been hyperextended. Two raters judged student performance using the TSAI.

Main Outcome Measure(s): Generalizability coefficients were calculated using variance scores for raters, participants,

and scenarios. A decision study was calculated to project the optimal number of raters and scenarios to achieve acceptable levels of reliability. Generalizability coefficients were interpreted the same as other reliability coefficients, with 0 indicating no reliability and 1.0 indicating perfect reliability.

Results: The result of our study design (2 raters, 1 standardized patient, 2 scenarios) was a generalizability coefficient of 0.67. Decision study projects indicated that 4 scenarios were necessary to reliably measure athletic taping skills.

Conclusions: We found moderate reliability coefficients. Researchers should include more scenarios to reliably measure athletic taping skills. They should also focus on the development of evidence-based practice guidelines and standards of athletic taping and should test those standards using a psychometrically sound instrument, such as the TSAI.

Key Words: validity, reliability, evidence-based practice, objective structured clinical examinations, measurement and evaluation, education

Key Points

- The generalizability coefficient indicated moderate reliability.
- More scenarios should be used to reliably test students using the Technical Skill Assessment Instrument.
- Researchers should develop evidence-based practice guidelines and standards of athletic taping and should test those standards with a psychometrically sound instrument, such as the Technical Skill Assessment Instrument.

thletic taping is a cornerstone of athletic therapy and training and has been useful in reducing the incidence of some injuries.¹ Further, it is a core competency in athletic therapy programs in Canada and the United States (Tables 1 and 2).²⁻⁵ Candidates for certification by the Canadian Athletic Therapists Association are required to complete 2 athletic taping techniques in a practical, performance-based examination that has not been psychometrically established.³ Candidates for certification by the Board of Certification are not required to demonstrate athletic taping skill proficiency in a practical, performance-based examination.⁶ Despite the importance and time dedicated to teaching athletic taping skills, few authors of peer-reviewed studies have measured Canadian or American professional standards. Furthermore, evidence of standards and expectations from accredited programs also appears to be lacking. In fact, only 3 peer-reviewed studies^{7–9} on the standards (content validation) or student performance expectations have been published. Two of the

articles^{7,8} on content validation and standards offer a very low level of evidence. Content validation is a process whereby expert consensus is sought for the content of a test or an examination.¹⁰ Expert consensus is at the lower end of the evidence-based practice scale based on current standards.^{11–13} The third article⁹ was related to intrarater reliability. When combined with the other studies, it was a good initial step to establish the validity and reliability of the Technical Skill Assessment Instrument (TSAI) as a measure of the technical taping skills of students in an academic program.^{14,15} However, overall, a substantial gap exists in evidence as it relates to evaluation of athletic taping.

Practical, objective, and structured performance-based examinations are considered the criterion standard in the medical profession to evaluate clinical competence, including its psychomotor (technical skills) aspects.^{16–19} However, the athletic therapy and training profession seems to be lagging behind medical education trends to assess

Table 1.	Athletic	Taping	Core	Competencies	From the	Canadian	Athletic	Therapists	Association ²
----------	----------	--------	------	--------------	----------	----------	----------	------------	--------------------------

Domain	Athletic Taping Competency			
I: Prevention				
Cognitive	Basic components of a comprehensive athletic injury/illness-prevention program, including (a) physical examinations and screening procedures; (b) physical conditioning; (c) fitting and maintenance of protective equipment; (d) application of taping, special pads, etc; and (e) control of environmental risks			
Psychomotor	Selection, fabrication and application of appropriate preventive taping, wrapping, splints, braces, and other special protective devices consistent with sound anatomical and biomechanical principles			
IV: Rehabilitation				
Cognitive	Contemporary immobilization devices (eg, casting materials, splints) and special protective/correction equipment (eg, braces, special pads, modified taping procedures, orthotics)			
	Comparative effectiveness of taping and bandaging, special padding, and standard protective equipment as related to the safe return of injured athletes to competition			
Psychomotor	Application of special protective devices (eg, braces, splints, special pads) and taping, bandaging, and wrapping procedures			

clinical competence, particularly as it relates to athletic taping competence. Perhaps a lack of evidence in this realm exists because of some of the shortcomings of performance-based examinations.²⁰ Criticisms of objective structured clinical examinations (ie, performance-based examinations) include cost and lack of validity, fidelity, and reliability.²⁰ Recently, more emphasis has been placed on a comprehensive evaluation plan for students in programs that may include performance-based examinations and workplace evaluation.^{21,22}

The TSAI was developed to assess the technical components of athletic taping. It has demonstrated content validity and intrarater reliability.7,9 To establish the construct validity of a measurement instrument, researchers must conduct a number of validation studies, the first of which is content validation.^{14,15} In addition, to measure clinical competence at a more global level (eg, Is student X a good athletic taper?), researchers need to complete a generalizability study whereby they test the tool for reliability among examiners, establish the optimal number of examiners, establish the optimal number of stations, and determine the total number of patients needed. Generalizability theory study design facilitates answers to those underlying questions so that valid and reliable examinations are implemented in medical and paramedical programs.²³⁻²⁵ Therefore, the purpose of our study was to test the reliability of the TSAI using a more robust measure of reliability, generalizability theory, and to hypothetically and mathematically project the optimal number of raters and scenarios to reliably measure athletic taping skills in the future.

METHODS

Design

We used a 2-facet, fully crossed, generalizability theory design for this study (Figure). The 2 facets of interest were

raters and scenarios. Specifically, 2 raters judged the performance of 29 participants on 2 ankle-taping scenarios. Generalizability theory is beneficial for evaluating the reliability of practical, performance-based examinations because it can measure the error associated with facets or variables thought to contribute to the overall error associated with measurement.²³⁻²⁶ Essentially, error is measured as a source of variance, and generalizability theory permits one to determine the amount of variance for which each facet is responsible in the total error in the examination.^{23–26} The other interesting aspect of generalizability theory is that after the generalizability coefficient has been calculated, researchers and educators can use those data, manipulating the number (ie, sample size) of scenarios or raters, to calculate or predict the optimal number of raters or scenarios necessary to achieve acceptable reliability coefficients that would make the examination psychometrically sound. These projections are called decision (D) studies.²³⁻²⁶

Participants

A total of 29 participants (8 men, 21 women; age = 20.79 ± 1.59 years) were chosen from a convenience sample of third-year undergraduate kinesiology students majoring in athletic therapy at Mount Royal University, which is a small (12 000 full-time students), publically funded program accredited by the Canadian Athletic Therapists Association. All participants provided written informed consent, and the study was approved by the Human Research Ethics Board of Mount Royal University.

Instrumentation

Two raters (M.R.L. and D.J.B.) used the TSAI to evaluate participant performance. They had been postsecondary educators for 18 and 28 years, respectively, and had gained much experience and exposure to the TSAI when using it for previous testing. The TSAI used to

Table 2. Athletic Taping Core Competencies From the National Athletic Trainers' Association⁴

Domain	Athletic Taping Competency			
PHP-23	Apply preventive taping and wrapping procedures, splints, braces, and other special protective devices.			
TI-16	Fabricate and apply taping, wrapping, supportive, and protective devices to facilitate return to function.			
CIP-2	Select, apply, evaluate, and modify appropriate standard protective equipment, taping, wrapping, bracing, padding, and other custom devices for the client/patient to prevent and/or minimize the risk of injury to the head, torso, spine, and extremities for safe participation in sport or other physical activity.			



Figure. Venn diagram representing a fully crossed generalizability theory design with 2 facets.

evaluate ankle and thumb taping consists of a 60-item checklist that samples such factors as materials used, starting position of the joint, taping techniques used, and posttaping effectiveness. Grading participants using the TSAI consists of removing a mark or point if the rater believes the student did not complete an item and leaving the item if the student completes it adequately based on the rater's professional judgment. The number of marks removed at the end is subtracted from the total number of items for each scenario. A minimal passing level was established when the TSAIs, including the scenarios, items, and weighting of each item, were content validated using a modified Ebel procedure, which is a weighting system of importance and difficulty for each item.⁷ The minimal passing level was 40/60 for the ankle scenario and 41.7/60 for the thumb scenario.⁷

Procedures

Participants were assigned randomly to testing time slots across a 2-day period during which they were required to complete athletic taping of the ankle and thumb in random order. One male, second-year graduate student served as the standardized patient for each scenario. The ankle and thumb scenarios had undergone content validation and intrarater reliability testing.^{7,9} For the ankle scenario, the standardized patient presented as a college soccer player who had sustained a second-degree sprain of the calcaneofibular and anterior talofibular ligaments 4 weeks earlier, was fully rehabilitated, and was preparing to participate in a game. For the thumb scenario, the standardized patient acted as a college football player (wide receiver) who had hyperextended his thumb within the year before presentation and wanted the thumb taped for prophylactic reasons. Participants were given the scenario information, were allotted 10 minutes to complete each scenario, and were stopped and graded accordingly at the 10-minute mark. The raters used their professional expertise and judgment to grade the participants over the 2-day period. They were blinded from each other when grading performance.

Data Analysis

An analysis of variance was used to estimate the variance in student scores because each variance component tested may contribute to the error in measurement. The 3 main effects in our study were raters, scenarios, and participants. The three 2-way interactions between main effects (raters \times scenarios, raters \times participants, scenarios \times participants) and the 3-way interaction effect (raters \times scenarios \times participants) were confounded with random error as a function of the fully crossed design. We used SPSS (version 17; IBM Corporation, Armonk, NY) to calculate the variance components. Generalizability coefficients and the D study were calculated manually using the following formula:

$$\mathrm{E} \mathbf{p}_{\delta}^{2} = \frac{\delta_{p}^{2}}{\delta_{p}^{2} + \left[\frac{\delta_{ps}^{2}}{\mathbf{n}_{s}} + \frac{\delta_{pr}^{2}}{\mathbf{n}_{r}} + \frac{\delta_{prs}^{2}}{\mathbf{n}_{s}\mathbf{n}_{r}}\right]},$$

where *p* indicates participants; *r*, raters; *s*, scenarios; Ep_{δ}^2 , generalizability coefficient; σ , variance; and n, the number of scenarios (n_s) or raters (n_r) .²⁵ A generalizability coefficient is interpreted in the same way other reliability coefficients are interpreted on a scale from 0 to 1.0, with 0.70 targeted as a minimal level for psychometric soundness.¹⁵ However, the generalizability coefficient is a much more robust statistic and, thus, represents a stronger indication of the tool's reliability.

RESULTS

The mean score for the ankle scenario across participants and raters was 69.47%. The mean score for the thumb scenario across participants and raters was 82.40%. The minimal passing level established in the content validation study was 66.7% for the ankle scenario and 69.5% for the thumb scenario.⁷ The variance components for testing the participants across 2 taping scenarios are listed in Table 3. The overall generalizability coefficient for testing taping clinical competence was $E\rho_{\delta}^2 = 0.67$ for the 2-rater, 2scenario design in this study. A D study was calculated to project reliability coefficients for rater or raters and scenario or scenarios (Table 4). As noted, the D study is a hypothetical calculation whereby the number of raters and scenarios is manipulated to achieve the 0.70 target reliability coefficient. Based on these hypothetical projections of the D study, 4 scenarios with 2 examiners would be needed in future testing to achieve a reliability coefficient of 0.70. Manipulation of the rater facet was less dramatic and, thus, not a factor for consideration in future studies.

DISCUSSION

Generalizability coefficients are interpreted in a similar fashion to other, more commonly used reliability coefficients, such as intraclass correlation coefficients or the Cronbach α reliability coefficient.²⁵ The scale ranges from 0 to 1.0, whereby 1.0 represents perfect reliability but scores ranging from 0.70 to 0.90 are optimal.^{14,15} The generalizability coefficient with our study design was 0.67, slightly missing the target of 0.70.

Our study had 2 facets of interest: raters and scenarios. The raters accounted for the least amount of total variance (ie, 5.03%). The D study demonstrated that increasing the

Table 3. Variance Component Results for 29 Participants, 2 Raters, and 2 Taping Scenarios

Source of Variation	Degrees of Freedom	Mean Squares	Variance Component	Variance Explained, %
Participants	28	124.905	35.512	12.73
Raters	1	21.882	14.027	5.03
Taping scenarios	1	794.708	189.209	67.84
Participants \times raters	28	47.210	19.166	6.87
Participants \times taping scenarios	28	33.089	12.105	4.35
Raters \times taping scenarios	1	4.090	-0.165	0.00
Participants × raters × taping scenarios ^a	28	8.879	8.879	3.18
Total			278.898	100.00

^a Indicates that the product is confounded by random error.

number of raters does not considerably improve the overall reliability. These results are consistent with the results others have found with practical, performance-based examinations, such as objective structured clinical examinations.²⁷

The data demonstrated that most (67.84%) of the total variance could be explained from the scenario facet. The benefit of generalizability theory is that it permits the researcher to hypothetically predict the effect of the various facets on the overall reliability of measurement.^{23–26} These are mathematical predictions and, thus, still need to be tested to confirm the results. However, they give investigators direction for future research study designs. To improve the reliability of our study, the results indicated that at least 4 scenarios should be used to reliably test participants using the TSAI (Table 4).

To truly test if a student is proficient at a technical skill or competency, longer examinations or more scenarios are required.²⁷ It is not good enough to merely test 1 or 2 athletic taping scenarios and expect to reliably predict if students can tape many joints or conditions as accurately as they did with the 1 or 2 taping scenarios on which they were tested in a single, summative examination. Practically, athletic training educators have 2 options: (1) test students on at least 4 taping scenarios in a summative examination to reliably measure their taping skill proficiency or competence and (2) test students throughout the semester in real-life settings using 2 raters and the TSAI with a minimum of 4 scenarios. Athletic therapy and training educators and administrators need to discuss the advantages and disadvantages of summative examination versus embedded examination in clinical rotations and then articulate their conclusions in an overall student-assessment plan.^{22,27,28} Researchers should focus on increasing the number of scenarios tested summatively or in a clinical placement to improve the overall reliability of the measurement.

Limitations

One major limitation of our study was the lack of peerreviewed, published standards or expectations of specific taping techniques. Drawing conclusions about a student's taping skill or performance without well-established, scientifically sound standards is challenging. Raters graded students based on their personal expertise and opinions. In addition, the TSAI has been content validated, but the science behind content validity is weak and tends to be biased to the local environment.⁷ In the content-validity study, a national group of experts from Canada agreed on the items that measured the taping technical skill for a number of body regions.⁷ However, the same consensus discussion revealed differences of opinion among experts as to the direction of ankle heel locks, for example.⁸ Expert consensus on all body region-specific TSAIs was achieved, but that does not mean the standards have clear evidence to demonstrate efficacy for their intended goal: injury prevention. This may also be part of the reason taping efficacy in the ankle has demonstrated mixed results in previous research.^{1,29,30} Therefore, the conclusions of our study need to be contextualized to the underlying purpose: (1) the number of raters needed to reliably measure technical skills using the TSAI and (2) the number of scenarios needed to reliably measure technical skills using the TSAI.

CONCLUSIONS

Athletic taping is a highly valued skill and perhaps one for which athletic therapists and trainers are best known in sport and athletic environments. However, few researchers have established professional standards and, thus, expectations for professors to teach at the preprofessional level. The TSAI was originally developed as a tool to measure athletic taping skills, but it has also served as a device that guided the standards and expectations for teaching taping skills through content validation. Investigators have provided content validation of the standards and expectations,⁷ but more research should be carried out to continue the quest of evidence-based practice and move beyond the lowest level of evidence.¹³ Through generalizability theory and a D study, we proposed the optimal number of raters and scenarios that would be required to reliably measure student performance of taping skills. However, the results need to be contextualized based on the TSAI's having been content validated by expert opinion. Testing students based on taping standards that have high levels of evidence associated with their efficacy should be a goal with researchers. Our study should be considered a starting point for determining the validity and reliability of testing taping skills in preprofessional students.

 Table 4.
 Generalizability and Decision Study Results for Various

 Combinations of Taping Scenarios and Raters

Taping Scenarios, n	Raters, n	$E\rho^2{}_\delta$
2	1	0.56
2	2	0.67 ^a
3	2	0.69
4	2	0.71
6	3	0.77

^a Indicates the $E\rho^2_\delta$ calculated for our study. All other $E\rho^2_\delta$ displayed in this table are hypothetical projections or decision study projections.

REFERENCES

- Handoll HH, Rowe BH, Quinn KM, de Bie R. Interventions for preventing ankle ligament injuries. *Cochrane Database Syst Rev.* 2001;(3):CD000018.
- Canadian Athletic Therapists Association Program Accreditation Manual: Self Study Report 5–40. Calgary, AB: Canadian Athletic Therapists Association; 2007. http://www.athletictherapy.org/pdf/ accreditation/5–40.pdf. Accessed July 18, 2013.
- The Certification Process. Canadian Athletic Therapists Association Web site. http://www.athletictherapy.org/en/educational_process. aspx. Accessed July 18, 2013.
- National Athletic Trainers' Association. *Athletic Training Education Competencies*. 5th ed. Dallas, TX: National Athletic Trainers' Association; 2011. http://www.nata.org/education/competencies. Accessed May 5, 2012.
- Board of Certification. 2009 Athletic Trainer Role Delineation Study. Omaha, NE: Board of Certification; 2010. http://kinrec.illinoisstate. edu/downloads/RD-PA6_Full_Version.pdf. Accessed July 23, 2013.
- Board of Certification Web site. http://www.bocatc.org/boc-partners/ nata. Accessed May 5, 2012.
- Butterwick DJ, Paskevich DM, Lagumen NG, Vallevand AL, Lafave MR. Development of content-valid technical skill assessment instruments for athletic taping skills. *J Allied Health*. 2006;35(3): 147–155.
- Lafave MR, Butterwick DJ. Ankle taping prophylaxis: does directionality matter? [abstract]. *Athl Train Sports Health Care*. 2011;3(3):150.
- Lagumen NG, Butterwick DJ, Paskevich DM, Fung TS, Donnon TL. Intra-rater reliability of nine content-validated technical skill assessment instruments (TSAI) for athletic taping skills. *Athl Train Educ J.* 2008;3(3):91–101.
- Lynn MR. Determination and quantification of content validity. *Nurs Res.* 1986;35(6):382–385.
- Hertel J. Research training for clinicians: the crucial link between evidence-based practice and third-party reimbursement. *J Athl Train*. 2005;40(2):69–70.
- Howick J, Chalmers I, Glasziou P, et al. The 2011 Oxford CEBM levels of evidence (introductory document). http://www.cebm.net/ index.aspx?o=5653. Accessed May 4, 2012.
- Kronenfeld M, Stephenson PL, Nail-Chiwetalu B, et al. Review for librarians of evidence-based practice in nursing and the allied health professions in the United States. *J Med Libr Assoc*. 2007;95(4):394– 407.

- Portney L, Watkins M. Foundations of Clinical Research: Applications to Practice. 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall; 2009.
- 15. Streiner D, Norman G. *Health Measurement Scales*. 3rd ed. New York, NY: Oxford University Press; 2003.
- Gormley G. Summative OSCEs in undergraduate medical education. Ulster Med J. 2011;80(3):127–132.
- Reznick RK. Teaching and testing technical skills. Am J Surg. 1993; 165(3):358–361.
- Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg.* 1997;173(3):226–230.
- 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.
- 20. Mavis BE, Henry RC, Ogle KS, Hoppe RB. The emperor's new clothes: the OSCE reassessed. *Acad Med.* 1996;71(5):447–453.
- Schuwirth LW, Van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach*. 2011; 33(6):478–485.
- Norcini JJ, Blank LL, Duffy FD, Forna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med.* 2003;138(6):476–481.
- 23. Brennan RL. Generalizability theory. *Educ Meas Issues Pract*. 1992; 11(4):27–34.
- 24. Brennan RL. A perspective on the history of generalizability theory. *Educ Meas Issues Pract.* 1997;16(4):14–20.
- Brennan RL. Performance assessments from the perspective of generalizability theory. *Appl Psychol Meas*. 2000;24(4):339–353.
- Shavelson RJ, Webb NM, Rowley GL. Generalizability theory. Am Psychol. 1989;44(6):922–932.
- van der Vleuten CP, Schuwirth LW, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol*. 2010;24(6):703–719.
- Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med.* 1995;123(10):795–799.
- Dizon JM, Reyes JJ. A systematic review on the effectiveness of external ankle supports in the prevention of inversion ankle sprains among elite and recreational players. *J Sci Med Sport.* 2010;13(3): 309–317.
- Verhagen EA, Bay K. Optimising ankle sprain prevention: a critical review and practical appraisal of the literature. *Br J Sports Med.* 2010;44(15):1082–1088.

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