

The Script Concordance Test as a Measure of Clinical Reasoning Skills in Geriatric Urinary Incontinence

Jorge G. Ruiz, MD,*†‡ Renuka Tunuguntla, MD,*† Bernard Charlin, MD, PhD,§
Joseph G. Ouslander, MD,†|| Stephen N. Symes, MD,†# Robert Gagnon, MSc,§ Francois Phancao, MD,†‡
and Bernard A. Roos, MD*†‡

OBJECTIVES: To validate the use of a script concordance test (SCT), a tool to assess clinical reasoning in contexts of uncertainty, which are common in clinical geriatrics practice, on geriatric urinary incontinence (UI) to discriminate levels of expertise in this content area.

DESIGN: A reference panel (15 geriatricians) and 12 respondents (10 senior geriatrics fellows and 2 interns) completed an online 100-item SCT test covering major topics in UI. The test was then optimized by discarding items with negative item-total correlation; the remaining 70 questions covered all major topics in UI. The test was then administered to a second group of participants with different levels of experience, mostly from the University of Miami: eight geriatricians, nine junior geriatrics fellows, 53 internal medicine residents, and 26 medical students. Investigators assessed test reliability and construct validity (to discriminate between levels of expertise).

SETTING: Tertiary academic medical center and affiliated medical school.

PARTICIPANTS: Medical students, internal medicine residents, geriatric medicine fellows, and practicing geriatricians.

MEASUREMENTS: Seventy-item SCT.

RESULTS: The Cronbach alpha for the 70-item test was 0.72. Mean scores were 75.3 ± 7.9 for geriatricians ($n = 23$), 69.0 ± 9.3 for senior geriatrics fellows ($n = 10$), 66.4 ± 6.8 for junior geriatrics fellows ($n = 9$), 66.1 ± 5.7 for residents ($n = 53$), and 65.6 ± 8.5 for students ($n = 26$).

From the *Geriatric Research, Education, and Clinical Center, Bruce W. Carter Veterans Affairs Medical Center, Miami, Florida; †Department of Medicine, Miller School of Medicine, University of Miami, Miami, Florida; ‡Stein Gerontological Institute, Miami, Florida; §Centre de Pédagogie Appliquée aux Sciences de la Santé, Faculty of Medicine, University of Montreal, Montreal, Canada; ||Charles E. Schmidt College of Biomedical Science, Florida Atlantic University, Boca Raton, Florida; and #Jackson Memorial Hospital, Miami, Florida.

Address correspondence to Jorge G. Ruiz, Associate Professor of Clinical Medicine, Director, Geriatric Medicine Fellowship Program, University of Miami Miller School of Medicine, GRECC Associate Director for Education/Evaluation, VA Medical Center GRECC (11 GRC), 1201 NW 16 Street, Miami, FL 33125. E-mail: jruiz2@med.miami.edu

DOI: 10.1111/j.1532-5415.2010.03136.x

Using analysis of variance, significant differences were found between the mean scores of the geriatricians and all other participants except senior fellows.

CONCLUSION: The geriatric UI SCT demonstrated moderate reliability and evidence of construct validity, discriminating between experienced and nonexperienced physicians. *J Am Geriatr Soc* 58:2178–2184, 2010.

Key words: script concordance test; clinical reasoning; urinary incontinence; medical education

Clinical reasoning depends on a group of cognitive skills whose mastery is essential for successful patient evaluation and management. Clinical reasoning and clinical skills are the major components of clinical competence.¹ Evidence indicates a linkage between poor clinical reasoning skills and diagnostic errors.² Nevertheless, clinical reasoning is an important competency domain that educators do not consistently assess.^{3,4} There is no single criterion-standard tool that is feasible, valid, and reliable for measuring clinical reasoning skills in medical trainees.^{4,5}

The script concordance test (SCT) is emerging as a reliable and valid tool to measure elements of clinical reasoning. A “script” is a goal-directed knowledge structure adapted by the individual to perform tasks efficiently.⁶ Thus, when a student encounters real situations with patients, illness scripts begin to appear. Over the course of the learner’s training and experiences, his or her scripts are developed and refined.⁶ These personalized scripts are used actively to process information and verify or discard hypotheses or discern differences between management alternatives.⁷ The SCT is based on the assumption that the development of clinical reasoning occurs through a series of qualitative judgments made by the individual.⁸ The test measures these judgments and compares them with those of a reference panel of experienced clinicians. This comparison provides a method of assessment through

reasoning of nonspecific problems and contexts of uncertainty, both of which are common in any clinical practice and even more common in geriatrics.⁹ By design, the test measures only one step in clinical reasoning—clinical data interpretation—a step early in the process and one that is considered crucial. Inferences can be made from SCT scores about clinical knowledge organization and clinical judgment.¹⁰

Geriatricians are experts at managing complex older patients with multiple, concurrent medical conditions and syndromes.¹¹ Widely used multiple-choice question examinations, by circumscribing assessment to clinically relevant knowledge and cueing examinees,^{4,12,13} probably cannot capture the complexity of geriatrics decision-making. Geriatrics educators need tools that assess a physician's ability to address ill-defined clinical contexts, in which the ability to interpret data relates to the physician's clinical experience, but sampling a broad and complex domain such as geriatrics creates its own challenges. One geriatrics SCT did not discriminate between medical students and practicing geriatricians.¹⁴ Possible explanations for this failure—in addition to the challenges associated with sampling the large geriatrics content domain—were the recruiting technique for experts, the predominance of family medicine practitioners, and the inclusion of test items deliberately built to ascertain differences in reaction time rather than different levels of experience.¹⁴ Given the broad scope of geriatrics practice, sampling more-limited areas of geriatrics content, such as urinary incontinence (UI), may help clarify the role of the SCT in measuring clinical reasoning skills. Although no studies have used the SCT to assess clinical reasoning in a highly specialized content area, UI is a good candidate for a geriatrics SCT. UI is a common and potentially disabling condition in the older population and a risk factor for nursing home placement and the resulting economic implications.^{15,16} It affects all four domains of a person's life (physical, psychological, functional, and social), with devastating effect on self-image and quality of life.¹⁷ Despite its high prevalence, many geriatric patients suffering from UI are not diagnosed or treated.¹⁵ The objective of the current study was to validate the use of the scores of a geriatric UI SCT to discriminate between levels of provider expertise. It was predicted that geriatricians would achieve the highest scores, with senior and junior fellows, internal medicine residents, and medical students achieving progressively lower scores.

METHODS

UI SCT Test Development

Three geriatricians (JGO, a national expert in geriatric UI at Florida Atlantic University, and two geriatricians, JGR and FP, from the Division of Gerontology and Geriatric Medicine, University of Miami Miller School of Medicine) developed an initial pool of 32 cases (155 questions). The experts developed a blueprint outline covering the major clinical topics in UI, with questions addressing diagnostic, intervention, or treatment options in geriatric UI. The tested domains were acute incontinence, urge incontinence and overactive bladder, stress incontinence, functional incontinence, overflow incontinence, and age-related changes including detrusor hyperactivity with impaired contractility,

urethral sphincter weakness, pelvic-floor muscle weakness, cystocele, atrophic vaginitis, and benign prostatic hyperplasia. An academic geriatrician with expertise in UI and a urologist reviewed all the cases to assess whether each question actually addressed a realistic geriatric UI dilemma as well as elements of clinical reasoning skills. After this expert review, the 25 cases (100 questions) remaining were still considered representative of the major clinical topics in UI (Figure 1). The final SCT contained the same blueprint domains, and the number of items and the proportions of questions for each domain were similar. The UI SCT test is available from the authors upon request. The investigators uploaded the questions for the UI SCT test to a secure SCT Web site. Participants received logins and passwords and could log in at their convenience to take the test.

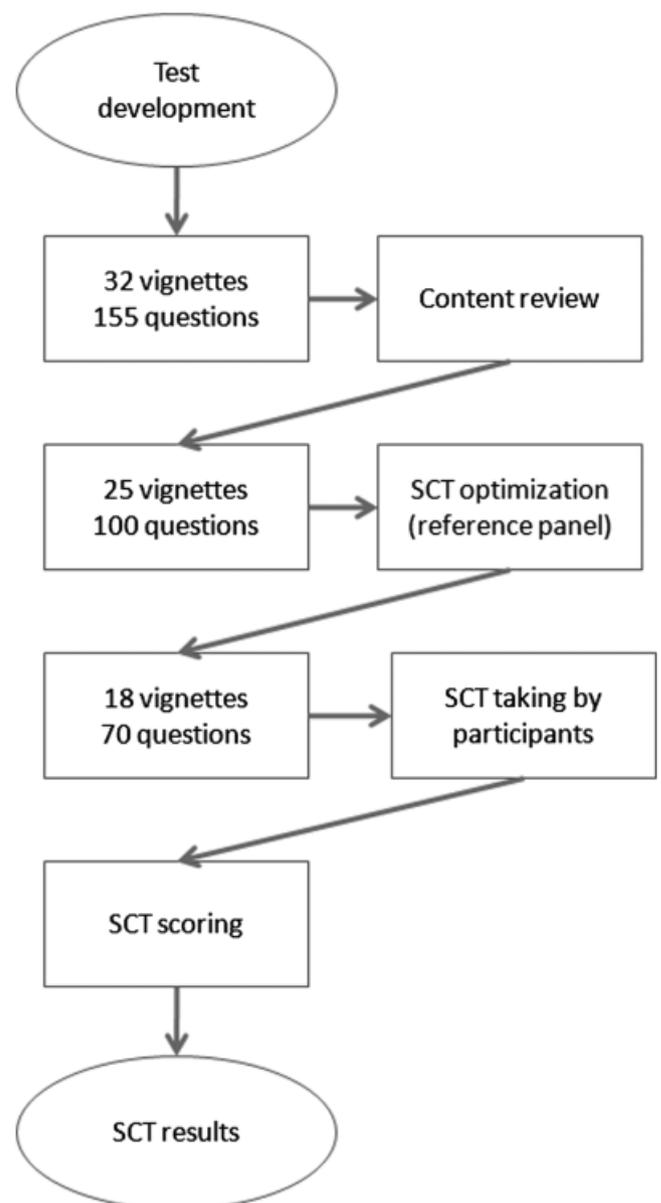


Figure 1. UI script concordance test (SCT) development flow-chart. The figure depicts the process for development of the SCT questions.

Reference Panel

As documented previously, any expert panel consisting of 15 or more members provides acceptable reliability estimates.¹⁸ Fifteen board-certified geriatricians constituted the reference panel for the current study (10 academic faculty geriatricians and 5 geriatricians practicing in the community who were former geriatric medicine fellows). Nine of the 15 geriatricians completed a 10-item pretest questionnaire on demographic data, including board certification, years in practice, type of practice, and percentage of elderly patients with UI, and seven questions asking them to grade their confidence level in managing different types of UI on a Likert scale (ranging from 1 being not confident at all to 10 being extremely confident). All panel members were board certified in geriatric medicine (4 internal medicine, 2 family medicine; 3 did not state their primary specialty). All members except one had been practicing geriatric medicine for more than 3 years, one was practicing only outpatient geriatric medicine, and one was a hospital-based geriatrician; the remaining panel members practiced outpatient and inpatient geriatric medicine. The mean confidence level of the composite score (mean of confidence level for managing different types of UI in the 7 questions) for these physicians was 7.5 (Likert scale ranging from 1 to 10). Six panel members reported seeing elderly patients with UI frequently or very frequently, and three reported seeing them occasionally. Physicians on the reference panel took the test online individually. Members of the reference panel had the option of leaving individual questions blank if they did not feel confident in answering. The analysis was based on the number of panel members answering the specific question.

Test Optimization

Initially, 15 practicing board-certified geriatricians (the reference panel) and 12 trainees took the test. Test optimization was then performed at the case level; cases with negative total correlation (7 cases) were discarded. The remaining 18 cases (70 questions) constituted the final version of the SCT. A score for each case was calculated using the mean SCT scores of the embedded questions (or items). Total score on the test is the summation of all 25 case scores. The correlation between each case score and the total score was used for optimization. To prevent the spuriousness of such a correlation, the case score was excluded from calculation of the total score for any given case with total correlation. Any case with a negative correlation with the total score was excluded from the set, starting with the case with the lowest correlation. This process was repeated in an iterative manner until case/total correlation under 0.10 was obtained. This process ensured homogeneity of the content of the final scale.

Scoring

The scoring key was created from the answers that individual panel members gave, following standard SCT methodology.⁸ The main idea behind the SCT was to compare the performance of students, residents, junior geriatrics fellows, and senior geriatrics fellows with that of a group (panel) of board-certified geriatricians practicing geriatrics in diverse clinical settings. The geriatricians were representatives of the field of the aspiring trainees in the study and

hence constituted the study's reference panel. This scoring method takes into account the variability demonstrated by experienced physicians (constituting the reference panel) in their reasoning process. The maximum score for any item was 1 for the modal answer, and the minimum score was 0. Partial credit was given to other panel member choices, proportional to the number of members who chose that answer on the Likert scale. Answers not chosen by any panel members received 0. For example, in the sample case described below (Figure 2), assuming a panel of 15 members, none chose a +2, +1, or -2 rating, 10 chose a 0 rating, and five chose a -1 rating. The modal answer in this example was 0. Participants choosing this rating would therefore receive 1, whereas those choosing -1 would receive 0.5 (5/10), and participants choosing a +2, +1, or -2 rating would not receive any points. The total score for the test was the sum of credits obtained on each item.

Subjects and Setting

Participants

The participant pool consisted of five groups with varying levels of clinical experience. The first group consisted of 26 medical students (3rd or 4th year) from the University of Miami Miller School of Medicine who were completing a geriatric medicine clerkship rotation. Medical students have a 1-hour session on voiding disturbances in Year 2 (Renal Module) and a 2-hour session devoted to UI during their obstetrics and gynecology clerkship. The second group consisted of 53 internal medicine residents (Postgraduate Years 1–3) from the Jackson Memorial Hospital/University of Miami training program, all of whom had

IF YOU WERE THINKING OF:	AND THEN YOU FIND:	THIS HYPOTHESIS BECOMES:
Benign prostatic hypertrophy and overactive bladder	A normal prostate in rectal exam	-2 -1 0 +1 +2
Overflow incontinence	Postvoid residual 200 mL	-2 -1 0 +1 +2
-2	Excluded or almost excluded	
-1	Less probable	
0	Neither less nor more probable	
+1	More probable	
+2	Certain or almost certain	

Figure 2. Sample script concordance test (SCT) question: A 66-year-old man with a history of hypertension comes to your office complaining of nocturia, hesitancy, frequency, and urinary incontinence for the last 30 days. A SCT item includes a case with its related questions. Case scenarios are followed by a series of questions, presented in three parts. The first part (“if you were thinking of”) contains a relevant diagnostic or management option. The second part (“and then you find”) presents a new clinical finding, such as a physical sign, a preexisting condition, an imaging study, or a laboratory test result. The third part (“this hypothesis becomes”) is a 5-point Likert scale that captures examinees' decisions.

completed a 1-month rotation in geriatric medicine during internship. Information was not obtained on the previous exposure of these trainees to UI. A third group consisted of nine junior geriatrics fellows who had completed residency training in internal medicine or family medicine at various U.S. programs and had joined the fellowship in geriatric medicine at the Jackson Memorial Hospital/University of Miami. These latter trainees completed the study during their sixth month of fellowship training. The fourth group consisted of 10 geriatric medicine fellows during their last month of training. Exposure to UI during training consists of a 1-hour didactic session and clinical experiences with geriatric patients in diverse care settings. The fifth group consisted of 23 board-certified geriatricians: the 15 geriatricians initially included in the reference panel and eight additional geriatricians (4 academic geriatricians from the University of Miami Miller School of Medicine and 4 geriatricians practicing in the community). For each panel member, his or her own SCT scores were calculated on the responses of all other panel members, excluding his or her own responses.

Similar to the reference panel, a pretest questionnaire was sent to all participants asking them to grade their confidence level in managing different types of UI on a Likert scale ranging from 1 being not confident at all to 10 being extremely confident. Participants received specific instructions on the format of the SCT. Participation was voluntary, and neither the participants nor the members of the reference panel received any compensation. The institutional review board at the University of Miami approved the study for exempt status. All data were decoded and anonymous.

Statistical Analysis

After the initial phase, item analysis was performed to detect problematic questions. Item–total item correlation was performed by obtaining the total score of the test followed by calculation of the correlation between each item (or question) and the total score. To reduce the spuriousness of this correlation, for any given item–total correlation, the score of the specific item was excluded from the total score. Questions with item–total item correlation less than 0.10 were eliminated. The reliability of the test was estimated using the Cronbach alpha internal consistency coefficient. The Levene test was used to evaluate the homogeneity of the variance between the five groups. The Levene statistic was 3.35 ($P = .08$). One-way analysis of variance (ANOVA) was used to study differences within and between groups' means, and the Tukey honestly significant difference (HSD) was used to discriminate the scores between the five groups. $P < .05$ was considered statistically significant. To avoid bias, when geriatricians were used as members of the panel, scores for each item were computed using a scoring key that excluded that individual's response to the question. When these geriatricians were included in the fifth group of experienced physicians, their scores were computed using the scoring key used for the other participants, similar to the method in other SCT studies.¹⁹ Spearman rank-order correlation was used to identify whether there was a significant correlation between the SCT scores and self-reported mean confidence in managing UI. The analysis was done using

SPSS software, version 14 for Windows (SPSS, Inc., Chicago, IL). Investigators were unaware of the participants' identities during the analysis.

RESULTS

Participants

One hundred twenty-three participants completed the study: 23 geriatricians, 10 senior geriatrics fellows, nine junior geriatrics fellows, 53 internal medicine residents, and 26 medical students. Eighty-six of the 123 completed and returned the pretest. The self-reported mean confidence level of the composite score (mean of confidence level for managing different types of UI in 5 questions) for the five groups (students, residents, junior geriatrics fellows, senior geriatrics fellows, and geriatricians) were 3.6 ± 1.2 , 5.1 ± 1.4 , 5.3 ± 1.4 , 6.5 ± 1.1 , and 7.4 ± 1.1 , respectively (Likert scale ranging from 1 to 10).

Scores

Mean scores were 75.3 ± 7.89 for geriatricians ($n = 23$), 69.0 ± 9.3 for senior geriatrics fellows ($n = 10$), 66.4 ± 6.8 for junior geriatrics fellows ($n = 9$), 66.1 ± 5.7 for residents ($n = 53$), and 65.6 ± 8.5 for students ($n = 26$) (Table 1). Using ANOVA, significant differences were found between the mean scores of the groups ($F = 7.6$, $P < .001$). Post hoc analysis using the Tukey HSD revealed significant differences between the mean scores of geriatricians and all other participants ($P < .05$) except senior fellows. No significant difference was found between junior fellows, residents, and students (Figure 3). The Spearman rho was 0.9, which is statistically significant ($P = .04$) for the correlation between the SCT scores and self-reported mean confidence in managing UI.

Item Analysis

In the initial phase, a 100-item test was administered to the 15 panel members and 12 respondents (10 senior geriatrics fellows and 2 interns). Item analysis was performed and resulted in the elimination of seven cases (30 questions) with item–total item correlation of less than 0.10. After test optimization, the final test consisted of 18 cases (70 questions).

Table 1. Script Concordance Test Scores of Different Groups

Group	n	%	
		Mean Score \pm Standard Deviation	Range
Geriatricians	23	75.3 ± 7.9	60.4–91.8
Senior fellows	10	69.0 ± 9.3	55.2–82.2
Junior fellows	9	66.4 ± 6.7	56.5–76.1
Residents	53	66.1 ± 5.7	53.6–77.5
Students	26	65.6 ± 8.5	47.9–78.1

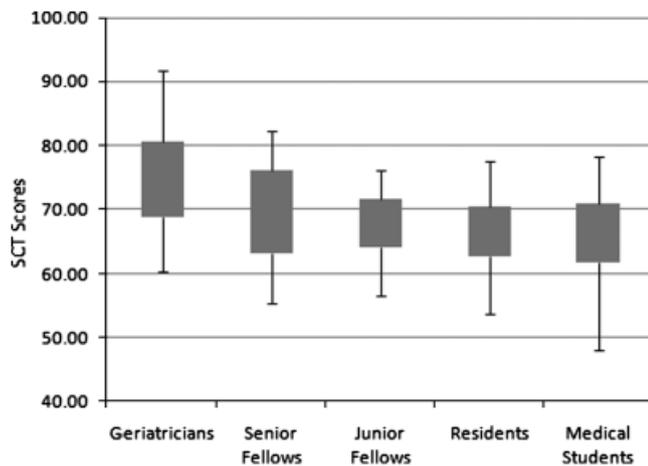


Figure 3. Script concordance test (SCT) scores according to level of experience. The box plot depicts the SCT results according to experience levels.

Reliability

The Cronbach alpha coefficient for the final optimized test was 0.72 (for 123 participants and 18 cases with 70 questions).

Feasibility of the Test

As an online test, the SCT was easy to administer, and participants could take it at their convenience, in one or more sittings. Collecting the data was convenient, because the answers could be exported to a spreadsheet.

DISCUSSION

The geriatric UI SCT demonstrated evidence of construct validity by discriminating experienced physicians such as geriatricians and senior fellows from less-experienced individuals such as junior fellows, residents, and medical students. No significant difference was observed between the test scores of junior fellows, residents, and students. Similarly, no significant differences were observed between experienced geriatricians and senior fellows. Furthermore, the self-reported confidence level was consistent with the SCT scores of the experienced practitioners (geriatricians and senior fellows), who reported higher confidence levels than the less-experienced groups. The SCT was also demonstrated to be moderately reliable.

These findings using an SCT to assess a highly specialized area such as geriatric UI are consistent with the results of other SCT studies conducted in broader content areas. Most of those studies showed that experienced physicians achieved higher mean scores than residents and students, supporting the construct validity of the SCT.^{19–21} Unlike previous studies, the participants' confidence levels in the assessment and management of geriatric UI were also measured, and a strong association was found with performance in the SCT. Some studies in healthcare professional education revealed similar associations between objective assessments of clinical reasoning skills and self-confidence,^{22,23} whereas others did not.^{24,25} More research is needed to shed light on this association.

In the current study, there were no significant differences between medical students, residents, and junior fellows. A plausible explanation is that all the participating medical students in this study had extensive, longitudinal geriatrics experience extending over the 4 years of their training.²⁶ Students learn about UI not only during ward rounds and didactics, but also through e-learning modules and virtual patients to which all the trainees have access.²⁶ Nevertheless, residents and fellows come from various medical schools, where they may or may not have had extensive geriatrics training. All residents during their internship rotated for 1 month in geriatrics, but this experience may not have been enough to make a substantial difference. In contrast to this study, previous studies have reported good discrimination between residents and students, with residents scoring higher.¹⁹

No differences were found in the current study between geriatricians and senior fellows, which is again consistent with the literature.^{19,27} Senior fellows in the last month of fellowship are close to becoming board-certified geriatricians and ready for independent practice in the tested domains. Geriatricians, alternatively, may not be real experts in UI, as previous surveys have indicated.²⁸ Geriatric UI is a highly specialized area, and although geriatricians may in general be more experienced in the broad area of geriatric medicine, they might not be as competent in a specialized area. In addition, their clinical practice may not be as diverse as it was during training.

These results indicate that the UI SCT discriminates between physicians according to level of experience and may represent a useful tool in assessing trainees' clinical reasoning skills, important for the assessment and management of patients.¹⁹ The implication for geriatrics educators is the clear potential of the SCT to measure domains of competence beyond knowledge—clinical reasoning skills that are more relevant to geriatrics practice. The SCT format measures a domain of clinical reasoning and knowledge that educators traditionally evaluate with multiple-choice question examinations.^{4,12,13} The SCT explores the interpretation of data in ill-defined clinical contexts, which are common in geriatrics practice,^{11,29} where ability relates to clinical experience in managing complex older patients. As geriatrics educators move toward more-comprehensive assessments of competencies beyond multiple-choice question examinations, remediation may become an important concern for training programs. Remediation is a complex and challenging process that lacks a consistent and systematic approach.³⁰ A rational remediation program was recently proposed, with four elements where the SCT can be a featured component.³⁰ First, the SCT can be part of a menu of multiple assessment tools for the early identification of learners in need of assistance with clinical reasoning skills. Second, the SCT can spur the development of an individualized remediation plan targeting deficiencies noted in the SCT. Third, the SCT can be a blueprint for deliberate practice, with feedback using clinical cases (paper cases or virtual patients) aiming at building clinical reasoning skills. Finally, the SCT can serve as a cost-effective approach for reassessment of competencies to ensure the learner's achievement of competence and performance.^{14,30} The SCT could also be used to assess practicing clinicians after a continuing medical education session. Participating

physicians who succeeded in passing the SCT (a sufficiently high score when compared with the reference panel) would receive continuing medical education credit. A physician who failed the test could be given a list of specific areas of weakness with access to relevant instructional resources to address gaps in clinical reasoning skills.²⁷

Limitations of this study were the small number of attending geriatricians recruited to serve as one of the five participant groups. As a result, geriatricians from the reference panel had to be included as part of this group of participants. Although this approach was adjusted for, it may not have been possible to eliminate bias. As documented previously, for high-stakes examinations, any number of panel members above 15 provides acceptable reliability estimates,¹³ whereas a smaller panel size may be enough for formative examinations. In the current study, having fewer than 15 panel members completing some test items may be associated with less precision in the assessment of the true reliability of the scores. Another potential limitation is the lack of a time limit for completing the test, which precludes the assessment of test efficiency.

Future studies should focus on psychometric subjects such as standard setting by defining scores at or above which examinees achieve minimum competency in UI. In terms of improving content validity, it may also be helpful to obtain the perspective of teachers and residents to learn why their performance did not differ significantly from that of the medical students. Other studies might objectively measure and compare expert panel members' quality indicators in UI such as Assessing Care of Vulnerable Elderly (ACOVE) with SCT scores.³¹ To confirm the robustness of the findings of interrater reliability, future studies may use panel members in other institutions or other specialized physicians such as practitioners in gynecology and urology, who also manage UI. Test-retest reliability is another area worth study and may include wait-list designs in which one group completes an intervention with related content in UI in which improvement in scores may be expected and there is comparison with a control group with no intervention whose scores should not change after retesting. The former may provide further evidence of the construct validity of the test. Another area to address is the matter of valid reference ranges and reliability scores when panel experts do not complete all the items in the SCT.

CONCLUSION

These results showed a difference in the mean scores on the UI SCT of groups with different clinical experience in this specialized content area. The UI SCT may be useful as formative feedback for trainees and practicing physicians.

ACKNOWLEDGMENTS

The authors thank Maria Van Zuilen, PhD, for reviewing the manuscript and Virginia Roos for editing the manuscript.

Conflict of Interest: The editor in chief has reviewed the conflict of interest checklist provided by the authors and has determined that the authors have no financial or any other kind of personal conflicts with this paper.

Funding Sources: Bruce W. Carter Veterans Affairs Medical Center Geriatric Research, Education, and Clinical Center and the Stein Gerontological Institute.

Author Contributions: Drs. Ruiz and Tunuguntla: concept and design, acquisition of subjects and data, analysis and interpretation of data, and preparation of the manuscript. Drs Charlin and Roos: concept and design, analysis and interpretation of data, and preparation of the manuscript. Dr. Ouslander: concept and design and preparation of the manuscript. Dr. Symes: acquisition of subjects and data and preparation of the manuscript. Dr. Gagnon: concept and design and analysis and interpretation of data. Dr. Phancao: concept and design.

Sponsor's Role: The funders had no role in the study design, analysis, or writing of this manuscript.

REFERENCES

1. Accreditation Council for Graduate Medical Education. ACGME Outcome Project: Enhancing residency education through outcomes assessment [online]. Available at <http://www.acgme.org/outcome> Accessed January 2010.
2. Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Acad Med* 1999;74(10 Suppl):S138–S143.
3. Barrows HS, Feltovich PJ. The clinical reasoning process. *Med Educ* 1987;21:86–91.
4. Epstein RM. Assessment in medical education. *N Engl J Med* 2007;356:387–396.
5. van der Vleuten CP, Newble DI. How can we test clinical reasoning? *Lancet* 1995;345:1032–1034.
6. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: Theory and implication. *Acad Med* 1990;65:611–621.
7. Charlin B, Boshuizen HP, Custers EJ et al. Scripts and clinical reasoning. *Med Educ* 2007;41:1178–1184.
8. Charlin B, Roy L, Brailovsky C et al. The script concordance test: A tool to assess the reflective clinician. *Teach Learn Med* 2000;12:189–195.
9. Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: The script concordance approach. *Eval Health Prof* 2004;27:304–319.
10. Lubarsky S, Chalk C, Kazitani D et al. The script concordance test: A new tool assessing clinical judgement in neurology. *Can J Neurol Sci* 2009;36:326–331.
11. Fried LP, Hall WJ. Editorial: Leading on behalf of an aging society. *J Am Geriatr Soc* 2008;56:1791–1795.
12. Neufeld V. Written examinations. In: Neufeld R, Norman G, editors. *Assessing Clinical Competence*. New York: Springer, 1985, pp 94–118.
13. Norcini J, Meskauskas J, Langdon L et al. An evaluation of a computer simulation in the assessment of physician competence. *Evaluation and the Health Professions* 1986;9:286–304.
14. Gagnon R, Charlin B, Roy L et al. The cognitive validity of the script concordance test: A processing time study. *Teach Learn Med* 2006;18:22–27.
15. Gibbs CF, Johnson TM II, Ouslander JG. Office management of geriatric urinary incontinence. *Am J Med* 2007;120:211–220.
16. Hu TW, Wagner TH, Bentkover JD et al. Costs of urinary incontinence and overactive bladder in the United States: A comparative study. *Urology* 2004;63:461–465.
17. Fung CH, Spencer B, Eslami M et al. Quality indicators for the screening and care of urinary incontinence in vulnerable elders. *J Am Geriatr Soc* 2007;55(Suppl 2):S443–S449.
18. Gagnon R, Charlin B, Coletti M et al. Assessment in the context of uncertainty: How many members are needed on the panel of reference of a script concordance test? *Med Educ* 2005;39:284–291.
19. Lambert C, Gagnon R, Nguyen D et al. The script concordance test in radiation oncology: Validation study of a new tool to assess clinical reasoning. *Radiat Oncol* 2009;4:7.
20. Meterisshian SH. A novel method of assessing clinical reasoning in surgical residents. *Surg Innov* 2006;13:115–119.
21. Carriere B, Gagnon R, Charlin B et al. Assessing clinical reasoning in pediatric emergency medicine: Validity evidence for a script concordance test. *Ann Emerg Med* 2009;53:647–652.
22. Sobral DT. Diagnostic ability of medical students in relation to their learning characteristics and preclinical background. *Med Educ* 1995;29:278–282.
23. Chatterjee S, Ng J, Kwan K et al. Assessing the surgical decision making abilities of novice and proficient urologists. *J Urol* 2009;181:2251–2256.

24. Elzubeir MA, Rizk DE. Assessing confidence and competence of senior medical students in an obstetrics and gynaecology clerkship using an OSCE. *Educ Health (Abingdon)* 2001;14:373–382.
25. Patterson JS. Increased student self-confidence in clinical reasoning skills associated with case-based learning (CBL). *J Vet Med Educ* 2006;33:426–431.
26. van Zuilen MH, Mintzer MJ, Milanez MN et al. A competency-based medical student curriculum targeting key geriatric syndromes. *Gerontol Geriatr Educ* 2008;28:29–45.
27. Sibert L, Darmoni SJ, Dahamna B et al. On line clinical reasoning assessment with Script Concordance test in urology: Results of a French pilot study. *BMC Med Educ* 2006;6:45.
28. Bland DR, Dugan E, Cohen SJ et al. The effects of implementation of the Agency for Health Care Policy and Research urinary incontinence guidelines in primary care practices. *J Am Geriatr Soc* 2003;51:979–984.
29. Steel K, Williams TF. Geriatrics. “Fruition of the clinician.” *Arch Intern Med* 1974;134:1125–1126.
30. Hauer KE, Ciccone A, Henzel TR et al. Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. *Acad Med* 2009;84:1822–1832.
31. Shekelle PG, MacLean CH, Morton SC et al. ACOVE quality indicators. *Ann Intern Med* 2001;135(8 Pt 2):653–667.