

ORIGINAL ARTICLE

Validation of the Script Concordance Test as an instrument to assess clinical reasoning of residents in pediatric emergency medicine in Mexico

Sergio Zavaleta-Hernández,¹ Magdalena Cerón-Rodríguez,² Víctor Olivar-López,² Rubén Espinoza-Montero,³ Antonio Rizzoli-Córdoba⁴

ABSTRACT

Background. Clinical reasoning is a crucial skill to be acquired during a residency training program. In pediatric emergency medicine, physicians are challenged by diagnostic, investigative, and treatment uncertainties. The Script Concordance Test (SCT) uses authentic clinical scenarios to compare trainee's judgement skills with those of experts. The purpose of this study was to evaluate the clinical reasoning using a pediatric emergency medicine SCT and to determine if it would be able to differentiate between different levels of residency training, validating it in Mexico.

Methods. A SCT containing 58 questions nested in 14 cases was administered to pediatric and senior residents at one academic institution. Fourteen experienced emergency medicine pediatricians were part of a reference panel to establish the basis for the scoring process.

Results. Ninety six residents consented to participate. Thirteen residents were fellows of pediatric intensive care medicine or pediatric emergency medicine, and there were 83 pediatric residents. Overall, there was a significant difference in performance across levels of training. The difference between all levels was significant.

Conclusions. This pediatric emergency medicine Script Concordance Test was useful to assess the progression of clinical reasoning during residency training in Mexico.

Key words: Script Concordance Test, clinical reasoning, pediatric emergency medicine, pediatric residents.

INTRODUCTION

Clinical reasoning is a crucial skill for all future physicians to acquire during their training. In pediatric emergency medicine, physicians and trainees are challenged by diagnostic, investigative, and treatment uncertainties. It has been noted that the clinical supervision of residents often takes place without direct observation of the trainee's history taking and examination skills and, after a brief

period of reflection, the trainee reports the findings to an attending physician. From these brief reporting encounters, the attending physician judges the clinical competence of each trainee and reports the judgements on clinical rating forms¹ based on the observation of the residents in the clinical setting over the length of a clinical rotation.² In the United States³ and Canada as in Mexico, these forms often represent the sole means to assess clinical reasoning. Although clinical rates are easy to apply, these result in many ratings based on subjective impressions, which do not discriminate well among trainees.⁴

The Script Concordance Test (SCT) presents another option. The assessment is based on the script theory of medical decision-making.⁵ The basic concept behind this theory has to do with knowledge organization and posits that medical expertise in decision-making is related to the progressive development of organized networks of knowledge called "scripts." These unique, individualized networks of experience and medical knowledge expand and deepen over time, eventually reaching the point that they can be activated with little effort by experts.^{1,6}

¹ Pediatric Department,
² Emergency Department, ³Neurology Department,
⁴ Research Division, Hospital Infantil de México Federico Gómez, Mexico, D.F., Mexico

Corresponding author: Antonio Rizzoli-Cordoba, MD
Research Division, Hospital Infantil de México Federico Gómez
Dr. Marquez 162, Col. Doctores
06720 Mexico, D.F., Mexico
Tel: +52-55-52289917 ext. 2365
E-mail:antoniorizzoli@hotmail.com

Received for publication: 11-07-11
Accepted for publication: 10-08-11

The SCT has been studied in various disciplines in the United States and Canada looking at performance of both residents and medical students. There is a clear improvement in performance as a student gains clinical experience.⁷

The purpose of this study was to evaluate the clinical reasoning using a pediatric emergency medicine SCT and to determine whether the SCT would be able to differentiate between different levels of residency training, providing evidence of the validity of the SCT in Mexico as a Spanish language country.

METHODS

Study Design

We carried out a cross-sectional observational study to compare the performance according to levels of residency training using the pediatric emergency medicine SCT.

Study Setting and Population

All pediatric, pediatric emergency and pediatric critical care residents from the Hospital Infantil de México Federico Gómez were recruited to voluntarily complete a pediatric emergency medicine SCT between March and May 2010. The expert panel was comprised of 14 attending pediatric emergency or critical care physicians who were at least 3 years postcertification and who regularly managed pediatric emergencies. All participants contributed their responses anonymously; only information regarding their level of training and gender was collected.

Study Protocol

The SCT was developed in Spanish by two of the authors who are pediatricians working at the emergency department of the hospital and who have significant clinical experience in the field (MC-R and VO-L). They developed 58 questions nested in 14 cases in the SCT format (without any external financial support).⁸ The questions were categorized as *diagnostic, investigational, and therapeutic*. Lickert-scale anchors were adapted from previous published papers on the SCT.⁹ The topics for the questions were mapped to the clinical content from the University Program for Medical Specialties (PUEM) from the National University of Mexico (UNAM). The development of test items started with commonly encountered clinical scenarios in pediatric

emergency medicine and continued with a determination of data that would seek to make decisions in that situation (Table 1).

Participants responded to each item using a 5-point Lickert scale (-2, -1, 0, +1, +2) to indicate the effect of the new information on the clinical decision confronting them. The scoring matrix was derived from the expert panel. The value for each answer was calculated by dividing the number of experts giving an answer by the total number of experts. The sum of the highest values for each question was considered as the value of 100% concordance. Each answer given by a participant was assigned to the value corresponding to the answer in the scoring matrix and transformed in a percentage of concordance.

The proposed hypothesis to evaluate the validity of the SCT in Mexican pediatric residents was that the SCT would be able to detect progressive differences in clinical reasoning according to the academic level of the residents.

Statistical Analysis

All data analyses were performed using SPSS v.16. Descriptive statistics and ANOVA with Scheffe post hoc analyses were applied to compare SCT scores of the four groups. Two-sample unpaired t-test was used to analyze the differences between specific groups. All *p* values were considered significant at $\alpha \leq 0.05$.

RESULTS

Ninety six residents consented to participate: 55 females (57.3%) and 41 males (42.7%). Thirteen residents (13.5%) were fellows of pediatric intensive care medicine or pediatric emergency medicine (senior residents) and there were 83 pediatric residents (86.5%). Thirty pediatric residents (31.3%) were in their first year of training (PRY-1); 30 (31.3%) in the second year (PRY-2), and 23 (23.9%) in the third and last year (PRY-3).

Global SCT Score

Overall there was a significant difference in performance across levels of training ($F = 39.48$; $df = 3, 92$; $p < 0.001$) (Figure 1). The difference between all categories was significant (Table 1). Senior residents were stronger than PRY-3 with a mean difference of 4.61 (95% CI: 1.45-7.75); and PRY-3 were stronger than PRY-1 with a mean difference of 10.03 (95% CI: 7.28-12.78).

Table 1. Example of one pediatric emergency Script Concordance Test with four questions*

If you are thinking of: ^a	And you find:	Your hypothesis or planned management becomes:				
Bronchiolitis	Inspiratory wheezing	-2	-1	0	+1	+2
Pneumonia	Report in the pulse oxymetry of 92%	-2	-1	0	+1	+2
Endotracheal intubation	Silverman-Andersen score of 3	-2	-1	0	+1	+2
Request a blood cell count	Hyperemic pharynx	-2	-1	0	+1	+2

*Case 1. An 18-month-old female patient who arrived for consultation due to fever of 37.5°C, cough and nasal flare.

-2 = ruled out or almost ruled out; -1 = less probable, 0 = neither more or less probable;

+1 = more probable; +2 = certain or almost certain.

^aThe first two questions correspond to *diagnostic* questions, the third to a *therapeutic* question, and the last to an *investigational* question.

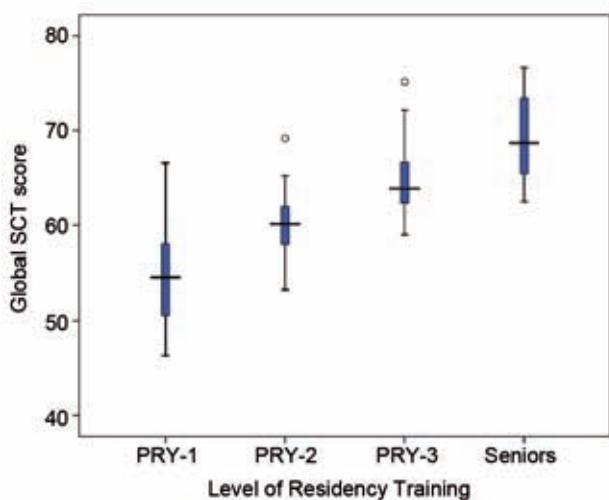


Figure 1. Boxplot of the global SCT scores according to levels of residency training.

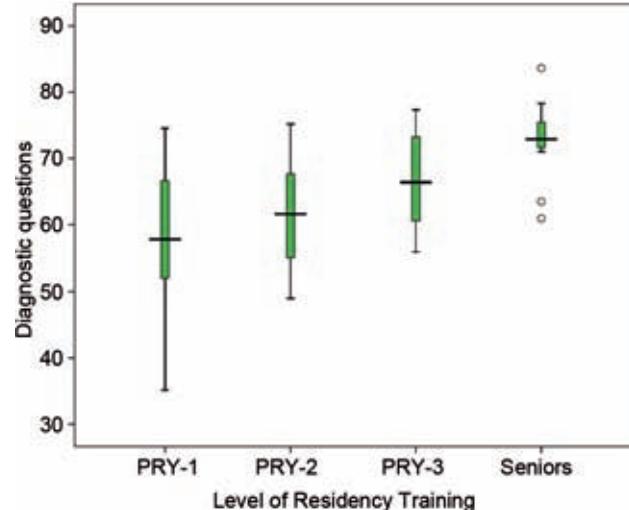


Figure 2. Boxplot of the *diagnostic* SCT scores according to levels of residency training.

SCT Score According to Type of Questions

In *diagnostic* questions, there was a significant difference in performance across levels of training ($F = 12.86$; $df = 3, 92$; $p < 0.001$). Senior residents were stronger than PRY-3 with a mean difference of 6.24 (95% CI: 1.52-10.96); and PRY-3 were stronger than PRY-1 with a mean difference of 9.23 (95% CI: 4.27-14.19). The difference between PRY-1 and second-year pediatric residents (PRY-2) was not statistically significant (Figure 2).

In *investigational* questions, there was a significant difference in performance across levels of training ($F = 15.64$; $df = 3, 92$; $p < 0.001$) (Figure 3). Senior residents were stronger than PRY-3 with a mean difference of 6.21 (95% CI: 0.683-11.74) and third-year residents were stronger than PRY-1 with a mean difference of 8.32 (95% CI: 4.42-12.22) (Figure 3).

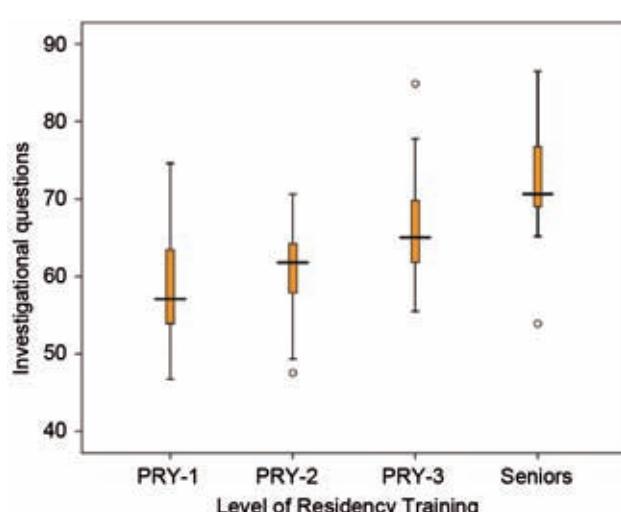


Figure 3. Boxplot of the *investigational* SCT scores according to levels of residency training.

In *therapeutic* questions, there was a significant difference in performance across levels of training ($F = 28.30$; $df = 3, 92$; $p < 0.001$). The difference between senior residents and PRY-3 was not statistically significant. PRY-3 were stronger than PRY-1 with a mean difference of 12.93 (95% CI: 9.67-16.2) (Figure 4). Every statistical result is condensed in Table 2.

DISCUSSION

Many existing assessment tools such as multiple choice questionnaires and short- or long-answer written exams

are often limited to probing pure factual knowledge. The SCT seeks to provide a practical, objective method for evaluating clinical judgment, a critical higher level competency that is currently assessed subjectively and rather informally in most training programs.²

The vast majority of the studies with the SCT have been done in English- or French (Canada)-speaking countries. We only found one article in Latin America evaluating the SCT in geriatric medicine in Brazil¹⁰ in Portuguese language in which the SCT was a useful tool in a Brazilian teaching institution. To our knowledge, this is the first study of SCT in the Spanish language and in Mexico.

We found that the performance of the residents significantly improved overall across levels of clinical experience, supporting the validity of the use of SCT in residency training programs. In the analysis according to the type of question, neither *diagnostic* nor *investigational* question scores had differences between PRY-1 and PRY-2 but were higher in PRY-3 and senior residents. This could be explained as the fact that more time is needed (2 years) to be able to think in those areas in a manner that is more similar to the experts. In *therapeutic* questions, we did not find a significant difference between PRY-3 and senior residents, although there were differences between PRY-1, PRY-2 and PRY-3, due to the fact that treatment of emergencies is the first learning priority in the residency program.

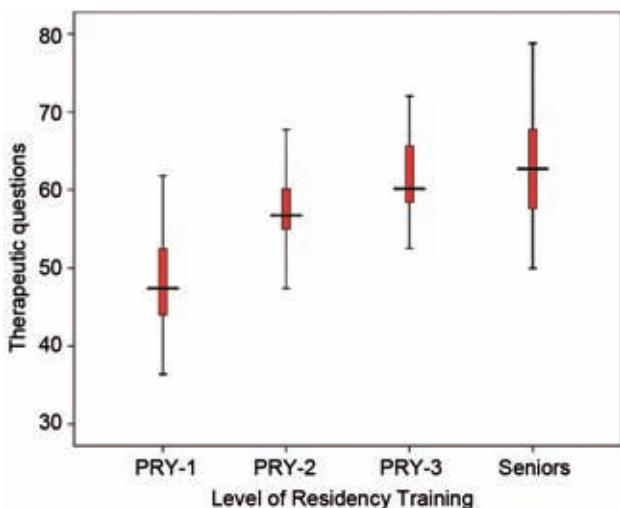


Figure 4. Boxplot of the *therapeutic* SCT scores according to levels of residency training.

Limitations

This study used a nonrandom group of residents as a convenience sample. Although this method is less ideal

Table 2. Mean SCT scores according to type of questions and according to levels of residency training

Level	n	Global SCT score ^a		Type of questions							
		Mean	SD	Diagnostic ^b	Mean	SD	Investigational ^c	Mean	SD	Therapeutic ^d	Mean
PRY-1	30	55.04	5.45	57.47	10.07	58.11	6.76	48.47	6.15		
PRY-2	30	60.11	3.52	61.99	7.44	60.26	6.66	57.40	4.57		
PRY-3	23	65.07	4.16	66.70	7.12	66.43	7.32	61.41	5.46		
Seniors	13	69.68	4.97	72.95	5.84	72.65	8.71	62.08	8.11		

SCT, Script Concordance Test; PRY, pediatric resident years 1, 2 or 3; senior; clinical fellow in pediatric emergency or critical care medicine.

^aMean differences for global SCT score: PRY-1 vs. PRY-2 $p < 0.001$; PRY-2 vs. PRY-3 $p < 0.001$; PRY-3 vs. seniors $p = 0.005$.

^bMean differences for *diagnostic* questions: PRY-1 vs. PRY-2 $p = 0.053$; PRY-2 vs. PRY-3 $p = 0.024$; PRY-3 vs. seniors $p = 0.011$.

^cMean differences for *investigational* questions: PRY-1 vs. PRY-2 $p = 0.150$; PRY-2 vs. PRY-3 $p = 0.003$; PRY-3 vs. seniors $p = 0.030$.

^dMean differences for *therapeutic* questions: PRY-1 vs. PRY-2 $p < 0.001$; PRY-2 vs. PRY-3 $p = 0.005$; PRY-3 vs. seniors $p = 0.768$.

than a stratified random sample, the study setting did not allow for random sampling. It was conducted at a single tertiary care medical center, and the results may not necessarily be generalized to other institutions. Another limitation is the fact that we did not include medical students as baseline comparisons. Because this study did not affect the final summative evaluation of the residents, it is unlikely that contamination occurred whereby residents would divulge the content of the test to other residents.

The results of this study contribute positively to the body of literature on the SCT approach to assess the progression of clinical reasoning during residency training. The test construction was feasible for a residency program and could be used as a formative test for students and residents in Mexico or in Spanish-language countries. Further studies are needed to evaluate the comparisons between the usual grading systems and the SCT, similar to the development of SCT in other residency training programs.

Competing interests

The authors declare that there are no competing interests.

REFERENCES

1. Carriere B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a Script Concordance Test. *Ann Emerg Med* 2009;53:647-652.
2. Hudle TS, Heudebert GR. Taking apart the art: the risk of anatomizing clinical competence. *Acad Med* 2007;82:536-541.
3. Bandiera GW, Morrison LJ, Regehr G. Predictive validity of the Global Assessment Form used in a final-year undergraduate rotation in emergency medicine. *Acad Emerg Med* 2002;9:889-895.
4. Gray JD. Global rating scales in residency education. *Acad Med* 1996;71(suppl 1):S55-S63.
5. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med* 1990;65:611-621.
6. Charlin B, Boshuizen HP, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ* 2007;41:1178-1184.
7. Lubarsky S, Chalk C, Kazitani D, Gagnon R, Charlin B. The Script Concordance Test: a new tool assessing clinical judgement in neurology. *Can J Neurol Sci* 2009;36:326-331.
8. Fournier JP, Desmeester A, Charlin B. Script Concordance Tests: guidelines for construction. *BMC Med Inform Decis Mak* 2008;8:e18. doi:10.1186/1472-6947-8-18.
9. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance Test: a tool to assess the reflective clinician. *Teach Learn Med* 2000;12:189-195.
10. Piovezan RD, Custodio O, Cendoroglo MS, Batista NA. Script Concordance Test: an approach to the evaluation of clinical reasoning in uncertain contexts. *Rev Bras Educ Med* 2010;34:5-12.