2019

Commentary: Expert responses in script concordance tests: A response process validity investigation

Siu Hong Wan
The University of Notre Dame Australia, michael.wan@nd.edu.au

Elina Tor
The University of Notre Dame Australia, elina.tor@nd.edu.au

Judith N. Hudson

This other contribution to a refereed journal was originally published as:
Wan, S. H., Tor, E., & Hudson, J. N. (2019). Commentary: Expert responses in script concordance tests: A response process validity investigation. *Medical Education, 53* (7), 644-646.

Original other contribution to a refereed journal available here:
https://doi.org/10.1111/medu.13889

This article first published in *Medical Education*:

Wan, S.H., Tor, E., and Hudson, J.N. (2019). Commentary: Expert responses in script concordance tests: A response process validity investigation. *Medical Education, 2019*. doi: 10.1111/medu.13889

This article has been published in final form at: -

https://onlinelibrary.wiley.com/doi/full/10.1111/medu.13889

Commentary: Expert responses in script concordance tests: A response process validity investigation

Siu Hong WAN[1], Elina Tor[1], Judith N Hudson[2]
[1]School of Medicine, University of Notre Dame. Australia
[2]Faculty of Health and Medical Sciences, University of Adelaide. Australia.

Email: michael.wan@nd.edu.au

There is substantial evidence that clinical decision making and medical problem solving by doctors depend to a large degree on probabilistic logic (1) and/or typicality of patient information with reference to doctors' activated illness scripts (2). The Script Concordance Test (SCT) is a written assessment format designed specifically to assess individuals' performance on probability related clinical information processing tasks. It presents candidates with a clinical scenario and requires them to consider a new piece of clinical information to determine the extent to which that information alters the probability of a particular diagnosis or appropriateness of a particular investigation or action.

The SCT is built upon sound conceptual and theoretical underpinnings (3). A number of studies have explored the validity of SCT score interpretation, mostly comprising systematic gathering and documenting of evidence that SCT scores are indeed indicative of the soundness of candidates' clinical judgement (4-6). The latest research using a 'de-constructed' approach to validation of SCT scores is a very positive trend that is helping shed some light on the many grey areas surrounding the validity of SCT scores (6-8).

One such grey area derives from the fact that there is still limited study of response process validity (whether or not the responses of test takers suggest they share the same conception of the construct being measured as do the assessors). This is true for assessment generally, but is a particularly important issue for SCT designers, because the very rationale for SCT use is based on the assumption that candidate answers reflect the cognitive operations involved in integrating newly presented patient information into existing medical knowledge structures to generate updated probabilities of a particular outcome.

Lineberry and colleagues' study (9), published in this issue, is an effort in this direction. The authors explored the response processes of experts to understand their divergent beliefs about how new clinical data alters the suitability of proposed actions and how they reacted to other experts' perspectives. Their study elicited varieties of expert responses other than those intended, providing evidence of construct irrelevant variance in the experts' response process (10). These findings corroborate recent literature outlining plausible validity threats for SCT score interpretation (11). In particular, empirical data from this study highlighted that typical SCT formats in which post-data belief changes by experts are interpreted without considering experts' pre-data belief runs the risk of masking underlying agreement/disagreement between experts. Other significant findings reported include (a) experts' disagreement with the proposed action in SCT items, raising concerns about the credibility or content accuracy of SCT items; and, (b) instability of experts' responses, indicating a threat to the test-retest reliability of SCT scores. The authors discuss the challenge of balancing the tension between maintaining authenticity in reflecting "uncertainty" in clinical decision making and, at the same time, ensuring content accuracy in SCT items.

In collecting these data, Lineberry et al. acknowledge their SCT cases were adapted from real patient histories, with rich details and findings. As such, they may have diverged from the usual SCT guidance to be "brief" and "ill-defined" (3, 12, 13). This could be a critically important feature determining the extent to which SCTs are implemented in a way that yields valid scores. In our experience constructing SCT items, using simple and ill-defined case scenarios to test core concepts in clinical reasoning in medicine, only 20–30% of SCT items have generally been discarded or modified because of discordance in response pattern amongst clinicians (i.e. extreme inconsistencies among experts) (14).

This issue is not raised in an effort to facilely dismiss Lineberry et al.'s results given that is purely a speculative hypothesis at the moment. Rather, it is mentioned because Lineberry et al.'s findings remind us of the broader issue that it is important to remain aware of the fact that all assessment tools, regardless of how well grounded they are on sound theoretical underpinning and empirical data, may demonstrate unintended issues on any given administration. Validity must be established repeatedly with adequate evidence collected on each administration and deliberate exploration of what might be causing problems when unexpected findings arise. Evidence supporting the use of test scores should be documented over time, from multiple sources, consistent with the contemporary conception of validity as a unitary construct (15-17). That is, validation should be an ongoing process forming part of the fabric of all assessment initiatives (18), but particularly in the context of high stakes summative assessments of learning.

At the structural level, recent calls for a move towards a programmatic perspective on assessing competence is a paradigm shift in the right direction towards a more sustainable and constructive landscape in medical education. This more continuous form of assessment makes it all the more imperative that we adopt a continuous form of validation practices. The rich information that SCTs can provide, as discussed by Lineberry et al., can be optimised for learning and be meaningfully aggregated to inform progress decisions for trainees, but only if care is put into ensuring that the scores reflect what the theory intends (19-21). The post-scoring de-brief and debate by the expert reference panel, used by Lineberry et al. in this paper, provides an excellent example of a counter-measure to be used against validity threats that could simultaneously serve as a useful continuing professional development activity for clinicians, test developers, and educators alike. Engaging in such activity may turn controversial SCT cases into valuable stimuli for learning, hence achieving and role-modelling the goal of authentically reflecting the complexity of medical decision making (21, 22).

In sum, while Lineberry et al.'s findings might be considered a negative mark on the validity evidence for SCT use, we argue that the authors' research approach more constructively provides a strategy for enabling a longitudinal exploration of the development of clinical reasoning in both learners and experts. Panels comprised of undergraduate or postgraduate learners can learn from reflection on proposed actions (pre and post) as well as the responses of peers and experts. This could provide valuable understanding of what stimulates clinical reasoning ability in learners during their professional development and what maintains or furthers clinical reasoning ability in those who are more well established.

References:

1. Kreiter CD. A Bayesian perspective on constructing a written assessment of probabilistic clinical reasoning in experienced clinicians: Assessing clinical reasoning. Journal of Evaluation in Clinical Practice. 2017;23(1):44-8.
2. Lubarsky S, Gagnon R, Charlin B. Script concordance test item response process: The argument for probability versus typicality. Advances in Health Sciences Education. 2012;17(1):11-3.
3. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance Test: A Tool to Assess the Reflective Clinician. Teaching and Learning in Medicine. 2000;12(4):189-95.
4. Lubarsky S, Vleuten CPMvd, Charlin B, Chalk C, Cook DA. Script concordance testing: a review of published validity evidence. Medical Education. 2011;45(4):329-38.
5. Subra J, Chicoulaa B, Stillmunkes A, Mesthe P, Oustric S, Bugat MER. Reliability and validity of the script concordance test for postgraduate students of general practice. EUROPEAN JOURNAL OF GENERAL PRACTICE. 2017;23(1):209-14.
6. Wan MS, Tor E, Hudson JN. Improving the validity of script concordance testing by optimising and balancing items. Medical education. 2018;52(3):336-46.
7. Lubarsky S, Dory V, Meterissian S, Lambert C, Gagnon R. Examining the effects of gaming and guessing on script concordance test scores. Perspectives on Medical Education. 2018;7(3):174-81.
8. Wan SH, Duggan P, Tor E, Hudson JN. Association between candidate total scores and response pattern in script concordance testing of medical students. Focus on Health Professional Education: A Multi-disciplinary Journal. 2017;18(2):26-35.
9. Lineberry M, Hornos E, Pleguezuelos E, Mella J, Brailovsky C, Bordage G. Experts' responses in script concordance tests: a response process validity investigation. Medical education. 2019.
10. Whitely SE. Construct validity: Construct representation versus nomothetic span. Psychological Bulletin. 1983;93(1):179-97.
11. Lineberry M, Kreiter CD, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. Medical Education. 2013;47(12):1175-83.
12. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. Medical Education. 2012;46(6):552-63.
13. Duggan P, Charlin B. Summative assessment of 5th year medical students' clinical reasoning by script concordance test: requirements and challenges. BMC MEDICAL EDUCATION. 2012;12(1):29-.
14. Wan SH. Using the script concordance test to assess clinical reasoning skills in undergraduate and postgraduate medicine. HONG KONG MEDICAL JOURNAL. 2015;21(5):455-61.
15. Crooks TJ, Kane MT, Cohen AS. Threats to the Valid Use of Assessments. Assessment in Education: Principles, Policy & Practice. 1996;3(3):265-86.
16. Kane MT. Validating the Interpretations and Uses of Test Scores. Journal of Educational Measurement. 2013;50(1):1-73.
17. Messick S. Meaning and Values in Test Validation: The Science and Ethics of Assessment. Educational Researcher. 1989;18(2):5-11.
18. Tor E, Steketee C, Mak D. Clinical audit project in undergraduate medical education curriculum: an assessment validation study. International journal of medical education. 2016;7:309-19.

19. Heeneman S, Pool AO, Schuwirth LWT, van der Vleuten CPM, essen EW. The impact of programmatic assessment on student learning: theory versus practice. Medical Education. 2015;49(5):487-98.
20. Schuwirth LWT, van der Vleuten CPM. Programmatic assessment and Kane's validity perspective. Medical Education. 2012;46(1):38-48.
21. Van Der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. Medical Education. 2005;39(3):309-17.
22. Power A, Lemay J-F, Cooke S. Justify Your Answer: The Role of Written Think Aloud in Script Concordance Testing. Teaching and Learning in Medicine. 2017;29(1):59-67.