Scientific Open Access Journals
SOAJ

Journal of Medical Education and Training

Research Article | Open Access

Script Concordance Testing (SCT): Development of a Novel SCT Assessing Clinical Reasoning for Lumbar Spine Management

Brett Windsor, PT, PhD, MPA[1]*, Taz Kicklighter, PhD, ATC[2], Paul R. Geisler, EdD, ATC[3], Malissa Martin, EdD, ATC/L[4], Borko Rodic, DSc, DPT, OCS, COMT, CMPT[5], Michael Lehr, DSc, DPT, OCS, COMT, CMPT[6]

[1]Rasmussen University, School of Health Sciences, 21300 Lancaster Run, Unit 925, Estero, FL 33928, USA
[2]Associate Professor of Athletic Training, Lee University, Cleveland, Tennessee, USA
[3]Associate Dean, Health Sciences, Professor Health Professions Education, College of Natural, Behavioral and Health Sciences, Simmons University
[4]Vice Provost of Academic Affairs, Health Science and Continuous Learning, PhD Health Science Athletic Training, Concentration Track Director, Rocky Mountain University of Health Professions, Provo, Utah, USA
[5]Assistant Professor of Physical Therapy, Messiah University, Mechanicsburg, PA, USA
[6]Associate Professor of Physical Therapy, Messiah University, Mechanicsburg, PA, USA

***Corresponding author: Brett Windsor, PT, PhD, MPA;** Email; Brett.windsor@rasmussen.edu

Received: 23 May 2022; **Accepted:** 24 June 2022; **Published:** 28 June 2022

Abstract

Introduction: There is little consensus in physical therapy (PT) on how to measure the development and progression of diagnostic clinical reasoning skills (DxR). This article investigates the development of a novel Script Concordance Test (SCT) as a possible tool to measure the clinical decision-making of PT clinicians treating patients with lumbar spine conditions.

Methods: A 26 case scenario, 99-item SCT, measuring the clinical decision making of the PT clinicians' evaluation of patients with lumbar spine conditions was developed. An expert reference panel (5 females, 12 males; mean age >50; mean clinical experience > 20 yrs. [+/- 15 years]) comprised of graduates of an accredited manual therapy fellowship program was created. Item analysis resulted in 47 of 99 SCT items being discarded, leaving 50 individual test items. Internal consistency was measured by Cronbach's alpha statistic.

Results: The SCT was found to have strong reliability coefficient levels (r = 0.82) of internal consistency and acceptable face validity. The mean score (81.89%; 95% confidence interval (CI) [80.08-83.69]) was within acceptable levels.

Discussion & Conclusion: The establishment of a valid and reliable SCT offers the potential to differentiate varying levels of clinical expertise. This could inform teaching of DxR and measuring skills progression. Future research could focus on the ability of the lumbar spine SCT to differentiate PT clinicians based on various levels of experience, or outcomes-based expertise.

Keywords: Script Concordance Testing; Clinical Reasoning; Clinical Decision Making; Expert, Diagnostic Reasoning

Background

In the PT literature, Higgs defined clinical reasoning as a process based on clinical data, patient choices, and professional judgment and knowledge [1]. Recent literature has focused on a dual process theory [2,3]. System I (fast-thinking, non-analytical) involves a rapid, unconscious, pattern recognition process involving long-term memory and the recall of previous precise experiences, while System II (slow-thinking, analytical) is conscious and is based on rules, algorithms, and structure [4]. Both systems work together in some way, although the question of whether these systems are binary or continuum based is a controversial one [5]. Experts appear to use both systems, however default to primarily using System I reasoning [6]. Within this continuum, two distinct types of clinical reasoning have emerged. Diagnostic Clinical Reasoning (DxR) focuses on decision-making processes used to achieve a correct diagnosis [7] and is distinct from Therapeutic Clinical Reasoning (TxR), which has received less emphasis in research but emphasizes processes used to implement a plan of care. Although DxR and TxR are related, emerging research is suggesting they are not interchangeable [8].

DxR processes in PT have been found to largely mirror those used in medicine [9] and have been described as central to both the pre-professional and post-professional education pathways [10-13]. Within the field of physical therapy, operational definitions of clinical reasoning have varied significantly among sources and there is no consensus within the profession as to how DxR should be taught and assessed, nor has PT thus far been able to demonstrate how DxR education can lead to an improvement in levels of expertise, or patient outcomes [14].

The Script Concordance Test (SCT) was developed in the late 1990's as research shifted away from the search for a generic problem-solving approach toward a focus on the organization of memory, knowledge use, and mental representations [15,16]. SCT uses 'illness scripts' as central tenets in the development of decision-making skills in clinical practice. Illness scripts are stored in long-term memory and developed by clinicians in response to patient encounters. With each new patient encounter, clinicians subconsciously access these scripts, matching earlier experiences with current circumstances to develop intuitive hypotheses about possible diagnoses and management strategies.

The SCT presents testers with a series of clinical problems, asking them to make clinical decisions after further elements of information are provided. The SCT has been developed to assess both DxR and TxR, but it can be tailored to a specific domain of interest [16]. The scoring system of the SCT measures the gap existing in scripts between examinee and a panel of experts. The format purportedly reflects real clinical decision-making and uses a scoring system allowing for variation in processes, even among experienced clinical experts [17].

The more closely the examinee 'agrees' with the expert panel, the higher the score will be, and vice-versa. The SCT has been validated as a tool for separating groups of participants with varying degrees of clinical expertise in gynecology [18], urology [19], internal medicine [20], optometry [21], radiotherapy [22], neurology [23], and nursing education [24]. Clinicians find SCT's appealing as a learning tool [25], and it also has potential as a self-reflective assessment tool used in conjunction with expert narrative feedback [26]. SCT's are well recognized as a teaching aide in continuing professional development courses [27,28]. The SCT has generally shown a high level of internal consistency between items across a wide variety of medical disciplines [29,30]. Despite being used in other medical and health professional disciplines, SCT have not been widely used in PT. The purposes of this study are to 1) show an expert reference panel and 2) develop a reliable and valid SCT to be used as a possible basis for distinguishing among PT clinicians with varying levels of education, expertise, and experience.

Methods

Study Design

This qualitative descriptive research study design was approved by the Institutional Review Board of Rocky Mountain University of Health Professions (IRB#: HRP-228). There is an established process described in the literature for building SCT's [21,31]. Using this process, a pilot SCT of 33 scenarios and 99 items was developed, focused on the lumbar spine. Second, an expert reference panel (5 females, 12 males; mean age 40; mean clinical experience > 20 +/- 15 yrs.) was assembled to show baseline reference scoring. Finally, the SCT was completed by the expert reference panel to establish a baseline reliability and validity of the instrument.

Subject Recruitment

Inclusion criteria were as follows: 1) currently practicing in musculoskeletal physical therapy and 2) a graduate of an American Board of Physical Therapy Residency and Fellowship Education (ABPTRFE) accredited Orthopedic Manual Physical Therapy (OMPT) fellowship program. An ideal size for an expert reference panel for SCTs has been described as twenty [32]. Thirty-three potential subjects were invited to take part in the study via email.

Twenty-one participants agreed to take part in the study and completed the informed consent. Among 21 participants, three completed the informed consent, but did not complete the SCT. One participant completed both the survey and the 99-item SCT but did not meet the inclusion criteria of completion of an APTA accredited manual therapy fellowship program. Therefore, these four participants were eliminated. Seventeen subjects (12 males and five females) completed the entire survey and were included in this study. A flow of diagram for subject recruitment is represented in Figure 1.

Data Obtaining via a Survey Instrument

Once assembled, all reference panel members were sent email containing a link to a survey (Survey Monkey™). The survey included informed consent, the demographics of participants, instructions for completing the SCT, and the 33 scenario, 99 item SCT. Participants were asked to complete the SCT in a single sitting of 90 minutes. Thirty-three (33) diagnostic scenarios originating from everyday clinical experiences were written by the primary investigator [17]. Each diagnostic scenario presented 1) a plausible diagnosis, 2) new clinical information, and 3) a five-point Likert scale representing 'significantly less-likely, -2' on the far left, through to neutral '0' and 'significantly more likely, +2, on the far right (Table 1) which together comprise a single test item. Each individual test item relates to the first scenario, but not the other two test items. The content of the SCT was based on patients presenting with commonly seen illness scripts of the lumbar spine.

In this scenario, the choice of +1 or +2 were most concordant with the experts' responses. Both responses would have shown a maximum score for this item. It is generally considered that the addition of bilateral paresthesia to a presentation already involving bilateral lower extremity pain would increase the likelihood of a diagnosis of cauda equina syndrome.

Results

Twenty-one participants agreed to take part in the study and completed the informed consent. Twelve males and five females completed the survey (Table 2).

Score Analysis

Upon completion of the SCT, a scoring key was constructed using the aggregate scoring method, a commonly accepted method of scoring [33]. Individual responses were downloaded to an excel

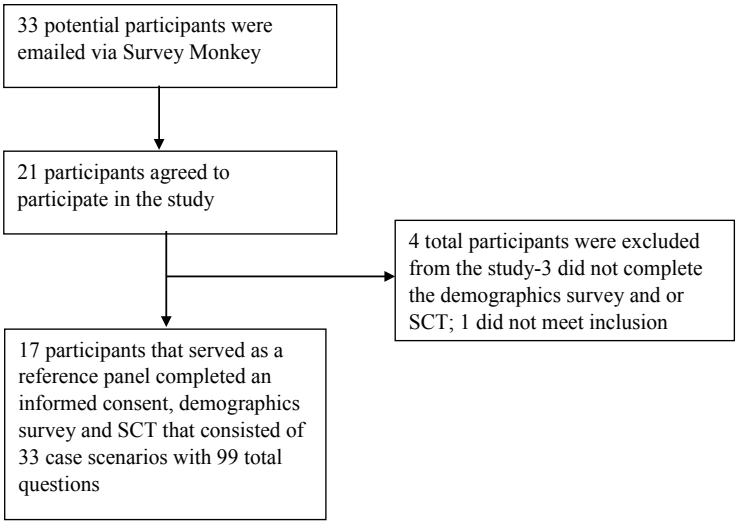

Figure 1: Recruitment of participants.

Table 1: An example SCT scenario and test items

A 36-year-old male is complaining of a 2-week history of acute central low back pain which extends into both buttocks and both posterior thighs.
If you were thinking of the following diagnosis: *Cauda Equina Syndrome*
And the following information became available: *There was paresthesia in both legs*
Then this hypothesis would become: -2 Significantly less likely -1 Slightly less likely 0 No change at all in the likelihood +1 Slightly more likely +2 Significantly more likely

Table 2: Age and Gender distribution of reference panel participants.

Gender	The number of participants
Female	5
Male	12
Age	
30-39	5
40-49	5
50-59	4
60+	3

spreadsheet (Mac, Version 15.31). The survey program assigned a 10-digit numeric identifier unique to each participant to de-identify participants. Each response was converted from the raw Likert format into a single numeric identifier (-2=5, -1=4, 0=3, 1=2, 2=1) and was then scored using the aggregate scoring method [33]. For each item, the maximum possible score was 1.0, fractioned according to the number of participants who chose a response. E.g., if 11/17 participants chose response -1, and 6/17 chose response -2, then the assigned score for any future participants taking the SCT would be either 11/17, = 0.65, or 6/17, = 0.35. In instances where no participant chose -1 or -2, a score of 0 would be assigned.

Scores were analyzed within Excel to decide those items best suited to differentiating performance. Cases with too little response variability (not representative of uncertainty or ambiguity) or too much variability (likely a poorly written item) are considered unlikely to contribute to the reliability and validity of the instrument [33], and were discarded. Three independent expert reviewers (individuals who met the reference panel inclusion criteria, but did not complete the SCT), were instructed to review each item and discard those items with too much or too little variance. Too little variance meant that >13 of the 17 expert panel members chose the same response to a scenario, showing little ambiguity. Too much variance meant that all five responses were chosen by the panel, or majorities were clustered around responses at each extreme. Ultimately, each expert reviewers' decisions were subjective. Final decisions about whether to keep or discard individual test items were made by a majority vote. This process resulted in 32 of the 99 individual SCT items being discarded, leaving 67 individual test items. Finally, Excel data was manually entered into SPSS and a Cronbach's Alpha analysis was performed on the remaining 67 items. Items with item-total correlations of <0.05 were discarded, resulting in the elimination of a further 17 individual test items. This left a 50-item SCT with a reliability co-efficient of 0.82, reflecting strong internal consistency.

SCT Scores

Within SPSS, descriptive statistics were developed for the SCT Scores. The mean score was 81.89% (range 76.16%-86.86%; 95% CI's [80.08%-83.69%]). The data was normally distributed with acceptable skewness in test scores (0.108). SCT developers suggest using a T-transformation of scores based on a distribution of panelists' scores with a mean of 80 and a standard deviation of 5 to help examinees and educators interpret test scores [33]. These results are within those estimates and, therefore, signify a potentially useful assessment tool.

Discussion

PT would benefit from a more standardized approach to assessing clinical reasoning, assisting with curriculum development informing professional growth and differentiate levels of expertise. This study demonstrated a reliable SCT (r=0.82) and acceptable construct validity with the number of case scenarios falling within previously published guidelines [29,30]. The relatively low number of test-items may raise concerns about whether a wide enough sampling of cases exists to draw conclusions about DxR skills. Earlier studies suggest around 100 items (25-30 cases), of which 25% are discarded after item analysis, being capable of reliable scores [31]. This SCT covered a narrow area of practice widely, sampling acute, sub-acute, and chronic cases, with/without leg symptoms, with/without neurological signs, as well as extension and flexion-based diagnoses, and instability. The higher number of discarded items may be due to relative inexperience with SCT item-writing, or a lack of familiarity on the part of the test-takers.

Another concern in this study is whether the reference panel can be considered a representative sample of experts. All panel members had significant patient care experience. Three members graduated from accredited orthopedic clinical residency programs and 11 members achieved orthopedic specialist status. However, those achievements have not been linked to superior patient outcomes [34-36]. All panel members graduated from accredited fellowship programs, which has been correlated with superior patient outcomes [34]. Therefore, it is reasonable to consider that the reference panel in this study is experts in DxR for the lumbar spine.

The SCT scores in this study (81.89%; 95% confidence interval (CI) [80.08-83.69]) appear to be reflective of expertise. Earlier research has suggested a mean score of 80 as capable of discriminating between separate groups of learners [33]. Reported 95% CI's (80.08%-83.69%) are tightly clustered, increasing chances for discrimination due to lower likelihoods of score range crossover between groups.

There are limitations to this study. First, this SCT is narrow in scope. It deals only with DxR and only patients with lumbar spine conditions. Second, this expert reference panel was homogenous in its educational preparation, all being graduates of the same accredited fellowship program. Therefore, it may not be representative of expertise across the wider profession. Finally, developing, implementing, and scoring SCT's is time- and resource-intensive.

Further research could investigate experts from a more diverse sample of educational preparations, which may facilitate the use of this panel as a proxy for the wider PT profession. Studies investigating discriminative abilities between PT clinicians and students of varying experiences would provide information informing curriculum development. Finally, research should investigate creating more scalable methods for administering, scoring, and interpreting SCT's. Together, these studies could create a foundation for the SCT to play a meaningful role in into methods of clinical reasoning assessment, as it does within other healthcare professions.

Conclusion

This study demonstrates a reliable DxR SCT with acceptable construct validity in a narrow area of clinical PT practice. This SCT may be a useful teaching tool capable of identifying differences in the clinical reasoning skills of clinicians and students with varying levels of experience and expertise. Further research should focus on the ability of the SCT to differentiate groups of clinicians based on their level of education, professional achievements, and/or their clinical outcomes.

Lessons for Practice

Lessons learned include the importance of incorporating clinical reasoning into pre and post professional physical therapy education.

Ability to accurately assess clinical reasoning is important to gauge the development of clinical reasoning in clinicians with different levels of experience.

SCT can be a valid and reliable tool for clinical reasoning assessment and can be easily administered to students.

References

1. Higgs J, Burn A, Jones M. Integrating clinical reasoning and evidence-based practice. AACN Clin Issues. 2001; 12:482-490.

2. Norman G. Taking stock. Adv Heal Sci Educ. 2014; 19:465-467.

3. Geisler PR, Lazenby TW. Clinical Reasoning in Athletic Training Education: Modeling Expert Thinking. Athl Train Educ J. 2009; 4:52-65.
4. Croskerry P, Petrie DA, Reilly JB, Tait G. Deciding About Fast and Slow Decisions. Acad Med. 2014; 89:197-200.
5. Norman G, Monteiro S, Sherbino J. Is clinical cognition binary or continuous? Acad Med J Assoc Am Med Coll. 2013.
6. Ericsson A. Development of Professional Expertise: Toward Measurement of Expert Performance. Cambridge University Press. 2009.
7. Croskerry P, Norman G. Overconfidence in Clinical Decision Making. Am J Med. 2008; 121:24-29.
8. McBee E, Ratcliffe T, Picho K, Schuwirth L, Artino Jr. AR, Yepes-Rios AM, et al. Contextual factors and clinical reasoning: Differences in diagnostic and therapeutic reasoning in board certified versus resident physicians. BMC Med Educ. 2017; 17:1-8.
9. Payton OD. Clinical Reasoning Process in Physical Therapy. Phys Ther. 1985; 65:9241-928.
10. Edwards I, Jones M, Carr J, Braunack-Mayer A, Jensen GM. Clinical Reasoning Strategies in Physical Therapy. Phys Ther. 2004; 84:312-330.
11. Jones M. Clinical reasoning: the foundation of clinical practice. Part 1. Aust J Physiother. 1997; 43:167-170.
12. Jones M. Clinical Reasoning & Pain. Man Ther. 1995; 1:17-24.
13. Sweeney A, Doody C. The clinical reasoning of musculoskeletal physiotherapists in relation to the assessment of vertebrobasilar insufficiency: A qualitative study. Man Ther. 2010; 15:394-399.
14. Christensen N, Black L, Furze J, Huhn K, Vendrely A, Wainwright S. Clinical Reasoning: Survey of Teaching Methods, Integration, and Assessment in Entry-Level Physical Therapist Academic Education. Phys Ther. 2017; 97:175-186.
15. Charlin B. Script questionnaires: their use for assessment of diagnostic knowledge in radiology. Med Teach. 1998; 20:567-571.
16. Lubarsky S, Durning S, Charlin B. AM Last Page: The Script Concordance Test: A Tool for Assessing Clinical Data Interpretation Under Conditions of Uncertainty. Acad Med. 2014; 89:1089.
17. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. BMC Med Inform Decis Mak. 2008; 8:18.
18. Park AJ, Barber MD, Bent AE, Dooley YT, Dancz C, Sutkin G, et al. Assessment of intraoperative judgment during gynecologic surgery using the Script Concordance Test. Am J Obstet Gynecol. 2010; 203:240.e1-6.
19. Sibert L, Darmoni SJ, Dahamna B, Hellot M-F, Weber J, Charlin B. On line clinical reasoning assessment with Script Concordance test in urology: results of a French pilot study. BMC Med Educ. 2006; 6:45.
20. Marie I, Sibert L, Roussel F, Hellot M-F, Lechevallier J, Weber J. [The script concordance test: a new evaluation method of both clinical reasoning and skills in internal medicine]. La Rev Med interne. 2005; 26:501-507.
21. Faucher C, Dufour-Guindon MP, Lapointe G, Gagnon R, Charlin B. Assessing clinical reasoning in optometry using the script concordance test. Clin Exp Optom. 2016; 99:280-286.
22. Lambert C, Gagnon R, Nguyen D, Charlin B. The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. Radiat Oncol. 2009; 4:7.
23. 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.
24. Deschênes M-F, Charlin B, Gagnon R, Goudreau J. Use of a script concordance test to assess development of clinical reasoning in nursing students. J Nurs Educ. 2011; 50:381-387.
25. Sibert L. Script Concordance Test and Continuing Education: a marriage that can only be successful. Rev Mal Respir. 2016; 33:329-331.
26. Brownell AKW. The Script Concordance Test. Can J Neurol Sci / Le J Can Des Sci Neurol. 2009; 36:272-273.
27. Hornos E, Pieguezuelos L, Harillo L, Gagnon R, Malagrino P, Charlin. Stability of SCT scores across panels from different countries in a continuing professional development environment. Can Fam Physician. 2013; 59:518-525.
28. Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CPM. Script concordance testing: A review of published validity evidence. Med Educ. 2011; 45:329-338.
29. Gagnon R, Charlin B, Lambert C, Carrière B, Van Der Vleuten C. Script concordance testing: More cases or more questions? Adv Heal Sci Educ. 2009; 14:367-375.
30. Ramaekers S, Kremer W, Pilot A, Beukelen P Van, Keulen H Van. Assessment of competence in clinical reasoning and decision making under uncertainty: the script concordance test method. Assess Eval High Educ. 2010; 35:661-673.
31. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no. 75. Med Teach. 2013; 35:184-193.
32. Gagnon R, Charlin B, Coletti M, Sauvé E, van der Vleuten C. 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.
33. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: Insights from a systematic review. Med Educ. 2012; 46:552-563.
34. Rodeghero J, Wang Y-C, Flynn T, Cleland JA, Wainner RS, Whitman JM. The impact of physical therapy residency or fellowship education on clinical outcomes for patients with musculoskeletal conditions. J Orthop Sports Phys Ther. 2015; 45:86-96.
35. Hart DI, Dobrzykowski EA, Atc2 P. Influence of Orthopaedic Clinical Specialist Certification on Clinical Outcomes. J Orthop Sport Phys Ther. 2000; 30:183-193.
36. Whitman JM, Fritz JM. The Influence of Experience and Specialty Certifications on Clinical Outcomes for Patients with Low Back Pain Treated Within a Standardized Physical Therapy Management Program.J Orthop Sports Phys Ther. 2004; 34:662-672.

Citation: Windsor B, Kicklighter T, Geisler PR, et al. Script Concordance Testing (SCT): Development of a Novel SCT Assessing Clinical Reasoning for Lumbar Spine Management. J Medic Educ Training 2022; 6:077.