Adv in Health Sci Educ (2012) 17:11–13
DOI 10.1007/s10459-011-9338-8

RESPONSE

Script concordance test item response process: The argument for probability versus typicality

Stuart Lubarsky · Robert Gagnon · Bernard Charlin

Received: 18 November 2011 / Accepted: 18 November 2011 / Published online: 29 November 2011

Kreiter's (2011) incisive commentary cuts to the quick of an unresolved issue pertaining to script concordance testing: what are examinees *thinking* as they work through the test items? One possibility, advanced by Kreiter, is that SCT examinees rely on Bayesian reasoning, whereby they engage in a series of probability assignations and computations. This account has examinees estimating the probability of the hypothesis provided in the first column (**P1**, per Kreiter), followed by the probability of the same hypothesis given the new piece of clinical information provided in the second column (**P2**). They are then presumed to calculate the difference between **P2** and **P1**, and to translate the result into an appropriate response (from −2 to +2) on the Likert scale provided in column 3.

This depiction of the thought and response processes of SCT examinees is unlikely to be accurate. Even experienced practitioners are notoriously inept at framing clinical problems in Bayesian terms. In a classic study, Eddy (1982) found that the vast majority of physicians made errors solving a problem in which probabilistic reasoning was required to determine a screened patient's risk of developing breast cancer. If SCT demanded pure Bayesian analysis, one would not expect for SCT performance to improve with experience and expertise, as has been consistently observed in participants tested across various medical disciplines.

Furthermore, carefully-constructed SCTs obviate the need for the type of probabilistic reasoning postulated by Kreiter (2011). Test-makers are deliberately instructed to avoid proposing hypotheses in column 1 that are not credible or relevant to a given vignette, i.e. ones with low *a priori* probability (Fournier et al. 2008). (If a hypothesis in the "If you were thinking…" column generates a reaction akin to "What? I would never have considered *that*!" in reasonable examinees, then that item probably should be considered unfair and discarded). Hypotheses with overly high pre-test probabilities are also meant to be eschewed, since they do not infuse SCT items with the requisite condition of

S. Lubarsky (✉)
Centre for Medical Education, Faculty of Medicine, McGill University, Montreal, QC, Canada
e-mail: stuart.lubarsky@mcgill.ca

R. Gagnon · B. Charlin
Centre de pédagogie appliquée aux sciences de la santé (CPASS), Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

uncertainty. Since the prior probability of SCT hypotheses is calibrated, by design, to a set upper-range level (at a value of, say, 0.7 or so), the Bayesian examinee would quickly realize that estimating and subtracting **P1** for each item is pointless.

There is an alternative account of the SCT reasoning process which does not assume that examinees engage in probability calculations and Bayesian inferences. By this account, script concordance solicits judgments about *typicality*, not probability. Typicality and probability are conceptually quite different animals. Consider the following example, adapted from work by Tversky and Kahnemann (1982):

> "A 66-year old man has been diagnosed with bacterial pneumonia." Please indicate the most likely alternative:
> [a] he is coughing
> [b] he is coughing and has fever

The knee-jerk reaction of most physicians is to respond [b], since the typical patient with pneumonia is febrile and coughs. However, from a probabilistic standpoint, [a] is always the more likely answer, since the co-occurrence of two likely events cannot be *more* likely than the probability of either event alone (Tversky and Kahnemann 1982). Experts are certainly better than novices at deciding whether information is typical or representative of a given hypothesis, but, as we have seen, they do not necessarily outperform novices in tests involving probability estimates.

The argument for typicality contends that the central challenge of SCT is for examinees to judge how well newly discovered data coheres with a plausible starting premise about a case. According to the typicality theory, the clinical scenario (**S1**) and given hypothesis (**S2**) trigger the mobilization of a relevant illness script from an examinee's mental database. The examinee's task is then to determine, in a single cognitive step that does not involve a calculation of **P1**, the extent to which a new piece of clinical information (**S3**) is (or is not) typical of, consistent with, or appropriate to the activated script. (The test's developers refer to this specialized task as 'clinical data interpretation.') Script concordance hinges on an inference that examinees with more evolved illness scripts interpret data and make judgments that increasingly concord with those of experienced practitioners given the same clinical scenarios (Charlin et al. 1998).

Some prior work supports the claim that the response process of SCT examinees includes an estimation of 'fit' between clinical data and activated scripts (Gagnon et al. 2006). For example, in one study that exploited the script concordance format, subjects were asked to gauge the impact of new pieces of information on a series of diagnostic hypotheses. Subjects' response times were significantly faster when they were provided clinical information that was either typical or incompatible with the given hypothesis than when they were presented information that was atypical. Subjects also responded more accurately when presented typical than atypical information. Processing time and accuracy of data interpretation on script concordance tasks, then, seem to be influenced by the degree of compatibility between new clinical information and relevant activated scripts, as the typicality argument would predict.

Ultimately, Kreiter (2011) is correct in reporting that a clear relationship between the purported construct of the SCT (clinical data interpretation) and the response process of examinees is not, as yet, rooted in firm empirical evidence (Lubarsky et al. 2011). He then lays the groundwork for further research with a testable hypothesis: '[W]hen **P1** is close to 1.0 and **S3** adds little or no additional information, the respondent must decide between using the scale to indicate [that] the diagnosis is almost certain or to indicate that **S3** adds no useful information.' The latter conclusion (in which $\mathbf{P2} - \mathbf{P1} = 0$) would be the predicted response of a Bayesian reasoner; the former conclusion would be expected from an examinee evaluating the typicality of the new piece of information with respect to the initial hypothesis. Finally, ‘think-aloud’ or concept mapping protocols might also help to shed further light on examinees’ use of probability- versus typicality-based reasoning strategies in responding to SCT items.

References

Charlin, B., Brailovsky, C. A., Leduc, C., & Blouin, D. (1998). The diagnostic script questionnaire: A new tool to assess a specific dimension of clinical competence. *Advances in Health Sciences Education, 3*, 51–58.

Eddy, D. (1982). Probabilistic reasoning in clinical medicine: Problems, opportunities. In D. Kahnemann, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Fournier, J. P., Demeester, A., & Charlin, B. (2008). Script concordance tests: Guidelines for construction. *BMC Medical Informatics and Decision Making, 8*, 18.

Gagnon, R., Charlin, B., Roy, L., St-Martin, M., Sauve, E., Boshuizen, H. P. A., et al. (2006). The cognitive validity of the script concordance test: A time processing study. *Teaching and Learning in Medicine, 18*(1), 22–27.

Kreiter, C. (2011). Commentary: The response process validity of a script concordance item. *Advances in health sciences education: Theory and practice.* 2011 Oct 1 [Epub ahead of print].

Lubarsky, S., Charlin, B., Cook, D. A., Chalk, C., & van der Vleuten, C. P. M. (2011). Script concordance testing: A review of published validity evidence. *Medical Education, 45*, 329–338.

Tversky, A., & Kahnemann, D. (1982). Judgments of and by representativeness. In D. Kahnemann, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.