Use of a Script Concordance Test to Assess Development of Clinical Reasoning in Nursing Students

Marie-France Deschênes, MSc (Nursing), RN; Bernard Charlin, PhD, MD; Robert Gagnon, MPsy; and Johanne Goudreau, PhD, RN

ABSTRACT

The methods for assessing and measuring clinical reasoning as it pertains specifically to nurses are inadequate, if not nonexistent. The purposes of this methodological study were to develop a script concordance test and conduct a preliminary validation of its psychometric qualities. A script concordance test was created and the test scoring grid was constructed using the combined-score method and based on the responses of a panel of 15 experts. Thirty first-year bachelor of nursing students completed the test. The scores for the experts and students were compared with a *t* test, and the reliability of the scores was measured by Cronbach's alpha coefficient. A statistically significant difference was found between the scores of the experts and novices. The scores' reliability is high. The script concordance test is an innovative instrument that provides a standardized method for assessing nurses' clinical reasoning.

Received: March 2, 2010
Accepted: August 31, 2010
Posted: March 31, 2011

Ms. Deschênes is nursing instructor, Collège de Maisonneuve, Dr. Charlin is Director of Research and Development and Dr. Gagnon is consultant, Centre de Pédagogie Appliquée Pour Les Sciences de la Santé (CPASS), and Dr. Goudreau is Associate Professor and Vice-Dean (Baccalaureate Programs), Faculty of Nursing, Université de Montréal, Montreal, Quebec, Canada. Dr. Goudreau is also Researcher, Équipe de Recherche en Soins de Première Ligne, Centre de Santé et de Services Sociaux de Laval, Laval, and Researcher, Center for Innovation in Nursing Education, Montreal, Quebec, Canada.

The authors have no financial or proprietary interest in the materials presented herein.

Address correspondence to Johanne Goudreau, PhD, RN, Associate Professor and Vice-Dean (Baccalaureate Programs), Faculty of Nursing, Université de Montréal, Pavillon Marguerite-d'Youville, 2375 Côte-Ste-Catherine, Montreal, Quebec, Canada H3T 1A8; e-mail: johanne.goudreau@umontreal.ca.

doi:10.3928/01484834-20110331-03

Constantly evolving health care needs and expectations, the growing diversity of clinical situations, and rapid technological change all play a part in making today's practice of nursing increasingly complex and, consequently, in making sound clinical reasoning a crucial skill for nursing students to develop (Kautz, Kuiper, Pesut, Knight-Brown, & Daneker, 2005; Simmons, Lanuza, Fonteyn, Hicks, & Holm, 2003). Therefore, teaching establishments have made major efforts to put effective pedagogical methods into place and enhance learning in the area of clinical reasoning. However, there remains the challenge of how to assess such reasoning. It is a complex phenomenon and, although it is an essential component of professional practice and its attributes are known, the process by which it is developed is not clear (Fonteyn, 1998; Fonteyn & Ritter, 2000; Kuiper & Pesut, 2004). The issue of how to assess students' clinical reasoning performance during their academic training arises continually in education curricula in the health sciences, including nursing.

In response to this problem, over the past several years researchers have developed and experimented with a new measurement instrument—the script concordance test. Based on cognitive script theory, script concordance tests initially emerged in the field of medicine (Charlin, Boshuizen, Custers, & Feltovich, 2007). A script concordance test is a written examination that presents candidates with several prototypical clinical practice situations in which information is ambiguous, complex, or incomplete. Scoring entails the establishment of a reference panel. The number of points awarded depends on the number of panel members who respond as the students do, hence the name *concordance test* (Charlin & van der Vleuten, 2004; Fournier, Demeester, & Charlin, 2008). The current article describes an experiment on the use of script concordance testing to assess human caring, a dimension of clinical reasoning particular to nurses.

LITERATURE REVIEW

Clinical reasoning is a function of knowledge, attitudes, and reflective professional practice. It is judged by the correctness of the interventions performed in the wake of a decision making process (Higgs & Jones, 2000). As it relates to clinical problem solving (Charlin, Bordage, & van der Vleuten, 2003; Nendaz,

Charlin, Leblanc, & Bordage, 2005), clinical reasoning is defined by processes of reflection and decision making that allow professionals to perform the intervention deemed to be most appropriate in a particular clinical context (Higgs & Jones, 2000). Clinical reasoning refers to any intellectual activity in which the information that has been collected is integrated, assimilated to the professional's knowledge and previous experience, and used to arrive at a highly effective decision (Charlin et al., 2003; Charlin, Boshuizen, et al., 2007; Nendaz et al., 2005). However, studies on clinical reasoning have demonstrated that although experts may come to the same clinical judgment, their reasoning processes can differ (Elstein, Schulman, & Sprafka, 1978; Grant & Marsden, 1988). The result is the continuing difficulty of conducting a clear-cut in-training assessment of students' clinical reasoning (Charlin et al., 2003; Charlin, Gagnon, Sibert, & van der Vleuten, 2002).

Several authors have described clinical reasoning in the nursing field in terms of recursive cognitive processes that nurses use to understand and give meaning to the clinical information that has been collected and to determine an appropriate treatment plan (Fonteyn & Cahill, 1998; Fonteyn & Ritter, 2000; Kautz et al., 2005; Simmons et al., 2003). In handling complex situations, expert nurses do not always take a systematic, linear approach to problem solving (Grobe, Drew, & Fonteyn, 1991; McCarthy, 2003a, 2003b). The strategies they follow often involve putting their clinical hypotheses into practice quickly and handling and processing problem-related information and the associated interventions jointly and simultaneously (Goransson, Ehnfors, Fonteyn, & Ehrenberg, 2008; Simmons, 2002; Simmons et al., 2003). They link data around various key concepts that guide and determine their reasoning process (Greenwood, Sullivan, Spence, & McDonald, 2000; Simmons et al., 2003).

Experience, a sound knowledge base, and intuition are the principal components of proven clinical reasoning in expert nurses (Benner, Tanner, & Chesla, 1996; Fonteyn & Ritter, 2000). McCarthy (2003a, 2003b) and Carr (2004) noted further that nurses' clinical reasoning is influenced by their philosophical perspectives; their way of thinking significantly influences their assessment and interpretation of clinical situations and their responses. Nurses use a combination of different knowledge and adapt it to the contingencies of the clinical situations they encounter. The care they provide is also a function of the values and beliefs that underlie their nursing practice (Kérouac, Pepin, Ducharme, & Major, 2003). However, although various schools of thought provide archetypes for practice, the application of assessment in the context of a cognitive perspective on learning remains underdeveloped, if not nonexistent. With this in mind, and address a specific type of reasoning particular to the discipline of nursing, Watson's (1988, 1997, 2002) theory of human caring was used as a frame of reference and the subject to be assessed by the script concordance test developed for the current study.

Watson's notion of human caring is a conception of nursing as a discipline founded on phenomenological, existential, scientific, and humanistic principles (Kérouac et al., 2003). It involves attaining a moral ideal with the aim of promoting the health and development of the individual while maintaining and respecting human dignity (Cara, 2003). In the context of this study, nursing clinical reasoning based on human caring manifests itself in every reflective process in which a nurse makes a decision about a clinical intervention and considers the unbreakable association between person and environment. The interventions favored are characterized by mobilization of the individual's resources. Expected outcomes center on the concerns and priorities of the person rather than on the application of predetermined, predictive approaches. Assessing a health situation is an iterative process that involves continual validation of the perception and meaning of the experience of the person being treated. All of these elements combine in an attentive presence and a spirit of consensual partnership with the other. Teaching establishments must seek to help students develop a mode of reflection and a practice of nursing based on caring and marked by diligence, competence, and conscientiousness (Duffy, 2005). The problems are how to support nursing students in this reflective process and how to assess their clinical reasoning.

It is difficult, if not impossible, to assess clinical reasoning through direct observation. Reasoning may be inferred from (observable) answers to a questionnaire or from decisions that are made. For example, an objective structured clinical examination may be used to effectively assess the clinical process by direct observation of behavior and of skills applied during simulated clinical situations. However, for an objective structured clinical examination to have good psychometric qualities, it would have to have 20 or more stations (Norman, 1993), a number that would make the evaluators' work considerably more burdensome. The general assessment guides used in clinical internships also suffer from significant limitations, notably in terms of interjudge reliability and the limited number of observations (Charlin et al., 2003; Norman, 1993). Page, Bordage, and Allen (1995) proposed the use of key feature examinations, written examinations setting out clinical situations followed by multiple choice and short-answer and long-answer questions. Although such examinations make it possible to properly assess decision making abilities, considerable time is needed to create them, and many of cases are required to ensure their reliability (Charlin et al., 2003). Reflective practice, using a log book, and mapping clinical situations provide qualitative data by which students' clinical reasoning may be assessed. Such instruments make it possible to substantiate the development of reasoning through cognitive and metacognitive information-processing activities (Fonteyn & Cahill, 1998; Kuiper & Pesut, 2004). Although useful, they fall short as reliable, valid, standardized tools to provide correct, accurate and adequate evidence of nursing clinical reasoning (Kautz, Kuiper, Bartlett, Buck, Williams, & Knight-Brown, 2009).

A script concordance test is a written examination by which clinical reasoning may suitably be assessed (Charlin, Boshuizen, et al., 2007). Script concordance testing is based on script theory and was developed in the late 1990s. Script theory (Charlin, Boshuizen, et al., 2007; Charlin, Tardif, & Boshuizen, 2000; Feltovich & Barrow, 1984; Schmidt, Norman, & Boshuizen, 1990) is derived from cognitive theories of learning and posits that, faced with a clinical situation and in a problem–resolution process, health care professionals use mental scripts (preestablished networks of knowledge stored in their long-term memory). The term *script* refers to the associative networks,

the forged linkages between knowledge elements that guide the selection, interpretation, and memorization of information (Charlin, Boshuizen, et al., 2007; Charlin, Tardif, et al., 2000; Schmidt et al., 1990). Scripts are essentially described as structures of associative linkages between knowledge elements specifically adapted to different clinical tasks. For experts coping with clinical situations, clinical hypotheses and their related scripts come to mind, usually emerging from an automatic, relatively unconscious thinking process. Experts may disagree somewhat on the details of their scripts, but they generally arrive at the same judgment, shaped by their personal experience (Charlin, Tardif, et al., 2000).

TABLE 1

Scenario Example

You are in charge of the diabetes clinic and are presenting an information session to a group of people who have been newly diagnosed with the disease.

If you were thinking of:	And one of the attendees said to you:	Your intervention becomes[a]:				
Starting the teaching session by inviting the participants to talk about their experiences/ concerns regarding diabetes	"I don't have a problem. You're taking care of me, and things are just fine that way."	–2	–1	0	1	2
Validating the attendees' knowledge about diabetes mellitus before starting to teach	"I'm not doing anything about changing my eating habits, and I still smoke two packs of cigarettes a day. I know I'm not helping myself."	–2	–1	0	1	2

[a] *–2* = totally or virtually contraindicated; *–1* = not very useful or possibly harmful; *0* = neither particularly useful nor useless; *1* = useful; *2* = necessary or absolutely necessary.

Script concordance testing enables us to see whether the decisions made through the students' reasoning processes are similar to those of a panel of experts in the field (Charlin et al., 2002; Charlin, Roy, Brailovsky, Goulet, & van der Vleuten, 2000; Charlin & van der Vleuten, 2004). In a script concordance test, a candidate is presented with several clinical scenarios that reflect situations similar to those experienced in real life, circumstances of the type genuinely faced by professionals, down to the element of uncertainty. The test situations pose a problem for experts also because some of the information is incomplete, ambiguous, or hard to interpret (Charlin, Roy, et al., 2000; Charlin et al., 2002; Charlin & van der Vleuten, 2004). Script concordance testing assesses the quality of clinical reasoning by judging whether students have refined and organized—or merely amassed—knowledge (Charlin et al., 2003; Charlin & van der Vleuten, 2004). The originality of script concordance tests lies in their use of the combined-score method (Norcini, Shea, & Day, 1990; Norman, 1985); there are no right answers in script concordance tests. Every answer an expert gives has intrinsic value and is considered valid, even if it conflicts with the responses of the other experts on the panel. In determining a student's score, script concordance tests account for the normal variability in experts' answers (Charlin et al., 2002; Charlin & van der Vleuten, 2004).

Studies of the use of script concordance tests with medical students have reported excellent psychometric qualities. The statistically linear progression of the scores of groups with differing levels of expertise indicates the construct validity of the instrument (Charlin & van der Vleuten, 2004; Gagnon, Charlin, Coletti, Sauvé, & van der Vleuten, 2005). This validity corroborates the theoretical foundations of script concordance testing—with time and practice, the experts' professional experience allows them to continually refine their mental scripts (Charlin, Tardif, et al., 2000; Charlin & van der Vleuten, 2004; Schmidt et al., 1990). These data emerge from fields ranging from urology (Sibert et al., 2001, 2002) and emergency pediatric medicine (Carrière, 2005) to radiooncology (Lambert, 2005) and emergency medicine (Fournier et al., 2006). Script concordance tests have also been tested several times for internal consistency, an index of reliability, with Cronbach's alpha coefficient. Values approaching 0.75 have been reported for instruments comprising approximately 60 questions (Charlin, Gagnon, Sauvé, & Coletti, 2007; Charlin & van der Vleuten, 2004; Gagnon et al., 2005). Another study found that the use of script concordance tests in conjunction with other didactic methods enhanced the ability of nurses in an intensive care unit to quickly detect signs of delirium in patients (Devlin et al., 2008). To date, no other nursing research has dealt with script concordance testing. However, the use of script concordance tests is highly appropriate for evaluating student reasoning in situations in which there is a degree of uncertainty and in which no single, universal response applies.

METHOD

Development of the Script Concordance Test

Development of the script concordance test involved the following four steps: selecting scenarios that require the application of nursing clinical reasoning based on human caring; writing the scenarios and the test items; having expert judges validate them; and constructing a scoring grid. First, scenarios were selected on the basis of the 14 reserved activities assigned to nurses under the Quebec Nurses Act to assess the student's judgment in prototypical, legislated nursing practice situations. Second, the scenarios were written so that each incorporated three or four items for assessment. The items were created to reflect Watson's (1988, 1997, 2002) carative factors, which characterize nursing reasoning based on human caring. Items were grouped together in relation to the three dimensions of human caring identified by Cossette, Cara, Ricard, and Pepin (2005): the human aspect, the therapeutic relationship, and clinical activities. Each item in turn had three parts: a hypoth-

TABLE 2

Cronbach's Alpha Coefficient

Dimensions of Human Caring	Number of Items	Cronbach's Alpha
Human aspect	31	0.72
Clinical activities	20	0.71
Therapeutic relationship	22	0.68
Total test	73	0.86

TABLE 3

Correlation Between Assessment Dimensions of Human Caring and the Test as a Whole

Dimensions of Human Caring	Total	Human Aspect	Clinical Activities
Human aspect	0.91*	–	–
Clinical activities	0.78*	0.54*	–
Therapeutic relationship	0.84*	0.67*	0.50*

*$p < 0.01$.

esis or option for a nursing intervention, new information, and a 5-point Likert-type scale. The students were asked to consider the effect of new information on a nursing-intervention option or hypothesis. It should be noted that the items in each individual scenario are interdependent; the student had to reflect on every item to respond (Charlin et al., 2002; Charlin, Roy, et al., 2000). **Table 1** provides a clinical scenario followed by two assessment items. Third, the script concordance test was presented to two collaborating expert judges who met the criteria of possessing good knowledge of the human caring school of thought and holding a master's or doctoral degree in nursing. These experts validated the appropriateness of the scenarios and of the test item options. Finally, a scoring grid was created based on the answers provided by a panel of 15 other expert collaborators who took the test. They were required to be experienced nurses who were knowledgeable in human caring.

The test was completed individually, and scoring was determined by the combined-score method (Norcini et al., 1990; Norman, 1985); for each answer, points were awarded based on the number of experts who selected it. Students who chose the response selected by more experts, the modal response, received one point; others received a partial mark (Charlin et al., 2002; Charlin & van der Vleuten, 2004). Thus, given a group of 10 experts, if eight chose –2 *(totally or virtually contraindicated)* and two chose –1 *(not very useful or possibly harmful)*, the value of the –2 response would be 8/8 = 1 point. Scores for other answers are expressed as a ratio of the number of experts who choose them to the number of experts who choose the modal answers; thus a –1 response that was chosen by two experts received 0.25 point (2/8). The score was calculated for each item, and the total score was the sum of the scores for all reported items divided by 100. For each participating expert, scores were determined by using the distribution of all the other experts' responses (i.e., excluding his or her own answers) for the reference panel scoring grid.

Setting and Sample

The script concordance test was tested with a convenience sample of 30 bachelor of nursing students in their first year of the program at the Université de Montréal. Participation was voluntary, and the test took approximately 1 hour to write.

Data Analysis

The data analysis was two-fold: analysis of the psychometric qualities of the script concordance test and the comparison of scores between groups of respondents. To optimize the value of the tool and establish the final reference panel, the distribution and variance of the panel's responses were analyzed and the items were studied to find any that might have been problematic. Homogeneity (internal consistency) was calculated and expressed by Cronbach's alpha coefficient. The Pearson correlation coefficient was used to determine the strength of the association between each human caring assessment dimension and the test as a whole. An analysis was made of the distribution (mean, range, and standard deviation [*SD*]) of the total scores of the two groups (experts and students). Normality of the distribution was tested with Kolmogorov-Smirnov. Effect size of comparison was done using Cohen's *d*. The mean scores of both groups were compared with an independent samples *t* test, and Levene's statistic was used to test the hypothesis of homogeneity of variance in both samples. A nonparametric Mann–Whitney test was performed to validate the results. Differences were considered statistically significant at a level of $p < 0.05$. The statistical analyses were performed with SPSS software version 13.

Ethical Considerations

The study was approved by the Health Sciences Research Ethics Board of the Université de Montréal. The participants were free to decide to take the script concordance test, and their anonymity and confidentiality were respected.

Optimization of the Script Concordance Test

In optimizing the script concordance test, the first step was to analyze the distribution and variance of the test responses given by the experts recruited for construction of the scoring grid. Three questionnaires were excluded from the statistical analyses as atypical; the responses were, in fact, the opposite of those provided by all of the other panel members because either the instructions or the script concordance test items were interpreted incorrectly. Optimization of the expert group resulted in the final reference panel of 12 members. The second step was to study the individual script concordance test items. Using the responses to the test provided by the two groups, the items were analyzed to see whether any had been problematic. Items with a weak ($r < 0.05$) or negative item total correlation were excluded from the statistical analyses. Of the 92 items in the preliminary form of the test, 19 were eliminated. After optimization of the panel and the script concordance test items, the test comprised 29 clinical scenarios and

73 items. The statistical analyses were then performed.

RESULTS

Sociodemographic Data

Thirty first-year bachelor of nursing students at the Université de Montréal participated in the study; 27% were younger than 20 years, 60% were between 20 and 25 years, and 13% were older than 25 years. Of the 12 experts whose responses were used to construct the scoring grid, 42% worked in direct patient care in various Montreal-area hospitals, 33% in teaching and education, 17% in management, and 8% in research.

TABLE 4

Descriptive Analyses of the Distribution of Script Concordance Test Scores for Students and Experts

Results	Mean	*SD*	Minimum	Maximum	Effect Size
Total test					
Students	53.3	7.2	35.8	64.2	1.5
Panel	61.6	3.1	54.9	65.4	
Human aspect					
Students	21.4	3.5	9.7	26.6	1.5
Panel	25.7	2.0	20.7	27.8	
Clinical activities					
Students	15.0	2.5	10.4	18.3	1.2
Panel	17.2	0.9	15.8	18.4	
Therapeutic relationship					
Students	16.9	2.4	11.9	20.2	1.0
Panel	18.7	1.1	16.9	20.7	

SD = *standard deviation.*

Psychometric Results

After optimization, the script concordance test scores were found to have a good index of reliability, with a Cronbach's alpha coefficient of 0.86. The alpha coefficient was also calculated for the three human caring dimensions (**Table 2**). A significant positive correlation was observed between the scores for each human caring assessment dimension and those for the test as a whole (**Table 3**). We also found a statistically significant linear relationship between the different human caring assessment dimensions. Thus, the human aspect correlated with the therapeutic relationship ($r = 0.67$, $p < 0.01$); the therapeutic relationship with clinical activities ($r = 0.50$, $p < 0.01$); and the human aspect with clinical activities ($r = 0.54$, $p < 0.01$).

An analysis was also made of the distribution and variance of the two groups' scores (**Table 4**). The distribution was found to be normal for the expert group ($p = 0.20$) but not for the student group ($p = 0.02$). Levene's test revealed a significant difference in group variance ($F = 8{,}5$; $p < 0.01$). Appropriate t test procedure was then used to compare mean scores. Students' scores (53.3; $SD = 7.2$) were significantly lower than the experts' scores (61.6; $SD = 3.1$) ($t = 5.2$; $p < 0.01$). The students' scores display greater variability, thus evidencing the capacity of the test to place the students along a continuum. The **Figure** clearly shows difference in both distributions.

The scores for the tests and for each human caring assessment dimension were also compared. There is a significant difference between the scores of the two groups for each dimension: the human aspect of care (25.7 versus 21.4; $t = 3.9$, $p < 0.01$), clinical activities (17.2 versus 15; $t = 4.2$ $p < 0.01$), and the therapeutic relationship (18.7 versus 16.9; $t = 3.2$, $p = 0.02$). The results were validated with a Mann–Whitney nonparametric test. Even in the presence of non-normality in the distribution of the students' concordance scores, the same statistically significant differences between the two groups were observed for the total score ($Z = -3.93$, $p < 0.01$), the human aspect ($Z = -2.84$, $p < 0.01$), clinical activities ($Z = -4.07$, $p < 0.01$), and the therapeutic relationship ($Z = -2.03$, $p = 0.04$).

DISCUSSION AND IMPLICATIONS FOR NURSING EDUCATION

The results of this study indicate that it is legitimate to assess nursing clinical reasoning even for attributes as hard to quantify as caring. A script concordance test is appropriate for situations of nursing practice in which there is a degree of uncertainty. It explores a student's capacity to generate hypotheses and to decide on intervention options in a context in which nurses cannot fall back on prescribed formulas or truths to justify their practices. In this study, the difference between the scores of the students and those of the experts was found to be statistically significant. Thus, a linear relationship exists between results in a script concordance test and in a clinical experience. These findings are similar to those of various prior medical studies. The large number of items retained for the statistical analysis ($n = 73$) helped optimize the measure's reliability, and the test scores have a high index of reliability (Cronbach's alpha = 0.86). The internal consistency of a script concordance test scores would be better if it includes 20 to 30 scenarios, each followed by three to five question items (Gagnon, Charlin, Lambert, Carrière, & van der Vleuten, 2009).

Calculation of the correlation coefficient revealed that the three human caring assessment dimensions correlate strongly with the test as a whole. Each dimension has a role in assessing human caring. The moderate positive correlations that were found between the dimensions are indicative of a moderate colinearity (i.e., a level of dependence among them) and suggest that the three dimensions of human caring are interdependent

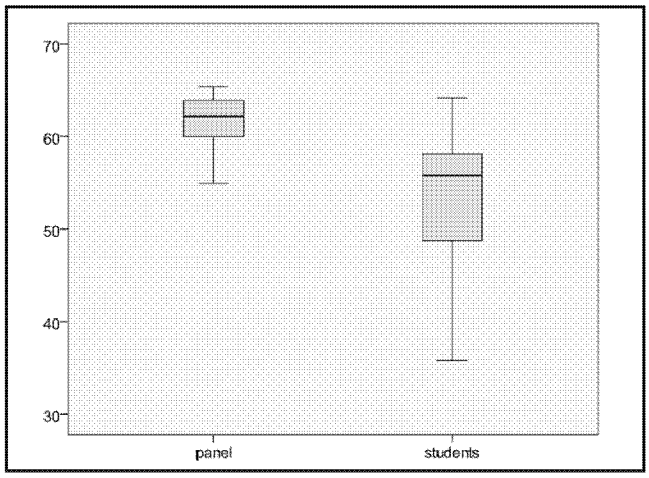

Figure. Distribution of panel members' and students' scores.

and go hand-in-hand with the delivery of quality care. This study reinforces the point that a nurses' values and beliefs, which are associated with a philosophy of care, will find practical expression in their attitudes, reflective processes, and actions. With a goal to humanize and improve the quality of care, teaching institutions and facilities ought to position themselves in a school of thought particular to the discipline and mold nursing care accordingly.

Furthermore, knowledge of the theoretical foundations of clinical reasoning provides a guide for introducing effective pedagogical methods to help students develop their mental scripts. When applied early and experientially based, such scripts will be stimulated and expanded by didactic methods that simultaneously promote analytical and hypothetic-deductive processes (Charlin, Boshuizen, et al., 2007; Nendaz et al., 2005). Examples of this include studying prototypical and atypical cases and elaborating on maps and abstracts, along with practicing early and regularly in clinical settings. Given the current constructivist trend in education, assessment is of great use as a strategy for learning and teaching and can be applied to determine how students process information using an integrated stock of knowledge (Nendaz et al., 2005). Script concordance tests are a good way to accomplish this task and expand the variety of measures available for assessing clinical competence.

Several limitations to this study should be noted. During the preliminary analyses, three of the collaborating experts had a low score, necessitating their exclusion from the study and raising questions as to the selection of the participants for the reference panel. A study by Gagnon et al. (2005) determined the minimum number of experts needed for such panels and found that 10 to 15 easily ensure the accuracy of the scoring grid. The variability in the answers provided by reference panel experts was analyzed by Charlin et al. (2006). They observed that the questions that make it possible to properly discern differing levels of expertise among study participants are the ones that display strong divergences in the panel members' responses ($SD = 1.5$; $F = 16.3$, $p < 0.001$). More recently, Charlin, Gagnon, et al. (2007) studied the makeup of reference panels and found a strong correlation between the data obtained from two panels (intraclass correlation coefficient = 0.98): one comprised of experts from a teaching faculty and a second with experts from private practice. Their findings make the inclusion of clinical experts on a reference panel legitimate inasmuch as their practice is constantly updated and evidences a high level of skill (Charlin, Gagnon, et al., 2007). However, other observations on setting the inclusion criteria for reference panels are readily justifiable; although experts vary somewhat in their reasoning in handling clinical situations, there should be a degree of homogeneity in panel membership. Measures will be needed to elucidate scientific and ethical questions surrounding the choice of reference panel experts.

Further studies on a range of questions are needed if this tool is to be recognized for the formative or summative assessment of student reasoning: How are a student's results to be interpreted during training? Is there a progression in students' performance on script concordance tests over the course of their years of study? At what value is student–expert concordance considered satisfactory? Does performance on a script concordance test correlate with other measures of clinical reasoning in nursing? In addition, clinical scenarios and test items should be written and reviewed to improve script concordance tests and apply them to particular fields of nursing practice. This study was limited to reasoning that is humanistic and altruistic. It would be useful to assess a nurse's reasoning in common practice situations in which he or she evaluates, analyzes, and decides on a treatment plan.

CONCLUSION

In-training assessment is no longer an occasional activity unrelated to teaching. It should address the way a student processes information by drawing on an integrated inventory of knowledge (Nendaz et al., 2005). Script concordance tests can be particularly useful for this purpose, a function that underscores the relevance and originality of this study. Script concordance tests are a valid, reliable, standardized tool for assessing reasoning based on human caring. They make it possible to assess the quality of students' organization of knowledge. More specifically, they evaluate their ability to make appropriate hypotheses and decisions about nursing interventions in professional practice. Script concordance testing in nursing is still in its infancy. Further study and analysis are needed to enhance its psychometric value and optimize its pedagogical utility and broader applicability.

REFERENCES

Benner, P., Tanner, C.A., & Chesla, C.A. (1996). *Expertise in nursing practice: Caring, clinical judgment and ethics.* New York, NY: Springer.

Cara, C. (2003). A pragmatic view of Jean Watson's theory. *International Journal of Human Caring, 7,* 51-61.

Carr, S.M. (2004). A framework for understanding clinical reasoning in community nursing. *Journal of Clinical Nursing, 13,* 850-857.

Carrière, B. (2005). *Development and initial validation of a script concordance test for residents in a pediatric emergency medicine rotation* (Unpublished master's thesis). University of Illinois, Urbana, IL.

Charlin, B., Bordage, G., & van der Vleuten, C. (2003). L'évaluation du raisonnement clinique. *Pédagogie Médicale, 4,* 42-52.

Charlin, B., Boshuizen, H.P., Custers, E.J., & Feltovich, P.J. (2007). Scripts and clinical reasoning. *Medical Education, 41,* 1178-1184.

Charlin, B., Gagnon, R., Pelletier, J., Coletti, M., Abi-Rizk, G., Nasr, C.,

et al. (2006). Assessment of clinical reasoning in the context of uncertainty: The effect of variability within the reference panel. *Medical Education, 40,* 848-854.

Charlin, B., Gagnon, R., Sauvé, E., & Coletti, M. (2007). Composition of the panel of reference for concordance tests: Do teaching functions have an impact on examinees' ranks and absolute scores? *Medical Teacher, 29,* 49-53.

Charlin, B., Gagnon, R., Sibert, L., & van der Vleuten, C. (2002). Le test de concordance de script: Un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale, 3,* 135-144.

Charlin, B., Roy, L., Brailovsky, C., Goulet, F., & van der Vleuten, C. (2000). The script concordance test: A tool to assess the reflective clinician. *Teaching and Learning in Medicine, 12,* 189-195.

Charlin, B., Tardif, J., & Boshuizen, H.P. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine, 75,* 182-190.

Charlin, B., & van der Vleuten, C. (2004). Standardized assessment of reasoning in contexts of uncertainty: The script concordance approach. *Evaluation and the Health Professions, 27,* 304-319.

Cossette, S., Cara, C., Ricard, N., & Pepin, J. (2005). Assessing nurse-patient interactions from a caring perspective: Report of the development and preliminary psychometric testing of the Caring Nurse–Patient Interactions Scale. *International Journal of Nursing Studies, 42,* 673-686.

Devlin, J.W., Marquis, F., Riker, R.R., Robbins, T., Garpestad, E., Fong, J.J., et al. (2008). Combined didactic and scenario-based education improves the ability of intensive care unit staff to recognize delirium at the bedside. *Critical Care, 12,* R19.

Duffy, J.R. (2005). Want to graduate nurses who care? Assessing nursing students' caring competencies. *Annual Review of Nursing Education, 3,* 59-76.

Elstein, A.S., Schulman, L.S., & Sprafka, S.A. (1978). *Medical problem solving: An analysis of clinical reasoning.* Cambridge, MA: Harvard University Press.

Feltovich, P.J., & Barrow, H.S. (1984). Issues of generality in medical problem solving. In H.G. Schmidt & M.L. De Volder (Eds.), *Tutorial in problem-based learning: A new direction in teaching the health professions* (pp. 128-142). Allen, Holland: van Gorcum.

Fonteyn, M. (1998). *Thinking strategies for nursing practice.* Philadelphia, PA: J.B. Lippincott-Raven.

Fonteyn, M.E., & Cahill, M. (1998). The use of clinical logs to improve nursing students' metacognition: A pilot study. *Journal of Advanced Nursing, 28,* 149-154.

Fonteyn, M.E., & Ritter, B.J. (2000). Clinical reasoning in nursing. In J. Higgs & M. Jones (Eds.), *Clinical reasoning in the health professions* (2nd ed, pp. 235-244). Oxford, England: Butterworth Heinemann.

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

Fournier, J.P., Thiercelin, D., Pulcini, C., Alunni-Perret, V., Gilbert, E., Minguet, J.M., et al. (2006). Évaluation du raisonnement clinique en médecine d'urgence: Les tests de concordance des scripts décèlent mieux l'expérience clinique que les questions à choix multiples à contexte riche. *Pédagogie Médicale, 7,* 20-30.

Gagnon, R., Charlin, B., Coletti, M., Sauvé, E., & van der Vleuten, C. (2005). Assessment in the context of uncertainty: How many members are needed on the panel of reference of a script concordance test? *Medical Education, 39,* 284-291.

Gagnon, R., Charlin, B., Lambert, C., Carrière, B., & van der Vleuten, C. (2009). Script concordance testing: More cases or more questions? *Advances in Health Science Education, 14,* 367-375.

Goransson, K.E., Ehnfors, M., Fonteyn, M.E., & Ehrenberg, A. (2008). Thinking strategies used by registered nurses during emergency department triage. *Journal of Advanced Nursing, 61,* 163-172.

Grant, J., & Marsden, P. (1988). Primary knowledge, medical education and consultant expertise. *Medical Education, 22,* 173-179.

Greenwood, J., Sullivan, J., Spence, K., & McDonald, M. (2000). Nursing scripts and the organizational influences on critical thinking: Report of a study of neonatal nurses' clinical reasoning. *Journal of Advanced Nursing, 31,* 1106-1114.

Grobe, S.J., Drew, J.A., & Fonteyn, M.E. (1991). A descriptive analysis of experienced nurses' clinical reasoning during a planning task. *Research in Nursing and Health, 14,* 305-314.

Higgs, J., & Jones, M. (2000). Clinical reasoning in the health professions. In J. Higgs & M. Jones (Eds.), *Clinical reasoning in the health professions* (2nd ed., pp. 3-14). Oxford, England: Butterworth Heinemann.

Kautz, D., Kuiper, R.A., Bartlett, R., Buck, R., Williams, R., & Knight-Brown, P. (2009). *Building evidence for the development of clinical reasoning using a rating tool with the Outcome-Present State-Test (OPT) model.* Retrieved from http://snrs.org/publications/SOJNR_articles2/Vol09Num01Art15.pdf

Kautz, D.D., Kuiper, R., Pesut, D.J., Knight-Brown, P., & Daneker, D. (2005). Promoting clinical reasoning in undergraduate nursing students: application and evaluation of the Outcome Present State Test (OPT) model of clinical reasoning. *International Journal of Nursing Education Scholarship, 2,* Article 1.

Kérouac, S., Pepin, J., Ducharme, F., & Major, F. (2003). *La pensée infirmière* (2nd ed). Laval, Quebec, Canada: Beauchemin.

Kuiper, R.A., & Pesut, D.J. (2004). Promoting cognitive and metacognitive reflective reasoning skills in nursing practice: Self-regulated learning theory. *Journal of Advanced Nursing, 45,* 381-391.

Lambert, C. (2005). *Le test de concordance de scripts: Un outil pour évaluer le raisonnement clinique des résidents en radio-oncologie* (Unpublished master's thesis). Université de Montréal, Montreal, Quebec, Canada.

McCarthy, M.C. (2003a). Detecting acute confusion in older adults: Comparing clinical reasoning of nurses working in acute, long-term, and community health care environments. *Research in Nursing and Health, 26,* 203-212.

McCarthy, M.C. (2003b). Situated clinical reasoning: Distinguishing acute confusion from dementia in hospitalized older adults. *Research in Nursing and Health, 26,* 90-101.

Nendaz, M., Charlin, B., Leblanc, V., & Bordage, G. (2005). Le raisonnement clinique: Données issues de la recherche et implications pour l'enseignement. *Pédagogie Médicale, 6,* 235-254.

Norcini, J.J., Shea, J.A., & Day, S.C. (1990). The use of aggregate scoring for a recertifying examination. *Evaluation and the Health Professions, 13,* 241-251.

Norman, G.R. (1985). Objective measurement of clinical performance. *Medical Education, 19,* 43-47.

Norman, G.R. (1993). *Theoretical and psychometric considerations. Report on the evaluation system for specialist certification. Task force of the evaluation committee.* Ottawa, Ontario, Canada: Royal College of Physicians and Surgeons of Canada.

Page, G., Bordage, G., & Allen, T. (1995). Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine, 70,* 194-201.

Schmidt, H.G., Norman, G.R., & Boshuizen, H.P. (1990). A cognitive perspective on medical expertise: Theory and implication. *Academic Medicine, 65,* 611-621.

Sibert, L., Charlin, B., Corcos, J., Gagnon, R., Grise, P., & van der Vleuten, C. (2002). Stability of clinical reasoning assessment results with the script concordance test across two different linguistic, cultural and learning environments. *Medical Teacher, 24,* 522-527.

Sibert, L., Charlin, B., Gagnon, R., Corcos, J., Khalaf, A., & Grise, P. (2001). Évaluation du raisonnement clinique en urologie: L'apport du test de concordance de script. *Progrès en Urologie, 11,* 1213-1219.

Simmons, B. (2002). *Clinical reasoning in experienced nurses* (Unpublished master's thesis). University of Loyola, Chicago, IL.

Simmons, B., Lanuza, D., Fonteyn, M., Hicks, F., & Holm, K. (2003). Clinical reasoning in experienced nurses. *Western Journal of Nursing Research, 25,* 701-719.

Watson, J. (1988). New dimensions of human caring theory. *Nursing Science Quarterly, 1,* 175-181.

Watson, J. (1997). The theory of Human Caring: Retrospective and prospective. *Nursing Science Quarterly, 10,* 49-52.

Watson, J. (2002). Intentionality and caring-healing consciousness: A practice of transpersonal nursing. *Holistic Nursing Practice, 16,* 12-19.