Adv in Health Sci Educ (2012) 17:7–9
DOI 10.1007/s10459-011-9325-0

Commentary: The response process validity of a script concordance test item

Clarence D. Kreiter

Received: 27 July 2011 / Accepted: 6 September 2011 / Published online: 1 October 2011

While the Script Concordance Test (SCT) item format has generated considerable research that has reflected positively on its reliability (Charlin et al. 1998) and ease of construction (Fournier et al. 2008), a recent review suggests the response process validity of the SCT is weakly supported (Lubarsky et al. 2011). The Gagnon et al. (2011) paper published in this issue of *Advances in Health Sciences Education* suggests that the validity of an SCT score as a measure of clinical reasoning hinges on the degree to which aggregate scoring conveys a concordance between experts' and examinees' scripts. However, the assumption that an examinee's solution process reflects their clinical scripts has not been confirmed. Since the validity of an SCT score derives from the congruence between current score interpretations and a logical analysis of the task presented by the item, a sound validity argument must consider what the SCT task is actually asking the examinee to do. While it is obviously important to consider the underlying cognitive processes used to respond to the item, an even more fundamental prerequisite is to logically decompose the SCT task into its component parts. By breaking down the SCT problem into its logical steps, it will be easier to understand the cognitive demands of an SCT item.

Figure 1 presents a flow diagram of the four task stimuli **(S1–S4)** and the three calculation steps **(P1, P2, {P2–P1})** required by an SCT-based question. Since Gagnon et al. (2011) already present an excellent example and description of the SCT item format, here it will be sufficient to simply note that an SCT first presents a clinical scenario **(S1)** to the examinee and then suggests a hypothesis **(S2)**. Next, the item presents a single discrete new piece of information **(S3),** and then asks how that new clinical information impacts the probability of the suggested hypothesis **(S4)**. As Fig. 1 indicates, after both **S2** and **S3** the examinee is implicitly asked to make a probabilistic calculation **(P1 & P2 respectively)** in order to answer the formal question **(S4).** The first probability **(P1)** represents the likelihood of the hypothesis **(S2)** given the clinical scenario **(S1)**. The second probability **(P2)** is the likelihood of the hypothesis **(S2)** given both the scenario **(S1)** and the new piece of

C. D. Kreiter (✉)
University of Iowa, Iowa City, IA, USA
e-mail: clarence-kreiter@uiowa.edu

$$\mathbf{S1} \rightarrow \mathbf{S2} \rightarrow \mathbf{P1} \rightarrow \mathbf{S3} \rightarrow \mathbf{P2} \rightarrow \mathbf{S4} \rightarrow \{\mathbf{P2} - \mathbf{P1}\}$$

Fig. 1

information **(S3)**. The final question **(S4)** explicitly asks examinees to characterize the difference between **P1** and **P2** on a Likert-type scale.

One can best represent this sort of problem in terms of a Bayesian equation like that shown in Eq. 1.

$$P(D|T) = P(T|D) * P(D)/P(T) \quad (1)$$

Using as an example an SCT item that asks about a diagnostic hypothesis, Equation 1 can be used to calculate the probability that a patient has a particular disease (D) given a certain test result (T). Equation 1 describes the probability that a patient has the disease given a positive test result [P(D|T)] **(P2)**, and equals the probability of a positive test result given the patient has the disease [P(T|D)—*sensitivity*] multiplied by the probability of the disease given the scenario [P(D)—*prior or subjective probability*] **(P1)** divided by the overall probability of a positive test result in the population referenced in the scenario [P(T)]. The explicitly asked question in the SCT item **(S4)** is really about the relationship between **P1** and **P2** [P(D) and P(D|T)] which can formally be expressed as the difference **{P2–P1}.** The SCT item simply requests that the examinee to subjectively rate the magnitude of this difference on a 5-point Likert scale.

Equation 1 has proven to be useful method for representing evidence-based medical decision problems. While it is common for physicians to deal with considerable uncertainty related to the subjective or prior probability P(D) **(P1)**, and hence generate differing values for P(D|T) **(P2),** there are few clinically meaningful instances in medicine where there is substantial disagreement regarding the direction of impact made by a single piece of medical evidence **(S3)**. While the 'true' magnitude of P(D|T) **(P2)** may be uncertain, **S3's** direction of impact on **P1** is seldom debatable. In fact, if there is medical evidence about which there is disagreement regarding whether it contributes positively or negatively towards making a particular diagnosis, it hardly seems reasonable to infer that it is a fundamental aspect of reasoning. In addition, the Likert-type response scale does not allow a clear logical response for many medically-related problems. For example, when **P1** is close to 1.0 and **S3** adds little or no additional information, the respondent must decide between using the scale to indicate the diagnosis is almost certain or to indicate that **S3** adds no useful information.

While it is clear that the direction of impact is the question of interest for an SCT item, this insight does not imply that the SCT is a failed format. Rather, it appears we need to reconsider the script-based interpretation and the 'no correct answer' Likert-type response format. Employing heuristics consistent with Bayes theorem is important in defining sound medical reasoning and may also offer a method for scoring SCT-like items (Wolf et al. 1985). If one looks closely at the SCT question, it is not so different from other item formats commonly used in medical education, and in fact SCT questions can usually be rephrased as True/False propositional statements (Bland et al. 2005). The special advantage of the SCT question format does not reside in its script-based measurement capabilities, or in not having a correct answer, for it usually does, but rather in its ability to easily ask clinical questions that are useful for assessing an examinee's clinical knowledge structure.

References

Bland, A. C., Kreiter, C. D., & Gordon, J. A. (2005). The psychometric properties of five scoring methods applied to the script concordance test. *Academic Medicine, 80*(4), 395–399.

Charlin, B., Brailovsky, C. A., Ledus, 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.

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

Gagnon, R., Lubarsky, S., Lambert, C., & Charlin, B. (2011). Optimization of answer keys for script concordance testing: Should we exclude deviant panelists, deviant responses, or neither? *Advances in Health Sciences Education*. doi:10.1007/s10459-011-9279-2.

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.

Wolf, F. M., Gruppen, L. D., & Billi, J. E. (1985). Differential diagnosis and the competing-hypotheses heuristic. *Journal of the American Medical Association, 253*(19), 2858–2862.