European Journal of Neurology 2014, **21:** 1419–1422 doi:10.1111/ene.12373

SHORT COMMUNICATION

Validating a script concordance test for assessing neurological localization and emergencies

K. Tan[a,b], N. C. K. Tan[a,b], N. Kandiah[a,b], D. Samarasekera[b] and G. Ponnamperuma[c]

[a]*Department of Neurology, Office of Neurological Education, National Neuroscience Institute;* [b]*Yong Loo Lin School of Medicine, National University of Singapore, Singapore; and* [c]*Faculty of Medicine, University of Colombo, Sri Lanka*

Keywords:
assessment tool, clinical judgment, neurological emergencies, neurological localization, script concordance test

Received 2 September 2013
Accepted 27 December 2013

Background and purpose: Clinical judgment is the ability to weigh clinical information and make decisions under conditions of uncertainty. Although neurological localization (NL) and neurological emergencies (NE) present such uncertainties, no validated method is reported to assess these decision-making skills. A script concordance test (SCT) was designed and validated to assess clinical judgment in NL and NE.

Methods: Our SCT comprised 14 clinical scenarios (53 questions). Candidates picked the response they considered the best for the questions in each scenario. Undergraduates and internal medicine residents completed the SCT; their responses were scored against the scoring key derived from an expert panel of accredited neurologists. Scores were expressed as a percentage of the maximum score.

Results: Mean total scores for undergraduates (n = 52), residents (n = 37) and experts (n = 15) were 61.0 ± 0.9, 68.3 ± 1.1 and 76.6 ± 1.1 (mean ± standard error of the mean, $P < 0.001$). Mean scores for undergraduates, residents and experts were 59.3 ± 1.1, 66.4 ± 1.4 and 76.1 ± 1.8 ($P < 0.001$) for NL, and 62.9 ± 1.3, 70.4 ± 1.3 and 77.2 ± 1.6 ($P < 0.001$) for NE. Senior residents scored higher than junior residents (postgraduate years 2–5 versus postgraduate year 1, 69.7 ± 1.4 vs. 65.3 ± 1.1, $P = 0.035$). The higher scores with increasing clinical experience supports the construct validity of the SCT. The SCT showed acceptable reliability (G coefficient 0.74 ± 0.05).

Conclusions: Our SCT is validated to reliably assess NL and NE in undergraduate and postgraduate learners; it is generalizable and feasible. It has potential as a valuable adjunct assessment tool for clinical judgment. Future plans to design SCTs to evaluate other topics in clinical neurology, as a multi-center study, are under way.

Introduction

Physicians need to learn to evaluate and manage clinical problems which often present with a continuum of uncertainty and several reasonable hypotheses and courses of action are acceptable. They decide by estimating the effect of relevant information on the hypotheses in question. Clinical judgment is the ability to weigh clinical information and make decisions under conditions of uncertainty. This is a crucial skill for practising neurologists but is difficult to test objectively [1].

Script theory, derived from cognitive psychology, provides an explanation of how information is mentally processed. In medicine, illness scripts develop from the application of biomedical and clinical sciences knowledge to real patients. Each encounter with a patient with a specific disease will add and reorganize information to the related illness script [2].

The script concordance test (SCT), an assessment tool based on script theory, is a validated, reliable and useful method for assessing clinical judgment under conditions of uncertainty [3]. It measures how closely the examinees' clinical decisions are in concordance with those of experts. It has been used to assess clinical judgment in several clinical disciplines including neurology [1,3]. In this study, an SCT focused on assessing clinical judgment in neurological localization (NL) and neurological emergencies (NE), key topics in undergraduate and postgraduate neurology training in Singapore, was designed and validated as there are currently few standardized assessment tools available.

Correspondence: K. Tan, Department of Neurology, National Neuroscience Institute, 11 Jalan Tan Tock Seng, Singapore 308433, Singapore (tel.: +65-6357-7171; fax: +65-6357-7137; e-mail: Kevin_Tan@nni.com.sg).

european journal of neurology
the official journal of the european academy of neurology

Methods

The SCT items using 14 clinical scenarios (seven on each topic) with three to five questions in each scenario were initially formulated by the authors NCKT and KT. A content blueprint ensured that NL (dysarthria, encephalopathy, visual symptoms, nystagmus and various patterns of weakness and numbness) and NE (acute stroke, headache, weakness and numbness, status epilepticus and central nervous system infections) scenarios were comprehensive. The scenarios were designed to reflect uncertainty in real-life clinical situations (Fig. 1). The final test was optimized to 53 items using item analysis methodology where overly concordant (poor discriminant value) or discordant (measurement error) questions were discarded.

Previously established scoring procedures were used for our SCT [1]. The scoring key was derived from an expert panel of 15 attending neurologists from the Department of Neurology, National Neuroscience Institute, Singapore; all had experience managing inpatients and outpatients, including acute emergencies. Third-year medical undergraduate students and internal medicine residents [postgraduate years (PGY) 1–5] completed the SCT and were scored against the scoring key; scores were expressed as a percentage of the maximum score. Approval was received from the National University of Singapore Institutional Review Board to conduct this study; all participants gave informed consent.

Reliability of scores was tested using generalizability (G) theory. The G-study design was 'persons (or candidates) fully crossed with questions nested within items (or scenarios) nested within sections' {persons into [questions within items (i.e. case scenarios) within sections (i.e. NL or NE)]; i.e. $p \times (q : i : s)$}. For multivariate analysis, sections were considered as a fixed facet. A one-way ANOVA with Bonferroni correction for pairwise comparisons was used to compare SCT scores amongst the three groups, at a significance of $P \leq 0.05$.

Results

All participants (52 students, 37 PGY1–5 residents and 15 experts) completed the test within 30 min.

Comparison of mean total scores for students, residents and experts yielded statistically significant differences (61.0 ± 0.9 vs. 68.3 ± 1.1 vs. 76.6 ± 1.1; mean ± standard error of the mean, $P < 0.001$) (Fig. 2a). Junior residents (PGY1, $n = 25$) had significantly lower mean total scores compared with senior residents (PGY2–5, $n = 12$) (65.3 ± 1.1 vs. 69.7 ± 1.4, $P = 0.035$). Comparison of mean scores for students, residents and experts in NL (59.3 ± 1.1 vs. 66.4 ± 1.4 vs. 76.1 ± 1.8, $P < 0.001$) (Fig. 2b) and in NE (62.9 ± 1.3 vs. 70.4 ± 1.3 vs. 77.2 ± 1.6, $P < 0.001$) (Fig. 2c) yielded similar statistically significant differences.

The higher scores with increasing clinical experience provide evidence that supports the construct validity of the SCT. The SCT also showed acceptable generalizability (G coefficient 0.74 ± 0.05). However, eight scenarios in each topic, with five questions in each scenario, are needed to achieve a reliability of 0.8.

Discussion

Medical accreditation bodies worldwide recognize the need to assess physician competence beyond factual

(a)

A 55y old male is referred for a 6 month history of dysarthria. Palatal movement is impaired. He has a history of diabetes.						
If you were thinking of this diagnosis…	*And you find…*	*This hypothesis would become*				
		Very unlikely	Unlikely	Neither likely nor unlikely	More likely	Very likely
a.Myasthenia gravis	Bilateral ptosis	–2	–1	0	+1	+2
b.Motor Neuron Disease	Normal deep tendon reflexes	–2	–1	0	+1	+2
c.Pseudobulbar palsy	Upgoing left plantar response	–2	–1	0	+1	+2
d.Multiple cranial nerve palsies	Bilateral tongue wasting	–2	–1	0	+1	+2
e.Cerebellar degeneration	Absence of dysdiadochokinesia	–2	–1	0	+1	+2

(b)

A 60 y old diabetic male on glipizide was last seen well at noon. He is brought into A&E at 3.15pm for drowsiness. He has a right hemiparesis, MRC grade 3.						
If you were considering the usefulness of this treatment…	*And this new information becomes available …*	*You would then consider this treatment*				
		Useless	Less useful	Neither more or less useful	Useful	Very useful
a. intravenous rTPA	Patient's hemiparesis improves to 4+ in A&E	–2	–1	0	+1	+2
b. intravenous rTPA	Past history of epilepsy	–2	–1	0	+1	+2
c. oral aspirin	BP 240/120 mmHg	–2	–1	0	+1	+2

Figure 1 Examples of SCT items in neurological localization (a) and neurological emergencies (b).

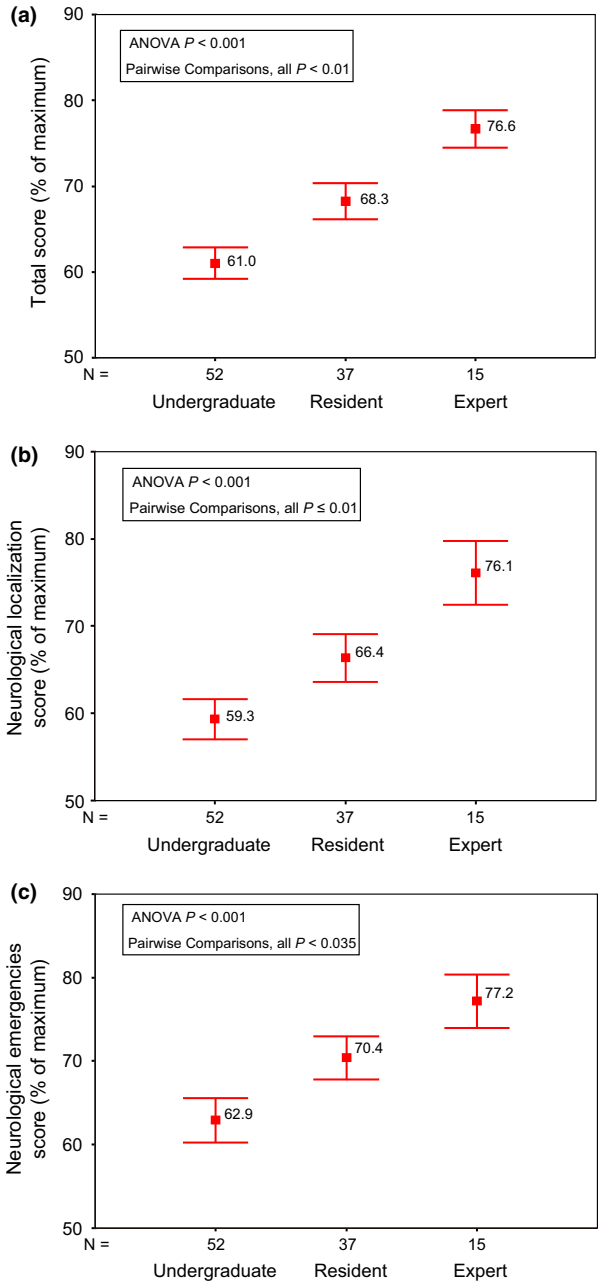

Figure 2 Comparison by level of training (undergraduate students versus residents versus experts) of total scores (a), neurological localization scores (b) and neurological emergencies scores (c). Mean scores and standard error of the mean are shown.

knowledge [1]. Commonly used assessment tools (multiple-choice questions, other written examinations and objective structured clinical examinations), however, are resource-intensive, troublesome to administer or score, difficult to standardize, or predominantly test pure factual knowledge [4]. Currently, clinical judgment is assessed subjectively and rather informally in most training settings. The SCT seeks to practically and more objectively evaluate this skill by mimicking the ambiguity seen in the real world.

Our SCT has several important advantages over other assessment tools. First, it did not suffer from an 'intermediate effect', a phenomenon whereby intermediate-level trainees outperform both novices and experts. Rather, residents scored lower than expert neurologists but higher than students, supporting test construct validity. Furthermore, junior residents scored lower than senior residents, displaying the potential discriminatory power of the test. Secondly, the test is generalizable (G coefficient was 0.74) and administering it was easy. Using Van der Vleuten's criteria for determining the utility of assessment methods [5], our SCT demonstrates reliability, construct validity and low cost. Thirdly, it has an impact on learning and relevant practice as NL and NE are key neurology topics. Further studies are required to determine the test's acceptability to learners and faculty.

There are several limitations in our study. First, our study was based on a moderate sample size. Secondly, generalizability may be optimized by increasing the number of questions to 80 to achieve a G coefficient of 0.8. Thirdly, the script concordance approach does not evaluate the examinees' reasoning process. Based on script theory, no two physicians will have the same repertoire of scripts. However, in similar clinical contexts, experts in medicine tend to arrive at conclusions that are largely concordant, despite using differing ways of reasoning and posing questions and examining patients differently [1]. Finally, our experts and trainees were drawn from a single institution and medical school.

In conclusion, our study shows that our SCT is a valid and reliable tool for assessing clinical judgment in NL and NE. It appears to be practical, impactful and generalizable. This SCT has potential as a valuable adjunct to other traditional standardized methods of evaluation of medical students and neurology trainees. Future plans to compare the performance of expert panels from different institutions and design SCTs which can be used to evaluate other clinical neurology topics are under way.

Acknowledgements

The authors wish to thank members of the Department of Neurology, National Neuroscience Institute, for their participation and Ms Kay Tan and Ms Ivane Chew for administrative support.

Disclosure of conflicts of interest

The authors declare no financial or other conflicts of interest.

References

1. 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.
2. Charlin B, Boshuizen HPA, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ* 2007; **41:** 1178–1184.
3. 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.
4. Epstein RM. Assessment in medical education. *N Engl J Med* 2007; **356:** 387–396.
5. Van Der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996; **1:** 41–67.