

Context impact of clinical scenario on knowledge transfer and reasoning capacity in a medical problem-based learning curriculum

A. Collard^a, S. Brédart^b and J.-P. Bourguignon^{a,c}

^aInstitute of Teaching and Research in Higher Education, University of Liège, Liège, Belgium; ^bDepartment of Psychology: Cognition and Behaviour, University of Liège, Liège, Belgium; ^cDepartment of Clinical Science, University of Liège, Liège, Belgium

ABSTRACT

Since 2000, the faculty of Medicine at the University of Liège has integrated problem-based learning (PBL) seminars from year two to seven in its seven-year curriculum. The PBL approach has been developed to facilitate students' acquisition of reasoning capacity. This contextualized learning raises the question of the de- and re-contextualization process and its development during medical education. The aim is to assess transfer capacity of knowledge acquired through contextualized learning. A script concordance test (SCT) to assess clinical reasoning was administered to 104 volunteer students from year three to six of the Faculty of Medicine, and to a reference panel. The SCT has been created to measure, in a standardized way, reasoning capacity for ill-defined clinical problems. We used clinical cases from PBL seminars related to endocrinology from the third- and fifth-year (APP [Apprentissage Par Problèmes] and ARC [Apprentissage du Raisonnement Clinique] seminars). Two SCT scenarios aiming at the same learning objectives were created: one in the original learning context and one in a different, new context. Irrespective of the learning context, the reference panel showed higher scores than all groups of students. Students in year five and six obtained higher scores than younger students. A significant difference was also found in relation to the context (original/new) of the question. Higher scores were unexpectedly obtained for questions related to ARC cases than APP cases and in new contexts versus original ones, irrespective of the study level. The highest score was obtained for fifth-year cases in a new context. This study shows that from third year, students are able to transfer the knowledge acquired throughout PBL seminars for application to questions with information out of the learning context. Moreover, reasoning capacity grows with experience, irrespective of the process of transfer.

ARTICLE HISTORY

Received 26 January 2014
Accepted 14 December 2014

KEYWORDS

assessment; knowledge;
medical education; problem-
based learning

Introduction

One of the most important learning aims is to develop the capacity to use knowledge acquired in one context to solve a new dissimilar problem in another context. This

phenomenon is identified in cognitive psychology as transfer capacity. It is a real difficulty to use basic knowledge acquired during the learning process to solve or explain a new problem. In accordance with this known difficulty, Norman (2009, p. 808) wrote 'typically, students who have learned a concept in one problem context will only have a 10–30% success rate in applying the concept to solve a new problem'. Johnson, Dixon, Daugherty, and Lawanto (2011) have reported the inability of students to recognize the transferability of concepts learned from solving well-structured problems in the classroom to ill-structured problems faced outside of the classroom. This observation raises the question of whether the lack of transfer capacity results from absent or insufficient training, or inadequacy of learning conditions as compared to problem-solving conditions. The latter hypothesis is supported by research on the capacity of generalization or transfer where the issue of factors influencing access to memory has been addressed. The results of these studies show that the environment in which learning occurs: the room; the time of the day; the surrounding noise, ... (McGeoch, 1932; Smith, 1988; Smith & Vela, 2001) or the characteristics of the learners, such as experience or knowledge (Eva, Neville, & Norman, 1998), as well as emotions, can influence the retrieval of information from memory. This phenomenon was identified during the 1970s by Tulving and Thomson (1973) as 'context specificity' (Smith & Vela, 2001).

In medical education, it is not uncommon that students, after passing exams successfully, are still unable to apply the acquired knowledge to solve a new problem (Norman, 2009). Such observations have led to revised educational strategies like the implementation of problem-based learning (PBL) throughout the medical curriculum with the aim of better integration of basic science with clinical problems. The specificity of PBL is learning from clinical cases, including basic sciences, in other words contextualization of learning. Using clinical cases in PBL gives students the opportunity to be closer to the real professional world and, therefore, can enhance students' general learning (Herbert & Burt, 2001) transfer and problem-solving skills (Hmelo-Silver, 2004). The cognitive approach has largely demonstrated the positive impact of integrated learning on storage and retrieval of information and also on reasoning and problem-solving capacity (Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Farr, 1988; DeGroot, 1965; Lesgold et al., 1988; Newell & Simon, 1972). When students are more actively involved in learning – exploring by themselves the learning material, collecting and analysing data – they tend to have more stable long-term memory structures and a higher level of memory performance (Herbert & Burt, 2001).

Since 2000, the faculty of Medicine at the University of Liège has integrated PBL seminars from year two to seven, in its seven-year curriculum (Boniver, 2004). The PBL approach has been developed to facilitate students' process of acquisition, organization and retrieval of knowledge (Norman & Schmidt, 1992; Regehr & Norman, 1996; Schmidt, 1983; Van Gessel, Nendaz, Vermeulen, Junod, & Vu, 2003).

PBL, however, still faces the question of the difficulty of knowledge transfer after being acquired in a specific context. Thus, it is up to the individual student to extract knowledge from clinical cases and transfer that knowledge to new contexts. The student has to retrieve a learned concept to apply that concept in understanding and solving a new problem. Thus, the difficulty not only involves extraction and transfer of information but also lies in the mental representation, the abstraction and organization of the knowledge base. To use a concept for solving a problem, the common features that the new

problem shares with a concept acquired must be recognized (Norman, 2009). With the experience of a concept in varying contexts, the knowledge is more likely to become decontextualized, not dependent on contextual cues, and understanding is at a more abstract level (Herbert & Burt, 2001). Expert problem-solvers remember the underlying conceptual structure that makes two problems similar although they have different surface features. By contrast, novice problem-solvers tend to prioritize surface similarities between problems (Sutton, 2003). This ability makes the transfer of concepts easier because of the conscious effort to abstract knowledge and concepts from one context for application to another (Johnson, 1995). Sometimes the new problem context has clues that do exactly match a problem already solved but more frequently, it does not look analogous on the surface. Then, the transfer task is more difficult because in order to identify the analogies, the process must take place at the level of the deep structure of the problem (Norman, 2009). As an example, a systolic murmur at the aortic valves will be interpreted by the novice as a turbulence of blood flow generating a murmur like in stenosis of other heart valves. In contrast, the expert will rapidly think in terms of consequences of malfunction of this particular valve for pressure upstream in the left heart and flow rate and blood supply downstream in organs including heart.

In the cognitive perspective of learning process, it is proposed that learning progresses through interpretation of new knowledge in the light of what is already known (Norman, 2009). A well-organized knowledge base and efficient integration of any new learning with what a student already knows are predictors of the transfer capacity (Dixon & Brown, 2012). Without an extensive and organized knowledge base, transfer strategies are less likely to operate outside the instructional context since there is nothing, or too little, to connect together (Haskell, 2001). Experts not only have an appropriate knowledge base in their long-term memory, but they also have the ability to access and use that knowledge. An important difference between experts and novices is how their knowledge base is organized (Glaser, 1987). A structured knowledge enables experts to 'see into' a problem more quickly than novices. Further, this difference enables them to more precisely encode and store information. As a consequence, access to and retrieval of relevant information is more efficient. Haskell considers that knowledge base is an absolute requirement not only for transfer but also for thinking and reasoning. Along the same line, Galloti (1989, p. 338) concluded:

it appears that some of what predicts good everyday reasoning is the breadth and depth of the knowledge base. The knowledge that one possesses affects the type of mental models that one can construct as well as the type of problems that can even be recognized.

In the medical domain, Regehr and Norman (1996) suggested that expertise is characterized by 'scripts', that is, construction of knowledge networks that connect specific experiences (contextualized information), and/or abstract concepts (de-contextualized information). The cognitive approach identifies major differences between experts' scripts and novices' scripts. With experience, physicians develop more extensive and better organized scripts (Custers, Boshuizen, & Schmidt, 1998; Feltovitch & Barrows, 1984; Schmidt, Norman, & Boshuizen, 1990). Expertise also enhances pattern of information perception, and then, activation of scripts (Custers, Boshuizen, & Schmidt, 1996). Some argued that experienced physicians show superior performance in relevant script activation from prior information available at the first step of clinical reasoning

(Elstein, Shulman, & Sprafka, 1978), even with limited information (Hobus, Boshuizen, & Schmidt, 1990; Hobus, Hofstra, Boshuizen, & Schmidt, 1988; Hobus, Schmidt, Boshuizen, & Patel, 1987; Hofstra, Hobus, Boshuizen, & Schmidt, 1988). Finally, with experience, the capacity of information processing progresses, in rapidity and accuracy. According to knowledge and experience, experts are able to filter relevant information and then prioritize it. With PBL settings, the aim is to help the student integrate new information in a rich and connected knowledge network that would be later activated (illness scripts) (Barrows, 1986). PBL promotes transfer skills, on top of clinical reasoning skills (Dixon & Brown, 2012; Hmelo-Silver, 2004).

The present study addresses, in a PBL setting, the de- and re-contextualization process and its development during medical education. The aim is to assess the capacity to use knowledge acquired through contextualized learning into a new reasoning task. We have hypothesized that, through the multiple opportunities to encounter knowledge in different contexts, PBL makes decontextualized concepts more usable and improves that transfer capacity. During the curriculum, students are exposed to different learning modules involving endocrinology both in APP (Apprentissage Par Problèmes) seminars starting from second year and through lectures (see Table 1). Are students able to use that knowledge in new tasks and from what stage in that cursus? The investigation tools involved a script concordance test (SCT) (Caire, Sol, Charlin, Isodori, & Moreau, 2004; Charlin, Gagnon, Sibert, Van der Vleuten, 2002; Charlin, Roy, Brailovsky, Goulet, & Van der Vleuten, 2000; Charlin & St-Jean, 2002; Charlin & Van Der Vleuten, 2004) to assess clinical reasoning throughout the curriculum (year two to six). The SCT has been created to measure, in a standardized way, reasoning capacity for ill-defined clinical problems.

Methods

Participants

The study was based on 104 volunteer students from years three to six of the Faculty of Medicine at the University of Liège (with a minimum of 20 students for each study year). Endocrinology was the selected topic of the investigations. The students differed according to the PBL seminars and the clerkship period they had attended (Table 1). In the second and third years, the students learned the structure and normal functioning of the human body systems, through APP seminars. In the clinical cases used for APP seminars, reasoning aimed at identifying among the observations (history, examination, complementary assessment, treatment) those that needed to be explained and providing relevant explanations. In the clinical cases used for ARC (Apprentissage du

Table 1. Characteristics of the participants.

Year of the curriculum	3rd	4th	5th	6th
Training completed (endocrine system) at the time of study	APP seminars	APP seminars	APP and ARC seminars	APP and ARC seminars
Lectures on diagnosis and treatment of endocrine disorders	–	–	Done	Done
<i>N</i> participants/ <i>N</i> total	35/114	20/100	25/96	24/92
Reference panel	9 tutors involved in APP and ARC seminars about the endocrine system			

Raisonnement Clinique) seminars, the aim of reasoning was much more focused: obtaining information (history, examination, complementary assessment) to raise diagnostic hypotheses as early as possible and prioritizing one hypothesis as a proposed diagnosis. In years four and five, the diagnostic and therapeutic bases of diseases of the different systems were learned through lectures and ARC seminars. Written informed consent was obtained from the participants who responded to the test questions anonymously.

Among the tutors handling the PBL seminars, nine served as a reference panel for SCT scoring (see below). Two teachers wrote the test questions.

Testing procedure

Originally, an SCT question consisted of a short and incomplete clinical scenario which is challenging even for an expert. Following the scenario, a diagnostic or management hypothesis was proposed, and then new information was provided. The responders have to assess the impact of this new information on the initial diagnostic or management hypothesis. The response is given on a 5-point Likert scale. Each individual response is credited by a score calculated on the basis of the distribution of responses given by a reference panel (Charlin et al., 2002; Charlin, Roy, et al., 2000; Charlin, Tardif, & Boshuizen, 2000).

While keeping the original SCT format, we adapted the test to assess skills of clinical reasoning and knowledge transfer in young students not yet familiar with the diagnostic process. We used clinical cases from PBL seminars related to endocrinology: five cases from third-year seminars (APP) and five cases from fourth-year seminars (ARC). For each clinical case, two SCT scenarios aiming at the same learning objectives were created: one in the original learning context and one in a different, new context. The 'Original Context' scenarios involved information coming only from the clinical cases of PBL seminars used for learning purposes, without any irrelevant information. In addition, the physiological function was disturbed in the same way as in the learning case. The 'New Context' scenarios involved information that was new (different from the learning clinical cases) and mixed with some irrelevant information (i.e., that was not usable for reasoning in this particular case). The physiological function was disturbed differently from the learning case. Table 2(a and b) gives an example of these two series of questions. For each scenario, three independent hypotheses were proposed, each followed by additional new information.

The present study was based on 48 SCT questions, including 22 APP and 26 ARC questions. Those 48 questions were validated by the reference panel, whereas 12 questions were discarded based on two exclusion criteria (ambiguous or confusing writing and/or content out of the learning objectives).

Scoring of SCT

The original scoring method developed by Charlin and collaborators was 'the aggregate scoring process' (Charlin et al., 2002; Charlin & Van Der Vleuten, 2004). The participants' response to each item is compared with the response of the panel members. The score was attributed to each responder based on how many of panel members have chosen the same

Table 2. Examples of SCT questions in 'original context' (a) and in 'new context' (b) with the element of the original context in italics.*(a) A young women with Hashimoto's thyroiditis treated by 75 µg/day of L-thyroxine complains about tiredness, loss of hair and intolerance to coldness*

If the hypothesis is	And you are informed that	The hypothesis becomes	Answer of ref. panel
The dose of thyroxine is insufficient	Reverse T3 serum levels are in the normal range	Much less likely	–
		Less likely	3
		Not affected	6
		More likely	–
		Much more likely	–

(b) A young man is admitted in the emergency room for drug intoxication. He could have taken 12 tablets (6 g as a whole) of aspirin and 10 tablets (2 mg as a whole) of L-thyroxine. His heart rate is 72/min.

If the hypothesis is	And you are informed that	The hypothesis becomes	Answer of ref. panel
The patient took less medication than suspected	The medications were taken 1 hour before admission	Much less likely	1
		Less likely	7
		Not affected	1
		More likely	–
		Much more likely	–

response. A maximum score of 1 was given for the modal answer. Other responses are given a partial credit depending on the number of panel members who provided that answer. For example, on a given item, if eight (out of 10) members of the reference panel responded 'much less likely' on the Likert scale, this option was credited by the maximum score of 1 point (8/8, i.e., the modal response). Then, if two members selected 'no effect' as response, this option was credited 0.25 point (2/8). The answers not selected by any of the panel members received no credit whether they were close to or far from the modal answer.

That scoring process was adapted for this study using the kappa coefficients methodology (Vanbelle, Massart, Giet, & Albert, 2007). Weighted kappa coefficients (Cohen, 1968) quantified the agreement between two raters on an ordinal scale and corrected for agreement due to chance. This methodology was extended to quantify the agreement between the reference panel and a responder. The weights permitted to assign higher scores to responses closer to the modal response(s) given by the reference panel. The weights were derived from the reference panel responses by computing the frequency distribution of the number of category differences separating two experts' responses of all possible pairs of experts. This led to a weight equal to 1 given for a category of the Likert scale consistent with the modal response of the reference panel, a weight of 0.96 for the adjacent categories and weights of 0.19, 0.06 and 0 at a distance of 2, 3 and 4 categories, respectively. Cohen's kappa coefficient in the reference panel was also calculated (Collard et al., 2009).

Statistical analysis

For each category of SCT question (APP Original Context, APP New Context, ARC Original Context and ARC New Context), the performances were determined using a modified scoring method (Vanbelle et al., 2007). The results were expressed as mean ($\pm 95\%$ CI). The effect of experience (curriculum level) of the responders was tested using a two-way ANOVA, with activity (APP vs. ARC) and context (Original vs. New) as within factors.

Planned comparisons were used to determine significant differences. The results were considered to be significant at the 5% probability level ($P < .05$).

Results

SCT scores

The mean scores of the five groups (students of years three to six and referents) for the four categories of questions are given in Table 3. The mean scores were significantly different among the five groups ($F(4, 108) = 12.56$; $P < .0001$). More specifically, the reference panel showed higher scores ($P < .05$) than all groups of students; students in years five and six obtained higher scores ($P < .05$) than younger students (years three and four). Significant differences were found based on the level in the curriculum and the type of seminar (APP/ARC) to which the question was related ($F(1, 108) = 5.28$; $P < .05$). A significant difference was also found in relation to the context (Original/New) of the question ($F(1, 108) = 7.97$; $P < .001$). Higher scores were unexpectedly obtained at ARC than at APP and in new contexts versus original ones, irrespective of the study level (Figure 1). Also, statistical analysis revealed a significant relation between the type of seminar (APP/ARC) and the context of the question (Original/New) ($F(1, 112) = 5.65$; $P < .05$). Planned comparisons showed a statistically significant difference between ARC-related questions in a new context and the three other categories of questions ($P < .05$).

Discussion

This study shows that as early as in the third year of the curriculum, students who have experienced PBL seminars for one and a half years already have good reasoning and transfer capacity. During the subsequent years, the SCT scores increase throughout the curriculum. The latter findings are consistent with the observations made later during training and showing an increase in SCT scores between postgraduate students and expert physicians (Caire et al., 2004; Charlin & St-Jean, 2002; Charlin, Roy, et al., 2000; Charlin et al., 2002; Charlin & Van Der Vleuten, 2004). We have reported previously the increase in reasoning maturation throughout the curriculum that was found to be correlated with knowledge base as well as the capacity of core knowledge self-estimation (Collard et al., 2009).

Table 3. Mean score and standard deviation (SD) in the different groups for the four categories of SCT questions.

Group	APP questions		ARC questions		All questions Mean (SD)
	Original context Mean (SD)	New context Mean (SD)	Original context Mean (SD)	New context Mean (SD)	
Year 3 ($n = 35$)	0.55 (0.09)	0.58 (0.05)	0.54 (0.10)	0.61 (0.07)	0.57 (0.05)
Year 4 ($n = 20$)	0.51 (0.11)	0.55 (0.08)	0.49 (0.11)	0.67 (0.08)	0.55 (0.07)
Year 5 ($n = 25$)	0.61 (0.07)	0.64 (0.07)	0.69 (0.08)	0.73 (0.05)	0.67 (0.04)
Year 6 ($n = 24$)	0.63 (0.08)	0.61 (0.07)	0.57 (0.11)	0.74 (0.06)	0.64 (0.05)
Reference panel ($n = 9$)	0.76 (0.09)	0.76 (0.08)	0.79 (0.08)	0.84 (0.06)	0.79 (0.03)

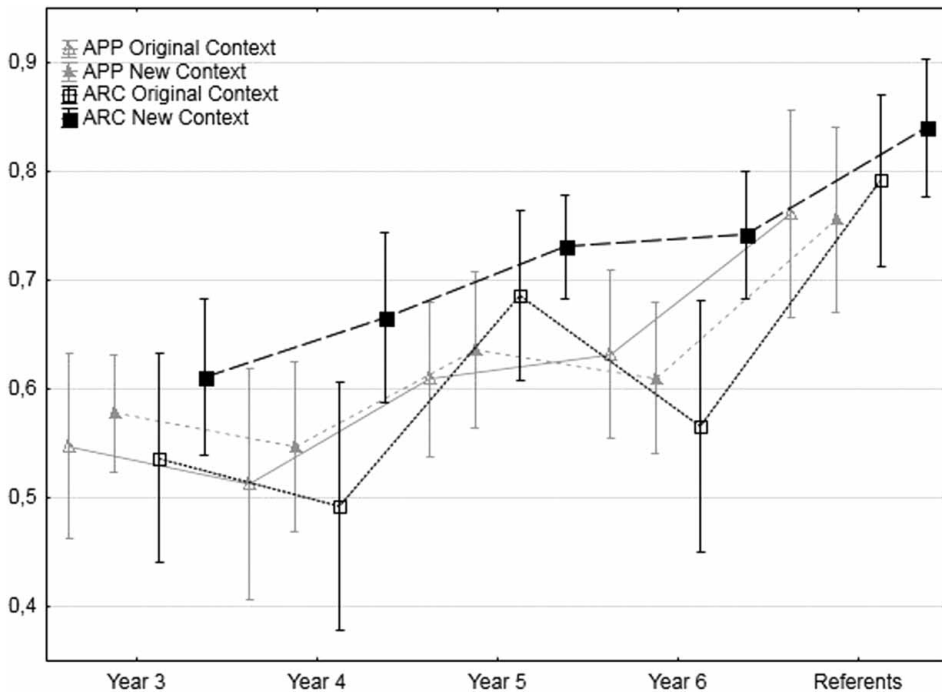


Figure 1. Changes in SCT Score (mean \pm % confidence interval) throughout the curriculum. Note: The data are shown for APP and ARC questions, and in each category the scores are compared for questions in the original learning context and in a new context.

At each level in the curriculum, the students are able to transfer knowledge acquired during learning into a new context. A specific effect of PBL seminars could not be determined since we could not compare with students trained without PBL seminars. Norman (2009) has argued that substantial gain can be observed when students are engaging in multiple teaching examples. The students are exposed to multiple contexts in learning when it includes examples demonstrating wide application of what is being taught, as in a PBL curriculum. Repeated experiences with the same information but in rather different context result in a strengthening of the knowledge that is common to the varying problem representation (Herbert & Burt, 2001, 2003, 2004). In such instances, students develop a flexible representation of knowledge and are likely to abstract the relevant features of concepts that make two unique problem scenarios similar (Gick & Holyoak, 1983; Herbert & Burt, 2001; Spiro, Vispoel, Schmitz, Samarapungavan, & Boerger, 1987). To train transfer capacity, PBL could be a relevant method because, as Norman stated: to improve transfer, you have to link the concept learned to what is already known, use multiple examples to identify common deep structure and use mixed practice with multiple examples to focus on identifying when a concept applies. According to Spiro, Feltovitch, Jackson, and Coulton (1991), the difficulty in transferring concepts that students learn in schools to the real world is due to the over simplification of educational material to make it easier to teach. PBL does not suffer from a lack of context or an oversimplification of content. Learning through problem-based activities can enhance students' general learning transfer and problem-solving skills (Dixon & Brown, 2012; Hmelo-Silver, 2004).

Besides the expected findings that were described above, some unexpected observations were made. First, the scores on ARC questions were higher than on APP questions for the younger students (years three and four) though they had never attended ARC seminars. Second, the scores obtained for questions in a new context were higher than questions in the original learning context. The highest scores were obtained for new context ARC questions whatever the group. Several explanations could be proposed. The ARC scenarios of SCT were written on the basis of clinical cases discussed in the ARC seminars centred on the diagnostic process while APP cases were constructed to meet the learning objectives of basic sciences. So, the realistic nature of ARC cases might result in greater activation of reasoning than APP cases.

The hypotheses proposed in SCT questions were, most of the time, different from the hypotheses discussed in PBL seminars. Moreover, the additional information to be evaluated was out of the original learning context for all questions. Thus, the lower scores obtained for original context questions could also involve possible interference created by new information in SCT questions. A proactive interference could be created by the first learning experience and entail difficulties for students when they analyse a familiar case with discordant elements compared to the learning context (Bernstein, Penner, Clarke-Stewart, & Roy, 2007). Transfer between tasks is a function of the similarity between the tasks and learning experiences (Dixon & Brown, 2012). Transfer is, therefore, affected by the context of the original learning; so processing new hypotheses and new additional information in an already known context (i.e., the original learning context) could be more difficult for students than treating new information in a new context without interference. With experience, the clinicians develop a more flexible representation of knowledge and are likely to abstract the relevant features of concepts. Thus, they are less influenced by contextual surface elements of a case. The interference phenomenon is, therefore, not so important for the tutor group.

In addition, even in a new context, if the responders generated a given hypothesis based on the SCT scenario while the SCT proposed another hypothesis less likely for the student, it could be difficult for them to ignore the hypothesis they generated when processing the new information. Considering these possible biases in writing test material, the SCT questions used in further research could be selected on the basis of realism or plausibility of both the SCT scenarios and hypotheses. A research design with think-aloud protocol or a writing justification of each response could provide an explanation of the features that influence reasoning and transfer capacity.

Conclusion

This study shows that from the third year, students are able to use the knowledge acquired through PBL seminars for application to questions with information out of the learning context. Moreover, reasoning capacity grows with experience, irrespective of the process of transfer.

Acknowledgements

The authors thank the reference panel, the medical students who kindly volunteered to participate in this study and Sophie Vanbelle for the supervision of statistical analysis.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors acknowledge the Belgian Fonds de la Recherche Fondamentale Collective who have supported the research project [grant numbers G.4571.06 and 2.4593.08].

References

- Barrows, H. (1986). The scope of clinical education. *Journal of Medical Education*, 61(9), 23–33.
- Bernstein, D., Penner, L.A., Clarke-Stewart, A., & Roy, E.J. (2007). *Psychology* (8th ed.) (pp. 262–263). Boston: Cengage Learning.
- Boniver, J. (2004). Les études de médecine à l'Université de Liège: Le renouveau pédagogique de la Faculté de Médecine. *Revue Médicale de Liège*, 59(12), 717–730.
- Caire, F., Sol, J.C., Charlin, B., Isodori, P., & Moreau, J.J. (2004). Le test de concordance de script (TCS) comme outil d'évaluation formative des internes en neurochirurgie: Implantation du test sur Internet à l'échelle nationale. *Pédagogie Médicale*, 5(2), 87–94.
- Charlin, B., Gagnon, R., Sibert, R., & Van der Vleuten, C. (2002). Le test de concordance de script, un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale*, 3(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(4), 189–195.
- Charlin, B., & St-Jean, M. (2002, Mai). Le test de concordance de script: Un outil pour évaluer le jugement en médecine. *Bulletin du CEFES, Université de Montréal*, 6, 4–5.
- Charlin, B., Tardif, J., & Boshuizen, P. (2000). Scripts and medical diagnosis knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine*, 75(2), 182–190.
- Charlin, B., & Van der Vleuten, C. (2004). Standardized assessment in context of uncertainty: The script concordance approach. *Evaluation in Health Profession*, 27(3), 304–319.
- Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152.
- Chi, M., Glaser, R., & Farr, M. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement of partial credit. *Psychological Bulletin*, 70(4), 213–220.
- Collard, A., Gelaes, S., Vanbelle, S., Bredart, S., Defraigne, J.O., Boniver, J., et al. (2009). Reasoning versus knowledge retention and ascertainment throughout a problem-based learning curriculum. *Medical Education*, 43(9), 854–865.
- Custers, E., Boshuizen, H., & Schmidt, H. (1996). The influence of medical expertise, case typicality, and illness scripts component on case processing and disease probability estimates. *Memory and Cognition*, 24(3), 384–399.
- Custers, E., Boshuizen, H., & Schmidt, H. (1998). The role of illness scripts in the development of medical diagnostic expertise: Results from an interview study. *Cognition and Instruction*, 16(4), 367–398.
- DeGroot, A. (1965). *Thought and action in chess*. New York: Basic Books.
- Dixon, R.A., & Brown, R.A. (2012). Transfer of learning: Connecting concepts during problem solving. *Journal of Technology Education*, 24(1), 2–17.
- Elstein, A.S., Shulman, L.S., & Sprafka, S. (1978). *Medical problem solving*. Cambridge, MA: Harvard University Press.
- Eva, K., Neville, A., & Norman, G. (1998). Exploring the aetiology of content specificity: Factors influencing analogic transfer and problem solving. *Academic Medicine*, 73(10), S1–S5.
- Feltovitch, P., & Barrows, H. (1984). Issues of generality in medical problem solving. In H. Schmidt & M. De Volder (Eds.), *Tutorials in problem-based learning: A new direction in teaching the health professions*, (pp. 128–142). Assen, The Netherlands: Van Gorcum.

- Galloti, K.M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin*, 105(3), 331–351.
- Gick, M., & Holyoak, K.J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15(1), 1–38.
- Glaser, R. (1987). Learning theories and theories of knowledge. In E. de Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and Instruction* (Vol. 1, pp. 397–413). Oxford: Pergamon Press.
- Haskell, R.E. (2001). *Transfer of learning: Cognition, instruction and reasoning*. San Diego, CA: Academic Press.
- Herbert, D.M., & Burt, J.S. (2001). Memory awareness and schematization: Learning in the university context. *Applied Cognitive Psychology*, 15(6), 617–637.
- Herbert, D.M., & Burt, J.S. (2003). The effect of different review opportunities on schematisation of knowledge. *Learning and Instruction*, 13(1), 73–92.
- Herbert, D.M., & Burt, J.S. (2004). What do students remember? Episodic memory and the development of schematization. *Applied Cognitive Psychology*, 18(1), 77–88.
- Hmelo-Silver, C.E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16(3), 235–266.
- Hobus, P., Boshuizen, H., & Schmidt, H. (1990). *Expert-novice differences in the mental representations of patient prototypes*. Paper presented at the 2nd International Symposium on the Problem-based Learning, Yogyakarta, Indonesia.
- Hobus, P., Hofstra, M., Boshuizen, H., & Schmidt, H. (1988). De context van de klacht als diagnosticum [The context of the complaint as a diagnostic tool]. *Huisarts en Wetenschap*, 31, 261–264.
- Hobus, P., Schmidt, H., Boshuizen, H., & Patel, V. (1987). Contextual factors in the activation of the first diagnostic hypotheses: Expert-novice differences. *Medical Education*, 21(6), 471–476.
- Hofstra, M., Hobus, P., Boshuizen, H., & Schmidt, H. (1988). De invloed van ervaring op diagnostische prestaties van huisartsen [The influence of experience on diagnostic accuracy of family physicians]. *Huisarts en Wetenschap*, 31, 282–284.
- Johnson, S.D. (1995). Transfer of learning. *The Technology Teacher*, 54(7), 33–35.
- Johnson, S.D., Dixon, R., Daugherty, J., & Lawanto, O. (2011). General versus specific intellectual competencies: The question of learning transfer. In M. Barack & M. Hacker (Eds.), *Fostering human development through engineering and technology education* (pp. 55–74). Rotterdam, The Netherlands: Sense.
- Lesgold, A., Rubinson, H., Feltovitch, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Lawrence Erlbaum Associates.
- McGeoch, J. (1932). Forgetting and the law of disuse. *Psychological Review*, 39(4), 352–370.
- Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Norman, G. (2009). Teaching basic science to optimize transfer. *Medical Teacher*, 31(9), 807–811.
- Norman, G.R., & Schmidt, H.G. (1992). The psychological basis of problem-based learning: A review of the evidence. *Academic Medicine*, 67(9), 557–565.
- Regehr, G., & Norman, G.R. (1996). Issues in cognitive psychology: Implications for professional education. *Academic Medicine*, 71(9), 988–1001.
- Schmidt, H.G. (1983). Problem-based learning: Rationale and description. *Medical Education*, 17(1), 11–16.
- Schmidt, H., Norman, G., & Boshuizen, H. (1990). A cognitive perspective on medical expertise: Theory and implication. *Academic Medicine*, 65(10), 611–621.
- Smith, S. (1988). Environmental context-dependent memory. In G. Davies & D. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13–43). New York: John Wiley.
- Smith, S., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin and Review*, 8(2), 203–220.
- Spiro, R.J., Feltovitch, P.L., Jackson, M.L., & Coulton, R.L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domain. *Educational Technology*, 31(5), 24–33.

- Spiro, R.J., Vispoel, W.P., Schmitz, J.P., Samarapungavan, A., & Boerger, A.E. (1987). Knowledge acquisition for application: Cognitive flexibility and transfer in complex domains. In B.K. Britton & S.M. Glynn (Eds.), *Executive control processes in reading* (pp. 177–199). Hillsdale, NJ: Erlbaum.
- Sutton, M.J. (2003). Problem representation, understanding, and learning transfer: Implications for technology education research. *Journal of Industrial Teacher Education*, 40(4), 47–61.
- Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352–373.
- Vanbelle, S., Massart, V., Giet, D., & Albert, A. (2007). Test de concordance de script: Un nouveau mode d'établissement des scores limitant l'effet du hasard. *Pédagogie Médicale*, 8(2), 71–81.
- Van Gessel, E., Nendaz, M.R., Vermeulen, B., Junod, A., & Vu, N.V. (2003). Development of clinical reasoning from the basic sciences to the clerkships: A longitudinal assessment of medical students' needs and self-perception after a transitional learning unit. *Medical Education*, 37(11), 966–974.