

ORIGINAL RESEARCH ARTICLE

Initial validation of a script concordance test to measure the development of clinical reasoning among physical therapy residents

Suzanne R. O'Brien*, Naomi Dillon, Mikkel Linskey, Kevin Lagueras, Julie Uhl, Sarah Conroy, Alex Kiley, Nicole Lanzafame, Erin States and Emma Sturm

Nazareth College, School of Health and Human Services, Department of Physical Therapy

Abstract

Rationale: Clinical reasoning (CR) skills of physical therapy (PT) residents may be limited by a lack of clinical experience, and the complexity of neurologic and elderly patients requires individualized approaches for prescribing walking assistive devices (ADs). Script concordance tests (SCTs), which are case-based assessments that present information that the test taker must examine before answering, can measure the growth in CR. The purpose was to validate a SCT designed to measure CR growth in residents for walking AD prescription.

Methods: Using a prospective cohort design, a panel of 20 experienced practicing clinicians reviewed the test content and provided the correct answers. Neurologic and geriatric residencies were identified from the American Board of Physical Therapy Residencies and Fellowships database. Residency directors were informed of the study, and 27 of them provided access to their residents. Residents gave assent in month 1 and were followed through month 6. The SCT and instructions were emailed at both time points. In total, 27 month-1 tests and 20 month-6 tests were returned. Fleiss kappa was used to evaluate the consistency of the clinician panel's and the residents' responses at both time periods. Mean scores for the panel and the residents at months 1 and 6 were calculated. Testing for differences between the clinician panel and the residents at months 1 and 6 was assessed using a Mann-Whitney U test, and testing for differences between the residents at the same time points was assessed using a Wilcoxon signed rank test.

Results: Demographics of the clinical panel and residents are presented. Internal consistency was moderate, and chance agreement was fair in both groups. Mean test scores were 34.8 (4.3), 34.5 (4.1), and 36.3 (3.9) points for the panel, and residents at months 1 and 6, respectively. Concordance of choices on the SCT between the clinician panel and residents was similar at month 1 and grew more concordant by month 6.

Conclusion: The SCT did not differentiate between the practicing clinicians' and residents' CR skills to prescribe ADs. Further research comparing the CR of PT residents to PT clinicians in early practice should be undertaken since the residents' responses were consistent with the practicing clinicians. The unique test construction or the clinician panel, whose members were not required to be board certified, may have introduced error.

Keywords: *MeSH*; clinical reasoning; self-help devices; internship and residency; author; ambulation; physical therapy

Received: 6 January 2022; Revised: 29 December 2022; Accepted: 15 January 2023; Published: 28 February 2023

The academic physical therapy (PT) community is interested in the development of clinical reasoning (CR), which is a skill that develops over time.¹⁻³ PT CR has been defined as '... integrating cognitive, psychomotor, and affective skills. It is contextual in nature and involves both the therapist and client perspectives. It is adaptive, interactive, and collaborative with the intended

outcome being a biopsychosocial approach to patient/client management'.⁴ There is consensus that CR is an essential skill for practicing PTs and that CR should be covered in all phases of clinical training. For instance, the American Council of Academic Physical Therapy (ACAPT) identified CR as essential beginning in the first year of professional training,⁵ and the American Board of

Physical Therapy Residency and Fellowship Education⁶ has identified CR as one of the seven Domains of Competence for post-professional training.

CR is difficult to measure since multiple interacting factors, thought processes, and preferences within the context of the patient's characteristics and environments must be considered during clinical situations. Healthcare practice requires clinicians to make decisions under uncertainty because each patient's situation is unique and needs to be taken into account while making decisions, and therefore, each must be individualized.^{7,8} The integration of the knowledge and judgment needed to become an effective clinician has drawn great interest to the development of CR skill during entry-level and post-professional PT training. Consequently, CR is essential to the physical therapist's ability to deliver efficient, effective, and individualized care.⁴

There is consensus that one developmental cornerstone of emerging CR skill is the presence of illness scripts (ISs). ISs are rooted in cognitive psychology and explain how information or data about a patient presentation is interpreted.^{9,10} As a result, ISs have been described as stored knowledge of clinical presentations used by clinicians when interacting with patients and, therefore, are derived from continual clinical practice that exposes learners to a variety of patients' presentations. The presence of ISs is considered a higher-order CR skill because knowledge is organized and retrieved rapidly when a pattern of illness or injury is recognized.¹¹ Because an IS contains more information than just case presentations and management strategies, skilled clinicians integrate diagnostic presentations and patient management approaches.^{11,12} When using an IS, the clinician has encapsulated information beyond the associated signs and symptoms to include other relevant information, such as the consequences of the diagnosis and the context in which it develops.¹² Early-career clinicians should also possess such knowledge; however, due to the use of the hypothetico-deductive method of reasoning, novice knowledge has been characterized as less efficient and error-prone.¹¹ Early learners lack the amount of exposure to a continuum of common presentations of a particular diagnosis and often take time to propose and test several options before drawing conclusions. Additionally, early learners may lack an individualized approach^{2,12} meaning that adapting plans to unique patients' attributes may be underdeveloped. Once an IS has been developed, CR is abridged, becoming faster, more efficient, and adaptable.^{11,13}

Script concordance tests (SCTs) emerged from script theory as a means to measure the growth of CR in learners.¹³⁻¹⁶ SCTs were developed to address the relative presence of 'illness scripts (ISs)' by a student or an early-career clinician.^{12,17} Since the presence of ISs in the clinical repertoire is considered a higher-order way of organizing and retrieving knowledge, it is an indication that the clinician

has advanced beyond the more basic methods of reasoning.¹¹ SCTs measure CR by utilizing clinical scenarios, called vignettes, which present commonly seen clinical presentations in routine clinical practice.¹⁸ The vignettes present uncertain and ambiguous information since clinicians often operate with incomplete information.^{3,8-10,13-16}

Several investigations have validated SCTs to differentiate the relative skill of learners. SCTs have differentiated students in geriatric medicine from clinicians,¹⁴ students from general practitioners,³ and residents in an emergency department from expert practicing clinicians.¹⁹ Traditionally, 'experts' are practicing clinicians in the subject area covered by the SCT with entry-level credentials expected for practice within the profession being investigated.^{9-17,20} To avoid confusion, we will refer to the practitioners used to validate the SCT for this study as 'practicing clinicians'. Differentiating residents from practicing clinicians in the same practice setting could serve to direct the development of CR skills and give insight into the continued growth in CR beyond the entry level. PT residents should develop their CR skills over time while there is exposure to the CR modeled by instructors and from interactions with patients' presentations and management. Promoting the additional growth of CR to approximate that of clinicians in practice is one of the core competencies of PT clinical residency.⁶

Possessing excellent CR is important for safe and effective assistive device (AD) prescription. According to a census estimate, 6.8% of Americans report an ambulatory disability,²¹ and ADs are commonly used in neurologic and geriatric populations. Ambulation deficits may increase the need for long-term care²² and lead to decreased participation in vocational and social situations.²³ The SCT used in this study incorporated the complexity of the interacting factors previously identified for walking AD prescription in stroke and brain injury.^{7,8} It is unknown how well-developed early clinicians' CR skills are for walking AD prescription in neurologic and geriatric clinical practice. Most PT residents are in the early stages of their career and may lack sufficient CR skill to make efficient and accurate decisions about walking ADs. This study's purpose was to investigate the reliability and validity of a SCT designed to measure whether improvement in residents' CR skill for walking AD prescription approached that of practicing clinicians during the first 6 months of an accredited neurologic and geriatric clinical residency.

Methods

Study setting and population

Between May and October 2019, participants were recruited from accredited neurologic and geriatric residency programs using the American Board of Physical

Therapy Residencies and Fellowships (ABPTRFE) database. At the time of data collection, there were 68 neurologic and geriatric residency programs. Programs were excluded if the residency did not begin within our time frame ($n = 6$) or was not offering a residency during 2019 ($n = 3$). The directors of the remaining 59 residencies were contacted via email and phone; of these, 27 programs consented to provide their residents' contact information. A total of 52 residents were contacted through email to obtain their assent, and if obtained, the test was emailed with identical instructions for test completion that was provided to the panel (Appendix). During the first month of residency, a weekly email reminder was sent to encourage test return, and 27 residents did so, representing a 51.9% response rate. Following the month 1 test administration, no residents received feedback or results. There were four residencies with two residents in the study who were instructed not to discuss the SCT while the study was being conducted. At the beginning of each resident's sixth month, the test and the instructions were emailed, and weekly email reminders continued until the test was returned or the month ended. A total of 20 month-6 tests were returned, yielding a 74% response rate. A \$5.00 gift card was offered to residents who completed both month-1 and month-6 tests.

Test development

The development of a SCT occurs in two stages. After being written (stage 1), the second stage is to convene a panel of practicing clinicians to take the test to provide content validity and the correct answers. The panel should contain between 15 and 20 practicing clinicians with experience in the SCT subject matter.^{9-10,16} The relative alignment (concordance) of answers between the panel and learners indicates how well the learners have used higher-order reasoning skills.^{9-10,16}

A SCT was developed, and it contained 54 items nested within 18 vignettes to measure the CR of walking AD prescription in neurologic and geriatric PT residents. Five authors wrote test items (SRO, ND, ML, KL, JU) with content informed by two recent descriptions of PT practice for walking AD prescription in inpatient rehab practice and homecare practice and by clinical experience.^{7,8} The test's broad scope was justified since many patients who receive inpatient rehabilitation also receive home care. Several sources guided test development.^{9-10,13} Test

items were distributed into three phases of care, namely, evaluation, treatment, or discharge, from both inpatient rehab and home care settings, with a focus on patients with cerebral vascular accident (CVA) or traumatic brain injury (TBI) (Table 1). These diagnoses were chosen because they were commonly seen in neurologic and geriatric practice and can present with similar cognitive and physical impairments.

SCTs have a unique question structure that consists of a case vignette followed by two additional sections. The vignette ends with a hypothesis for either evaluation, intervention or discharge planning and finishes with the phrase, 'if you were thinking ...'. The first additional section presents the learner with a new piece of information that must be integrated with the vignette. Lastly, in the second additional section, the learner answers whether the additional information has altered the hypothesis or not. On a 3-point Likert scale, the learner selects one of the three options: -1 (the hypothesis was unsupported by the new information); 0 (the new information did not affect the hypothesis); or 1 (the hypothesis was supported by the new information) (Appendix). A 3-point scale can be used for those at an early stage of CR development and to set standards¹³ without loss of reliability.^{9,24} In addition, a 3-point scale can reduce noise at the extreme ends of the scale, which have been described as representing fact-based knowledge rather than clinical judgment.^{16,20} Studies that used a 5-point scale for the SCT [-2, -1, 0, 1, 2] offered more decision-making possibilities; however, there is evidence that some are reluctant to choose the extreme ends of the scale (-2/2), and therefore, the selection of answers may be related to test-taking style, which the 3-point scale may help to mitigate.^{9,13,16}

Practicing clinicians panel: content validity

In SCT development, expert panels serve two purposes: first, they provide content validity and input on the test content, and second, they provide the test answers.^{9,20} A convenience sample of 26 physical therapists with at least 5 years of experience and actively employed in the fields of neurologic, geriatric, or home care PT was recruited between May and August 2018, with 20 tests completed and returned.¹³ This number exceeded the criteria for panel membership (at least 15 members), which has been found to provide adequate reliability.^{9,16,19} The modal score for each item was determined (there were no ties),

Table 1. Content and distribution of items ($n = 54$) for a SCT to measure CR for walking assistive device prescription

Examination		Treatment		Discharge	
Acute rehabilitation	Home care	Acute rehabilitation	Home care	Acute rehabilitation	Home care
12	6	15	9	9	3

SCT, script concordance test; CR, clinical reasoning.

which became the answer that received full credit.^{13,16} The modal scoring method was chosen to grade the residents' tests because this method allows for a single best answer, making scoring simple. The panel took the SCT separately and did not interact with other members. The final version of the SCT that was sent to residents included suggestions from panel members. The panel's consensus was that the test accurately reflected standard practice for walking AD prescription for inpatient rehabilitation and for home care settings.

Analysis

Tests were scored using the modal method,^{13,16} where residents were given 1 point for each answer that was congruent with the modal score from the experts. Frequencies were used to describe the panel and the residents' characteristics, as well as to obtain the mode for each test question. The overall reliability of the SCT was determined by Cronbach's alpha.¹⁰ Reliability of the clinician panel's answers and the residents' answers at months 1 and 6 was assessed using a Fleiss kappa. This analysis is designed for categorical data with multiple raters and examines the proportion of agreement over the amount of chance agreement.²⁵ Testing for differences between the clinician panel and the residents at months 1 and 6 was assessed using a Mann-Whitney U test for independent groups,²⁶ and testing for differences between the residents at the same time points was assessed using a Wilcoxon signed rank test for related groups.²⁷ Data analysis was performed using SPSS 28 (IBM, Chicago, IL). Human subjects' approval for the panel and resident sections of this study was granted by the Nazareth College Human Subjects Review Committee, HSRC #: SP2019-10.

Results

Panel characteristics/resident characteristics

The clinician panel ($n = 20$) had been in practice for a mean of 15.7 (7.0) years. Six participants were employed in skilled nursing facilities, 12 participants were employed in inpatient rehabilitation facilities, and nine participants were employed in home care (Table 2A). The majority of the 27 residents who enrolled in the study were studying in neurologic residencies ($n = 21$) and female ($n = 24$), with less than 1 year in practice ($n = 18$) (Table 2B).

Reliability

Internal consistency of the test for the panel, residents at month 1, and residents at month 6 was 0.73, 0.58, and 0.71, respectively. The Fleiss kappa results indicated that there was a fair amount of agreement over the possibility of chance for answers given on the SCT among the practicing clinicians and the residents at each time point [Clinicians: $K = 0.26$ (95% CI, 0.247 to 0.275), $P < 0.0005$;

Table 2. Panel and resident characteristics

2A: Panel characteristics (n = 20)	
Years in practice [Mean (sd)]	15.7 (7.0)
Area of practice	Frequency
Inpatient rehabilitation facility	12
Skilled nursing facility	6
Home care	9
Highest degree	
DPT	8
MS/BS	12
2B: Resident characteristics (n = 27)	
Type of residency	Frequency
Neurologic	21
Geriatric	6
Years in practice	
<1	18
1–2	6
>2	3
Age (years)	
24–25	13
26–29	11
>30	3

sd, standard deviation.

Residents at month 1: $K = 0.307$ (95% CI, 0.297 to 0.318), $P < 0.0005$; Residents at month 6: $K = 0.325$ (0.312 to 0.339), $P < 0.0005$].

SCT scores

Final test scores were calculated for the practicing clinicians and the residents at months 1 and 6 after the mode of each item was determined. Out of 54 items, the mean test scores for the clinicians were 34.8 (4.3), and the mean test scores for the residents at month 1 and month 6 were 34.5 (4.3) and 36.3 (3.9), respectively. Table 3 presents the calculated test scores for both groups. When comparing the clinicians' and month-1 residents' scores, there were 12/54 test items with different results; however, by the month-6 test results, there were only 5/54 test items with different results (Table 4). Over the 6 months of residency training, calculated scores increased in 12/20 residents, decreased in 7/20 residents, and remained the same in one resident. Statistically, one resident's raw scores increased on the month-1 and month-6 tests (Table 5).

Discussion

This study was designed to validate a SCT to measure the changes in CR ability for walking AD prescription in neurologic and geriatric PT residents. Inexperienced clinicians, like the residents included in this study, have been found to have less well-developed ISs than experts

Table 3. Calculated SCT scores for practicing clinicians and residents at month 1 and month 6

Practicing clinician	Score	Residents	Month 1	Month 6
1	34	1	36	42
2	42	2	41	38
3	34	3	30	–
4	29	4	38	42
5	40	5	35	37
6	35	6	34	–
7	27	7	38	38
8	42	8	38	–
9	33	9	35	32
10	36	10	34	39
11	29	11	43	–
12	41	12	31	39
13	35	13	43	38
14	36	14	35	31
15	34	15	31	43
16	36	16	37	–
17	34	17	38	35
18	29	18	28	–
19	32	19	33	35
20	38	20	30	32
		21	37	36
		22	33	31
		23	32	40
		24	33	34
		25	31	34
		26	25	30
		27	32	–

SCT, script concordance test.

within the same area of practice. The presence of ISs should broaden, solidify, and bring about the refinement of thought as a learner experiences continual clinical interactions and encounters with patients.^{12,17} ISs facilitate the development of the CR process and are considered by some as the hallmark of expertise in a specialty.^{10–12,17}

Internal consistency and reliability

The SCT in this study had moderate internal consistency, and the clinician panel deemed the content valid. The practicing clinicians were found to have fair agreement among them at a similar rate as other clinician panels.^{28,29} One reason posited for practicing clinicians having only fair agreement may be related to the gaps in information purposefully left by SCT writers to reflect the actual uncertainty in clinical practice and that each clinician ‘fills in’ differently, which leads to different responses to the same case. Another reason for different answer choices among the practicing clinicians may be related to how each clinician views the interaction of details present in

the vignette. Lastly, varied responses upon a second test administration of the same SCT among practicing clinicians were discovered, which may indicate that measuring uncertainty consistently is difficult and should warrant further investigation.²⁸ Additionally, over the first 6 months of their residency training, the reliability of the resident’s scores also remained fair and consistent.

Panel issues

Although guidance was followed for the development of our practicing clinician panel,^{9,16} the panel is a known contributor of error in SCT development. The practice areas of the SCT were represented in our panel (Table 2A), and having such clinical representation has been found to contribute to more reliable panel scores,⁹ which our sample achieved. Our sample also averaged over 15 years of clinical experience; however, demographic data on formal specialization in either home care PT, neurologic PT or geriatric PT were not collected. Limiting the panel to clinical specialists may have led to higher panel scores and thus may have yielded greater differences when compared to the residents’ scores.¹⁶ In a systematic review on SCT construction, Dory et al.⁹ found that panel scores are not associated with panel members being generalists versus specialists in their field. However, these authors admitted that a definitive definition of ‘expert’ clinician is lacking.

Resident scores

Several plausible reasons for the residents’ results should be considered.²⁸ First, since the PT residents were working in a focused practice area, the expertise gained during their first 6 months may have driven greater concordance with the choices made by the practicing clinicians on the SCT. Second, the overall test-taking strategies of the residents may have been better since they were all recent graduates from entry-level doctor of physical therapy (DPT) training. A third reason that could explain the similarity of scores between the residents and the practicing clinician panel is that those who seek a residency may not be representative of the entry-level skill of DPT graduates who do not seek a postgraduate residency. New graduates who seek a residency may possess an affinity for the specialty area of practice in which they have sought to continue to study and thereby have well-developed CR as they begin a residency program. To assess this possibility, an investigation comparing the CR of early clinicians who enter residency training and those who do not should be undertaken.

The neurologic and geriatric residents’ choices on the SCT were closely concordant with the practicing clinician panel at the beginning of their residency, and though not statistically significant, the mean scores of the residents at the mid-point of their 1-year residency began to

Table 4. Mann-Whitney U test results for differences between the practicing clinicians and residents at month 1 and month 6

Item	Practicing clinicians and residents at month 1			Practicing clinicians and residents at month 6		
	U	Z score	P	U	Z score	P
1a	276.0	0.267	0.79	210.5	0.62	0.78
1b	367.0	2.28	0.02*	159.0	-1.25	0.28
1c	293.5	0.55	0.58	166.0	0.30	0.37
2a	225.0	-1.11	0.27	143.0	0.6	0.13
2b	239.0	-0.76	0.45	219.0	0.53	0.62
2c	241.0	-1.29	0.20	170.0	-1.78	0.43
3a	186.5	-2.31	0.02*	152.5	-1.57	0.20
3b	242.5	-0.83	0.41	209.5	0.66	0.80
3c	277.0	0.19	0.85	187.5	-0.44	0.74
4a	402.0	3.14	0.002*	269.0	2.11	0.06
4b	229.0	-1.16	0.24	187.0	-0.44	0.74
4c	164.0	-2.80	0.005*	126.0	0.01	0.046*
5a	240.0	-0.87	0.38	149.0	-2.09	0.17
5b	289.5	0.47	0.64	218.5	0.56	0.62
5c	309.5	0.98	0.39	187.0	-0.44	0.74
6a	256.0	-0.46	0.65	187.0	-0.53	0.74
6b	262.0	-0.22	0.83	210.0	0.37	0.80
6c	164.0	-2.62	0.009*	136.0	-1.93	0.09
7a	266.0	-0.13	0.90	180.0	-0.87	0.60
7b	346.0	1.84	0.07	266.0	0.05	0.08
7c	227.5	-1.04	0.30	200.0	0.00	1.00
8a	256.0	-0.32	0.74	135.5	-1.87	0.08
8b	226.0	-1.129	0.26	148.5	-1.69	0.17
8c	272.0	0.05	0.96	161.0	-1.21	0.30
9a	274.0	0.25	0.81	210.0	0.32	0.80
9b	282.0	0.45	0.66	201.5	0.95	0.97
9c	249.5	-0.52	0.61	215.0	0.65	0.70
10a	349.0	2.12	0.03*	234.5	0.28	0.36
10b	334.5	1.71	0.09	204.5	0.89	0.90
10c	195.0	-1.96	0.06	117.0	-2.96	0.02*
11a	279.5	0.24	0.74	183.0	0.57	0.66
11b	186.0	-2.12	0.03*	122.0	-2.70	0.007*
11c	186.5	-2.11	0.04*	146.5	-1.75	0.15
12a	235.0	-1.06	0.29	192.0	-0.29	0.84
12b	240.0	-0.75	0.45	224.0	0.45	0.53
12c	268.0	-0.05	0.96	185.5	-0.48	0.70
13a	305.0	0.85	0.40	220.5	0.53	0.58
13b	258.0	-0.28	0.78	181.0	0.58	0.62
13c	222.5	-0.83	0.41	182.0	-0.24	0.84
14a	273.0	0.19	0.85	179.5	0.29	0.58
14b	278.5	0.20	0.84	194.5	-0.16	0.88
14c	231.0	-0.96	0.34	166.0	-1.03	0.37
15a	311.0	1.24	0.22	159.5	-1.28	0.28
15b	292.0	0.53	0.60	229.0	0.38	0.45
15c	346.5	1.90	0.06	181.0	-0.60	0.62
16a	223.0	-1.27	0.21	148.5	-1.79	0.17
16b	177.0	-2.32	0.02*	91.5	-3.39	0.003*

Table 4. (Continued)

Item	Practicing clinicians and residents at month 1			Practicing clinicians and residents at month 6		
	U	Z score	P	U	Z score	P
16c	240.0	-0.71	0.48	162.0	-1.12	0.31
17a	292.0	0.53	0.60	245.5	0.18	0.22
17b	250.0	-0.46	0.64	193.0	0.84	0.86
17c	165.0	-2.55	0.01*	100.0	-3.11	0.006*
18a	199.0	-2.48	0.01*	150.0	-2.05	0.18
18b	185.0	-2.04	0.04*	148.0	-1.57	0.17
18bc	299.5	1.31	0.19	168.5	-1.17	0.40

*Significance: $p < 0.05$.

Table 5. Wilcoxon signed rank results for differences between residents' scores at month-1 and month-6

Resident	Result
1	0.93
2	0.17
3	0.24
4	0.56
5	0.11
6	0.76
7	0.79
8	0.08
9	0.39
10	0.09
11	0.67
12	0.30
13	0.07
14	0.46
15	0.77
16	0.28
17	0.77
18	0.93
19	0.01*
20	0.11

*Significance: $p < 0.05$.

exceed the mean scores of the practicing clinician panel by month 6. Data collection ended at the residents' 6-month mark, leaving 6 months remaining before the residency year was completed; therefore, it is unknown whether residents' scores would have continued to increase significantly beyond the panel's scores. Further investigation is warranted into the growth of CR among PT residents during a full residency year.

A SCT presents learners with a challenging and clinically relevant form of testing. The test structure presents commonly seen cases with ambiguity, reflecting the actual skill needed in clinical practice, that is, making decisions

under uncertainty. Questions can be developed for any part of PT patient management, including examination, prognosis, intervention, and discharge planning.¹⁶ One way that a SCT could be valuable to a residency program is by giving the programs the potential to show growth in CR. For instance, the results of scaffolded teaching can be measured using a SCT either within a module or over a series of modules or over the entire residency. A SCT could also be valuable to a residency program by using the results as a tool to promote discussion between the instructor and the resident about clinical approaches that could be considered using the vignettes aligned with the practice setting.^{8,16,30} Additional research using the SCT in entry-level and postgraduate PT education could contribute to understanding how and when growth in CR skill advances in entry-level and postgraduate learners. A longitudinal study of learners may discern what stimulates growth and when advancement in CR occurs during training. This type of study could also identify and describe behaviors indicative of advancement in CR ability.

Walking AD prescription is complex and requires a multifactorial approach, and the process differs somewhat based on where a patient encounter occurs, for example, inpatient rehabilitation or in home care.^{7,8} The environment and patient preference are central to CR in home care; while these factors are also addressed during inpatient care, they often cannot be thoroughly replicated.^{7,8} Furthermore, there is a lack of high-level evidence to inform practice for the use and progression of patient training in walking ADs. Indeed, there is no clinical prediction tool to guide clinicians as to the best timing for patient progression to another device or from a device. A lack of such clinical guidance may have led to greater uncertainty than expected when test takers attempted to synthesize the ambiguous nature of a SCT item, and such uncertainty may have led to error in scores for both the panel and the residents. Additional investigation into methods of addressing walking AD progression would benefit the PT profession. Furthermore, follow-up with the panel members or residents to understand their rationale for test answers or for test-taking strategy may have revealed issues in CR within these realms, yet this was beyond the scope of this study. New studies should include post-testing follow-up for this purpose.

Limitations

The results of this study pertain only to neurologic and geriatric residencies. Furthermore, while residency programs have the same basic requirements, variations in curricula and mentors would be expected. One limitation is the potential for multiple testing errors created by the number of items assessed for significance; to address this, Table 4 provides *z*-scores to show the magnitude of the mean differences. Another limitation may stem from inexperience

with this still relatively novel form of testing and could have played a role in our findings. Future SCT developers should invest in carefully training all test takers to avoid error in test scores from a lack of comfort in the testing method. Additionally, including items from both inpatient rehab and home care may have made the SCT in this study too broad in scope. Although many patients, such as those with stroke and brain injury, receive care from inpatient rehabilitation and from home care settings, including items across the continuum of care may have required a different knowledge base. Furthermore, we did not ask residents whether a home care setting was part of the residency curriculum, so if residents had not or would not have home care experience within the residency, then the test may have exceeded their knowledge. Finally, we were unable to carry out the study for the entire 1-year residency; therefore, we may not have captured the entire growth in CR for this sample.

Conclusion

A SCT is a clinically relevant test meant to closely resemble the CR used in routine clinical practice. The SCT was reliable and valid yet did not discern differences in CR between the practicing clinicians and the residents at the beginning or at 6 months into the residency. In addition, no significant growth was found in the CR for AD prescription during the first 6 months among neurologic and geriatric residents. Resident scores closely approximated panel scores, indicating that the resident sample had similar CR skills as the experienced clinicians. The use of SCTs to measure the growth in CR for PT residencies requires more investigation.

Acknowledgment

Statistical assistance received from Binghamton University Statistical Consulting Services Director, Dr. Mei-Hsui Chen PhD.

Conflict of interest and funding

The authors have not received any funding or benefits from industry or elsewhere to conduct this study.

This study received the following institutional approval: Nazareth College's Human Subjects Research Committee #: SP2018-23/SP2019-10.

References

1. Gilliland S. Clinical reasoning in first- and third-year physical therapist students. *J Phys Ther Ed* (2014) 28(3): 64–80. doi: 10.1097/00001416-201407000-00009
2. Gilliland S, Wainwright SF. Patterns of clinical reasoning in physical therapist students. *Phys Ther* (2017) 97: 499–511. doi: 10.1093/ptj/pzx028
3. Wan MSH, Tor E, Hudson JN. Construct validity of script concordance testing: progression of scores from novices to experienced clinicians. *Int J Med Ed* (2019) 10: 174–9. doi: 10.5116/ijme.5d76.1eee

4. Huhn K, Gilliland SJ, Black LL, et al. Clinical reasoning in physical therapy: a concept analysis. *Phys Ther* (2019) 99: 440–56. doi: 10.1093/ptj/pzy148
5. American Council of Academic Physical Therapy. Student readiness for the first full-time clinical experience. Available from: https://acapt.org/docs/default-source/public-docs/ksas-and-levels-of-competency-for-considering-student-readiness.pdf?s-fvrsn=47da8bd8_2 [cited 1 October 2022].
6. American Board of Physical Therapy Residency and Fellowship Education. Quality standards for clinical residency and fellowship programs. 2022. Available from: <https://abptrfe.apta.org/for-programs/clinical-programs/quality-standards-clinical> [cited 29 September 2022].
7. O'Brien SR, Durr K, Laubisch E, et al. Every person is an individual: physical therapist clinical reasons used in inpatient rehabilitation for walking assistive device prescription in patients with stroke and brain injury. *Disab Rehabil Assist Technol* (2021) 16(1): 1–8. doi: 10.1080/17483107.2019.1647568
8. O'Brien SR, Barry M, Davidson E, et al. Physical therapist clinical reasoning in home care for walking assistive device prescription: a description of practice. *Physiother Theory Pract* (2023) 39(1): 80–8. doi: 10.1080/09593985.2021.1996495
9. Dory V, Gagnon R, Vanpee D, et al. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ* (2012) 46(6): 552–63. doi: 10.1111/j.1365-2923.2011.04211.x
10. Fournier JP, Demeester A, Charlie B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak* (2008) 8: 18. doi: 10.1186/1472-6947-8-18
11. Schmidt HG, Rikers RMJP. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ* (2007) 41: 1123–39. doi: 10.1111/j.1365-2923.2007.02915.x
12. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med* (2000) 75(2): 182–90. doi: 10.1097/00001888-200002000-00020
13. Kelly W, Durning S, Denton G. Comparing a script concordance examination to multiple-choice examination on a core internal medicine clerkship. *Teach Learn Med* (2012) 24(3): 187–93. doi: 10.1080/10401334.2012.692239
14. Piovezan RD, Custodio O, Cendoroglo MS, et al. Assessment of undergraduate clinical reasoning in geriatric medicine: application of a script concordance test. *J Am Geriatr Soc* (2012) 60(10): 1946–50. doi: 10.1111/j.1532-5415.2012.04152.x
15. Esteves JE, Bennison M, Thomson OP. Script concordance test: insights from the literature and early stages of its implementation in osteopathy. *Int J Osteopath Med* (2013) 16(4): 231–9. doi: 10.1016/j.ijosm.2013.05.001
16. Lubarsky S, Dory V, Dugan P, et al. Script concordance testing: from theory to practice: AMEE guide no. 75. *Med Teach* (2013) 35: 184–93. doi: 10.3109/0142159X.2013.760036
17. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory and implications. *Acad Med* (1990) 65(10): 611–21. doi: 10.1097/00001888-199010000-00001
18. Script concordance tests. University of Montreal School of Medicine. Available from: [https://www.cpass.umontreal.ca/recherche/axes-de-recherches/concordance/sct/](https://www.cpass.umontreal.ca/https://www.cpass.umontreal.ca/recherche/axes-de-recherches/concordance/sct/) [cited 11 August 2021].
19. Steinberg E, Cowan E, Lin MP, et al. Assessment of emergency medicine residents' clinical reasoning: validation of a script concordance test. *Western J Emerg Med* (2020) 21(4): 978–84. doi: 10.5811/westjem.2020.3.46035
20. Lineberry M, Kreiter CD, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. *Med Educ* (2013) 47(12): 1175–83. doi: 10.1111/medu.12283
21. Erickson W, Lee C, von Schrader S. 2018 Disability status report: United States. Ithaca, NY: Cornell University Yang Tan Institute on Employment and Disability (YTI); 2021.
22. Hubbard Winkler S, Cowper Ripley D, Wu S, et al. Demographic and clinical variation in veterans health administration provision of assistive technology devices to veterans poststroke. *Arch Phys Med Rehabil* (2010) 91: 369–77. doi: 10.1016/j.apmr.2009.10.028
23. Marzolini S, Balitsky A, Jagroop D, et al. Factors affecting attendance at an adapted cardiac rehabilitation exercise program for individuals with mobility deficits poststroke. *J Stroke Cerebrovasc Dis* (2016) 25: 87–94. doi: 10.1016/j.jstrokecerebrovasdis.2015.08.039
24. Bland AC, Kreiter CD, Gordon JA. The psychometric properties of five scoring methods applied to the script concordance test. *Acad Med* (2005) 80: 395–9. doi: 10.1097/00001888-200504000-00019
25. Laerd Statistics. Fleiss kappa using SPSS statistics. Statistical tutorials and software guides. Available from: <https://statistics.laerd.com/>
26. Laerd Statistics. Mann-whitney U test using SPSS statistics. Statistical tutorials and software guides. Available from: <https://statistics.laerd.com/>
27. Laerd Statistics. Wilcoxon signed rank test using SPSS statistics. Statistical tutorials and software guides. Available from: <https://statistics.laerd.com/>
28. Lineberry M, Hornos E, Pleguezuelos E, et al. Exerts' responses in script concordance tests: a response process validity investigation. *Med Educ* (2019) 53: 710–22. doi: 10.1111/medu.13814
29. Peyrony O, Hutin A, Truchot J, et al. Impact of panelists' experience on script concordance test scores of medical students. *BMC Med Educ* (2020) 20(1): 313. doi: 10.1186/s12909-020-02243-w
30. Subra J, Chicoulaa B, Stillmunkès A, et al. Reliability and validity of the script concordance test for postgraduate students of general practice. *Eur J Gen Practice* (2017) 23(1): 209–14. doi: 10.1080/13814788.2017.1358709

***Suzanne R. O'Brien**

Division of Physical Therapy
 School of Rehabilitation Science
 Decker College of Nursing and Health Sciences
 Binghamton University
 48 Corliss Avenue, Office 621
 Johnson City, NY 13790 USA
 Tel: (585) 747-7310
 Email: sobrien@binghamton.edu

Appendix

EXAMPLE

Directions: Each vignette has 3 *unrelated* questions. Read each vignette, then read the first column “If you were thinking of”, which begins each question. Reading across, consider the new information presented in the “And then you find” column. Answer the question by circling the number that BEST answers the question in the “This thinking is” column using the Legend below:

NOTE: Questions may be similar, but are *unrelated* and should be considered mutually exclusive.

Legend: -1: Unsupported 0: Neither supported nor unsupported 1: Supported
--

<p>1. This column begins each of the 3 questions. Read this information after the clinical vignette.</p>	<p>1. Clinical Vignette: The patient is an 80 year old female who lives alone in a 2-story colonial with 2 steps to enter and 13 steps to the 2nd floor. She sustained a left CVA one month ago and has been discharged home. Since being home, she has not been using the wide based quad cane and cruises walls and furniture. How would you treat this patient?</p>		
	If you were thinking of...	And then you find...	This hypothesis is...
	Practicing with a wide based quad cane.	She is independent throughout her home, including stairs.	-1 0 1
	Practicing with a straight cane.	Her unilateral stance time is 2 seconds on the left and 0 seconds on the right.	-1 0 1
	Having her perform stairs without a railing.	She has arthritis in her left knee.	-1 0 1
	<p>2. Then, read the new information in this column and use it to determine your decision in the last column.</p>		
			<p>3. This is where you choose your answer to the question. Highlight ONE.</p>

THANK YOU!