

Script concordance test to assess diagnostic and management reasoning in acute medicine

Dear Editor,

Clinical reasoning, an essential skill for patient care, can be difficult to assess. We created and validated a script concordance test (SCT) to assess clinical reasoning in acute medicine. This tool was used to provide feedback and targeted remediation for Postgraduate-Year-1 (PGY1) doctors, guide teaching and learning, and facilitate programme evaluation.

The SCT is a validated method for assessing the interpretation of medical data, hypothesis evaluation, and clinical judgement under conditions of uncertainty.¹ An SCT presents brief clinical scenarios and asks examinees to evaluate diagnostic or management options while considering new information (Supplementary Fig. S1). Questions have a degree of uncertainty with no single correct answer. Responses are compared with a scoring key derived from a reference panel of experts.²

Clinical reasoning research has largely focused on diagnostic reasoning. With increasing recognition that management reasoning is a distinct skill³ that is less well understood, calls have been made for more studies in the area.⁴ We thus studied differences between diagnostic and management reasoning across internal medicine (IM) doctors of varying seniority. We obtained Institutional Review Board approval (SingHealth CIRB 2020/2591).

We developed an Acute Internal Medicine SCT (AIM-SCT) blueprinted for common acute medical conditions. From 50 clinical scenarios, item analysis was performed to discard overly concordant (i.e. poor discriminant value) or discordant (i.e. measurement error) questions. Final question selection ensured representation across all medical specialties, in accordance with the blueprint. Questions were classified as “diagnostic reasoning” or “management reasoning” by three of our authors independently; there was 100% agreement. The final AIM-SCT consisted of 73 items (26 scenarios), with 37 items (13 scenarios) testing diagnostic reasoning and 36 items (13 scenarios) testing management reasoning.

We used published processes⁵ to derive the scoring key. The reference panel (20 senior residents, equivalent to specialty registrars or fellows) was selected based on the following criteria: (1) currency in on-site senior IM call duties within the past four months, (2) currency in intensive care medicine within the past four months, (3) completed formal IM training, and (4) relevant

postgraduate qualifications (e.g. American Board of Internal Medicine and Membership of the Royal College of Physicians). All major medical specialties were represented.

PGY1s, medical officers (postgraduate years 2–5) and senior residents completed the SCT asynchronously using a web response form with a 75-minute time limit. Responses were scored against the scoring key and expressed as a percentage of the total score.

Statistical analysis was conducted with SPSS Statistics version 28.0. (IBM Corp, Armonk, NY, US). Test reliability was assessed using Cronbach's alpha. SCT scores between seniority levels were compared using one-way analysis of variance (ANOVA), followed by pairwise comparisons with Bonferroni correction. Two independent sample t-tests were conducted to compare scores between PGY1s in their first posting (i.e. first 4 months of PGY1 training) and PGY1s in their second or third postings (i.e. latter 8 months of PGY1 training). Diagnostic and management reasoning scores were compared within each seniority level using Wilcoxon signed-rank test. Correlations between diagnostic and management reasoning scores were explored through Pearson correlation. Significance was set at $P < 0.05$.

Thirty-three PGY1s, 22 medical officers and 20 senior residents completed the SCT. All participants completed the test asynchronously using a web response form within a 75-minute time limit. Test reliability was high (Cronbach's $\alpha = 0.85$); all items showed item-total correlations of > 0.05 . Total SCT scores increased with seniority; mean percentage scores (standard deviation) increased from 60.3 (8.9) in PGY1s to 70.4 (7.5) in medical officers, and 78.2 (4.5) in senior residents (all pairwise $P < 0.01$; Table 1). First-posting PGY1s (those in their first 4 months of PGY1 training, $n = 10$) scored significantly lower than second and third-posting PGY1s (those in their latter 8 months of PGY1 training, $n = 23$), with mean percentage (standard deviation) scores: 54.3 (8.7) versus 62.9 (7.8), $P = 0.015$.

PGY1s scored significantly higher in diagnostic reasoning 62.0 (8.9) than in management reasoning 58.6 (10.4), mean score difference 3.34, 95% confidence interval (CI) 0.707–5.99, $P = 0.024$. No significant differences between diagnostic and management reasoning scores were found for medical officers and senior residents (Table 1). Overall, combining all

Table 1. Comparison of total, diagnostic and management reasoning scores.

	PGY1 (n=33)	Medical Officer (MO) (n=22)	Senior Resident (SR) (n=20)	<i>P</i> value ^a	Pairwise <i>P</i> value ^b
Total SCT scores					
Mean (SD)	60.3 (8.9)	70.4 (7.5)	78.2 (4.5)	<0.001	PGY1 vs MO: <0.001
Median (Q1, Q3)	62.8 (50.9, 68.0)	71.4 (65.6, 75.6)	77.8 (75.2, 82.3)		PGY1 vs SR: <0.001 MO vs SR: 0.004
Diagnostic versus management reasoning scores					
Diagnostic score					
Mean (SD)	62.0 (8.9)	70.9 (8.6)	78.6 (7.3)	<0.001	PGY1 vs MO <0.001
Median (Q1, Q3)	63.4 (53.3, 70.0)	71.0 (63.6, 77.4)	76.4 (72.7, 85.9)		PGY1 vs SR <0.001 MO vs SR 0.013
Management scores					
Mean (SD)	58.6 (10.4)	70.0 (8.5)	77.7 (6.3)	<0.001	PGY1 vs MO <0.001
Median (Q1, Q3)	57.0 (48.6, 67.6)	72.4 (63.5, 77.5)	77.7 (72.4, 83.6)		PGY1 vs SR <0.001 MO vs SR 0.018
Score difference (diagnostic minus management)					
Mean (SD)	3.34 (1.30)	1.00 (1.80)	0.87 (2.32)	N/A	N/A
(95% CI of Mean)	(0.71, 5.99)	(-2.74, 4.75)	(-4.00, 5.73)		
Median (Q1, Q3)	4.74 (-3.17, 4.74)	2.59 (-5.19, 2.59)	0.56 (-7.91, 10.1)		
<i>P</i> value for score difference^c	0.024	0.322	0.681		

PGY1: Postgraduate-Year-1; SCT: script concordance test; SD: standard deviation

^a Comparing the 3 groups using 1-way ANOVA.

^b Group pairwise comparisons using 2 independent sample t-test.

^c Using the Wilcoxon signed-rank test.

seniority levels, there was a moderate correlation between diagnostic and management reasoning scores ($r=0.688$, $P<0.001$).

Our AIM-SCT demonstrates construct validity with good discrimination between seniority levels among IM doctors. It is a feasible tool to assess clinical reasoning using a resource-lean online examination without direct examiner involvement. To the best of our knowledge, this is the first SCT developed in acute IM.

Our SCT is the first, as far as we know, to explore differences in diagnostic and management reasoning performance.⁴ We postulate several explanations for our finding that PGY1s, but not more senior clinicians, performed better in diagnostic than management reasoning. First, management reasoning may be more complex than diagnostic reasoning, requiring the weighing of testing and treatment thresholds, consideration of the value of care, and management of uncertainty,⁶ and thus may take more time and experience to develop. Our finding may support the hypothesis of a separate management script.^{6,7} Second, management reasoning requires shared decision-making and follow-up⁴—skills that may not develop until later in postgraduate training

as junior doctors learn at the workplace via situated learning.⁸ Third, undergraduate medical curricula may prioritise diagnostic over management reasoning. Singapore's national outcomes framework for medical graduates includes a list of conditions for which both diagnostic and management competencies are required, and a separate list of conditions where graduates are only expected to diagnose but not manage.⁹

Our study has several limitations. First, all reference panel members come from a single academic medical centre. A prior study found local experience to be associated with improved SCT performance,¹⁰ suggesting that clinical reasoning may be context-specific and not immediately transferrable to other settings. Second, SCTs do not test features of management reasoning such as the dynamic interplay between people, systems and priorities, as well as communication and shared decision-making,^{4,6} which are perhaps better assessed with workplace-based assessments.

In conclusion, our AIM-SCT is a reliable, valid and feasible tool to assess clinical reasoning in acute medicine. Plans are underway to incorporate our SCT into the formative assessment of PGY1s to provide

individualised feedback and plan targeted interventions. Finally, our finding that management reasoning may develop differently from diagnostic reasoning requires empirical validation, and a better understanding of how management reasoning develops will be valuable in its teaching and assessment.

Disclosure

No author has any affiliation or financial involvement with any commercial organisation with a direct financial interest in this manuscript.

REFERENCES

1. Lubarsky S, Charlin B, Cook DA, et al. Script concordance testing: a review of published validity evidence: Validity evidence for script concordance tests. *Med Educ* 2011;45:329-38.
2. Dory V, Gagnon R, Vanpee D, et al. How to construct and implement script concordance tests: insights from a systematic review: Construction and implementation of script concordance tests. *Med Educ* 2012;46:552-63.
3. Cook DA, Sherbino J, Durning SJ. Management Reasoning: Beyond the Diagnosis. *JAMA* 2018;319:2267-8.
4. Cook DA, Stephenson CR, Gruppen LD, et al. Management Reasoning: Empirical Determination of Key Features and a Conceptual Model. *Acad Med* 2023;98:80-7.
5. Lubarsky S, Dory V, Duggan P, et al. Script concordance testing: From theory to practice: AMEE Guide No. 75. *Med Teach* 2013; 35:184-93.
6. Wijesekera TP, Parsons AS, Abdoler EA, et al. Management Reasoning: A Toolbox for Educators. *Acad Med* 2022;97:1724.
7. Cook DA, Stephenson CR, Gruppen LD, et al. Management reasoning scripts: Qualitative exploration using simulated physician-patient encounters. *Perspect Med Educ* 2022;11:196-206.
8. Abdoler EA, Parsons AS, Wijesekera TP. The future of teaching management reasoning: important questions and potential solutions. *Diagnosis* 2022;10:19-23.
9. Ministry of Health, Singapore. National Standards for Basic Medical Education, National Outcomes Framework for Medical Graduates, and National Quality Improvement and Quality Assurance Framework. MOH Training Circular No. 16/2018. Annex B, pp. 29-32. Singapore: Ministry of Health.
10. Loh KW, Rotgans JI, Tan K, et al. Predictors of clinical reasoning in neurological localisation: A study in internal medicine residents. *Asia Pac Sch* 2020;5:54-61.

Jie Ming Nigel Fong ¹*MRCP (UK)*,
 Rebecca Hui Min Hoe ²*MRCP (UK)*,
 Dorothy Hannah Huang ¹*MRCP (UK)*,
 Jianlin Cheney Wong ³*MRCP (UK)*,
 Xiang Lee Jamie Kee ⁴*MRCP (UK)*,
 Kok Long Albert Teng ³*MRCP (UK)*,
 Rilong Hong ⁵*MRCP (UK)*, Seyed Ehsan Saffari ^{2,6}*PhD*,
 Kevin Tan ^{*2}*MS-HPEd*, Nigel Choon Kiat Tan ^{*2}*FAMS*

¹ Department of Renal Medicine, Singapore General Hospital, Singapore

² Department of Neurology, National Neuroscience Institute, Singapore

³ Department of Internal Medicine, Singapore General Hospital, Singapore

⁴ Department of Geriatric Medicine, Singapore General Hospital, Singapore

⁵ Department of Cardiology, National Heart Centre Singapore

⁶ Health Services & Systems Research, Duke-NUS Medical School, National University of Singapore, Singapore

Correspondence: Dr Nigel Choon Kiat Tan, 11 Jln Tan Tock Seng, Singapore 308433.

Email: nigel.tan.c.k@singhealth.com.sg

* Joint last authors