Duggan P, Charlin B
MedEdPublish
http://dx.doi.org/10.15694/mep.2016.000060

Original article **Open Access**

Introducing recent medical graduates as members of Script Concordance Test expert reference panels: what impact?

Paul Duggan[1], Bernard Charlin[2]

Corresponding author: Prof Paul Duggan paul.duggan@adelaide.edu.au
Institution: 1. Discipline of Obstetrics & Gynaecology, The University of Adelaide, 2. Centre for Pedagogy Applied to the Health Sciences (CPASS), University of Montreal
Categories: Assessment, Students/Trainees

Received: 02/08/2016
Published: 08/08/2016

Abstract

The Script Concordance Test (SCT) is being increasingly used in professional development in clinical reasoning, with linear progression in performance in SCT's observed with increasing clinical experience. One of the limiting factors for the SCT is potential burnout in expert reference panel (ERP) members, which we have attempted to address by the introduction of recent medical graduates as panel members. We sought to evaluate the effect of introducing recent medical graduates in to our ERP's on pass/fail decisions in the final clinical reasoning examination of the 6-year undergraduate program of the University of Adelaide, Australia. We engaged an ERP comprising 50 faculty members from three collaborating universities and 13 recent medical graduates to answer on line an identical 20 case scenario, 50 question multidisciplinary SCT twice 6 months apart. The questions were used in high stakes end of year assessment of 5th year medical students (n=132). The pass mark set by the experienced, specialist members of the panel was 49.6% and this increased to 50.4% by addition of recent medical graduates to the panel. This difference would have had no effect on fail rates estimated from the data from the cohort of 132 medical student candidates. In the context of assessment of clinical reasoning in medical programs, recent medical graduates are suitable members of SCT ERP's, and their contribution can enrich the panel and might help to minimise risk of burnout of more experienced faculty.

Keywords: Script Concordance Test, Recent Medical Graduates, Expert Reference Panel Membership, Summative Assessment, Medical Students

Article

Introduction

"Clinical reasoning" may be defined as the process by which a clinician arrives at a diagnosis and/or plan of management. There is a substantial literature on the subject, recently summarized by Durning

and colleagues (2013). Developing the clinical reasoning abilities of medical students is a key function in medical education and, perhaps in recognition of the importance of diagnostic errors, (Saber Tehrani et al 2013) tests of clinical reasoning are becoming increasingly important. To assess clinical reasoning, Charlin developed the Script Concordance Test (SCT) - a written assessment based on Script Theory, which hypothesises how physicians progressively acquire knowledge adapted to their clinical tasks. (Charlin et al 2000a, Charlin et al 2000b) The SCT utilises clinical scenarios designed to measure data interpretation under conditions of uncertainty, which is germane considering that uncertainty in medical diagnosis and treatment is a common state in clinical practice. (Beresford 1991) The SCT is unique in that members of expert reference panels sit the test in advance of its deployment and that answer keys from those data are used to calculate the subsequent test scores of candidates.

The SCT has been used in assessment in a wide range of clinical disciplines. (Brailovsky et al 2001, Brazeau-Lamontagne et al 2004, Lambert et al 2009, Lubarsky et al 2009, Meterissian et al 2006, Park et al 2010) Tests have been shown to be statistically reliable and have showed construct validity. (Lubarsky et al 2011) The SCT is increasingly being used in continuing professional development in medicine. (Ahmadi et al 2014) (Hornos et al 2013)

A few studies have described the use of the SCT to assess reasoning among medical students in specific domains. (Collard et al 2009, Duggan 2007, Duggan and Charlin 2012, Monnier et al 2011, Duggan et al 2016)

Since 2008 the undergraduate medical program of the University of Adelaide has used the SCT as a summative test of clinical reasoning in its end of year written assessments. (Duggan and Charlin 2012) This multidisciplinary assessment, comprising questions from our six core clinical disciplines, has required up to 180 items. Our experience is that about twice that number of items must be submitted to and answered by our expert reference panel to obtain sufficient good quality questions with an appropriate range of preferred Likert responses, necessary to minimise benefit to candidates from the practice of strategic answering, i.e. strategic selection of the less extreme descriptors from the range or Likert responses. (Duggan and Charlin 2012) Gagnon et al (2005) report that for optimal reliability up to 20 members of the SCT expert reference panel are recommended. We have found it difficult to recruit and train that number of expert panelists across all clinical disciplines and have had to develop strategies to avoid burnout in our experienced, volunteer members of expert reference panels. In undergraduate education, this burden tends to fall disproportionately on colleagues in General Practice, who are the most versatile and thus most called-upon members of our ERP's. Strategies to avoid expert burnout have included sharing work with sister universities, replacing single discipline panels with multidisciplinary panels, and engaging recent medical graduates as members of ERP's, which is the focus of this research. Lubarsky et al (2011) note that in SCT's in the postgraduate domain there is statistically linear progression of scores with clinical experience. However, for an undergraduate medical program it is uncertain if the addition of recent medical graduates to expert reference panels would have an effect on the outcome and in particular the crucial pass/fail cut point for SCT's used in summative assessment. We report here our experience with this approach. The research questions were: 1) Would an expert reference panel comprising recent medical graduates lower the standard required for demonstration of competence in an undergraduate multidisciplinary SCT? 2) Is the performance of our multi-disciplinary expert reference panel, including our recent medical graduates, stable over time?

Methods

Structure, production and scoring of the SCT

The SCT format is shown in Figure 1. This provides a clinical scenario, a hypothesis or plan of action based on the scenario, and some additional information that may or may not have an effect on the hypothesis or plan. Each scenario is followed by a number of questions. For each question, the participant selects the single best Likert response that describes the effect of the additional information that has been given. In contrast to many conventional forms of written testing (e.g. multiple choice questions), there is no single correct answer; several responses to each question may be considered acceptable. Credit is assigned to each response based on the proportion of experts on the reference panel choosing that response. A maximal score of 1 is given for the response chosen by most of the experts (i.e., the modal response). Other responses are given partial credit in proportion to the number of experts choosing them. (Lubarsky et al 2013)

Figure 1: an example of a SCT vignette with two questions.

The information in each question stands alone – i.e. when considering the answer to Q1 there is no oxygen saturation result available and for Q2 there is no chest X-ray result available.

You are called to a hospital ward to evaluate a 74-year-old woman three days following vaginal hysterectomy and anterior repair for prolapse. She is complaining of a sore leg and now feels short of breath whilst sitting in a chair.

	If you are considering the following investigation ...	and then you find ...	you would then consider the proposed investigation to be …
Q1	A ventilation-perfusion scan to rule out pulmonary embolism	Her chest X-ray demonstrates areas of collapse	• much less useful • slightly less useful • neither less nor more useful • slightly more useful • much more useful
Q2	An arterial blood gas	Her oxygen saturation whilst breathing room air is 96%	• much less useful • slightly less useful • neither less nor more useful • slightly more useful • much more useful

Development and selection of SCT questions

The questions were developed by teaching faculty in the University of Adelaide for deployment in end of year summative assessments in the final examination of the medical program held at the end of Year 5 of the 6-year undergraduate program. We use up to 180 questions in this type of assessment. A sub-group of 50 questions based on 20 case scenarios was selected for this research. Twelve questions were diagnostic, 7 questions on investigation, and 31 questions on management. The 50 questions were chosen to reflect a representative range of conditions from our assessment blueprint (see table 1). The number 50 was selected as a reasonable optimal number for our volunteer panel members to sit twice, with each session expected to take up to 1 hour to complete.

Table 1: Outline of the 20 clinical cases used in the 50-question Script Concordance Test. In this research, between 2-4 questions were provided per clinical case.

Case	TYPE	SUBJECT	Setting	Onset
1	Diagnosis	Breast lump 46 year old woman	Community	Acute
2	Diagnosis	Chest pain, fever 65 year old man	Emergency Department	Acute
3	Diagnosis	Cough, wheeze 3 year old girl	Emergency Department	Acute
4	Diagnosis	Enuresis 8 year old boy	Community	Chronic
5	Diagnosis	Upper abdominal pain 47 year old woman	Community	Acute
6	Investigation	Diabetic foot	Community	Acute
7	Investigation	Painless jaundice 64 year old man	Community	Acute
8	Management	Painful preterm antepartum haemorrhage	Rural hospital	Acute
9	Management	Contraception 32 year old woman	Community	Chronic

10	Management	Bleeding in early pregnancy	Emergency Department	Acute
11	Management	Preterm premature rupture of the membranes	Maternity hospital	Acute
12	Management	Acute shortness of breath, chronic asthma 55 year old woman	Community	Acute
13	Management	Intermittent vomiting 25 year old woman	Community	Acute
14	Management	Chest pain ST elevation 65 year old woman	Emergency Department	Acute
15	Management	Depression, weight loss in pregnancy	Community	Acute
16	Management	Hypertension, obesity 32 year old man	Community	Chronic
17	Management	Threatened suicide 25 y old man	Emergency Department	Acute
18	Management	Depression 52 year old man	Community	Chronic

19	Management	Urethral discharge following recent overseas travel 26 year old man	Community	Acute
20	Investigation	Abdominal pain after recent surgery 40 year old woman	Emergency Department	Acute

Composition of the multidisciplinary expert reference panel

We recruited a multidisciplinary expert reference panel comprising 50 teaching staff from the University of Adelaide and two other collaborating Australian universities, with whom we share questions and undertake benchmarking, and 13 recent medical graduates of the University of Adelaide with affiliate status in the university. On two separate occasions 6 months apart all panel members sat on line the same 50-item test by accessing a secure, anonymous on line facility provided by the University of Montreal. All panelists were volunteers with prior experience either as SCT expert reference panel members, or, for our recent graduates, as prior SCT candidates. All were actively engaged in clinical practice and teaching, with the recent graduates in their 2nd year after graduation from medical school. The choice of a 6 month interval between tests was arbitrary, and based on assumptions including that the panelists, who had no access to the questions between tests, would likely have forgotten the questions in that time, and that the time interval was too short for there to have been large shifts in either medical practice or reasoning abilities of the panelists. The make up of the panel is shown in table 2.

Ethical approval was not required for this work.

Table 2: Composition of the multidisciplinary expert reference panel.

Discipline	Number of panellists
Recent Medical Graduates	13
General Practitioners	17
Physicians	16
Obstetricians and Gynaecologists	11
Psychiatrists	5
Surgeons	1

Calculation of pass/fail cut score and effect of panel variation

The cut score for our medical student candidates (n=132) was calculated using the simple formula (expert reference panel mean – 4SD), which we have previously validated in an equivalent undergraduate cohort. (Duggan and Charlin 2012) Our medical program has a non-graded pass and the potential effect on pass/fail numbers was calculated between panels for the test-retest data of the expert reference panel. In the real examination only the ERP data from Test 1 (the panels' first round attempt) counted.

Method for estimation of agreement of panel performance over time

The difference in means was estimated using the paired t-test function and rater test-retest agreement was estimated using the correlation function and the kappa statistic (Descriptive Statistics Crosstabs function), using the software IBM SPSS Version 20 for Mac. A suggested interpretation of the kappa statistic is shown in table 3.

Table 3: The interpretation of agreement measured by kappa – Altman 1991 cited by Kwiecien et al (2011).

Value of kappa	Strength of agreement
<0.2	Poor
.21-.40	Fair
.41-.60	Moderate
.61-.80	Good
.81-1.0	Very good

Results

The overall panel mean (SD) scores were 77.2 (6.7) % and 77.3 (6.5) % for Test 1 and Test 2, respectively ($p < 0.001$). The correlation between Test 1 and Test 2 was significant (Pearson correlation coefficient 0.67, $p<0.001$). The pass mark calculations from the data are shown in table 4. For the first attempt made by the reference panel members (Test 1), the pass mark set by the experienced, specialist members of the panel was 49.6% and this increased to 50.4% by addition of recent medical graduates to the panel. This difference would have had no effect on fail rates estimated from the data from the cohort of 132 medical student candidates. When the same panel members repeated the reference panel work 6 months later (Test 2), the pass mark set by the senior members of the panel was 50.7%. Addition of recent medical graduates resulted in a change in pass mark to 51.2%. In addition to the recent medical graduates, there were 3 discipline groups with sufficient numbers of panelists to undertake sub-group analysis. This table shows that the performance of the recent medical graduate group was within the range of our experienced panel members, and also the most stable.

Table 4. Paired t-test data, Pearson correlation coefficients, and pass/fail cut points applying the panel (mean - 4SD) formula for the 50-item SCT. RMG = Recent Medical Graduates; GP = General Practitioner; Int. Med = Internal Medicine Specialists; O&G = Obstetricians and Gynaecologists. “Experts” means all panel contributors except RMG’s.

		Mean %	N	Std. Deviation %	P value	Correlation	P value	Pass/Fail Cut %
Pair 1	Test1 All	77.2	63	6.69				50.4
	Test2 All	77.3	63	6.53	.944	.670	< .001	51.2
Pair 2	Test1 Experts	76.6	50	6.76				49.6
	Test2 Experts	76.5	50	6.45	.920	.640	< .001	50.7
Pair 3	Test1 RMG	79.5	13	6.10				55.1
	Test2 RMG	80.1	13	6.26	.671	.740	.004	55.0

Pair 4	Test1 GP	77.8	17	7.11				49.3
	Test2 GP	78.2	17	5.11	.794	.643	.005	57.7
Pair 5	Test1 Int. Med	79.1	16	5.08				58.7
	Test2 Int. Med	78.6	16	6.04	.713	.550	.027	54.4
Pair 6	Test1 O&G	75.2	11	6.72				48.3
	Test2 O&G	74.1	11	7.54	.612	.534	.09	43.9

Figure 2 shows the cross tabulation for the entire 63-member panel between the first (Test 1) and second (Test 2) rounds, for the selection of the range of Likert responses (A-E) in the 50-item test. The kappa value is 0.46, indicating moderate agreement between Test 1 and Test 2. Sub-group analysis showed

test-retest agreement of the recent graduate panel also to be moderate and with a kappa value of 0.54. Kappa for the 63 panel members ranged from 0.25 - 0.75. Only 7 of 63 individuals were categorised as "good" in this scale (i.e. having "good" agreement with self), with the majority in the "moderate" range.

Figure 2: Crosstab data for the 5 Likert options (A-E) selected in identical 50-item tests sat 6 months apart by the same members of our 63-member expert reference panel.

		Test 2					Total
		A	B	C	D	E	
Test 1	A	703	150	56	25	27	961
	B	152	281	134	32	28	627
	C	51	152	527	89	35	854
	D	18	43	88	200	61	410
	E	23	21	35	75	140	294
Total		947	647	840	421	291	3146

Discussion

We primarily sought to establish the effect of introducing recent medical graduates as members of our ERP's on the pass/fail decisions in our end of year clinical reasoning assessment in the 5th year of our medical program. Literature from the postgraduate domain has consistently reported a linear progression in performance in SCT's with increasing clinical experience. (Lubarsky at al 2011) Our concern in the approach to using recent (inexperienced) graduates as members of our SCT expert reference panel was the potential for significantly altering the pass fail cut score using experienced graduates as panelists.

Our data show this concern appears to be unfounded. However, it is important to appreciate that our questions were developed specifically for use in assessment of medical students, which is a different application to those previously reported by Lubarsky at al (2011). There are a number of good reasons to include recent medical graduates in this work. An important motivation for us was to reduce the burden on a relatively small group of dedicated teaching members of faculty, and in recognition that this burden disproportionately fell on our general practitioner faculty members. We felt that our recent graduates might be very well placed to participate as “multidisciplinary experts” in this context, and that there was face validity in relating the standard of performance of senior medical students directly to the standard determined from data obtained from recent medical graduates.
In our large group (63 experts), there was only moderate test-retest agreement, but due to averaging out of errors in agreement this would not not have affected the fail rate of the 5th year cohort. This emphasises that for high stakes assessments, relatively large numbers of panel members are required, as is consistent with the work of Gagnon and colleagues (2005).
An incidental finding in this research was that at the level of individual items, test-retest agreement with self was seldom good, typically fair to moderate. This is consistent with literature on the phenomenon of context specificity, in which a practitioner can see patients with the same symptoms, findings and diagnosis, yet come up with two different diagnostic conclusions. (Eva 2005) It seems likely that the uncertainty requirement in the SCT format lends itself to revealing this phenomenon.

Take Home Messages

- In the context of assessment of clinical reasoning in medical programs, recent medical graduates are suitable members of Script Concordance Test expert reference panels.
- Script Concordance Tests developed for summative purposes can place significant demands on the members of expert reference panels and the contribution of recent medical graduates should help to minimise risk of burnout of more experienced faculty.

Notes On Contributors

Paul Duggan is Head of the Discipline of Obstetrics and Gynaecology and Chair of the MBBS Assessment Committee of the University of Adelaide, and a Gynaecologist based at the Central Adelaide Local Health Network, Adelaide, Australia.
Bernard Charlin is Professor and Head of Research and Development in the Centre for Pedagogy Applied to the Health Sciences (CPASS), University of Montreal, Montreal, Canada.

Acknowledgements

We would like to thank Noor Abdul Azidah Aziz for assistance with data collection and analysis.

Bibliography/References

Ahmadi S-F, Khoshkish S, Soltani-Arabsahi K, Hafezi-Moghadam P, Zahmatkesh G, Heidari P, Baba-Beigloo D, Baradaran HR, Lot pour S. Challenging script concordance test reference standard by evidence: do judgments by emergency medicine consultants agree with likelihood ratios? International Journal of Emergency Medicine, 7:34
http://dx.doi.org/10.1186/s12245-014-0034-3
Beresford, EB. Uncertainty and the shaping of medical decisions. The Hastings Center Report, Vol. 21, No. 4 (Jul. - Aug., 1991); 6-11
http://dx.doi.org/10.2307/3562993
Boursicot K, Roberts T and Pell G. Standard Setting for Clinical Competence at Graduation from Medical School: A Comparison of Passing Scores Across Five Medical Schools. Advances in Health Sciences Education 2006 11:173–183
http://dx.doi.org/10.1007/s10459-005-5291-8
Brailovsky C, Charlin B, CoteÂ S, and Van der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. Medical Education 2001; 35:430±436
Brazeau-Lamontagne L, Charlin B, Gagnon R, Samson L and Van Der Vleuten C. Measurement of perception and interpretation skills during radiology training: utility of the script concordance approach. Medical Teacher, Vol. 26, No. 4, 2004, pp. 326–332
http://dx.doi.org/10.1080/01421590410001679000
Charlin B, Tardif J and Boshuizen H. Scripts and Medical Diagnostic Knowledge: Theory and Applications for Clinical Reasoning Instruction and Research. Acad. Med. 2000; 75:182–190
http://dx.doi.org/10.1097/00001888-200002000-00020
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 Fall; 12(4):189-9
http://dx.doi.org/10.1207/S15328015TLM1204_5
Collard A, Gelaes S, Vanbelle S, Bredart S, Defraigne J, Boniver J and Bourguignon J. Reasoning versus knowledge retention and ascertainment throughout a problem-based learning curriculum. Medical Education 2009: 43: 854–865. http://dx.doi.org/10.1111/j.1365-2923.2009.03410.x
http://dx.doi.org/10.1111/j.1365-2923.2009.03410.x
Duggan P. Development of a Script Concordance Test using an electronic voting system. Ergo 1, 1 December 2007; 35-41
Duggan, Paul and Charlin, Bernard. Summative assessment of 5th year medical students' clinical reasoning by script concordance test: requirements and challenges. BMC Med Ed 2012 12:29
Duggan Paul, Monnier Patricia, Roex Alphonse, Bedard Maree Josee, Charlin Bernard. Bi-cultural bi-national benchmarking and assessment of clinical reasoning in Obstetrics and Gynaecology. MedEdPublish 2016
http://dx.doi.org/10.15694/mep.2016.000025
Eva Kevin W. What every teacher needs to know about clinical reasoning. Medical Education 2005. 39; 98-106
http://dx.doi.org/10.1111/j.1365-2929.2004.01972.x
Durning, Steven J. Artino, Anthony R. Jr., Schuwirth, Lambert and van der Vleuten, Cees. Clarifying Assumptions to Enhance Our Understanding and Assessment of Clinical Reasoning. Academic Medicine 2013, 88 (4):1-7

http://dx.doi.org/10.1097/ACM.0b013e3182851b5b
Gagnon R, Charlin B, Coletti M, Sauve E and 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? Medical Education 2005; 39: 284–291
http://dx.doi.org/10.1111/j.1365-2929.2005.02092.x
Hornos EH, Pleguezuelos EM, Brailovsky CA, Harillo LD, Dory V, Charlin B. The practicum script concordance test: an online continuing professional development format to foster re action on clinical practice. J Contin Educ Health Prof. 2013 Winter; 33(1):59-66
http://dx.doi.org/10.1002/chp.21166
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 Feb 9; 4:7
http://dx.doi.org/10.1186/1748-717X-4-7
Lubarsky S, Chalk C, Kazitani D, Gagnon R, Charlin B. The script concordance test: a new tool and assessing clinical judgment in neurology. Can J Neurol Sci. 2009; 36:326-31
http://dx.doi.org/10.1017/S031716710000706X
Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CPM. Script Concordance Method: A Review of Published Validity Evidence. Medical Education 2011; 45(4):329-38
http://dx.doi.org/10.1111/j.1365-2923.2010.03863.x
Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. (2013). Script concordance testing: From theory to practice: AMEE guide No. 75. Med Teach, 35(3): 184-93
http://dx.doi.org/10.3109/0142159X.2013.760036
Meterissian S. A Novel Method of Assessing Clinical Reasoning in Surgical Residents. Surg Innov 2006; 13; 115. http://dx.doi.org/10.1177/1553350606291042
http://dx.doi.org/10.1177/1553350606291042
Monnier P, Bédard M-J, Gagnon R, Charlin B. The relationship between script concordance test scores in an obstetrics-gynecology rotation and global performance assessments in the Curriculum. International Journal of Medical Education. 2011; 2:3-6
http://dx.doi.org/10.5116/ijme.4d21.bf89
Park AJ, Barber MD, Bent AE et al. Assessment of intraoperative judgment during gynecological surgery using the Script Concordance Test. Am J Obstet Gynecol 2010;203:240.e1-6
http://dx.doi.org/10.1016/j.ajog.2010.04.010
Saber Tehrani AS, Lee H, Mathews SC, Shore A, Makary MA, Pronovost PJ, Newman-Toker DE. 25-Year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National Practitioner Data Bank. BMJ Qual Saf. 2013 Aug; 22(8):672-80. doi: 10.1136/bmjqs-2012- 001550. Epub 2013 Apr 22. http://dx.doi.org/10.1136/bmjqs-2012-001550

Appendices

Declaration of Interest

The author has declared that there are no conflicts of interest.