

# The Practicum Script Concordance Test: An Online Continuing Professional Development Format to Foster Reflection on Clinical Practice

EDUARDO H. HORNOS, MD; EDUARDO M. PLEGUEZUELOS, MD; CARLOS A. BRAILOVSKY, MD, MA Ed; LEANDRO D. HARILLO, ENG; VALÉRIE DORY, MD, PHD; BERNARD CHARLIN, MD, PHD

**Introduction:** Judgment in the face of uncertainty is an important dimension of expertise and clinical competence. However, it is challenging to conceive continuing professional development (CPD) initiatives aimed at helping physicians enhance their clinical judgment skills in ill-defined situations. We present an online script concordance-based CPD program (the Practicum Script Concordance Test, copyright ©2006 by Practicum Foundation), a tool that can be used to support health professionals in the development of their reflective clinical reasoning ability. We describe the rationale and principles and report on the implementation of 2 online programs based on this new CPD initiative.

**Method:** The Practicum Script Concordance Test program consists of daily testing and feedback over the course of a year using SCT items. Feedback is both global (eg, health professionals are told their cumulative mean score) and specific (eg, they can view the expert panel's responses together with their justifications for their answers). Participants have the option of contacting a personal tutor, to whom they can send questions. Data regarding feasibility, participation, and acceptability were collected.

**Results:** Initial implementation took place in Mexico where 1901 physicians (1349 paediatricians, 552 cardiologists) were enrolled in Practicum programs. Around 70% of those enrolled pursued the program and were very satisfied with its format and content. The online format was an important factor in the development and maintenance of the programs. Dropouts had issues with the SCT concept and the time required to participate.

**Discussion:** The on-line Practicum Script Concordance Test program was designed to foster expertise development based on practice, reflection and feedback. Although further research is needed to examine its impact physicians' practice and ultimately on patient outcomes, it is an original and promising development in CPD.

**Key Words:** continuing professional development, script concordance, clinical reasoning, reflection, online learning method, evaluation, practicum

---

Disclosures: The authors report none.

*Dr. Hornos:* President, Practicum Institute of Applied Research in Health Sciences Education; *Dr. Pleguezuelos:* Academic Secretary, Practicum Institute of Applied Research in Health Sciences Education; *Dr. Brailovsky:* Director of the Psychometrics Department, Practicum Institute of Applied Research in Health Sciences Education, and Consultant on Testing and Assessment, College of Family Physicians of Canada; *Mr. Harillo:* CIO, Practicum Institute of Applied Research in Health Sciences Education; *Dr. Dory:* Post-doctoral Researcher, FNRS and Institute of Health and Society (IRSS), Université catholique de Louvain; *Dr. Charlin:* Director of the Unit of Research, Practicum Institute of Applied Research in Health Sciences Education, and Centre de Pédagogie Appliquée aux Sciences de la Santé, Faculty of Medicine, University of Montreal.

*Correspondence:* Bernard Charlin, CPASS (Centre de pédagogie appliquée aux sciences de la santé), Faculty of Medicine, University of Montreal, CP 6128 Succursale centre-ville, Montreal, Qc, H3C 3J7, Canada; e-mail: bernard.charlin@umontreal.ca.

## Introduction

Continuing professional development (CPD) programs provide opportunities for physicians to keep abreast of new developments in their field. Generally, these include didactic lectures to inform participants about new knowledge gained from research, and workshops to train them in new procedural skills. However, CPD is also concerned with developing expertise. Expertise comprises more than knowing the latest facts or being proficient in a recently-developed technique. Expertise in medicine involves demonstrating judgment in the indeterminate areas of practice where clear-cut

© 2013 The Alliance for Continuing Education in the Health Professions, the Society for Academic Continuing Medical Education, and the Council on Continuing Medical Education, Association for Hospital Medical Education.

• Published online in Wiley Online Library (wileyonlinelibrary.com).  
DOI: 10.1002/chp.21166

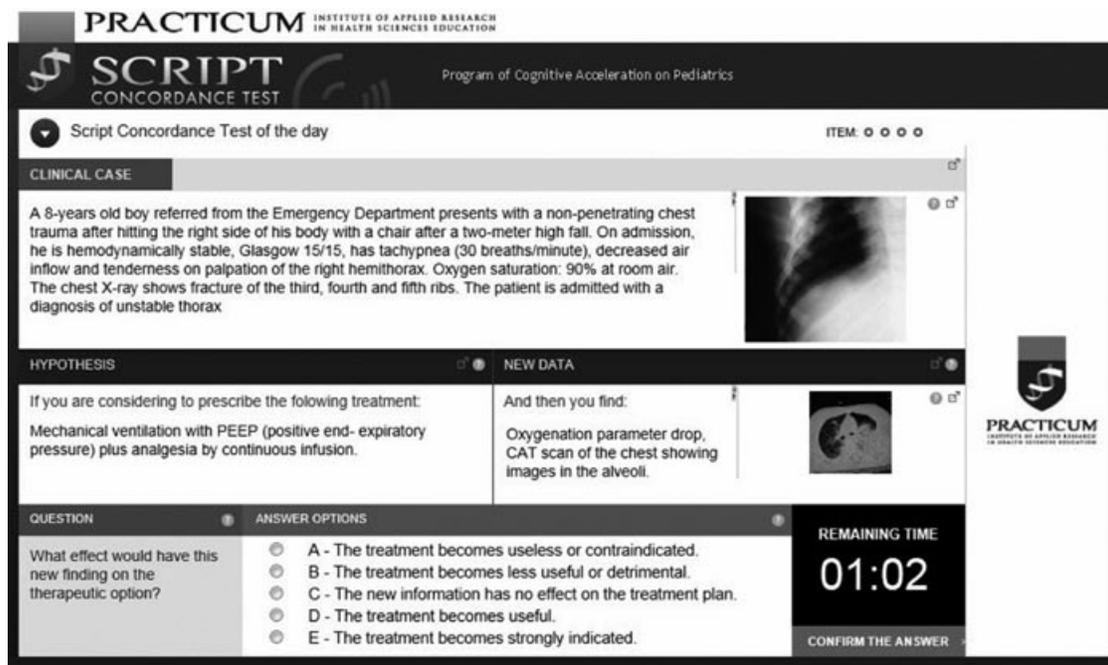


FIGURE 1. Screenshot of a Question Seen by Participants

solutions cannot be simply applied.<sup>1,2</sup> To do this requires more than factual knowledge. Studies on expert knowledge have found that experts have qualitatively different knowledge architectures.<sup>3,4</sup> Biomedical knowledge becomes “encapsulated,” or subsumed into more efficient knowledge structures called *illness scripts*, which contain the clinical knowledge required to categorize a patient’s presentation and guide decision making.<sup>3–5</sup> Scripts become refined with clinical experience.<sup>5</sup> Constant and deliberate fine-tuning of performance is the hallmark of true expertise,<sup>6</sup> and reflection is one of the key strategies practitioners use to learn from experience.<sup>2,7</sup>

Reflection is triggered by surprise or perplexity.<sup>7,8</sup> When physicians face a complex, ill-defined case for which they cannot find a routine solution, they must “reflect-on-action”; that is, envisage one or several potential solutions, try them out, and see what happens.<sup>2,8</sup> To make the most of this experience, they can later “reflect-on-action, analyzing the situation and how it unfolded to draw conclusions about how they could transfer insights from the experience to deal with a similar problem in the future.<sup>2,8</sup> Reflection requires an awareness of the challenge posed by the situation that created the sense of surprise or perplexity. This ability to self-monitor in action and self-assess one’s skills more broadly is far from perfect,<sup>9,10</sup> which poses a threat to the quality of learning by reflection. In order to inform their reflection, experts have been found to supplement their self-assessments by deliberately seeking valid external data, a process referred to as “self-directed assessment seeking.”<sup>10</sup>

CPD can facilitate reflective practice by providing physicians with external feedback on their performance. In an effort to contribute to this goal, the Practicum Institute of Applied Research in Health Sciences Education (Madrid, Spain) designed an online platform to provide physicians with opportunities to self-test and receive feedback on their clinical judgment in uncertain situations using the Script Concordance Test (SCT). The script concordance test was developed precisely to measure this important aspect of clinical expertise.<sup>11,12</sup> Other written tests of clinical reasoning require learners to select the single best answer. This constrains the testing domain to areas where a single best answer can be agreed upon.<sup>13</sup> By contrast, the SCT focuses on ill-defined problems where judgment is required. This is made possible by the aggregate scoring method used. The answer key to SCTs is computed by asking a panel of 10 to 20 experts to respond to the questions. Answers that no panel member selects are given no credit. Other answers are awarded credit depending on the fraction of experts selecting that answer. Higher scores indicate higher “concordance” with the expert panel. This allows the testing domain to be extended to the murky, uncertain areas of practice most relevant to the development of expertise.

The structure of the SCT is another of its key features (FIGURE 1). Research on clinical reasoning has found that clinicians generate hypotheses early on in an encounter and use these hypotheses to guide their data collection. They then iteratively assess the fit between the data acquired and the hypotheses.<sup>5</sup> This general process of clinical reasoning

is called *hypothetico-deductive reasoning*.<sup>14</sup> The SCT aims to simulate this process. It presents a brief clinical scenario and several plausible hypotheses (related to diagnosis, investigations, or management). Following each hypothesis is an additional piece of information. The question pertains to the effect of the new piece of information on the initial hypothesis; for example, its impact on the likelihood of the diagnostic hypothesis, or on the appropriateness of the management option. Each case is followed by 3 to 5 questions. Respondents provide their answer on a 5-point Likert scale. The validity argument of the SCT is strong, particularly in undergraduate and postgraduate medical education.<sup>12</sup> It has also been successfully used with physicians for assessment purposes<sup>15</sup> and as an adjunct to CPD activities to stimulate reflection and peer-group discussions.<sup>16,17</sup>

We report on the development and initial implementation of an online CPD program (Practicum Script Concordance Test) that used SCT and feedback to stimulate reflection. Guiding principles of this approach were: to engage learners in authentic, complex, clinical reasoning tasks rather than to provide facts; to provide feedback in a way that encourages reflection-on-action from a cognitive perspective; and to offer a learning format that is convenient for both physicians and CPD providers.

## Method

### *Overview of the Program*

The Practicum Institute of Applied Research in Health Sciences Education works with various scientific societies in Spanish-speaking countries to design online CPD programs in several medical specialties. Each Practicum SCT program consists of a yearlong cycle of daily testing and feedback. Once enrolled, participants are sent one SCT case with a set of 4 to 5 questions per day, 5 days a week. In total, a yearly cycle comprises 240 cases. Participants can choose to respond on a daily basis or whenever is most convenient. Questions are displayed one by one on their computer screen. Multimedia resources such as images or videos are incorporated in most case scenarios to enhance their authenticity. Participants are given up to 90 seconds to select their answer. In order to ensure that the questions present an appropriate challenge to participants, the difficulty of questions is adapted to each participant, initially according to their level of clinical experience, and subsequently according to their performance on the 10 previous SCT cases. When a case is failed it can be retaken after a lapse of 5 weeks or more.

Feedback is a crucial feature of the program and is provided in several formats. On completing each case, participants receive their score together with the responses of the panel of experts. To facilitate reflection, the justifications provided by panel experts are included with their responses

(see FIGURE 2). Participants are also offered more global feedback in terms of their global mean score. Furthermore, by means of graphics and charts, they can consult their performance in the different types of medical decisions or components of clinical reasoning (diagnosis, investigations, and treatment) as well as in the different content areas of their specialty (for instance, within cardiology, the mean score in areas such as coronary artery disease, heart failure, etc, see FIGURE 3). This gives them information on their relative strengths and weaknesses. Participants are also assigned a personal tutor to whom they can send questions. Tutors can access participants' professional profile data, answer history, as well as previous queries and can personalize their interaction with participants accordingly. The Web site also provides a link to a help desk for technical and basic methodological support.

CPD credits are provided to the physician if, at the end of the year, their performance, reaches a predetermined level. The minimum performance needed to obtain CPD credits and the number of CPD credits granted to the participants are regulated by the local accreditation bodies of the specialty in each country (eg, scientific society or medical council) according to their official CPD and recertification policies. For example, in Mexico, CPD for pediatricians is regulated by the Mexican Council of Certification in Pediatrics. Pediatricians must obtain a total of 100 credits over a 5-year period. Satisfactory participation in the Practicum Script Concordance Test program for 1 year earns 10 credits. For cardiologists, recertification is governed by the Mexican Council of Cardiology. Cardiologists must acquire 40 credits over a 5-year period. Participation in the Practicum Script Concordance Test program for 1 year earns 1 credit. Both organizations require that participants respond to a minimum of 75% of the cases and that they reach a minimum total score of 60 out of 100.

### *Development of the Item Bank*

Blueprints were developed by a group of experts within each specialty according to rules defined for content validity of exams as described by Brailovsky and GrandMaison.<sup>18</sup> They were based on the learning outcomes defined by the scientific societies of each discipline. Blueprints included content areas, key clinical situations and the proportion of cases to be created per area.

Item writing committees were set up in each discipline. They included experts from several countries fulfilling the following inclusion criteria:

- General specialist (not subspecialized in a specific area) certified by a recognized licensing body in his/her country.
- At least 10 consecutive years experience as a physician and 8 years experience as a specialist.

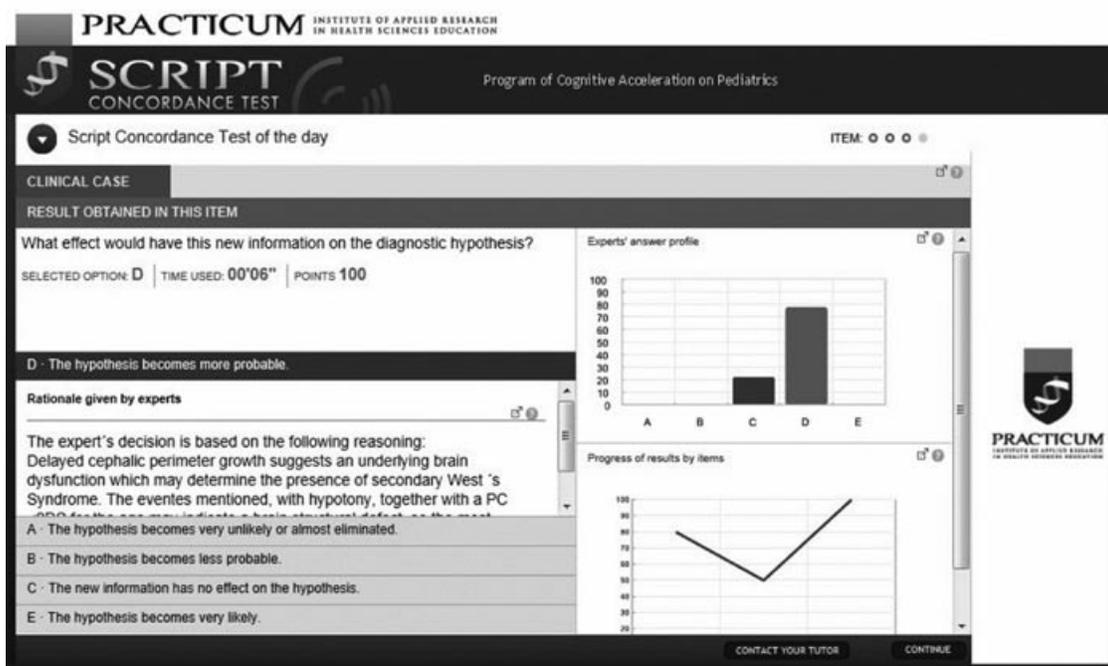


FIGURE 2. Screen Displaying the Answer Chosen by the Participant (D) and the Answers Given by the Reference Panel (The most frequently selected answer is D which is therefore credited with a score of 100%; C was selected by one-fourth as many experts and so is awarded a 25% credit). If participants click on D or C they can see the justifications given by experts for those answers

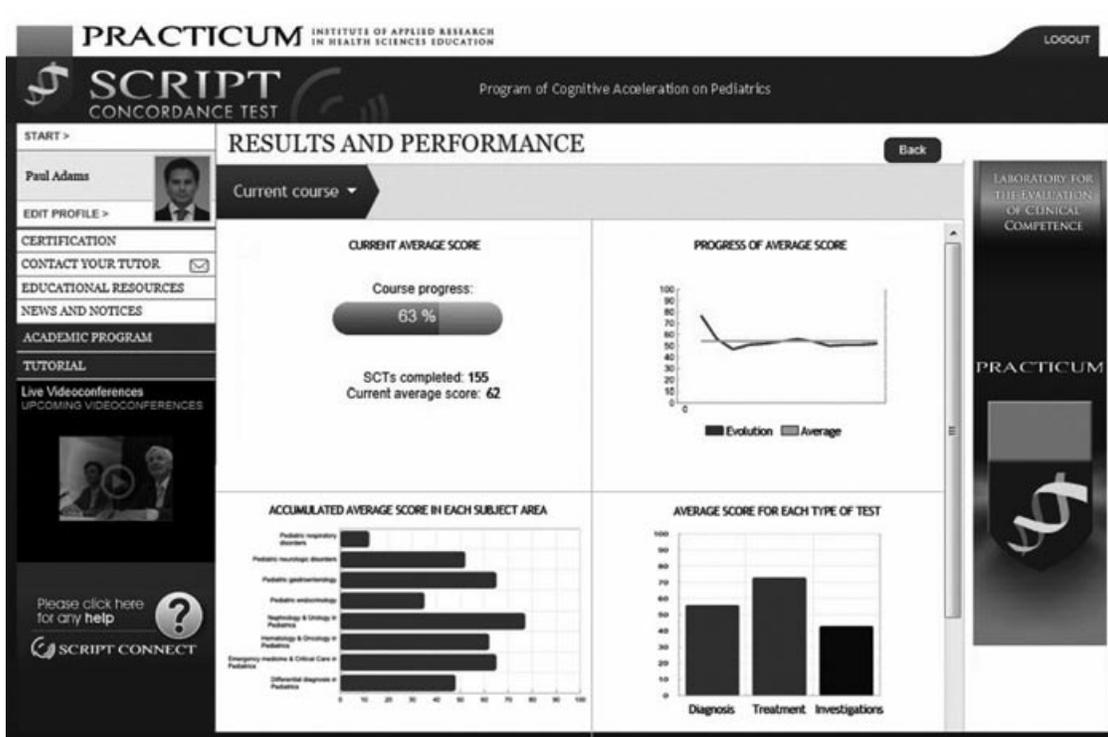


FIGURE 3. Global Feedback Including Average Score and Progress as Well as Results in the Different Types of Medical Decisions (Diagnosis, Investigations, and Treatment) and in the Different Content Areas of the Specialty

- Involvement in postgraduate training in the specialty in the previous 5 years.
- 50% of working time involved in patient care (currently and in the past 5 years).
- Evidence of relevant involvement in academic activities, medical conferences, and scientific publications (journal articles or monographs).

Training for item construction followed published guidelines<sup>19</sup> and was done in face-to-face full-day workshops. From then on, item writers did their tasks online using a dedicated electronic platform (<http://experts.script.md>) specially designed to facilitate the process of SCT banking (eg, items construction, collection of panel experts' answers and justifications, examination and optimization of the panel's responses) for Practicum programs. A personal tutor accessible through the platform provided methodological guidance. Items were reviewed by at least two other experts.

Scoring keys were then developed by submitting items to a reference panel composed of other experts in the same specialty and nominated by local scientific societies. These experts fulfilled the same criteria as item writers and also completed a mandatory workshop in which they were trained in the principles of script concordance testing and the use of the digital resources needed to perform their tasks. They could not, however, have participated in item writing. They were asked to answer the cases online as if they were being tested themselves and their answers were used to compute each item's scoring key. Outlier answers were not taken into account in the answer key computation and items that elicited highly variable answers from the panel were eliminated. Exclusion of deviant answers from discordant panellists was made by the the judgment-by-experts method.<sup>20</sup> Experts were asked to provide a justification for each of their answers to serve as feedback and as an aid for reflection for participants. They were invited to comment on the appropriateness and clarity of items. Finally, they assessed item difficulty as easy (1 point), medium (2 points), or difficult (3 points). These difficulty ratings were averaged to determine the level of difficulty of the case.

After they are deployed, items are continuously reviewed. Participants' responses are analyzed to determine the psychometric properties of items. Participants are also invited to submit their comments regarding the quality and relevance of items for their clinical practice. Participants' feedback is subjected to analysis by a Practicum review committee and represents an important part of the optimization process of the content for subsequent yearly cycles.

#### Technical Aspects

A multidisciplinary team composed of medical educators, clinical specialists, psychometricians, and software engi-

TABLE 1. Characteristics of Enrolled Learners

		N	%
Age (years)	<30	163	8.5
	30–39	488	25.7
	40–49	509	26.8
	50–59	538	28.3
	≥60	203	10.7
Gender	Male	1420	74.7
	Female	481	25.3
Years of experience as a specialist	<5	549	28.9
	5–9	0	0
	10–14	403	21.2
	15–19	298	15.7
	≥20	651	34.2

neers designed the online platform, which is used to facilitate the processes of item writing and translation, item banking (each program includes over 1000 individual items), item validation, item optimization, compilation of experts' justifications, computation of answer keys, and test administration.

#### Data Collection—Initial Implementation

Implementation began in 2010 in Mexico and Chile with one program in cardiology and one in pediatrics (TABLE 1). In this article, we report on the Mexican experience where 552 cardiologists and 1349 pediatricians were enrolled (total of 1901 participants). Evaluation of this experience was undertaken in the middle of the first 2 cycles. It consisted of (1) an automated collection of data regarding tutor use and technical problems; (2) an automated gathering of demographic characteristics of learners enrolled in the programs; and (3) a survey questionnaire sent to all enrolled participants, dropouts included, to collect information about acceptability, level of satisfaction with the Web-based SCT learning model, and their perceptions of its utility. Participants were defined as having dropped out if they had ceased to complete cases for at least 4 weeks before the survey. TABLE 2 presents the survey questions.

#### Results

Feasibility data: The process of recruiting experts to write items and serve as the reference panel took around 1 to 2 months. Having an electronic platform was instrumental in allowing us to recruit a large number of experts from several

TABLE 2. Data Obtained From Participants by the Survey

Survey Questions		Cardiology (n = 168)*					Pediatrics (n = 332)*				
		5	4	3	2	1	5	4	3	2	1
Overall format	Satisfaction with website's ease of use and navigation	59.3%	39.6%	0.0%	1.1%	0.0%	61.0%	38.0%	1.0%	0.0%	0.0%
	Satisfaction regarding the relevance of the clinical scenarios	53.3%	42.4%	1.1%	1.1%	2.1%	57.7%	37.1%	1.1%	2.4%	1.7%
	Satisfaction with SCT format and response methodology	38.5%	48.4%	2.1%	4.4%	6.6%	32.6%	46.9%	8.1%	6.8%	5.6%
	Satisfaction with time required	54.4%	38.9%	1.1%	3.4%	2.2%	54.9%	38.2%	1.1%	2.4%	3.4%
	Compared to other CPD formats, I found the SCT format...	57.8%	27.8%	12.2%	2.2%	0.0%	53.4%	35.7%	9.2%	0.6%	1.1%
Perceived benefits	It is useful for my clinical practice.	69.2%	19.8%	8.8%	2.2%	0.0%	65.3%	23.9%	7.4%	2.3%	1.1%
	It helps me to strengthen my clinical reasoning abilities	67.0%	18.7%	11.0%	3.3%	0.0%	71.1%	20.9%	6.8%	0.6%	0.6%
	It encourages me in terms of medical information search to broaden my knowledge of the specialty	73.6%	16.5%	7.7%	2.2%	0.0%	67.0%	18.7%	11.0%	3.3%	0.0%
	It gives me the opportunity to get CPD points.	59.3%	16.5%	14.3%	3.3%	6.6%	67.0%	18.7%	11.0%	3.3%	0.0%
	It promotes changes in behavior in my clinical practice	53.8%	31.9%	7.7%	6.6%	0.0%	—	—	—	—	—

\* Ratings on the questionnaires: 5 = very satisfied, 4 = satisfied, 3 = neither, 2 = dissatisfied, 1 = very dissatisfied)

countries. Item writing took around 3 to 4 months, including several rounds of validation. The electronic platform helped a great deal. In particular, we found that by providing item writers and reviewers with online help, we were able to increase the quality of items written and hence reduce the rate of item exclusion, which ultimately led to cost savings. It also helped in the psychometric analysis of items and item optimization process.

Characteristics of enrolled learners are presented in TABLE 1. Six hundred ninety-two participants sent a total of 3820 queries to get help or provide comments. The tutors responded personally to 42% of these, with the help desk providing support for the remaining 58%. The help desk was staffed by nonphysicians trained to provide technical assistance regarding difficulties in subscription, recovery of user names and passwords, and Web site functionalities. They were also trained to provide support regarding basic methodological aspects of the Practicum SCT program, such as how to answer the SCTs, how to review the results, how the scoring system works, and so on. Technical difficulties represented only 2.6% of queries, with methodological questions accounting for 22.3% and questions or comments regarding the clinical content of items and expert responses for another 33.2%.

Dropout rates in the yearlong programs were 29.7% among cardiologists and 31.1% among pediatricians. Only 10.7% of dropouts in cardiology and 13.8% in pediatrics responded to our survey. The main reasons they invoked for not continuing with the program were lack of time (53.3% and 52.4% in cardiology and pediatric respectively) and

difficulty grasping SCT methodology (46.7% and 52.4%). Among dropouts in pediatrics, 38.1% were not satisfied with the SCT format and response methodology. In free-text comments, many commented negatively on the lack of a consensus among experts. Response rates among active participants were higher (43.2% in cardiology and 35.7% in pediatrics). Among those who were active participants, satisfaction with the format and content of the program was high (TABLE 2). Over 96% said they would recommend the program to colleagues.

## Discussion

We designed the Practicum Script Concordance Test format to help practicing physicians develop their clinical expertise through reflection. The key features of the program were its online platform, focus on clinical judgment in uncertain clinical situations, self-testing on authentic clinical scenarios, and detailed feedback intended to stimulate reflection-on-action.

The online platform was a significant factor in the development, implementation, and maintenance of the program as a whole. It facilitated the construction of the large item banks required (over 1000 individual items in each program) and allowed us to recruit experts from several countries to write, respond to, and comment on test questions. As such, it played a key part in making this international collaborative endeavour possible. It was also a key component in the item quality control processes by providing psychometric data on individual items and by allowing participants in the

program to offer comments and suggestions regarding the items.

Of course, online formats are also thought to make participating in CPD more convenient: participants can learn from the comfort of their homes or offices whenever they like. Online testing also affords the possibility of giving learners specific feedback. By giving participants the panel's answers together with a rationale for each question, Practicum SCT should facilitate reflection by helping participants understand the reasoning behind each expert's response. However, receiving feedback is not sufficient in itself to produce change: learners must assimilate it and this is not always as straightforward as it sounds, particularly if the feedback is discordant with one's own perceptions.<sup>21</sup> Sargeant et al, for example, found that reflection was a key process in physicians assimilating feedback.<sup>21</sup> Although their findings related to a different CPD format—multisource feedback—the general processes are likely to be the same. They also found that facilitation of reflection was important.<sup>21</sup> We provided participants with access via the Web site to a personal tutor who could serve as a reflection facilitator. However, only around one-third of participants submitted queries and in fact most queries were not related to learning issues. It is unclear whether participants did not actually require such facilitation or whether they were reluctant to call upon their tutors for whatever reason.

Another key feature of the program was the use of the script concordance test as a stimulus for reflection. Among dropouts who responded to our survey, around half were dissatisfied with the script concordance format, which does not provide a single-best-answer solution. While others have also criticized the script concordance tests for this reason,<sup>22</sup> it is one of its core strengths. By not requiring a single-best solution, it allows item developers to include the more complex cases faced in the murky waters of real-life clinical practice. Nevertheless, we intend to take these comments into account by providing more information on script concordance methodology to participants.

Another significant issue raised by dropouts was the issue of time. Although we designed the program to provide small "daily doses" of learning, it is possible that the sum of these repeated tasks (240 cases per year, each including 4 to 5 questions) may become overwhelming. However, most of those enrolled persisted with the program, and 96% of those who responded to our survey would recommend the program to their colleagues.

Further evaluation is needed to examine whether the program actually leads to better practice and patient outcomes. Traditional didactic CPD activities have been found to be poor in this respect,<sup>23</sup> but it is unclear how our program would compare to other modern CPD formats such as those using audit and feedback, multisource feedback, portfolios, or various reminders. Nevertheless, we believe the Practicum

### Lessons for Practice

- CPD can facilitate reflective practice by providing health professionals with external feedback on their performance. The *Practicum Script Concordance Test* (SCT) has been designed to contribute to this goal. This online platform provides professionals with opportunities to self-test and to receive feedback on their clinical judgment. The answer key to SCTs is computed by asking a panel of experts to respond to the questions. Participants are awarded credit depending on how many experts selected that answer. This allows the testing domain to be extended to the murky, uncertain areas of practice most relevant to the development of expertise.
- Each Practicum SCT program consists of a yearlong cycle of daily testing and feedback. Once enrolled, participants are sent one SCT case with a set of 4 to 5 questions per day, 5 days a week. In total, a yearly cycle comprises 240 cases. Feedback is both global (eg, they are told their cumulative mean score) and specific (eg, they can view the expert panel's responses together with their justifications for their answers). When a case is failed it can be retaken after a lapse of time.
- Two programs were implemented in Mexico, where 1901 physicians (1349 pediatricians, 552 cardiologists) were enrolled in Practicum programs. Around 70% of those enrolled pursued the program and were very satisfied with its format and content. The online format was an important factor in the development and maintenance of the programs. This learning format appears convenient for both physicians and CPD providers. Among active participants, satisfaction with the format and content of the program was high. Over 96% said they would recommend the program to colleagues.

Script Concordance Test program is a promising and original development that can provide physicians with a convenient educational experience having the potential to help them

continuously fine-tune their judgment in uncertain clinical situations through practice, feedback, and reflection.

## Acknowledgments

The authors would like to thank Pablo Malagrino, MD, for his help in preparing this manuscript and the Fonds de la Recherche Scientifique (FNRS, a regional fund for scientific research), which funded Valerie Dory's postdoctoral research on assessment of clinical reasoning.

## References

1. Epstein RM HE. Defining and assessing professional competence. *JAMA* 2002; 287(2):226–235.
2. Schön DA. *The Reflective Practitioner: How Professionals Think in Action*. New ed. Aldershot, ON: Ashgate; 1995.
3. Schmidt HG, Boshuizen HP. On acquiring expertise in Medicine. *Educ Psychol Rev*. 1993;5(3):205–221.
4. Schmidt HG, Rikers RMJP. How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Med Educ*. 2007;41(12):1133–1139.
5. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med*. 2000;75(2):182–190.
6. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med*. 2004;79(10)(suppl):S1–S12.
7. Mann K, Gordon J, Macleod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract*. 2009;14(4):595–621.
8. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ*. 2004;38(12):1302–1308.
9. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296(9):1094–1102.
10. Eva KW, Regehr G. “I’ll never play professional football” and other fallacies of self-assessment. *J Contin Educ Health Prof*. 2008;28(1):14–19.
11. Charlin B, van der Vleuten C. Standardized assessment in context of uncertainty: the script concordance approach. *Eval Health Prof*. 2004;27(3):304–319.
12. Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CP. Script concordance testing: a review of published validity evidence. *Med Educ*. 2011;45(4):329–338.
13. Elstein AS. Beyond multiple-choice questions and essays: the need for a new way to assess clinical competence. *Acad Med*. 1993;68(4):244–249.
14. Elstein AS, Schulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press; 1978.
15. Goulet F, Jacques A, Gagnon R, Charlin B, Shabah A. Poorly performing physicians: does the Script Concordance Test detect bad clinical reasoning? *J Contin Educ Health Prof*. 2010;30(3):161–166.
16. Labelle M, Beaulieu M, Paquette D, et al. An integrated approach to improving appropriate use of anti-inflammatory medication in the treatment of osteoarthritis in Quebec (Canada): the CURATA model. *Med Teach*. 2004;26(5):463–470.
17. Petrella RJ, Davis P. Improving management of musculoskeletal disorders in primary care: the Joint Adventures Program. *Clin Rheumatol*. 2007;26(7):1061–1066.
18. Brailovsky, CA, Grand'Maison, P. Using evidence to improve evaluation: A comprehensive psychometric assessment of a SP-based OSCE licensing examination. *Adv Health Sci Educ*. 2000;5:207–219.
19. Fournier JP, Demeester A, Charlin B. Script Concordance Tests: guidelines for construction. *BMC Med Inform Decis Mak*. 2008; 8:18.
20. Gagnon R, Lubarsky S, Lambert C, Charlin B. Optimization of answer keys for script concordance testing: should we exclude deviant panelists, deviant responses, or neither? *Adv Health Sci Educ*. 2011; 16(5):601–608.
21. Sargeant J, Mann K, Van der Vleuten C, Metsemakers J. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ*. 2009;14(3):399–410.
22. Bland AC, Kreiter C, Gordon JA. Investigating the psychometric properties of five scoring methods applied to the Script Concordance Test. *Acad Med*. 2005;80:395–399.
23. Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *JAMA*. 1999;282(9):867–874.